

Poplar oxidoreductases involved in the oxidative stress response : a crystallographic snapshot towards the understanding of the catalytic mechanism

Cha San Koh

▶ **To cite this version:**

Cha San Koh. Poplar oxidoreductases involved in the oxidative stress response : a crystallographic snapshot towards the understanding of the catalytic mechanism. Other. Université Henri Poincaré - Nancy 1, 2008. English. NNT : 2008NAN10028 . tel-01748531

HAL Id: tel-01748531
https://hal.univ-lorraine.fr/tel-01748531

Submitted on 29 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

U.F.R. Sciences & Techniques de la Matière et des Procédés
Ecole doctorale Lorraine de Chimie et Physique Moléculaires (SESAMES)

FACULTE DES SCIENCES ET TECHNIQUES

Co-tutelle thesis
with the Universiti Sains Malaysia, Penang, Malaysia

Presented to obtain the title of

Doctor of the Université Henri Poincaré, Nancy 1

In Chemistry and Molecular Physico-Chemistry

by KOH Cha San

Poplar oxidoreductases involved in the oxidative stress response: A crystallographic snapshot towards the understanding of the catalytic mechanism

Public thesis defense on the 29th May 2008

with the presence of the commision board

Rapporteurs :
Mr Masami KUSUNOKI, Professor, University of Yamanashi, Japan
Mr David COBESSI, Chargé de Recherches au CNRS, HDR, ESBS Strasbourg, France

Examiners :
Madam Aida BAHARUDDIN, Lecturer, HDR, Universiti Sains Malaysia, Penang
Madam Catherine CORBIER, Professor, FST, UHP Nancy-Université
Mr Claude DIDIERJEAN, Lecturer, HDR, FST, UHP Nancy-Université
Mr Razip SAMIAN, Associate Professor, Universiti Sains Malaysia, Penang
Mr Shukri SULAIMAN, Associate Professor, Universiti Sains Malaysia, Penang

Laboratoire de Cristallographie et Modélisation des Matériaux Minéraux et Biologiques,
Groupe Biocristallographie (LCM3B, UMR 7036)

ACKNOWLEDGEMENTS

First, I am very grateful to the commission board members of my thesis who accept and willing to read my thesis. I would like to thank two of my main supervisors, Prof. Catherine Corbier from France, and Associate Prof. Dr Razip Samian from Malaysia, who spent a lot of time reading my thesis and gave me a lot of useful suggestions and criticisms, and also for being patient with me through all the times. I really appreciate the guidance and assistances that they have provided and very grateful them to accept me as their student under the UHP-USM cotutelle program. I also like to express my gratitude to Dr Claude Didierjean who shared with me a lot of his expertise and research insight. He quickly became for me the role model of a successful researcher in the field. I could not have imagined having a better mentor, and without his enthusiasm, his inspiration, his unselfish passion and his encouragement, I would never have finished.

The members of my supervisory committee provided with many useful suggestions and comments to me. This includes Prof. Omar Shawkataly, Associate Prof. Dr. Sheila Nathan and Dr. André Aubry. I want to thank Dr. Fréderique Favier and Dr. Guillermo Mulliert, who taught me the knowledge of protein X-ray crystallography and protein modeling, which they have provided me a much broader understanding of these subjects. Many thanks go to the director of LCM^{3}B Prof. Claude Lecomte who welcomed me in the laboratory and to the present and past LCM^{3}B members particularly Dr. Moutsé Ranaivoson, Sébastien Moniot, Dr. Hélène Dubourg, Dr. Fabien Pascale, Dr. Sébastien Lebègue, Sandrine Mathiot, Alexandre Bouché, Dorothee Liebschner, Mikael Elias etc. who offered countless aides, discussions and provided a friendly working environment in the laboratory.

Special thanks to the team of Prof. Jean-Pierre Jacquot, particularly Prof. Jacquot himself, Dr. Nicolas Rouhier, Dr. Eric Gelhaye, Dr. Nicolas Navrot and Dr. Alison Winger, who provided proteins supply generously and had given me some useful suggestions about this project from biochemical aspects.

The financial support of the Universiti Sains Malaysia (ASTS fellowship) and the French Government (financial aids from EGIDE, CUGN and CNRS) is gratefully acknowledged. I would like to acknowledge the FIP-team in ESRF, Grenoble, France and the DESY-team in EMBL-Hamburg Outstation, Germany, for providing access to synchrotorn beamlines.

I would also like to give special thanks to Nicolas who had given me uncountable useful advices and have comforted me when I faced some setbacks in my research work. Thank you for sharing your opinions with me and I am glad to have a best friend like you. I will remember you forever and will not forget the good time we have together. Special thanks also dedicated to Moutsé and Alison who encourage me to be strong and optimistic in life. I also want to thank you my friends in Nancy who have given me tremendous spiritual support throughout my study in France; they are Thu, Cao Lei, Na Du, Andrew, Hana, Ezatul, Syukri and Nazeer not forgetting my friends in Malaysia; they are Bee Ben, Kay Keat, Kue Peng, Hooi Ling, Yu Ching and many more. Because the list is too long, I will simply say thank you very much to all of my friends which, from my childhood until graduate school, have joined me in the discovery of what is life about and how to make the best out of it.

Finally, and most importantly, I give my distinguished appreciation to my family. They have always supported and encouraged me throughout my postgraduate research life. To them I dedicate this work.

TABLE OF CONTENTS

		Page
Acknowledgements		ii
Table of contents		iv
List of Tables		xii
List of Figures		xiv
List of Abbreviations		xvii
Abstract (Bahasa Malaysia)		xix
Abstract (English)		xx
Abstract (French)		xxi
CHAPTER 1 - INTRODUCTION		**1**
1.1	An overview of X-ray crystallography and protein crystallography	2
1.1.1	Nobel prize winners associated with crystallography	4
1.1.2	Integration and applications of crystallography in structural biology today	7
1.1.3	Understanding the basis of crystallography	10
1.1.4	Protein crystals	13
1.1.4.1	Preparation and handling of biological macromolecules for crystallization	13
1.1.4.2	Properties of protein crystals	15
1.1.4.3	Evidence that solution and crystal structures are similar	17
1.1.4.4	Growing protein crystals	18
1.1.4.4.1	The crystallization phase diagram	18
1.1.4.4.2	Factors influencing protein crystal growth	21
1.1.4.4.3	Crystallization techniques	23
1.1.4.4.3.1	Batch crystallization	23
1.1.4.4.3.2	Vapor diffusion	25
1.1.4.4.3.3	Dialysis method	25

1.1.4.4.3.4 Liquid-liquid diffusion / counter-diffusion crystallization 28
1.1.4.4.4 Strategies and approaches in growing crystals 30
1.1.4.4.4.1 Available screening matrices in the market today 32
1.1.4.4.4.2 Robotic apparatus 33
1.1.4.4.4.3 Some difficulties with screens and optimization of conditions for crystal growth 33
1.1.4.4.4.4 Seeding methods and other approaches to induce nucleation 35
1.1.4.4.4.5 Crystallization in gels 36
1.1.4.4.4.6 Crystallization in space 38
1.1.4.5 Judging crystal quality, mounting crystals and post-crystallization treatments 40
1.1.5 Diffraction data collection: basis of structure determination 43
1.1.5.1 Introduction 43
1.1.5.1.1 Importance of atomic resolution precision 43
1.1.5.1.2 Diffraction of a simple object 45
1.1.5.1.3 Arrays of simple objects: real and reciprocal lattices 46
1.1.5.1.4 Intensity of reflections 46
1.1.5.1.5 Arrays of complex objects 47
1.1.5.1.6 Three-dimensional arrays 47
1.1.5.1.7 Coordinate systems in crystallography 49
1.1.5.2 Geometric principles of diffraction 50
1.1.5.2.1 The generalized unit cell 50
1.1.5.2.2 Indices of the atomic planes in a crystal 51
1.1.5.2.3 Bragg’s law in reciprocal space 53

1.1.5.2.4 Number of measurable reflections 54
1.1.5.2.5 Unit-cell dimensions 56
1.1.5.2.6 Unit-cell symmetry 57
1.1.5.2.7 Friedel pairs 58
1.1.5.3 Collecting X-ray diffraction data 59
1.1.5.3.1 X-ray sources 59
1.1.5.3.2 Synchrotron radiation 60
1.1.5.3.3 Determining unit-cell dimensions 63
1.1.5.3.4 Symmetry and the strategy of collecting data 65
1.1.5.3.5 Scaling and post refinement of intensity data 67
1.1.6 From diffraction data to electron density 68
1.1.6.1 The Fourier transform: general features 69
1.1.6.2 From data to density 71
1.1.7 Obtaining phases 72
1.1.7.1 Introduction 72
1.1.7.2 Isomorphous replacement phasing 73
1.1.7.3 Anomalous scattering 74
1.1.7.4 Direct methods in protein crystallography 77
1.1.7.5 Molecular replacement: related proteins as phasing models 80
1.1.7.6 Iterative improvement of phases 82
1.1.8 Obtaining and judging the molecular model 82
1.1.8.1 Iterative improvement of maps and models 82
1.1.8.2 First maps 84
1.1.8.3 The model becomes molecular 84
1.1.8.4 Structure refinement 86
1.1.8.5 Convergence to a final model 87
1.1.8.6 Judging the quality and usefulness of the refined model 89
1.1.8.6.1 Structural parameters 89
1.1.8.6.2 Resolution and precision of atomic positions 90
1.1.8.6.3 Vibration and disorder 92
1.1.8.6.4 Disordered regions and unexplained density 93
1.1.8.6.5 Distortions due to crystal packing 94

1.1.8.6.6 Functional unit versus asymmetric unit 94
1.2 Oxidative stress in organisms 95
1.3 Overview of ROS in plants 96
1.3.1 ROS: friend or foe? 96
1.3.2 ROS cause oxidative damages 96
1.3.2.1 Production of ROS and scavenging of ROS in cells 96
1.3.2.2 Detoxification system of ROS 100
1.3.2.2.1 Non-enzymatic reduction system 100
1.3.2.2.2 Enzymatic reduction system 100
1.3.3 ROS signaling in plants: efficient secondary messengers 101
1.3.3.1 Modulation of gene expression 102
1.3.3.2 Influence signal transduction pathways 104
1.3.4 Thioredoxin fold 105
1.3.4.1 Overview on protein folding 105
1.3.4.1.1 The relationship between folding and amino acid sequence 106
1.3.4.1.2 Different types of protein folding 107
1.3.4.1.2.1 The thioredoxin fold 108
1.3.4.1.2.2 The Rossmann fold 108
1.3.4.1.2.3 The globin fold 110
1.3.4.1.2.4 TIM beta/alpha-barrel 110
1.3.4.1.2.5 Oligonucleotide-binding fold 111
1.3.4.1.2.6 The up-and-down β barrel 112
1.3.4.2 Definition of the canonical thioredoxin fold 113
1.3.4.2.1 Protein examples which contain Trx-fold 114
1.3.4.2.2 Sequence conservation 120
1.3.4.2.3 Disulfide bond formation 120
1.4 Protein of interest 121
1.4.1 Glutathione peroxidase (Gpx) 121
1.4.1.1 Biological theme: 121
1.4.1.1.1 General functions in living organisms 121
1.4.1.1.2 Gpx families in animals and plants 122

1.4.1.1.3 Plant Gpxs 123
1.4.1.1.4 Poplar Gpxs 124
1.4.1.2 Structural biology: 126
1.4.1.2.1 Gpx structures known to date 126
1.4.2 Thioredoxin (Trx) 126
1.4.2.1 Biological theme 126
1.4.2.1.1 General functions 126
1.4.2.1.2 Distinct families in all organisms 129
1.4.2.1.3 Plant thioredoxins 129
1.4.2.1.4 Poplar thioredoxin *h* 130
1.4.2.2 Structural biology 131
1.4.2.2.1 Thioredoxin structures known to date 131
1.4.3 Glutaredoxin (Grx) 132
1.4.3.1 Biological theme 132
1.4.3.1.1 General functions in living organisms 132
1.4.3.1.2 Plant Grx system 136
1.4.3.1.3 Poplar Grx 140
1.4.3.2 Structural biology 141
1.4.3.2.1 Grx structures known to date from all organisms 141
1.4.3.2.2 Background on known plant Grxs 141

CHAPTER 2 – RESEARCH OBJECTIVES 144
2.1 Overall research objectives 145
2.1.1 Gpx 147
2.1.2 Trx 148
2.1.3 Grx 150

CHAPTER 3 – METHODS AND MATERIALS 152
3.1 Crystallogenesis 154
3.2 Data collection and reduction 157
3.2.1 Beamline used 157

3.2.2 Crystal handling 162
3.2.3 Data collection 163
3.2.4 Data reduction 168
3.3 Structure solving 168
3.3.1 Phasing method using particular SAD approach: a case study for poplar thioredoxin-dependent peroxidase 5 168
3.3.2 Phasing method using conventional sulfur-SAD approach: a case study for poplar thioredoxin *h*4 172
3.3.3 Phasing method using molecular replacement: a case study for poplar glutaredoxin S12 174
3.4 Structure refinement and validation 175
3.4.1 Refinement using REFMAC5 176
3.4.2 Refinement using CNS suite 177
3.4.3 Structure validation 178
3.5 PDB deposition 178
3.6 Structure superimpositions 179
3.7 Structure modeling and molecular dynamic simulations 179

CHAPTER 4 – RESULTS AND DISCUSSION 182

Part I: Gpx

A. **RESULTS**
1. Specific objectives 183
2. Results obtained
2.1 Sequence alignment 184
2.2 Crystal description and structure refinement statistics 186
2.3 Quality of the model 188
2.4 Overall fold and secondary structure of PtGpx5 subunit 189
2.5 Oligomerization state of PtGpx5 192
2.6 Active-site of PtGpx5 194
2.6.1 Environment of the peroxidatic Cys in the reduced form

of the enzyme (rPtGpx5) 194
2.6.2 Environment of the peroxidatic Cys in the oxidized form of the enzyme (oPtGpx5) 197
2.7 Proposed model of the oPtGpx5-Trx *h*1 complex 198
2.8 Comparison to homologous structures 201
2.9 Cadmium-binding sites 203

B. DISCUSSION

1.1 Re-evaluation of PtGpx5 denomination 204
1.2 Oligomerization mode of PtGpx5 206
1.3 Comparison between reduced and oxidized form PtGpx5: redox-driven conformational changes 209
1.4 Active site differences in comparison to mammalian Gpxs 212
1.5 Possible interactions between PtGpx5 and thioredoxin *h*1 from poplar 214
1.6 Protective role of PtGpx5 against Cd^{2+} 216

Part II: Trx

A. **RESULTS**

1. Specific objectives 218
2. Results obtained
2.1 Sequence alignment 219
2.2 Crystal description and structure refinement statistics 221
2.3 Quality of models and overall description of structures 223
2.4 Structural comparison with other thioredoxins 226
2.5 Comparison with other Trxs 229

B. DISCUSSION

1.1 The uniqueness of PtTrx*h*4 231
1.2 PtTrx*h*4: formation of the classical C58-C61 disulphide bond 231
1.3 Cys4 of PtTrx*h*4 acts as the second resolving cysteine 233
1.4 PtTrx*h*4 fails to use the typical NTR reducing system 234

1.5 PtTrx*h*4 is regenerated by GSH/Grx reducing system 239

Part III: Grx

A. RESULTS

1. Specific objectives 243
2. Results obtained
   - 2.1 Sequence alignment 244
   - 2.2 Crystal description and structure refinement statistics 247
   - 2.3 Quality of models 247
   - 2.4 Overall structures 249
   - 2.5 Active site 252
   - 2.6 Glutathione binding site 254
   - 2.7 Structural evaluation of conserved residues in the PtGrxS12 and orthologs 255

B. DISCUSSION

1.1 PtGrxS12: an example of glutaredoxins subclass 1 of photosynthetic organisms 256
1.2 Environment of the active site of PtGrxS12 258
1.3 GSH binding footprint and comparison with other Grxs 261
1.4 Is there a role for the additional active-site cysteine? 264

CHAPTER 5 - CONCLUSION & PERSPECTIVES 267

5.1 Gpx 268
5.2 Trx 272
5.3 Grx 275

REFERENCES 278

PUBLICATIONS 324

LIST OF TABLES

		Page
Table 1.1	Nobel Prize laureates in scientific fields associated with crystallography.	6
Table 1.2	Important factors affecting macromolecular crystallization.	23
Table 1.3	Resolution and structural information.	44
Table 1.4	The seven crystal systems.	58
Table 1.5	Mechanisms for production and scavenging of ROS in plant cells.	97
Table 3.1	Summary of X-ray crystal structures of poplar oxidoreductases solved in this study.	154
Table 3.2	Crystallization conditions of obtaining crystals from poplar proteins of interests suitable for X-ray diffraction studies.	156
Table 3.3	Important specifications on macromolecular crystallography beamlines used for data collections in this study.	158
Table 3.4	Data collection and processing statistics.	165
Table 3.5	Solved protein crystal structures in this study and their corresponding PDB codes and references.	179
Table 4.1	Data refinement statistics for WT-PtGpx5$_{red}$ and WT-PtGpx5$_{oxd}$ crystals.	187
Table 4.2	Hydrogen bonding interactions at the dimerization interface.	194
Table 4.3	Coordination geometry of the active-site cadmium ion.	197
Table 4.4	Hydrogen bonding interactions between oPtGpx5 and TRX *h*1 in both complex 1 and complex 2 models.	200
Table 4.5	Data refinement statistics for WT PtTrx*h*4 and PtTrx*h*4 C61S crystals.	222
Table 4.6	Data refinement statistics for WT-PtGrxS12$_{oxd[GSH]}$ and WT-PtGrxS12$_{oxd[GSH+HED]}$ crystals.	248
Table 4.7	Secondary structure elements in PtGrxS12.	250

Table 4.8 Hydrogen bonding interactions at the PtGrxS12-GSH interface. 255

LIST OF FIGURES

		Page
Figure 1.1	Action of a simple lens.	10
Figure 1.2	Schematic illustration of a protein crystallization phase diagram (top panel) and various crystallization set-ups explained in terms of phase diagrams (bottom panel).	20
Figure 1.3	Protein crystallization techniques.	27
Figure 1.4	Counter-diffusion crystallization under restricted geometry.	29
Figure 1.5	Crystallographic data collection, showing reflections measured at one particular crystal orientation (solid, on the film) and those that could be measured at other orientations (hollow, within the sphere but not on the film).	48
Figure 1.6	This crystal has a 2-fold axis along *c*.	51
Figure 1.7	The intersection of three (234) planes with a unit cell.	52
Figure 1.8	The geometry of diffraction and its relationship to Bragg's law.	53
Figure 1.9	Limiting sphere.	55
Figure 1.10	The unit cell with edges, a, b, c and angles, α, β and γ.	57
Figure 1.11	Reflection spacings on the film are directly proportional to reciprocal-lattice spacings, and so they are inversely proportional to unit-cell dimensions.	63
Figure 1.12	The basic ROS cycle.	98
Figure 1.13	The extended ROS cycle.	99
Figure 1.14	A hypothetical model of the signaling pathway activated in plants in response to external application of oxidants.	105
Figure 1.15	Cartoon representation of selected protein folds: (a) the thioredoxin fold, (b) the Rossmann fold, (c) the globin fold, (d) TIM barrel, (e) the oligonucleotide-binding fold (OB fold), (f) the up-and-down β barrel.	109

Figure 1.16 Structure of Trx and typical architecture of Trx-like proteins. 116

Figure 1.17 The reaction scheme of Trx catalyzed protein disulfide reduction. 128

Figure 1.18 2-Cys and 1-Cys mechanisms of action of Grxs for reduction of dithiol groups. 133

Figure 1.19 Confirmed or proposed roles for plant glutaredoxins. 138

Figure 4.1 Multiple sequence alignment of 14 representative glutathione peroxidase proteins from mammals and plants (including the PtGpx5 as compared to the available secondary structure Elements). 185

Figure 4.2 (a) and (b) Topological diagrams and (c) and (d) cartoon representations showing the overall structures of the rPtGpx5 (left two panels) and the oPtGpx5 (right two panels). 191

Figure 4.3 Stereoscopic views of superimpositions between the Cα traces of the rPtGpx5 (red), the oPtGpx5 (blue) and a typical mammalian Gpx, the human plasma Gpx-3 (yellow). 192

Figure 4.4 Cartoon representation of dimerization interfaces of the rPtGpx5 (monomers A and B). 193

Figure 4.5 Stereoview details of the active site cleft of both (a) rPtGpx5 and (b) oPtGpx5 with final $3F_o$-$2F_c$ electron densities (1.2 σ level) covering chosen residues for clarity. 196

Figure 4.6 Cartoon representation of the oPtGpx5 structure with the molecular surface. 198

Figure 4.7 Molecular model of the oPtGpx5-Trx *h*1 complex. 200

Figure 4.8 Local surface charges of α-helix 2 of rPtGpx5 (d) and its corresponding helices of other mammalian Gpx structures ((a) the monomeric human Gpx-4, PDB code 2GS3; (b) the classical tetrameric human Gpx-3, (c) human plasma Gpx-3). 202

Figure 4.9 The 32 independent cadmium atoms in the rPtGpx5 crystal structure are displayed. 204

Figure 4.10 Cartoon representations of different oligomerization states of enzymes. 208

Figure 4.11 Cartoon representations of rPtGpx5 (left) and oPtGpx5 (right). 210

Figure 4.12 Cartoon representations explaining the driving force of the unwinding of α-helices 2. 212

Figure 4.13 Active sites of the reduced poplar GPX5 (rPtGpx5) and the reduced *Bos taurus* selenoenzyme glutathione peroxidase 1 (PDB code 1GP1). 213

Figure 4.14 Multiple alignment of poplar Trx*h*4 related sequences. 220

Figure 4.15 Cartoon representation showing overall structures of the WT PtTrx*h*4 (blue) and the PtTrx*h*4 C61S (orange). 223

Figure 4.16 Electrostatic potential surface charges of WT PtTrxh4 (top panel a and b) and PtTrx*h*4C61S mutant (bottom panel, c and d), viewing from both sides of the molecule (difference by 180°). 225

Figure 4.17 Superimposition of the active site cysteines and tryptophan from poplar WT PtTrx*h*4 (blue) (PDB code 3D21), poplar PtTrx*h*4C61S (orange) (PDB code 3D22), *C. reinhardtii* (green) (PDB code 1EP7), *T. brucei* (yellow) (PDB code 1R26) and spinach Trx*f* short form (pink) (PDB code 1F9M). 227

Figure 4.18 Activity of PtPrxQ in presence of DTT and PtTrx*h*4. 232

Figure 4.19 Activity of PtPrxQ in presence of PtGrx C4 and PtTrx*h*4. 234

Figure 4.20 Cartoon representation of the interaction interfaces of the Trx-TrxR from *E. coli* (PDB code 1F6M). 238

Figure 4.21 A proposed four-step mechanism for Grx-dependent PtTrx*h*4 catalysis is presented in a schematic pathway. 242

Figure 4.22 Multiple sequence alignment of 9 representative Grx proteins from mammals and plants (including the PtGrxS12) as compared to selected available secondary structure elements. 246

Figure 4.23 Cartoon representations of WT-PtGrxS12$_{oxd[GSH]}$ (left) and WT-PtGrxS12$_{oxd[GSH+HED]}$ (right) structures. 249

Figure 4.24 Topological diagram of PtGrxS12. 251

Figure 4.25 Cartoon representation of the PtGrxS12 active site. 253

Figure 4.26 Multiple sequence alignment of PtGrxS12 and its strict homologs. 256

Figure 4.27 Cartoon representations of different glutathione binding sites of glutaredoxins. 262

LIST OF ABBREVIATIONS

CCD	Charge-Coupled Device
COX	Cytochrome c oxidase
DHAR	Dehydroascorbate reductase
DTT	Dithiothreitol
FT	Fourier Transform
FTR	Ferredoxin/Thioredoxin Reductase
Gpx	Glutathione peroxidase
GR	GSH-reductase
Grx	Glutaredoxin
GSH	Glutathione
GST	Glutathione-*S*-transferase
HED	β-hydroxyethyl disulfide
IR	Infrared
MAD	Multi-wavelength anomalous diffraction
MAPK	Map Kinase
MAPKKK	Map Kinase Kinase Kinase
MD	Molecular Dynamics
MIR	Multiple Isomorphous Replacement
MPD	2-Methyl-2,4-Pentanediol
NMR	Nuclear Magnetic Resonance
NTR	NADPH/ Thioredoxin reductase
OB fold	Oligonucleotide-binding fold

PDB	Protein Data Bank
PDI	Protein Disulfide Isomerase
PEG	Polyethylene glycol
PICOT	Protein Kinase C-interacting cousin of thioredoxin homology domain
Prx	Peroxiredoxin
ROS	Reactive Oxygen Species
SAD	Single-wavelength anomalous diffraction
SIRAS	Single isomorphous replacement with anomalous scattering
S*n*B	Shake-and-Bake
TLS	Translation, Libration, Screw
Trx	Thioredoxin
TrxR	Thioredoxin reductase
VMD	Visual Molecular Dynamics
WT	Wild-type

OKSIDOREDUKTASE DARIPADA KAYU KERAS JENIS POPULUS (POPLAR) YANG TERLIBAT DALAM RESPONS TERHADAP TEKANAN OKSIDATIF: KAJIAN KRISTALOGRAFIK KE ARAH PEMAHAMAN MEKANISME PEMANGKINAN ENZIM

ABSTRAK

Penghasilan species oksigen reaktif (ROS) merupakan akibat daripada persekitaran sentiasa dalam keaadaan aerobik yang tidak boleh dielakkan dan ia boleh disingkirkan oleh beberapa jenis sistem enzim sebaik sahaja dihasilkan. Tiga jenis oksidoreduktase (Glutathion peroksidase, Gpx; thioredoksin, Trx; and glutaredoksin, Grx) daripada *Populus trichocarpa* × *deltoides* (kayu keras jenis poplar) telah dicirikan melalui pendekatan biokimia dan kristalografi X-ray. Gpxs terdiri daripada sekumpulan enzim yang mengawal tahap ROS dalam sel dan melindunginya daripada kerosakan oksidatif. Dalam kajian ini, saya telah mencirikan struktur kristal bagi Gpx5 poplar (PtGpx5) dalam bentuk terturun dan teroksida. Perbandingan antara kedua-dua struktur redoks menunjukkan bahawa perubahan konformasi yang drastik adalah satu keperluan untuk menghampirkan dua residu sistin yang berjauhan supaya membentuk ikatan disulfida intra-molekul. Trxs ialah sekumpulan enzim yang mengawalatur pelbagai protein melalui proses penurunan thiol-disulfida. Objektif kajian ini adalah untuk menerangkan proses pemangkinan sejenis isoform Trx baru poplar, PtTrx*h*4, memandangkan enzim ini hanya boleh diturunkan oleh Grx dan bukannya oleh sistem NADPH:thioredoksin reduktase yang biasanya digunakan oleh Trx. PtTrx*h*4 mengandungi tiga sistin; satu di bahagian terminal N protein (Cys4) dan dua lagi (Cys58 and Cys61) di tapak aktif tipikal Trx (WC_1GPC_2). Analisis dua struktur kristal PtTrx*h*4 dalam kajian ini, jenis liar dan mutan C61S, membolehkan kita mencadangkan satu mekanisme pemangkinan empat langkah yang juga disokong oleh kajian enzimatik. Grxs pula adalah protein redoks yang menggunakan elektron daripada GSH untuk pemangkinan proses penukaran thiol-disulfida. Di sini, saya mencirikan struktur PtGrxS12 yang berkompleks dengan GSH, di mana ia adalah struktur pertama dari kumpulan Grx tumbuhan jenis sub-kelas 1 yang memiliki tapak aktif atipikal $_{28}WCSYS_{32}$. Protein ini mempunyai sistin tambahan (Cys87) dan peranan residu tersebut masih tidak diketahui. Maklumat tentang tapak ikatan GSH juga dirumuskan dalam kajian ini. Semua struktur protein yang dicirikan di sini mendalami pengetahuan kita tentang pemangkinan redoks dalam sel umbuhan dan interaksi enzim-substrat.

POPLAR OXIDOREDUCTASES INVOLVED IN THE OXIDATIVE STRESS RESPONSE: A CRYSTALLOGRAPHIC SNAPSHOT TOWARDS THE UNDERSTANDING OF THE CATALYTIC MECHANISMS

ABSTRACT

The production of reactive oxygen species (ROS) is an unavoidable consequence of living in an aerobic environment and once produced, it can be removed by several different enzyme systems. Three oxidoreductases (glutathione peroxidase, Gpx; thioredoxin, Trx and glutaredoxin, Grx) from *Populus trichocarpa* × *deltoides* (poplar tree) were characterized using biochemistry and X-ray crystallography approaches. Gpxs are a group of enzymes that regulate the levels of ROS in cells, and protect them against oxidative damage. In this study, I have determined the crystal structures of the reduced and oxidized form of poplar Gpx5 (PtGpx5). Comparison of both redox structures indicates that a drastic conformational change is necessary to bring the two distant cysteine residues together to form an intramolecular disulfide bond. Trxs are a family of ubiquitous enzymes which regulate various protein partners through the thiol-disulfide(s) reduction. The aim of this study is thus to precisely describe the catalytic mechanism of a new isoform of Trx that has been characterized in poplar, PtTrx*h*4, since it has been demonstrated recently to be reduced by Grx but not by the typical NADPH:thioredoxin reductase reducing system. PtTrx*h*4 contains three cysteines; one localized in an N-terminal extension (Cys4) and two (Cys58 and Cys61) in the usual Trx active site (WC_1GPC_2). Analyses of two PtTrx*h*4 crystal structures solved in this study, wild-type and C61S mutant, allow us to propose a four-step disulfide cascade catalytic mechanism in accordance with enzymatic studies. Grxs are highly conserved redox-proteins that utilize electrons from GSH particularly to catalyze thiol-disulfide exchange reactions. Here, I present the structure of glutathionylated PtGrxS12, the first structure of plant Grx of subclass 1 with an atypical $_{28}WCSYS_{32}$ active site. This protein possesses an additional cysteine (Cys87) in which the role of this extra active-site cysteine remains obscure. Details of the GSH binding site are summarized in this study. Protein structures solved here shed lights to our understanding of the redox mechanism in plants and to the enzyme-substrate interactions.

ÉTUDE CRISTALLOGRAPHIQUE DES OXYDORÉDUCTASES IMPLIQUÉES DANS LA RÉPONSE AU STRESS OXYDATIF CHEZ LE PEUPLIER POUR UNE MEILLEURE COMPRÉHENSION DES MÉCANISMES CATALYTIQUES

RÉSUMÉ

La production des espèces réactives de l'oxygène (ROS) est une conséquence inévitable de la vie dans un milieu aérobie. Une fois produites, ces espèces peuvent être éliminées par différents systèmes enzymatiques. Les structure de trois oxydoréductases (la glutathion peroxydase (Gpx), la thiorédoxine (Trx) et la glutarédoxine (Grx)) de *Populus trichocarpa* × *deltoides* (le peuplier) ont été résolues par diffraction des rayons X. Les Gpxs forment un groupe d'enzymes qui régulent la concentration des ROS dans les cellules, et qui les protègent des effets d'un stress oxydant. Contrairement à leurs homologues d'origine animale, les Gpxs végétales ne dépendent pas du glutathion (GSH) mais des Trx pour leur fonctionnement. De plus, elles possèdent une cystéine dans leur site actif au lieu d'une sélénocystéine. Dans cette étude, j'ai résolu les structures cristallines des formes réduite et oxydée de la Gpx5 de peuplier (PtGpx5) et montré que des changements conformationnels drastiques sont nécessaires pour permettre la formation d'un pont disulfure intramoléculaire. Les Trxs constituent une famille de petites protéines ubiquitaires, régulant diverses protéines cibles par la réduction de leur pont disulfure. Par cette étude structurale, mon objectif était de comprendre précisément le mécanisme catalytique d'une nouvelle isoforme caractérisée chez le peuplier, la PtTrx*h*4, dont la capacité à accepter des électrons de la Grx a été récemment démontrée. A la différence des autres Trxs *h* caractérisées jusqu'ici, la PtTrx*h*4 ne peut pas être réduite par la voie typique qui implique la NADPH-Trx réductase. Cette PtTrx*h*4 contient trois cystéines, la première localisée dans une extension en position N-terminale (Cys4) et deux (Cys58 et Cys61) situées dans le site actif classique (WC_1GPC_2) de la Trx. Les résolutions des structures de l'enzyme sauvage et du mutant C4S sous forme oxydée m'ont permis de proposer un mécanisme catalytique en quatre étapes en accord avec les études enzymatiques. Les Grxs sont des protéines qui utilisent des électrons de GSH en particulier pour catalyser des réactions d'échange de thiol-disulfure. Ici, je présente la structure de la PtGrxS12 (en complexe avec le GSH), la première structure de la Grx végétale sous-classe 1 avec un motif du site actif atypique $_{28}WCSYS_{32}$. Cette protéine possède une autre cystéine (Cys87), dont le rôle reste obscur. Le site de fixation du GSH est détaillé ultérieurement dans ce manuscrit. Ces structures éclaircissent notre compréhension du mécanisme d'oxydoréduction chez les végétaux et nous permettent de mieux comprendre les interactions enzyme-substrat.

Introduction

CHAPTER 1

INTRODUCTION

1.1 An overview of X-ray crystallography and protein crystallography

X-ray crystallography is the science of determining the arrangement of atoms within a crystal from the manner in which a beam of X-rays is scattered from the electrons within the crystal. The method produces a three-dimensional picture of the electron density within the crystal, from which the mean atomic positions, their chemical bond, their disorders and sundry other information can be derived. By definition, a crystal is a solid in which a particular arrangement of atoms (its unit cell) is repeated indefinitely along three principal directions known as the basis (or lattice) vectors, which are not necessarily perpendicular. A wide variety of materials can form crystals; such as salts, metals, minerals, semiconductors, as well as various inorganic, organic and biological molecules, which has made X-ray crystallography fundamental to many scientific fields.

The oldest and most precise method of X-ray crystallography is single-crystal X-ray diffraction, in which a beam of X-rays is reflected from evenly spaced planes of a single crystal, producing a diffraction pattern of spots called reflections. Each reflection corresponds to one set of evenly spaced planes within the crystal. The density of electrons within the crystal is determined from the position and intensity of the various reflections observed as the crystal is gradually rotated in the X-ray beam; this density allows the atomic positions to be inferred. For single crystals of sufficient purity and regularity, X-ray diffraction data can determine the mean chemical bond lengths and angles to within a few thousandths of an Angstrom and to

within a few tenths of a degree, respectively. The data also allow the static and dynamic disorder in the atomic positions to be estimated, which is usually less than a few tenths of an Angstrom. At its most basic level, X-ray crystallography is useful in identifying known materials, characterizing new materials and in discerning materials that appear similar by other experiments. However, X-ray crystal structures have many other applications; for example, they can account for unusual electronic or elastic properties of a material, shed light on chemical interactions and processes, or serve as the basis for understanding enzymatic mechanisms and designing inhibitors of therapeutic roles with hope of preventing diseases. The term "X-ray crystallography" is also sometimes applied to methods that involve X-ray diffraction from polycrystalline materials, such as powders of small crystals studied by X-ray powder diffraction.

Early protein crystallographers, proceeding by analogy with studies of other crystalline substances, examined dried protein crystals and obtained no diffraction patterns. Thus X-ray diffraction did not appear to be a promising tool for analyzing proteins. In 1934, J. D. Bernal and Dorothy Crowfoot (later Hodgkin) measured diffraction from pepsin crystals still in the mother liquor. Bernal and Crowfoot recorded sharp diffraction patterns, with reflections out to distances in reciprocal space that inversely correspond in real space to the distances between atoms. The announcement of their success was the birth announcement of protein crystallography. Ever since, X-ray crystallography has been the most prolific technique for the structural analysis of proteins and protein complexes, and is still the “gold standard” in terms of accuracy and resolution (Russell et al. 2004). Each new structure provided a specific step forward, each a harder puzzle due to its size, the resolution needed or any number of other complicating factors. To list but a few,

GroEL-GroES (Xu et al. 1997), F_1-ATPase (Abrahams et al. 1994, Leslie & Walker 2000), the nucleosome (Davey et al. 2002, Schalch et al. 2005), a bacterial potassium ion channel (Doyle et al. 1998), photosystem I (Ben-Shem et al. 2003, Amunts et al. 2007), the ribosome (Ban et al. 2000), the plasma membrane proton pump (Pedersen et al. 2007) and a bacterial multidrug ABC transporter (Dawson & Locher 2006), certainly make the milestones seem closer as the protein structures proliferate. Nevertheless, the number of structures of macromolecular assemblies solved by X-ray crystallography is still quite small compared to that of the individual proteins and it will probably be many years before we have a complete repertoire of high-resolution structures for the hundreds of complexes in a typical cell. This discrepancy is due mainly to the difficult production of sufficient quantities of the sample and its crystallization (Russell et al. 2004).

1.1.1 Nobel prize winners associated with crystallography

The beauty and regularity of crystals impressed people such an extent that, in the past, crystals were regarded as products of nature with mysterious properties. Scientific investigation of crystals started in 1669, when Nicolaus Steno, a Dane working as a court physician in Tuscan, proposed that during crystal growth, the angles between the faces remained constant. For a given crystal form, individual crystals might differ in shape (for example, in the development of their faces), but they always have identical angles between the same faces.

Before the famous first X-ray crystallographic diffraction experiment by von Laue, Friedrich and Knipping in 1912, the internal regularity of a crystal was suggested but never proven. X-ray crystallography has dramatically changed this

situation. The idea that crystals could be used as a diffraction grating for X-rays arose in 1912 in a conversation between Paul Peter Ewald and Max von Laue in Munich. Ewald had proposed a resonator model of crystals for his thesis, but this model could not be validated using visible light, since the wavelength was much larger than the spacing between the resonators. Von Laue realized that electromagnetic radiation of a shorter wavelength was needed to observe such small spacings, and suggested that X-rays might have a wavelength comparable to the unit-cell spacing in crystals. Working with Friedrich and Knipping, they shined a beam of X-rays through a sphalerite crystal and recorded its diffraction (a large number of well-defined spots arranged in a pattern of intersecting circles around the spot produced by the central beam) on a photographic plate. Von Laue developed a law that connects the scattering angles and the size and orientation of the unit-cell spacings in the crystal, for which he was awarded the Nobel Prize in Physics in 1914. Table 1.1 summarized the Nobel laureates with their contributions in the domain of crystallography.

Crystal structures of proteins began to be solved in the late 1950's, beginning with the structure of sperm whale myoglobin (Kendrew et al. 1960) and hemoglobin (Perutz et al. 1960) by Max Perutz and Sir John Cowdery Kendrew, for which they were awarded the Nobel Prize in Chemistry in 1962. Since then, the progress of protein crystallography has followed an evolutionary rather than a revolutionary path. Many important advances have been achieved on the way, but most of the methods used by Perutz are still valid today and still useful, albeit often in modified versions (Dauter 2005). Since that success, over 50000 X-ray crystal structures of proteins, nucleic acids and other biological molecules have been determined (according to PDB statistics in June 2008).

Table 1.1 Nobel Prize laureates in scientific fields associated with crystallography.

Year	Domain	Nobel Laureate(s)	Findings / Scientific contributions
1901	Physics	W.C.Röntgen	Discovery of X-rays
1914	Physics	M Von Laue	Diffraction of X-rays by crystals
1915	Physics	W.H.Bragg & W.L.Bragg	Use of X-rays to determine crystal structure
1929	Physics	L-V de Broglie	The wave nature of the electron
1937	Physics	C.J.Davisson & G.Thompson	Diffraction of electrons by crystals
1946	Chemistry	J.B.Sumner	Discovery of enzymes that can be crystallized
1954	Chemistry	L.C.Pauling	Research of the nature of the chemical bond and its application to the elucidation of the structure of complex substances
1962	Physiology or Medicine	F.Crick, J.Watson & M.Wilkins	Discovery of the helical structure of DNA Books about the double helix
1962	Chemistry	J.C.Kendrew & M.Perutz	Studies of the structures of globular proteins
1964	Chemistry	D.Hodgkin	Determination of many biochemical substances structures including Vitamin B12
1976	Chemistry	W.N.Lipscomb	Elucidation of the structure of boranes
1982	Chemistry	A.Klug	Development of crystallographic electron microscopy and discovery of the structure of biologically important nucleic acid-protein complexes
1982	Physics	K.G.Wilson	Conceptualizing the theory of critical phenomena in connection with phase transitions
1985	Chemistry	H.Hauptmanm & J.Karle	Development of direct methods for the determination of crystal structures
1988	Chemistry	J.Deisenhofer, R.Huber & H.Michel	Determination of the three-dimensional structure of a photosynthetic reaction centre
1991	Physics	P-G de Gennes	Development of the methods of discovering order in simple systems that can be applied to polymers and liquid crystals
1992	Physics	G.Charpak	Discovery of the multi wire proportional chamber
1994	Physics	C.Shull & N.Brockhouse	Neutron diffraction
1996	Chemistry	R.Curl, H.Kroto & R.Smalley	Discovery of the fullerene form of carbon
1997	Chemistry	P.D.Boyer, J.E.Walker & J.C.Skou	Elucidation of the enzymatic mechanism underlying the synthesis of adenosine triphosphate (ATP) and discovery of an ion-transporting enzyme
2002	Chemistry	J.B.Fenn, K.Tanaka & K.Wüthrich	Development of methods for identification and structure analyses of biological macromolecules
2002	Physics	R.Davis Jr., M.Koshiba & R.Giacconi	Pioneering contributions to astrophysics (detection of cosmic neutrinos and the discovery of cosmic X-ray sources)
2003	Chemistry	P.Agre & R.Mackinnon	Discoveries concerning (water and ion) channels in cell membranes
2006	Chemistry	R.D.Kornberg	Studies of the molecular basis of eukaryotic transcription

1.1.2 Integration and applications of crystallography in structural biology today

The development and application of technologies in structural biology (availability of complete genome sequences, automation of cDNA cloning, automated protein expression screens, affinity tags and parallel purification strategies, high-throughput crystallization procedures, usage of third-generation synchrotron, automated sample changing and the use of high-energy sources and CCD detector technology) do accelerate the transition from gene to structure (Dry et al. 2000). X-ray crystallography has benefited from several technological advances in recent years that make the genome-wide protein structure determination a practical reality. Indeed, it is now used routinely by scientists to determine how a pharmaceutical interacts with its protein target and what changes might be advisable to improve it (Scapin 2006). The number of protein structures deposited in the Protein Data Bank now (June 2008) is almost 48 000, with the vast majority (85 %) determined using X-ray crystallographic methods. Among the available protein crystal structures, only 34 % (~14000 structures) of them actually represent unique proteins (structures with similar sequences at 90 % identity were excluded). Thousands of studies describing such structures have been published in the scientific literature, and 14 Nobel prizes in chemistry or medicine have been awarded to protein crystallographers (see Table 1.1).

There is a wide spectrum of experimental and computational methods for the identification and structural characterization of macromolecular complexes. These methods need to be combined into hybrid approaches to achieve greater accuracy, coverage, resolution and efficiency than any of the individual methods. New methods must be capable of generating possible alternative models consistent with

information such as stoichiometry, interaction data, homology to known structures, docking results and low-resolution images. There is a need to describe the structures and dynamics of stable and transient complexes (Russell et al. 2004). As motions of macromolecules throughout their conformational landscapes generate biological activity, these motions can be investigated by NMR, neutron scattering, molecular dynamics, UV to IR spectroscopy, cryo-electron microscopy, and visible-ray or X-ray techniques (Bourgeois & Royant 2005). X-ray crystallography provides insight into protein dynamics *via* the analysis of mean-square atomic displacements and TLS (translation, libration, screw) parameters (Chaudhry et al. 2004), or when a series of static structures assigned to various states along a reaction pathway is morphed into movies (Echols et al. 2003).

Structure-based drug design has contributed to the discovery of a number of drugs and late-stage clinical candidates. It is now common for a series of ligand-protein structures to be available in discovery projects. The availability of X-ray derived structural information on protein-ligand complexes is increasing, and this is a useful tool in lead optimization (Davis et al. 2003). Indeed, knowledge of 3-dimensional structure based function properties of a drug target is very essential for a successful *in silico* designing of drugs (Kishan 2007). However, some difficulties during structure determination processes and lack of knowledge of conformational freedom associated with available protein structures often hurdle the structure based drug designs.

Structural genomics may contribute to a comprehension and efficient structural description of complexes in an additional way. Although structural genomics currently focus on single proteins or their domains, it could be expanded to the sampling of domain-domain interactions (Aloy et al. 2003, Apic et al. 2001, Sali

2003). Indeed, recent developments in the experimental and computational techniques have allowed structural biology to shift its focus from structures of individual proteins to the structures of large assemblies (Sali et al. 2003, Baumeister 2002, Sali & Kuriyan 1999). In contrast to structure determination of individual proteins, structural characterization of macromolecular assemblies usually poses a more difficult challenge. Nevertheless, a comprehensive structural description of large complexes generally requires the use of several experimental methods in order to maximize efficiency, completeness, accuracy and resolution (Sali et al. 2003). Coupled with X-ray crystallography, hybrid approaches have been successfully elucidating models of the actin-myosin complex (Pellecchia et al. 1999), the yeast ribosome (Spahn et al. 2001, Beckmann et al. 2001) and 20S proteasome (Groll & Huber 2005, Borissenko & Groll 2007), pre-mRNA splicing complex SF3b (Golas et al. 2003) and complex virus structures (Zhou et al. 2001, Baker et al. 2003).

Structural biology is a great unifying discipline of biology. Thus, structural characterization of many protein complexes may be the way to bridge the gaps between genome sequencing, functional genomics, proteomics and system biology. The goal seems daunting, but the prize will be commensurate with the effort invested, given the importance of molecular machines and functional networks in biology and medicine (Russell et al. 2004). It seems that in the near future, macromolecular crystallography will move along two parallel paths, system-oriented and discovery-oriented (Stevens 2004). The system-oriented approach corresponds to the traditional, slower, but more focused way of solving crystal structures of macromolecules and the discovery-oriented approach is, in other words, the structural genomics, speedier and wider track.

1.1.3 Understanding the basis of crystallography

When we see an object, light rays bounce off (are diffused by) the object and enter the eye through the lens, which reconstructs an image of the object and focuses it on the retina. In a simple microscope, an illuminated object is placed just beyond one focal point of a lens, which is called the objective lens. The lens collects lights diffused from the object and reconstructs an image beyond the focal point on the opposite side of the lens, as shown in Figure 1.1.

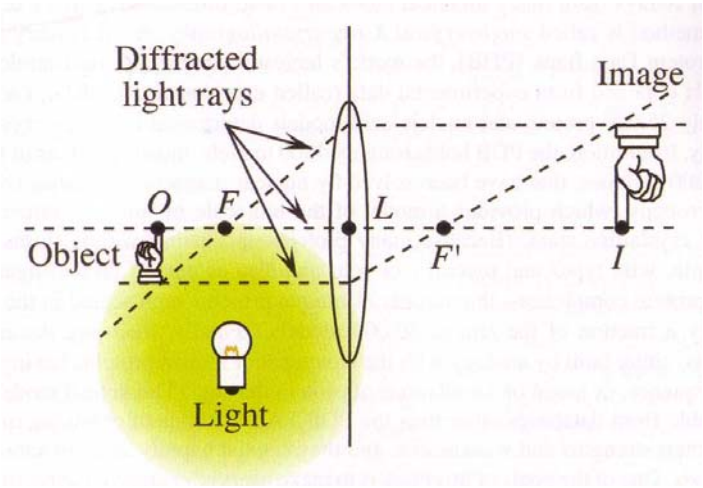


Figure 1.1 Action of a simple lens. Rays parallel to the lens axis strike the lens and are refracted into paths passing through a focus (*F* or *F'*). Rays passing through a focus strike the lens and are refracted into paths parallel to the lens axis. As a result, the lens produces an image at I of an object at *O* such that (*OF*) (*IF'*) = (*FL*) (*F'L*) (Rhodes 2006).

For a simple lens, the relationship of object position to image position in Figure 1.1 is (*OF*) (*IF'*) = (*FL*) (*F'L*). Because the distances, *FL* and *F'L* are constant (but not necessarily equal) for a fixed lens, the distance *OF* is inversely proportional to the distance *IF'*. Placing the object just beyond the focal point *F* results in a magnified image produced at a considerable distance from *F'* on the other side of the

lens, which is convenient for viewing. In a compound microscope, the most common type, an additional lens, the eyepiece, is added to magnify the image produced by the objective lens.

In order for the object to diffract light and thus be visible under magnification, the wavelength (λ) of the light must be, roughly speaking, no larger than the object. Visible light, which is electromagnetic radiation with wavelengths of 400 - 700 nm (nm = 10^{-9} m), cannot produce an image of individual atoms in protein molecules, in which bonded atoms are only about 0.15 nm or 1.5 angstroms (Å = 10^{-10} m) apart. Electromagnetic radiation of this wavelength falls into the X-ray range, so X-rays are diffused by even the smallest molecules.

Even though individual atoms diffract X-rays, it is still not possible to produce a focused image of a single molecule, for two reasons. First, X-rays cannot be focused by lenses. Crystallographers sidestep this problem by measuring the directions and strengths (intensities) of the diffracted X-rays and then using a computer to simulate an image-reconstructing lens. In short, the computer acts as the lens, computing the image of the object and then displaying it on a screen. Second, a single molecule is a very weak scatterer of X-rays. Most of the X-rays will pass through a single molecule without being diffracted, so the diffracted beams are too weak to be detected. Analyzing diffraction from crystals, rather than individual molecules, solves this problem. A crystal of a protein contains many ordered molecules in identical orientations, so each molecule diffracts identically, and the diffracted beams for all molecules augment each other to produce strong, detectable X-ray beams.

In brief, determining the structure of a protein by X-ray crystallography entails growing high-quality and well-ordered crystals (that will diffract X-rays

strongly) of the purified protein, measuring the directions and intensities of X-ray beams diffracted from the crystals (regular and repeating array of many identical molecules), and using a computer to simulate the effects of an objective lens and thus produce an image of the crystal's contents. Finally, the crystallographer must interpret that image, which entails displaying it by computer graphics and building a molecular model that is consistent with the image.

Electron densities can be probed by the scattering of X-rays. This is usually done by creating crystals, containing many trillions of molecules arranged in a regular three-dimensional pattern. When this pattern is illuminated with X-rays, diffraction of the X-rays occurs. The diffraction pattern encodes the information about the electron density distribution in the crystal. The diffracted rays are themselves, like the incident X-rays, electromagnetic waves with amplitude and phase and form a three-dimensional pattern, known as the reciprocal lattice. They are an indirect image of the electron density and are related to it by a Fourier transform. To recover the electron density from this diffraction pattern by inverse Fourier transform, both amplitudes and phases are needed. Experimentally usually only the amplitudes can be measured. The phase information remains hidden within these amplitude data and has to be recovered by computational techniques, often requiring additional measurements. This situation is known as the "Phase Problem in Crystallography". It is indeed a problem, because the phases influence the electron density to a much larger extent than the amplitudes do.

1.1.4 Protein crystals

1.1.4.1 Preparation and handling of biological macromolecules for crystallization

In the crystallization of biological macromolecules the quality (purity and homogeneity) and quantity of the required material is important. Difficulties in crystal growth sometimes may be linked to the nature or source of the biological material. Better crystallization conditions or diffracting crystal habits could be found by switching from one organism to another. This is because variability in sequences between heterologous species may lead to different conformations, and consequently to different crystallization behaviors. In practice, proteins isolated from eukaryotes are frequently more difficult to crystallize than their prokaryotic counterparts. Often their degree of structural complexity is higher. They can possess additional domains that may contribute to less compact and/or more flexible structures. Post-translational modifications (addition of other biochemical functional groups, changing the chemical nature of an amino acid or making structural changes like the formation of disulfide bridges) are often responsible for structural or conformational microheterogeneity. Purification, stabilization, storage and handling of macromolecules are therefore essential steps prior to crystallization. The “freshness” of the starting material and physiological state of cells is very important. Some proteins from unicellular organisms have been isolated in their native state only when cells are in exponential or pre-stationary growth phase (for an example, see Lorber & DeLucas 1990).

Special cares are needed in handling pure macromolecules in order to ensure that they are not damaged or lost before or during the crystallization trials and for

their reproducibility. To achieve this, several guidelines should be taken into account (see Ducruix & Giegé 1999 for details).

1. Diluted protein solutions are concentrated by ultrafiltration in devices using pressure (optimizing stir rate to prevent adsorption on to membrane surfaces and damage by shearing) or centrifugal force (remove aggregates formed as a result of a decrease in pH, of oxidation, or of an increase in salt or protein concentration), by dialysis against hygroscopic compounds (e.g. PEG or gel-filtration matrices), or by precipitation (addition of ammonium sulphate).
2. High concentrations of denaturing agents should be avoided because they inactivate or unfold proteins. In contrast, mild non-ionic detergents (e.g. octyl glucoside, heptyl thioglucoside and alkyl thiomaltosides are very useful in membrane protein biochemistry) help to solubilize proteins.
3. For better reproducibility, the pH of buffers after mixing of all ingredients should be adjusted since it may change after dilution or in the presence of other compounds.
4. Freeze-thawing of macromolecules should never be repeated to avoid denaturation. Therefore, experimentation on aliquots to limit repeated handling of stock solutions is highly recommended.
5. Any undesired molecules that might hinder crystallization should be removed by dialysis (e.g. glycerol, excess ligands), ultrafiltration, or size-exclusion chromatography.
6. It is also worthwhile to prepare macromolecules with or without their ligands (e.g. coenzyme, metal ions) or try additives (e.g. ions, reducing agents, chelators) because one or the other form may be more able to crystallize.

1.1.4.2 Properties of protein crystals

Whereas inorganic crystals can often be grown to dimensions of several centimeters or larger, it is frequently impossible to grow protein crystals as large as 1 mm in their shortest dimension. In addition, larger crystals are often twinned (two or more crystals grown into each other at different orientations) or otherwise imperfect and not usable. Roughly speaking, protein crystallography requires a crystal of at least 0.2 mm in its shortest dimension, although modern methods of data collection can sometimes succeed with smaller crystals, and modern software can sometimes decipher data from twinned crystals.

Inorganic crystals derive their structural integrity from the electrostatic attraction of fully charged ions. On the other hand, protein crystals are held together by weaker forces, primarily hydrogen bonds between hydrated protein surfaces. In other words, proteins in the crystal stick to each other primarily by hydrogen bonds through intervening water molecules. Protein crystals are thus much more fragile than inorganic crystals; gentle pressure with a needle is enough to crush the hardiest protein crystal. Growing, handling, and mounting crystals for analysis thus require very gentle techniques. If possible, protein crystals are often harvested, examined, and mounted for crystallography within their mother liquor, the solution in which they formed. This is because protein crystals are not only susceptible to chemical but also physical degradation, therefore they need to be maintained and stabilized in an equilibrated liquid environment as how they were formed and yet crystals will not be resolved (some sort of encapsulating effect).

Real macroscopic crystals are actually mosaics of many submicroscopic arrays in rough alignment with each other. The result of mosaicity is that an X-ray

reflection actually emerges from the crystal as a narrow cone rather than a perfectly linear beam. Thus the reflection must be measured over a very small range of angles, rather than at a single, well-defined angle. In protein crystals, composed as they are of relatively flexible molecules held together by weak forces, this mosaicity is more pronounced than in crystals of rigid organic or inorganic molecules, and the reflections from protein crystals therefore suffer greater mosaic spread than do those from more ordered crystals. The crystal mosaicity is a direct indicator of the physical perfection of the macromolecular crystal and it provides a simple measurement of crystal quality independent of many experimental parameters. It has been used to characterize successfully the improvement seen in some microgravity samples, for example, a reduction in the reflection mosaic spread providing a corresponding increase in the signal-to-noise ratio of the reflection (Snell et al. 1995, Ng et al. 1997).

Careful analysis of electron-density maps usually reveals many ordered water molecules on the surface of the protein in the structure. Additional disordered water is presumed to occupy regions of low density between the ordered particles. Ordered water molecules refer to water molecules that occupy the same site on every protein molecule in every unit cell (or a high percentage of them) and thus show up clearly in electron-density maps. Disordered water molecules refer to bulk water molecules that occupy the spaces between protein molecules which are in different arrangements in each unit cell, and thus show up only as uniform regions of low electron density. The quantity of water varies among proteins and even among different crystal forms of the same protein. The number of detectable ordered water molecules averages about one per amino-acid residue in the protein. Both the ordered and disordered water are essential to crystal integrity and maintaining the activity of

protein molecules in crystalline form (Frey 1994, Timasheff 1995), so drying destroys the crystal structure. For this reason, protein crystals are subjected to X-ray analysis in a very humid atmosphere or in a solution that will not dissolve them, such as the original mother liquor or a protective harvest buffer (with a higher concentration of precipitant or supplemented with cryoprotective agents such as PEG 400, PEG 600, glycerol or MPD) (Heras & Martin 2005).

NMR analysis of protein structure suggests that the ordered water molecules seen by X-ray diffraction on protein surfaces have very short residence times in solution. Thus most of these molecules may be of little importance to an understanding of protein function. However, ordered water molecules are of great importance to the crystallographer. As the structure determination progresses, ordered water molecules become visible in the electron-density map. Assignment of water molecules to isolated areas of electron density (small regions of disconnected density) improves the overall accuracy of the model, and improvements in accuracy in one area of the model give accompanying improvements in all other regions.

1.1.4.3 Evidence that solution and crystal structures are similar

Well-ordered crystals are difficult to grow because globular protein molecules are large, spherical, or ellipsoidal objects with irregular surfaces, and it is impossible to pack them into a crystal without forming large holes or channels between the individual molecules. These channels, which usually occupy more than half the volume of the crystal, are filled with disordered solvent molecules. The protein molecules are in contact with each other at only a few small regions, and even in these regions many interactions are indirect, through one or several layers of solvent

molecules. This is one reason why structures of proteins determined by X-ray crystallography are the same as those for the proteins in solution.

In a few cases, the structure of a protein has been obtained from more than one type of crystals (in different space groups). The resulting subunit models were identical, suggesting that the molecular structure was not altered by crystallization. One of the recent examples is the crystal structure of the Type IIP restriction endonuclease MspI bound to DNA (Xu et al. 2005) containing its cognate recognition sequence that has been determined in both monoclinic and orthorhombic space groups. These two independent crystal forms present an identical structure of a novel monomer-DNA complex, suggesting that this is not merely a crystallographic artifact.

1.1.4.4 Growing protein crystals

1.1.4.4.1 The crystallization phase diagram

The crystallization process can be illustrated by a phase diagram, which indicates which state (liquid, crystalline or amorphous solid [precipitate]) is stable under a variety of crystallization parameters. It provides a mean of quantifying the influence of certain parameters, such as the concentration of protein, precipitant(s), additive(s) and so on, on the production of crystals. Thus, phase diagrams form the basis of the design of crystal optimization experiments (Ataka 1993, McPherson 1999, Ducruix & Giegé 1999).

Figure 1.2 illustrates a typical crystallization phase diagram, consisting of four zones representing different degrees of supersaturation: a zone of high

supersaturation, where the protein will precipitate; a zone of moderate supersaturation, where spontaneous nucleation will take place; the metastable zone (just below the nucleation zone) of lower supersaturation, where crystals are stable and may grow but no further nucleation will take place (the conditions in this region are the best for the growth of well-ordered crystals); and a zone of undersaturation, where the protein is fully dissolved and will never crystallize (Chayen 2004, Ducruix & Giege 1999).

To obtain good structural data, crystals need to be single and have dimensions of at least 100 μm^3, preferably much larger. In an ideal experiment, once nuclei have formed, the concentration of protein in the solute will drop, thereby leading the system into the metastable zone, where few single crystals will grow (Figure 1.2). More often than not, either no crystals form at all or excess nucleation occurs, whereby numerous clusters of tiny crystals are formed instead of a few sizeable ones. An additional frustrating problem is the formation of large single crystals that do not diffract (Chayen 2004).

By examining the crystallization phase diagram and solubility properties of the sample, an understanding may be gained to optimize the process of crystal growth. The aim is to devise methods that will enable the experimenter to manipulate the phase diagram and actively control the crystallization environment in order to lead to crystal growth in the direction that will produce the desired results (Chayen 2005). This can be achieved in various ways (see section below), for example, control of the nucleation stage; dynamic separation of the nucleation and growth phases; and influencing the kinetics of the crystallization process.

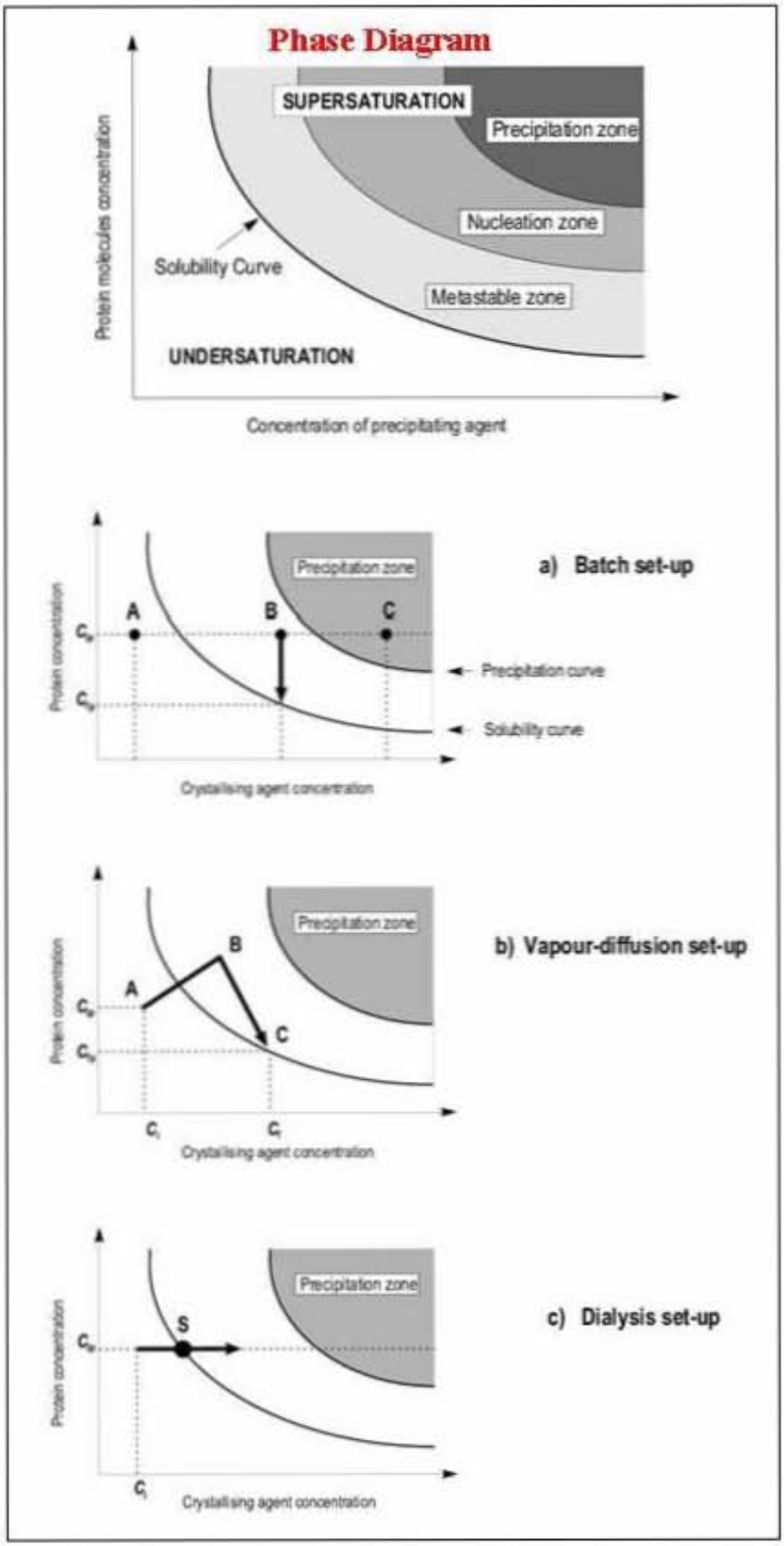

Phase Diagram
SUPERSATURATION
Precipitation zone
Nucleation zone
Metastable zone
Solubility Curve
UNDERSATURATION
Protein molecules concentration
Concentration of precipitating agent
Precipitation zone
A
B
C
Precipitation curve
Solubility curve
Protein concentration
Crystallising agent concentration
a) Batch set-up
Precipitation zone
A
B
C
Protein concentration
Crystallising agent concentration
b) Vapour-diffusion set-up
Precipitation zone
S
Protein concentration
Crystallising agent concentration
c) Dialysis set-up

Figure 1.2 Schematic illustration of a protein crystallization phase diagram (top panel) and various crystallization set-ups explained in terms of phase diagrams (bottom panel). The adjustable parameter can be precipitant or additive concentration, pH, temperature and so on. The three major crystallization methods are represented, showing their different routes to the nucleation and metastable zones, assuming the adjustable parameter is crystallizing agent concentration. The black circles represent the starting conditions. The solubility is defined as the concentration of the protein in the solute that is in equilibrium with crystals. The supersolubility curve is defined as the line separating conditions where spontaneous nucleation (or phase separation, precipitation) occurs from conditions where the crystallization solution remains clear if left undisturbed. In the bottom panel is the schematic representation of solubility phase diagram and correlation between protein and crystallizing agent concentrations in (a) batch, (b) vapour-diffusion and (c) dialysis crystallization experiments. C_{ip} and C_i are the initial concentrations of protein and crystallizing agent respectively, C_{fp} and C_f are their final concentrations (Figure adapted from Ducruix & Giegé 1999).

1.1.4.4.2 Factors influencing protein crystal growth

The two most important keys to success of a crystallographic project are purity and quantity of the macromolecule under study. Impure samples usually will not make suitable crystals, and even for proteins of the highest purity, repeated trials will be necessary before good crystals result.

Many variables influence the formation of protein crystals and each protein is unique. It is not possible to foresee the conditions that can cause or promote crystallization of a protein. The various parameters that affect crystallization are not independent of each other and their interrelation may be complicated and difficult to distinguish (McPherson 1999). There is no universal guideline to crystallize macromolecules successfully. The only way to do this is to identify the important components and refine each of them distinctively, based on knowledge on protein properties and its behavior observed from purification experiences. Again, these

components vary with proteins. In general, protein purity and concentration, precipitant type and concentration, buffer type and pH, and temperature are the most important factors for protein crystallization. They are normally being considered first when determine its effect on crystallization. More subtle ones like cleanliness, vibration and ultrasound (Edwards & Palmer 1990), convection, source and age of the protein and the presence of ligands are also among parameters that should also be taken into considerations. Clearly, the problem of developing a reliable source of crystals entails controlling and testing a large number of parameters. Indirectly, the difficulty and importance of obtaining good crystals has prompted the invention of crystallization robots that can be programmed to set up many trials under systematically varied conditions. However, when varying the more conventional parameters fails to produce good crystals, the crystallographer may take more drastic measures, e.g. mutagenesis. Sometimes limited digestion of the protein by a proteolytic enzyme removes a disordered surface loop, resulting in a more rigid, hydrophilic, or compact molecule that forms better crystals. A related measure is adding a ligand, such as a cofactor, that is known to bind tightly to the protein. The protein-ligand complex may be more likely to crystallize than the free protein, either because the complex is more rigid than the free protein or because the cofactor induces a conformational change that makes the protein more amenable to crystallizing. In a nutshell, Table 1.2 summarizes the factors (physical, chemical and biochemical factors) which affect the crystallization of macromolecules.

Table 1.2 Important factors affecting macromolecular crystallization (McPherson 1990; McPherson 1999).

Physical	Chemical	Biochemical
Temperature / temperature fluctuation	Buffer pH	Purity of macromolecule
Vibration / sound / mechanical perturbation	Precipitant type	Substrate / coenzyme / ligand / inhibitor / effectors
Time / rate of growth	Precipitant concentration	Inherent symmetric of the macromolecule
Equilibrium rate	Macromolecule concentration	Biochemical modification
Dielectric constant of medium	Ionic strength	Genetic / post-transitional modification
Viscosity of medium	Additive / specific ions	Isoelectric point
Pressure	Metal ions	Macromolecule stability
Gravity	Detergent / surfactant	Aggregation state of macromolecule
Homogeneity of macromolecule	Degree of supersaturation	Storage time of macromolecule
Electric / magnetic fields	Reducing / oxidizing environment	Source of macromolecule / history of sample
Volume of crystallization sample drop	Present of amorphous substances / impurities	Proteolysis / hydrolysis
Methodology / approach of crystallization	Cross-linker	Microorganism contamination

1.1.4.4.3 Crystallization techniques

1.1.4.4.3.1 Batch crystallization

The batch technique is attractive because of its inherent simplicity and reproducibility. It requires nothing more than the combination of two or more solutions (precipitating reagents and protein solution) into one, and a period of time until spontaneous nucleation commences, suddenly bringing the solution to a state of higher supersaturation. With luck, crystals grow gradually from the supersaturated solution without further processing. An automated system for microbatch

crystallization has been designed by Chayen et al. (1990, 1992). In the micro-batch technique (Figure 1.3a), smaller volumes, as little as 0.5 μl, can be used. But usually people tend to grow protein crystals in 2 to 3 μl drop, containing the protein and the precipitant, which is dispensed into the well of a “Terazaki-type” microtiter plate, covered with paraffin oil. The oil acts as a sealant to prevent evaporation. It does not interfere with the common precipitants, but it does interfere with organic compounds that dissolve in the oil (Chayen 1997, Chayen 1998). During the incubation period, the concentration of a precipitant agent remains constant since evaporation is limited and, therefore, the volume of the drop remains the same during the experiment. On the other hand, the concentration of the protein changes on formation of either crystals or amorphous precipitant. If the concentration of precipitant agent is chosen in such a way that the solution is in an undersaturated state, crystallization will never occur.

The main disadvantage of this method could be that the equilibration occurs very rapidly, thus affecting the rate of crystal growth (increases nucleation rate) and consequently decreases the size and the quality of crystals obtained. The manipulation of the crystals from the drop covered by oil could be difficult too. However, since the use of very small volumes (up to 50 nanoliters by using robotic system) of protein solution can be made, the micro-batch technique is quite useful as an initial screening method. Although the evaporation of water from the drop covered by oil is negligible, it does occur, and therefore the 'life-time' of micro-batch trials is usually about 2 to 3 weeks (see http://www-cryst.bioc.cam.ac.uk/ for details).

1.1.4.4.3.2 Vapor diffusion

Another widely used crystallization technique is vapor diffusion, in which the protein / precipitant solution is allowed to equilibrate in a closed container with a larger aqueous reservoir whose precipitant concentration is optimal for producing crystals. One of many examples of this technique is the hanging-drop method. In this method, drops are prepared on a siliconized microscope glass cover slip by mixing 0.5 - 10 μl of protein solution with the same volume of precipitant solution. The slip is placed upside down over a depression in a tray; the depression is partly filled with the required precipitant solution (~500 μl). The chamber is sealed by applying oil or grease to the circumference of the depression before the cover slip is put into place (Figure 1.3b). If the protein solution has a low surface tension, it tends to spread out over the cover slip in the hanging drop method. In such cases, the sitting drop method is preferable. A third approach in which the protein sample droplet is simultaneously in contact with both an upper and lower surface, called a sandwich drop, is sometimes used, but generally for special purposes such as making optical measurements.

1.1.4.4.3.3 Dialysis method

As with the other methods for achieving protein crystallization, many variations of dialysis techniques exist. Dialysis has the advantage that liquid-liquid diffusion through a semipermeable membrane permits exposure of a protein solution to a continuum of potential crystal-producing conditions without physically perturbing the mother liquor. Diffusion of molecules through the membrane is slow and can

even be controlled by the pore size of the membrane (McPherson 1999). Another advantage of dialysis is that the precipitating solution can be easily changed. For moderate amounts of protein solution (> 0.1 ml), dialysis tubes can be used (Figure 1.3c). The dialysis membrane is attached to a tube by means of a rubber ring. The membrane should be rinsed extensively with water before use or, preferably, boiled in water for about 10 min. For microliter amounts of protein solution, one can use either a thick-walled microcapillary (Zeppezauer method, see Zeppezauer et al. 1968) or a plexiglass “button” covered with a dialysis membrane (Figure 1.3c). These buttons, which can be reused indefinitely, are now in wide use and have proven quite successful despite its disadvantage that the protein crystal in the button cannot be observed with a polarizing microscope. Another microdialysis procedure has also been introduced although the method is rarely used by protein crystallographers. Five microliters of protein solution is injected into a capillary, which is covered by a dialysis membrane. The membrane can be fastened with a piece of tubing. The protein solution is spun down in a simple centrifuge and the capillary closed with modeling clay. The capillary is then placed in a microtube containing the dialysis solution.

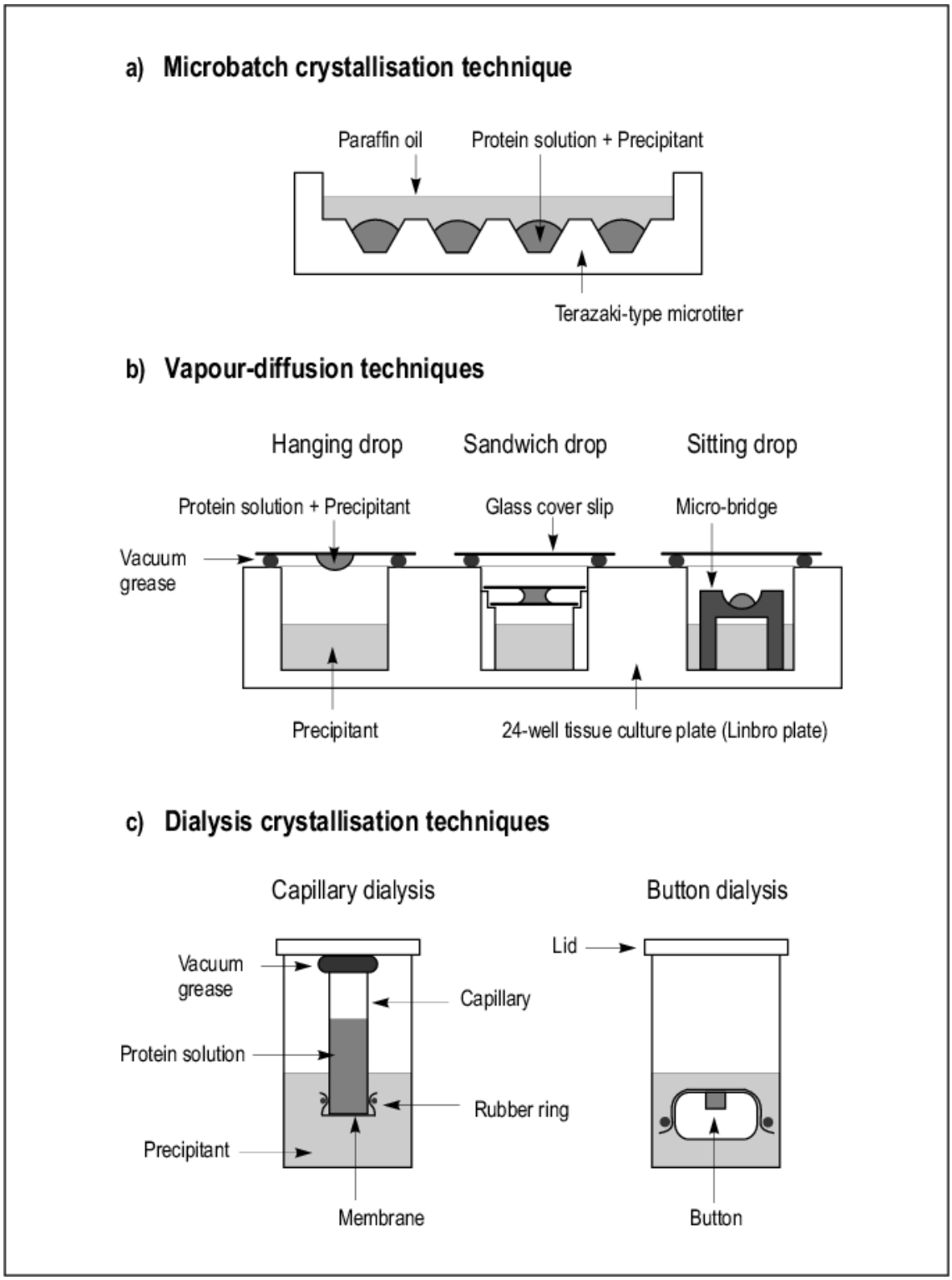


Figure 1.3 Protein crystallization techniques. Schematic representations of (a) microbatch, (b) vapour-diffusion and (c) dialysis crystallization techniques widely used in growing protein crystals. This figure is taken from the website of the Crystallography and Bioinformatics Group of University of Cambridge (http://www-cryst.bioc.cam.ac.uk/).

1.1.4.4.3.4 Liquid-liquid diffusion / counter-diffusion crystallization

The principles of counter-diffusion used for crystal growth have been well known for over a century (for review see Ng et al. 2003, García-Ruiz et al. 2002). Its usages for protein crystallization in a capillary had first been proposed by García-Ruiz (1991). In this method, the protein solution and the solution containing the precipitant are layered on top of each other in a small-bore capillary (restricted geometry configuration); a melting-point capillary can conveniently be used (Figure 1.4). The lower layer is the solution with higher density (e.g., a concentrated ammonium sulfate or PEG solution). If an organic solvent such as MPD is used as precipitant, it forms the upper layer. For a 1:1 mixture, the concentration of the precipitant should be two times its desired final concentration. The two solutions (up to 50 nanoliters by using the robotic sytem, 100 trials in one run) are introduced into the capillary with a syringe needle, beginning with the lower one. Spinning in a simple swing-out centrifuge removes air bubbles. The upper layer is added and a sharp boundary is formed between the two layers. They gradually diffuse into each other. As the two liquids contact each other, a gradient of supersaturation is formed over time along the length of the capillary thus allowing optimal conditions to be found for nucleation and crystal growth (all in one experiment). The counter-diffusion crystallization method in capillary has been effectively used to produce protein crystals in the presence of cryogenic reagents and strong scattering atoms for *in situ* data collection for *ab initio* structure determination (Gavira et al. 2002). This technique is, therefore, very effective for screening, optimizing and preparing protein crystals for direct crystallography without ever having to handle the crystal. Contrary to conventional crystallization methods, optimal supersaturation conditions leading to protein

precipitation, nucleation and crystal growth can be achieved concurrently in the capillary chamber. The structures of a number of proteins (~9 protein structures were well-characterized) have been determined using crystals obtained and prepared for direct data collection and phasing (Pusey et al. 2005).

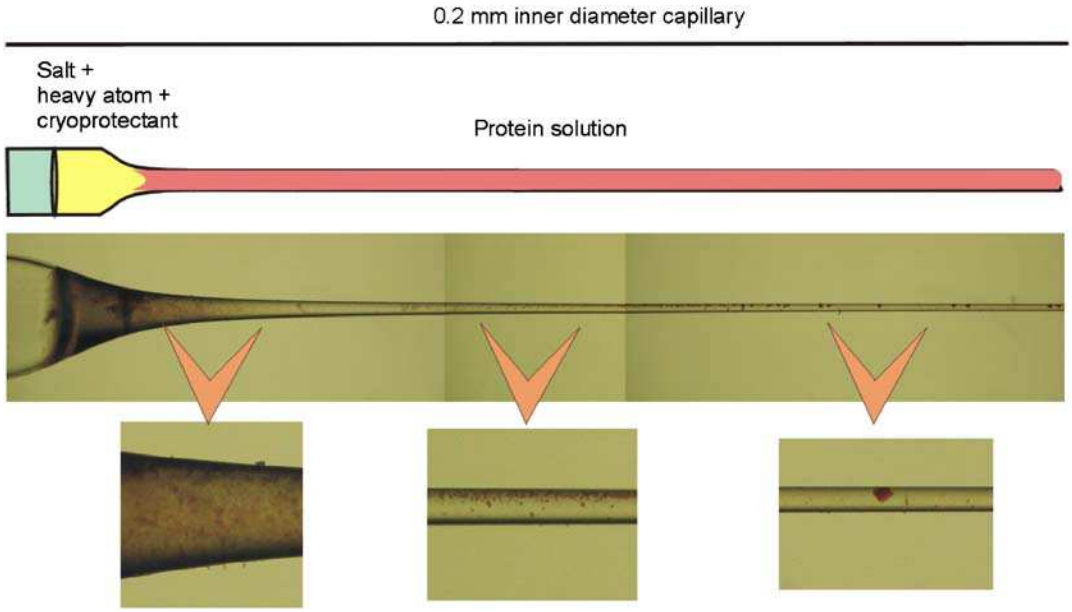


Figure 1.4 Counter-diffusion crystallization under restricted geometry. Protein crystallization takes place in an X-ray quartz capillary. Under this configuration the protein solution (about 5 µl) is loaded along the length of the capillary (pink) against an equal volume of solution containing a precipitating agent, anomalous scattering atom and cryopreservative (yellow). The ends of the capillary are sealed with nail varnish and clay. During the equilibration process, the precipitating and protein solution will diffuse into each other such that the salt solutions will initially mix into the protein layer, followed by the slower diffusion of the cryoprotectant reagent. Consequently, a self-forming supersaturation gradient will be created along the length of the capillary where initially the highest supersaturation will be at the protein-precipitant solution interface. As a function of time, a supersaturation wave will move across the capillary tube giving rise to amorphous precipitation (left end of the tube) and eventually separate crystals growing across the capillary (middle and right end of the tube). There will be a position along the length of the capillary where a crystal of optimal quality will be produced. (Pusey et al. 2005)

1.1.4.4.4 Strategies and approaches in growing crystals

Crystallization of a protein is a multiparameter problem in which the parameters are varied in the search for optimal crystallization conditions. The most common parameters that are changed include protein concentration, the nature and concentration of the precipitant, pH, and temperature. Specific additives that affect the crystallization can also be added in low concentration (Ducruix & Giegé 1999, McPherson 1999).

There are two strategies for identifying initial crystallization conditions. The first is to systematically define and sample the most important variables that determine whether crystals nucleate, called the systematic method (Kingston et al. 1994), and the second is to apply in a more arbitrary way various screening procedures and sets of crystallization conditions that have proven successful in the past, called the screening method (Carter 1997, Kundrot 1997). At the start of crystallization of a new protein, one is faced with the choice of where to begin the trial and error process. In the former method, the search for crystallization conditions was initially based upon biochemical properties learned during the purification and general characterization of that macromolecule (McPherson 1982). Although this method has been the accepted approach for many years, the extraordinary success of screening formulations in recent years has made them currently the most popular choice, at least for the initial phase. Even when they are successful for the initial identification of crystallization conditions, however, a systematic approach is generally required for optimization.

Although screening procedures are attractive for their convenience, speed, and simplicity, they have drawbacks. For example, they are discontinuous in many

parameters and can miss useful conditions lying between screening points. The use of these kits also locks the experimenter to a relatively narrow set of historical conditions. It has been suggested that totally random screening approaches to identify crystallization conditions will work equally well, if not better (Segelke 2001). Interestingly, large number of screens is now appearing at increasing rates but an investigator may be bewildered as to which screen he should invest with his precious material. The most serious shortcoming of the screens however, is that they offer rather little opportunity to integrate the investigator's knowledge of the macromolecule's physical and chemical behaviors into the search for crystals. Systematic approaches, on the other hand, are more satisfying in a sense because they utilize the investigator's knowledge and expertise, and they embody the traditional scientific philosophy. They are also more complete and continuous, and even when they fail; they yield additional, valuable information about the macromolecule's solubility behavior. Their major disadvantages are that they are slower, more tedious, and, most important; they consume a large amount of valuable material. These drawbacks are, fortunately, compensated by the usage of robotic systems.

Despite their shortcomings mentioned above, the most expedient approach to identifying initial conditions for crystallization is first to use established screens, and if they fail, to utilize more rational and systematic means. This provides the most efficient, and at the same time, the most economical route to success. Nevertheless, after some number of crystallization trials one should look at modifications to the protein that may increase its probability of crystallization, rather than blindly carrying out more trials (Segelke 2001, Pusey et al. 2005).

1.1.4.4.4.1 Available screening matrices in the market today

Because a prediction of the best crystallization conditions is not possible, a huge number of crystallization experiments (different conditions) are required to screen all the parameters for the few combinations that might give high-quality crystals suitable for x-ray diffraction analysis in a reasonable time. The screening technique is called sparse matrix sampling. Specifically, crystallization conditions that produced diffraction quality crystals were compiled from the literature and the subset which sampled the widest range of buffers and precipitants was identified (Jancarik & Kim 1991). Since its initial release, this screen has been used by hundreds of crystallographers to identify initial crystallization leads.

One of the first large-scale efforts to accumulate and mine crystallization data was initiated over 15 years ago and led to the creation of the Biological Macromolecule Crystallization Database (BMCD; http://xpdb.nist.gov:8060/BMCD4/), an extensive archive of the experimental crystallization details of published protein structures (Gilliland et al. 2002). Crystallization screens can be homemade or obtained commercially. Suppliers of screens and other useful gadgets for crystallization are, for instance; Hampton Research (http://www.hamptonresearch.com), Molecular Dimensions Ltd. (www.moleculardimensions.com), Jena Bioscience GmbH (http://www.jenabioscience.com), Emerald BioSystems (http://www.emeraldbiosystems.com) and Qiagen (http://www1.qiagen.com/Products/ByApplication/Protein/Crystallization/).

1.1.4.4.4.2 Robotic apparatus

Manual methods are slow and error-prone. The time pressure exerted by high-throughput structural genomics requires speed and efficiency in scanning the enormous number of crystallization conditions. The solution is by automation of the process. Recently crystallization robots have been built to automate and speed up the laborious manual operations of reproducibly setting up large numbers of crystallization experiments.

With latest robotic system, once a target protein has been chosen, a few hundreds of different variations of growth solution can be coarse screened all at once, each in its own tiny reservoir. The smaller the drop size, the quicker the crystallization. Therefore, most of the robots developers are trying to get crystallization with a few nanoliters of protein instead of microliters – a reduction by a factor of a thousand so that crystals can be obtained within a few hours to a few days, and yet are experimentally manipulatable. Good robotic systems have to come with a high-resolution CCD camera to detect the growth of even tiny crystals in the transparent reservoirs and original software to control the numerous delicate steps in the operation.

1.1.4.4.4.3 Some difficulties with screens and optimization of conditions for crystal growth

Although attractive for their simplicity, convenience and their notable success rates, commercially available screening kits are still not ideal. Actually, the idea of screen is to identify a starting point for optimization. In order to produce a “hit” of the modest success of a microcrystalline shower or a few crystals among the precipitate

with maximum probability, the precipitate concentrations of the screening solutions are usually very high by most standards. Many trials precipitate at once, or within hours, and many possibly productive combinations of buffer, precipitant and additive are obscured. This shortcoming can be overcome by diluting the reagents in the screening kit by half or even one-third, thereby making two or three kits from one. The protein is then screened three times with the same reagents, but at three different concentrations. This tests not only different starting and final concentrations, but also different kinetic pathways to equilibrium.

A second problem with screens is that although they may be very good at identifying starting conditions for proteins that are less challenging, they don't address well the more problematic cases such as those that may require detergents (e.g. membrane proteins) or biological ligands. This can also be remedied to some extent by recycling trials. For example, a series of samples, hanging- or sitting-drop vapor diffusion trials, may be deployed and monitored for several weeks. If at the end of that time no hits are obtained, the samples can be opened and an aliquot (2-5 μl) of a detergent solution (0.2 – 1 % β-octylglucoside) can be added to every sample and the screen reinitiated. Clearly, this is not as good as beginning with detergent present, but it may produce the needed hit. Similarly, instead of a detergent, a cofactor or substrate may be deemed more likely to be effective, and that may be added instead. Multiple temperatures may also be explored by moving the samples between 4°C, 20°C, and 37°C over the course of 1 or 2 months.

Another problem to watch for is that the screens tend to include too many samples containing Ca^{2+}, Mg^{2+}, other transition metal ions, or Li^{2+}. All these can produce salt crystals, hence false positives, with irritating frequency. However, if you use a particular screen over a long period with many macromolecules, you learn to

recognize, and even expect, these red herrings. It may not always be clear exactly what a particular screening sample contains, as the descriptions of the solutions may be ambiguous. This produces difficulties when one then tries to optimize a hit, in that the solution yielding the success is not what one believes. Consultation with the manufacturer of the screening kit is the best course should suspicions arise.

In a nutshell, there may be just too many screens available today. It is almost fair to say that by the time all of the tempting and available screens have been tried, one might as well have taken a slow, steady, systematic approach from the start. At least, by choosing that more demanding option a good collection of useful solubility data would be in hand.

1.1.4.4.4.4 Seeding methods and other approaches to induce nucleation

It is known that the protein should be in the metastable phase for crystal growth, but that higher levels of saturation are needed for nucleation. In many crystallization experiments the required saturation levels are not reached, so that nucleation does not occur. Nucleation is the rate-limiting step in protein crystallization. There are two types of nucleation: homogeneous (in the bulk of the solution) and heterogeneous, which can be caused by a host of factors, such as solid material in the crystallization medium, charged surfaces and even the walls of the crystallization vessel. Some of these factors can be controlled if the experiment is designed with care (Chayen 2004). Introducing heterogeneous substrates may in some cases create a higher local concentration of macromolecules and thus lower the energy barrier for nucleation and thereby facilitate crystal growth (Thakur et al. 2007). Currently, the mechanism of heterogeneous protein nucleation remains poorly understood even though a search

for a “universal nucleant” has been ongoing for two decades. To date, the following lines of research have been pursued.

(i) McPherson introduced the idea of controlling nucleation by using mineral substrates as epitaxic nucleants for protein crystallization (McPherson & Shlichta 1987). His initiative has been pursued for more than 15 years by employing a variety of substrates, but so far none of them have been generally adopted.

(ii) Subsequently, the idea of using lipid layers and protein monolayers of two-dimensional crystals was introduced, which also improved the crystallization of three-dimensional protein crystals (Hemming et al. 1995). More recently, Fermani et al. (2001) demonstrated that substrates containing ionizable surface groups can also enhance the crystallization of certain proteins.

(iii) The idea of using natural seedlings (whiskers, seeds, fibers etc.) to generate nucleation in crystallization experiments is another approach that has been successfully applied in protein crystallography (D’Arcy et al. 2003).

(iv) The development of lithography techniques allowed the fabrication of a variety of silicon substrates with different surface characteristics: terraces, steps and even pores of a few nanometers in size that can match a crystal lattice. Sanjoh et al. (2001) and Chayen et al. (2001) have explored this field intensively and showed that in general structured surfaces appear to be more efficient than unstructured.

1.1.4.4.4.5 Crystallization in gels

Gel media reduce convection and crystal sedimentation, and somehow, they can mimic microgravity (García-Ruíz et al. 2001a). Thus, it provides an attractive growth environment (homogenous and a diffusive regime molecular transport) for

optimizing biological crystals (García-Ruíz et al. 2001b). In addition, crystals can no longer move or sedimentate, since nuclei are trapped in the pores of the matrix. They remain in the mother liquor and are able to grow in three dimensions (García-Ruíz et al. 2001b, Lorber & Giegé 2001, Sauter et al. 2002). Such a more stable growth environment has been shown to prevent twinning of crystals (Sica et al. 1994) and to reduce the incorporation of impurities into the crystals (Chernov et al. 2001). Furthermore, it may favor higher internal order and generate crystals with lower mosaicity and a more homogeneous lattice. This has actually been observed for lysozyme (Vidal et al. 1999) but also for larger macromolecules crystallized in gel (Lorber et al. 1999a). On a practical point of view, the gel matrix provides a mechanical protection of the crystals. The samples can be transported without the risk of undergoing severe damages. The soft gel matrix allows seeding and soaking experiments, and can easily be dissected for crystal mounting. The addition of agarose gels is fully compatible with classical crystallization techniques, for example, in vapor diffusion (Bernard et al. 1994, Sica et al. 1994, Thiessen 1994), microdialysis (Thissen 1994, Lorber et al. 1999b) or microbatch (Moreno et al. 2002). A simpler technique is to set up gelled trials in microbatch, which has enabled, for the first time, the automatic dispensing of 0.3 - 2.0 µl gel drops in high-throughput mode (Chayen & Saridakis 2002). The gel is dispensed while still a liquid and, after a given time, polymerization occurs. Comparison of crystals grown in gels with those grown in standard trials shows a significant improvement in crystal order (Chayen 2004). Furthermore gellified media can be used to perform counter-diffusion experiments either by gel-acupuncture (García-Ruíz and Moreno 1994, Moreno et al. 1999) or by using gel inside capillaries (Lopez-Jaramillo et al. 2001).

Regardless of their favorable properties, the application of gels to crystallization was under-used because of complicated procedures and the requirement for large quantities (>10 μl) of sample (Chayen 2004). The past few years have seen major improvements in the use of gels by miniaturization and automation of gel trials. Crystals can be grown in small volumes of gel inside capillaries, thus combining the advantages of growth in gels with those of the liquid-liquid free interface diffusion method (García-Ruíz et al. 2002).

1.1.4.4.4.6 Crystallization in space

With the recent advances in powerful X-ray sources and supercomputers for solving the complex three-dimensional structures, the limiting step in this technology (after of course the protein overexpression and purification steps) is the ability to crystallize specific proteins with sufficient size and crystalline quality to obtain the high resolution X-ray diffraction patterns needed to solve their structure. For reasons that are not totally clear, crystals often will grow larger and with much better internal order, if made in microgravity. Furthermore, in microgravity, crystallization methods such as liquid-liquid diffusion offer opportunities for many interesting configurations simply not possible on earth, for example new crystal structures, crystals with less thermal motion and crystals with resolution improvement of 0.3 – 1.0 Å (McPherson 1999). For reasons that are not totally clear, crystals often will grow larger and with much better internal order, if made in microgravity. Improved X-ray diffraction data from these space-grown crystals could be collected and has allowed researchers to refine structures in much more detail than ever before, and has allowed the structure of some proteins to be determined for the first time

(http://science.nasa.gov/msl1/pcg_why.htm). Indeed, microgravity research has enriched the field through new concepts, new benchmarks, new methods, new apparatus, novel protein crystal structures and better resolved existing protein structures. This method is however of course difficult i.e. not having easy access and the dangers of space flight to the astronauts flying on the flights are well known (Helliwell 2005).

The key concepts that attracted crystal growers, protein or non-protein (*e.g.* semiconductors), to microgravity research is that in essentially weightless conditions density difference fluid flows can be eliminated and sedimentation of the growing crystals is also eliminated (McPherson 1996). Thus, defect and flaws in the crystals can be reduced, even eliminated. Defects in crystals however can affect not only crystal properties but also the diffraction data from which protein structures are elucidated. Overall a principal effect witnessed with microgravity grown crystals has been a sharpened mosaicity. This has increased the peak height of reflections above background. A sharper mosaicity makes sense because it in essence means that the crystal as an ordered array has a longer spatial coherence length. This can be imagined to be readily linked to a more orderly growth environment from a non-convective fluid flow situation in microgravity (Helliwell 2005).

Much interest is also currently focused on gravity as an obstacle to the formation of well-ordered crystals, and a number of costly protein crystallization experiments have been performed in space, under microgravity conditions. However, it still remains to be shown that these conditions have more than a marginal effect on protein crystallization.

1.1.4.5 Judging crystal quality, mounting crystals and post-crystallization treatments

Using equipment typical of today's crystallography laboratories, researchers can collect preliminary diffraction data quickly and decide whether to obtain a full data set. However, a brief inspection of crystals under a low-power light microscope can also provide some insight into quality and can help the crystallographer pick out the most promising crystals. Desirable visible characteristics of crystals include optical clarity, smooth faces, and sharp edges. Broken or twinned crystals sometimes exhibit dark cleavage planes within an otherwise clear interior. Depending on the lattice type and the direction of viewing relative to unit-cell axes, some crystals strongly rotate plane-polarized light. This property is easily observed by examining the crystal between two polarizers, one fixed and one rotatable, under a microscope. Upon rotation of the movable polarizer, a good-quality crystal will usually brighten and darken sharply. Once crystals have been chosen, the major concern is to verify if they are made of macromolecules or merely small molecules (e.g. salts) that present in the crystallization solutions. The X-ray diffraction pattern is of course the most convincing proof that crystals are of macromolecular nature. In this way, information about crystal quality (resolution and space group) can be obtained spontaneously (Ducruix & Giegé 1999).

Crystal mounting for crystallography can be carried out using capillary tubes, fine glass capillary tubes, thin glass spatulas and thin loop films (Rodgers 1997). The classical method of mounting crystals is to transfer them into a fine glass capillary along with a droplet of the mother liquor (Bernal & Crowfoot 1934). The capillary is then sealed at both ends and mounted onto a goniometer head, a device that allows control of the crystal's orientation in the X-ray beam. The droplet of mother liquor

keeps the crystal hydrated. While very widely used, this method of mounting has several disadvantages, for example, deforms fragile crystals and substantially increased mosaic spread (Przybylska 1988), showed non-uniform X-ray absorption effects (Teng 1990), and hindering the shock freezing method (Haas & Rossman 1970). As macromolecular crystallographers move to solve more and more structures each year, the limitations of current technology for handling and mounting protein crystals are becoming evident. In the late 1980s, the versatile loop mounting method had been introduced (Teng 1990). Ever since, it has had few rivals and surprising longevity. Even so, new methods for the emerging high-throughput environment can improve upon loops in many ways. For example, the free-standing loop mounting technique has become the prevalent procedure for crystal mounting at cryogenic temperatures (Rodgers 1997). Originally the loops were made from copper or tungsten wire with 1 − 2 mm diameters and 25 – 75 μm thickness, but now, they are made from fine nylon fibers with 0.05 – 1.0 mm diameters and 10 & 20 μm thickness. These types of cryo-loops show minimal background diffraction due to the optically clear environment and the loops are tiny enough to be convenient for fast freezing. The major advantages of using this approach to lift up, transfer, flash cool the crystal are the production of lower X-ray scatter and lower X-ray absorption. Uniform results can be obtained compared to the conventional mounting techniques such as capillary mounting (Teng, 1990). The loop itself serves as a platform to support the crystal in place and to keep it away from foreign material which has detrimental effects to the crystal. It is a good idea to select a loop size which is just wide enough to keep the crystal from dropping off. For instant data collection, the crystal is picked up and flash-cooled immediately. The cold stream is first deflected

from the nitrogen gas nozzle, the crystal is placed on the goniometer, and then the obstruction of the flow is removed immediately.

Despite technical and methodological advances in the field, including the development of highly focused X-rays, high-throughput crystallization techniques and automated X-ray data analysis, crystal growth still remains an empirical and tedious process and a common occurrence is the production of well formed crystals that are not suitable for diffraction studies. Loose packing of molecules and large solvent volume are common problems that result in low-resolution and poor-quality diffraction (Heras & Martin 2005). What are the options if crystals form but do not diffract well? Strategies that have been used in the past include improving protein purification steps, searching for new crystallization conditions to identify a new crystal form, crystallizing the protein of interest from a different organism, crystallizing a different form of the protein by using proteases that produce smaller protein fragments, addition of additives into the crystallization drops, generating new constructs encoding a truncated form of the protein or mutating surface amino acids to enhance protein crystallization (Longenecker et al. 2001). These methods all have one thing in common: they each give up on the crystals that have already been grown. However, sophisticated post crystallization treatments (for example, post-crystallization soaking, cross-linking, crystal annealing and controlled dehydration), far beyond simple cryoprotection, have been recently shown to be very important for many crystal structure determinations particularly of very large macromolecules (See Heras & Martin 2005 for a comprehensive overview of post crystallization methods that can significantly improve diffraction quality of crystals). Post crystallization methods to improve diffraction quality are more crucial for large macromolecular complexes since a new and possibly better diffracting crystal form is usually more

difficult to obtain compared to small protein targets (Mueller et al. 2007). More recently, the use of electron and radical scavengers to reduce radiation damage (Southworth-Davies & Garman 2007, Kauffmann et al. 2006) has been reported once for very large macromolecules (Kettenberger et al. 2003). It is possible that this method will become more widely used during data collection for crystals of large macromolecular complexes and in crystallography in general.

1.1.5 Diffraction data collection: basis of structure determination

1.1.5.1 Introduction

Crystal structure analysis using X-ray diffraction is, in many cases, the most advanced method available for obtaining high-resolution structural information about biological macromolecules. The ways in which X-ray diffraction data are collected and refined have a strong impact on the final quality of the structural models and the type and magnitude of their associated errors (Acharya & Lloyd 2005).

1.1.5.1.1 Importance of atomic resolution precision

The accuracy of the atomic positions is defined by the strength of crystal diffraction and thus by the resolution to which the X-ray data have been measured. The crystallographic resolution is approximately equal to the minimum distance between two points that can still be resolved in the electron density map (Schmidt & Lamzin 2007). This in turn defines the amount of detail that can be obtained from the resulting macromolecular model. Table 1.3 gives us an idea of how accurate we can

expect an X-ray crystal structure to be at various resolutions. During the crystal structure refinement the protein stereochemistry is usually restrained to the target values that are ordinarily taken from the library of small molecule compounds (Schmidt & Lamzin 2007). At a resolution of 3.0 Å or lower the amount of the observed data is very limited. The individual atoms and often even the side chains cannot be resolved in the electron density map. Consequently, the derived models are heavily influenced by predefined stereochemical restraints and the coordinate error in the atomic positions can be in the range of 0.5 Å. The resolution of 2.2 Å corresponds to about "typical" structures available from PDB, where chemical groups (phenyl rings, carboxylates, etc) and isolated atoms (e.g. solvent or metals) are already resolved. The coordinate error is of an order of 0.2 Å. At atomic resolution (1.2 Å or higher [Morris & Bricogne 2003]) not only the bonded atoms are resolved and their types become distinguishable but also the atomic positions can be accurately determined to 1/100th of an Ångstrom. At ultra-high resolution (0.8 Å or higher), the bonding electrons and hydrogen atoms can be detected and thus the chemical properties can be directly inferred from the crystal structure (Schmidt & Lamzin 2007).

Table 1.3 Resolution and structural information (adapted from Fersht 1985).

Resolution (Å)	Structural features observable in a good map
5.5	Overall shape of the molecule. Helices as rods of strong intensity.
3.5	The main chain (usually with some ambiguities).
3.0	The side chains partly resolved.
2.5	Side chains well resolved. The plane of the peptide bond resolved. Atom located to about ± 0.4 Å.
1.5	Atoms located to about ± 0.1 Å (the limit of protein crystallography as of 1985 – it's a little bit lower now).
0.77	Bond lengths in small crystals measured to 0.005 Å.

Note: (i) At moderate resolutions (such as 2.5 Å), the atoms are located only within 0.4 Å or so. This means that the lengths of hydrogen bonds calculated from the PDB file have at least this much uncertainty. (ii) Some information is available at resolution that doesn't resolve individual atoms. Exciting new structures appear at > 5 Å resolution are very informative in which they are often later superseded by structures at atomic resolution (For examples, see Fu et al. 1999, Ban et al. 1999, Clemons et al. 1999).

Atomic resolution and ultra-high resolution structures allow determination of its protonation state, prediction of the pKa value (using Amber9 program), or H-bonding interactions in general. Since most of the catalytic reactions involve hydrogen transfer of some sort, such information can be of high importance for establishing the catalytic mechanism, for example the importance of direct observation of hydrogen atoms has been extensively illustrated for a highly relevant drug target – the aldose reductase (Howard et al. 2004, Ruiz et al. 2004, Muzet et al. 2003). Determination of atomic resolution structures of complexes that represent various stages of the catalytic reaction can help in mapping the complete reaction path and understanding the catalytic mechanism. The high accuracy of the atomic positions allows reliable interpretation of ligand-induced structural changes, even if they seen small. This was beautifully demonstrated in a study of horseradish peroxidase, where X-ray radiation during the diffraction data collection was used to create various binding states of oxygen, hydrogen peroxide and water to the haem group, thus mapping the catalytic pathway (Berglund et al. 2002). High resolution structure data is also one of the key elements needed for drug design.

1.1.5.1.2 Diffraction of a simple object

Visual intuitions can be developed for the information available from X-ray diffraction by examining the diffraction patterns of simple objects like spheres or

arrays of spheres. If we examining the diffraction by a single sphere (this example is a simplified one and is used here for explaining purpose since the diffraction patterns involve arrays of simple objects), the diffraction pattern exhibits high intensity at the center and smoothly decreasing intensity as the diffraction angle increases.

1.1.5.1.3 Arrays of simple objects : real and reciprocal lattices

The diffraction pattern consists of reflections (spots) in an orderly array on the film. The spacing of the reflections varies with the spacing of the spheres in their array. In details, although the lattice spacing of the crystal is smaller vertically, the diffraction spacing is smaller horizontally. In fact, there is a simple inverse relationship between the spacing of unit cells in the crystalline lattice, called the real lattice, and the spacing of reflections in the lattice on the film, which, because of its inverse relationship to the real lattice, is called the reciprocal lattice. Thanks to this inversely proportional relationship, crystallographers can calculate the dimensions, in Ångströms, of the unit cell of the crystalline material from the spacings of the reciprocal lattice on the X-ray film. The simplicity of this relationship is a dramatic example of how the macroscopic dimensions of the diffraction pattern are connected to the submicroscopic dimensions of the crystal (Rhodes 2006).

1.1.5.1.4 Intensity of reflections

The variations in intensity of reflections contain important information. The diffraction pattern of spheres in a lattice is simply the diffraction pattern of the average sphere in the lattice, but this pattern is incomplete. The pattern is sampled at

points whose spacings vary inversely with real-lattice spacings. The pattern of varied intensities is that of the average sphere because all the spheres contribute to the observed pattern. To put it another way, the observed pattern of intensities is actually a superposition of the many identical diffraction patterns of all the spheres (Rhodes 2006).

1.1.5.1.5 Arrays of complex objects

The relationship between (1) diffraction by a single object and (2) diffraction by many identical objects in a lattice holds true for complex objects also. The reciprocal-lattice spacing (the spacing of reflections in the diffraction pattern) is characteristic of (inversely related to) the spacing of identical objects in the crystal, whereas the reflection intensities are characteristic of the shape of the individual objects (Rhodes 2006). From the reciprocal-lattice spacing in a diffraction pattern, we can compute the dimensions of the unit cell. From the intensities of the reflections, we can learn the shape of the individual molecules that compose the crystal. It is actually advantageous that the object's diffraction pattern is sampled at reciprocal-lattice positions. This sampling reduces the number of intensity measurements we must take from the film and makes it easier to program a computer to locate and measure the intensities.

1.1.5.1.6 Three-dimensional arrays

Unlike the two-dimensional arrays, a crystal is a three-dimensional array of objects. If we rotate the crystal in the X-ray beam, a different cross section of objects will lie

perpendicular to the beam and we will see a different diffraction pattern. In fact, just as the two-dimensional arrays of objects discussed above are cross sections of objects in the three-dimensional crystal, each two-dimensional array of reflections (each diffraction pattern recorded on film) is a cross section of a three-dimensional lattice of reflections. Figure 1.5 shows a hypothetical three-dimensional diffraction pattern, with the reflections that would be produced by all possible orientations of a crystal in the X-ray beam.

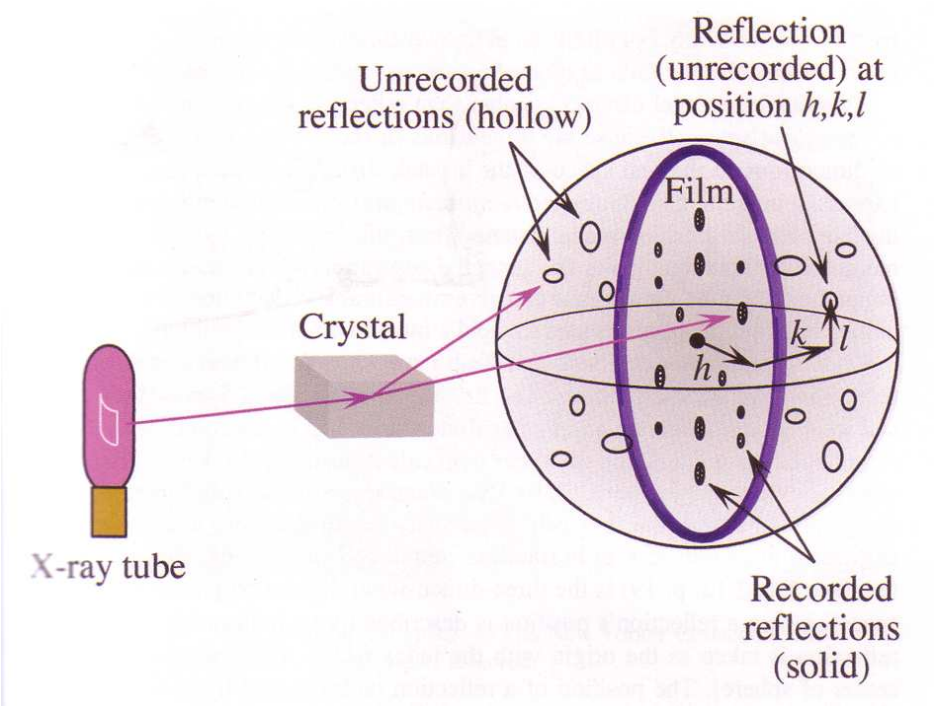


Figure 1.5 Crystallographic data collection, showing reflections measured at one particular crystal orientation (solid, on the film) and those that could be measured at other orientations (hollow, within the sphere but not on the film). Each reflection is located by its indices *h*, *k* and *l*. The relationship between measured and unmeasured reflections is more complex than shown here (Rhodes 2006).

Notice that only one plane of the three-dimensional diffraction pattern is superimposed on the film. With the crystal in the orientation shown, reflections

shown in the plane of the film (solid spots) are the only reflections that produce spots on the film (data measured). In order to measure the directions and intensities of all additional reflections (shown as hollow spots), the crystallographer must collect diffraction patterns from all unique orientations of the crystal with respect to the X-ray beam. The direct result of crystallographic data collection is a list of intensities for each point in the three-dimensional reciprocal lattice. This set of data is the raw material for determining the structures of molecules in the crystal. The spatial relationship involving beam, crystal, film and reflections is more complex than shown here (Rhodes 2006).

1.1.5.1.7 Coordinate systems in crystallography

Crystallographers work back and forth between two different coordinate systems. The first system is the unit cell (real space), where an atom's position is described by its coordinates *x*, *y*, *z*. A vertex of the unit cell, or any other convenient position, is taken as the origin, with coordinates *x*, *y*, *z* = (0, 0, 0). Coordinates in real space designate real spatial positions within the unit cell. Real-space coordinates are usually given in Ångströms or nanometers, or in fractions of unit cell dimensions. The second system (Figure 1.5) is the three-dimensional diffraction pattern (reciprocal space), where a reflection's position is described by its indices *hkl*. The central reflection is taken as the origin with the index *hkl* = 000 (round black dot at center of sphere). The position of a reflection is designated by counting reflections from 000, so the indices *h*, *k* and *l* are integers. Distances in reciprocal space, expressed in reciprocal angstroms ($Å^{-1}$) or reciprocal nanometers (nm^{-1}), are used to judge the potential resolution of the model that the diffraction data can yield.

1.1.5.2 Geometric principles of diffraction

1.1.5.2.1 The generalized unit cell

The position of a point *P* in the unit cell is given by its position vector r (Figure 1.6). In terms of its fractional coordinates *x*, *y*, and *z* with respect to the crystal axes a, b, and c, **r** is given by

$$\mathbf{r} = \mathbf{a}x + \mathbf{b}y + \mathbf{c}z.$$

The position of *P* can thus be described by its fractional coordinates (i.e., by its coordinate triplet *x*, *y*, *z*). The coordinate triplets of the points *P* and *P'*, related by the 2-fold axis along c in Figure 1.6, are *x*, *y*, *z* and –*x*, -*y*, *z*. If a molecule occupies position *x*, *y*, *z*, then an identical molecule occupies position –*x*, -*y*, *z*. These molecules are said to occupy general positions.

If, however, the molecule itself has a 2-fold axis that coincides with the crystallographic 2-fold axis, half of the molecule is mapped onto the other half by the symmetry operation. This molecule occupies a "special" position. Each special position has certain point symmetry. In the present example, the point symmetry would be given by the symbol 2.

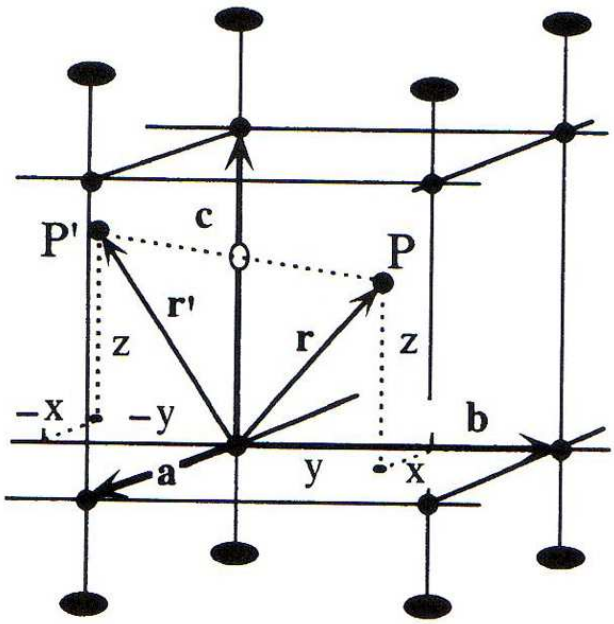


Figure 1.6 This crystal has a 2-fold axis along *c*. The point *P* with coordinate triplet *x, y, z* is related by the symmetry operation to point *P'* with coordinate triplet *–x, -y, z* (Drenth, 2007).

1.1.5.2.2 Indices of the atomic planes in a crystal

The most readily apparent sets of planes in a crystalline lattice are those determined by the faces of the unit cells. The orientation of a surface or a crystal regularly spaced plane may be defined by considering how the plane (or indeed any parallel plane that can be drawn through lattice points) intersects the main crystallographic axes of the solid. The application of a set of rules leads to the assignment of a set of three numbers which quantify the intercepts and thus may be used to uniquely identify the plane or surface, called the *lattice indices* or the *Miller indices*. Three indices *hkl* identify a particular set of equivalent, parallel planes. The index *h* gives the number of planes in the set per unit cell in the x direction or, equivalently, the number of parts into which the set of planes cut the *a* edge of each cell (for an example, see below). The indices *k* and *l* specify how many such planes exist per unit

cell in the y and z directions. An equivalent way to determine the indices of a set of planes is to start at any lattice point and move out into the unit cell away from the plane cutting that lattice point. If the first plane encountered cuts the *a* edge at some fraction $1/n_a$ of its length, and the same plane cuts the b edge at some fraction $1/n_b$ of its length, then the *h* index is n_a and the *k* index is n_b. Indices are written in parentheses when referring to the set of planes; hence, the planes having indices *hkl* are the (*hkl*) planes.

All planes perpendicular to the xy plane have indices *hk*0. Planes perpendicular to the xz plane have indices *h*0*l*, and so forth. Many additional sets of planes are not perpendicular to x, y, or z. For example, the (234) planes cut the unit cell edges ***a*** into two parts, ***b*** into three parts, and ***c*** into four parts (Figure 1.7). Miller indices can be negative as well as positive. Sets of planes in which all indices have opposite signs are identical.

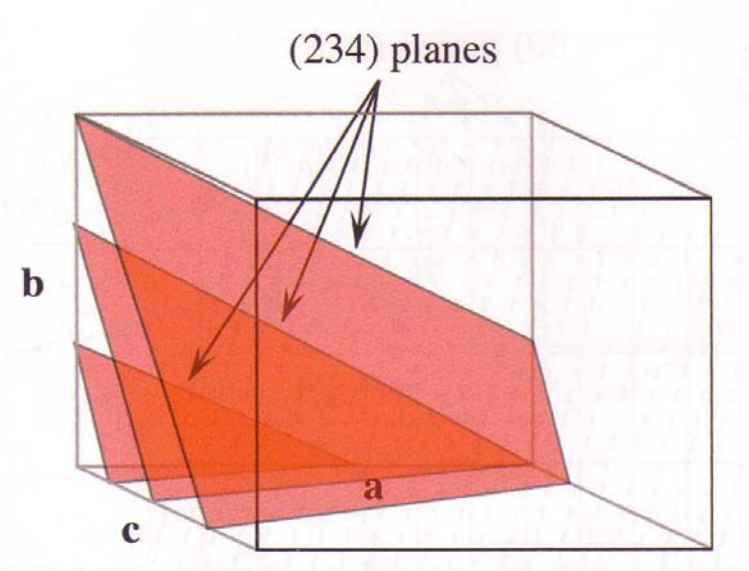


Figure 1.7 The intersection of three (234) planes with a unit cell. Note that the (234) planes cut the unit-cell edges **a** into two parts, **b** into three parts, and **c** into four parts (Rhodes 2006).

1.1.5.2.3 Bragg's law in reciprocal space

W. L. Braggs managed to visualize the scattering X-rays from a crystal by considering that the diffracted beams were reflected by planes with index *hkl* passing through points of a crystal lattice. The diffracted X-rays are scattered by the crystal at a certain angle of reflection (θ). This reflection is analogous to that from a mirror, for which the angle of incident X-ray beam is equal to the angle of diffracted X-ray beam. The incident and the diffracted X-ray beam are in the same plane and the X-rays of wavelength (λ) are normal to a set of diffracting planes (Figure 1.8). The constructive interference between X-rays scattered from successive planes in the crystal will only take place if the path difference ($2d_{hkl}$) between the X-rays is equivalent to an integral number (n) of wavelength (λ). That is the Bragg's law equation (Glusker & Trueblood 1985, Rhodes 2006):

$$n\lambda = 2d_{hkl} \sin \theta$$

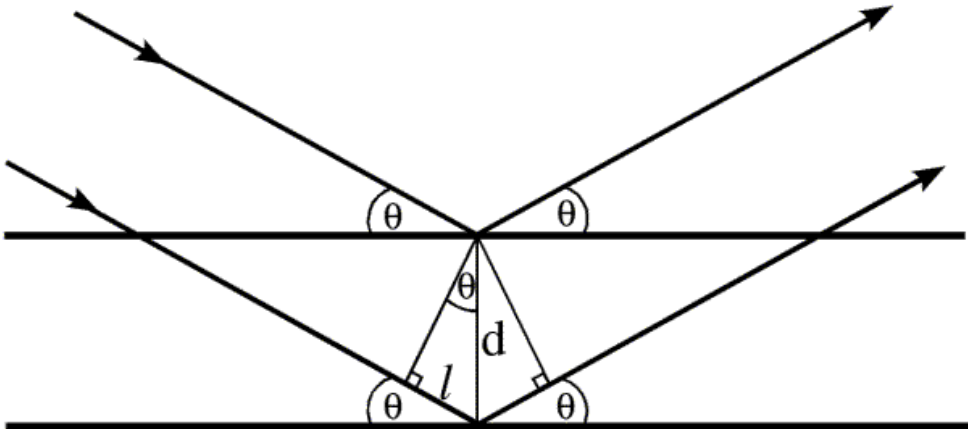


Figure 1.8 The geometry of diffraction and its relationship to Bragg's Law (Glusker & Trueblood 1985).

In Bragg's law, if the wavelength and the diffraction angle of a reflection are known, the perpendicular distance between the lattice planes in the crystal (interplanar spacing, d_{hkl}) can be easily calculated. As the angle increases, d_{hkl} must become smaller for the path length to remain equal to one wavelength. The equation can be rearranged as:

$$d_{hkl} = \frac{n\lambda}{(2 \sin\theta)}$$

The minimum interplanar spacing ($d_{hkl\min}$), where $d_{hkl\min} = \lambda/2\ (\sin\theta_{max})$, is also interpreted as the resolution of an electron density map. Since the maximum possible value of sinθ is 1, so the smaller the $d_{hkl\min}$ value, the higher the resolution will be of the X-ray diffraction pattern. For instance, if the radiation used for X-ray generation has a wavelength of 1.54 Å, then the maximum resolution that can be observed with this radiation would be 0.77 Å (Blundell & Johnson 1976, Glusker et al. 1994).

1.1.5.2.4 Number of measurable reflections

If the sphere of reflection has a radius of 1/λ, and the crystal is rotated about origin 000 on the surface of this sphere, then any reciprocal-lattice point within a distance 2/λ of the origin can be rotated into contact with the sphere of reflection (Figure 1.9). This distance defines the limiting sphere. The number of reciprocal lattice points within the limiting sphere is equal to the number of reflections that can be produced by rotating the crystal through all possible orientations in the X-ray beam. This demonstrates that the unit-cell dimensions and the wavelength of the X-rays determine the number of measurable reflections. Shorter wavelengths make a larger

sphere of reflection, bringing more reflections into the measurable sphere. Larger unit cells mean smaller reciprocal unit cells, which populate the limiting sphere more densely, also increasing the number of measurable reflections.

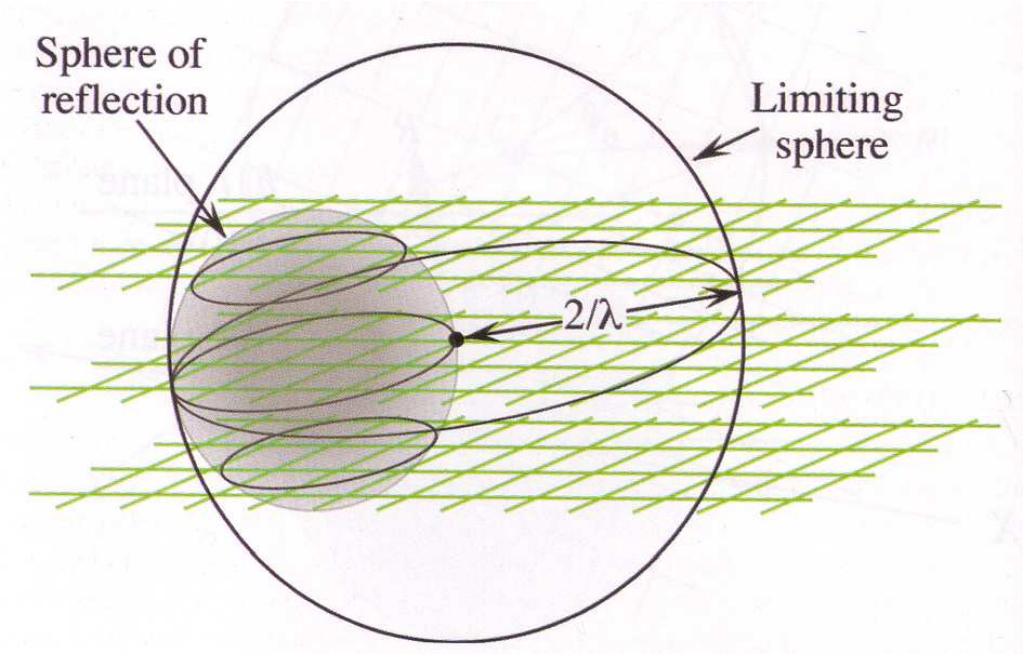


Figure 1.9 Limiting sphere. All reciprocal-lattice points within the limiting sphere of radius $2/\lambda$ can be rotated through the sphere of reflection (Rhodes 2006).

Because there is one lattice point per reciprocal unit cell (one-eighth of each lattice point lies within each of the eight unit-cell vertices), the number of reflections within the limiting sphere is approximately the number of reciprocal unit cells within this sphere. So the number N of possible reflections equals the volume of the limiting sphere divided by the volume V_{recip} of one reciprocal cell. The volume of a sphere of radius r is $(4\pi/3)r^3$, and r for the limiting sphere is $2/\lambda$, so

$$N = \frac{\left(4\frac{\pi}{3}\right)\frac{2}{\lambda^3}}{V_{recip}}$$

The volume V of the real unit cell is V_{recip}^{-1}, so

$$N = \frac{33.5\ V}{\lambda^3}$$

The latter equation shows that the number of available reflections depends only on V and λ, the unit-cell volume and the wavelength of the X radiation. Due to the cell and reciprocal-lattice symmetry, not all the reflections are unique. Nevertheless, getting most of the available information from the diffraction experiment with protein crystals may requires measuring somewhere between one thousand and one million reflections (Rhodes 2006).

1.1.5.2.5 Unit-cell dimensions

Because reciprocal-lattice spacings determine the angles of reflection, the spacings of reflections on the detector are related to reciprocal-lattice spacings. The exact relationship depends on the geometry of recording the reflections. Reciprocal-lattice spacings, in turn, are simply the inverse of real-lattice spacings. So the distances between reflections on the detector and the dimensions of the unit cell are closely connected, making it possible to measure unit-cell dimensions from reflection spacings (Rhodes 2006). The unit cell, basic building block of the crystal, is repeated infinitely in three dimensions. The directions of constructive interference depend only on the size and shape of the unit cell. The dimensions of a unit cell are designated by six parameters: the length of three unique edges (a, b, c) which run along x, y and z coordinates respectively, and three unique angles (α, β, γ) as indicated in Figure 1.10.

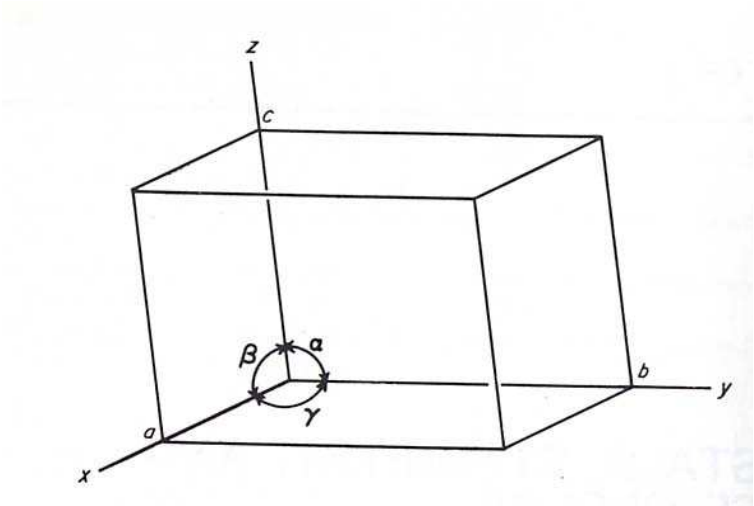


Figure 1.10 The unit cell with edges, a, b, c and angles, α, β and γ (Stout & Jensen 1989).

1.1.5.2.6 Unit-cell symmetry

There are 14 allowable unit cell types classified as Bravais lattice to distinguish their characteristic. The Bravais lattices themselves can be divided into five types of lattice, which are primitive (**P**), centered (**C**), body-centered (**I**), face-centered (**F**) and rhombohedral (**R**) (McPherson 2003, Stout and Jensen 1989). Any crystal can be regarded as being established by consecutively translational repetition of the unit cell and its content along a, b, c by distance |a|, |b|, |c| respectively, until a continuous three-dimensional array of repeated unit cells in a regular manner has been created (Glusker et al. 1994, Glusker & Trueblood 1985, Rhodes 2006).

The simple symmetry operations and elements needed to describe unit cell symmetry are translation, rotation and reflection. The symmetry of a unit cell is described in 230 space groups (like $P2_12_12_1$). The space group is a group of symmetry operations consistent with an infinitely extended, regularly repeating pattern. Protein is an asymmetric object since all amino acids except glycine have

chirality. However, the D-form of amino acids does not exist in proteins and only the L form does. Thus, there are less symmetry elements (mirror planes, inversion centers and glide planes) involved within the unit cell, and less space groups can be used to designate the protein. This limits the possible space groups to 65 out of the 230 mathematically possible ones (McRee 1999).

There are seven crystal systems used to classify the symmetry of the crystal, this corresponds to the seven fundamental shapes for the unit cells, consistent with the 14 Bravais lattices as displayed in the Table 1.4 (Glusker et al. 1994, Stout and Jensen 1989).

Table 1.4 The seven crystal systems (Glusker et al. 1994, Stout and Jensen 1989).

	Crystal System	Bravais Lattices	Lattice	Angle
1	Triclinic	P	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$
2	Monoclinic	P, C	$a \neq b \neq c$	$\alpha = \gamma = 90° \neq \beta$
3	Orthorhombic	P, C, I, F	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90°$
4	Tetragonal	P, I	$a = b \neq c$	$\alpha = \beta = \gamma = 90°$
5a	Trigonal	P	$a = b \neq c$	$\alpha = \beta = 90°, \gamma = 120°$
5b	Rhombohedral	R	$a = b = c$	$\alpha = \beta = \gamma < 120°, \neq 90°$
6	Hexagonal	P	$a = b \neq c$	$\alpha = \beta = 90°, \gamma = 120°$
7	Cubic	P, I, F	$a = b = c$	$\alpha = \beta = \gamma = 90°$

1.1.5.2.7 Friedel pairs

Friedel's law states that the intensities of the *h*, *k*, *l* and $\bar{h}, \bar{k}, \bar{l}$ reflections are equal (in other words members of a Friedel pair have equal amplitude and opposite phase). The couple of reflections *h*, *k*, *l* and $\bar{h}, \bar{k}, \bar{l}$ is called a Friedel pair or Bijvoet pair. They are Bragg reflections related by inversion through the origin. This is only true if there is no absorption. It is in that case not possible to tell by diffraction whether an

inversion center is present or not. Generally speaking, when absorption is present, equivalent reflections generated by the symmetry elements in the crystal have intensities different from those of equivalent reflections generated by the introduction of an additional inversion centre in normal scattering. Friedel or Bijvoet pairs are used in the resolution of the phase problem for the solution of crystal structures. The use of Friedel's pairs for helping in phase determinations was suggested by Bijvoet *et al.* (1951). Friedel's law is broken whenever there is anomalous scattering. If all atoms scatter equally, then the amplitudes remain equal but the phase relationship no longer holds. If some atoms scatter anomalously and some don’t, then both the amplitude and phase relationships are broken.

1.1.5.3 Collecting X-ray diffraction data

1.1.5.3.1 X-ray sources

The mounted crystal (see previous section 1.1.4.4) is irradiated with a beam of monochromatic X-rays. The brightest and most useful X-ray sources are synchrotrons; their much higher brightness allows for better resolution. They also make it convenient to tune the wavelength of the radiation, which is useful for multi-wavelength anomalous diffraction (MAD) phasing, described below. Synchrotrons are generally national facilities, each with several dedicated beamlines where data are collected around the clock, seven days a week. Crystallographers apply for a slot of time, which they must use whenever it is granted, even at 3am on a national holiday. Crystallographers will sometimes stay awake for days, collecting data continuously until their allotted time runs out.

Smaller, weaker X-ray sources are often used in laboratories to check the quality of crystals before bringing them to a synchrotron and sometimes to solve a crystal structure. In such systems, electrons are boiled off of a cathode and accelerated through a strong electric potential of roughly 50 kV; having reached a high speed, the electrons collide with a metal plate, emitting *bremsstrahlung* (free-free emission) and some strong spectral lines corresponding to the excitation of inner-shell electrons of the metal. The most common metal used is copper, which can be kept cool easily, due to its high thermal conductivity, and which produces strong K_α and K_β lines. The K_β line is sometimes suppressed with a thin layer (0.0005 in. thick) of nickel foil. The simplest and cheapest variety of sealed X-ray tube has a stationary anode and produces *circa* 2 kW of X-ray radiation. The more expensive variety has a rotating-anode type source that produces *circa* 14 kW of X-ray radiation.

X-rays are generally filtered to a single wavelength (made monochromatic) and collimated to a single direction before they are allowed to strike the crystal. The filtering not only simplifies the data analysis, but also removes radiation that degrades the crystal without contributing useful information. Collimation is done either with a collimator (basically, a long tube) or with a clever arrangement of gently curved mirrors. Mirror systems are preferred for small crystals (under 0.3 mm) or with large unit cells (over 150 Å).

1.1.5.3.2 Synchrotron radiation

Synchrotrons are devices for circulating electrically charged particles (negatively charged electrons or positively charged positrons) at nearly the speed of light

(Helliwell 1992). The particles are injected into the storage ring directly from a linear accelerator or through a booster synchrotron. When the particle beam changes direction (the trajectory of the particles is determined by their energy and by the magnetic field), the electrons or positrons are accelerated toward the center of the ring and therefore emit electromagnetic radiation and consequently, lose energy. This energy loss is compensated for by a radio-frequency input at each cycle. Ever since the radiation has been discovered as a useful and extremely powerful tool for crystallographic studies by chemists and molecular biologists (Rosenbaum 1971), radiation dedicated synchrotrons have been constructed and the use of synchrotron X-ray sources for macromolecular crystallography has increased enormously.

The main advantage of synchrotron radiation for X-ray diffraction is its high intensity, which is five orders of magnitude stronger than for a conventional X-ray tube, at least for radiation from a bending magnet (magnetic device that guide the electrons in their orbit), and a few orders of magnitude stronger for radiation from a multipole wiggler (magnetic device that produces a number of sharp extra bends in the electron trajectory) or an undulator (similar to a wiggler but with the difference that interference effects cause the emission of radiation at more specific wavelengths). This high intensity is profitably used by protein X-ray crystallographers for data collection. Another advantage is the low divergence of the beam resulting in sharper diffraction spots. Synchrotron radiation also differs from tube radiation in its tunability. Any suitable wavelength in the spectral range can be selected with a monochromator. This property is used in MAD and for Laue diffraction studies. In terms of time structure, synchrotron radiation, in contrast to X-ray tube radiation, is produced in flashes by the circulating bunches of charged particles in the picoseconds range. This allows structural changes in the nanosecond

timescale to be observed (Drenth 2007). The X-ray beam from an X-ray tube is not polarized whereas synchrotron radiation is highly polarized. However, the polarization of the X-ray beam from latter source has not yet found extensive application in X-ray diffraction. Nevertheless, it must be considered when applying the correct polarization factor because of its effect on the so-called anomalous X-ray scattering of atoms, which occurs when the X-ray wavelength approaches an absorption-edge wavelength.

Currently, the highest energy synchrotron in the world is the Tevatron, at the Fermi National Accelerator Laboratory, in the United States. It accelerates protons and antiprotons to slightly less than 1TeV of kinetic energy and collides them together. On the other hand, the largest of those 3rd generation synchrotron light sources are the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, the Advanced Photon Source (APS) near Chicago, USA, and Spring-8 in Japan, accelerating electrons up to 6, 7 and 8 GeV, respectively. Synchrotrons which are useful for cutting edge research are large machines, costing tens or hundreds of millions of dollars to construct, and each beamline (there may be 20 to 50 at a large synchrotron) costs another two or three million dollars in average. These installations are mostly built by the science funding agencies of governments of developed countries, or by collaborations between several countries in a region, and operated as infrastructure facilities available to scientists from universities and research organizations throughout the country, region, or world.

1.1.5.3.3 Determining unit-cell dimensions

A simple method for determining the unit-cell dimensions (for example, measuring reflection spacings from an orthorhombic crystal on an image of a zero-level plane; this example is chosen for its simplicity to the understanding of the concept) is described here. Figure 1.11 shows the geometric relationship between reflection spacings on the film and actual reciprocal-lattice spacings.

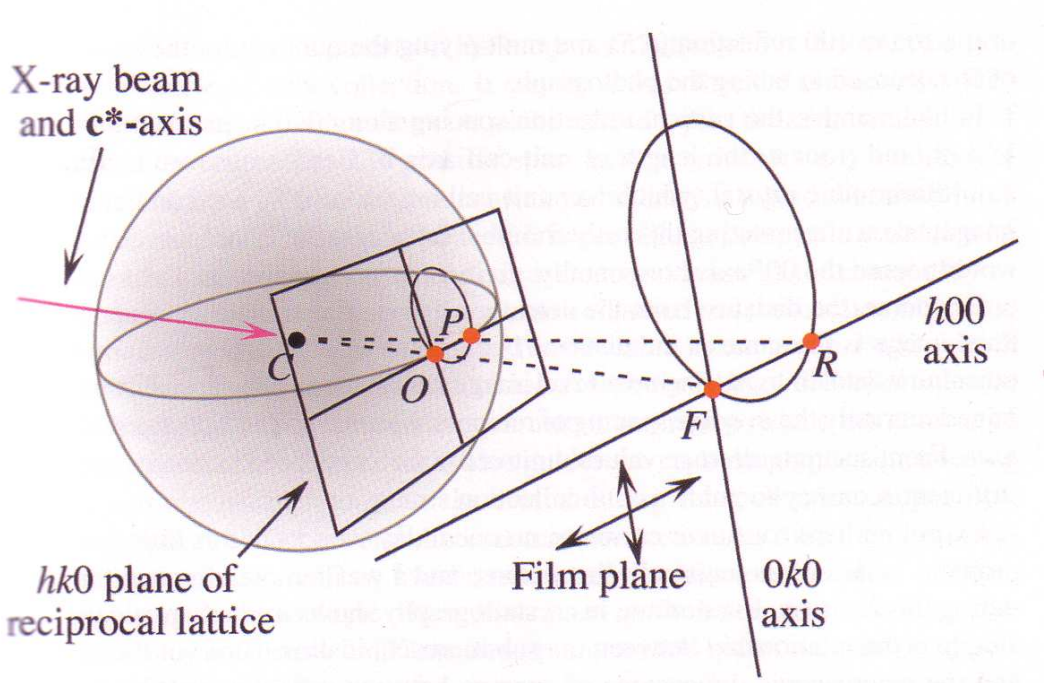


Figure 1.11 Reflection spacings on the film are directly proportional to reciprocal-lattice spacings, and so they are inversely proportional to unit-cell dimensions (Rhodes 2007).

The crystal at *C* is processing about its *c**-axis, and therefore recording hk0 reflections on the detector, with the *h*00 axis horizontal and the 0*k*0 axis vertical. Point *P* is the reciprocal-lattice point 100, in contact with the sphere of reflection, and *O* is the origin. Point *F* is the origin on the detector and *R* is the recording of

reflection 100 on the detector. The distance *OP* is the reciprocal of the distance d_{100}, which is the length of unit-cell edge a. Because *CRF* and *CPO* are similar triangles (all corresponding angles equal), and because the radius of the sphere of reflection is $1/\lambda$,

$$\frac{RF}{CF} = \frac{PO}{CO} = \frac{PO}{\frac{1}{\lambda}} = PO\,\lambda\,.$$

Therefore,

$$PO = \frac{R\,F}{CF\,\lambda}$$

Because $d_{100} = 1/PO$,

$$d_{100} = \frac{CF\,\lambda}{R\,F}$$

In other words, the axial length a (length of unit-cell edge *a*) can be determined by dividing the crystal-to-detector distance (*CF*) by the distance from the detector origin to the 100 reflection (*RF*) and multiplying the quotient by the wavelength of X-rays used in taking the photograph.

In like manner, the vertical reflection spacing along 0*k*0 or parallel axes gives $1/d_{010}$, and from it, the length of unit-cell axis *b*. Because we are considering an orthorhombic crystal which has the angles of 90°, a second zero-level image, taken after rotating this orthorhombic crystal by 90° about its vertical axis, would record the 00*l* axis horizontally, giving $1/d_{001}$, and the length of *c*.

Of course, the distance from the detector origin to the 100 reflection on a zero-level image is the same distance as the distance between any two reflections along this or other horizontal lines, so one zero-level image allows many measurements to determine accurately the average values, unit-cell-axis lengths can be determined with sufficient accuracy to guide a data-collection strategy.

1.1.5.3.4 Symmetry and the strategy of collecting data

If the unit-cell contents are symmetric, then the reciprocal lattice is also symmetric and certain sets of reflections are equivalent. In theory, only one member of each set of equivalent reflections need to be measured, so awareness of unit-cell symmetry can greatly reduce the magnitude of data collection. In practice, redundancy of measurements improves accuracy, so when more than one equivalent reflection is observed (measured), or when the same reflection is observed more than once, the average of these multiple observations is considered more accurate than any single observation.

Large unit cell dimensions are in general inherent to crystals of large macromolecular complexes. Therefore, starting with data collection, various crystallographic challenges imposed by large unit cells have to be tackled. Crystals with large unit cells diffract more weakly because fewer unit cells per crystal volume contribute to scattering. Furthermore, as a consequence of long unit cell axes, the distances between reflections in reciprocal space are small, leading to closely spaced diffraction spots on the detector (Mueller et al. 2007).

For successful data collection, one needs the best possible instrumentation as well as a thoughtful data collection strategy. Among the general considerations that should be taken into account (Dauter 1999), the orientation of unit cell axes is of special importance. If one of the unit cell axes is substantially longer than the others it should be aligned with the spindle axis of crystal rotation to avoid overlapping reflections from neighboring lunes. This strategy was crucial for the collection of diffraction data from crystals of, for example, the large ribosomal subunit (Ban et al. 2000) and yeast fatty acid synthase, FAS (Leibundgut et al. 2007). Precisely aligning

unit cell axes with the spindle axis using a kappa goniostat can be beneficial for the collection of anomalous data. This way, Friedel pairs are recorded on the same image, limiting the effect of radiation damage and systematic errors on the anomalous signal. This procedure was used for many of the experimentally phased structures of very large macromolecules (Abrahams & Ban 2003), for example, the 30 S ribosomal subunit (Clemons et al. 2001) and the RNA polymerase II (Cramer et al. 2000). Modern beamlines with fast and precise goniometers now usually allow data collection in inversed beam mode, which records Friedel pairs on two consecutive images.

The strong beam intensity and long exposures necessary to collect high-resolution diffraction data from crystals of large macromolecular complexes causes severe radiation damage and frequently limits the amount of data that can be collected from one crystal even at cryogenic temperatures. For a number of structures, it has been reported that data was collected from many crystals, for example, for bacterial ribosomes (Berk et al. 2006, Petry et al. 2005) and tetranucleosomes (Schalch et al. 2005). However, potential problems with non-isomorphism, crystal quality, and twinning have to be sorted out before this approach can be used. One should keep in mind that careful optimization of post crystallization treatments will reduce problems with non-isomorphism and variations in cell dimension (Clemons et al. 2001, Cramer et al. 2000). When diffraction data need to be phased experimentally radiation damage becomes even more problematic and the strategies for phasing and data collection have to be chosen carefully (Ravelli et al. 2005, González 2003). Phasing by MAD has rarely been successful for very large macromolecules (Cramer et al. 2000, Cate et al. 1999), most probably because the accumulation of radiation damage during data collection at three different

wavelengths. Therefore, it seems more promising to aim for single-wavelength anomalous diffraction (SAD) phasing and to collect a redundant dataset at the absorption peak wavelength, a strategy that can also be useful when phasing crystals with smaller asymmetric units (Rice et al. 2000). Datasets from different isomorphous derivatives collected this way might also allow phasing by single or multiple isomorphous replacements and anomalous scattering.

1.1.5.3.5 Scaling and post refinement of intensity data

The goal of data collection is a set of consistently measured, indexed intensities for as many of the reflections as possible. After data collection, the raw intensities must be processed to improve their consistency and to maximize the number of measurements that are sufficiently accurate to be used.

A complete set of measured intensities often includes many frames, as well as distinct blocks of data obtained from several (or many) crystals in different orientations. Because of variability in the diffracting power of crystals, the difference in the length of the X-ray path through the crystal in different orientations, and the intensity of the X-ray beam, the crystallographer cannot assume that the absolute intensities are consistent from one frame or block of data to the next. An obvious way to obtain this consistency is to compare reflections of the same index that were measured from more than one crystal or in more than one frame and to rescale the intensities of the two blocks of data so that identical reflections are given identical intensities. This process is called scaling. Scaling is often preliminary to a more complex process, post refinement, which recovers usable data from reflections that were only partially measured.

Primarily because real crystals are mosaics of submicroscopic crystals, a reciprocal-lattice point acts as a small three-dimensional entity (sphere or ovoid) rather than as an infinitesimal point. As a reciprocal-lattice point moves through the sphere, diffraction is weak at first, peaks when the center of the point lies precisely on the sphere, and then weakens again before it is extinguished. Accurate measurement of intensity thus entails recording the X-ray output during the entire passage of the point through the sphere. Any range of oscillation will record some reflections only partially, but these may be recorded fully at another rotation angle, allowing partial reflections to be discarded from the data. The problem for partial reflections is serious for large unit cells, where smaller oscillation angles are employed to minimize overlap of reflections. In such cases, if partial reflections are discarded then a great deal of data is lost. Data from partial reflections can be interpreted accurately through post refinement of the intensity data. This process produces an estimate of the partiality of each reflection. Partiality is a fraction p ($0 \geq p \geq 1$) that can be used as a correction factor to convert the measured intensity of a partial reflection to an estimate of that reflection's full intensity. Scaling and post refinement are the final stages in producing a list of internally consistent intensities for as many of the available reflections as possible.

1.1.6 From diffraction data to electron density

The three main qualities of a diffraction data set which need to be optimized during data collection are resolution, completeness and accuracy. However, in reality one will have to make compromises between these three due to probably time and resource limitations, crystal decay, etc. Knowing the crystals, equipment and,

importantly, the problem we want to solve with the data will certainly aid to make the best trade-off.

1.1.6.1 The Fourier transform : general features

What does the Fourier sum model add to the Bragg model? Bragg's model tells us where to look for the data. The Fourier sum model tells us what the data has to say about the molecular structure, that is, about where the atoms are located in the unit cell. Each atom or each small volume element of electron density is treated as an independent diffractor and the model is represented by one term in a Fourier sum that describes each reflection.

Fourier demonstrated that for any function $f(x)$, there exists another function $F(h)$ such that

$$F(h) = \int_{-\infty}^{+\infty} f(x)\, e^{2\pi i(hx)} dx$$

where $F(h)$ is called the Fourier transform (FT) of $f(x)$, and the units of the variable h are reciprocals of the units of x. For example, if x is time in seconds (s), then h is reciprocal time, or frequency, in reciprocal seconds (s^{-1}). So, if $f(x)$ is a function of time, $F(h)$ is a function of frequency. Taking the FT of time-dependent functions is a means of decomposing these functions into their component frequencies and is sometimes referred to as Fourier analysis. The FT in this form is used in infrared (IR) and nuclear magnetic resonance (NMR) spectroscopy to obtain the frequencies of many spectral lines simultaneously.

On the other hand, if x is a distance or length in Å, h is reciprocal length in $Å^{-1}$. The Fourier transform is a precise mathematical description of diffraction. Thus,

this highly general mathematical form is naturally adapted for relating real and reciprocal space. The diffraction patterns are Fourier transforms of the corresponding simple objects and arrays. According to the equation above, to compute $F(h)$, the Fourier transform of $f(x)$, just multiply the function by $e^{2\pi i(hx)}$ and integrate (by computer) the combined functions with respect to x. The result is a new function $F(h)$, which is the FT of $f(x)$. Computer routines for calculating FTs of functions are widely available, and form one of the vital internal organs of crystallographic software. The Fourier transform operation is reversible. That is, the same mathematical operation that gives $F(h)$ from $f(x)$ can be carried out in the opposite direction, to give $f(x)$ from $F(h)$ specifically.

There is one added complication. The preceding functions $f(x)$ and $F(h)$ are one-dimensional. Fortunately, the Fourier transform applies to periodic functions in any number of dimensions. To restate Fourier's conclusion in three dimensions, for any function $f(x, y, z)$ there exists the function $F(h, k, l)$ such that

$$F(h, k, l) = \sum_h \sum_k \sum_l f(x,y,z) e^{-2\pi i(hx+ky+lz) + i\alpha(hkl)}\, dx\, dy\, dz.$$

As before, $F(h, k, l)$ is called the Fourier transform of $f(x, y, z)$, and in turn $f(x, y, z)$ is the Fourier transform of $F(h, k, l)$. Recapitulating all the Fourier terms in brief,

(i) A Fourier sum is a sum of simple wave equations or periodic functions that describes or approximates a complicated periodic function.

(ii) Constructing a Fourier sum – that is, determining the proper F, h and α values to approximate a specific function – is called Fourier synthesis.

(iii) Decomposing a complicated function into its components is called Fourier analysis.

(iv) The Fourier transform is an operation that transforms a function containing variables of one type (e.g. time) into a function whose variables are reciprocals

of the original type [in this case, 1/(time) or frequency]. The function $f(x)$ is related to its Fourier transform $F(h)$ by equation shown above. The term transform is commonly used as a noun to refer to the function $F(h)$ and also loosely as a verb to denote the operation of computing as a Fourier transform.

(v) A Fourier series is an infinite sum based on some iterative formula for generating each term. So, a sum of experimental terms (such as X-ray reflections), or a sum of individual atomic or electron-density contributions to a reflection is a Fourier sum, not a series.

1.1.6.2 From data to density

One common use of a dataset is to calculate an electron density map. The electron density at a point x, y, z is given by:

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} |F_{hkl}| e^{-2\pi i(hx+ky+lz) + i\alpha(hkl)}$$

The density is the sum of the contributions of each reflection and the importance of each reflection is proportional to its amplitude. Low resolution reflections tend to be strong and therefore make a large contribution to the map. High resolution reflections are weaker, but there are many more of them and they provide the detail needed for many interpretations. Therefore, for good electron density maps data with as high resolution as possible but without losing the low resolution ones at the same time is needed. Every reflection not measured causes a loss of signal proportional to the amplitude of the reflection. In other words, a missed reflection is equivalent to a measurement with a 100 % error. Errors in the measurement correspond to some noises in the density map. In a typical dataset measurement error is just several percent, increasing to perhaps 50 % or so at high resolution. Hence a poor

measurement is normally still a lot better than no measurement at all. It is also useful to realize that even for a highly refined structure map noise due to errors in F_{obs} measurement is small compared to errors in F_{calc} and phase errors. In conclusion, for density maps, resolution and data completeness are most important.

1.1.7 Obtaining phases

1.1.7.1 Introduction

The molecular image that the crystallographer seeks is a contour map of the electron density $\rho(x, y, z)$ throughout the unit cell. The electron density, like all periodic functions, can be represented by a Fourier sum. The representation that connects $\rho(x, y, z)$ to the diffraction pattern is

$$\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| e^{-2\pi i(hx+ky+lz)}.$$

This equation tells us how to calculate $\rho(x, y, z)$: simply construct a Fourier sum using the structure factors F_{hkl}. For each term in the sum h, k and l are the indices of reflection *hkl*, and F_{hkl} is the structure factor that describes the reflection. Each structure factor F_{hkl} is a complete description of a diffracted ray recorded as reflection *hkl*. Being a wave equation, F_{hkl} must specify frequency, amplitude and phase. Its three frequency terms *h*, *k* and *l* are the indices of the set of parallel planes that produce the reflection. Its amplitude is proportional to $(I_{hkl})^{1/2}$, the square root of the measured intensity I_{hkl} of reflection *hkl*. Its phase is unknown and is the only additional information the crystallographer needs in order to compute $\rho(x, y, z)$ and thus obtain an image of the unit cell contents. Therefore, desired electron density as a function of the known/measurable amplitudes $|F_{hkl}|$ (measurable because it can be

computed from the reflection intensity I) and the unknown phases α'_{hkl} of each reflection can be expressed in the following equation:

$$\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| e^{-2\pi i(hx+ky+lz-\alpha'_{hkl})},$$

with α is the phase α_{hkl} of the specific reflection *hkl*. Each term in the series is a three-dimensional wave of amplitude $|F_{hkl}|$, phase α'_{hkl}, and frequencies *h* along the *x*-axis, *k* along the *y*-axis, and *l* along the *z*-axis.

The most demanding element of macromolecular crystallography is the so-called phase problem that of determining the phase angle α_{hkl} for each reflection. Each reflection has its own phase so the phase problems must be solved for every one of the thousands of reflections used to construct the Fourier sum that approximates $\rho(x, y, z)$. There are some common methods, for example heavy-atom method (also called isomorphous replacement), anomalous scattering (also called anomalous diffraction), and molecular replacement, to overcome this obstacle. Each of these techniques yield only estimates of phases, which must be improved before an interpretable electron-density map can be obtained. In addition, these techniques usually yield estimates for a limited number of phases, so phase determination must be extended to include as many reflections as possible.

1.1.7.2 Isomorphous replacement phasing

Isomorphous replacement phasing compares F_{native} and $F_{derivative}$, where the latter is the diffraction of a crystal with a bound heavy atom. Binding of the derivative normally leads to structural changes which lead to non-isomorphism. This effect is stronger at high resolution. Therefore, one can normally not use the high resolution data. In addition, medium or low resolution (3 to 5 Å) is often sufficient to solve the

phase problem as additional phase information can be obtained by density modification and, especially powerful, from a partial model. As for the data completeness, if 10% of reflections are randomly missing from both native and derivative datasets, then only 80% of reflections will have both an F_{native} and $F_{derivative}$ and can be used for phasing. Low levels of (random) incompleteness can however be tolerated as missing phases can be retrieved by density modification. The difference between F_{native} and $F_{derivative}$ is generally on the order of 10 to 25%, considerably more than the normal measurement error. In addition, non-isomorphism often makes a significant contribution to errors. Nevertheless, since we are interested in the difference between amplitudes measurement errors become more significant.

1.1.7.3 Anomalous scattering

Phasing based on anomalous diffraction compares F+ and F- which are the Friedel mates (or Bijvoet pairs). Unlike isomorphous replacement phasing, anomalous diffraction does not suffer from non-isomorphism and in theory the anomalous signal even increases with resolution. Collecting high resolution data is thus useful. However, it is still true that ~3 Å resolution phases are often sufficient so resolution should not be pushed to the limit. In terms of data completeness, what was mentioned before for isomorphous replacement phasing applies here as well. The difference between F+ and F- is generally on the order of 1 to 3%, which is normally less than the measurement error. This is made worse by the fact that the error in ||F+| - |F-|| is larger than the errors in F+ and F- themselves. We are therefore interested in a signal that is small compared to measurement error. Accordingly, accuracy is paramount to successful anomalous diffraction phasing. Errors leading to outliers in

the data are particularly damaging as it can generate anomalous differences that are several times what they should have been. Detecting heavy atom positions happens in Patterson space and is thus sensitive to the strongest coefficients, which would be due to these outliers. The best way to reduce/prevent outliers is to collect a highly redundant dataset and reduce radiation damage by not pushing for the highest possible resolution.

A second means of obtaining phases from heavy-atom derivatives takes advantage of the heavy atom's capacity to absorb X-rays of specified wavelength. As a result of this absorption, Friedel's law does not hold and the reflections $h\ k\ l$ and $\bar{h}\ \bar{k}\ \bar{l}$ are not equal in intensity. This inequality of symmetry-related reflections is called anomalous scattering or anomalous diffraction.

Under anomalous scattering, the members of a Friedel pair can be used to establish the phase of a reflection in the heavy atom derivative data, thus establishing the phase of the corresponding reflection in the native data. For example, in obtaining the initial structure factors by SIR with anomalous scattering (called SIRAS, single isomorphous replacement with anomalous scattering), a complete data set with native crystals is collected, giving us the amplitudes $|F_P|$ for each of the native reflections. Then we find a heavy-atom derivative and collect a second data set at the same wavelength, giving amplitudes $|F_{PH}|$ for each of the reflections in the heavy-atom in the heavy-atom data. Next we collect a third data set at a different X-ray wavelength, chosen to maximize anomalous scattering by the heavy atom. The non-equivalence of Friedel pairs in the anomalous scattering data is then used to establish phases of reflections in the heavy-atom data while the phased heavy-atom derivative structure factors is used to establish the native phases. In this method (SIRAS), only one heavy atom data set is required. In practice, several of the most

commonly used heavy atoms (including mercury and platinum) give strong anomalous scattering with Cu-K_α radiation.

Three developments (third generation synchrotron X-rays, cryocrystallography and the production of proteins containing selenomethionine instead of the normal sulfur-containing methionine) have recently allowed rapid progress in maximizing the information obtainable from anomalous diffraction. Proteins lacking functional heavy atoms can be expressed in *E. coli* containing exclusively selenomethionine. The selenium atoms then serve as heavy atoms in a protein that is essentially identical to the "native" form. Isomorphism is not a problem with these proteins because the same protein serves as both native and derivative forms. The power of multiwavelength radiation (the MAD method, see below) is that data sets from a heavy-atom derivative at different wavelengths are in many respects like those from distinct heavy-atom derivatives. Data sets at different wavelengths from a single crystal can contain sufficient phasing information to solve a structure if the molecule under study contains one or more atoms (usually 1 % of peptide length) that give anomalous diffusion. This method is called multiple wavelength anomalous diffusion, or MAD phasing, and it has become one of the most widely used phasing methods (For details see Ealick 2000, Rhodes 2006).

MAD phasing has become a routinely used tool for determining new macromolecular structures. The MAD method has stringent data-collection requirements, typically necessitating radiation-resistant crystals and access to a tunable synchrotron beamline. The increasing availability of tunable synchrotron data-collection facilities prompted the use of SAD phasing in conjunction with a phase-ambiguity resolving method such as solvent flattening. In conjunction with solvent flattening, it has been shown that diffraction data from the peak anomalous

wavelength alone can produce interpretable electron-density maps of comparable quality to those resulting from full MAD phasing. SAD phasing can therefore be a time-efficient alternative to MAD. The data also show that radiation damage can have a significant effect on the quality of SAD/MAD diffraction data. These results may be useful in the design of optimal strategies for collection of the diffraction data (Rice et al. 2000).

1.1.7.4 Direct methods in protein crystallography

The phase problem of X-ray is defined as the problem of determining the phases, φ, of the structure factors, $F = |F| \exp(i\varphi)$, from measurements of intensities alone. The phase information, which is lost in the diffraction experiment, is in fact recoverable from the measurable intensities. The methods devised to achieve this goal are known as direct methods, a class of *ab initio* methods in which probabilistic phase relations are used to derived reflection phases (Hauptman 1986, Xu & Hauptman 2006). The methods generally exploit constraints or statistical correlations between the phases of different Fourier components that result from the fact that the scattering density must be a positive real number. In two dimensions, it is relatively easy to solve the phase problem directly, but not so in three dimensions. The key step was taken by Hauptman and Karle, for which they were awarded the 1985 Nobel prize in Chemistry. At present, direct methods are the preferred method for phasing crystals with high diffraction power of small molecules having up to 1000 atoms in the asymmetric unit. However, they are generally not feasible by themselves for larger molecules such as proteins (Uson & Sheldrick 1999, Hauptman 1997).

It is pointed out that the 'direct methods' of phase determination for small-structure crystallography do not have immediate applicability to macromolecular structures. The term 'direct methods in macromolecular crystallography' is suggested to categorize a spectrum of approaches to macromolecular structure determination in which the analyses are characterized by the use of two-phase and higher-order-phase invariants. The evaluation of the invariants is generally obtained by the use of heavy-atom techniques (Karle 1989). One form of direct phasing relies on the existence of mathematical relationships among certain combinations of phases. From these relationships, a sufficient number of initial phase estimates can be obtained to begin converging toward a complete set of phases. One such relationship, called a triplet relationship, relates the phases and indices of three reflections as follows:

$$\alpha_{hkl} + \alpha_{h'k'l'} \approx \alpha_{(h+h')(k+k')(l+l')}$$

In words, if the indices of the reflections on the left sum to the indices of the reflection on the right, then their phases sum approximately. As a specific example, the plane of reflection $\bar{1}20$ is always approximately the sum of the phases of reflections 110 and $\bar{2}10$. If the number of reflections is not too large, so that a large percentage of these relationships can be examined simultaneously, they might put enough constraints on phases to produce some initial estimates. For macromolecules, the number of reflections is far too large to make this method useful (Rhodes 2006).

Another form of direct phasing, executed by a program called *Shake-and-Bake*, in essence tries out random arrangements of atoms, simulates the diffraction patterns they would produce, and compares the simulated patterns with those obtained from the crystals. The widely used *Shake-and-Bake* algorithm has implemented in S*n*B (Miller et al. 1994, Weeks & Miller 1999) and B*n*P (Weeks et al. 2002) software and adapted in the *SHELXC/D/E* (Sheldrick 2006) and *PHENIX*

(Adams 2006) software. The procedure is based on a minimal function which is optimized through alternate cycles of reciprocal space phase refinement and real-space filtering (Ealick 1997, Hauptman 1986, Xu & Hauptman 2006). Even though the trial arrangements are limited to those that are physically possible (for example, having no two atoms closer than bonding or van der Waals forces allow), the number of trial arrangements can be too large for computation if the number of atoms is large. But this "try-everything" method is enormously powerful for any number of atoms that computation can handle and this number grows with the rapidly growing capacity of computers. Recently, a more powerful statistical *Shake-and-Bake* procedure (Xu & Hauptman 2006) has been implemented in the latest version of the computer programs S*n*B and B*n*P. These programs can be downloaded from the websites http://www.hwi.buffalo.edu/SnB/ or http://www.hwi.buffalo.edu/BnP/, respectively.

Direct methods work if the molecules, and thus the unit cells and numbers of reflections, are relatively small. Isomorphous replacement works if the molecules are large enough that a heavy atom does not disturb their structures significantly. The most difficult structures for crystallographers are those that are too large for direct methods and too small to remain isomorphous despite the intrusion of a heavy atom. If a medium-size protein naturally contains a heavy atom, like iron or zinc, or if a selenomethionine derivative can be produced, the structure can often be solved by MAD phasing. As the MAD method is much easier and faster to implement as compared to the isomorphous replacement, more and more researchers tend to damper the usage of the latter method, regardless the size of the protein.

1.1.7.5 Molecular replacement : related proteins as phasing models

As more protein structures are solved, the likelihood that current structural investigations will involve proteins for which there exists no homologous structure continually decreases. The extraction of phase information from diffraction experiments is one of several great barriers that crystallographers must overcome on the path to structure solution. One means to overcome this obstacle, the technique of molecular replacement, uses the structural similarity between proteins with similar sequences to give a good first estimate of the phases for the diffraction data of the protein of interest (Toth 2007). Homology in the amino acid sequence is an indication of whether a model is suitable. The solutions of the rotation and translation functions are not always found in a straightforward way. It can be necessary to modify the model, for instance, by ignoring side chains and deletions/additions in the model and to play with the resolution range of the X-ray data. With the rapid increase in the number of successful protein structure determinations, molecular replacement has become an extremely useful technique for protein phase angle determination.

The programs that execute this technique currently come in many flavors, from traditional Patterson-based methods, to stochastic searches in greater than three dimensions, to maximum likelihood-enhanced molecular replacement, each possessing unique advantages that can shake loose a recalcitrant solution (Toth 2007). In a nutshell, molecular replacement needs to find six parameters to define the position and orientation of a search model in the new unit cell. When the search model is rather different than the protein of interest it will not be able to give a good correspondence to the higher resolution data. So in practice we tend to only use the

low resolution data, say 4 to 10 Å. With only six parameters to determine we clearly don't need lots of data. Incompleteness due to randomly missed reflections is not a serious problem. Missing large chunks of reflections should however be avoided. Molecular replacement works in Patterson space, e.g. on intensities rather than amplitudes. Since the intensity is the square of the amplitude, molecular replacement is dominated by the strongest reflections. These are also the easiest ones to measure accurately so normally no special precautions are needed, except perhaps the need to prevent overloading strong low resolution reflections.

It is noteworthy to note that when combined with measured amplitudes, model-based phases yield electron-density maps with features of the correct structure but with a significant bias towards features of the model. Model bias is a very significant problem in macromolecular crystallography because it can lead to misinterpretation of electron-density maps, even when the maps are relatively accurate overall (Hodel et al. 1992, Adams et al. 1999, Kleywegt 2000). It can occur in many stages of crystallographic analysis and is particularly important in molecular replacement (Rossman 1972), model building (Perrakis et al. 1999), ligand-binding and conformation-change studies and structure validation (Kleywegt 2000). Methods for reducing model bias in electron-density map calculations have been developed (Hodel et al. 1992), but none are entirely satisfactory because each either removes only part of the effect of model bias or substantially increases the overall error in the maps. Recently, a map-probability function that can be used to calculate phase probabilities based only on the presence of such features in the electron-density map (Terwilliger 2001) and a “prime-and-switch” phasing technique (see Terwilliger 2004 for details) have been described to reduce model bias in molecular replacement.

1.1.7.6 Iterative improvement of phases

The phase problem greatly increases the effort required to obtain an interpretable electron-density map. In whichever methods of obtaining phases, the phases obtained are estimates, and often the set of estimates is incomplete. Calculated electron-density maps, using measured amplitudes and first phase estimates, are often difficult or impossible to interpret. However, as phase improvement and extension proceed, electron-density maps become clearer and easier to interpret as an image of a molecular model. The iterative process of structure refinement eventually leads to a structure that is in good agreement with the original data.

1.1.8 Obtaining and judging the molecular model

The final stages of structure determination involve the obtaining and improving of the electron-density map that is based on the first phase estimates, interpreting the map to produce an atomic model of the unit-cell contents, and refining the model to optimize its agreement with the original native reflection intensities.

1.1.8.1 Iterative improvement of maps and models

In brief, obtaining a detailed molecular model of the unit-cell contents entails calculating $\rho(x, y, z)$ from the second equation in section 1.1.7.1, using amplitudes ($|F_{hkl}|$) computed from measured intensities in the native data set and phases (α'_{hkl}) computed from heavy-atom data, anomalous scattering, or molecular replacement. Because the phases are rough estimates, the first map may be uninformative and

disappointing. Crystallographers improve the map by an iterative process in which the principle is the following: any features that can be reliably discerned in, or inferred from, the map become part of a phasing model for subsequent maps. Without the input of new information, the map will not improve. In another words, crystallographer cannot get any new phase information without modifying the model in some way. The possible modifications include solvent flattening, noncrystallographic symmetry averaging, or introducing a partial atomic model.

The model can be improved in another way: by refinement of the atomic coordinates. This method entails adjusting the atomic coordinates to improve the agreement between amplitudes calculated from the current model and the original measured amplitudes in the native data set. In the latter stages of structure determination, the crystallographer alternates between map interpretation and refinement of the coordinates. At its heart, refinement is an attempt to minimize the differences between (a) measured diffraction intensities and (b) intensities calculated by the current model, which, in intermediate stages, is incomplete and harbors errors that will eventually be removed.

1.1.8.2 First maps

Today's computers can display the equivalent of a minimap and allow more informative first glimpses of the electron density and the contour level of the map can be adjusted for maximum detail. If the phase estimates are sufficiently good, the map will show some of the gross features of the unit cell contents. Contrary, at the worst, the first map is singularly uninformative, signaling the need for additional phasing information, perhaps from another heavy-atom derivative. Usually the

minimum result that promises a structure from the existing data is that protein be distinguishable from bulk water, and that dense features like alpha helices can be recognized. If the boundary of each molecule, the molecular envelope, shows some evidence of recognizable protein structure, then a full structure is likely to come forth.

The crude molecular image seen in the F_0 map, which is obtained from the original indexed intensity data ($|F_{obs}|$) and the first phase estimates (α'_{calc}), serves now as a model of the desired structure. A crude electron density function is calculated to describe the unit-cell contents as well as they can be observed in the first map. Then the function is modified to make it more realistic in the light of known properties of proteins and water in crystals. This map improving process is called, depending on the exact details of procedure, density modification, solvent leveling (determine the molecular envelope using the reciprocal-space equivalent), solvent flattening (by drawing a "mask" around the molecule, a low-resolution model of the structure can be created and the density peaks outside of it is canceled to improve the quality of the phases and the electron density), or solvent flipping (the density corresponding to solvent is inverted to enhance contrast between solvent and protein and to counteract bias in the model) (Rhodes 2006).

1.1.8.3 The model becomes molecular

At some critical point in the iterative improvement of phases, the map becomes clear enough that we can trace the protein chain through it. In the worst circumstances at this stage, we may only be able to see some continuous tubes of density. We can skeletonize the map by drawing line segments along lines of maximum density. The

positioning of these rough lines along which the molecular chain is likely to lie can somehow help to locate both the main chain and the branch points of side chains. In a clearer map on the other hand, we may be able to recognize alpha helices, one of the densest features of a protein, or sheets of beta structure. Now, a partial molecular model (as opposed to an electron-density model) of the protein can be constructed using computer graphics to build and manipulate a stick model of the known sequence within small sections of the map. This technique is called map fitting. Then, new structure factors are calculated from the resulting, usually crude, molecular model (which may harbor many errors and undefined regions) rather than from an approximation of $\rho(x, y, z)$. Additional iterations may improve the map further, allowing more features to be constructed therein.

During the density modification, we are actually improving the resolution of the model by making it structurally realistic: which means assigning atoms and connecting them in proper geometries that we know must be correct. Successive models give us phases for reflections at higher and higher resolution. Electron-density maps computed from these phases and, as always, the original native amplitudes $|F_{obs}|$ become more and more detailed. As the map becomes clearer, even more specific prior knowledge of protein structure, can be applied. For example, in pleated sheets, successive carbonyl C=O bonds point in opposite directions. Side chains in pleated sheets also alternate directions, and are roughly perpendicular to the carbonyls. Near the ends of pleated sheet strands, if electron density is weakening, this knowledge of pleated-sheet geometry can often guide the placement and orientation of extra residues that are not as well-defined by the current map (Rhodes 2006).

1.1.8.4 Structure refinement

During structure refinement we use our data as observations to fit our model against. An important factor in refinement is the observation/parameter ratio. The larger the ratio is, the more stable the refinement is and the lesser the problems with model bias will exist. Large numbers of observations also allow you to refine a larger number of parameters. The best way to get a large number of observations is to obtain high resolution data. High resolution reflections are also sensitive to small errors in atomic parameters and therefore allow precise positioning during refinement. A reduced completeness clearly leads to fewer reflections to refine against. Since errors in F_{calc} tend to be significantly larger than errors in F_{obs}, measurement accuracy is not as critical.

Similar to density maps, resolution and completeness are more important than accuracy. There is, however, a difference. For the electron density map, weaker reflections are not as important as stronger reflections. However, during refinement we minimize the difference between observed and calculated structure factors. For least squares the target function Φ is:

$$\Phi = \Sigma_{hkl}\, w_{hkl}\, (|F_{obs}\,(hkl)| - |F_{calc}\,(hkl)|)^2.$$

The sum is taken over all reflections *hkl* currently in use. Each difference is weighted by the term w_{hkl}, a number that depends on the reliability of the corresponding measured intensity.

So, in refinement the comparison of the observed and calculated structure factors is the one which should be taken into account but not the absolute value of each structure factor. Therefore, a weak reflection can definitely "tell" the refinement program that the model needs adjustment; namely if F_{calc} is large. Hence for

refinement, the more the merrier and every reflection that can lay the hands on should be collected.

1.1.8.5 Convergence to a final model

In the last stages of structure determination, the crystallographer alternates computed, reciprocal-space refinement with map fitting, or real-space refinement. The most powerful software packages can start with the first map that shows any sort of structural details, and cycle between automated map fitting and refinement cycles until the most of the model is complete. The progress of iterative real- and reciprocal-space refinement is monitored by comparing the measured structure-factor amplitudes $|F_{obs}|$ (which are proportional to $(I_{obs})^{1/2}$) with amplitudes $|F_{calc}|$ from the current model. In calculating the new phases at each stage, we learn what intensities our current model, if correct, would yield. As we converge to the correct structure, the measured Fs and the calculated Fs should also converge. The most widely used measure of convergence is the residual index, or *R*-factor:

$$R = \frac{\Sigma \, ||F_{obs}| - |F_{calc}||}{\Sigma \, |F_{obs}|} .$$

In this expression, each $|F_{obs}|$ is derived from measured reflection intensity and each $|F_{calc}|$ is the amplitude of the corresponding structure factor calculated from the current model. Values of R range from zero, for perfect agreement of calculated and observed intensities, to about 0.6, the R-factor obtained when a set of measured amplitudes is compared with a set of random amplitudes. Generally, an R-factor greater than 0.5 implies that agreement between observed and calculated intensities is very poor and many models with R of 0.5 or greater will not respond to attempts at improvement unless more data are available. An early model with R near 0.4 is

promising and is likely to improve with the various refinement methods available. A desirable target R-factor for a protein model refined with data to 2.5 Å, for instance, is 0.2. Very rarely, small, well-ordered proteins may refine to R-values as low as 0.1, whereas small organic molecules commonly refine to R values below 0.05.

A more demanding and revealing criterion of model quality and of improvements during refinement is the free R-factor, R_{free}. R_{free} is computed with a small set of randomly chosen intensities, the "test set", which are set aside from the beginning and not used during refinement. They are used only in cross-validation, a quality control process that entails assessing the agreement between calculated (from the model) and observed data. At any stage in refinement, R_{free} measures how well the current atomic model predicts a subset of the measured intensities that were not included in the refinement, whereas R measures how well the current model predicts the entire data set that produced the model. A sort of circularity in R that is avoided in R_{free} can be seen. Many crystallographers believe that R_{free} gives a better and less-biased measure of overall model. In Bayesian methods, R_{free} is replaced by free log-likelihood gain, L_{free}, calculated over the same test data set as R_{free}.

A stereochemically reasonable model has no inverted centers of chirality (for instance, no D-amino acids). A conformationally reasonable model meets several criteria. (1) Peptide bonds are nearly planar, and nonproline peptide bonds are trans-, except where obvious local conformational constraints produce an occasional cis-peptide bond, which is usually followed by proline. (2) The backbone conformational angles Φ and Ψ fall in allowed ranges, as judged from Ramachandran plots of these angles. And finally, (3) torsional angles at single bonds in side chains lie within a few degrees of stable, staggered conformations. During the progress of refinement, all of these structural parameters should continually improve. And unlike

bond lengths and bond angles, conformational angles are usually not restrained during refinement, so these parameters are better indicators of model quality.

1.1.8.6 Judging the quality and usefulness of the refined model

1.1.8.6.1 Structural parameters

Crystallographers monitor the decrease in some type of R-factor (or the increase in a likelihood factor) as an indicator of convergence to a final, refined model, with a general target for R of 0.20 or lower for proteins, and adequate additional cycles of refinement to confirm that *R* is not still declining. In addition, various constraints and restraints are relaxed during refinement, and after these restricted values are allowed to refine freely, they should remain in, or converge to, reasonable values. Among these are the r.m.s.d. of the model's bond lengths, angles, and conformational angles from an accepted set of values based upon the geometry of small organic molecules. A refined model should exhibit r.m.s.d. of no more than 0.008 Å for bond lengths and 4° for bond angles. These values are routinely calculated during refinement to be sure that all is going well. Because they are restrained in earlier stages, they are not as valuable as quality indicators as parameters that are allowed to vary freely throughout the refinement. Spectroscopic evidence abundantly implies that peptide bonds are planar, and some refinements constrain peptide geometry. If unconstrained, peptide bonds should settle down to within a few degrees of planar.

Apart from peptide bonds, the other backbone conformational angles are Φ, along the N – C_α bond and Ψ, along the C_α – C bond. Φ is the torsional angle of the N – C_α bond, defined by the atoms C – N – C_α – C (C is the carbonyl carbon) and Ψ

is the torsional angle of the C_α – C bond, defined by the atoms N – C_α – C – N. Φ = Ψ = 180°. Model studies show that, for each amino acid, the pair of angles Φ and Ψ is greatly restricted by steric repulsion. The allowed pairs of values are depicted on a Ramachandran diagram. The shapes of allowed regions are slightly different for each of the 20 common amino acids. During the final stages of map fitting and crystallographic refinement, Ramachandran diagrams are a great aid in spotting conformationally unrealistic regions (any residues that lie in high-energy areas or the disallowed regions) of the model. The conformations of amino-acid side chains are unrestrained during refinement. Nevertheless, in well-refined models, side-chain single bonds end up in staggered conformations (commonly referred to as rotamers).

1.1.8.6.2 Resolution and precision of atomic positions

Atomic resolution macromolecular crystallography has become a powerful and versatile tool in structural biology as we can see that the number of atomic resolution structures is steadily increasing. One major reason is the growing number of available beamlines at synchrotron radiation facilities. Crossing the border from medium resolution to atomic resolution often means crossing over from hypothesis to fact, from (theoretical) models to observations and, finally, from structure to function (Schmidt & Lamzin 2002). Atomic resolution is defined as crystallographic data extending to a resolution higher than 1.2 Ǻ, with at least 50% of the theoretically possible diffraction intensities in the highest resolution shell exceeding twice the value of their estimated standard uncertainty (Sheldrick 1990). X-ray data to 0.8 Ǻ set the limit for ultrahigh or subatomic resolution, at which details finer than the roughly spherical shape of individual atoms become visible.

The benefits of atomic and ultrahigh resolution have been outlined (Dauter et al. 1997, Longhi et al. 1998, Schmidt & Lamzin 2002). One of the most rewarding is that an atomic resolution structure can be refined with only weak stereochemical restraints. The stereochemical characteristics of the resulting structure are essentially unbiased towards idealized library values. With a high data to parameter ratio, which, at 1.2 Å resolution, reaches about 50 reflections per atom, the refinement is well-conditioned numerically. The standard uncertainties for each of the refined parameters (e.g. atomic coordinates) can, in principle, be deduced directly from the inversion of the normal matrix.

Determination of atomic resolution structures of complexes that represent various stages of the catalytic reaction can help in mapping the complete reaction path and understanding the catalytic mechanism. The high accuracy of the atomic positions allows reliable interpretation of ligand-induced structural changes, even if they seem small (Schmidt & Lamzin 2007). This was beautifully demonstrated in a study of horseradish peroxidase, where the X-ray radiation during the diffraction data collection was used to create various binding states of oxygen, hydrogen peroxide and water to the haem group, thus mapping the catalytic pathway (Berglund et al. 2002). Obviously, atom types can be distinguished in the atomic resolution electron density and so mismodelling, for example, the incorrect orientation of histidine, glutamine or asparagine side-chain, can be largely avoided or easily corrected during the refinement. In a protein crystal, solvent accounts for commonly 30 – 70 % of the unit cell content. Crystals diffracting to atomic resolution, solvent molecules can be placed accurately within the well-defined ordered shells. Not only can more well-separated solvent molecules be detected, but also disorder can be accounted for and

relative occupancies refined, which enhances the overall accuracy of the model (Schmidt & Lamzin 2002).

The determination of an atomic resolution protein structure always creates new challenges. In parallel to the increase in the number of structures, new methods are constantly being developed to even more benefit from atomic resolution structures. Structure analysis and validation have shifted away from the classical Ramachandran plot towards analysis of the deviation of main-chain parameters from idealized library values (Ridder et al. 1999, Esposito et al. 2000), mapping of hydrogen-bonding contacts (Esposito et al. 2000, Wang et al. 1997, Heikinheimo et al. 2001), direct identification of atom types (Schmidt et al. 2002) and exploiting the information stored in the anisotropic atomic displacement parameters (Harata & Kanai 2002, Longhi et al. 1997). This rapidly growing number of atomic resolution protein structures and the advancement of new combination methods of protein crystallography with complementary techniques (Schmidt & Lamzin 2002) will certainly promote further interest in atomic resolution structural biology, as fine structural details, invisible even in the recent past, now become attainable reality.

1.1.8.6.3 Vibration and disorder

Thermal motion and disorder are two important physical reasons for uncertainty in atom positions. Thermal motion refers to vibration of an atom about its rest position. Disorder refers to atoms or groups of atoms that do not occupy the same position in every unit cell, in every asymmetric unit, or in every molecule within an asymmetric unit (Rhodes 2006). The temperature factor B_j obtained during refinement reflects

both the thermal motion and the disorder of atom *j*, making it difficult to sort out these two sources of uncertainty.

Occupancies n_j for atoms of the protein (but not necessarily its ligands, which may be present at lower occupancies) are usually constrained at 1.0 early in refinement, and in many refinements are never released, so that both thermal motion and disorder show their effects upon the final *B* values. In some cases, after refinement converges, a few *B* values fall far outside the average range for the model. This is sometimes an indication of disorder. Careful examination of sigma-A weighted maps or comparison of $2F_o$-F_c and F_o-F_c maps may give evidence for more than one conformation in such a troublesome region. If so, inclusion of multiple conformations followed by refinement of their occupancies may improve the *R*-factor and the map, revealing the nature of the disorder or the flexibility of concerned region.

1.1.8.6.4 Disordered regions and unexplained density

Occasionally, portions of the known sequence of a protein are never found in the electron-density maps, presumably because the region is highly disordered or in motion, and thus invisible on the time scale of crystallography. The usual procedure is simply to omit these residues from the deposited model. It is not at all uncommon for residues at termini, especially the N-terminus, to be missing from a model. Discussions on these structure-specific problems are included in a thorough refinement paper, and lists of missing residues are provided in PDB header.

Electron-density maps sometimes show clear, empty density after all known contents of the crystal have been located. Apparent density can appear as an artifact

of missing Fourier terms, but this density disappears when a more complete set of data is obtained. Among the possible explanations for density that is not artifactual are ions like phosphate and sulfate from the mother liquor; reagents like mercaptoethanol, dithiothreitol, or detergents used in purification or crystallization; or cofactors, inhibitors, allosteric effectors, or other small molecules that survived the protein purification. Later discovery of previously unknown but important ligands has sometimes resulted in subsequent interpretation of empty density.

1.1.8.6.5 Distortions due to crystal packing

Packing effects may be evident in the model itself and should be mentioned in the refinement papers and PDB headers. For example, packing may induce slight differences between what are otherwise expected to be identical subunits within an asymmetric unit. Examination of the neighborhood around such differences may reveal that intermolecular contact is a possible cause. In areas where subunits come into direct contact or close contact through intervening water, surface temperature factors are usually lower than at other surface regions.

1.1.8.6.6 Functional unit versus asymmetric unit

The symmetry of functional macromolecular complexes in solution is sometimes important to understanding their functions. As mentioned before, in the unit cell of a crystal, the largest aggregate of molecules that possesses no symmetry elements, but can be juxtaposed on other identical entities by symmetry operations, is called the crystallographic asymmetric unit. Users of models should be careful to distinguish

the asymmetric unit from the functional unit, which the PDB currently calls the "biological molecule". For example, the functional unit of mammalian hemoglobin is a complex of four subunits, two each of two slightly different polypeptides, called α and β. Hemoglobin functions as an $\alpha_2\beta_2$ tetramer. In some hemoglobin crystals, the twofold rotational symmetry axis of the tetramer corresponds to a unit-cell symmetry axis, and the asymmetric unit is a single αβ dimer. In other cases, the asymmetric unit may contain more than one biological unit. Therefore, it is safe to think of the non-crystallographic symmetry as not necessarily having anything to do with crystallographic symmetry.

1.2 Oxidative stress in organisms

All aerobic organisms are exposed to reactive oxygen species (ROS), such as H_2O_2, the superoxide anion and the hydroxyl radical, during the course of normal aerobic metabolism or after exposure to radical-generating compounds. Such ROS can cause wide-ranging damage to macromolecules and have been recognized for some time as being associated with various disease processes including cancer, ageing and neurodegenerative disorders (Halliwell and Gutteridge 1989, Gutteridge 1993). To protect against oxidant damage, cells contain effective defense mechanisms including enzymes, such as catalase, superoxide dismutase and glutathione peroxidase, and antioxidants, such as glutathione (GSH) and vitamins C and E (Yu 1994).

1.3 Overview of ROS in plants

1.3.1 ROS: friend or foe?

Ever since the introduction of molecular oxygen (O_2) into our atmosphere, by O_2-evolving photosynthetic organisms, about two billion years ago, reactive oxygen species (ROS) have been the unwelcome companions of aerobic metabolism. In contrast to O_2, these partially reduced or activated derivatives of oxygen (O_2^1, O_2^-, H_2O_2, and HO·) are highly reactive and toxic, and can lead to the oxidative destruction of cells. Consequently, the evolution of all aerobic organisms has been dependent upon the development of efficient ROS-scavenging mechanisms. In recent years a new role for ROS was identified: the control and regulation of biological processes such as programmed cell death, hormonal signaling, stress responses, and development. These studies extend our understanding of ROS and suggest a dual role for ROS in plant biology: (a) toxic byproducts of aerobic metabolism and (b) key regulators of metabolic and defense pathways.

1.3.2 ROS cause oxidative damages

1.3.2.1 Production of ROS and scavenging of ROS in cells

The steady state level of ROS in the different cellular compartments is determined by interplay between multiple ROS-producing pathways, and ROS-scavenging mechanisms. These are controlled by the ROS-signal transduction pathway and constitute the “basic ROS cycle”. During normal growth and development this

pathway monitors the level of ROS, produced by aerobic metabolism, and controls the expression and activity of ROS-scavenging pathways (Figure 1.12). The basic ROS cycle may also perform fine metabolic tuning, e.g., suppression of photosynthesis, to reduce the production rate of ROS. There are many potential sources of ROS in plants (Table 1.5). Some are reactions of normal aerobic metabolism, such as photosynthesis and respiration, while others belong to pathways enhanced during abiotic stresses, such as photorespiration. In recent years new sources of ROS were identified in plants, including NADPH oxidases, amine oxidases, and cell wall-bound peroxidases. These are tightly regulated and participate in the control of processes such as programmed cell death, stress response, and pathogen defense.

Table 1.5 Mechanisms for production and scavenging of ROS in plant cells. AOX, alternative oxidase; APX, ascorbate peroxidase; CAT, catalase; GPX, glutathione peroxidase; SOD, superoxide dismutase.

	Localization	**Primary ROS**
Production of ROS		
Photosynthesis ET and PSI or II	Chl	$O_2^{\cdot}$
Respiration	Mit	$O_2^{\cdot}$
Glycolate oxidase	Per	H_2O_2
Excited chlorophyll	Chl	O_2^{1}
NADPH oxidase	PM	$O_2^{\cdot}$
Fatty acid β-oxidation	Per	H_2O_2
Oxalate oxidase	Apo	H_2O_2
Xanthine oxidase	Per	$O_2^{\cdot}$
Peroxidases, Mn^{2+} and NADH	CW	H_2O_2, $O_2^{\cdot}$
Amine oxidase	Apo	H_2O_2
Scavenging of ROS		
SOD	Chl, Cyt, Mit, Per, Apo	$O_2^{\cdot}$

APX	Chl, Cyt, Mit, Per, Apo	H_2O_2
CAT	Per	H_2O_2
GPX	Cyt	H_2O_2, ROOH
Peroxidases	CW, Cyt, Vac	H_2O_2
Thioredoxin peroxidase	Chl, Cyt, Mit	H_2O_2
Ascorbic acid	Chl, Cyt, Mit, Per, Apo	H_2O_2, $O_2^{\cdot}$
Glutathione	Chl, Cyt, Mit, Per, Apo	H_2O_2
α-Tocopherol	Membranes	ROOH, O_2^1
Caretenoids	Chl	O_2^1
AOX	Chl, Mit	$O_2^{\cdot}$

Abbreviations : Apo, apoplast ; Chl, chloroplast ; CW, cell wall; Cyt, cytosol; ET, electron transport; Mit, mitochondria; O_2^1, singlet oxygen; Per, peroxisome; PM, plasma membrane; PS, photosystem; Vac, vacuole.

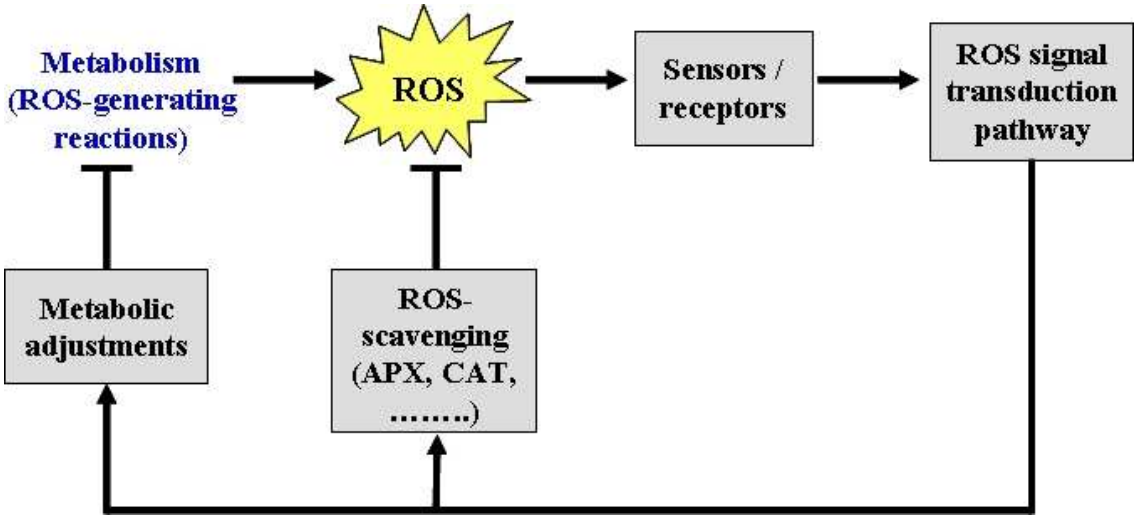


Figure 1.12 The basic ROS cycle. This cycle modulates the cellular levels of ROS during normal metabolism. Some of the key ROS scavenging enzymes of plants, ascorbate peroxidase (APX) and catalase (CAT) are indicated (Figure adapted from Pitzcheke et al. 2006).

Under optimal growth conditions the production of ROS in cells is estimated at a constant rate of 240 μM s^{-1} O_2^-, and a steady state level of 0.5 μM H_2O_2.

However, stresses that disrupt the cellular homeostasis of cells result in the enhanced production of ROS (up to 720 μM s^{-1} O_2^-, and a steady state level of 5-15 μM H_2O_2). These include drought and desiccation, salt, chilling, heat shock, heavy metals, UV radiation, air pollutants such as ozone and SO_2, mechanical stress, nutrient deprivation, pathogen attack, and high light. The enhanced production of ROS during stress can pose a threat to cell, and many stress conditions enhance the expression of ROS-scavenging enzymes. However, it is also thought that during stress ROS are actively produced by cells (e.g., by NADPH oxidase), and act as signals for the induction of defense pathways. Thus, ROS may be viewed as cellular byproducts of stress metabolism, as well as secondary messengers involved in the stress-response signal transduction pathway. This view, of the "extended ROS cycle", is presented in Figure 1.13.

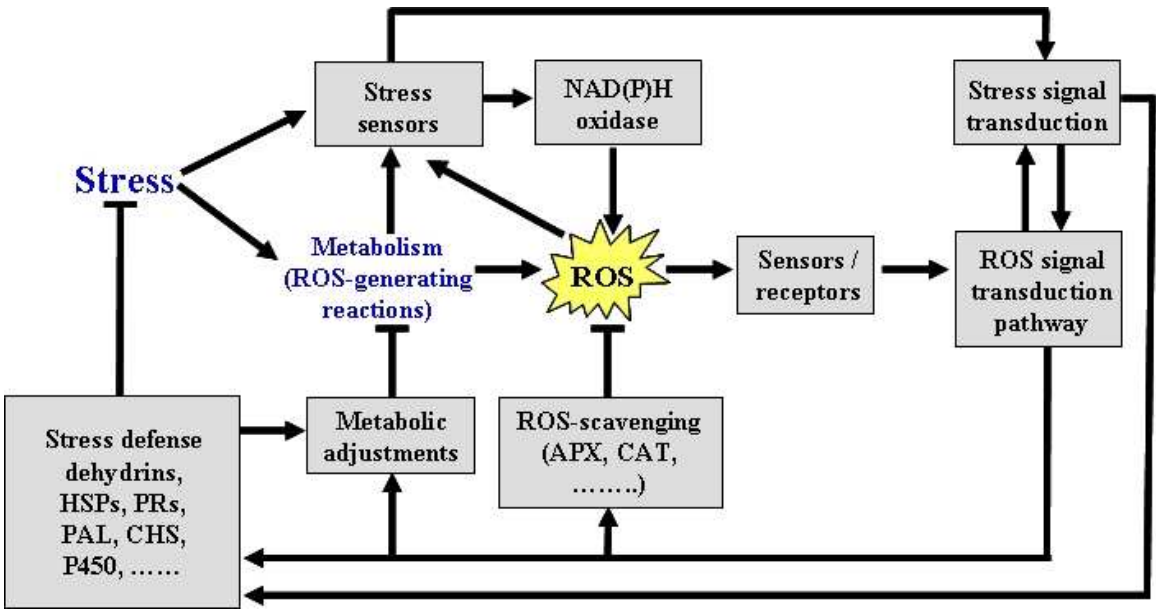


Figure 1.13 The extended ROS cycle. This cycle operates in plants during biotic and abiotic stresses. HSPs, heat shock proteins; PR, pathogenesis related proteins; PAL, phenylalanine ammonia-lyaze; CHS, chalcone synthase; P450, cytochrome P450 (Figure adapted from Pitzcheke et al. 2006).

1.3.2.2 Detoxification system of ROS

1.3.2.2.1 Non-enzymatic reduction system

Major non-enzymatic antioxidant compounds including carotenoids, tocopherols, acid lipoic, caratenoids, flavonoids, alkaloids, reduced glutathione (GSH) and ascorbate are found in plants to fight against the oxidative damage caused by ROS. Ascorbate and GSH are major cellular redox buffers and they are found at very high concentrations in chloroplasts and other cellular compartments (5-20 mM ascorbic acid and 1-5 mM GSH). They are also important for the defense of plants against oxidative stress. Consequently, mutants with suppressed ascorbic acid levels, and transgenic plants with suppressed ROS-scavenging enzymes, are hypersensitive to pathogen attack and abiotic stress conditions (Pitzschke et al. 2006).

1.3.2.2.2 Enzymatic reduction systems

Because ROS are toxic but also participate in key signaling events, plant cells require different mechanisms to regulate their intracellular ROS concentrations by scavenging of ROS. Major ROS scavenging mechanisms of plants are summarized in Table 1.5. Enzymatic ROS scavenging mechanisms in plants include superoxide dismutase (SOD), ascorbate peroxidase (APX), glutathione peroxidase (Gpx), catalase (CAT), thioredoxin reductase (TrxR), glutaredoxin reductase (GrxR), sulfiredoxin (Srx), peroxiredoxin (Prx), thioredoxin (Trx) and glutaredoxin (GRX). SODs act as the first line of defense against ROS, dismutating superoxide to H_2O_2. APX, Gpx, Prx, Srx and CAT subsequently detoxify H_2O_2 while Trx, Grx and their

reducing partners complete the regeneration cycle of the H_2O_2 detoxifying enzymes mentioned above. Unlike most other organisms, plants have multiple genes encoding ROS scavenging enzyme isoforms that are specifically targeted to chloroplasts, mitochondria, peroxisomes, as well as to the cytosol and apoplast (Pitzschke et al. 2006). The balance between SOD and APX (and/or CAT) activity in cells is considered to be crucial for determining the steady state level of O_2^- and H_2O_2. This balance, together with sequestering of metal ions such as Fe and Cu by ferritin and copper-binding proteins, is thought to be important to prevent the formation of the highly toxic HO· via the metal-dependent Haber-Weiss or the Fenton reactions. Over-expression of ROS-scavenging enzymes increases the tolerance of plants to abiotic stresses. ROS production can also be decreased in cells by the alternative channeling of electrons in the electron transport chains of the chloroplasts and mitochondria by a group of enzymes called alternative oxidases.

1.3.3 ROS signaling in plants: efficient secondary messengers

The cellular thiol redox state is a crucial mediator of multiple metabolic, signaling and transcriptional processes in cells and a fine balance between oxidizing and reducing conditions is essential for the normal function and survival of cells. ROS, particularly H_2O_2, are now known to be important second messengers in intracellular signaling. With respect to proteins, the thiol group of cysteinyl side chains is susceptible to a number of oxidative modifications, for instance, the formation of inter- or intramolecular disulfides between protein thiols [protein (P)-S-S-P] or between protein thiols and low-molecular-weight thiols such as glutathione (P-S-SG), the oxidation to sulfenic (P-SOH), sulfinic ($P\text{-}SO_2H$) and sulfonic ($P\text{-}SO_3H$)

acid and S-nitrosylation (P-S-NO) (Berndt et al. 2007). These modifications can alter the function of numerous proteins that contain cysteines of structural importance, within their catalytic centers or as part of protein-protein interaction interfaces. To a greater extent, the redox state of these cysteinyl residues is counterbalanced by the Trx and Grx systems. Besides cysteine, methionine is also very sensitive to oxidation by H_2O_2 within the protein structure (Jones et al. 2004). Oxidation of methionine is another reversible modification. An interesting example is the small heat shock protein in chloroplasts, which is inactivated by methionine sulfoxidation, but reactivated by reduction catalysed by the enzyme peptide methionine sulfoxide reductase (Gustavsson et al. 2002). These reversible oxidations of sulfur-containing side-chains are an important mechanism for regulating enzyme activity thus acting as a stress indicator (Møller & Kristensen 2004). To a lesser extent or rather a selective process, oxidative post-translational modification of tryptophan residue has also been reported to sense the oxidative environment in plant cells (Møller & Kristensen 2006).

1.3.3.1 Modulation of gene expression

All cells sense and respond to their environment. Often this function is performed using ligand receptors, which trigger the activation of cellular kinases that in turn result in changes in enzymatic activity or in the induction of transcription factors. Cells must also sense the presence of free radicals and respond appropriately to them. Plants, like other organisms, do respond to reactive oxygen intermediates in their environment (McCurbrey & Franklin 2006). Plant thioredoxin genes reveal distinct expression patterns as a function of organ type and stress conditions. For instance, *A.*

thaliana Trx *h*8 and *h*5 genes are strongly induced by biotic and abiotic stress and most genes encoding plastidic Trxs are preferentially expressed in leaves in response to variations in light intensity and quality, which probably indicates a function in the redox balance related to photosynthetic metabolism (Reichheld et al. 2002, Collin et al. 2004). The chloroplast is the main site of ROS production in plant cells because of photosynthetic activity, particularly during environmental constraints (Mittler et al. 2004).

In several plant species, Trx gene expression is often associated with physiological contexts characterized by increased ROS levels. For example in potato, severe abiotic constraints and agents that directly generate a high level of ROS production, such as methyl viologen, result in the increased abundance of a peculiar two-module thioredoxin termed CDSP32 (Chloroplastic Drought-induced Stress Protein) (Rey et al. 1998, Broin et al. 2000). Methyl viologen also induces an increase in Trx *h* transcript abundance in rice seedlings (Tsukamoto et al. 2005). Trx gene expression can be controlled at a transcriptional step as demonstrated by the presence of antioxidant responsive *cis*-elements in several promoters of Trx *h* genes (Lee et al. 2005, Tsukamoto et al. 2005). Interestingly, rice Trx *h*, GRX and SOD gene promoters contain a novel *cis*-element regulating expression in response to methyl viologen (Tsukamoto et al. 2005). This finding reveals the coordinated participation of disulfide reductases and superoxide dismutase in defense mechanisms against high ROS levels.

1.3.3.2 Influence signal transduction pathways

One mechanism by which cells will respond to reactive oxygen intermediates is to modify the activity of certain cellular signaling pathways. Recent studies identified a number of components that may be involved in the ROS signal transduction of plants (Pitzcheke et al. 2006, McCurbrey et al. 2006, McCurbrey & Franklin 2006). These include the MAPKKK, AtANP1 (also NPK1), the MAPKs, AtMPK3/6, and Ntp46MAPK, and calmodulin. A hypothetical model depicting some of the players involved in this pathway is shown in Figure 1.14 (Pitzcheke et al. 2006). A sensor that might be a two component Histidine-kinase, or a receptor-like protein kinase, is thought to sense H_2O_2. Calmodulin and a MAPK cascade are then activated resulting in the induction/ activation/ suppression of a number of transcription factors. These regulate the response of plants to oxidative stress. Cross-talk with the pathogen-response signal transduction pathway (gene-for-gene) also occurs and may involve interactions between different MAPK pathways, feedback loops, and the action of salicylic acid and nitric oxide.

Pitzcheke et al. (2006) review the effects of the reactive oxygen intermediates on cellular signaling in plants, where they mediate responses such as stromal closure, root hair development, and pathogen defense. The majority of the ROS in plants are produced in the peroxisomes and chloroplasts. In pathogen defense, ROS appear to activate signaling pathways that lead to apoptosis or cell death much as they do in animal cells. The apoptotic response is thought to prevent the spread of the pathogen and when the respiratory burst oxidase homolog (*rboh*) gene is knocked out in *A. thaliana*; increases in ROS and apoptosis are noted in response to bacterial challenge (Torres et al. 2002).

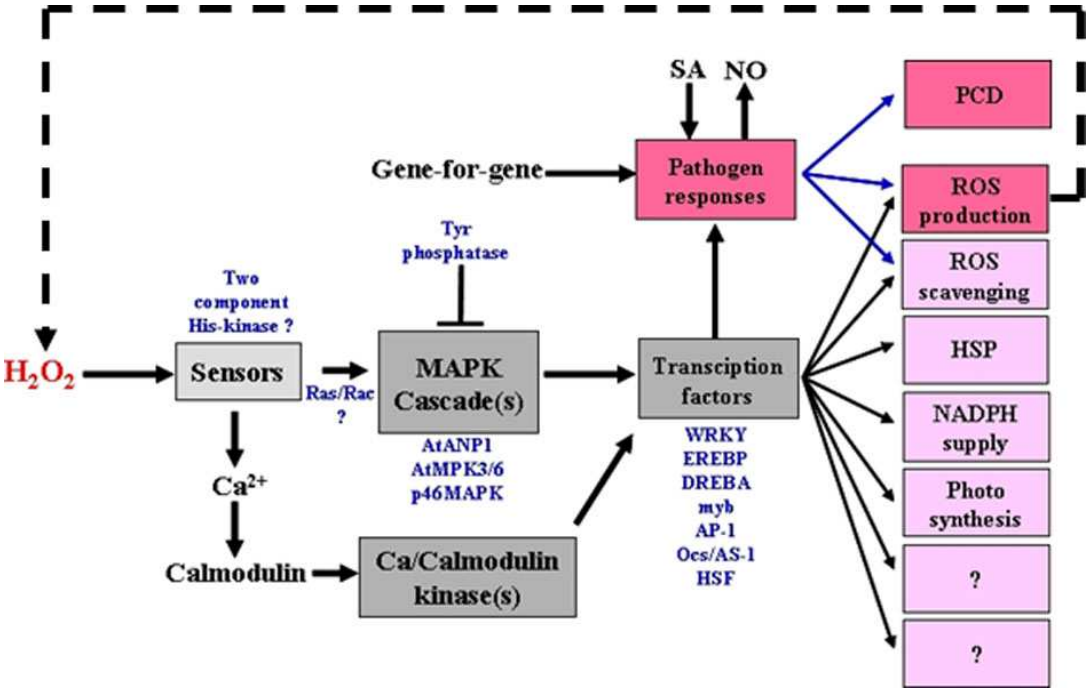


Figure 1.14 A hypothetical model of the signaling pathway activated in plants in response to external application of oxidants. SA, salicylic acid; NO, nitric oxide; PCD, programmed cell death; HSP, heat shock protein; MAPK, mitogen-activated protein kinase; Tyr, tyrosine; HSF, heat shock transcription factor (Figure adapted from Pitzcheke et al. 2006).

1.3.4 Thioredoxin fold

1.3.4.1 Overview on protein folding

Origami is the Japanese art of folding a piece of paper into complex shapes and forms. Much like origami of paper, nature (with energetic driven) has taken polypeptides and also folded them into very complicated structures. Unlike folded paper, proteins are capable of performing countless numbers of reactions, essential to life. Nature currently uses numerous folds. Extrapolating back to the origin of life, it seems likely that all proteins were derived from one or a very limited number of primordial protein folds. These folds then diverged from each other so much that

there is often no detectable resemblance that can be found between protein families (Pan & Bardwell 2006).

1.3.4.1.1 The relationship between folding and amino acid sequence

The amino acid sequence (or primary structure) of a protein predisposes it towards its native conformation. It will fold spontaneously during or after synthesis, with mechanism depends equally on the characteristics of the cytosol, including the nature of the primary solvent (water or lipid), the concentration of salts, the temperature and molecular chaperones.

Most folded proteins have a hydrophobic core in which side chains packing stabilizes the folded state, and charged or polar side chains on the solvent-exposed surface where they interact with surrounding water molecules. It is generally accepted that minimizing the number of hydrophobic side-chains exposed to water is the principal driving force behind the folding process (Pace et al. 1996), although a recent theory has been proposed which reassesses the contributions made by hydrogen bonding (Rose et al. 2006).

For the most part, scientists have been able to study many identical molecules folding together *en masse*. At the coarsest level, it appears that in transitioning to the native state, a given amino acid sequence takes on roughly the same route and proceed through roughly the same intermediates and transition states. Often folding involves first the establishment of regular secondary and supersecondary structures, particularly α-helices and β-sheets, and afterwards tertiary structure. Formation of quaternary structure usually involves the “assembly” or “coassembly” of subunits that have already folded. The regular α-helix and β-strand structures fold rapidly

because they are stabilized by intramolecular hydrogen bonds, as was first characterized by Linus Pauling (Pauling et al. 1951). Protein folding may involve covalent bonding in the form of disulfide bridges formed between two cysteine residues or the formation of metal clusters. Shortly before settling into their more energetically favourable native conformation, molecules may pass through an intermediate “molten globule” state.

The essential fact of folding, however, remains that the amino acid sequence of each protein contains the information that specifies both the native structure and the pathway to attain that state. This is not to say that identical amino acid sequences always fold similarly (Alexander et al. 2007). Conformations differ based on environmental factors as well; similar proteins fold differently based on where they are found. Folding is a spontaneous process independent of energy inputs from nucleoside triphosphates. The passage of the folded state is mainly guided by hydrophobic interactions, formation of intramolecular hydrogen bonds, and van der Waals forces, and it is opposed by conformational entropy, which must be overcome by extrinsic factors such as chaperones (not all the time).

1.3.4.1.2 Different types of protein folding

All protein molecules are heterogeneous unbranched chains of amino acids. By coiling and folding into a specific three-dimensional shape, they are able to perform their biological function. Many protein structures have now been determined and reveal that protein molecules can adopt the same fold despite having very different sequences. It has been suggested that, owing to different stereochemical constraints, the number of ways that a sequence can fold may be limited (Wang 1996). In

general, protein folding can be classified into three main structural groups: α structures, where the core is built up exclusively from α-helices; β structures, which comprise antiparallel β-sheets; and α/β structures, where combinations of βαβ motifs form a predominantly parallel β-sheet surrounded by α helices (Branden & Tooze 1999). Only a few common yet important examples of these folds have been illustrated in greater details here (see below).

1.3.4.1.2.1 The thioredoxin fold

The Trx fold (Figure 1.15a), named after the protein in which it was first observed, is a distinct structural motif consisting of a four-/five-stranded β-sheet and three/four flanking α-helices. These secondary structures of the fold can be subdivided into an N-terminal βαβ motif and a C-terminal ββα motif, connected by a loop of residues that incorporates a third helix. The active site of this fold usually contains a dithiol/disulfide group which protrudes from the protein surface. The Trx fold that is common to other protein families (see below) is somewhat smaller than Trx itself, having one less β-strand and one less α-helix (Martin 1995).

1.3.4.1.2.2 The Rossmann fold

The Rossmann fold (Figure 1.15b) is the most populated fold among α/β-folds and is one of the three most highly represented folds in the Protein Data Bank (PDB). The Rossmann fold is known to have a super-site with functional residues located at the C-termini of the β-strands (Branden & Tooze 1999, Russell *et al*. 1998). A subset of the proteins that adopt the Rossmann fold also binds to nucleotide cofactors such as

FAD and NAD(P) and function as oxidoreductases. These Rossmann folds can often be identified by the short amino acid sequence motif, GX(1-2)GXXG. However, recently another motif (GXXXG/A) has been identified as an extended motif, V/IXGX(1-2)GXXGXXXG/A, that is more strongly indicative than previously described motifs of Rossmann folds that bind FAD or NAD(P) (Kleiger & Eisenberg 2002). The V/IXGX(1-2)GXXGXXXG/A motif can be used to search genomic sequence data and to annotate the function of proteins containing the motif as oxidoreductases, including proteins of previously unknown function.

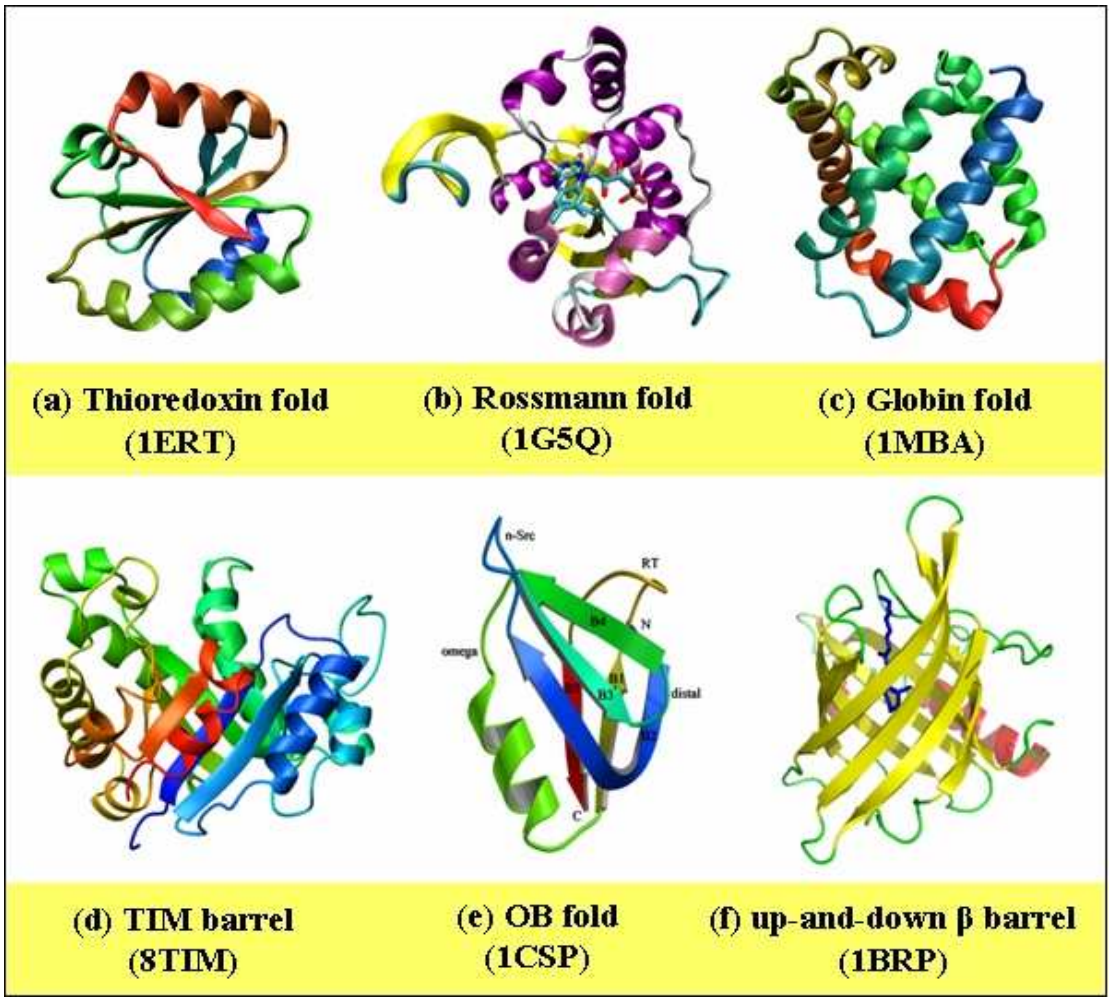


Figure 1.15 Cartoon representations of selected protein folds: (a) The thioredoxin fold, (b) The Rossmann fold, (c) The globin fold, (d) TIM barrel, (e) The oligonucleotide-binding fold (OB fold), (f) The up-and-down β barrel. PDB codes corresponding to structures are labelled (see text for details).

1.3.4.1.2.3 The globin fold

One of the most important α structures is the globin fold (Figure 1.15c). This fold has been found in a large group of related proteins, including myoglobin, hemoglobins, and the light-capturing assemblies in algae, the phycocyanins. The pairwise arrangements of the sequential α helices in the globin fold are quite different from the antiparallel organization found in the four-helix bundle α structures. The globin structure is a bundle of eight α-helices, A-H, connected by rather short loop regions and arranged so that the helices form a pocket for the active site, which in myoglobin and the hemoglobins binds a heme group. In the globin fold, the α-helices wrap around the core in different directions so that sequentially adjacent α-helices are usually not adjacent to each other in the structure. This fold therefore is not built up from an assembly of smaller motifs. Thus it is quite difficult to visualize conceptually in spite of the fold's relatively small size and simplicity. The only exceptions are the last two α helices G and H, which form an antiparallel pair with extensive packing interactions between them. All other packing interactions are formed between pairs of α-helices that are not sequentially adjacent (Branden & Tooze 1999).

1.3.4.1.2.4 TIM beta/alpha-barrel

The most frequent and most regular of the domain structures are the alpha/beta (α/β) domains, which consist of a central parallel or mixed β-sheet surrounded by α-helices. There are two main classes of α/β proteins. In the first class, there is a core of eight twisted parallel β-strands arranged close together, like staves, into a barrel

(known as the TIM barrel from the structure of the enzyme triosephosphate isomerase, where it was first observed). The α-helices that connect the parallel β-strands are all on the outside of this barrel (Branden & Tooze 1999). TIM barrel (Figure 1.15d) is the third most populated α/β-fold. Sequences and functions of the proteins sharing this fold are very diverse. Very little is known about stability of the TIM barrel proteins and no data are available regarding their folding kinetics. On the other hand, functions of the majority of TIM barrel proteins are well known. This fold has a distinctive super-site at the loops on the top of the barrel (Russell et al. 1998, Branden & Tooze 1999).

The second class contains an open twisted β sheet surrounded by α helices on both sides of the β sheet. A typical example, a nucleotide-binding domain found in a number of dehydrogenases and kinases (see below).

1.3.4.1.2.5 Oligonucleotide-binding fold

The oligonucleotide-binding (OB) (Figure 1.15e) fold is a β-fold with a barrel topology. The OB fold is built up from βαβ motifs that are linked such that the β-strands are parallel. In this fold (also known as open-sheet structures), there are α-helices on both sides of the β sheet and such alignments apply the geometric rule because virtually all βαβ motifs are right-handed (Branden & Tooze 1999). Proteins belonging to this fold have diverse sequences and functions. About 20 non-homologous protein families share this fold, for example, (i) nucleotide binding (transcription signals, RNA binding, tRNA synthetase, staphylococcal nucleases); (ii) inorganic pyrophosphates; (iii) tissue inhibitor of metalloproteinases; or (vi) toxins.

The binding site of single-stranded DNA and RNA is localized mostly on the face of the molecule in the strands 2, 3 and on the surface loops (Newkirk et al. 1994).

1.3.4.1.2.6 The up-and-down β barrel

Antiparallel beta (β) structures comprise the second large group of protein domain structures. Functionally, this group is the most diverse; it includes enzymes, transport proteins, antibodies (known as immunoglobulin fold), and virus coat proteins. The cores of these domains are built up by a number of β-strands that can vary from four or five to over ten. The β-strands are arranged in a predominantly antiparallel fashion and usually in such a way that they form two β-sheets that are joined together and packed against each other. The β-sheets have the usual twist, and when two such twisted β-sheets are packed together, they form a barrel-like structure. Antiparallel β structures, therefore, in general have a core of hydrophobic side chains inside the barrel provided by residues in the β-strands. The surface is formed by residues from the loop regions and from the strands. The three most frequently occurring groups: up-and-down barrels (Lalonde et al. 1994), Greek key (Vaughn & Bjorkman 1996) and jelly roll barrels (Adachi et al. 2001) – can all related to simple ways of connecting antiparallel β-strands arranged in a barrel structure.

The up-and-down beta-barrel (Figure 1.15f) is a common folding motif found frequently in proteins that bind and transport hydrophobic ligands. It is formed by an array of β-strands arranged in an antiparallel manner with each strand hydrogen-bonded to neighboring strands nearly always adjacent in the amino acid sequence. The arrangement is completed by forming hydrogen bonds between the first and last strands. The barrel motif so formed produces interior and exterior components.

Proteins belonging to this class of up-and-down beta-barrels are found typically to be lipid-binding proteins in which the interior surface forms a cavity or pit that serves as the ligand binding region. Two evolutionarily distinct but structurally related families of such carriers have been identified by comparing known crystal structures. One group found intracellularly uses a 10-stranded β-structure and a second family of proteins typically found extracellularly utilizes an 8-stranded motif. The 10-stranded β-barrels have a large, hydrophilic water-filled interior cavity that serves as the ligand-binding domain. Hydrophobic lipids such as fatty acids and retinoids bind within the cavity, totally sequestered from the external milieu. The 8-stranded beta-barrel proteins have a hydrophobic pit, which serves as the ligand-binding domain for compounds such as bilins and retinoids. The up-and-down β-barrel motif appears to be one of nature's primary choices for hydrophobic ligand transport proteins (Lalonde et al. 1994). The first reported example of this simplest topology is the plasma-borne retinol-binding protein, RBP, which is a single polypeptide chain of 182 amino acid residues. This protein is responsible for transporting the lipid alcohol vitamin A (retinol) from its storage site in the liver to the various vitamin-A-dependent tissues (for details see (Branden & Tooze 1999, Newcomer & Ong 2000, Zanotti & Berni 2004).

1.3.4.2 Definition of the canonical thioredoxin fold

The thioredoxin fold is a protein fold common to enzymes that catalyze disulfide bond formation and isomerization. The fold is named for the canonical example thioredoxin and is found in both prokaryotic and eukaryotic proteins. It is an example

of an alpha/beta protein fold that has oxidoreductase activity. The fold's spatial topology consists of a four-stranded β-sheet sandwiched between two α-helices.

1.3.4.2.1 Protein examples which contain TRX-fold

Thioredoxin family proteins are involved in various biologic processes by regulating the response to oxidative stress. The Trx-fold is a characteristic protein structural motif (Martin 1995) that has been found in at least other twelve classes of proteins that have the common property of interacting with cysteine-containing substrates. These proteins include the glutaredoxin, Dsb protein, protein disulfide isomerase, glutathione S-transferase, glutathione peroxidase, peroxiredoxins, PICOT (protein kinase C-interacting cousin of thioredoxin), nucleoredoxin (a nuclear thioredoxin-like protein with oxidoreductase activity), cytochrome c oxidase (COX, proteins conserved among prokaryotes and eukaryotes) assembly factors and evolutionary conserved Dim protein (Stefanková et al. 2005). This fold is modified in above mentioned proteins by additional amino acid residues at the N and/or C terminus of the Trx fold and some insertion in the specific regions of this fold. The overall structural similarity between the Trx fold in all proteins is very interesting, despite it, proteins suggest functional differences and low sequence identity (Martin 1995, Kemmink et al. 1996).

The Trx-fold comprises about 80 residues, but each of the proteins containing it has extra residues in addition to the fold (Figure 1.16). Glutaredoxin (87 residues) and thioredoxin (108 residues) are single-domain monomeric proteins: glutaredoxin has little structure in addition to the basic Trx fold, whereas Trx has an extra β-strand and α-helix at the N-terminus (Figure 1.16).

In bacteria, the periplasm contains a family of proteins that catalyze disulfide bond formation in newly translocated proteins. These proteins belong to the “Dsb” family (Dsb stands for disulfide bond). Most periplasmic Dsb proteins contain a thioredoxin fold and a catalytic CXXC motif. Depending on the function of the Dsb protein, the CXXC motif is kept in the reduced form (Dsb-$(SH)_2$; with a dithiol) or in the oxidized form (Dsb-S_2; with an intramolecular disulfide bond). The Dsb proteins are involved in two major pathways: an oxidation pathway (DsbA and DsbB) and an isomerization pathway (DsbC, DsbG, DsbD) (Messens & Collet 2005).

Protein disulfide isomerase (PDI), a multifunctional protein of the endoplasmic reticulum, catalyzes the formation, breakage and rearrangement of disulfide bonds during protein folding. Dissection of this protein into its individual domains has confirmed the presence of the a and a' domains, which are homologous to thioredoxin, having related structures and activities. The a and a' domains both contain a -Cys-Gly-His-Cys- active-site sequence motif. The remainder of the molecule consists primarily of two further domains, designated b and b' which are thought to be sequence repeats on the basis of a limited sequence similarity. The global fold of the b domain of PDI by NMR reveals that, like the a domain, the b domain contains the thioredoxin motif, even though the b domain has no significant amino-acid sequence similarities to any members of the thioredoxin family. This observation, together with indications that the b' domain adopts a similar fold, suggests that PDI consists of active and inactive thioredoxin modules. These modules may have been adapted during evolution to provide PDI with its complete spectrum of enzymatic activities (Kemmink et al. 1997).

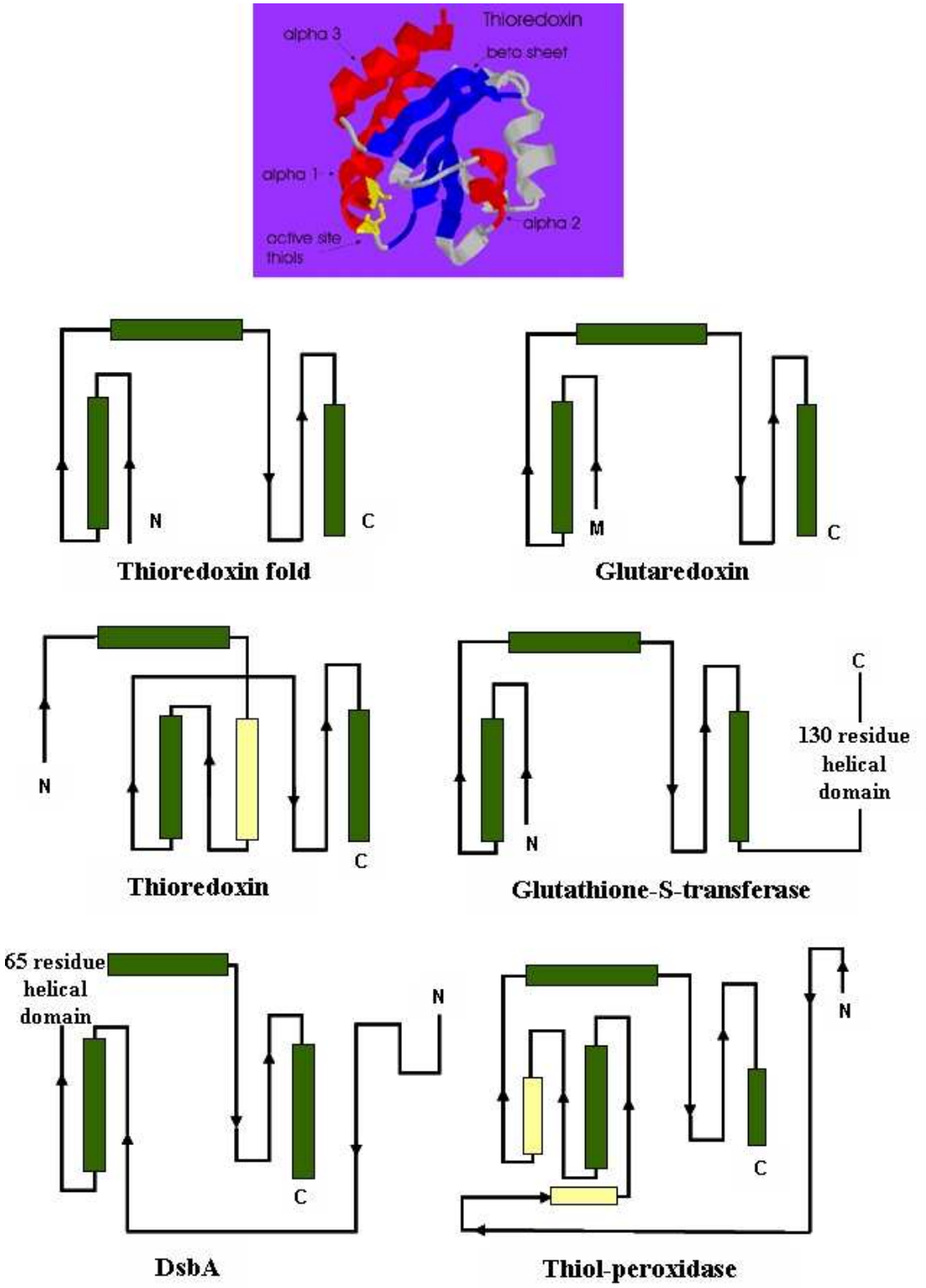


Figure 1.16 Structure of Trx and typical architecture of Trx-like proteins. The Trx fold, named for the founding member of the family Trx, is a protein fold consisting of four β-sheets surrounded by three α-helices. The Trx family of protein is typically involved in redox active reactions and usually contains a CxxC motif, where x is any amino acid. Lower panel of the figure displays the two-dimensional representation of the Trx-fold in representative proteins in the Trx superfamily. β-Sheets are

drawn as arrows and a-helices are drawn in boxes (dark green). A typical topology of Trx-fold is shown where two parallel β-sheets are intervened by an α-helix. In addition, between the N-terminal and C-terminal portion is an α-helix that connects these two domains. Helices that are not part of the Trx-fold are drawn colored in beige (Figure is adapted from Pan & Bardwell 2006).

Glutathione peroxidase (Gpx) belongs to the family of selenoproteins and plays an important role in the defense mechanisms of mammals, birds and fish against oxidative damage by catalyzing the reduction of a variety of hydroperoxides, using glutathione as the reducing substrate (Ren et al. 1997). The subunit structure of animal Gpx shows the typical structure motif of the thioredoxin fold consisting of a central beta-sheet and several flanking alpha-helices. The active site selenocysteine residue is situated in the loop between β1 and α1 and is located in a pocket on the protein surface. The overall structures of the Gpx enzymes are similar among homologues. The main difference in their subunit structures is an extended N terminus in certain isoforms.

Peroxiredoxins (Prxs) are a ubiquitous family of antioxidant enzymes that also control cytokine-induced peroxide levels which mediate signal transduction in mammalian cells. Prxs can be regulated by changes to phosphorylation, redox and possibly oligomerization states. Prxs are divided into three classes: typical 2-Cys Prxs; atypical 2-Cys Prxs; and 1-Cys Prxs. All Prxs share the same basic catalytic mechanism, in which the peroxidatic cysteine is oxidized to a sulfenic acid by the peroxide substrate. The recycling of the sulfenic acid back to a thiol is what distinguishes the three enzyme classes (Wood et al. 2003). Since 1998, the crystal structures of six Prxs have been published, including four typical 2-Cys Prxs (PrxI, PrxII, TryP and AhpC; Hirotsu et al. 1999, Schröder et al. 2000, Alphey et al. 2000,

Wood et al. 2002), one atypical 2-Cys Prx (PrxV, Declercq et al. 2001) and one 1-Cys Prx (PrxVI; Choi et al. 1998). These structures reveal Prxs to be very similar, each containing a thioredoxin fold with a few additional secondary-structure elements present as insertions. The most striking differences involve their oligomeric states. The structure and sequence of the peroxidatic active site is highly conserved among the Prx classes.

PICOT protein plays a negative-regulatory role in cellular stress responses associated with activation of the transcription factors AP-1 and NF-κB (Witte et al. 2000). PICOT possesses an N-terminal thioredoxin (Trx)-homology domain, followed by two (human and animal) or three (plant) tandem repeats of a highly conserved sequence motif of 84 residues, termed as PICOT-homology domain (HD) (Isakov et al; 2000). Duplication of PICOT-HD (generally localized at the C-terminus of the protein) was found in mammalian-derived PICOT-proteins (human, mouse and rat), whereas three tandem repeats of PICOT-HD were found in plants (*A. thaliana* and poplar tree). There are also proteins which contain only a single PICOT-HD (yeast). Predicted secondary structure (obtained using the PHD program) of PICOT-HD suggests a globular topology, which includes three α helices and an intervening β-strand between each pair of helices. It is interesting to note that the overall predicted structure of PICOT-HD resembles that of Grx, which is predicted to possess just one extra short β-strand before the C-terminal α-helix (Aslund et al. 1996). To date, there is only a solution structure (by NMR) of the PICOT homology 2 domain of the mouse PKC-interacting cousin of thioredoxin protein is available in the Protein Data Bank (PDB code 1WIK).

Nucleoredoxin (Nrx), a relatively uncharacterized member of the Trx family protein, has recently been reported to regulate the Wnt/β-catenin pathway, which

itself regulates cell fate and early development (cell proliferation and differentiation), in a redox-dependent manner. Although Nrx possesses a conserved Trx domain and a catalytic motif for oxidoreductase activity, its sequence homology to Trx is not as high as that of the close relatives of Trx. The sequence of Nrx is more similar to that of tryparedoxin (TryX), a Trx family member originally identified in parasite trypanosomes (Funato & Miki 2007).

Cytochrome c oxidase (COX) is a multi-subunit terminal oxidase of the eukaryotic respiratory chain involved in the reduction of oxygen to water. Numerous lines of evidence suggest that the assembly of COX is a multi-step, assisted process that depends on several assembly factors with largely unknown functions (Chinenov 2000). Sco1, a protein required for the proper assembly of COX, has a soluble domain anchored to the cytoplasmic membrane through a single transmembrane segment. The solution structure of the soluble part of apoSco1 from *Bacillus subtilis* has been solved by NMR and the internal mobility characterized. Its fold places Sco1 in a distinct subgroup of the functionally unrelated thioredoxin proteins. *In vitro* Sco1 binds copper (I) through a CXXXCP motif and possibly His135 and copper (II) in two different species, thus suggesting that copper(II) is adventitious more than physiological. The Sco1 structure represents the first structure of this class of proteins, present in a variety of eukaryotic and bacterial organisms, and elucidates a link between copper trafficking proteins and thioredoxins (Balatri et al. 2003).

The Dim protein family is composed of two classes, Dim1 and Dim2, which share a common thioredoxin-like fold. They were originally identified for their role in cell cycle progression and have been found to interact with Prp6, an essential component of the spliceosome, which forms the bridge of U4/U6.U5-tri-snRNP. In spite of their biological and structural similarities, Dim1 and Dim2 proteins differ in

many aspects. Dim1 bears distinctive structural motifs responsible for its interaction with other spliceosome components. Dim2 forms homodimers and contains specific domains required for its interactions with partners. This originality suggests that although both proteins are involved in pre-mRNA splicing, they are likely to be involved in different biological pathways (Simeoni & Divita 2007).

1.3.4.2.2 Sequence conservation

All Trx have similar three-dimensional structures despite the large variation in amino acid sequences (Eklund et al. 1991, Martin 1995). Despite sequence variability in many regions of the fold, Trx proteins share common active site sequence with two cysteine residues: CxxC, where x is often but not necessary hydrophobic amino acid. The reduced form of the protein contains two free thiol groups at the cysteine residues, whereas the oxidized form contains a disulfide bond between them.

1.3.4.2.3 Disulfide bond formation

Different Trx fold-containing proteins vary greatly in their reactivity and in the pKa of their free thiols, which derives from the ability of the overall protein structure to stabilize the activated thiolate. Although the structure is fairly consistent among proteins containing the thioredoxin fold, the pKa is extremely sensitive to small variations in structure, especially in the placement of protein backbone atoms near the first cysteine. Apart from the primary functions of various Trx fold-containing proteins, disulfide bond formation can also play a regulatory role. Native disulfide

bond formation is not trivial, so it is often the rate-limiting step of protein folding both *in vivo* and *in vitro* (Hatahet & Ruddock 2007).

1.4 Protein of interest

As the production of reactive oxygen species (ROS) is an unavoidable consequence of living in an aerobic environment, photosynthesizing plant cells have their mechanisms for regulating this production. Once produced, ROS can be removed by several different enzyme systems. Finally, should these strategies fail; the ROS produced can act as a signal to the rest of the cell and/or cause damage to DNA, lipids and proteins. In this study, I will be focusing on some oxidoreductases (glutathione peroxidase, thioredoxin and glutaredoxin) which are involved in the first line of defense against ROS in the poplar tree (of economic interest).

1.4.1 Glutathione peroxidase (Gpx)

1.4.1.1 Biological theme:

1.4.1.1.1 General functions in living organisms

In aerobic organisms, ROS such as superoxide radicals, hydroxyl radicals or hydrogen peroxide are generated during the incomplete reduction of molecular oxygen to water (Farr & Kogoma 1991), following exposure to environmental factors (Dat et al. 2000) or during several metabolic processes (e.g. photosynthesis, respiration) (Milla et al. 2003). The accumulation of ROS can gives rise to oxidative

stress and leads to cell damage(s), mutation(s), or even death. The ROS challenge to land plants is made difficult because they are immobile in a constantly changing environment and, besides consuming O_2 during respiration they also generate it during photosynthesis (Foyer et al. 1994). Plants have developed several non-enzymatic and enzymatic systems to withstand the oxidative damage caused by these ROS. While the major non-enzymatic antioxidant compounds include carotenoids, tocopherols, reduced glutathione (GSH) and ascorbate (Noctor & Foyer 1998), the enzymatic systems rely on superoxide dismutases, catalases, ascorbate peroxidases, peroxiredoxins (Prxs) and glutathione peroxidases (Gpxs). In mammals, Gpxs that form a group of enzymes with an important role in protecting cells against ROS, using GSH as the reducing substrate, have been studied extensively (Roy et al. 2005). In plants, different isoforms of Gpx are found to be expressed in various subcellular compartments. Recently, one of these Gpxs was proposed to play a role in signal transduction during stress conditions (Miao et al. 2006, Eshdat et al. 1997). Likewise, *Saccharomyces cerevisiae* has also been shown to possess three plant-type Gpxs, with the Gpx-3 functioning as a redox sensor involved in gene activation (Delaunay et al. 2002).

1.4.1.1.2 Gpx families in animals and plants

Depending on the active site amino acid, mammalian Gpx enzymes are generally classified into two groups: one contains SeCys at its active center (at least four out of five isozymes belong to the selenium-containing Gpxs in mammals), the other contains a regular cysteine (the mammalian epididymis-specific isoform, Gpx-5) (Grignard et al. 2005). The non-selenium type Gpx-5 is speculated to serve as the

backup enzyme for the selenium-containing Gpx to counteract ROS-mediated damages of sperm cells, since selenium deficiency causes male infertility and selenium-containing Gpx is suspected to be a candidate for the defective molecule (Grignard et al. 2005, Fujii et al. 2003). The activity of selenium-independent Gpx is low and, hence, its contribution as a GSH-dependent peroxide scavenger is ambiguous (Fujii et al. 2003).

1.4.1.1.3 Plant Gpxs

Typical plant enzymes display sequences similar to those of animal Gpx enzymes, except that the SeCys in the catalytic site of most animal Gpx is replaced by Cys in plant Gpx, which in general show reduced catalytic efficiency (Navrot et al. 2006, Jung et al. 2002, Kang et al. 2004, Holland et al. 1993, Sugimoto & Sakamoto 1997, Ren et al. 2001). Interestingly, recent studies have demonstrated that some plants (Navrot et al. 2006, Jung et al. 2002, Herbette et al. 2002) and yeast (Tanaka et al. 2005) Gpxs can reduce peroxides, much more efficiently or sometimes exclusively, using Trx instead of GSH as an electron donor. In the unicellular parasite *Plasmodium falciparum*, a GPX-like protein closely related to plant Gpxs is also specific for Trx as a reductant (Sztajer et al. 2005). Most animal Gpxs catalyse the reduction of hydrogen peroxide by consuming two GSH molecules to reduce the oxidized form of SeCys (selenenic acid) that is generated upon catalysis (H_2O_2+2GSH→GS-SG+2H_2O) (Prabhakar et al. 2005, 2006, Flohé et al. 1972, Mugesh et al. 2001). On the contrary, plant Gpxs possess a second Cys residue to complete the reduction of ROS. Navrot *et al*. (2006) have recently proposed a three-step reaction mechanism for plant Gpxs similar to that used by peroxiredoxin Q (Prx

Q) (Navrot et al. 2006, Rouhier & Jacquot 2005), with two Cys forming an intramolecular disulfide bridge in the oxidized state. These steps are: (i) a nucleophilic attack of the peroxidatic cysteine (Cys107, numbering in PtGpx3.2, see Navrot et al. 2006) on the peroxide with the release of an alcohol and the concomitant formation of a sulfenic acid; (ii) an attack of the sulfenic acid by the resolving Cys (Cys155, numbering in PtGpx3.2, see Navrot et al. 2006) that leads to the formation of an intramolecular disulfide bridge between Cys107 and Cys155; and finally (iii) the reduction of disulfide by Trx. Plant Gpxs have low substrate specificity, in which they are capable to reduce a wide spectrum of peroxides, including those more complex hydroperoxides (Navrot et al. 2006). On the basis of biochemical evidences, rather than on phylogenetic classification, Rouhier and Jacquot (2005) suggested that plant Gpxs should be included in the large thioredoxin-dependent peroxidase family. This is relevantly true because if taking account only the primary structure homologies, the levels of sequence identity between Gpxs and PRXs are relatively low (ranging from 1 % with cytosolic peroxiredoxin type IIB to 16 % with the Prx Q; see Rouhier and Jacquot (2005) for the nomenclature of peroxiredoxin) but they are actually very closely-related in terms of function and 3-dimensional structures.

1.4.1.1.4 Poplar Gpxs

Black cottonwood (poplar) has been chosen as the model plant in this study. Very recently, an exhaustive overview was done on poplar Gpxs and the annotation of the first release of the poplar genome indicates the presence of six complete GPX genes (Navrot et al. 2006). The poplar Gpxs were initially defined as GSH dependent, in

fact they use only reduced Trx for their regeneration and do not react with GSH or glutaredoxin. The two chloroplastic Gpxs display a marked selectivity toward their electron donors, being exclusively specific for Trxs of the *y*-type (plastidial Trx) for their reduction. In contrast, poplar Gpxs are much less specific with regard to their electron-accepting substrates, reducing hydrogen peroxide and more complex hydroperoxides equally well. Site-directed mutagenesis indicates that the catalytic mechanism and the Trx-mediated recycling process involve only two (Cys-107 and Cys-155) of the three conserved Cys, which form a disulfide bridge with an oxidation-redox midpoint potential of -295 mV. The reduction/formation of this disulfide is detected both by a shift on sodium dodecyl sulfate-polyacrylamide gel electrophoresis or by measuring the intrinsic tryptophan fluorescence of the protein. The six genes identified coding for Gpxs are expressed in various poplar organs, and two of them are localized in the chloroplast, with one colocalizing in mitochondria, suggesting a broad distribution of Gpxs in plant cells. The abundance of some Gpxs is modified in plants subjected to environmental constraints, generally increasing during fungal infection, water deficit, and metal stress, and decreasing during photooxidative stress, showing that Gpx proteins are involved in the response to both biotic and abiotic stress conditions. Poplar Gpx5 is a putatively secreted protein and it is particularly interesting because it may involve in the first line defense of plant cells during oxidative stresses.

1.4.1.2 Structural biology

1.4.1.2.1 Gpx structures known to date

To date, there are nine Gpx crystal structures available, all of them from mammals. These crystal structures include one bovine Gpx-1 (PDB code 1GP1) structure (Epp et al. 1983) and five human Gpx structures (Gpx-1, PDB code 2F8A; Gpx-2, PDB code 2HE3; Gpx-3, courtesy from Professor Rudolf Ladenstein (Ren et al. 1997); Gpx-3 selenocysteine to glycine mutant, PDB code 2R37; Gpx-4, PDB code 2GS3; Gpx-4 selenocysteine to cysteine mutant, PDB code 2OBI (Scheerer et al. 2007); Gpx-5, PDB code 2IY3 and Gpx-7, PDB code 2P31). These Gpxs are homotetrameric enzymes, except that human Gpx-4 (an isoform that specifically detoxifies phospholipid hydroperoxides) is reported as a monomeric enzyme. The overall structures of all mammal Gpx monomers are similar (all possesses the Trx-fold), except that the environment around their active site catalytic residues is quite different (Prabhakar et al. 2005, Epp et al. 1983, Ren et al. 1997).

1.4.2 Thioredoxin (Trx)

1.4.2.1 Biological theme

1.4.2.1.1 General functions

Thioredoxins are small molecular weight proteins found in all organisms from prokaryotes to higher eukaryotes, considered to be the best reductants (low redox

potential confers strong reductive properties) in the cell. It behaves as an electron transporter containing a disulfide bridge (-S-S-) in its oxidized from (Trx-S_2) and a dithiol in its reduced form (Trx-$(SH)_2$); both forms are catalytically redox-active (Holmgren 1985). Apart from few exceptions, majority of Trxs have the highly conserved sequence -Cys-Gly-Pro-Cys- (-CGPC-) located in their dithiol/disulfide active site (Arnér & Holmgren 2000). The redox reaction that is catalyzed by thioredoxin reductase (TrxR) is shown in Figure 1.17. The oxidized form of Trx is reduced by NADPH and TrxR to form Trx-$(SH)_2$ and through the reversible oxidation of the Trx-$(SH)_2$ active site dithiol, to a disulfide again (Trx-S_2). The Trx, NADPH and TrxR together form a system called “the Trx system” or NTR.

The main function of thioredoxin is to reduce the disulfide bonds of target proteins to change the conformation and activity of these proteins. The active form of Trx is Trx-$(SH)_2$, the two Cys at the active site are in the thiol form. The Trx-$(SH)_2$ behaves as a reducing agent to reduce disulfide bonds in the target proteins. After the reduction of the target proteins, the two Cys form a disulfide bond between them. The mechanism of protein disulfide reduction catalyzed by the Trx system is schemed as below (Figure 1.17, Arnér & Holmgren 2000, Holmgren 1981, 1989, Minarik 1997).

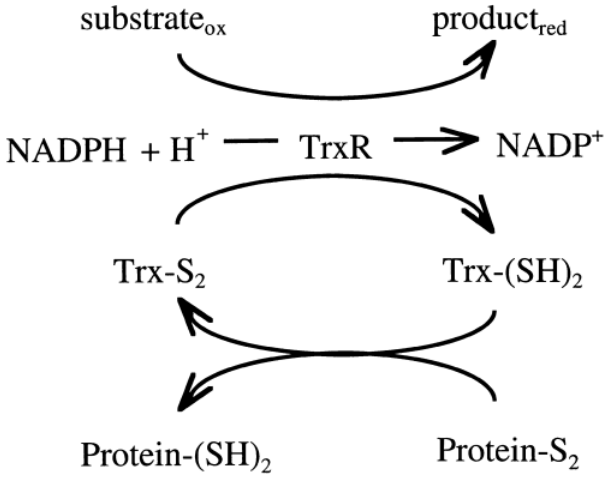


Figure 1.17 The reaction scheme of Trx catalyzed protein disulfide reduction (Arnér & Holmgren 2000).

Trxs are involved in many cellular processes, dealing primarily with cell redox regulation (Nakamura 2005, Santos & Rey 2006, Powis & Kirkpatrick 2007). Trx can be described as a hydrogen donor or disulfide reductase and function as a regulatory protein to regulate the thiol-disulfide status of proteins. During the last decade, the functional roles of Trx have continued to expand, also including novel functions such as a secreted growth factor or a chemokine for immune cells (Lillig & Holmgren 2007). Trx is also involved in many other biological activities such as an antioxidant (Yoshida et al. 2003), cofactor (Masutani et al. 2004), growth factor (Oblong et al. 1994) and regulator (Tanaka et al. 2002). It also activates and regulates the DNA binding transcription factor NF-κB, and bind with various proteins (Powis & Montfort 2001). The role of Trx in antioxidant defense and in controlling recruitment of inflammatory cells offers potential use in clinical therapy (Lillig & Holmgren 2007). In addition, its small size, compact globular structure and high stability have prompted its widespread utilization in biotechnology, especially as a

fusion module that enhances both the solubility and stability of recombinant proteins (Dyson et al. 2004).

1.4.2.1.2 Distinct families in all organisms

In animal and bacterial cells there are multiple thioredoxins which are reduced *via* NADPH and a thioredoxin reductase (Holmgren 1989, Lillig & Holmgren 2007). In plants, Trxs display a particularly large diversity in comparison to non-photosynthetic organisms, using the same reduction system.

1.4.2.1.3 Plants thioredoxins

In plants, numerous Trx isoforms have been reported. For example, at least 20 genes encoding "true thioredoxins" with a conserved catalytic site WC[G/P]PC are present in the completely sequenced genome of *A. thaliana* (Gelhaye et al. 2005, Meyer et al. 2006, Jacquot et al. 2002). The Trxs *f*, *m*, *x* and *y* are present in chloroplasts (Schürmann 2003, Gelhaye et al. 2005), whereas the thioredoxins *o* are localized in mitochondria (Laloi et al. 2001) and the *h* type in the cytosol and mitochondria (Gelhaye et al. 2005). In chloroplasts, besides their function in the ROS detoxification system, Trxs are involved in the light regulation of carbon metabolism through fine-tuning of the reducing pentose phosphate pathway and also of the C4 pathway of photosynthesis (Rouhier & Jacquot 2002). In this regulatory pathway, the reducing power provided by light is transferred through ferredoxin and FTR to Trx. In addition, a fusion protein between a thioredoxin reductase and a Trx module called NTRc has been shown to be expressed in plastids, which increases even more

the complexity of the plastidial Trx system (Pérez-Ruiz et al. 2006). In contrast, in cytosolic and mitochondrial systems, the reducing power from NADPH is mainly transferred to thioredoxins *via* a specific NTR (Jacquot et al. 1994, Laloi et al. 2001).

1.4.2.1.4 Poplar thioredoxin *h*

Plants contain several genes encoding thioredoxin *h*. These proteins have been reported to be important in stimulating cell growth, germination, during early seedling development and are implicated in different aspects of plant life (including adaptation to environmental changes and stresses). Nevertheless the mechanism regulating the expression of these genes and their specific function is still poorly unknown (Cazalis et al. 2006, Meyer et al. 2007). Likewise in other plants, the thioredoxins *h* of poplar constitute a large disparate group including cytosolic and mitochondrial thioredoxins (Gelhaye et al. 2004a, Laloi et al. 2001, Bréhélin et al. 2004, Gelhaye et al. 2004b). Thioredoxins *h* have been divided in three distinct subgroups in a classification based on protein primary structure and position of the introns (Gelhaye et al. 2004a). Members of the first and second groups are reduced by NADPH in a reaction catalyzed by NADPH-Thioredoxin Reductase (NTR). NTR is a homodimeric flavoprotein, with each subunit containing one NADPH-binding domain and one flavin adenine dinucleotide (FAD)-binding molecule. Members of the first and second thioredoxins *h* subgroups contain a conserved WC[G/P]PC catalytic site. The first cysteine is the one involved in the nucleophilic attack on disulfide bridge presents in target proteins, leading to the formation of a disulfide bridge between the target protein and thioredoxin. This intermolecular disulfide is

then reduced by the second cysteine, leading to the release of reduced target protein and oxidized thioredoxin.

It is only recently that members of plant thioredoxin *h* subgroup 3 have been detected. Poplar is among the plants in which this type of thioredoxin has been detected (Gelhaye et al. 2003, Juttner et al. 2000). While the poplar thioredoxin PtTrx*h*4 belongs to this subgroup, it contains the usual catalytic site WCGPC but differs from previously characterized thioredoxins *h* in the reduction pathway used. This isoform is not reduced *in vitro* by NTR but by glutaredoxins (Gelhaye et al. 2003). This unique feature raised several questions about the mechanisms of reactions involving PtTrx*h*4. For example, as most characterized Trxs have a redox midpoint potential of ca -290 mV while Grxs are more electropositive (-200 mV), questions regarding the thermodynamic favourability of reduction of a Trx-like molecule by Grx naturally arise.

1.4.2.2 Structural biology

1.4.2.2.1 Thioredoxin structures known to date

The crystal structures of several thioredoxins have been solved, from *E. coli* (Holmgren et al. 1975, Katti et al. 1990), human (Weichsel et al. 1996, Smeets et al. 2005), *Anabaena* (Saarinen et al. 1995), *Chlamydomonas reinhardtii* (Menchise et al. 2000, Menchise et al. 2001), spinach (Capitani et al. 2000), *Trypanosoma brucei brucei* (Friemann et al. 2003), Drosophila (Wahl et al. 2005) and yeast (Bao et al. 2007). However, structural information on the plant thioredoxin type-*h* was not known until recently (Coudevylle et al. 2005, Peterson et al. 2005). Nevertheless,

there is no *h*-type plant Trx of subgroup 3 structures had been solved to date. Since the latter exhibits an N-terminal extension (as compared to subgroup 1) and possesses a conserved cysteine in the forth position among homologs, thus essential questions regarding the role(s) and effect(s) of the presence of the extension and an additional cysteine (Cys4) to the protein structure and functions arise. It was of interest to know how this extension alters the formation of a Trx fold and most importantly to understand why this protein does not bind its typical reducing partner, NTR.

1.4.3 Glutaredoxin (Grx)

1.4.3.1 Biological theme

1.4.3.1.1 General functions in living organisms

Glutaredoxins (Grxs) are thiol-disulfide oxidoreductases and involved in the maintenance of cellular redox homeostasis. They are glutathione (GSH)-dependent enzymes, first discovered as electron donor for ribonucleotide reductase (Holmgren 1976, 1979). Grxs belong to the thioredoxin family of proteins, with whom they share a similar structure and overlapping functions (Holmgren 1989). Grxs usually reduce disulfides *via* two distinct mechanisms (Figure 1.18). The dithiol mechanism requires both active site Cys residues, whereas the monothiol mechanism relies only on the more N-terminal active site cysteine (Bushweller et al. 1992). Using these mechanisms, Grxs catalyze the glutathione-dependent reduction of disulfides and glutathione-mixed disulfides, serving as electron donor, e.g., for ribonucleotide or

sulfate reduction (Holmgren 1979, Lillig et al. 1999) and controlling the levels of protein GSH-mixed disulfides (Holmgren & Åslund 1995). In comparison, the structurally and functionally related Trxs reduce a wide range of protein disulfides, but have low or no activity with mixed disulfides (Fernandes & Holmgren 2004). Grx is then recycled by the ubiquitous GSH and GSSG in turn is reduced by NADPH and glutathione reductase (together called the GSH/Grx reducing system).

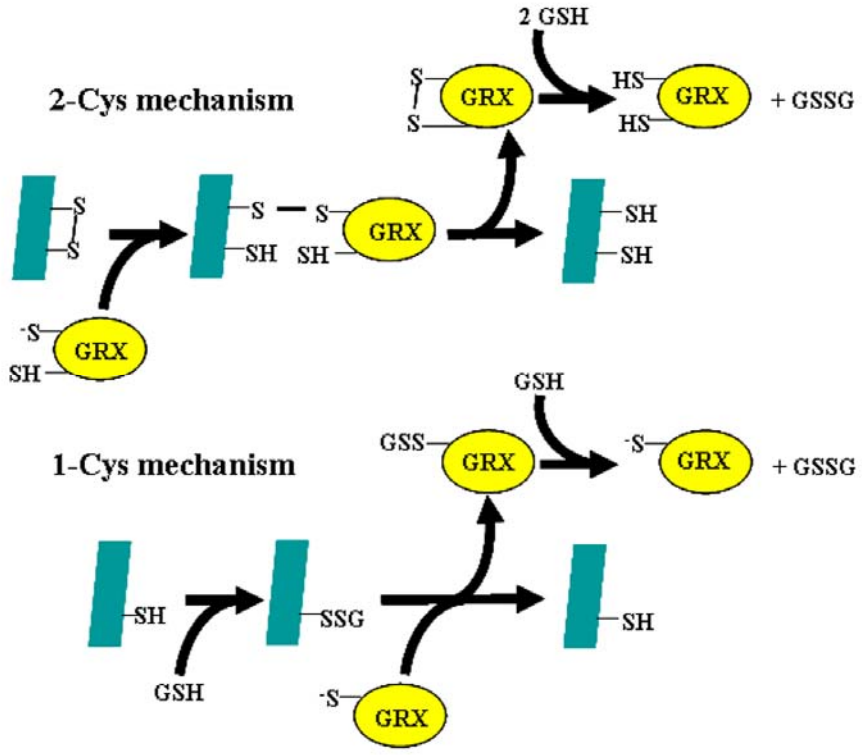


Figure 1.18 2-Cys and 1-Cys mechanisms of action of Grxs for reduction of dithiol groups. Active cysteine residues in glutaredoxin proteins are indicated as SH or S– depending on the redox state. (Grx, glutaredoxin; GSH, reduced glutathione; GSSG, glutathione disulfide) See text for details. Figure adapted from Herrero & de la Torre-Ruiz 2007).

The Grx family has grown during the last years, and there are today numerous isoforms known in different organisms with quite different structures and catalytic

activities. In terms of their structure and catalytic properties, Grxs can today be classified in three categories (Vlamis-Gardikas & Holmgren 2002) and this classification is slightly different from that of plant Grxs (see following section). The first group is exemplified by the classical 10-kDa Grx proteins with the CxxC motif (usually CPYC) as their active site and present in thioredoxin/glutaredoxin fold. Belonging to this first classical category, both Grx1 and Grx3 of *E. coli* (10-kDa proteins) have 33 % sequence identity and share similar structure topology (Aslund et al. 1996, Bushweller et al. 1992, Martin 1995). These enzymes are electron donors for reductive enzymes like ribonucleotide reductase. The second category is structurally related to the glutathione-*S*-transferases (GSTs), but with Grx oxidoreductase activity. Common structural characteristics are a two-domain structure, the first domain having a Trx/Grx fold containing the active-site residues and the second domain having an α-helical structure. This class of Grxs is defined by E. coli Grx2, which has a three-dimensional structure, highly similar to GSTs (Xia et al. 2001). This type of Grxs differ from the GSTs as in they contain the dithiol active-site sequence CPYC in the Grx domain and thus has Grx activity. All of these proteins, nevertheless, are detoxifying or stress response proteins. The third category of Grxs is defined by having a monothiol active site (normally with CGFS sequence). Monothiol Grxs, widespread among prokaryotes and eukaryotes, have so far been identified in bacteria (Fladvad et al. 2005), yeast (Rodríguez-Manzaneque et al. 2002), plants (Lemaire 2004, Rouhier et al. 2004, 2006, Cheng et al. 2006) and man (Witte et al. 2000). Studies in *S. cerevisiae* suggest that monothiol Grxs have a protective role against oxidative stress, very specific for their substrates and their functions cannot be replaced by their dithiol counterparts (Rodríguez-Manzaneque et al. 1999).

S-glutathionylation is a reversible post-translational modification with critical roles in sulfhydryl homeostasis and signal transduction. Protein glutathionylation increases globally during overt oxidative stress [e.g. cardiac ischemia-reperfusion (Eaton et al. 2002)], but selectively in the presence of ROS that generated during physiological signaling (e.g. by growth factors or angiotensin II (Adachi et al. 2004). Reversible glutathionylation modulates the activities of diverse protein substrates *in situ*, with effects on multiple physiological events (Shelton et al. 2005, Shelton et al. 2007, Adachi et al. 2004, Reynaert et al. 2006). Reversible protein S-glutathionylation (protein-SSG) is an important post-translational modification, providing protection of protein cysteines from irreversible oxidation and serving to transducer redox signals. Protein deglutathionylation is efficiently catalyzed by the thiol-disulfide oxidoreductase glutaredoxin (Grx) which is analogous to phosphatases and accounts for most of the deglutathionylating activity in mammalian cells (Chrestensen et al. 2000, Jung & Thomas 1996) as well as regulating diverse intracellular signaling pathways (Gallogly & Mieyal 2007). Grx is the most thoroughly characterized deglutathionylating enzyme, with the mammalian cytosolic form, that is, Grx1, serving as the prototype. Grx1 operates *via* a nucleophilic, double displacement (ping-pong) mechanism with both steps displaying high commitment to catalysis. Rate enhancement by Grx1 has been attributed to the unusually low p*K*a (3.5) of the Cys22-thiol, which serves as the leaving group in the second, rate-limiting step (Srinivasan et al. 1997, Jao et al. 2006). Although other factors may influence the redox and catalytic properties of the CxxC/S motif (Hennecke et al. 1997), the p*K*a value of the N-terminal cysteine clearly contributes to these properties (Chivers et al. 1997).

In Grxs, the catalytic cysteine (with a lowered p*K*a value) is the key for nucleophilic attacks on various disulfide substrates, initiating the thiol-disulfide exchange reactions catalyzed by these enzymes. This leads to an intermediate mixed-disulfide where the substrate is covalently linked to the Grx. Then, release of the reduced substrate can proceed either *via* a dithiol or a monothiol mechanism. In the dithiol mechanism there is an intramolecular attack of the mixed-disulfide by the second Grx active-site cysteine, yielding an oxidized Grx, which can then be reduced by GSH. In the monothiol mechanism, the mixed-disulfide is reduced directly by GSH. The details of these mechanisms and their many cellular functions have been reviewed recently (Fernandes & Holmgren 2004).

1.4.3.1.2 Plant Grx system

Genome sequencing projects of photosynthetic organisms revealed the presence of abundant Grxs in plants (~27-38 isoforms) as compared to bacteria genomes (1-5 isoforms), animal genomes (2-3 isoforms) and mushroom genomes (3-6 isoforms), thus raising the question of the redundancy and of the specificity of these isoforms (Rouhier et al. 2004). The high number of genes coding for Grx in plants suggests that these proteins should have many important biological functions (Rouhier et al. 2006). Figure 1.19 summarizes established roles for plant Grxs and putative functions based on available data in other organisms. When submitted to adverse environmental conditions (biotic or abiotic stresses), plants very often react by generating oxidative bursts. To cease severe biological damages, the concentration of the oxidizing species must be kept under control. One of the most documented functions of Grxs in plants is their involvement in the oxidative stress response. They

are implicated in many different ways, for example by directly reducing peroxides or dehydroascorbate (DHA), by reducing peroxiredoxins (Prx), and also by protecting thiol groups on other enzymes via glutathionylation/deglutathionylation mechanisms (for a review, see Rouhier et al. 2004). Grx is indeed a specific and efficient catalyst of protein-glutathione mixed disulfide reduction. Upon deglutathionylation of protein-glutathione, Grxs are then recycled by the GSH/Grx reducing system in order to become reduced again. A recent finding, not linked to oxidative stress, is that some Grxs of subgroup 2 (see classification below) (AtGrx S14 and AtGrx S16) interact with Ca^{2+} transporters and probably regulate their activity (Cheng & Hirschi 2003). Indirectly, this suggests that Grxs could be involved in cell signaling by regulating the Ca^{2+} concentration. Plant Grxs could play additional roles, uncharacterized so far, in signaling, such as the regulation of phosphatases or transcription factors, two well-known reactions in other organisms. For example, a protein tyrosine phosphatase 1B from soybean was shown to be regulated *via* cysteine glutahionylation (Dixon et al. 2005).

Despite the availability of genome sequences from different oxygenic photosynthetic organisms, the number of Grxs and the different groups present in these organisms are still rarely described. The sizes of these proteins range between 10-15 kDa (Fomenko & Gladyshev 2002). Based on sequence alignments, active site sequences and construction of unrooted phylogenetic trees, three Grx subgroups can be distinguished. It is important to mention here that the Grxs classification for plants slightly differs from the classification of Grxs from other organisms described in the previous section. This is because plants most likely possess Grx isoforms with active site (see below) which are specific for higher plants and appears later in the evolution (Rouhier et al. 2004). These subgroups (numbering in poplar tree) are similar to

those already described for *A. thaliana* Grx (Rouhier et al. 2004, 2006, Lemaire 2004).

Subgroup 1 contains "classical" Grxs with CPYC, CGYC, CPFC, and CSY[C/S] active sites. This group comprises five different classes of Grx (Grx C1-C4 and S12) which differ in their active site sequences. The nomenclature used (C or S) is based on the presence of a cysteine or a serine in the fourth position of the active site (CxxC or CxxS). Nevertheless the presence of isoforms varies with the higher plants.

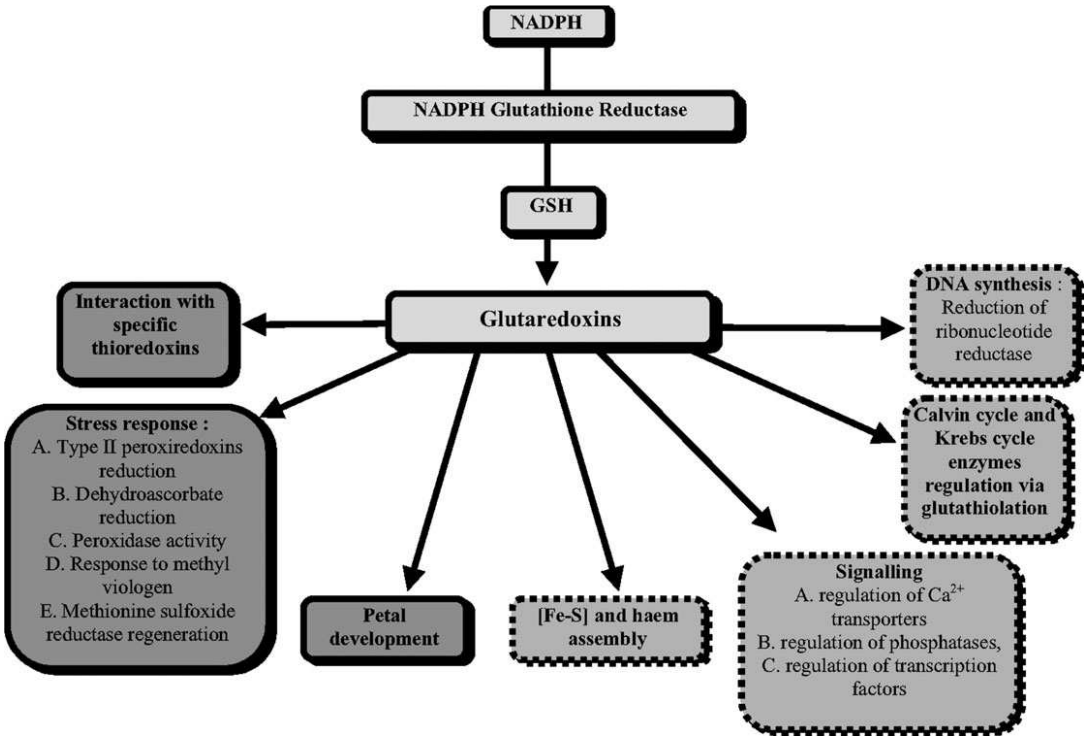


Figure 1.19 Confirmed or proposed roles for plant glutaredoxins. Grxs are mostly reduced by glutathione, which is in turn maintained in a reduced state with the help of glutathione reductase at the expense of NADPH. Established functions are on the left of the figure and are included within dark grey panels. Putative functions are on the right of the figure and are deduced from the known roles of Grx in other organisms. They are represented in dashed frames light grey panels (Rouhier et al. 2006).

The proteins of subgroup 2 possess CGFS active sites, but they differ in the number of repeated modules (one in Grx S14, S15 and S16 and three in GrxS17) and thus in their size, ranging from 170 to 492 amino acids. In addition to the Grx domain, Grx S16 possesses an additional module in the N-terminal part, which is related to a Trx domain, as it presents a WCDAS sequence, close to some atypical Trx active sites. Grxs from this subgroup were previously identified as a PICOT HD domain (for protein kinase C-interacting cousin of thioredoxin homology domain (Isakov et al. 2000), but some characteristics other than the active site sequences and especially a conserved motif [K/G]GE[L/F][I/V]GG[C/S], specific to Grx and present in the C-terminus part of these proteins, allow them to be identified as Grxs.

In *A. thaliana*, all the proteins of subgroup 3 possess active sites of the CC[M/L][C/S] form, except one with a CCLG active site (At1g03850) (Rouhier et al. 2004). Likewise in *P. trichocarpa*, only one sequence is divergent, with a CYMS active site. In *Oryza sativa*, the active site sequences vary greatly compared with *A. thaliana* and *P. trichocarpa*. Nevertheless, these atypical sequences are not restricted to *O. sativa*, since similar active site sequences are also found in Poaceae such as *Hordeum vulgare*, *Triticum aestivum* or *Zea mays* (Rouhier et al. 2006). Most of the proteins of subgroup 3 are extremely small, as they possess only 99-102 amino acids. This suggests that they should be essentially located in the cytosol, unless they can be imported into organelles without an N-terminal extension or signal peptides (Rouhier et al. 2006). Interestingly, one of these Grxs (*A. thaliana* monothiol Grx S15, accession number At3g15660), devoid of an N-terminal extension, was found in a mitochondrial proteomic study devoted to the identification of proteins able to bind divalent cations (Herald et al. 2003).

1.4.3.1.3 Poplar Grx

From the fully sequenced genome of *P. trichocarpa* (http://genome.jgi-psf.org/poplar0/poplar0.home.html), even though not fully annotated, there are at least 19 different Grx have been identified in the GenBank database by similarity search with *A. thaliana* Grxs (Rouhier et al. 2004). Thus far, surprisingly, one major difference is the lack of many isoforms with a -CCMS- active site. Otherwise, numerous classical Grxs of the CxxC and the CxxS groups are present, with most of them are putatively localized in the cytosol. Very few data are available concerning the distribution of Grx in the different plant organs, and even less is known about their intracellular localizations. However, based on the abundance of each Grx among the poplar expressed sequence tag (EST) in the database, the level of expression and the organ localization of each isoform can be estimated *in silico* (Rouhier et al. 2004). For example, PtGrx CxxC2 (32 ESTs) and CxxC4 (20 ESTs) are the two most abundantly expressed isoforms with CxxC2 predominantly expressed in flowers and CxxC4 in roots.

There is also one isoform (PtGrxS12) similar to Arabidopsis plastidial GrxC5 (AtCxxC5) or GrxS12 (AtCxxS12), but only one EST which is incomplete was found (Rouhier et al. 2004). Therefore, resulted incomplete protein of PtGrxS12 is 113 residues instead of 174 or 179 residues as compared to its corresponding homologs, AtCxxC5 and AtCxxC12, respectively. This isoform, one of the proteins of interest in this study, possesses of a WCSYS active site motif and an extra active-site cysteine at the C terminal of protein that is also conserved in certain plant Grxs isoforms regardless of their active site motif.

1.4.3.2 Structural biology

1.4.3.2.1 Grx structures known to date from all organisms

Many structures of dithiol Grx from human (PDB codes 1JHB, 2CQ9, 2FLS, 2HT9, 1B4Q), pig (PDB code 1KTE), mouse (PDB codes 1T1V, 1WJK), plant (PDB codes 1Z7P, 1Z7R, 2E7P), yeast (PDB code 2JAC), bacteria (PDB codes 1EGR, 1EGO, 1FOV, 1GRX, 1G7O, 1H75, 2AYT, 1J08, 1R7H, 2YWM), virus (PDB codes 2HZE, 2HZF) or T_4 bacteriophage (PDB codes 1DE1, 1DE2, 1AAZ, 1ABA, 1QFN, 3GRX) in oxidized and reduced forms have been resolved by nuclear magnetic resonance (NMR) spectroscopy or X-ray crystallography. Some of these proteins are ligated with its substrates or other proteins in the structures. Despite the abundance of monothiol Grxs reported, only one NMR structure is available in the databank (PDB code 1YKA). All models show a similar sequence of secondary elements. The active site is consistently situated at the beginning of the α-helix at the N-terminal of the enzyme. In all sequences, the amino acids proposed to be responsible for the binding of glutathione are conserved. However, when the Grx models are compared with those of Trx, it is apparent that the short Grxs lack the first β-strand and α-helix found at the N-terminus of Trx, but both proteins nevertheless share the same overall fold (Rouhier et al. 2006, 2008), named the Trx fold (Martin 1995).

1.4.3.2.2 Background on known plant Grxs

Very few biochemical and structural information are available for plant Grxs. Indeed, only two structures of plant Grxs of subgroup I (GrxC1 and C4) have been solved to

date both by NMR and X-ray crystallography (PDB codes 1Z7P, 1Z7R, 2E7P) (Rouhier et al. 2007, Feng et al. 2006, Noguera et al. 2005). The arrangement of secondary structures is quite similar to those of Trx, with an α1, β1, α2, β2, α3, β3, β4, α4, α5 organisation and the four β-strands forming a β-sheet flanked by five helices (Feng et al. 2006, Noguera et al. 2005). Compared to most prokaryotic Grxs, the poplar Grxs and other eukaryotic Grxs possess two supplementary α-helices in the N and C terminal regions. As in thioredoxin, the first active site cysteine is found at the beginning of the α-helix 2 and exposed to the solvent while the second cysteine, when present, is more buried. The Grx C4 can specifically self-associate into a monomer-dimer equilibrium, with the auto-association surface comprising the active site and the GSH binding site (Noguera et al. 2005). This dimer interface is different from the only one described so far in the *Haemophilus influenzae* Prx-Grx (*Hi*PrxGrx) hybrid protein, essentially because of the presence of this extra N-terminal helix (Kim et al. 2003, Noguera et al. 2005). The Prx-Grx interface of the *Hi*PrxGrx protein is stabilized by electrostatic interactions involving negatively charged residues in the Prx domain and positively charged residues in the Grx domain. Those amino acid residues are conserved both in the poplar Grx C4 and Prx IIB, but the two important residues in Grx C4, a lysine in position 24 and an arginine in position 69, are located at the auto association surface, probably preventing the interaction with Prx when and if the protein is organized as a dimer. Although the formation of a homodimer is possibly not favored in the cell, the presence of this newly described interface supports other results obtained later with Grx C1, which bridged an iron sulphur cluster into a homodimer *via* two catalytic cysteines and two external glutathione molecules (Feng et al. 2006, Rouhier et al. 2007). Nevertheless, the quaternary structure is different since in Grx C4, the monomers are arranged in a

head-to-tail orientation, while in Grx C1 they are in a face-to-face conformation (Rouhier et al. 2008b). Among the poplar subgroup I Grxs, only Grx C1 forms an iron–sulfur cluster. From mutagenesis results and analysis of the structures, it appeared that the side chain of the active-site proline residue in Grx C2, C3, or C4 (C**P**FC or C**P**YC active sites) likely prevents the incorporation of an iron sulfur cluster (Rouhier et al. 2007, 2008). This suggests that the proline residue may have effect on the geometry of the active site which hinders the binding of iron. Indeed, the replacement of this proline by a glycine in Grx C2, C3 and C4, allows incorporation of the cluster as in Grx C1 (Roubier et al. 2007). In light of these results, the presence of a small non-bulky residue (notably a glycine) and the two tyrosine residues surrounding the active site, allow to circumscribe a small pouch essential for Fe–S cluster incorporation.

Research Objectives

CHAPTER 2

RESEARCH OBJECTIVES

2.1 Overall research objectives

Maintenance of normal intracellular redox status plays an important role in processes such as DNA synthesis, gene expression, enzymatic activity and others. Alteration in the redox status of intracellular content and individual molecules, resulting from stress or intrinsic cellular activity, are involved in the regulation of different processes in cells. The challenge is even greater for land plants when submitted to adverse environmental conditions (biotic and abiotic stresses). In response to the oxidative stress, low molecular weight compounds such as GSH and ascorbic acid, and other cellular redox detoxification proteins such as Trx, Grx and various oxidoreductases in plants have been shown to be functionally involved to avoid excessive biological damages and maintaining cell redox homeostasis. In line with this, LCM^{3}B has established a long-term collaboration with the laboratory of Professor Jacquot in Nancy, France (Unité Mixte de Recherche INRA-UHP 1136, Interactions Arbes/Micro-organismes IFR 110 GEEF, Faculté des Sciences et Techniques, Nancy Université), which the latter group is particularly interested in the redox regulation systems in plants especially for poplar hybrids. My role in this collaborating project is to contribute to the elucidation of the structure function relationships (machinery/interaction with their potential *in vivo* partners) *via* resolution of X-ray crystal structures of three proteins (see below) involved in redox-

dependent systems for the elimination of ROS for a better understanding of reactional mechanisms within the network of cellular detoxification in poplar.

In plants, key proteins involved in the defense mechanism normally belong to multigenic families which are present in various cellular compartments. The complete sequencing of the genome of *A. thaliana* and the very recent poplar genome indicates that there are at least 26 genes encoding for thioredoxins (Trxs), 29 for glutaredoxins (Grxs) and at least 10 for peroxiredoxins (Prxs) (Meyer et al. 2007, Rouhier et al. 2004, 2006, 2008). One may wonder the need and necessity of such redundancy among these proteins, whereas bacterial and animal genomes have only a few isoforms (3 to 6 depending on the enzymes considered). A draft response could be that specific recognition between various physiological reductants and their targets is important for plant cells, taking into consideration that the expression and localization of these isoforms are different. Moreover, there is a necessity for land plants to have these various enzymes to protect them against the oxidative burst in a constantly changing environment. To provide answers regarding possible specific interactions, extensive biochemistry work had been carried out in recent years by the team of Professor Jacquot in order to characterize the enzymatic mechanisms employed by different physiological reductants including Trxs and Grxs, concomitantly with their potential target enzymes (for examples, Prxs and methionine sulfoxide reductases), by using the poplar hybrid as a study model. Briefly, the approaches include the production of recombinant proteins in *E. coli* system and site-directed mutagenesis studies to determine the catalytic mechanism of these proteins and their interaction with other proteins. Nevertheless, the clarity of the understanding of these life mechanisms seems incomplete without the structural

information of proteins involved. Therefore, the work of my thesis strengthens our knowledge on the subject of interest.

The current study is focusing on the 3-dimensional X-ray structural study of the glutathione peroxidase 5 (PtGpx5), thioredoxin *h*4 (PtTrx*h*4) and glutaredoxin S12 (PtGrxS12) from *Populus trichocarpa* × *deltoides* hybrid.

2.1.1 Gpx

"Glutathione peroxidases" (Gpxs) are a group of non-heme enzymes that play an important role in protecting cells against oxidative damage generated by ROS. Plant Gpxs possess similar primary structures to those of mammalian Gpx enzymes, except that most of the mammalian Gpxs contain selenocysteine at their active sites whereas most plant Gpxs do not (Ren et al. 1997). Interestingly, the plant Gpxs are not dependent on glutathione but rather on Trx as their *in vitro* electron donor. Therefore, we are interested particularly in understanding the properties and the catalytic mechanism of plant "glutathione peroxidase" with above-mentioned particularity. From the mutagenesis study, the catalytic cysteine (Cys44) of poplar Gpx5 was identified (Navrot et al. 2006). Amino acids sequence of this enzyme shows the existence of two other cysteine residues (Cys73 and Cys92) which only one of them (Cys73) is conserved in its homologues of all organisms. In solution, this protein has been shown to be able to form different intra- and inter- disulfide bonds. Hence, the resolving cysteine can be indentified from biochemical approaches with incertitude. One thing that can be concluded from the biochemical study by our collaborators is that this enzyme does form an intra-disulfide bond during the catalytic mechanism.

Therefore, the main objective of this sub-project is to identify, in subatomic structural level, the resolving cysteine of this enzyme.

Navrot et al. (2006) have reported that the reduction/formation of the intramolecular disulfide bond in poplar Gpx5 is detected by a shift on sodium dodecyl sulfate-polyacrylamide gel electrophoresis and by measuring the intrinsic tryptophan fluorescence of the protein. These results suggested that there might be conformational changes during the redox catalysis of the enzyme. Thus, it is also the aim of this study to identify the conformational changes, if any, associated with the catalysis (transition from reduced to oxidized states or vice versa). Thus, the resolution of both reduced and oxidized forms structures of the enzyme will considerably clarify the incertitude arise from biochemical study. Initial biochemistry work also has doubts on the oligomerization state of this protein. Therefore, it was also initially the aim of this structural study to provide some clues on the layout of the quaternary structure of the enzyme that may represent its physiological functional nature. Interaction of this enzyme with its reductant (e.g. Trx) in structural level is also the interest of this project.

2.1.2 Trx

Thioredoxin (Trx), which contains an active site with the sequence motif WCxxC, catalyzes fast and reversible dithiol-disulfide exchange reactions between the active-site cysteine residues and half-cysteines of the disulfide substrates. In classical Trx, the two active-site cysteines are non equivalent, with the N-terminus one (next to the conserved Trp residue) being the catalytic cysteine (Rouhier et al. 2008a). The heterodisulfide is reduced by the back-up C-terminus active-site cysteine, leaving the

target in the dithiol form whereas Trx in oxidized state with a disulfide bond. The reduced state of Trx is then regenerated from NADPH-thioredoxin reductase (NTR) (Sandalova et al. 2001, Lennon et al. 2000, Dai et al. 1996) or photoreduced ferredoxin in the plastids (Dai et al. 2007).

PtTrx*h*4 belongs to the plant *h*-type Trx of subgroup 3 and it contains the usual catalytic site ($WC_{58}GPC_{61}$) but has been shown to differ from previously characterized Trx *h* in terms of enzyme recycling pathway used. Instead of using the classical NTR reducing system, this isoform is reduced *in vitro* by GSH/Grx reducing system (Gelhaye et al. 2003). In addition to the two active-site cysteines, PtTrx*h*4 possesses a third cysteine localized in the N-terminal extension (Cys4). Cys4 is conserved in all other PtTrx*h*4 homologs. Site-directed mutagenesis study by Gelhaye et al. (2003) suggested that Cys4 intervenes in the catalytic mechanism of this atypical Trx isoform. However, the exact mechanism or the intervention order of each cysteine during the catalysis remains unclear. Therefore, structural data are needed to further explain the mechanism in line with concurrent biochemical studies. In order to achieve this, PtTrx*h*4 and its mutant PtTrx*h*4C61S proteins had been provided by Dr. Eric Gelhaye (member of Prof. Jacquot's group) and subjected to X-ray crystallization study (part of my thesis work) so that we are able to have a clearer view of the catalytic mechanism of this unusual biological reductant by combining information from different research approaches. Resolution of the wild type and the mutant of PtTrx*h*4 structures should allow us to localize the peculiar N-terminal extension and at the same time to identify, if any, the alternative intramolecular disulfide of the enzyme. From the redox titration experiments, it is not possible to attribute the exact redox potential to each disulfide bond because in the solution PtTrx*h*4 seems to form various different intra- and inter-molecular disulfide bonds

(E. Gelhaye, *unpublished results*). Thus, it is important to have some structural data particularly of the PtTrx*h*4C61S mutant structure to justify the possibility of any intra- (Cys4-Cys58) or intermolecular (Cys4-Cys4 / Cys58-Cys58) disulfide formation. Combining biochemistry and crystallographic studies, we hope to have a better understanding of the catalytic mechanism of PtTrx*h*4.

2.1.3 Grx

Grx, a structural and functional homolog of Trx, displays a common Trx-fold and an active site involving one or two conserved cysteines. Compared to Trxs, Grxs are considered to be less efficient reductants, because they possess a higher redox potential (approximately -170 mV compared to -300 mV for Trxs) (Rouhier et al. 2007). However, Grxs have the ability to catalyze very efficiently deglutathionylation reactions, for example the reduction of a protein-GSH mixed disulfide, reactions for which Trxs are poor catalysts (Rouhier et al. 2008a).

The classical CxxC motif in Trx and Grx has been previously characterized in great details. Interestingly, some Grxs are known to contain only the N-terminus cysteine in the CxxC motif, whereas the second cysteine is replaced with serine (forming natural CxxS active site). This variation of active site sequence does exist in plants (Rouhier et al. 2004, 2006). In this study, we are interested in deciphering the structure of a Grx isoform (PtGrxS12) from *P. trichocarpa* which possesses unique active site sequence, $WC_{29}SYS$. In addition of having a natural CxxS active site motif, there is an additional cysteine found to be present at 54 residues after the active site of PtGrxS12. The role of this additional active-site cysteine in the monothiol- or dithiol-containing Grx remains obscure (Rouhier et al. 2004). This

protein is supplied by Dr. Nicolas Rouhier (from Prof. Jacquot's group). To date, there is no published biochemical data on this isoform. However, from the preliminary biochemistry data Cys29 has been identified as the catalytic cysteine. In general, this enzyme is regenerated by GSH/GSH reductase system. In some "monothiol" Grxs, the formation of an intramolecular disulfide bond between the catalytic cysteine and additional active-site cysteine had been reported. Thus, it is of interest to know if this is also the case for PtGrxS12. In parallel with the biochemistry study by Rouhier and co-workers, it is one of the objectives of this structural study to investigate the possible role(s), if any, of the additional cysteine (Cys87) of this Grx isoform from crystal structures. In addition, the interaction of this Grx isoform with its substrates *in vitro* (HED) and *in vivo* (GSH) is also the interest of this study so that we can have a better understanding of protein-substrate interactions in subatomic level.

Methods and Materials

CHAPTER 3

METHODS AND MATERIALS

In this study, eight crystal structures of three different proteins have been solved (summarized in Table 3.1). All the purified proteins used in this crystallographic study were provided by the plant biology group from Prof. Jean-Pierre Jacquot's laboratory in Nancy, France (Unité Mixte de Recherche INRA-UHP 1136, Interactions Arbes/Micro-organismes IFR 110 GEEF, Faculté des Sciences et Techniques, Nancy Université). The production protocols of each studied proteins is published or will be published elsewhere: poplar "glutathione peroxidase" 5 (Navrot et al. 2006, Koh et al. 2007), poplar thioredoxin *h*4 (Gelhaye et al. 2003, Koh et al. *submitted to J. Biol. Chem.*) and poplar glutaredoxin S12 (Koh et al. *in preparation*), hence will not be mentioned in this manuscript since the works were done by the collaborators laboratory members. Precise protein purification steps can also be found in papers mentioned above. All protein samples are provided in the TE solution (30 mM Tris-HCl pH 8.0, 1 mM EDTA). The complexes of poplar glutaredoxin with GSH or GSH/HED (HED, β-hydroxyethyl disulfide is a conventional substrate of Grxs) were prepared upon purification steps. Briefly, the complex with GSH was prepared by the addition of 2 mM of GSH to the purified recombinant active protein upon purification for 30 minutes at room temperature prior to extensive dialysis procedures with TE buffer to remove unbound GSH. As for the latter complex, the same preparation with GSH was performed before the addition of HED in excess (10 mM) to the glutathionylated wild-type PtGrxS12

mixture for two hours incubation time at room temperature. Then the mixture was extensively dialyzed against the TE buffer.

Table 3.1 Summary of X-ray crystal structures of poplar oxidoreductases solved in this study.

Proteins of poplar	State of protein	Size (molecular weight)	Ligand	Denomination in this study
"Glutathione peroxidase" 5 (Gpx5)				
SeMet – Gpx5	Reduced form	170 amino acids (19.38 kDalton)	N. A.	SeMet-PtGpx5$_{red}$
Wild-type Gpx5	Reduced form	170 amino acids (19.36 kDalton)		WT-PtGpx5$_{red}$
Wild-type Gpx5	Oxidized form			WT-PtGpx5$_{oxd}$
Thioredoxin *h*4 (Trx*h*4)				
SeMet - Trx*h*4	Oxidized and truncated form	115 amino acids (13.09 kDalton)	N. A.	SeMet-PtTrx*h*4$_{oxd}$
Wild-type Trx*h*4	Oxidized form	139 amino acids (15.65 kDalton)		WT-PtTrx*h*4$_{oxd}$
C61S Mutant of Trx*h*4	Oxidized form	139 amino acids (15.63 kDalton)		C61S-PtTrx*h*4$_{oxd}$
Glutaredoxin S12 (GrxS12)				
Wild-type GrxS12	Oxidized form 1	113 amino acids (12.49 kDalton)	GSH	WT-PtGrxS12$_{oxd[GSH]}$
Wild-type GrxS12	Oxidized form 2		GSH + HED	WT-PtGrxS12$_{oxd[GSH+HED]}$

Abbreviation: GSH, glutathione; HED, 2-hydroxyethyl disulfide; N. A., not applicable

3.1 Crystallogenesis

In this study, random crystallization screening method was used for initial attempts to obtain crystals. Various crystallization screens were used: commercial screening matrices from Hampton Research (SalTrx™, PEG/Ion™, PEG/Ion 2 Screen™, Crystal Screen™ and Crystal Screen 2™); from Jena Biosciences (JBScreen

Classic); Emerald BioSystems (Wizard I™ and Wizard II™). All crystallization conditions used in this study are summarized in Table 3.2.

All preliminary crystallization screenings were carried out by using the microbatch under oil (paraffin) method. Then, if needed, further optimizations to obtain crystals suitable for X-ray diffraction studies were carried out by employing the hanging drop method. In the microbatch method, paraffin oil (Merck KGaA, Germany) was filled into the wells of a 72-well Terasaki/HLA plate for microbatch (Molecular Dimensions Limited, England or Greiner Bio One, Dutscher, France) with a disposable Pasteur pipette (each well should be fully filled with paraffin). Then 1.5 µl of commercial precipitating solution was added to a well prior to the addition of the same amount of protein solution with known concentration. Once the screen was completed, the microbatch plate was centrifuged for 1 minute under 4000 *g* in order for the drops to settle down at the bottom of the wells. The plate was stored in the 20°C room for crystal growth.

For the optimization stage, hanging drop vapor diffusion method was used. A 24-well XRL™ ("Linbro style" raised lid) plate (Molecular Dimensions Limited, England) that had a fine bead of vacuum grease applied around the edge of the well was utilized. Commercial or homemade precipitating solutions (500 µl each) that were composed of precipitant, buffer and/or additive were pipette dispensed into the reservoirs of the crystallization plate individually. Then, different ratio of protein-reservoir solutions were pipette dispensed and mixed on the center of a clean siliconized 22 mm circle glass cover slide. Immediately, the cover slide was inverted with care and sealed over the reservoir with gentle pressure to ensure a proper seal. The plate was stored in the 20°C room for crystal growth.

Table 3.2 Crystallization conditions of obtaining crystals from poplar proteins of interests suitable for X-ray diffraction studies.

	PtGpx5		
	SeMet-PtGpx5$_{red}$	WT-PtGpx5$_{red}$	WT-PtGpx5$_{oxd}$
Initial condition to obtain crystals	Hampton Crystal Screen 2; solution 34 (0.1 M Hepes (pH7.5), 0.05 M cadmium sulphate hydrate, 1.0 M sodium acetate)	Hampton Crystal Screen 2; solution 34 (0.1 M Hepes (pH7.5), 0.05 M cadmium sulphate hydrate, 1.0 M sodium acetate)	JBScreen 2; solution C4 (0.1 M Tris-HCl (pH 8.5), 25 % (w/v) PEG 4000, 0.2 M calcium chloride)
Duration to obtain crystals	1 week	1 week	2 days
Drop buffer to protein ratio	1 : 1	1 : 1	1 : 1
Initial protein concentration (mg/ml)	20	20	20
	PtTrx*h*4		
	SeMet-PtTrx*h*4$_{oxd}$	WT-PtTrx*h*4$_{oxd}$	C61S-PtTrx*h*4$_{oxd}$
Initial condition to obtain crystals	Hampton SalTrx Screen 2; solution 54 (1.0 M sodium / potassium (pH 6.9)) with 1 mM DTT added to the initial protein solution	JBScreen 2; solution D2 (0.1 M Na HEPES (pH 7.5), 30 % (w/v) PEG 4000, 0.2 M $CaCl_2$) with 1 mM DTT added to the initial protein solution	Hampton SalTrx Screen 1; solution 55 (1.0 M sodium / potassium (pH 8.2))
Duration to obtain crystals	1 month	1 month	1 week
Drop buffer to protein ratio	1 : 1	1 : 1	1 : 1
Initial protein concentration (mg/ml)	20	20	25
		PtGrxS12	
		WT-PtGrxS12$_{oxd[GSH]}$	PtGrxS12$_{oxd[GSH+HED]}$
Initial condition to obtain crystals		JBScreen 5; solution B4 (0.1 M Na HEPES (pH 7.5), 20 % PEG 8000)	JBScreen 1; solution C3 (0.1 M Na HEPES (pH 7.5), 25 % PEG 1000)
Duration to obtain crystals		1 day	1 day
Drop buffer to protein ratio		1 : 1	1 : 1
Initial protein concentration (mg/ml)		15	15

3.2 Data collection and reduction

3.2.1 Beamline used

In recent years, number of X-ray synchrotron beamlines dedicated to collecting diffraction data from macromolecular crystals has exceeded 50 (Dauter 2005). Indeed, today most protein crystal structures are solved and refined based on the synchrotron data as the synchrotron allows its users to perform experiments not possible otherwise. Each beamline has different properties with regards to intensity and tunability. Collecting diffraction data from synchrotron facilities has become significantly easier in recent years. Several factors such as popularity of crystal cooling and transporting of frozen crystals, the advent of the automatic crystal-mounting robots, the routine use of the fast CCD detectors as well as the availability of fast and powerful computers, contributed to this progress. For this study, data collections have been carried out in two European synchrotron facilities: the ESRF in Grenoble (France) and the DESY/EMBL in Hamburg (Germany). Here, I summarize several important blueprints of five different beamlines (which I've used throughout the study; at the time of writing this manuscript) at the mentioned synchrotrons in Table 3.3.

Table 3.3 Important specifications on macromolecular crystallography beamlines used for data collections in this study.

Location	DESY / EMBL Hamburg, Germany http://www.embl-hamburg.de/services/mx/index.html		
Beamline	X11	X13	BW7A
Synchrotron source	Bending magnet	Bending magnet	1 Tesla multi-pole wiggler (28 poles, 14 cm period)
Mirror 1	N. A.	N. A.	Rh coated, zerodur, vertical focusing possible
Monochromator	Si [111], horizontally focusing	Si [111], horizontally focusing	Fixed exit double crystal Si [111], horizontally focusing
Mirror 2	Bent, vertically focusing	Bent, vertically focusing	Rh coated, zerodur, vertical focusing
Wavelength	0.81 Å	0.81 Å	0.7 to 1.8 Å
Focal spot	2 mm x 0.4 mm	2 mm x 0.4 mm	2 mm x 0.4 mm (varies with energy)
Slits	0.1 x 0.1 to 2 x 2 mm	0.1 x 0.1 to 2 x 2 mm	0.05 x 0.05 to 1 x 1 mm
Cryo-system	Oxford 600 series	Oxford 700 series	Oxford 600 series
MAD facility	No	No	Yes (range of 7000 - 18000 eV, encompassing elements Fe to Y)
Detector	MAR CCD 165 mm	MAR CCD 165 mm	MAR CCD 165 mm
Goniometer	Single axis	Single axis	Single axis
Cryo-shutters for crystal annealing	Yes	Yes	No
Online spectroscopic measurements	No	No	Yes
Robotic sample charger	No	No	No
Crystal centering	Manual	Manual	Manual
Maximum resolution achievable	d_{min} = 0.8 Å	d_{min} = 0.8 Å	d_{min} = 0.6 Å (without vertical left of the detector)
Detector swing-out angle	N. A.	N. A.	Vertical lift up to 8 cm
Ancilliary equipment	No	No	Online microspectrometer

Table 3.3. Continued.

Location	ESRF Grenoble, France http://www.esrf.eu/UsersAndScience/Experiments/Beamlines	
Beamline	BM30A	ID14-3
Synchrotron source	Bending magnet	Three undulators in tandem on a high β section
Mirror 1	vertical focusing	Toroidal mirror
Monochromator	Double crystal monochromator [Si (111), Si (311)]: horizontally focussing by sagittal curved 2nd crystal	Double crystal monochromator [Diamond (111), Ge (220)]
Mirror 2	vertical focusing	Toroidal mirror
Wavelength	0.59 - 1.77 Å	0.931 Å
Focal spot	0.3 mm x 0.3 mm	338 µm x 23 µm
Cryo-system	Oxford 600 series	Oxford 700 series
MAD facility	Yes (range of 7.0 – 21.0 keV)	Yes (range of 7.0 – 14.5 keV)
Detector	MAR 345 image plate or MAR 165 CCD	ADSC Q4R
Goniometer	5 circle diffractometer	Microdiffractometer with on axis visualization and semiautomatic sample centering
Cryo-shutters for crystal annealing	Yes	No
Robotic sample charger	Yes (sample has to be mounted on magnetic caps)	Yes (storage capacity of 50 samples)
Crystal centering	Yes (automated crystal centering device)	Yes (semi-automatic sample centering device)
Maximum resolution achievable	0.59 Å	1.0 Å
Ancilliary equipments	Roentec fluorescence detector, 3 stereo microscopes LEICA, cold room on beamline, Hampton cryo tools, magnetic mini-agitator etc.	microdiffractometer with on axis visualization and semi or fully automatic sample centering

Abbreviation: N. A., not applicable

The European Synchrotron Radiation Facility, ESRF (or *Installation Européenne de Rayonnement Synchrotron*) is among the three largest and most important functional third-generation synchrotrons in the world currently, in line

with the APS (Advanced Photon Source) in Argonne (Chicago, USA) and the Spring-8 in Harima Science Garden City (Hyogo, Japan). This facility is situated at the scientific polygon of Grenoble (France). The ESRF is Europe's leading synchrotron light source with the participation of 18 countries. To meet growing demand, ever more beamlines have been added over the years – now reaching a total of 43 (for details, see http://www.esrf.eu/UsersAndScience/Experiments/Beamlines) – some general-purpose, but mostly designed to carry out more sophisticated kinds of measurements developed for particular fields of study. Five areas of growing importance have been identified in terms of user demand and science delivery: (i) nanoscience and nanotechnology, (ii) pump-probe experiments and time-resolved diffraction, (iii) X-ray imaging, (iv) structural and functional biology and soft matter, and (v) science at extreme conditions. With the present configuration, there is no further space around the ring to add more. In ESRF, the macromolecular crystallography group supports the study and structure determination of biological macromolecules. The facilities available consist of three fixed energy beamlines (ID14-1, ID14-2 and ID14-3), three tunable MAD capable beamlines (ID14-4, ID23-1 and ID29) and a fixed energy microfocus beamline (ID23-2). A number of laboratories and facilities are available to the community. Of particular interest is The Partnership for Structural Biology (PSB) which is the collaboration between ESRF, the European Molecular Biology Laboratory (EMBL), Institut Laue-Langevin (ILL, a world-leading neutron-beam facility) and Institut de Biologie Structurale (IBS) to bring together a set of complementary technologies for structural biology. Other collaborating research groups and beamlines include the beamlines ID13, BM14, BM16, BM26, BM30A and BM30B.

The German Synchrotron Research Centre, DESY (Deutsches Elektronen-Synchrotron), member of the Helmholtz Association, is one of the leading accelerator centers in the world. DESY is a national research center supported by public funds and has locations in Hamburg and Zeuthen (Brandenburg). DESY conducts basic research in the natural sciences with special emphasis upon accelerators, particle physics and research with photons. DESY hosts leading facilities for synchrotron radiation (namely DORIS-III, in operation; PETRA-III, under construction, user operation planned for 2009) and electron lasers (namely VUV-FEL, commissioned; X-FEL, planned). The EMBL Hamburg unit is situated on the campus of DESY. This unit is part of the EMBL with the main laboratory located in Heidelberg and research units in Hamburg, Grenoble, Hinxton Hall (Cambridge) and Monterotondo (Rome). Activities at EMBL Hamburg focus on state-of-the art structural biology methods using synchrotron radiation. At present, EMBL Hamburg operates seven experimental stations using synchrotron radiation from the DORIS-III ring. Five of these stations are for applications in Macromolecular Crystallography (MX); two of them are energy-tunable (beamlines X12 and BW7A), two of the MX end stations are equipped with automatic sample changers (beamlines X12 and BW7B) while two others with cryo-shutters for crystal annealing (beamlines X11 and X13). The beamline BW7A is the only one that is furnished with a UV/VIS microspectrophotometer for online spectroscopic measurements. The other two experimental stations are for applications in Small Angle X-ray Scattering (SAXS) of biological material (beamline X33) and in X-ray Absorption Spectroscopy (XAS). In addition, the EMBL Hamburg Unit has set up a biochemistry laboratory hosting a pipeline for sample preparation and

characterization. This includes facilities for high-density cell fermentation, semi-automated protein purification and mass spectrometry.

3.2.2 Crystal handling

All data collection experiments have been carried out at 100K using the synchrotron beamline facilities mentioned above. Normally, protein crystals grow to a limiting size and subsequently degrade. Therefore, crystals need to be cryo-protected or flash-cooled at their finest conditions until data collection is performed. The main challenge of cooling the crystal to cryogenic temperatures is to avoid the formation of crystalline ice within the sample. The goal is to obtain amorphous ice rather than crystalline ice and to reduce the damage on the crystal by the cryofreezing process. Usually, cryoprotectants are added to promote the result. However, one have to bear in mind that in practice, cryoprotectants that work with one protein do not work with another, thus requiring a trial-and-error search. This process is often time consuming and in some cases unsuccessful.

Cryoprotectant used in this study is glycerol where 20 % of glycerol is added into the mother liquor of the crystallization solution that crystals were obtained initially. The free-standing loop mounting technique was used in this study. Cryo-loops from different companies (mounted Litholoops from Molecular Dimensions Limited, England and mounted Cryoloops from Hampton Research, USA) were used, made from fine nylon fibers, copper or tungsten wires with diameters ranging from 0.05 mm to 0.5 mm depending on the mounted crystal size. Briefly, the technique I've employed was as follows: with a loop, crystal was "fished out" from the microbatch well. It was preferably to use a loop smaller than the crystal to spoon

it easily. In this way, less of the mother liquor is sticking to the crystal or the loop. If the crystallization condition contains appropriate amount of cryoprotectant, such as glycerol, ethylene glycol, PEG and MPD, then crystal can normally be flash-cooled directly under nitrogen gas stream of the diffractometer just before data collection. Another approach used in this study was the use of immiscible oil such as Paratone-N to replace the external liquid around the crystal. It was not too difficult to get rid of residual mother liquor just by passing the crystal back and forth though the cryoprotectant or oil interface. Basically, the oil "glues" the crystal to the loop. After crystal mounting, the loop was plunged in liquid nitrogen or freeze in a stream, ready for data collection.

In principle, once the crystals have been successfully cooled to cryogenic temperatures, they can be kept for an indefinite time. Sometimes, crystals (still in its crystallization condition) were brought directly to synchrotron facilities for data analysis without additional cryo-protection procedure beforehand. In these cases, crystals were rapidly plunged beneath the surface of liquid nitrogen or subjected to other cryoprotectants on-the-spot at the experimental hall of the synchrotron. From experiences, on-the-spot flash-cooling usually, but not all the time, may increases the diffraction quality of the crystals (higher resolution data with lower mosaicities can be normally collected).

3.2.3 Data collection

Cryo-protected crystals were diffracted on the synchrotron X-ray diffraction facilities either at the ESRF, Grenoble (France) or at the DESY/EMBL, Hamburg (Germany) synchrotron beamlines (Details in section 3.2.1 and Table 3.4). The crystals were

mounted on the goniometer head using a cryo-loop. All data collections were carried out at 100K.

After crystal mounting, the crystal was aligned to the center of the X-ray beam and the detector was placed to an appropriate distance. Most of the synchrotrons provide automated crystal centering and automated detector displacement facilities. For X-ray diffraction experiments, the crystals were exposed to X-ray for 5 seconds to 60 seconds to collect an image, depending on the beam intensity and crystal diffraction power (diffraction patterns). If there was a diffraction pattern shown on the image (with distinct diffraction spots), an initial image was collected to determine the cell dimensions of the crystal. The exposure time of a crystal to the X-ray beam differed with the crystal and beamline. It can be increased to intensify the diffraction spots. The stronger the intensities of the diffraction spots, the more reliable of the unit cell dimensions could be determined. However, one should take into consideration the overloading of each spot (peak) in a diffracted image which could complicate the data processing later. After the exposure time, the diffraction resolution limit of crystal was determined and analysis of diffraction pattern was analyzed by XDS (Kabsch 1993) or *HKL*2000 softwares (Otwinowski & Minor 1997) to determine the unit cell dimensions and space group of the crystal. Table 3.4 summarized all the data collection and processing statistics of protein crystals studied.

Table 3.4 Data collection and processing statistics.

Datasets	**PtGpx5**		
	SeMet-PtGpx5$_{red}$	WT-PtGpx5$_{red}$	WT-PtGpx5$_{oxd}$
Data collection date	17 November 2005	15 December 2004	21 February 2006
Data collection site	BM30A ESRF-Grenoble (France)	X11 DESY/EMBL-Hamburg (Germany)	BM30A ESRF-Grenoble (France)
Wavelength (Å)	0.9800	0.8156	0.8063
Detector to crystal distance (mm)	169.99	180.00	200.00
Oscillation range (°)	1.0	1.0	1.0
Number of images collected	200	160	120
Space group	R3	R3	$P3_121$
Unit cell dimensions (Å) (a, b, c)	222.19 222.19 48.36	221.67 221.67 48.14	71.58 71.58 117.82
Asymmetric unit	4	4	2
Resolution range (Å) (outer shell)	50.00 – 2.70 (2.85 – 2.70)	50.00 – 1.95 (2.02 – 1.95)	50.00 – 2.45 (2.49 – 2.45)
Redundancy (outer shell)	6.50 (6.50)	5.04 (4.52)	7.60 (7.79)
Completeness (%) (outer shell)	100.00 (100.00)	99.8 (99.0)	99.8 (100.0)
I / σ I (outer shell)	21.30 (7.70)	11.48 (2.20)	17.24 (2.91)
R_{sym} (outer shell)	0.079 (0.300)	0.066 (0.409)	0.062 (0.335)
Figure of merit	0.342/0.665 (MLPHARE/DM)	-	-

Table 3.4. Continued.

Datasets	SeMet-PtTrx*h*4$_{oxd}$	PtTrx*h*4 WT-PtTrx*h*4$_{oxd}$	C61S-PtTrx*h*4$_{oxd}$
Data collection date	26 April 2006	16 March 2005	15 March 2005
Data collection site	BM30A ESRF-Grenoble (France)	X11 DESY/EMBL-Hamburg (Germany)	BW7A DESY/EMBL-Hamburg (Germany)
Wavelength (Å)	0.9805	0.8123	1.2400
Detector to crystal distance (mm)	200.00	179.89	60.00
Oscillation range (°)	1.0	0.2	0.5
Number of images collected	180	720	360
Space group	$P4_12_12$	$P6_1$	$P2_12_12_1$
Unit cell dimensions (Å) (a, b, c)	44.89 44.89 131.74	47.36 47.36 196.00	31.78 44.10 85.68
Asymmetric unit	1	2	1
Resolution range (Å) (outer shell)	32.94 – 2.46 (2.60 – 2.46)	50.00 – 2.15 (2.23 – 2.15)	30.00 – 1.60 (1.66 – 1.60)
Redundancy (outer shell)	7.0 (3.6)	28.31 (37.63)	5.20 (5.32)
Completeness (%) (outer shell)	93.4 (65.7)	99.8 (100.0)	98.5 (98.4)
I / σ I (outer shell)	17.85 (3.10)	22.98 (2.60)	16.03 (6.20)
R_{sym} (outer shell)	0.089 (0.588)	0.053 (0.392)	0.092 (0.266)
Figure of merit	0.34824 / 0.06270 (acentric / centric)	-	-

Table 3.4. Continued.

Datasets	PtGrxS12 WT-PtGrxS12$_{oxd[GSH]}$	PtGrxS12$_{oxd[GSH+HED]}$
Data collection date	22 August 2007	23 February 2006
Data collection site	X11 DESY/EMBL-Hamburg (Germany)	X13 DESY/EMBL-Hamburg (Germany)
Wavelength (Å)	0.8150	0.8063
Detector to crystal distance (mm)	151.71	161.01
Oscillation range (°)	0.5	1.0
Number of images collected	224	180
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å) (a, b, c)	39.03 47.27 55.62	38.83 46.82 55.36
Asymmetric unit	1	1
Resolution range (Å) (outer shell)	50.0 – 1.7 (1.73 – 1.70)	50.0 – 1.8 (1.86 – 1.8)
Redundancy (outer shell)	4.49 (4.16)	6.69 (6.85)
Completeness (%) (outer shell)	99.7 (95.0)	99.6 (100.0)
I / σ I (outer shell)	13.37 (2.26)	23.26 (16.95)
R_{sym} (outer shell)	0.045 (0.239)	0.048 (0.107)
Figure of merit	-	-

3.2.4 Data reduction

The diffraction images of the different crystals were indexed, integrated and scaled using either the *HKL*2000 program or the XDS program package then analyzed using the CCP4 software package version 6.0.2 (CCP4: Collaborative Computational Project Number 4 1994).

3.3 Structure solving

3.3.1 Phasing method using particular SAD approach: a case study for poplar thioredoxin-dependent peroxidase 5

In this study, the crystal structure of PtGpx5 was solved by SAD approach using the Auto-Rickshaw platform. Auto-Rickshaw is the EMBL-Hamburg automated crystal structure determination platform (Panjikar et al. 2005) which combines a number of existing macromolecular crystallographic computer programs and several decision-makers into a software pipeline for automated and efficient crystal structure determination. The objective of this pipeline is to make the “structure on the fly” concept possible at the beamlines as soon as X-ray data from protein crystals have been collected and processed. It is controlled by a web-based graphical user interface (http://webapps.embl-hamburg.de/cgi-bin/Auto-Rick/arinitAR.cgi) for data and parameter input and for monitoring the progress of structure determination. A large number of possible structure solution paths are encoded in the system and the optimal path is selected by the decision-makers as the structure solution evolves. The processes have been optimized for speed so that the pipeline can used effectively for validating the X-ray experiment at a synchrotron beamline. Currently, the platform is

restricted to SAD, single isomorphous replacement and anomalous scattering (SIRAS), 2 wavelengths-MAD (2W-MAD), 3 wavelengths-MAD (3W-MAD) or 4 wavelengths-MAD (4W-MAD) phases determination. Auto-Rickshaw comes in two flavors: a “Beamline Version” and an “Advanced Version”. Both versions use a similar GUI input, but the “Advanced Version” requires the sequence information for the protein target in order to build the side chains. The “Beamline Version” is explicitly geared towards its use for validation of the X-ray experiment at the synchrotron as soon as the data have been collected and processed. Once the X-ray experiment is validated, the “Advanced Version” can be used for a more complete model building if the resolution of the data permits. The “Advanced Version” of the platform uses ARP/wARP (Perrakis et al. 1999, Morris et al. 2004). The platform is installed on a 16-processor Linux cluster to ensure an efficient use of the beamline available. Despite its robustness and convenience, this platform is, however, only available to EMBL staff and Hamburg beamline users.

MAD data were initially collected from PtGpx5 crystal to 2.2 Å resolution at the selenium edge. The 3W-MAD, 2W-MAD and SAD approaches were tested using Auto-Rickshaw at webserver (http://www.embl-hamburg.de/Auto-Rickshaw/), however only the single wavelength anomalous diffraction approach using the “Beamline Version” of Auto-Rickshaw was successful and resulted in an interpretable map with a partial model. The “Beamline Version” uses the data to 3.0 Å resolution whereas the “Advanced Version” of Auto-Rickshaw includes maximum resolution of the data. The version was nevertheless not successful with any phasing protocol; this observation indicated that the crystal might have suffered severely from radiation damage. High-resolution reflections are more sensitive to radiation damage and such damage causes problems for scaling and merging the datasets, and

can prevent measurement of anomalous and dispersive differences that consequently affects the experimental phasing. Therefore, the peak data were reprocessed and rescaled. Data quality was found to be good to 2.7 Å for phasing. Overall processes that have taken place in the pipeline included: SHELXD (Schneider & Sheldrick 2002), which located eight expected selenium atoms present in the asymmetric unit; MLPHARE (CCP4: Collaborative Computational Project Number 4 1994) for initial phasing; DM (Cowtan 1994) for density modification and twofold NCS averaging; and ESSENS (Kleywegt & Jones 1997) to position α-helices in the structure. The initial phases were extended to 2.7 Å resolution using RESOLVE version 2.02 (Terwilliger 2000). Apart from above mentioned programs, Auto-Rickshaw also invokes other programs (which are not applicable for current case study) for its decision makers: CCP4 (CCP4: Collaborative Computational Project Number 4 1994), SHELXC (Sheldrick et al. 2001), SHELXE (Sheldrick 2002), NANTMRF (Smith 2002), BP3 (Pannu et al. 2003, Pannu & Read 2004), SHARP (La Fortelle & Bricogne 1997), MAPMAN (Jones & Thirup 1986), ARP/wARP and HELICAP (Morris et al. 2004). Initially, molecular replacement (MOLREP) using the closest structure (bovine erythrocyte GPX, PDB code 1GP1) was unsuccessful against the native dataset; however, the model was placed successfully in electron density using phased molecular replacement. The resultant model was checked on graphics for the crystal packing, which clearly showed that a long loop (19 residues) clashes with the neighbouring residues; therefore, the loop was removed and the phased molecular replacement was repeated. This gave the best fit to the experimental electron density. The model was used as a starting point for further manual model building using the graphic program Coot (Emsley & Cowtan 2004).

Once the model was 90 % complete, positional and B factor refinements were performed using the protocols implemented in CNS, version 1.1 (Brünger et al. 1998). The partially refined SeMet-PtGpx5$_{red}$ model was used as a search model in a molecular replacement procedure using MOLREP (Vagin & Teplyakov 1997) against the native data, and the resultant model was used for automatic model building using the program ARP/wARP: 99% of the model was built automatically. Further, the model was improved by manual building with Turbo-Frodo (Roussel & Cambillau 1989) interspersed with refinement using CNS. At this stage, it became evident that the PtGpx5 crystal contained additional heavy atoms. The calculation of anomalous difference Fourier map indicated 32 cadmium ions bound to the protein in one asymmetric unit. Some cadmium were expected to bind the protein, as the cadmium sulphate hydrate was used during crystallization; however such a large number of cadmium ions bound to the protein is generally not common (see results chapter). In addition, eight acetate anions, which was also a component of the crystallization solution, were added to the final WT-PtGpx5$_{red}$ structure before several refinement steps using both REFMAC version 5.4 (CCP4: Collaborative Computational Project Number 4 1994) and CNS version 1.1.

A model of the reduced form of native enzyme was used for molecular replacement to solve the structure of the oxidized form of enzyme. The solution was further refined using CNS, and iteratively combined with manual building with the program described above. Throughout the model building and refinement process, the qualities of the models were assessed using the program PROCHECK (Laskowski et al. 1993).

3.3.2 Phasing method using conventional sulfur-SAD: a case study for poplar thioredoxin *h*4

Initial molecular replacement attempts using relatively high sequence homology protein model as template was fruitless. It was thought that perhaps the presence of the extended N-terminal of PtTrx*h*4 might induce a large conformational rearrangement which may consequently precluded structure solution by molecular replacement. Therefore, a SeMet-labeled protein was prepared and a SAD dataset was collected. The SAD data was input into SHELX-97 (Sheldrick & Schneider 1997) for evaluation and phasing in space group $P4_12_12$. Although three selenium atoms per protein were expected, only two were found ($SeMet_{79}$ and $SeMet_{134}$). Subsequent mass spectroscopy analysis allows us to identify the truncated protein (absence of the N-terminal part) which is presumed to be truncated during its production in *E. coli* system. The truncated SeMet-PtTrx$h4_{oxd}$ protein contains 113 instead of 139 amino acids (two residues at the C-terminal were absent). The solution with the highest correlation coefficient in heavy atom position determination was fed into SHARP (de La Fortelle & Bricogne 1997) for further refinement of SeMet-sulfur positions, where all parameters, including those modeling residual errors, were refined using a fully fledged maximum-likelihood approach (de La Fortelle & Bricogne 1997). After refinement, SHARP reported that occupancies of heavy atoms were 1.0 respectively, and calculated experimental phase had an overall figure of merit (FOM) of 0.35 (acentric reflections) and 0.06 (centric reflections) for 35.0 – 2.5 Å diffraction data. After the phase improvement was made using SOLOMON (Abrahams & Leslie 1996), the FOM gradually increased to 0.79 and the quality of the electron density map was improved substantially. This interpretable map was submitted to the automatic model-building program ARP/wARP, version 6.1 and 95

residues of the total 113 (truncated form SeMet-labeled protein) amino acids were built by the iterative free-atom density-modification and model-building procedure. The initial model was further improved by manual building with Turbo-Frodo interspersed with refinements using both CNS version 1.1 and REFMAC5.

Models of the wild type and the PtTrx*h*4C61S mutant were solved by molecular replacement (MOLREP, Vagin & Teplyakov 1997) using the SeMet-PtTrx*h*4$_{oxd}$ as initial search model. The WT dataset was integrated in a primitive hexagonal lattice. The highest symmetry space group in which the data could be scaled with better statistics is $P6_1$. Manual inspection of the molecular replacement solution in this space group suggested that this is the only system with plausible crystal packing. The final structure of the WT-PtTrx*h*4$_{oxd}$ was obtained after manual rebuilding of many parts of the structure (mainly near the active site and the α3 helix regions) and refinements using CNS and REFMAC5. As for the C61S-PtTrx*h*4$_{oxd}$ structure, the resulting solution from MOLREP was used as the input model for automatic model building using ARP/wARP and 99 % of the model was built automatically. Sequence coverage of resulted model is 93 % (the N-terminal region from residues 1 to 24 were not modeled). The model was further refined using CNS, REFMAC5 and iteratively combined with manual building by Turbo-Frodo and Coot. Structure refinement of the mutant structure involved substantial manual building intervention, not only due the presence of 23 additional residues (Met1 is not modeled), but also to significant structural differences (localize particularly at the interaction surface between the extension and the active site of the protein) and to a slightly lower quality of the electron density along the N-terminal extension. Throughout the model building and refinement processes, qualities of all models were assessed using the program PROCHECK.

3.3.3 Phasing method using molecular replacement: a case study for poplar glutaredoxin S12

The crystal structure of WT-PtGrxS12$_{oxd[GSH]}$ was determined by the molecular-replacement method with MOLREP (Collaborative Computational Project, Number 4, 1994; Vagin & Teplyakov, 1997) using the coordinates of the monomer without GSH of the human glutaredoxin 2 complexed with glutathione crystal structure (PDB code 2FLS; Johansson et al. 2007) as a model. Molecular replacement was performed using data to the maximum resolution of 3.0 Å and a distinguish rotation and translation function solution (rotation function/sigma value, 6.89; translation function/sigma value, 10.36; R factor, 0.522 and correlation value of 0.332) was found with one monomer in the asymmetric unit. The resultant solution coordinates were used for automatic model building using the program ARP/wARP (Lamzin & Perrakis 2004). After 75 cycles of autobuilding, 99 % of the model was built automatically. Hence, the initial model of WT-PtGrxS12$_{oxd[GSH]}$ comprise 100 residues with the connectivity index of 0.96 and R factor of 19.5 %. This structure was then refined using REFMAC version 5.4 interspersed with manual inspection using Coot. A GSH molecule was added almost towards the end of the refinement of the WT-PtGrxS12$_{oxd[GSH]}$ model.

The coordinates of WT-PtGrxS12$_{oxd[GSH]}$ were then used to solve the structure of the WT-PtGrxS12$_{oxd[GSH+HED]}$, by molecular replacement method. In these cases, the rotation and translation function solutions were easily identified as the template model used is the wild-type protein (assuming that all share similar 3-dimensional structure and no drastic conformational changes are present which might obstruct the structure solving using this method). Crystallographic refinements were performed with REFMAC5, alternating with manual rebuilding of the model and electron

density interpretation using Coot and/or Turbo-Frodo. Protein structures were initially refined without the presence of substrate(s) and water molecules. As soon as all the main-chain and side-chain of residues have been correctly positioned in the electron density maps and the overall structure is in accordance with the R factor (as this value is tied to the quality of the model and the data), ligands such as GSH and HED were added into the model. To create a test set that was used for R_{free} calculation, 5 % of the reflections were randomly selected. Positions of water molecules were identified with ARP/wARP and were also checked manually. Structure validation was performed with PROCHECK.

3.4 Structure refinement and validation

Refinement and model building of all structures were carried out using the REFMAC program version 5.4 of the CCP4 package, CNS suite version 1.1, Coot and Turbo-Frodo as visualization tools. For those liganded complexes or heavy atoms contained proteins, usually after first round of refinement, the $F_o - F_c$ difference electron density maps revealed clear density for bound ligands. These ligand molecules (such as cadmium, selenium, GSH and HED) were built in the model based on the difference densities. Solvent molecules were added subsequently using Coot and Turbo-Frodo. All crystallographic data and refinement statistics of the final models described here are presented in the result section.

3.4.1 Refinement using REFMAC5

REFMAC version 5.4 is a macromolecular refinement program which is comprised in the CCP4 package. This program can be run in two modes: the “review” mode which only checks and updates the input model to establish that the geometric restraints can be properly set up and the standard “refine” mode which carry out rigid body, translation-libration-screw (TLS), restrained or unrestrained refinements against X-ray data, or idealization of a macromolecular structure (CCP4: Collaborative Computational Project Number 4 1994). It minimizes the coordinate parameters to satisfy either a Maximum Likelihood or Least Squares residual. TLS parameters describe the possible mean square anisotropic displacements of rigid bodies and this refinement is often useful when there is a non-crystallographic symmetry (NCS) operator. It is often the case that different copies of a molecule in the asymmetric unit have different overall displacements. These can be accounted for by refining TLS parameters for such molecule. The residual atomic displacement parameters (B factors) should then be similar between molecules, and NCS restraints can be applied between them. For more details of TLS parameters see Martyn Winn’s page on TLS in REFMAC5 (http://www.ccp4.ac.uk/martyn/tls_research.html). REFMAC5 can also refine the isotropic displacement parameters of the models.

In terms of output files, the program usually generates necessary .log files. All statistics and lots of useful information about the behavior of the program are printed out to these standard output .log files. Besides a pdb file containing refined coordinates of the model, REFMAC5 also produces an MTZ output file (a flat-file representation of a particular data model or storage of reflection data) containing data

such as weighted coefficients for SigmaA weighted mF_o-DF_{calc} and 2mF_o-DF_{calc} maps, a crystallographic information file (CIF, pure ASCII text file) containing all relevant data for submitting structural data and an output TLS file (ASCII file) containing details of the TLS groups and the values of the corresponding T, L and S tensors. There is also the possibility to generate weighted difference maps files in either CCP4 or XtalView formats.

3.4.2 Refinement using CNS suite

Crystallography & NMR System or rather known as CNS, the result of an international collaborative effort among several research groups, is designed to provide a flexible multi-level hierarchical approach for the most commonly used algorithms in macromolecular structure determination. With the command line *cns_solve*, this program executes various jobs covering from heavy atom searching, experimental phasing (including MAD and MIR), density modification, crystallographic refinement with maximum likelihood targets, up to NMR structure calculation using NOEs, J-coupling, chemical shift, and dipolar coupling data. For some of the refinements done using CNS in this study, I've used the following tasks (listed in terms of input files): make_cv.inp (setup test array for cross-validation [free R] using a random selection of data), generate_easy.inp (generate coordinate and structure file for simple models), optimize_wa.inp (optimize Wa [X-ray/geometry weight]), rigid.inp (crystallographic rigid-body refinement), refine.inp (combined simulated annealing, energy minimization, B-factor refinement, and map calculation), anneal.inp (crystallographic simulated annealing refinement), minimize.inp (crystallographic conjugate gradient minimization refinement),

bgroup.inp (grouped, 2-groups per-residue, unrestrained B-factor refinement), bindividual.inp (restrained, individual B-factor refinement), water_pick.inp (pick water molecules in electron density map), sa_omit_map.inp (make annealed omit map with spherical omitted region), fo-fc_map.inp (make an electron density map using phase information from a model and 2 experimental amplitude data sets) and xtal_pdbsubmission.inp (produces partial information required for PDB submission). Details of the program and more executable jobs descriptions can be obtained from the website http://cns.csb.yale.edu/v1.1/.

3.4.3 Structure validation

Before deposition of final structures to the Protein Data Bank, the quality of each structure solved in this study were evaluated using the program PROCHECK which is one of the supported program in CCP4 suite. By executing this program, the stereochemical quality of the protein model is checked, producing a number of PostScript plots that analyzed its overall and residue-by-residue geometry.

3.5 PDB deposition

Only the coordinates and structure factors of all the native and mutant protein structures, liganded and non-liganded (reduced and oxidized forms), have been deposited in the RCSB Protein Data Bank (http://www.rcsb.org/pdb/home/home.do), with stated accession codes and references (Table 3.5). Both SeMet-PtGpx5$_{red}$ and SeMet-PtTrx*h*4$_{oxd}$ structures have not deposited into the PDB because these crystal structures were obtained merely for the purpose of solving the phase problem in

crystallography (see Introduction). At the time of this manuscript submission, PtGrxS12 structures have not been deposited into the PDB therefore no PDB code is available.

Table 3.5 Solved protein crystal structures in this study and their corresponding PDB codes and references.

Proteins (non-SeMet-type)	PDB code	Reference
SeMet-PtGpx5$_{red}$	[not deposited]	Koh et al. 2007
WT-PtGpx5$_{red}$	2P5Q	
WT-PtGpx5$_{oxd}$	2P5R	
SeMet-PtTrx*h*4$_{oxd}$	[not deposited]	Koh et al. 2008
WT-PtTrx*h*4$_{oxd}$	3D21	
C61S-PtTrx*h*4$_{oxd}$	3D22	
WT-PtGrxS12$_{oxd[GSH]}$	[not deposited]	Koh et al. *in preparation*
PtGrxS12$_{oxd[GSH+HED]}$	[not deposited]	

3.6 Structure superimpositions

Superimpositions of structures solved in this study with their corresponding homologous structures obtained from PDB were performed using LSQMAN program from the DEJAVU package (Kleywegt & Jones 1997) and the Lsqkab (superpose) program of the CCP4 suite with default parameters proposed by the authors.

3.7 Structure modeling and molecular dynamic simulations

There is no complex structure of peroxidase-thioredoxin available to date. Thus, the initial model for the oxidized PtGpx5-thioredoxin (oPtGpx5-1TI3) complex was constructed as follow. The initial model of the complex was generated using the

program Turbo-Frodo (http://www.afmb.univ-mrs.fr/-TURBO-) by simply superimposing the minimized average structure of 1TI3 (as the structure of TRX *h*1 was solved by a nuclear magnetic resonance approach and there is actually an ensemble of 20 structures) to the nearest symmetry mate of the oPtGpx5 molecule (monomer A), paying special attention to Trp124 of the symmetry mate (Trp124′). The complex was built to calculate the most energy-favourable complex in terms of its stability. All minimizations and MD simulations of the complex were performed using the AMBER8.0 simulation package (http://www.amber.scripps.edu) and the AMBER parm99 force field (Cornell et al. 1995) with a TIP3P water model (Jorgensen et al. 1983). The complex molecules were explicitly solvated in a truncated octahedral box and all counter-ions were excluded from all the calculations. The calculations were performed using two processors (Intel P4 Xeon) on the local in-house (LCM^3B, Nancy Université, France) cluster with 3 GHz CPUs. The non-bonded cutoff was set to 12.0 Å in all calculations unless mentioned otherwise. Before the MD simulations, all structures were energy-minimized using the SANDER module in AMBER 8.0, first by holding the complex fixed while minimizing the positions of water molecules using the conjugate gradient method for 1000 cycles and then minimizing the entire molecular system for 2500 cycles. MD simulations were carried out thereafter. The temperature of the system was raised gradually from 0 K to 300 K (with weak restraint on the complex) in 20 ps followed by 100 ps equilibration (without any restraint) at 300 K. All MD simulations were performed with a time-step of 2.0 fs. Various parameters (density, temperature, pressure, kinetic energy and potential energy) were monitored during the MD simulations, and generally were stable during the 100 ps equilibration.

Using the molecular mechanics Poisson–Boltzmann/ surface area (MM-PB/SA) method (Srinivasan et al. 1998), snapshots for energy analyses were obtained from the MD trajectories of both complexes. The water molecules were deleted for energy analyses. We have selected 100 snapshots from each equilibrated trajectory in 100 ps intervals. Snapshots from the trajectories were visually examined and illustrated with visual molecular dynamics (VMD) program (Humphrey et al. 1996).

Results and Discussion

CHAPTER 4

RESULTS AND DISCUSSION

Part I: Gpx

A. RESULTS

1. Specific objectives

I describe here the crystal structures of the Gpx5 of *P. trichocarpa* × *deltoides* (PtGpx5) in its oxidized (oPtGpx5) and reduced (rPtGpx5) states, revealing that PtGpx5 is a Gpx-type thioredoxin dependent peroxidase with the unique dimerization pattern mainly depending on hydrophobic contacts. The differences between the structures are discussed, with emphasis on the drastic transition between the oxidized and the reduced forms. Results obtained also clarified the doubts from biochemical experiments regarding the exact resolving cysteine of this enzyme and its functional quaternary state. In addition, a computer model of a complex of oPtGpx5 (PDB code 2P5R) and its *in vitro* recycling partner Trx *h*1 of poplar (PDB code 1TI3) is proposed on the basis of the crystal packing of the oxidized form enzyme. A possible role of PtGpx5 as a heavy-metal sink is also discussed.

2. Results obtained

2.1 Sequence alignment

In the search of protein sequence homologues using protein-protein BLAST (blastp) at the National Center for Biotechnology Information (NCBI) website against the nucleotide databank, PtGpx5 is found to share high identities with various plant Gpxs, ranging from 72 % to 91 %. However, none of these plant Gpxs had any structural reference. The BLAST search was then performed against structural databank (PDB) and mammal Gpxs (nine structures available to date) were highlighted as the closest hits of PtGpx5. Alignments of selected structures' sequences (based on the closest animal structure hits) and secondary structures are shown in Figure 4.1. From the sequence alignment, the absence of the region coded for the classical oligomerization loop clearly distinguished PtGpx5 and its plant homologues from the mammalian Gpxs. Another major difference that can be pointed out from the sequence alignment is the shortening of α2 helix by 5 residues in plant Gpxs as compared to the mammalian Gpxs.

During the time when I was first trying to solve the structure of PtGpx5, molecular replacements with available structures at that time apparently were unsuccessful due to the limit in sequence homologies, as expected from their sequence identities (34 % with 1GP1, 33 % with 2F8A, 29 % with 2HE3, and 26 % with human plasma GPX-3 and 32 % with 2I3Y, respectively). Thus, a SeMet-labelled protein was prepared in order to solve the structure of PtGpx5. However, the recently solved structure of human Gpx-4 (PDB code 2GS3) is structurally very similar to the rPtGpx5. Their sequences identities are 40 % and indeed, both

structures superimposed well (see results part I section 2.7). Therefore, molecular replacement could be possible to solve the reduced form structure in this study.

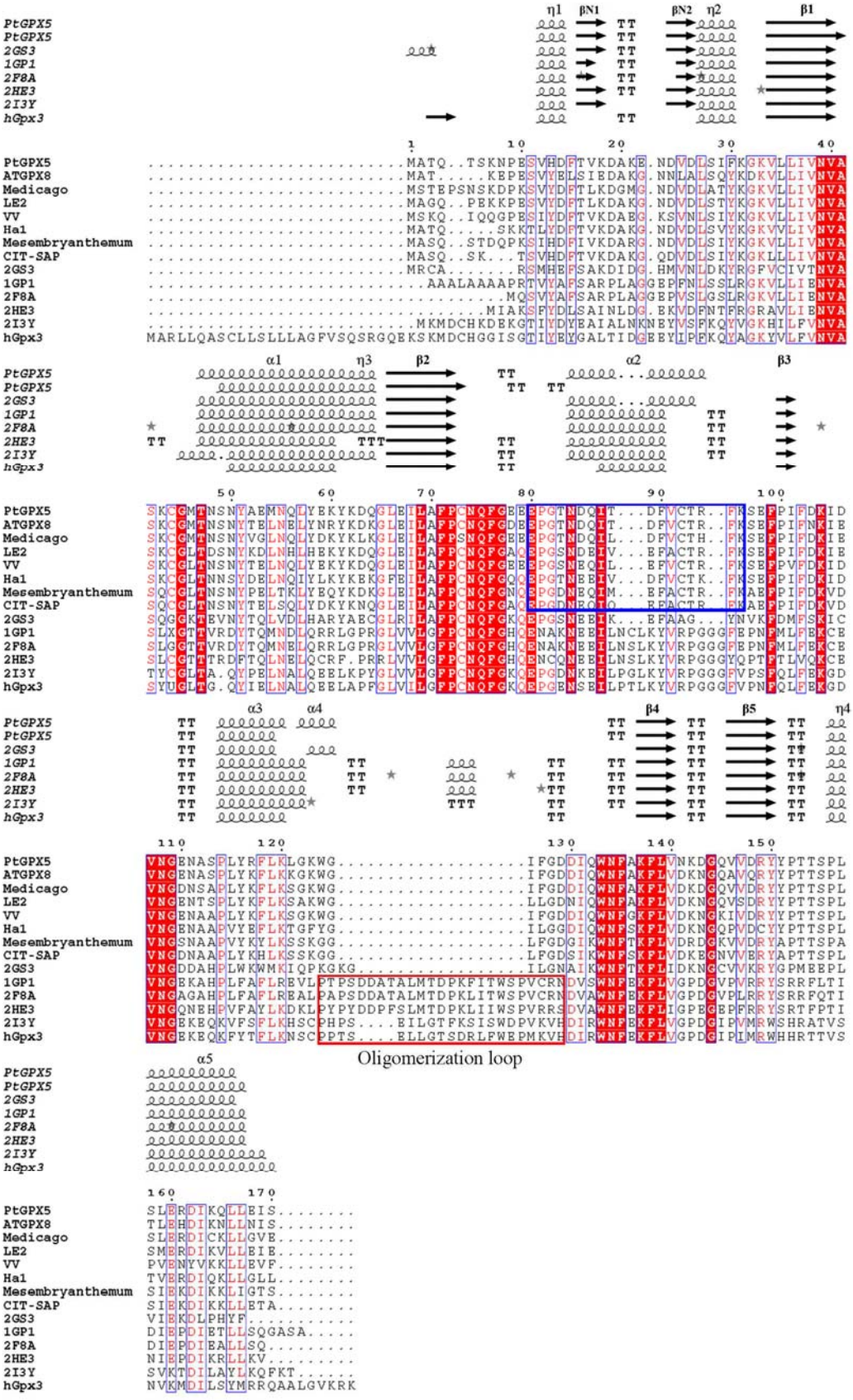


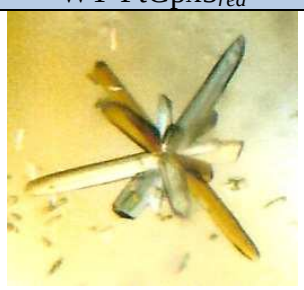
Figure 4.1 Multiple sequence alignment of 14 representative glutathione peroxidase proteins from mammals and plants (including the PtGpx5) as compared to the available secondary structure elements. Mammalian structures compared are bovine erythrocyte Gpx-1 (PDB code 1GP1), human Gpx-1 (PDB code 2F8A), human Gpx-2 (PDB code 2HE3), human plasma Gpx-3 (courtesy from Professor Rudolf Ladenstein), human Gpx-4 (PDB code 2GS3) and human Gpx-5 (PDB code 2IY3). Both reduced and oxidized states secondary structures of PtGpx5 are shown in comparison to the mammalian Gpx structures. Residues that are identical are boxed in red and displayed in white letters, while homolog residues are framed in blue and shown in red letters. For secondary structure representation, α and 3_{10} helices are displayed in coils with α and η labels, respectively, β strands are shown as arrows labeled β, and strict α and β turns are marked TT. Contacts between protein residues and supported hetero-compounds are shown as black stars. Sequence comparison between plant Gpxs and mammalian Gpxs (except mammal Gpx-4) show a large gap in plant Gpx sequences, while this region (framed in red) codes for an oligomerization loop in animal Gpxs. The accession numbers for other plant GPX sequences displayed in the final alignment are as follows: *Arabidopsis thaliana*, ATGPX8: NP_564813; *Medicago truncatula*, Medicago: ABE92132; *Solanum lycopersicum*, LE2: AI898013; *Vitis vinifera*, VV: CB978870; *Helianthus annuus*, Ha1: CAA74775; *Mesembryanthemum crystallinum*, Mesembryanthemum: CAC83045; *Citrus sinensis*, CIT-SAP: CAA47018. A motif from residue Glu80 to Lys96 (numbering in PtGpx5) in plant Gpxs is proposed (framed in dark blue) to be the recognition site for Trx (see results part I section 2.6). The Figure was prepared using ESPript (Gouet et al. 1999).

2.2 Crystal description and structure refinement statistics

Both SeMet-PtGpx5$_{red}$ and WT-PtGpx5$_{red}$ crystals are rhombohedral (designated form I) while the WT-PtGpx5$_{oxd}$ crystal is trigonal (designated form II). Form I crystals belong to the space group R_3 with cell dimensions of a = b = 222.7 Å and c = 48.1 Å. Assuming four subunits per asymmetric unit, the V_M value of the crystal is 2.93 Å^3/Da. Form II exhibits in the space group $P3_121$ with cell dimensions of a = b = 71.6 Å and c = 117.8 Å. The presence of two subunits per asymmetric unit in form

II crystal gives the V_M value of 2.24Å^3/Da. Details of data refinement statistics are summarized in Table 4.1.

Table 4.1 Data refinement statistics for WT-PtGpx5$_{red}$ and WT-PtGpx5$_{oxd}$ crystals.

Data Set	WT-PtGpx5$_{red}$	WT-PtGpx5$_{oxd}$
		
Resolution range (Å)	30.0 - 2.00	50.0 – 2.45
Reflections used	59452	12907
R_{cryst}[a] (R_{free})[b]	17.85 (20.04)	21.08 (28.03)
Number of atoms (Protein / waters / Cd^{2+} / Ca^{2+} / acetate ions)	5187 / 668 / 32 / 0 / 8	2650 / 224 / 0 / 5 / 0
Average *B*-value for individual chain (Å^2)	16.32 / 25.41 / 21.08 / 29.68 (subunit A / B / C / D)	39.15 / 53.58 (subunit A / B)
Mean *B*-factor		
Main-chain (Å^2)	22.10	46.24
Side-chain (Å^2)	24.12	46.49
Water (Å^2)	33.19	42.08
All (Å^2)	24.40	46.07
Rms deviation from ideal geometry		
Bond lengths (Å)	0.005	0.006
Bond angles (°)	1.3	1.3
Dihedral angles (°)	22.9	23.9
Improper angles (°)	0.67	0.72
Ramachandran plot		
Residues in most favored regions (%)	90.5	81.4
Residues in additionally allowed regions (%)	9.3	16.2
Residues in generously allowed regions (%)	0.2	2.4

[a] $R_{cryst}=\Sigma|F_o-F_c|/\Sigma F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.
[b] The R_{free} value was calculated from 10 % of all data that were not used in the refinement.

2.3 Quality of the model

PtGpx5 is a 170 residues protein with a monomeric molecular mass of 19.36 kDa and exists as a dimer in solution (Navrot et al. 2006). The reduced form of the enzyme (rPtGpx5) was crystallized as two dimers (dimer 1, subunits A and B; dimer 2, subunits C and D) in the asymmetric unit and this structure was solved from rhombohedric selenium-substituted protein crystal dataset (2.7 Å) using the single wavelength anomalous diffraction method. The structure was completed using the high-resolution data (1.95 Å) collected from the native rPtGpx5. The final model of rPtGpx5 structure, with R and R_{free} factors of 17.85 % and 20.04 %, respectively, contains 160 or 161 amino acid residues corresponding to residues from Glu10 to Ile169 in monomers B, C, and D and from Glu10 to Ser170 in monomer A. The overall structure is well defined in the electron density map except for the missing portions mentioned above. Very few residues have poorly defined side-chains and they are usually located in flexible loops on the protein surface. In addition, there are 32 cadmium atoms, eight acetate anions and 668 water molecules found in the asymmetric unit of the reduced form crystal. The r.m.s.d. overlaps of all the subunits fall in the range from 0.15 Å – 0.33 Å based on alignments of Cα positions, thus suggesting that all subunits are equivalent. The most ordered with well-defined electron density subunit, as indicated by the average B-value (Table 4.1), is subunit A that participates most highly in the subunit–subunit interactions. Hence, unless indicated otherwise, our discussion will focus on this subunit.

The oxidized form of the enzyme (oPtGpx5) was crystallized as a dimer in the asymmetric unit. The final model of oPtGpx5 contains residues from Ser6 to Ile169 in each subunit with an R-value of 21.08 % (R_{free} = 28.03 %) at 2.45 Å

resolution. Compared to the structure of the reduced form, oPtGpx5 structure has more poorly defined regions. Most of the residues located in a long flexible loop (residues 73– 100) are found in the generously allowed regions of the Ramachandran plot (Table 4.1). The electron density of the region between residues 77 to 84 can hardly be seen. Among the subunits in the asymmetric unit, subunit A shows the clearest densities for the main-chain atoms. Thus, we have adopted the conformation of the flexible loop of subunit A for the loop positioning of the subunit B as this monomer has more poorly defined regions. In the model of the oxidized form, we have assigned five calcium atoms per asymmetric unit as calcium chloride was present in the crystallization condition.

2.4 Overall fold and secondary structure of PtGpx5 subunit

The subunit structure shows a thioredoxin fold (Martin 1995), which consists mainly of a central twisted β-sheet and several flanking α-helices, with few additional secondary structure elements, similar to that described for the Gpx (Epp et al. 1983, Ren et al. 1997) and Prx (Choi et al. 2003, Hirotsu et al. 1999, Schröder et al. 2000, Declercq et al. 2001) families. The reduced form (rPtGpx5) subunit structure consists of four α helices and seven β strands (Figure 4.2). Helices α1, α2 and α4 are located on one side of the 45° twisted central β sheet (β3, β2, β1, β4 and β5) and α3 is on the other side. Helices α1 and α4 are nearly parallel and their axes are roughly parallel with the central β strands. The two other α helices, α2 and α3, are oriented perpendicular to each other and to the helices α1 and α4 (Figure 4.2(c)). The N-terminal end exhibits an extra chain folding into a β-hairpin designated as βN1 and βN2, respectively.

Superimposition of both reduced and oxidized form of PtGpx5 subunit structures shows a r.m.s.d. of 4.02 Å (based on alignments of 160 Cα positions) (Figure 4.3). This high r.m.s.d. value is due to the large local differences observed in α1 and α2 helices between both structures. The beginning of α1 helix in the oPtGpx5, where the peroxidatic Cys44 is located, undergoes a local unwinding and rearrangement. The local flipping of this small loop towards the direction of α2 helix, however, is rather small as compared to the complete unwinding of the α2 helix, where the resolving Cys92 is situated. The latter motion results in a flexible extra long loop (residues 73–100), connecting β2 and β3 strands. These conformational changes lead to the loss of 58 interactions, which were involved in the stabilization of the reduced form enzyme, whereas 28 interactions are newly created to stabilize the oxidized form. If the two structures are superimposed, excluding the conformational changed regions, the r.m.s.d. is only 0.525 Å (based on alignments of 135 Cα positions) (Figure 4.3). Indeed, the rest of both rPtGpx5 and oPtGpx5 structures superimpose well.

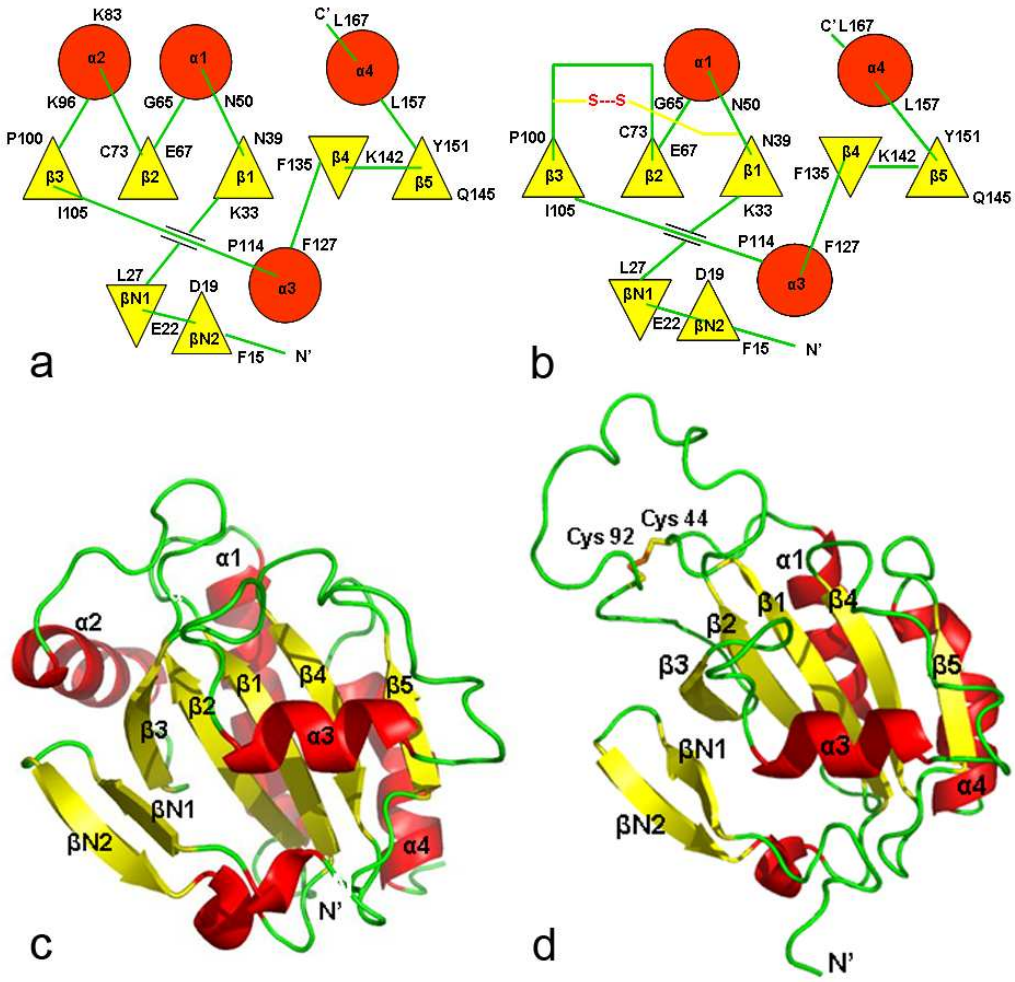


Figure 4.2 (a) and (b) Topological diagram and (c) and (d) cartoon representations showing the overall structures of the rPtGpx5 (left two panels) and the oPtGpx5 (right two panels). All α-helices are shown in red while β-strands are in yellow. In the topological diagrams, the helices are represented as circles and the β-strands as triangles. The beginning and the ending residues with numbering of each secondary structural element are labeled. The intramolecular disulfide bond between Cys44 and Cys92 is highlighted.

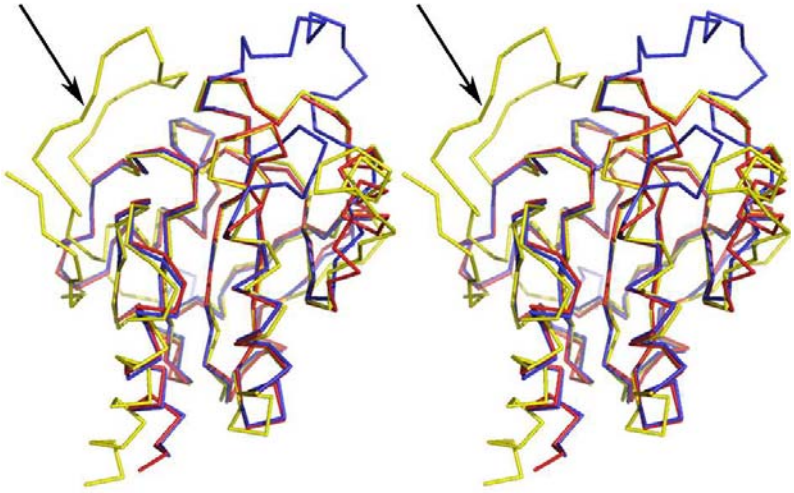


Figure 4.3 Stereoscopic views of superimpositions between the Cα traces of the rPtGpx5 (red), the oPtGpx5 (blue) and a typical mammalian Gpx, the human plasma Gpx-3 (yellow) (Ren et al. 1997). The major difference between PtGpx5 and the classical mammalian Gpx is the absence of the oligomerization loop (indicated by the arrow) in the PtGpx5.

2.5 Oligomerization state of PtGpx5

One of the inter-subunit contact types observed within the crystal lattice of PtGpx5 likely represents the subunit interaction that stabilizes the dimeric enzyme in solution. Indeed, from two different experiments, using different crystallization conditions, we observed the same dimerization pattern in the crystal structures regardless of their redox states. Dimer interface areas in both reduced and oxidized structures range from 822 $Å^2$ to 847 $Å^2$, involving 60 % of non-polar atoms and 40 % of polar atoms. In this study, we highlight a new mode of dimerization pattern that differs from previously reported Gpxs and Prxs. The dimerization interface localizes at the C-terminal region of PtGpx5. The main dimer building block consists of two anti-parallel β-sheet ($\beta5_A$-$\beta5_B$) positioned side-by side in a head-to-tail manner and are related by local 2-fold symmetry axis. Formation of the PtGpx5 dimer results in a

central pseudo-10-pleated β-sheet core through the dimer (Figure 4.4). However, although the two β5 strands are adjacent to each other, the –CO groups of one $\beta5_A$ strand and the -NH groups on the adjacent $\beta5_B$ strand are 6 Å apart. In fact, the dimer interface is stabilized by van der Waals interactions which involve the participation of a hydrophobic and aromatic cluster (Trp124, Ile126, Phe127, Trp133, Tyr150, Tyr151, Pro152, Cγ of Thr153 and Cγ of Thr154) and several hydrogen bonds involving polar side-chains of residues located in the loop that connects α3 to β4, β5 and α4 (summarized in Table 4.2).

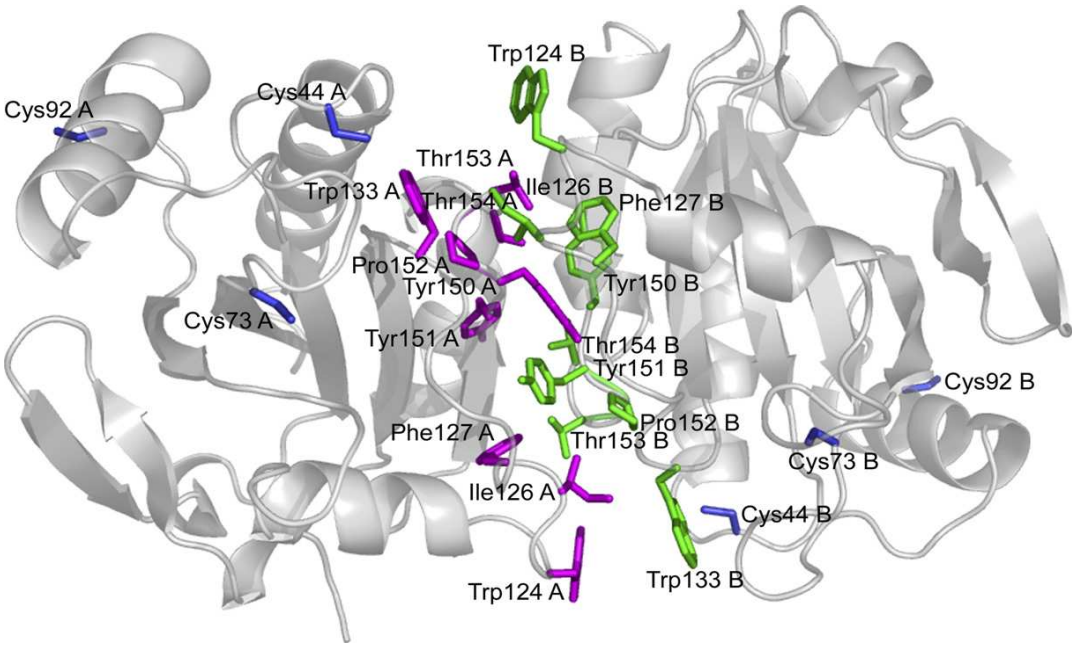


Figure 4.4 Cartoon representation of dimerization interfaces of the rPtGpx5 (monomers A and B). This figure highlights the dimerization interface involving the hydrophobic and aromatic residues (shown as sticks). The side-chains of monomer A and monomer B, which are involved in the dimerization, are coloured magenta and green, respectively. The locations of the peroxidatic Cys44, the highly conserved Cys73 and the resolving Cys92 are coloured blue.

Table 4.2 Hydrogen bonding interactions at the dimerization interface.

Monomer A	Monomer B	Distance (Å)	
		rPtGpx5	oPtGpx5
Loop between α3 and β4 (124-133)			
Asp 129 $O_{\delta1}$	Gln 132 $N_{\varepsilon2}$	2.67	2.44
Asp 129 $O_{\delta1}$	Tyr 151 OH	2.82	3.00
Gln 132 $N_{\varepsilon2}$	Asp 129 $O_{\delta2}$	2.71	2.82
Gln 132 $N_{\varepsilon2}$	Tyr 151 OH	2.84	2.82
β5 (147-155)			
Arg 149 NH_1	Tyr 151 OH	2.94	(3.73)
Arg 149 NH_2	Tyr 151 OH	3.17	(3.55)
Tyr 151 OH	Asp 129 $O_{\delta2}$	2.65	3.05
Tyr 151 OH	Arg 149 NH_1	3.14	2.98
Tyr 151 OH	Gln 132 $N_{\varepsilon2}$	2.89	2.85
Ser 155 Oγ	Arg 161 NH_2	3.28	3.37
α4 (156-166)			
Ser 158 Oγ	Asp 162 $O_{\delta2}$	2.81	3.05
Ser 158 Oγ	Asp 162 $O_{\delta1}$	2.93	2.41
Ser 158 Oγ	Arg 161 NH_1	2.73	3.03
Arg 161 NH_1	Ser 155 Oγ	2.66	3.21
Arg 161 NH_1	Ser 158 Oγ	2.92	2.72
Asp 162 $O_{\delta1}$	Ser 158 Oγ	2.83	2.52
Asp 162 $O_{\delta2}$	Ser 158 Oγ	2.87	3.19
Distance >3.5 Å is not considered as a hydrogen bonding interaction and is indicated for comparison purposes (marked in parentheses).			

2.6 Active-site of PtGpx5

2.6.1 Environment of the peroxidatic Cys in the reduced form of the enzyme (rPtGpx5)

The active site residues in the neighbourhood of the peroxidatic Cys44 are well-defined in ($3F_o$–$2F_c$) electron density map. The active site pocket is formed by residues in the long loop between α3 and β4 (Phe127–Phe135), the loops between β2 and α2 (Cys73–Thr83), β5 and α4 (Tyr151–Leu157) and β1 and α1 (Val40–Met46), while the assumed peroxidatic Cys44 of PtGpx5 (Navrot et al. 2006, Rouhier & Jacquot 2005) is positioned on the seven residue-long loop just before the α1 helix

(near the interface of a dimer of the asymmetric unit). The side-chain of Cys44 is pointing towards the interior of the protein, surrounded mainly by non-polar groups (Met46, Phe76, Phe135, Trp133, Pro152 and the aliphatic chain of Lys43), while the remaining residues are hydrophilic (Glu79, Asn134) (Figure 4.5(a)). The peroxidatic Cys44 remains accessible to solvent because of its strategic location near the surface of the subunit.

The active-site cleft appears to be occupied by a cadmium ion (Cd^{2+}) that originated from the crystallization solution, ligated by three residues and well coordinated with three water molecules. The coordination of this cadmium ion displays a distorted octahedron, ligated with Cys44-S_{γ}, Glu79-$O_{\varepsilon 1}$, Trp133-$N_{\varepsilon 1}$ and three water molecules (see Table 4.3 and Figure 4.5(a)). The positions of Cd^{2+}-surrounding ligands show no significant geometry changes compared with the active sites described for mammalian Gpxs (which do not possess any cadmium ion) (Epp et al. 1983, Ren et al. 1997).

Further, the active site of rPtGpx5 is shielded by the adjacent subunit of the dimer (residues Gly122– Ile126 at the C terminus of the α3), which suggests that the dimer in PtGpx5 and probably in all plant Gpxs is of catalytic significance. In the structure of the reduced form, peroxidatic cysteines (Cys44) of a dimer are ~38 Å apart and occupy opposite sites (Figure 4.4). In addition to the catalytic site Cys44, there are two other cysteine residues present in each PtGpx5 subunit. Cys73 is located on the loop that connects β2 and α2, about 11 Å apart from Cys44, while Cys92 is located on α2 and is 21 Å away from the peroxidatic residue.

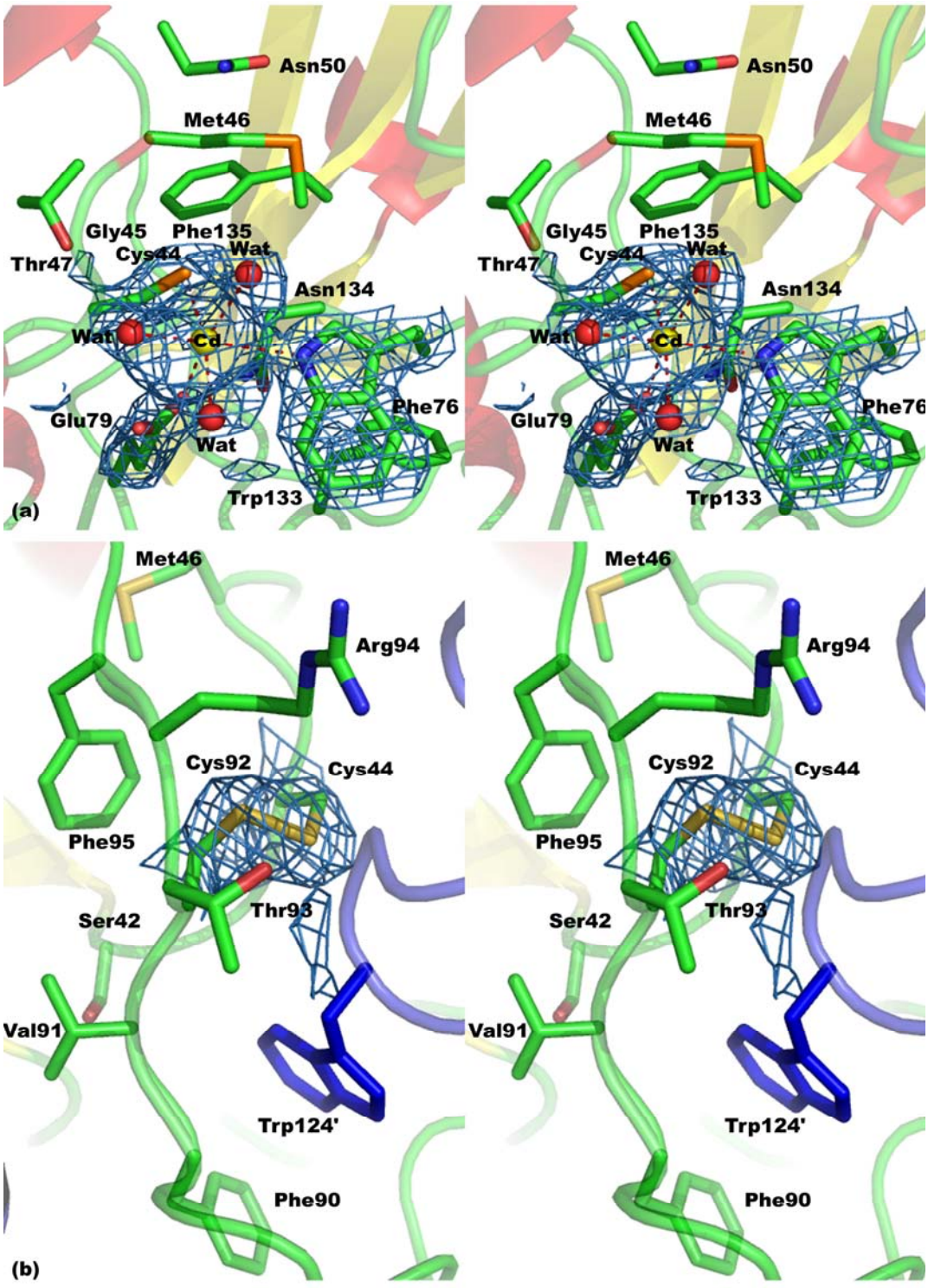


Figure 4.5 Stereoview details of the active site cleft of both (a) rPtGpx5 and (b) oPtGpx5 with final $3F_o–2F_c$ electron densities (1.2 σ level) covering chosen residues for clarity. Relevant residues are labeled. In the rPtGpx5 active site, a cadmium ion is coordinated to the peroxidatic Cys44, Trp133, Glu79 and three water molecules, forming an octahedral geometry. In the oPtGpx5 active site, an intramolecular disulfide bond is formed between Cys44 and Cys92. All the residues shown here (in sticks, colored according to atom type) are from monomer A. (b) The Trp124 residue, colored blue, stacks from a symmetry-related subunit.

Table 4.3 Coordination geometry of the active-site cadmium ion.

Bond lengths (Å)		Bond angles (°)	
Cys44 S^{γ} - Cd	2.59	Cys44 S^{γ} – Cd – Glu79 $O^{\varepsilon 1}$	90
Glu79 $O^{\varepsilon 1}$ - Cd	2.44	Cys44 S^{γ} – Cd – Trp133 $N^{\varepsilon 1}$	105
Trp133 $N^{\varepsilon 1}$ - Cd	2.75	Cys44 S^{γ} – Cd – Wat318	167
Wat318 - Cd	2.65	Cys44 S^{γ} – Cd – Wat482	96
Wat482 - Cd	2.71	Cys44 S^{γ} – Cd – Wat626	103
Wat626 - Cd	2.81	Glu79 $O^{\varepsilon 1}$ – Cd – Trp133 $N^{\varepsilon 1}$	92
		Glu79 $O^{\varepsilon 1}$ – Cd – Wat318	93
		Glu79 $O^{\varepsilon 1}$ – Cd – Wat482	171
		Glu79 $O^{\varepsilon 1}$ – Cd – Wat626	111
		Trp133 $N^{\varepsilon 1}$ – Cd – Wat318	88
		Trp133 $N^{\varepsilon 1}$ – Cd – Wat482	79
		Trp133 $N^{\varepsilon 1}$ – Cd – Wat626	143
		Wat318 – Cd – Wat482	84
		Wat318 – Cd – Wat626	64
		Wat482 – Cd – Wat626	75

2.6.2 Environment of the peroxidatic Cys in the oxidized form of the enzyme (oPtGpx5)

In the structure of the oxidized form (oPtGpx5), an intramolecular disulfide bond between the peroxidatic Cys44 and the resolving Cys92 is clearly visible. The formation of the disulfide bond causes two regions in oPtGpx5 to rotate toward each other, decreasing the Cα–Cα distance between residues 44 and 92 by 12.1 Å relative to the reduced molecule. The torsion angles for the bonds constituting the disulfide bond between Cys44 and Cys92 are as follows: Cys44, x1=−58°, x2=−60°, x3=+96°, x2′=+84°, x1′=−165°, Cys92.

Examining the surroundings of the disulfide bond in the oPtGpx5 reveals that one side is freely exposed to the solvent while the other side faces an aromatic residue (Phe95). Close to the disulfide bond, there are several other residues, including Met46, Thr93 and Arg94 (Figure 4.5(b)). In addition, a spoon-shaped, surface-exposed loop (Figure 4.6) connecting β2 and β3 (residues 77–100) is

observed facing the oxidized active site. This long loop results from the unwinding of α2 helix mentioned earlier. The conformation of this loop is an obtuse angle between protein core and the surface loop (~120°).

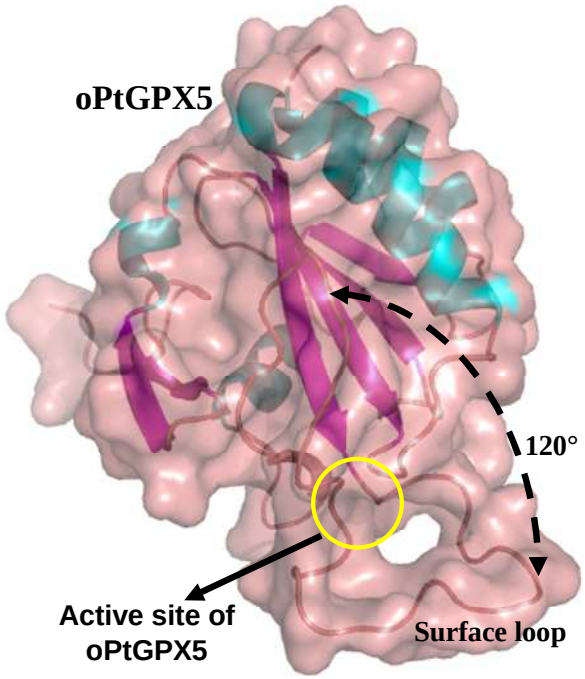


Figure 4.6 Cartoon representation of the oPtGpx5 structure with the molecular surface. The active site region is circled in yellow for clarity. All α-helices are shown in magenta while β-strands are in magenta and connecting loops are in pink. The surface loop (see text) makes a 120° angle with respect to the core of the enzyme in its oxidized state.

2.7 Proposed model of the oPtGpx5-TRX *h*1 complex

As a result of crystal packing, a Trp residue (Trp124′) from a crystallographically symmetric neighboring molecule is observed in the vicinity of the intramolecular disulfide bond (Figure 4.5(b)). As mentioned before, oPtGpx5 is regenerated by Trx (Navrot et al. 2006), and the Trp residue at the catalytic site of Trx plays an important role in Trx-substrate recognition (see Discussion). It is tempting to

speculate that a similar situation at the active site may occur when the oPtGpx5 binds the reducing redox partner, Trx. Therefore, we were prompted to model the complex that enables us to visualize the possible interaction between the ROS scavenger (oPtGpx5) and its *in vitro* recycling partner (Trx *h*1, Coudevylle et al. 2005, PDB code 1TI3). Several molecular dynamics (MD) processes were performed by placing the Trx molecule next to oPtGpx5, in such a position that Trp124′ of the symmetry-related oPtGpx5 is replaced by Trp37 of Trx *h*1 (see Materials and Methods). Two complex models can be built with a transient intermolecular disulfide bond formed between the catalytic Cys38 of poplar Trx *h*1 and either the peroxidatic Cys44 (complex 1) or the resolving Cys92 (complex 2) of oPtGpx5. The complex 1 exhibits a plus left-handed hook disulfide bond, while the complex 2 possesses a minus right handed hook disulfide bond. In terms of calculated free energy, however, complex 2 reveals a more stable complex as compared to complex 1.

In both complex models, several edge-to-face aromatic interactions were essential at the protein– protein interfaces involving especially the Trp37 of Trx *h*1 (Figure 4.7). The $N_{\varepsilon 1}$ atom of this residue is hydrogen bonded with Gln86-O in complex 1 or with Asn84-O in complex 2. Otherwise, this residue is always located between a Phe residue and a Pro residue. In both cases, the Phe ring interacts perpendicularly, with its edge pointing towards the Trp face, while Pro tends to make CH–π interaction on the Trp face through major contributions from its C_{γ} and C_{δ}. In complex 1, both of these Phe and Pro residues are from oPtGpx5 (Phe90 and Pro81), whereas in complex 2, Phe90 of oPtGpx5 and Pro39 of Trx *h*1 are involved. Nevertheless, all these residues are conserved in plant Gpx homologues and Trx *h*1 subgroup. Details of the hydrogen bonding interactions involved in both complexes are summarized in Table 4.4.

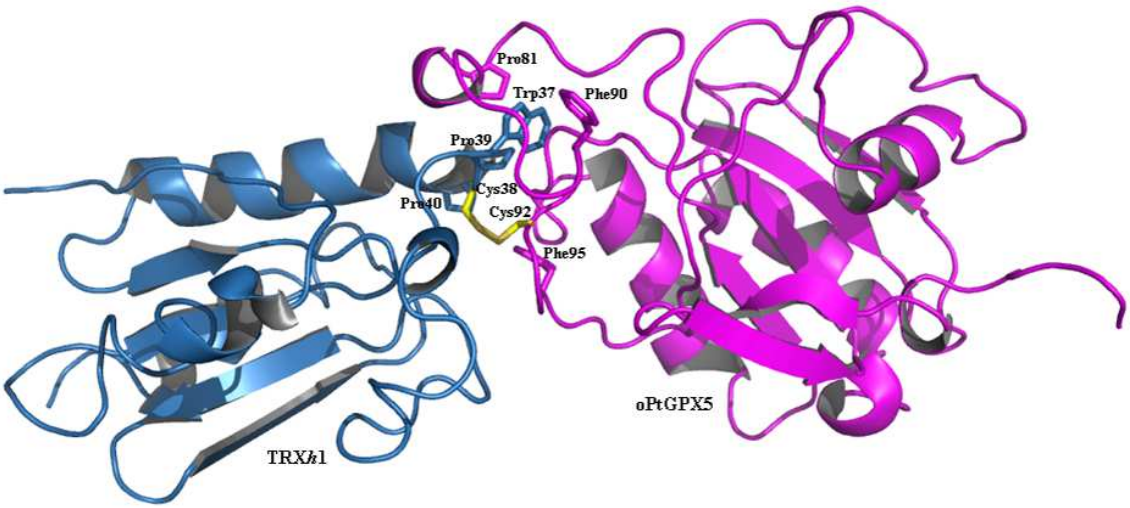


Figure 4.7 Molecular model of the oPtGpx5-Trx *h*1 complex. The modeled complex 2 (see Results) structure shown in cartoon representation displays an intermolecular disulfide bond (yellow) between the catalytic Cys38 of Trx *h*1 and the resolving Cys92 of oPtGpx5. Protein–protein interactions between the oPtGpx5 (magenta) and the Trx *h*1 (blue) involving several aromatic residues at the interface are shown as sticks and are coloured according to respective molecules.

Table 4.4 Hydrogen bonding interactions between oPtGpx5 and Trx *h*1 in both complex 1 and complex 2 models.

Molecule	oPtGpx5	TRX *h*1	*d* (Å)
A. Complex 1	Cys44 N	Trp37 O	3.24
	Gln86 $N^{\varepsilon 2}$	Asp64 $O^{\delta 2}$	3.25
	Gln86 $N^{\varepsilon 2}$	Asp66 $O^{\delta 2}$	2.88
	Gln86 O	Trp37 $N^{\varepsilon 1}$	3.22
	Thr88 $O^{\gamma 1}$	Ser36 O^{γ}	3.29
	Thr88 O	Lys67 N^{ζ}	2.81
B. Complex 2	Glu80 O	Lys42 N^{ζ}	2.89
	Thr83 $O^{\gamma 1}$	Ser36 O	3.00
	Asn84 O	Trp37 $N^{\varepsilon 1}$	3.03
	Gln86 $N^{\varepsilon 2}$	Ser36 O^{γ}	2.76
	Arg94 NH_1	Val77 O	2.71

Complex 1; disulfide bond between Cys44 of oPtGpx5 and Cys38 of Trx *h*1.
Complex 2; disulfide bond between Cys92 of oPtGpx5 and Cys38 of Trx *h*1.

2.8 Comparison to homologous structures

PtGpx5 shares a high level of identity with various plant Gpxs, ranging from 72 % to 91 %. However, none of these plant Gpxs has any structural reference. To date, there are nine Gpx crystal structures available, all of them from mammals. Among them are the bovine Gpx-1 (PDB code 1GP1) structure (Epp et al. 1983) and five human Gpx structures (Gpx-1, PDB code 2F8A; Gpx-2, PDB code 2HE3; Gpx-3, courtesy from Professor Rudolf Ladenstein (Ren et al. 1997); Gpx-3 selenocysteine to glycine mutant, PDB code 2R37; Gpx-4, PDB code 2GS3; Gpx-4 selenocysteine to cysteine mutant, PDB code 2OBI (Scheerer et al. 2007); Gpx-5, PDB code 2IY3 and Gpx-7, PDB code 2P31). All the classical mammalian Gpx structures known so far are in the tetrameric form, except for human Gpx-4, which is a monomeric enzyme (PDB code 2GS3). Gpx-5 from poplar presents as a dimer with interactions that differ from those existing in other known Gpx or 2-Cys Prx structures (see Discussion).

The superimposition of rPtGpx5 subunit with the other mammal Gpxs leads to r.m.s.d. varying from 0.72 Å with human Gpx-4 (based on alignments of 152 Cα positions) to 1.08 Å with human plasma Gpx-3 (based on alignments of 144 Cα positions). Structure-based alignments demonstrated that the main differences between PtGpx5 and other mammalian Gpxs are the absence of an extended N terminus, the deletion of the region coding for the classical oligomerization loop, and the shortening of α2 helix (by five residues) (Figure 4.1). In terms of the N terminus, PtGpx5 does not possess the extra N-terminal residues (the region before the N-terminal β-hairpin) observed in the human Gpx-3 enzyme structure. Likewise, the other mammalian Gpx-1, Gpx-2, Gpx-4 and Gpx-5 also lack this extra feature. In addition, a large part of the classical mammalian sequence (positions 125–142 in bovine Gpx-1 structure numbering) is deleted in PtGpx5 and in human phospholipid hydroperoxide Gpx-4 (PDB code 2GS3). This missing part corresponds precisely to the subunit interaction sites of the tetrameric mammalian Gpxs.

Another important difference is that PtGpx5 displays an overall negatively charged protein surface, while the rest of the mammalian Gpxs show more evenly distributed charged protein surfaces. Sequence analysis also demonstrated that PtGpx5 contains 15 % of negatively charged residues (Asp and Glu), while the mammalian Gpxs contain 11 %. Remarkably, the α-helix 2 in PtGpx5, where the resolving Cys92 is situated, exhibits a highly negatively charged region (Figure 4.8(d)) as compared to the corresponding regions in all mammalian Gpxs. Further, active site residues of rPtGpx5 and their geometries are similar to those described for the most classical Gpx enzymes. However, a peroxidatic Cys (Cys44) and a Glu (Glu79) are found in the active site pocket of the present structure instead of SeCys and Gln (Gln83; numbering in human plasma Gpx-3 structure) residues in classical mammalian Gpx structures. These differences may suggest enzyme adaptation to substrate binding and specificity or to the adapted redox partner (see Discussion).

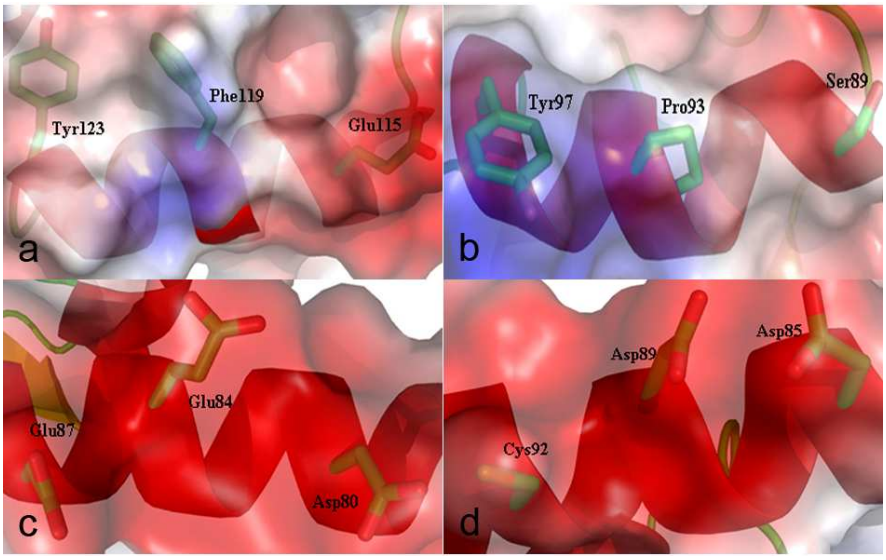


Figure 4.8 Local surface charges of α-helix 2 of (d) rPtGpx5 and its corresponding helices of other mammalian Gpx structures ((a) the monomeric human Gpx-4, PDB code 2GS3; (b) the classical tetrameric human Gpx-3, human plasma Gpx-3). The corresponding α2 helix in calmodulin (PDB code 1CLL) that exhibits a helix–coil transition is shown in (c). Negatively and positively charged regions are shown in red and blue, respectively, while the rest are shown in grey. Three residues, situated

at each helix turn and facing the same side of the helix, are displayed as sticks.

2.9 Cadmium-binding sites

There are 32 cadmium ions present in the asymmetric unit of PtGpx5 in its reduced form (Figure 4.9(a)). We have distinguished four potential cadmium-binding sites per monomer (16 sites per asymmetric unit) on the basis of the following three prerequisites: (i) the cadmium-binding site is present in the four independent subunits; (ii) at least two non-water ligands are involved in ion-binding; and (iii) the ion-binding site is independent of the crystal packing. The first cadmium-binding site is situated at the active site cavity, displaying the coordination geometry of an octahedron as described before (see the previous section). The second binding site located at the dimer interface also shows the typical octahedron geometry by binding to six ligands, including Asp148-$O^{\delta 1}$ of one subunit, Thr153-O of the adjacent subunit and four water molecules (Figure 4.9(b)). Anchored to the subunit interface, the third cadmium-binding site (Figure 4.9(c)) is hepta-coordinated by two bidentate residues (Asp85 and Asp89) and water molecules. Similar coordination geometry is found for the fourth cadmium-binding site (Figure 4.9(d)) at the β-hairpin loop of each subunit, except that the two bidentate residues involved are Glu22 and Asp103.

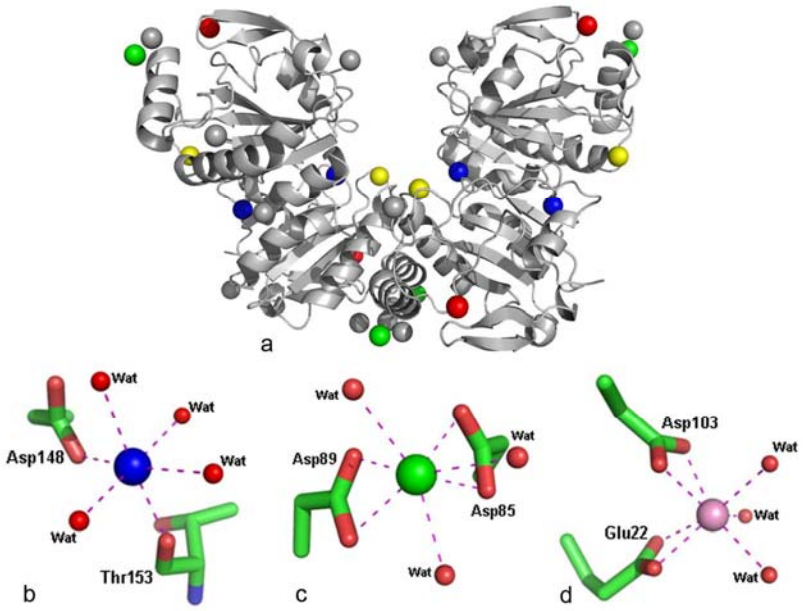


Figure 4.9 The 32 independent cadmium ions in the rPtGpx5 crystal structure are displayed. (a) The two dimers of rPtGpx5 are shown in cartoon representation and Cd^{2+} involved in potential cadmium-binding sites are highlighted (yellow, blue, green and pink for sites 1, 2, 3 and 4, respectively). The sites 1 (Figures 4.5(a)), 2(b), 3(c) and 4(d) are shown in detail with Cd^{2+} in the same colour code as in (a).

B. DISCUSSION

1.1 Re-evaluation of PtGpx5 denomination

Sequence comparison revealed that PtGpx5 and other plant homologues share relatively high levels of sequence similarities with mammal Gpxs. Contrary to the majority of mammalian Gpxs, which involve only a SeCys residue in the catalytic mechanism, the higher plant Gpx homologues identified so far possess two cysteine residues (peroxidatic Cys, C_P and resolving Cys, C_R) that take part in the reduction of peroxide. These mammalian Gpxs use GSH for their regeneration of the SeCys catalytic residue. However, it has been suggested that Gpx-type enzymes are not

completely specific for GSH but may react also with CxxC motifs present in Trx or tryparedoxins (Flohé et al. 2003). Indeed, most of the plant Gpxs probably use reduced Trx for their regeneration (Navrot et al. 2006, Jung et al. 2002, Kang et al. 2004, Herbette et al. 2002, Tanaka et al. 2005), and do not react with GSH or Grx, as demonstrated recently for PtGpx5 (Navrot et al. 2006). The reaction mechanism of PtGpx5 was shown to be homologous to that of atypical 2-Cys Prxs, in which (i) the reduction of the peroxide is accompanied by the formation of a sulfenic acid on the catalytic cysteine, (ii) a second Cys residue forms an intramolecular disulfide bridge with the C_P and acts as a C_R in being indispensable for the reduction by Trx (Navrot et al. 2006). Thus, in a classification based on existing biochemical evidence rather than on phylogenetic linkage, the plant Gpxs constitute a fifth group of plant Prxs (Rouhier & Jacquot 2005). On the other hand, a reclassification of the group of enzymes in which SeCys is replaced by Cys in Gpx homologs is being suggested on the basis of biochemical properties, as Gpx-type Trx peroxidases (Maiorino et al. 2007). We describe here for the first time in plants, two structures of PtGpx5 in two different redox states that give insights for the denomination of this enzyme from a structural point of view.

Gpx and Prx present the same overall fold, namely the Trx fold, which consists of a central β-sheet and flanking α helices. However, with respect to the nature of their redox centers, the SeCys residue in mammal Gpxs is at hydrogen bonding distance from a Trp and a Gln residue (Ren et al. 1997) (active site signature of mammalian Gpxs, which is highly conserved), while the immediate neighbouring residues of the catalytic Cys at the active site pocket of Prxs are Thr and Arg (Wood et al. 2003). As expected, the rPtGpx5 displays the same overall fold as classical mammal Gpxs and Prxs. Interestingly, the rPtGpx5 possesses an active site cleft architecture (consisting of Cys, Glu and Trp) very similar to that reported for

mammal SeCys-Gpx structures, which clearly distinguishes rPtGpx5 from Prxs. On the other hand, when rPtGpx5 is superimposed onto the thiol peroxidase structures (which belong to 2-Cys Prxs), the locations of both C_P and C_R are similar in structure, although the proteins have a very low level of sequence identity. This observation suggests the importance of the positions of these Cys residues in order to facilitate the peroxide reduction and the regeneration process. Flohé et al. (2003) suggested that thiol-dependent peroxidases care little about homology-based assignments of function (Flohé et al. 2003). Taking into account previous studies (Navrot et al. 2006, Maiorino et al. 2007) and the present crystallographic evidence, we conclude that PtGpx5 is actually a thioredoxin peroxidase, structurally related to glutathione peroxidases but exhibiting catalytic and Trx-dependent recycling mechanisms of peroxiredoxin. The historical term of Gpx can be used accurately only to describe a subgroup of Gpx family (Maiorino et al. 2007) that uses GSH as electron donor despite the high level of similarity between the sequences among the family members.

1.2 Oligomerization mode of PtGpx5

Recent biochemical studies (Navrot et al. 2006) showed that poplar Gpxs are non-covalent dimeric enzymes in solution. This dimeric arrangement, which is found in both the reduced and oxidized form, exhibits a pseudo antiparallel β-sheet that is stabilized mostly by hydrophobic clusters. Most of the residues involved in the dimerization are conserved in plant Gpxs, suggesting that similar form of enzyme could be observed for all plant Gpxs. Among the residues involved at the dimer interface of PtGpx5, Tyr151 is structurally an important residue (Table 4.2). It makes

several critical contacts with residues (particularly Asp129, Gln132 and Arg149) from another subunit and is also a part of an aromatic cluster (see Results).

To date, all classical mammalian Gpxs characterized are tetrameric enzymes (Figure 4.10(b)) except PHGpx, a mammal Gpx-4 that is monomeric. The sequence alignment (Figure 4.1) indicates that the region coding for the oligomerization loop (located in the core of the tetramer, Figure 4.10(b)) in typical tetrameric mammal Gpxs (positions 125–142 in the bovine Gpx-1 sequence) is absent from all plant Gpxs as well as mammal Gpx-4, suggesting that these enzymes might not be tetrameric. Our structures reveal that the PtGpx5 dimer cannot be formed by other mammal Gpxs due to the structural differences that would cause steric hindrances at the dimerization interface. For example in mammal Gpx-1, Gpx-2, Gpx-3 and Gpx-5, part of the oligomerization loop covers the C-terminal part of the β5 strand that is the dimerization interface in PtGpx5. In addition, mammal Gpx-3 possesses a longer N-terminal chain that completely covers the β5 strand, making this enzyme unable to form the same dimer as PtGpx5. However, in the mammal monomeric Gpx-4 (Ursini et al. 1995) there is neither the oligomerization loop nor the extra-long N-terminal chain, yet this enzyme cannot form the dimeric pattern of PtGpx5. Further structural comparison reveals that a loop (residues Lys111–Ala126) in mammal Gpx-4 obstructs the formation of a β-sheet interaction between two subunits and hinders the dimerization.

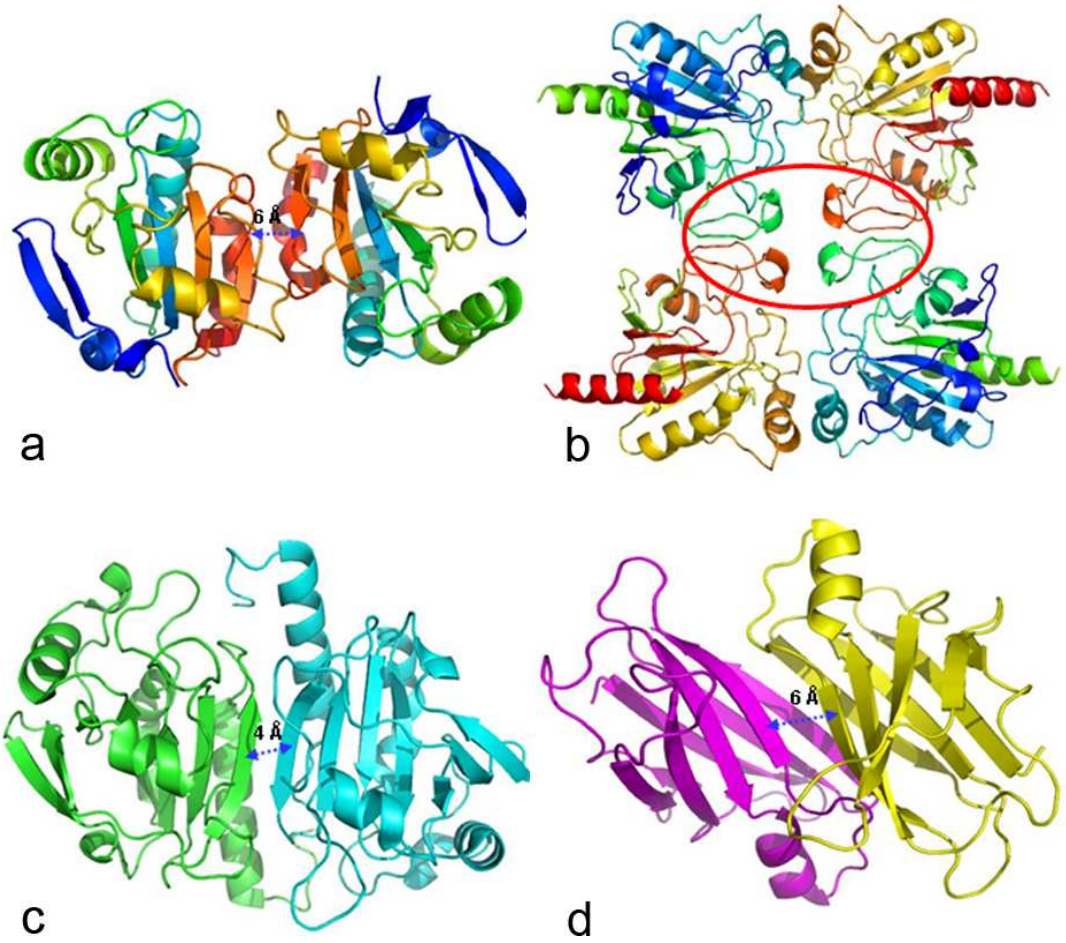


Figure 4.10 Cartoon representation of different oligomerization states of enzymes. These structures include (a) the dimeric plant PtGpx5 in its reduced state (PDB code 2P5Q), (b) the classical tetrameric human plasma Gpx-3 (from Professor Rudolf Ladenstein), (c) the mammalian 2-Cys peroxiredoxin heme-binding protein, HBP23 (PDB code 1QQ2) and (d) the human transthyretin or prealbumin (PDB code 2PAB). Note that the presence of oligomerization loops, circled in red, in the human plasma Gpx-3 structure in (b) is likely to cause steric hindrance and obstruct the dimer formation if this feature exists in PtGpx5. In the 1QQ2 structure in (c), a true ten-stranded β-sheet is observed (the intermolecular β-sheet distance is indicated) while in (a) the rPtGpx5 dimer a pseudo ten-stranded β sheet similar to that observed (d) for the 2PAB structure is seen (see the text).

Similar dimer formation by antiparallel association of two β-strands has been reported for typical 2-Cys and 1-Cys Prxs (Declercq et al. 2001, Wood et al. 2003). However instead of a central true ten-stranded β-sheet as observed in Prx (Figure 4.10(c)), dimerization of PtGpx5 results in a central pseudo ten-stranded β-sheet

(Figure 4.10(a)) due to the relatively large distance between two β-strands (~ 6 Å apart) at the dimer interface that does not allow backbone hydrogen bond formation.

Therefore, the hydrophobic and aromatic networks play an important role in forming the dimer of PtGpx5 independently of its redox state. This rather unusual kind of β-sheet interactions has already been described in an amyloid sheet (Serag et al. 2002, Hwang et al. 2004, Lei et al. 2006), and in the prealbumin protein (Blake et al. 1978). Sharing similar dimer properties with prealbumin, PtGpx5 dimer also produces an apparently more open quaternary structure due to the greater separation of the β-sheet at the dimerization site (Figure 4.10(d)).

1.3 Comparison between reduced and oxidized form PtGpx5: redox-driven conformational changes

Resolution of the reduced and oxidized structures of PtGpx5 allowed us to examine how the intramolecular disulfide bond can be formed in plant Gpxs. Upon scavenging peroxide, the resulting peroxidatic Cys44-SOH intermediate will react rapidly with the resolving Cys92 to generate a stable disulfide bond at the active site of oPtGpx5. This reaction involves two conformational changes (Figure 4.11): (i) first, the bowing of the loop that connects β1 and α1 (where peroxidatic Cys44 is located) by about 90° towards Cys92 (in α2 helix). The flipping of this helical turn causes the peroxidatic Cys44 in rPtGpx5 to move ~ 10 Å before the formation of an intramolecular disulfide bond. (ii) Second, the unwinding of the whole α2 helix bringing Cys92 closer to Cys44. The absence of α2 helix in the oPtGpx5 structure (Figure 4.2) results in a long flexible loop that now connects β2 directly to β3. Due to its flexibility, a loop motion seems possible, leading to a “closed” conformation where the active site of oPtGpx5 is covered. We speculate that this motion is needed

to cover/ uncover the active site cavity of oPtGpx5 during the recycling process of this enzyme by reduced Trx.

Similar extensive conformational changes could occur for the other plant Gpxs, since they share closely similar amino acid sequences and biochemical properties. Comparatively, the conformational changes observed in the present study are much more drastic than those described for atypical 2-Cys Prxs and prokaryotic thiol peroxidases so far. Indeed, in atypical 2-Cys Prxs, the change of redox state is accompanied by limited conformational changes, a helical turn of peroxidatic-Cys- (C_P-) and resolving-Cys- (C_R-) loops (Choi et al. 2003).

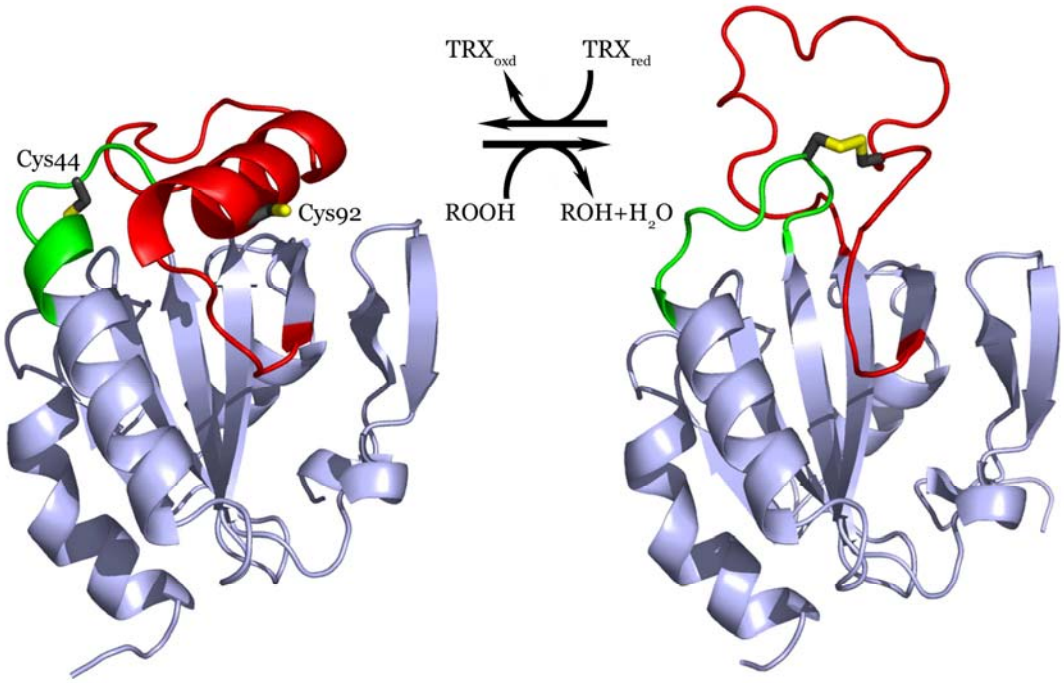


Figure 4.11 Cartoon representations of rPtGpx5 (left) and oPtGpx5 (right). A remarkable redox-driven conformational change is a crucial step for thioredoxin-mediated recycling process in poplar thioredoxin peroxidase. Two helical regions that undergo drastic structure rearrangements are coloured in red and green, respectively (see text). Both cysteines (Cys44 and Cys92) involving in the catalytic mechanism are highlighted in sticks.

The α2 helix unwinding of rPtGpx5 shows a noticeable analogy with ribosomal protein L20 (Timsit et al. 2006) and calmodulin (CaM) (Babu et al. 1988), which exhibit a helix–coil transition. The flexible regions of L20 and CaM share a similar sequence motif that confers helical instability, a cluster of charged amino acid residues distributed on the same helical face and spanning three helical turns (Figure 4.12). In the α2 helix of rPtGpx5, Cys92 and two negatively charged residues, Asp85 and Asp89, are located at each helix-turn, with all side-chains oriented facing the same side of the α2 helix. Once Cys92 is present in its thiolate form during catalysis, the emerging negatively charged cluster of α2 helix (Figure 4.12(c)) tends to make it unstable. This may lead to the complete unwinding of the α2 helix, as observed from the differences between the reduced and the oxidized form of PtGpx5. Interestingly, these two negatively charged residues are conserved or replaced by homologue residues in plant Gpxs, suggesting that similar conformational changes could happen for the other Gpx homologues.

In all three cases (PtGpx5, L20 and CaM), sidechain electrostatic repulsion could be responsible for helical instability, and therefore constitute a hot spot for α helix unwinding. This helix–coil transition was suggested to confer to CaM the required plasticity for binding to an exceptional variety of distinct partners (Hoeflich & Ikura 2002). We suggest here that the drastic structural rearrangements observed in PtGpx5 are needed to allow disulfide bond formation and facilitate the recognition of its reducing partner, as also proposed for methionine sulfoxide reductases (Boschi-Muller et al. 2005), enzymes that repair the oxidized methionine sulfoxide derivatives generated from ROS (Boschi-Muller et al. 2005, Moskovitz et al. 2001).

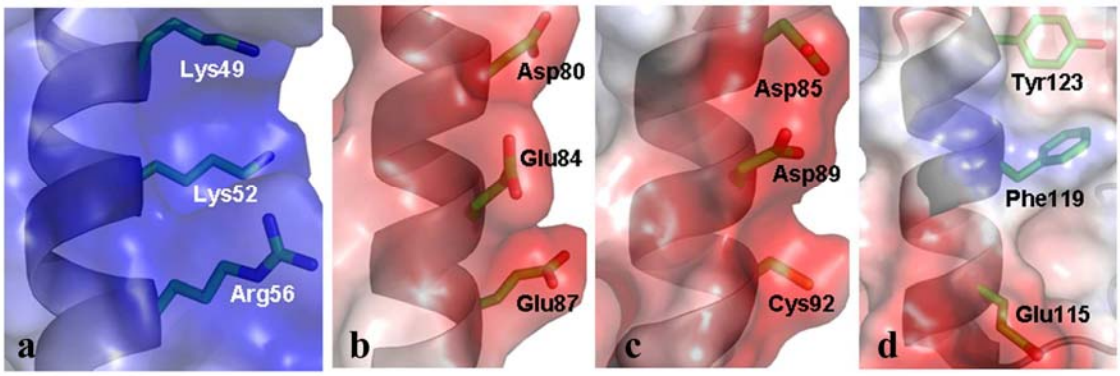


Figure 4.12 Cartoon representations explaining the driving force of the unwinding of α-helices 2. Local surface charges of α-helix 2 of (c) rPtGpx5 and its corresponding helix of a mammalian Gpx structure example ((d) the monomeric human Gpx-4, PDB code 2GS3). The corresponding α2 helices in the ribosomal protein L20 (PDB code 2GHJ) and in the calmodulin (PDB code 1CLL) that also exhibit helix–coil transitions are shown in (a) and (b), respectively. Negatively and positively charged regions are shown in red and blue, respectively, while the rest is shown in grey. Three residues, situated at each helix turn and facing the same side of the helix, are displayed as sticks.

1.4 Active site differences in comparison to mammalian Gpxs

The rPtGpx5 possesses a Trp (Trp133), a Cys (Cys44) instead of a SeCys and a Glu (Glu79) instead of a Gln at the same positions and geometries as observed in the classical mammalian Gpx active site (Figure 4.13). There is studies show that plant Gpxs are far less reactive that their animal counterparts, most likely due to the lower reactivity of Cys, compared with SeCys (Eshdat et al. 1997). However, replacing the Cys by SeCys did not restore a high Gpx activity, as demonstrated in a citrus Gpx (Hazebrouck et al. 2000). Hence, other differences between animal and plant enzymes could be responsible for their different biochemical properties.

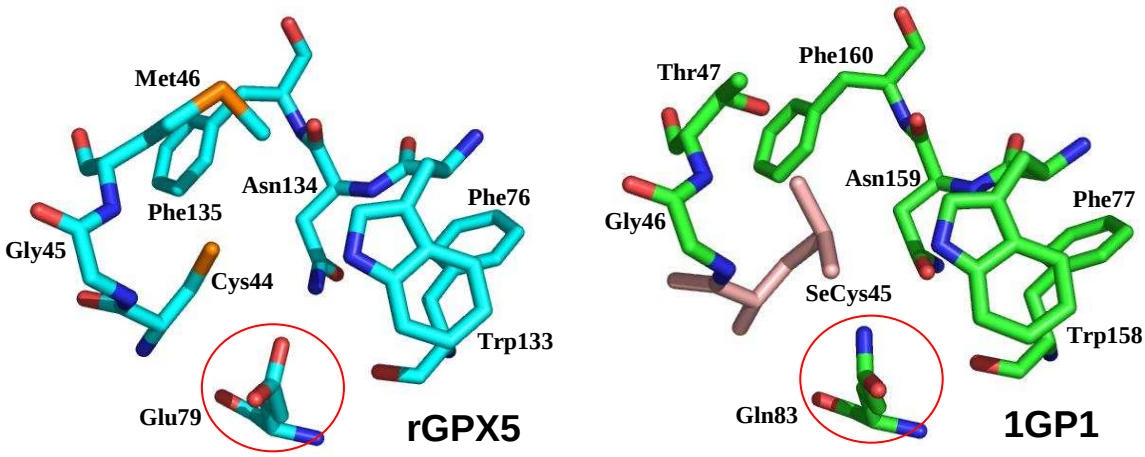


Figure 4.13 Active sites of the reduced poplar Gpx5 (rPtGpx5) and the reduced *Bos taurus* selenoenzyme glutathione peroxidase 1 (PDB code 1GP1). Residues involved are labelled and coloured according to the residue type for clarity. An important residue suggested to be involved in the catalytic mechanism of mammalian Gpxs (conserved Gln residue) is substituted with a Glu residue in plant Gpxs (circled in red).

In animal Gpx, the conserved active site Gln residue (Gln83, numbering in 1GP1; Figure 4.13) has been proposed to participate in the catalytic cycle based on experimental evidence (Dietz 2003). However, discrepancy does exist in the recent propsal based on a computational study (Sztajer et al. 2001) which suggested that the Gln83 plays a key role as a proton acceptor. The fact that Gln is a polar uncharged residue by nature, it can neither donates or accepts a proton. But the presence of a Glu residue, replacing a Gln at the same position in PtGpx5 suggest that this residue might play the role of a general base catalyst in the catalytic process since (i) it is within the hydrogen bonding interaction with the peroxide Cys and Trp in the active site of rPtGpx5, and (ii) Glu is potentially deprotonated at basic pH. Nevertheless, the possibility of Glu deprotonation caused by a pKa shift at the active site local environment upon reactive substrate binding is not excluded.

In this study, we have shown that a conformational change has taken place (see previous section) during the transition of the rPtGpx5 active site to the oPtGpx5 active site by the intervention of Cys92 (the resolving Cys). Residues around the resolving Cys92 in PtGpx5 are functionally conserved in plant Gpxs, but share very low similarity with the corresponding region in mammalian Gpxs. Indirectly, the low homology at the α2 helix where Cys92 is situated, between plant and mammal Gpxs could explain why the plant Gpxs prefer Trx as reducing substrate (see discussion below).

1.5 Possible interactions between PtGpx5 and thioredoxin *h*1 from poplar

Oxidized PtGpx5 is regenerated by Trx (Navrot et al. 2006) and the Trp residue adjacent to the catalytic Cys of Trx plays an important role in Trx-substrate recognition. Indeed, Menchise et al. (2000-2001) suggested that the Trp residue at the active site of Trx has a dual function both to force the active site in the bioactive conformation and to mediate protein–protein recognition. We therefore simulated two model complexes of oPtGpx5-Trx *h*1. The poplar Trx *h*1 enzyme exhibits an unusual active site with the sequence WCPPC rather than the typical WCGPC motif. The Trx *h*1 structure (PDB code 1TI3), however, adopts the classical Trx fold, in general, although it appears to be more rigid (Coudevylle et al. 2005). The major interaction force in Gpx5-Trx *h*1 complex is derived from two charge interactions: negative on the oPtGpx5 domain and positive on the Trx *h*1 domain.

Smeets et al. (2005) have proposed a simulated complex structure between both a reduced form of human mitochondrial Trx 2 (hTXN2) and hybrid human Prx 5 (hPRDX5). However in their proposed complex structure, a direct interaction between two cysteines (catalytic $Cys31_{hTXN2}$ and catalytic $Cys47_{hPRDX5}$) seems

impossible because of the large distance between them (17 Å). Hence, an unwinding of the N-terminal part of the α2 helix containing $Cys47_{hPRDX5}$ is postulated to be necessary to bring the two cysteine residues much closer to each other in order to facilitate a mixed disulfide bond formation (Choi et al. 2003, Smeets et al. 2005, Evrard et al. 2004).

Our molecular dynamics simulations suggest that the complex that contains an intermolecular disulfide bond between the resolving Cys92 of oPtGpx5 and catalytic Cys38 of Trx *h*1 (complex 2) is more favourable in terms of calculated binding free energy. This result corroborates recent evidence that a disulfide can be formed between the resolving Cys (Cys91) of the oxidized *Drosophila melanogaster* (Dm) Gpx and the catalytic Cys (Cys32) of a reduced Trx C35S mutant (Maiorino et al. 2007). In this study, we provide structural information describing for the first time the Gpx–Trx interface and the possibility of a transient intermolecular disulfide bond formation. Taking both complexes into account, we propose a Trx recognition motif found in the majority of plant Gpx sequences, the -$_{80}$EPGxx(D/E)xIxx(F/M)(V/A)CT(R/K)FK$_{96}$- motif (where x represents any residue, Figure 4.1), whereas the identified Gpx recognition motifs in Trx, -$_{36}$SWC(P/G)PC(K/R)$_{42}$- and -$_{64}$D(V/T)DELK$_{69}$- (numbering based on the 1TI3 structure), is conserved also in the Trx *h*1 subgroup.

The protein–protein interaction illustrated in the model of oPtGpx5 and Trx *h*1 complex is in agreement with the suggestion that a cluster of charged groups on one side of the S-S bond of Trx participates in the mechanism of the redox reactions or to substrate recognition (Eklund et al. 1984). In contrast, there is also a flat and hydrophobic surface on the other side of the S-S bond in Trx, which is suggested to be the interaction area with thioredoxin reductases (Slaby & Holmgren 1979, Thelander 1974) via ping-pong mechanisms (Holmgren 1979, Fortuniak et al. 1996).

Hence, based on the model of complex oPtGpx5-Trx *h*1 in this study, we conclude that the molecular surface of Trx interacting with its substrates, for example plant Gpxs, is different from the site of Trx interacting with its regenerating partner, Trx reductase.

1.6 Protective role of PtGpx5 against Cd^{2+}

Cadmium is non-redox-reactive but strongly hazardous to metal-sensitive enzymes, and its toxicity is believed to be due to the depletion of GSH and sulfhydryl groups in proteins (Fortuniak et al. 1996, Zenk 1996, Stohs et al. 2001), contributing indirectly to oxidative stress by affecting the cellular thiol redox balance. Vido et al. (2001) showed that the expression of proteins involved in protection of the cell against peroxides and superoxides is increased upon exposure to Cd^{2+}. In response to Cd^{2+} stress, higher plants utilize a number of defense systems, such as retention in cell walls, binding by organic molecules in the cytosol and sequestration in the vacuole (Vázquez et al. 2006). Retention of Cd^{2+} in the cell wall is the first barrier against Cd^{2+} stress, operating mainly at the root level (Nishizono et al. 1989). Several studies have described the key role of plant cell walls in reducing Cd^{2+} accumulation in the protoplast (Vázquez et al. 2006, Al-asheh & Duvnjak 1999, Khan et al. 1984, Lozano-Rodríguez et al. 1997). Once Cd^{2+} enters the cytosol, phytochelatins are believed to efficiently sequestrate the heavy metals (Cobbett & Goldsbrough 2002), thus conferring protection to heavy-metal-sensitive enzymes.

Thiol-rich compounds such as GSH are also proposed as Cd^{2+}-detoxifying compounds in plants (Howden et al. 1995, Vogeli-Lange & Wagner 1990). However GSH depletion in response to Cd^{2+} has been reported in several plant species (Steffens 1990). Alternatively, Trx from the thiol-disulfide oxidoreductase family has

been proposed to be an efficient Cd^{2+} chelator upon acute Cd^{2+} exposure (Rollin-Genetet et al. 2004, Lemaire et al. 1999). In a more relevant context, yeast GPX-3 over-expression was shown to be effective in raising the threshold of cadmium tolerance (Avery et al. 2004). Moreover, expression of Gpx-3^{C82S} did not complement the cadmium sensitivity of the Gpx3Δ strain, corroborating the notion that a peroxidase activity of Gpx-3 is required for cadmium resistance. Beside cadmium, yeast Gpx-3 also displays phospholipids-hydroperoxidase-dependent activities for Cu resistance (Avery et al. 1996). However, none of the Gpx-3 dependent activities is important for Cr resistance (Sumner et al. 2005).

The present reduced form structure (rPtGpx5) displays the ability of two dimers to interact with 32 atoms of Cd^{2+} in the asymmetric unit (Figure 4.9(a)) in which eight of them are trapped independently by a dimer, suggesting that this enzyme could play an additional role as a Cd^{2+}-sink in defence against heavy metals, particularly Cd^{2+}. Four potential cadmium-binding sites are identified in each subunit (see Results). It is worth mentioning that the first three cadmium-binding sites, including the one at the active site, involve residues that are conserved in other plant Gpx homologues. The cadmium ion captured at the active site also has an important role in locking the reduced conformation of PtGpx5 (Figure 4.5(a)) since this enzyme oxidizes spontaneously in solution. The fourth cadmium binding site, however, involves non-conserved residues in plant Gpxs, thus suggesting that PtGpx5 may be an efficient Cd^{2+}-sink as compared to other isoforms. Preliminary biochemical assay had been performed to testify this proposition and initial data showed that *E. coli* carrying recombinant PtGpx5 gene is more resistant towards Cd^{2+} toxicity (Navrot et al., *unpublished results*). Recently, Navrot et al. (2006) have demonstrated that when poplar leaves are subjected to metal stresses such as the presence of high concentrations of cadmium or copper, the levels of Gpx are modified. Likewise,

plant Gpx5 homologues could play a broader role as a heavy-metal sink than just as a peroxide scavenger.

Part II: TRX

A. RESULTS

1. Specific objectives

PtTrx*h*4 belongs to the third subgroup of Trx *h* and exhibits the usual WCGPC active site. Besides these two cysteinyl residues (Cys58 and Cys61), members of this subgroup exhibit an additional conserved cysteine (Cys4) found in the fourth position in an N-terminal extension (comprising the 24 most N-terminus residues). Since members of this subgroup exhibit an N-terminal extension containing a conserved cysteine in the fourth position as compared to subgroup 1, essential questions regarding the role(s) and effect(s) of this extension and its additional cysteine (Cys4) for protein structure and functions arise. It was of interest to know how this extension is positioned with respect to the conservation of the Trx fold and more importantly to understand why this protein does not react with its typical reducing partner, NTR. The unique properties of PtTrx*h*4 are analyzed in this manuscript and it is demonstrated that, in contrast to other Trxs, three cysteines rather than two are involved in the catalytic mechanism. Resolution of wild type and mutant PtTrx*h*4 structures allowed us to localize the peculiar N-terminal extension and to clearly identify the formation of an alternative intramolecular disulfide bridge between the cysteines in position 4 and 58. Several amino acids near the active site which might be important for Trx-target protein interactions are also highlighted.

2. Results obtained

2.1 Sequence alignment

Sequence analysis (Figure 4.14) using several localization prediction programs suggested that the N-terminal extension does not correspond to a signal peptide and that PtTrx*h*4 would be cytosolic. From the sequence alignment (Figure 4.14), it is obvious that PtTrx*h*4 and its orthologs share very high sequences identity, especially the active site sequences ($_{57}$WCGPC$_{61}$) and a suggested loop for substrate binding ($_{76}$DVD[D/E]L$_{80}$) which is located in between β3 and α3 (Koh et al. 2007). Notice that PtTrx*h*4 contains a conserved aromatic sequence motif ($_{93}$PTFFF$_{97}$) at the C-terminus of β4. This motif is absent in non-PtTrx*h*4 type Trxs (data not shown).

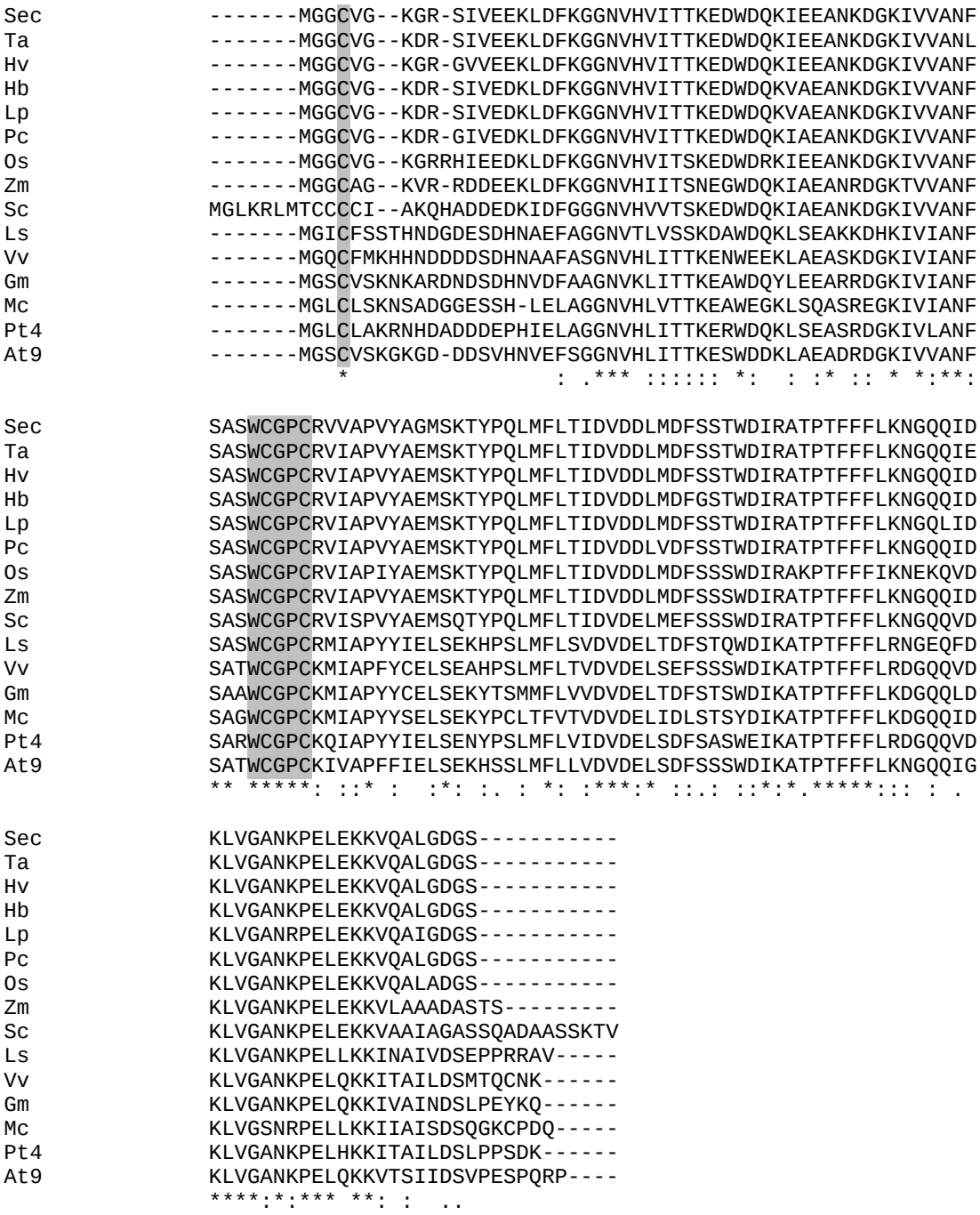


Figure 4.14 Multiple alignment of poplar Trx*h*4 related sequences. Amino acid sequences were aligned using CLUSTALW software. Accession numbers are as follows: At9: *Arabidopsis thaliana* (At3g08710); Gm : *Glycine max* (CA799351) ; Hb : *Hordeum bulbosum* (AF159385) ; Hv : *Hordeum vulgare* (AF435815) ; Lp : *Lolium perenne* (159387) ; Ls : *Lactuca sativa* (TC9851) ; Mc : *Mesembryanthemum crystallinum* (CA838461) ; Os : *Oryza sativa* (AF435817) ; Pc : *Phalaris coerulescens* (AF159388) ; Pt : *Populus trichocarpa* (BU867450) ; Sb : *Sorghum bicolor* (TC72759) ; Sc : *Secale cereale* (AF159386) ; Ta : *Triticum aestivum* (AF438359) ; Vv : *Vitis vinifera* (CB004453) ; Zm : *Zea mays* (AF435816). The asterisk corresponds to strict identity, the colon corresponds to functional homology, and the period corresponds to structural homology.

It was thought that perhaps the presence of the extended N-terminal of PtTrx*h*4 might have some structural effects, e.g. introduce some minor conformational changes, to the overall folding which may or may not consequently precluded structure solution by molecular replacement. Apparently, initial molecular replacement attempts using relatively high sequence homology protein models (PDB codes 1XFL, 1WMJ, 1TI3) as templates were fruitless. Therefore, a SeMet-labeled protein was prepared and a SAD dataset was collected.

2.2 Crystal description and structure refinement statistics

SeMet-PtTrx*h*4$_{oxd}$ and WT-PtTrx*h*4$_{oxd}$ crystals are tetragonal and hexagonal, respectively (both designated form A) while the C61S-PtTrx*h*4$_{oxd}$ crystal is orthorhombic (designated form B). Based on the Matthew's coefficient calculation using the cell content analysis program in the CCP4 software package version 6.0.2, one and two subunits per asymmetric unit were identified for SeMet-PtTrx*h*4$_{oxd}$ and WT-PtTrx*h*4$_{oxd}$ crystals, respectively, with the V_M value of 1.75 Å^3/Da (solvent content 29.62 %) and 2.03 Å^3/Da (solvent content 39.36 %). Form B belongs to space group $P2_12_12_1$, with unit-cell parameters a = 31.78, b = 44.10 and c = 85.68. The presence of one subunit per asymmetric unit in form B crystal gives the V_M value of 1.91 Å^3/Da (solvent content 35.78%). Crystallographic data were refined only for non-SeMet-labeled protein structures. Details of data refinement statistics are summarized in Table 4.5.

Table 4.5 Data refinement statistics for WT PtTrx*h*4 and PtTrx*h*4C61S crystals.

Data Set	**WT PtTrx*h*4**	**PtTrx*h*4 C61S**
		
Resolution range (Å)	41.0 - 2.15	30.0 - 1.6
Reflections used	12725	15470
R_{cryst}[a] (R_{free})[b]	20.97 (28.53)	19.06 (19.63)
Protein / waters / PO_4	2 x (111 residues) / 63 / 0	129 residues / 115 / 1
Average *B*-value for individual chain (Å^2)	56.27 / 56.61 (subunit A / B)	20.46 (subunit A)
Mean B factor		
Main-chain (Å^2)	55.20	19.40
Side-chain (Å^2)	57.67	21.52
Water (Å^2)	59.29	32.47
All (Å^2)	56.53	21.69
Rms deviation from ideal geometry		
Bond lengths (Å)	0.030	0.010
Bond angles (°)	2.53	1.40
Dihedral angles (°)	26.30	24.10
Improper angles (°)	5.59	1.46
Ramachandran plot		
Residues in most favored regions (%)	91.5	92.8
Residues in additionally allowed regions (%)	7.5	7.2
Residues in generously allowed regions (%)	1.0	0.0

[a] $R_{cryst}=\Sigma|F_o-F_c|/\Sigma F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.
[b] The R_{free} value was calculated from 5 % of all data that were not used in the refinement.

2.3 Quality of models and overall description of structures

The two PtTrxh4 models presented in this study display intramolecular disulfide bridges between the catalytic Cys58 with either Cys61 (WT PtTrx*h*4, despite the crystallization in the presence of DTT) or with Cys4 (PtTrx*h*4C61S mutant). Here, unless indicated otherwise, form A structure refers to the WT PtTrx*h*4 while form B structure refers to the PtTrx*h*4C61S mutant. Generally, both forms of PtTrx*h*4 structures retain the overall Trx fold, which is shared by all members of the thiol-disulfide oxidoreductase family, with a five-stranded β sheet surrounded by four α helices in a βα βα βα ββα topology (Figure 4.15).

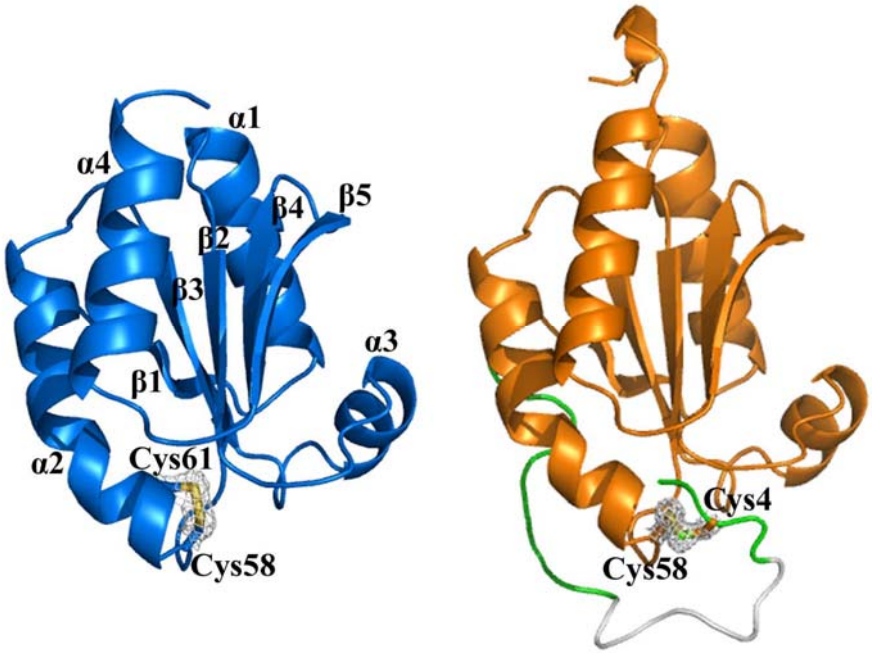


Figure 4.15 Cartoon representation showing overall structures of the WT PtTrx*h*4 (blue) and the PtTrx*h*4C61S (orange). The intramolecular disulfide bond in each structure is highlighted, with final $2F_o$-F_c electron density (1.2 σ level) covering chosen residues for clarity. The N-terminal extension (residues 1 - 24) is not modeled in the WT PtTrx*h*4 structure (right) due to the lacking of electron density, whereas in the PtTrx*h*4C61S partial of the N-terminal extension (residues 2 – 6 and 16 – 24, colored in green) can be assigned to the structure (left, see text for details). In the latter structure, lacking residues (7 – 15) are shown in an imaginary gray loop for a better overview of the structure. Relevant residues and secondary structures are labeled. The figure was prepared using PyMOL.

In the form A structure, the electron density is well-defined along the main-chain and for most of the side-chains except that the 24 most N-terminal amino acids of the WT enzyme could not be located due to weak electron density, thereby suggesting a high degree of structural flexibility of this part of the molecule. Hence, this structural model comprises only residues 25 – 133 of PtTrx*h*4 (109 amino acids). Nevertheless, the extended N-terminal residues are present in the form A crystal as verified by mass spectroscopy of the initial protein sample. The oxidized form A is present as a dimer in the asymmetric unit with active site regions facing each other. The active site architecture is similar to other known Trxs. One side of the redox-active disulfide bond is flat and rather hydrophobic (formed by residues Ala55, $_{59}$GP$_{60}$, $_{99}$ATP$_{101}$, $_{116}$VGA$_{118}$; named as hydrophobic site 1; Figure 4.16), while the other side is more shielded with a pitcher-like cleft predominantly hydrophobic (this cleft is formed by Val26, Leu28, Asn52, Ser54, $_{64}$IAP$_{66}$, $_{68}$YI$_{69}$, Val82, Phe103, Leu115, Leu123; named as hydrophobic site 2; Figure 4.16). At the edge of the pitcher-like cleft lies a positively-charged patch, $_{56}$RW$_{57}$ and Lys62. The hydrophobic site 2 is accommodating hydrophobic residues from a symmetry-related molecule.

On the other hand, form B model consists of 129 amino acids (residues 2 – 139, excluding residues from 7 to 15), including the presence of the N-terminal extension (Gly2 – Gly24) and the protein crystallizes as a monomer in the asymmetric unit. Upon formation of Cys4-Cys58 disulfide bond, the active site of form B is covered by the N-terminal extension. The electron density between residues 7 and 15 was barely visible, so no residue was assigned to this assumed highly flexible loop region. Nevertheless, the disulfide bond is undoubtedly present (Figure 4.15). The N-terminal is kinked from the rest of the molecule at the double-glycine region ($_{23}$GG$_{24}$) and this extension is further stabilized by making extensive hydrophobic interactions which involve residues

I19 and $_{21}$LAGG$_{24}$ with the residues located at hydrophobic site 2. Apart from the hydrophobic interactions, the N-terminal extension is also stabilized by hydrogen bonding (Val26N-Gly23O, Ile19N-Pro17O, Thr100N-Cys4O, Cys4N-Thr100O, Ala118N-Gly2O, Ile19O-Arg56N$_{\eta 1}$). Upon closure of the N-terminal towards the core of the protein, the molecular surface property of hydrophobic site 1 remains unchanged (just as described in form A structure, see Figure 4.16a and 4.16c for comparison) while the initially deep hydrophobic surface of hydrophobic site 2 is now shielded by residues from the N-terminal extension (Pro17 – Gly24) and concomitantly the depth of the pocket is greatly reduced see Figure 4.16b and 4.16d for comparison). In other words, half of the pitcher-like cleft wall (hydrophobic site 2) has now become more hydrophilic in the form B structure (Figure 4.16d).

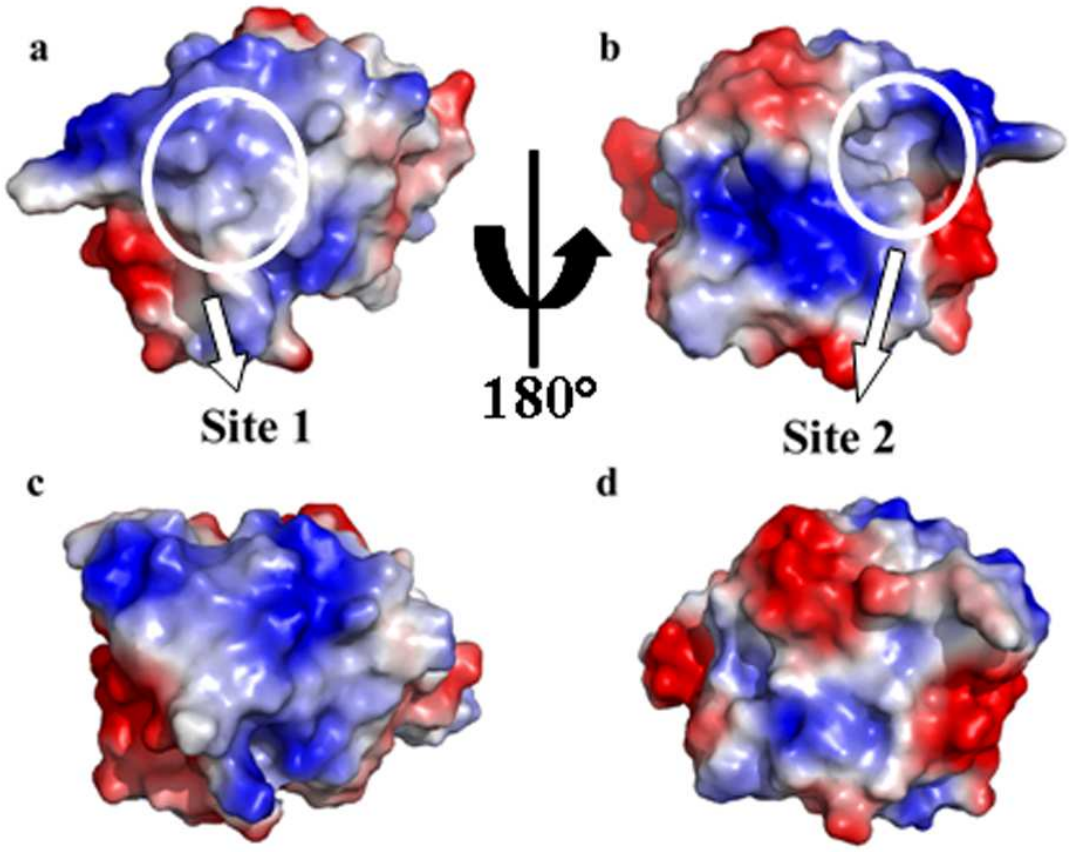


Figure 4.16 Electrostatic potential surface charges of WT PtTrx*h*4 (top panel, a and b) and PtTrx*h*4C61S mutant (bottom panel, c and d), viewing from both sides of the molecule (difference by 180°). These molecules are corresponding to the view from the bottom of the molecule (orthogonal to the view of the molecule at the right panel in Figure 4.15). Two hydrophobic sites, site 1 (left, a) and site 2 (right, b), are circled for clarity. Corresponding sites in the mutant can be found in the bottom panel (c and d). Notice that site 1 cleft is shallower than the pitcher-like hydrophobic site 2 (for WT PtTrx*h*4). The N-terminal extension (residues 1 - 24) is not modeled in the WT PtTrx*h*4 due to lacking of electron density at the corresponding region. Likewise for PtTrx*h*4C61S, only partial N-terminal extension (excludes residues 1 and 7 - 15) can be modeled for the same reason (see text for details). The figure was prepared using PyMOL.

2.4 Structural comparison with other thioredoxins

If the first 24 amino acids are excluded, superimposition of form A and form B structures yield r.m.s.d. values of 0.6 Å (for 110 common C^{α} atoms). The positions of the active site cysteines (or Ser61 in form B structure) remain unchanged in both structure forms. Form A and form B structures exhibit different conformations in their active site regions with the most obvious difference observed at residue Trp57. In the form B structure, the position of Trp57 (χ_1 value = +50.01°, residue stabilized by making a hydrogen bond with the $O_{\delta 2}$ atom of a conserved residue Asp86) is the same as in most of other classical Trxs (χ_1 values between +40° and +55°) and its side-chain covers an important part of the active site (Cys4-Cys58 disulfide bond) surface. This residue has been shown to be important for the protein-protein interaction or recognition (Menchise et al. 2000-2001). On the contrary, the plane of the indole ring of Trp57 (χ_1 value = -81.87°) in the form A structure has to “flip out” by about 90° with respect to its usual position as described in form B structure and orients towards the surface of the protein. In the form A structure, Trp57 of two independent monomers are facing each

other. This residue possesses an average elevated B factor (72.55 Å^2) and the indole side-chain lacks of electron density, suggesting the flexibility of the side-chain. The flipping out of Trp57 and the active site conformation in form A structures is similar to that observed in Trx *f* (short form) from spinach chloroplast (PDB code 1F9M) and in the Trx from *Trypanosoma brucei* (PDB code 1R26). Figure 4.17 displays a superimposition of the active-site region close-up view of form A PtTrx*h*4, form B PtTrx*h*4 and other mentioned Trx structures, highlighting the position and orientation of the active site Trp residue. The atypical position of Trp57 in form A is not retained in form B mainly due to the constraint of the N-terminal.

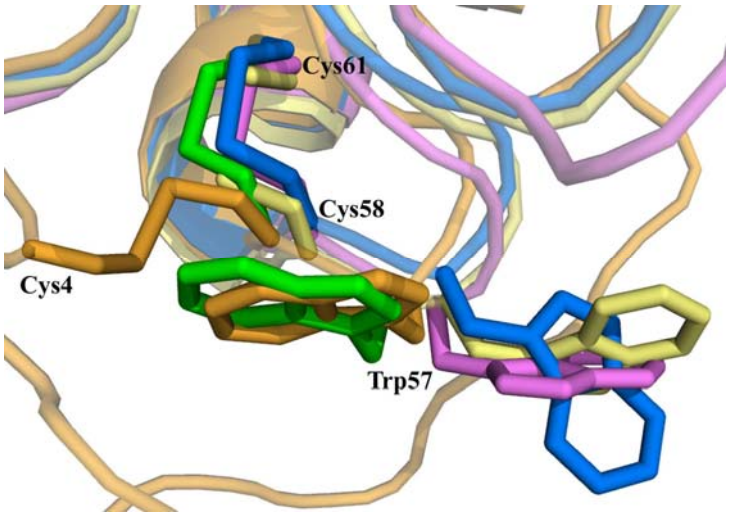


Figure 4.17 Superimpostion of the active site cysteines and tryptophan from poplar WT PtTrx*h*4 (blue) (PDB code 3D21), poplar PtTrx*h*4C61S (orange) (PDB code 3D22), *C. reinhardtii* (green) (PDB code 1EP7), *T. brucei* (yellow) (PDB code 1R26) and spinach Trx*f* short form (pink) (PDB code 1F9M). For PtTrx*h*4, Cys58 is the catalytic cysteine (corresponding to $Cys36_{1EP7}$, $Cys46_{1F9M}$ and $Cys30_{1R26}$), Cys61 is the first resolving cysteine (corresponding to $Cys39_{1EP7}$, $Cys49_{1F9M}$ and $Cys33_{1R26}$) while Cys4 is the additional N-terminal cysteine or the second resolving cysteine in the proposed atypical mechanism of PtTrx*h*4 (see text for details). Contrary to the typical position in most Trxs, exemplified by Trp35 of green alga Trx *h* (green) and Trp57 of poplar WT PtTrx*h*4 (blue), the Trp57 of the PtTrx*h*4C61S mutant has a flipped-out position. This latter unusual Trp side-chain position is also adopted in two other Trx structures mentioned ($Trp45_{1F9M}$ and $Trp29_{1R26}$). The figure was prepared using PyMOL.

Apart from the flexibility of Trp57, superimposition of monomers of form A and form B also revealed that its neighbour residue, Arg56, also possesses certain degree of flaccidity. This latter side-chain has moved in towards the molecule by about 40° when comparing the A and B structures. The movements of these two residues actually make way for the N-terminal extension of PtTrx*h*4 to gain access to the active site of the enzyme. Another discrepancy between the two forms is the direction of the side-chain of Ser54. Contrary to the A form where it is more solvent-exposed, the O_{γ} atom of Ser54 is hydrogen bonded to the $O_{\delta 1}$ atom of Asn52 (~ 2.7 Å) in the B form. In the B form structure, the O_{γ} atom of Ser54 is also hydrogen bonded to the O_{γ} atom (~ 3 Å) of Ser61 (this residue is originally a cysteine residue). This serine residue is highly conserved in PtTrx*h*4 homologues, but usually replaced by a threonine in other plant *h*-type Trxs or an aromatic residue (Phe or Trp) in other non-*h*-type Trxs.

Another apparent interesting difference between PtTrx*h*4 and other Trx structures is that Val82 of PtTrx*h*4 which is situated in the hydrophobic site 2 mentioned earlier (see previous section) is substituted by a lysine residue in all other Trxs (corresponding to Lys57 in *E. coli* Trx). In all plant Trxs *h* of subgroup 3, a hydrophobic residue is always present at this position in order to maintain the hydrophobic environment of hydrophobic site 2 whereas the substitution of a positively charged residue as observed for other Trxs most likely changes the properties of the cavity. In classical Trxs, the corresponding cavity is solvent-exposed whereas in PtTrx*h*4, Ser54 is the most hydrophilic and polar residue embedded at the bottom of the hydrophobic site 2. All these results suggest that the “hydrophobic cavity” at the hydrophobic site 2 may be involved in regulating the flipping of the N-terminal extension of PtTrx*h*4 (since this cavity is complementarily covered by the N-terminal

extension as observed in form B structure), influencing the catalytic mechanism and/or harboring another protein.

2.5 Comparison with other Trxs

Numerous Trx structures (from different organisms) had been determined, and all display a high degree of structural homology. Among the plant *h*-type Trxs, there are 4 structures of the plant subgroups 1 and 2 available in the PDB to date (they comprise both NMR and X-ray structures), which show very little difference with other known Trx structures. Therefore, we limit the structure comparison here with *h*-type Trxs unless mentioned otherwise. Poplar Trx *h*4 is a noteworthy example of plant Trx of subgroup 3 and it is particularly interesting because of the extra N-terminal cysteine residue at position 4 of the peptide and its atypical catalytic mechanism linked to the presence of its N-terminal extension.

Structures alignment with 100-109 C^{α} showed the r.m.s.d. values ranging from 1.0 to 1.6 Å, with Trx *h* of *Chlamydomonas reinhardtii* (green alga, PDB code 1EP7), Trx *h*2 of *Hordeum vulgare* (barley, PDB code 2IWT), Trx *h*1 of *A. thaliana* (mouse-ear cress, PDB code 1XFL) and Trx *h*1 of *P. tremula* (poplar tree, PDB code 1TI3). Sequence identity among plant *h*-type Trxs mentioned above ranged between 30% and 35% (50% to 57% of sequence homology). The main difference lies on the extended N-terminal sequence of PtTrx*h*4 which is absent in *h*-type Trxs of subgroups 1 and 2. Nevertheless the canonical active site motif of Trx, WCxxC (x can be any amino acid), is in view. The loop region between β_3 and β_4 of all Trxs compared seems to be quite flexible and this region shares low sequence homologies among Trxs. Interestingly, a

motif that involves a conserved cis-proline residue, $_{79}$AMP$_{81}$ (numbering in poplar Trx *h*1), is well-conserved in all plant Trxs mentioned above and also in most other Trxs from different organisms, but is replaced by an $_{99}$ATP$_{101}$ motif in PtTrx*h*4 and its plant orthologs. The positioning of the substituted Thr100 in PtTrx*h*4 at the proximity of the disulfide bond (the $O_{\gamma 1}$ of Thr100 is less than 4 Å away from the S_{γ} of the catalytic Cys58) and the hydrophobic site 2, as observed in the present structure, certainly plays a role which needs to be identified by further mutagenesis and structural analyses. This motif could be of interest since it is located at the proximity of the disulfide bond in both forms.

Another noticeable peculiarity of the active sites of other Trxs is the absence of a buried carboxylate behind the active-site cysteines. Instead of the common aspartate residue, this corresponding position is occupied by Asn52 in PtTrx*h*4. The buried Asp26 (using the *E. coli* Trx numbering) is likely to facilitate the reaction by providing a proton carrier in the hydrophobic environment of the complex between Trx and its substrates (Jeng & Dyson 1996). In other words, this asparte may play a role in the activation of the second cysteine at the active site, either directly or indirectly. Among the Trx crystal structures containing the conserved buried Asp behind the less exposed Cys, a majority of the structures contain an ordered water molecule, hydrogen bonded to one of the oxygens on the carboxylate. The substitution of this common Asp by another residue (as in the case of PtTrx*h*4) has also been reported for *Anabaena* Trx-2 (replaced by Tyr26) (Saarinen et al. 1995) and for *Trypanosoma* Trx (replaced by Trp24) (Friemann et al. 2003). This ordered water molecule is also present in both form A and form B of PtTrx*h*4 and is hydrogen bonded to the $N_{\delta 2}$ atom of Asn52 (~ 3.2 Å).

B. DISCUSSION

1.1 The uniqueness of PtTrx*h*4

All the other Trxs characterized up to now have been shown to accept electrons from either NTR or FTR except for the poplar thioredoxin PtTrx*h*4 (Bréhélin et al. 2004). To date, PtTrx*h*4 is the only characterized Trx that is reduced by the GSH/Grx system *in vitro*. It was thus of interest to examine the reasons for the original properties of this Trx. Three conserved cysteines are present in PtTrx*h*4, two found in the classical Trx active site (WCGPC) (Cys58 and Cys61 respectively) and one localized in the first part of the protein (Cys4). Cys4, shown to play a role in the catalytic mechanism in this study, is also a well-conserved residue at the same position in all other Trxs of subgroup 3 to which PtTrx*h*4 belongs (Gelhaye et al. 2003).

1.2 PtTrx*h*4: formation of the classical C58-C61 disulphide bond

Several mutants have been generated to investigate the role of the two classical active-site cysteines (Cys58 and Cys61) in these proteins. The PtTrx*h*4C58S mutant exhibits a completely inactive enzymatic profile, suggesting that this residue acts as the catalytic cysteine (as also demonstrated for other Trxs) that starts off the substrate reduction activity. Moreover in the structure, Cys58 is located at the N-terminus of α2 helix (the same position as found in other Trxs) which is favourable to initiate a nucleophilic attack towards an oxidized substrate. The helix dipole effect has been reported in Trx to stabilize the thiolate form of the catalytic cysteine during catalysis (Miranda 2003). On the other hand, the replacement of Cys61 by serine decreases the catalytic efficiency of

the enzyme by approximately 75 %, regardless of which reduction system is used (GSH/Grx or DTT). The fact that the PtTrx*h*4C4S mutant remains fully active in presence of DTT suggests that Cys61 functions as the resolving cysteine, as demonstrated for classical Trx catalytic mechanism. Excluding the N-terminal extension, the overall structure of PtTrx*h*4 shows no significant differences from other classical Trx structures. The active-site geometry of PtTrx*h*4 in the presence of the C58-C61 disulphide bond is highly conserved as seen for other Trx known structures. Thus, in agreement with the biochemical studies (Figure 4.18), the structural information reported here strongly suggests that the first reduction step of the catalytic mechanism is very similar to other classical Trxs mechanism.

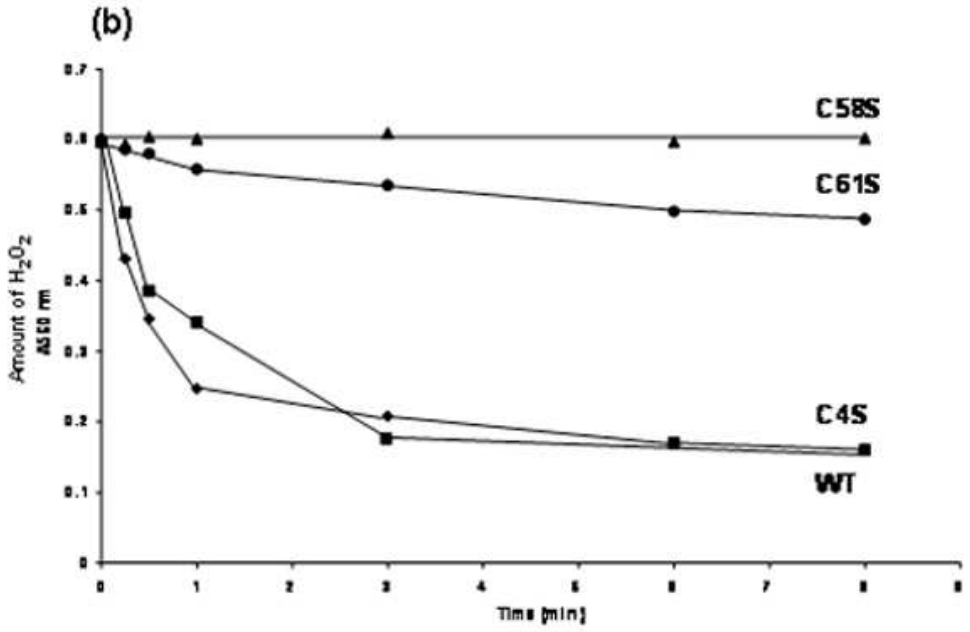


Figure 4.18 Activity of PtPrxQ in presence of DTT and PtTrx*h*4. The reduction of H_2O_2 by PtPrxQ (5 µM) was measured by following the disappearance of 500 µM H_2O_2 using 1 mM DTT in presence of either PtTrx*h*4WT or the different mutants (30 µM).

1.3 Cys4 of PtTrx*h*4 acts as the second resolving cysteine

From the activity assay (Figure 4.19), we know that PtTrx*h*4C4S is almost inactive as the PtTrx*h*4C58S when the Grx/GSH system is used as the source of electrons. This observation is consistent with the hypothesis that Cys4, present in the atypical N-terminal extension of PtTrx*h*4, is directly involved in the catalytic mechanism of this glutaredoxin-dependent type of Trx. The catalytic mechanism proposed for PtTrx*h*4 is related to the one proposed previously for methionine sulfoxide reductase A, a mechanism that involves three cysteines (each of them displaying the propensity to form disulfide) and the formation of two successive intramolecular disulfide bridges during catalysis before the regeneration of the enzyme by Trxs (Boschi-Muller et al. 2000). PtTrx*h*4 contains an N-terminal extension containing Cys4 that is not present in other Trxs *h* of subgroups 1 and 2 (Meyer et al. 2002, Gelhaye et al. 2004, Bréhélin et al. 2004). The form B clearly shows that Cys4 is able to form a disulfide with Cys58. From the same structure, we also observed that i) the N-terminal extension fits perfectly to the hydrophobic site 2 pocket by complementary hydrophobic matching, and ii) a kinked signature ($_{23}$GG$_{24}$) exists at the end of the N-terminal extension, conserved in Trxs *h*-type of subgroup 3. These observations strongly suggest that Cys4-Cys58 can be formed *in vitro* and most likely *in vivo* for PtTrx*h*4 and its homologs.

Evidences from this study suggest that Cys4 acts as the second recycling cysteine by reducing the Cys58-Cys61 disulfide bond, leading to the formation of the second intramolecular disulfide bond, i.e., Cys4-Cys58. With the presence of the N-terminal in a “closed” form (form B), Cys58 is not accessible to any substrate thus cannot be attacked by the reducing partner. Cys4 is the only solvent-accessible cysteine, suggesting that it is the one which interacts with the electron donor. In the Cys4-Cys58

conformation, Cys61 is unlikely to attack the disulfide bond since it is totally buried in the enzyme. The inefficiency of Cys61 at this step may also be related to the substitution of the conserved aspartate with Asn52 (see results).

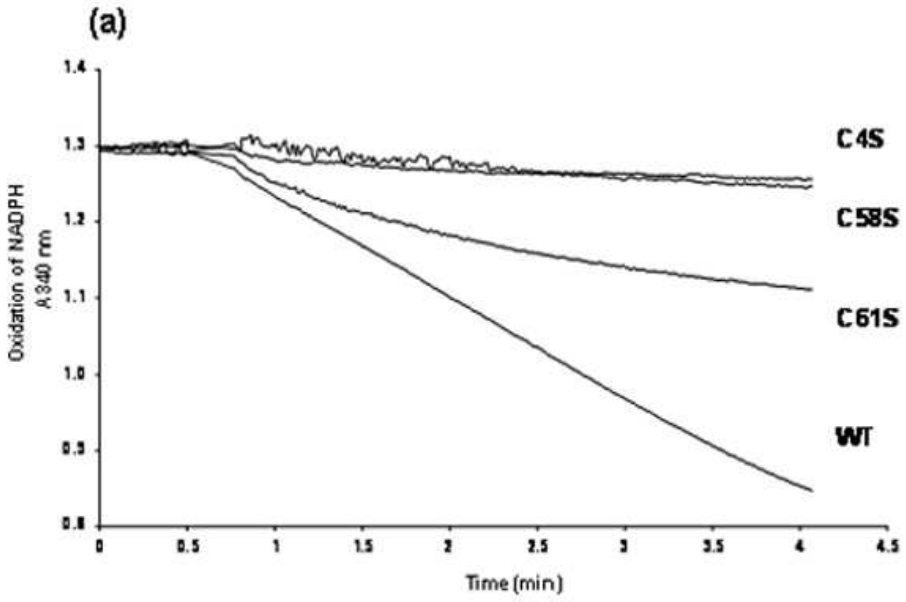


Figure 4.19 Activity of PtPrxQ in presence of PtGrx C4 and PtTrx*h*4. PtPrxQ (2 µM) was incubated with either PtTrx*h*4WT or the different mutants (16 µM) in presence of the GSH/Grx system (200 µM NADPH, 1 mM GSH, 0.5 IU glutathione reductase, 6 µM PtGrx C4, 500 µM H_2O_2). NADPH oxidation was measured at 340 nm as absorbance change per min (Koh et al. 2008).

1.4 PtTrx*h*4 fails to use the typical NTR reducing system

Instead of using the conventional NTR reducing system for its regeneration, PtTrx*h*4 uses the GSH/Grx reducing system, as demonstrated *in vitro* (Gelhaye et al. 2003). The apparent E_m value of -165 mV measured for WT PtTrx*h*4 at pH 7.0 in single-component redox titrations might represent an over-simplified picture of the situation and might in

fact reflect the presence of more than one disulfide bridge with E_m values of different components being too close to one another to be resolved as separate components (simulations suggest that the E_m values would have to differ by at least 30 mV to be resolved in these titrations). While not ignoring these possible complications, the simplest assumption to make is that the component with an E_m value of -165 mV at pH 7.0 seen in titrations of WT PtTrx*h*4 can be attributed either entirely or in large part to the active-site disulfide. This value is considerably more positive than E_m values in the -280 mV to -310 mV range observed for other Trxs from oxygenic photosynthetic organisms that have been characterized to date (Bréhélin et al. 2004, Gelhaye et al. 2004, Wangensteen et al. 2001). The other two components of the Trx/NTR system, i.e., thioredoxin reductase (TrxR) and the NADPH/NADP+ couple, both have E_m values that are very much more negative than the likely value for the active site of PtTrx*h*4-NTR has macroscopic E_m values between -270 to -300 mV (based on a *Drosophila melanogaster* system) and the NADP/NADPH has a E_m value of -327 mV (Cheng et al. 2007). In terms of redox potential, WT PtTrx*h*4 should be reducible by an NTR reducing system as demonstrated for majority of the Trxs. In fact, given the large favorable driving force for reduction of PtTrx*h*4 by NADPH that arises from the relatively positive E_m value of PtTrx*h*4, it is surprising that the PtTrx*h*4 could not be reduced by NADPH *in vitro* in the presence of NTR from either *E. coli* or *A. thaliana*. In contrast other poplar Trx *h* isoforms of subgroups 1 and 2 can be reduced by TrxR (Gelhaye et al., *unpublished results*).

Thus, there must be reasons other than thermodynamic driving force for the inability of NTR to catalyze PtTrx*h*4 reduction by NADPH. As the most likely explanation lies in the specificity of protein/protein interactions between NTR and PtTrx*h*4, it seemed particularly important to compare the structure of PtTrx*h*4 to

structures of Trxs that do serve as good electron-accepting substrates for NTR. The obvious unique feature of PtTrx*h*4 is the presence of an N-terminal extension, prompting us to investigate the possibility that this extension might hinder the interaction with NTR. The form B structure exhibits a highly negatively charged molecular surface around the active site which may unfavorable its interaction with NTR that is also possesses a negatively charged active site surface. Nevertheless, the presence of the N-terminal is not the major determinant of deterrent because our biochemical data clearly showed that the SeMet-PtTrx*h*4 (truncated form, consists residues from Asn25 to Met134) too was reduced neither by the NTR of *E. coli* nor the NTR of *A. thaliana*.

The conformation of the side chain of Trp57 may also play an important role in preventing a productive interaction between PtTrx*h*4 and TrxR. In both forms A and B, the side-chain of Trp57 at the active site (WCxxC) shows a high flexibility, so that it can either be close to the disulfide bond or at a considerable distance from the disulfide bond. A similar flexibility of the corresponding tryptophans has been reported the spinach chloroplastic Trx *f* (short form) and for the Trx of *T. brucei* (Figure 4.17). Interestingly, spinach Trx *f* uses the ferredoxin / thioredoxin reductase (FTR) system for its reduction (Crawford et al. 1989) while the Trx of *T. brucei* is spontaneously reduced by trypanothione under the parasite-specific trypanothione / trypanothione reductase system (Schmidt & Krauth-Siegel 2003). The corresponding Trp residue in classical Trxs has been proposed to possess a structural function both to force the active site in the bioactive conformation, and to mediate the protein-protein recognition (Menchise et al. 2000-2001). For WT PtTrx*h*4, we observed that the side-chain of Trp57 is moving away from the first intramolecular disulfide bond (Cys58-Cys61), which results in a very different position of this residue in PtTrx*h*4 compared to the majority of Trxs that

utilize NTR reducing system. The demonstration of this atypical positioning of the Trp side chain naturally raised the possibility that a productive interaction with NTR would be prevented. Strikingly, following the formation of the Cys4-Cys58 disulfide bond, the side chain of Trp57 returns to its "traditional" position (position superimposable to the corresponding Trp in other Trxs), making hydrogen bonding with residue Asp86. The indole ring of Trp57 at the latter position not only covers the hydrophobic pocket which is suggested to harbor the NADPH domain of TrxR (Lennon et al. 2000) but also to stabilize the disulfide bond by van der Waals interactions.

Based on the only complex structure between Trx and TrxR of *E. coli* (PDB code 1F6M), it has been proposed that the interaction of Trx with the NADPH domain (but not the FAD domain) of TrxR is responsible for recognition of the protein substrate (Lennon et al., 2000). These interactions involve a loop in Trx (residues $_{70}$YGIRGI$_{75}$ in *E. coli* Trx, corresponding to the residues $_{95}$WEIKAT$_{100}$ in PtTrx*h*4) that occupies a complementary groove on the surface of the NADPH domain of TrxR (Figure 4.20). Arg73 which bulges from the surface of Trx (Eklund et al. 1991) interacts with Arg130 and its nearby residues on one side of the groove in TrxR. Along the other side of the groove, Phe141 and Phe142 from *E. coli* TrxR (corresponding to the residues Pro155 and Ile156 of *A. thaliana* TrxR, PDB code 1VDC) fit into a "hydrophobic" pocket on Trx formed by Trp31, Ile60, Gly74, and Ile75 (Lennon et al. 2000). The fact that in PtTrx*h*4, the residue corresponding to the Arg73 of *E. coli* Trx is replaced by a lysine (Lys98) could decrease the "tightness" of this part of Trx to hook onto TrxR. Moreover, the corresponding so-called "hydrophobic" pocket in PtTrx*h*4 is assembled by Trp57 (with a flexible side-chain), Val85, Gly99, and Thr100. These latter residues are very similar to the hydrophobic pocket described for Trx by Lennon et al. (2000), except that in PtTrx*h*4 a threonine (Thr100) instead of an isoleucine (Ile75) is located in the pocket.

The presence of a more polar residue at the "hydrophobic" pocket could be the major decisive deterrent for its binding to TrxR (regardless from *E. coli* or from *A. thaliana*) *in vitro*. This residue, with the $O_{\gamma 1}$ distant to the S_{γ} of the catalytic Cys58 of less than 4 Å, is just one residue before the *cis*-Pro101 and situated at the N terminus of β_4. The positioning of this Thr100 residue at the proximity of the disulfide bond and the "hydrophobic" pocket, as observed in the present structure, certainly plays a role which needs to be identified by further mutagenesis and structural analyses. Moreover, all the plant PtTrx*h*4 homologs found to date possess this peculiarity, which is always a hydrophobic and non-polar residue in any other Trx isoforms. Thr100 of PtTrx*h*4 is substituted by either a methionine residue in majority of plant Trxs (for example, the Met82 in Trx*h*1 of *A. thaliana*, PDB code 1XFL) or retained as an isoleucine residue as found in most of the prokaryotic Trxs.

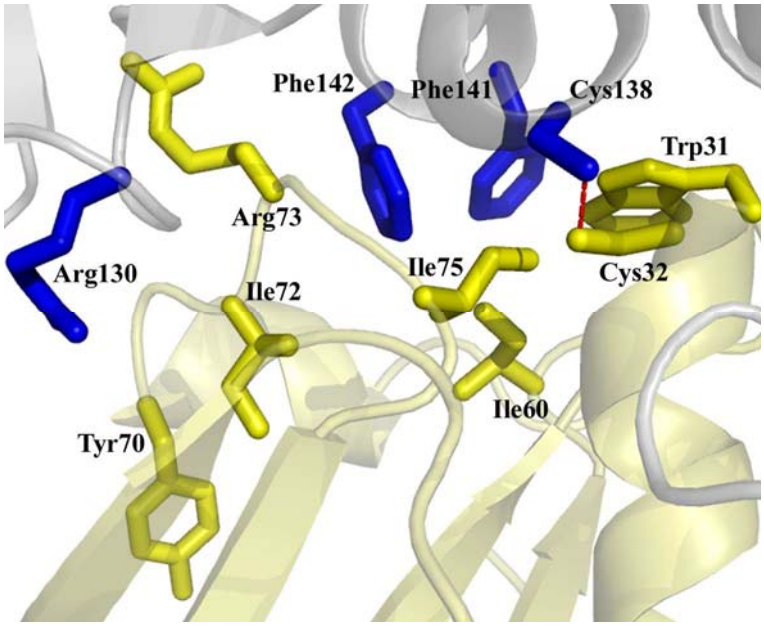


Figure 4.20 Cartoon representation of the interaction interfaces of the Trx-TrxR from *E. coli* (PDB code 1F6M). This figure highlights the complementary pocket involving most of the hydrophobic and aromatic residues (shown in sticks). The side-chain of Trx and TrxR, involving in the interaction, are colored in yellow and blue, respectively. Only relevant residues are displayed and labeled for clarity. The intermolecular disulfide bond between the two proteins is presented here in red dashes.

On the other hand, a novel monothiol *E. coli* glutaredoxin (known as Grx4) has recently been shown to be a direct substrate for NADPH and *E. coli* TrxR, but is inactive as a GSH-disulfide oxidoreductase in a standard glutaredoxin assay (Fernandes et al., 2005). Analyses of sequence alignments of the Grx4 (also sharing a Trx-fold) with the PtTrx*h*4 and Trx of *E. coli* reveal the conservation of hydrophobic residues around the active site in both *E. coli* Trx and Grx4 (especially the "hydrophobic" pocket mentioned above), but not in PtTrx*h*4, in such a way that these residues may facilitate their binding to TrxR. If residues involved in the formation of the "hydrophobic" pocket are indeed responsible for the recognition of its reducing partner, then the groove which contains residues Pro56 and Phe71 in *E. coli* Grx4 should be accommodating the two Phe residues (Phe141 and Phe142) of *E. coli* TrxR.

1.5 PtTrx*h*4 is regenerated by GSH/Grx reducing system

One point that is still unclear is whether the Cys4-Cys58 disulfide bond, once formed, is reduced by GSH or by Grx. In terms of redox potential, an E_m value of -165 mV for PtTrx*h*4 is fully compatible with its thermodynamically favorable complete reduction by either GSH (GSH/GSSG, E_m value of -241 mV) (Cheng et al. 2007) or Grxs (E_m values of -233 mV, -198 mV and -170 mV ± 10 mV for *E. coli* Grx1, *E. coli* Grx3 and plant Grx C4, respectively) (Rouhier et al. 2004, Aslund et al. 1997, Rouhier et al. 2007). From our experiments, we clearly showed that neither GSH nor Grx alone can complete the regeneration of the enzyme but both together are essential for the catalytic mechanism. The fact that we have not been able to obtain a complex between a mutated PtGrx C30S and the PtTrx*h*4C61S suggests that the possibility of direct reduction by Grx of the second intramolecular disulfide bond of PtTrx*h*4C61S (Cys4-Cys58) is

unlikely. Hence, we postulate that the reaction with a GSH molecule occurs prior to the reaction with Grx in the catalytic cycle.

Examining the accessibility of both Cys4 and Cys58 in the PtTrx*h*4C61S mutant for reduced GSH reveals that Cys4 seems to be the only plausible route to reduce the Cys4-Cys58 disulfide and return PtTrx*h*4 to its original redox state. As shown above, PtTrx*h*4C4S is almost inactive in the activity assay with the GSH/Grx reduction system, implying that Cys4 plays an important role in the interaction between the protein itself and the GSH/Grx molecules. Of considerable importance is the demonstration (using mass spectrometry to characterize tryptic digests of glutathionylated WT PtTrx*h*4) that PtTrx*h*4 can be easily glutathionylated *in vitro*. Furthermore the fact that glutathionylation occurs at Cys4 but not either Cys58 or Cys61, suggests that a late step in the catalytic cycle must involve the reduction of Cys4-glutathionylated PtTrx*h*4 by Grx. The fact that both a wild-type PtGrx and a PtGrx C30S mutant are equally efficient in the PtTrx*h*4 reduction suggested that monocysteinic Grxs are also able to reduce PtTrx*h*4.

In summary, PtTrx*h*4 is the first example of a Trx reduced by GSH and Grx. These observations shed light to the possibility that a direct interconnection / crosstalk between these two major redox regulating systems may be of physiologically importance. Further *in vivo* investigations are certainly needed in order to understand the physiological role of this unique Trx and to explore its possible role as an important link between different pools of regulatory reductants. Summarizing all the biochemical and structural evidences, we propose here (see Figure 4.21) a possible scheme for the reaction of reduced PtTrx*h*4 (a four-step mechanism) with an oxidized target protein (e.g., PrxQ). As is the case for other Trxs, catalysis proceeds through the formation of a heterodimer between the target protein and PtTrx*h*4 involving the catalytic Cys58. The

hetero-disulfide is then cleaved by Cys61, leading to the release of the reduced target protein and the formation of a disulfide bridge between Cys58 and Cys61 (this step is identical to one that has been documented for other Trxs). Then, the N-terminal Cys4 is involved in a nucleophilic attack on the Cys58-Cys61 disulfide bond, leading to the formation of a second intramolecular disulphide bond, this time between Cys4 and Cys58. The oxidized protein can then be reduced, with the Grx/GSH system, serving as the source of reductant. In the atypical Trx mechanism, the thiolate of Grx is proposed to initiate a nucleophilic attack on the Cys4-glutathionylated Trx (subgroup 3 of type *h* Trx), leading to the release of fully reduced Trx.

Orthologs of PtTrx*h*4 are also present in different plants. Based on amino acid sequence comparisons, it seems quite reasonable to assume that these proteins should, like PtTrx*h*4, be able to use Grx as an electron donor. This assumption is consistent with the report that the *Phalaris cerelescens* ortholog is not reduced by NTR (Juttner et al. 2000). It has been shown that ribonucleotide reductase cannot use monocysteinic Grx as donor where other reactions catalyzed by Grx require either monothiol or dithiol Grxs (Rouhier et al. 2004). Based on these considerations, it is tempting to speculate that both kinds of Grx may be able to promote the activity of PtTrx*h*4 and its orthologs. Interestingly, there are many naturally occurring monocysteinic Grx in plants. It will be of considerable interest to compare the relative efficiencies of the corresponding proteins for PtTrx*h*4 catalysis.

The physiological role of this particular kind of Grx-reducible Trxs remains to be elucidated (the interaction of PtTrx*h*4 with the chloroplastic PrxQ used in this study is a non-physiological reaction and similar studies need to be carried out with a physiological reaction partner). Although PtTrx*h*4 is unable to reduce several isoforms of poplar glutathione peroxidases (Navrot et al., *unpublished results*), it is able to reduce

a cytosolic methionine sulfoxide reductase A and a Prx II from poplar (Gelhaye et al. 2003). In looking for other possible targets, it will be necessary to keep in mind the fact that the E_m value of PtTrx*h*4 is approximately 130 mV more positive than that of most Trxs and thus thermodynamic considerations will limit the number of protein disulfides that PtTrx*h*4 is capable of reducing.

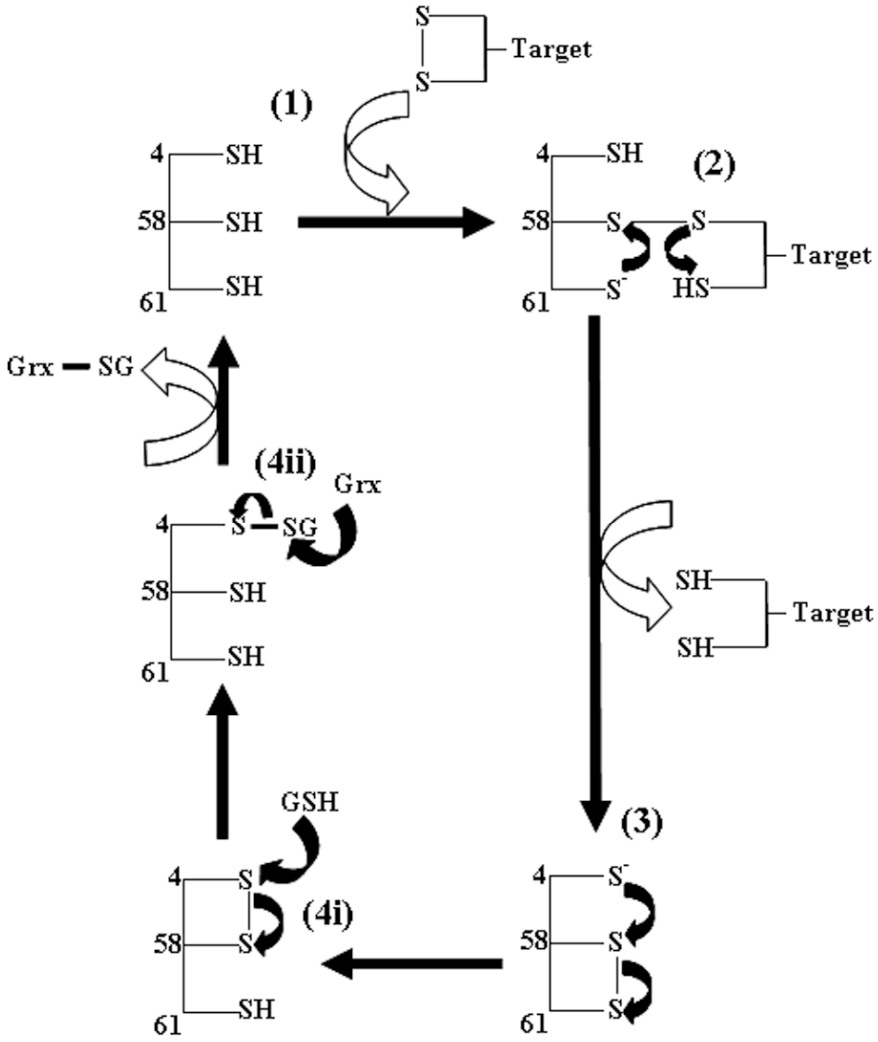


Figure 4.21 A proposed four-step disulfide cascade mechanism for Grx-dependent PtTrx*h*4 catalysis. 1) Cys58 attacks initially the Sγ atom of an oxidized target protein, 2) the heterodisulfide is then reduced by the active site recycling cysteine (Cys61), 3) the resulting disulfide bond (Cys58-Cys61) is further reduced by the second recycling cysteine, Cys4, and 4i) a glutathione molecule (GSH) reduces the Cys4-Cys58 disulfide bond prior to 4ii) the final nucleophilic attack of the Trx-SG adduct by the catalytic Cys of monothiol or dithiol Grx.

Part III: Grx

A. RESULTS

1. Specific objectives

For Grx with a “monothiol”-like active site (CxxS) like PtGrxS12, it cannot be excluded that extra active-site cysteine(s) can sometimes be involved. The deglutathionylation reaction necessitates either one or two cysteines. The well-established process (involving only one Cys) is the removal of the GSH moiety from the target protein *via* the first Grx active site cysteine (Cys29), which becomes itself glutathionylated, before another GSH molecule reduces this mixed-disulfide. An alternative (involving two Cys) has recently been proposed for a plant Grx with a monothiol CGFS active site, but with an additionally conserved C-terminal cysteine (Rouhier et al. 2008a). In this case, the C-terminal cysteine attacks the active site glutathionylated Grx, leading to the formation of an intra-molecular disulfide, which is then reduced by FTR. Therefore, it was very tempting to study the mechanism of PtGrxS12 since this isoform contains an extra active-site cysteine at its C-terminal. It was of interest to anticipate the role of this extra C-terminal cysteine (Cys87) in addition to the catalytic cysteine at the natural -$_{29}$CSYS$_{32}$- active site. Resolution of liganded PtGrxS12 structures allowed the identification of the GSH binding site in this enzyme. In comparison with other GSH binding proteins, several common residues around the active site which could represent the common GSH recognition motif in Grxs will be also highlighted.

In general, Grxs are categorized into two groups if depending just to the active site structure: a division has been established between dithiol (CPY/FC motif) and

monothiol (CGFS motif) Grxs (Fernandes & Holmgren 2004, Rodríguez-Manzaneque et al. 1999). The increasing number of Grxs with primary structures that deviates from standard dithiol and monothiol Grxs (Lemaire 2004, Deponte et al. 2005, Rouhier et al. 2006, Herrero et al. 2006) underlines the complexicity of the Grx family. This prompts a more appropriate and thorough classification of Grxs, as proposed recently for plant Grxs by Rouhier et al. (2004). In the latter classification of plant Grxs, three subclasses are identified (based on their active site structures and biochemical properties): the subclass 1 with the CxxC/S active site motif, the subclass 2 contains uniquely CGFS motif and the subclass 3 Grxs possess the atypical CCxx active sites (see Results section 1.4.3.1.2). Therefore in this study, unless specified, otherwise monothiol Grx will be refered as the Grx with CGFS active site motif.

2. Results obtained

2.1 Sequence alignment

PtGrxS12 belongs to the well-defined subclass 1 of plant Grxs, characterized by a Cxx[C/S] or more precisely [Y/W]C[G/P/S]Y[C/S] active site (Rouhier et al. 2004, 2006). Most of the Grx characterized so far in other organisms belong to this group. In the search of protein sequence homologues using protein-protein BLAST (blastp) at the NCBI website against the non-redundant protein sequence database, PtGrxS12 is found to share relatively high sequence identities with various plant Grxs, ranging from 40 % to 74 %. The BLAST search was also performed against structural databank (PDB) and a few human Grx2 structures (PDB codes 2FLS, 2HT9 and 2CQ9) were highlighted as the closest hits of PtGrxS12. These structures share 44 % of sequence identity with

PtGrxS12. To lesser extent as compared to the human Grx2, PtGrxS12 also shares structural homology with the holo form of Grx C1 (PDB code 2E7P) from the poplar same poplar hybrid origin (despite possessing higher sequence identity with PtGrxS12: ~ 50 % sequence identity). Alignments of selected structure sequences (based on the closest structure hits) and secondary structures are shown in Figure 4.22. From the sequence alignment, apart from the N-terminal region which varies there is no significant dissimilarity observed in terms of both primary and secondary structures of this enzyme and their closest homologues. Based on the alignment, the active site motif (CxYC/S, x can be any amino acids) is well-conserved in Grx across the species. However, the extra active-site cysteine (Cys87 in PtGrxS12) is conserved only in plant GrxS12 orthologs but appears diverse in other species (both plant and non-plant origin of Grxs).

Careful examination of the sequence alignments revealed an interesting discrepancy among the Grx sequences shown in Figure 4.22. For PtGrxS12 and its plant orthologs, there is a glycine residue (Gly55) located in between a 3_{10} helix and α-helix 3. This one-residue insertion is not observed in other plant non-S12-type Grx isoforms (Figure 4.22) and is totally absent in any other animal and bacterial Grxs (data not shown). The presence of this residue could be unique to this Grx isoform in plants and could have its importance in the functionality or the overall structure stabilization of this enzyme.

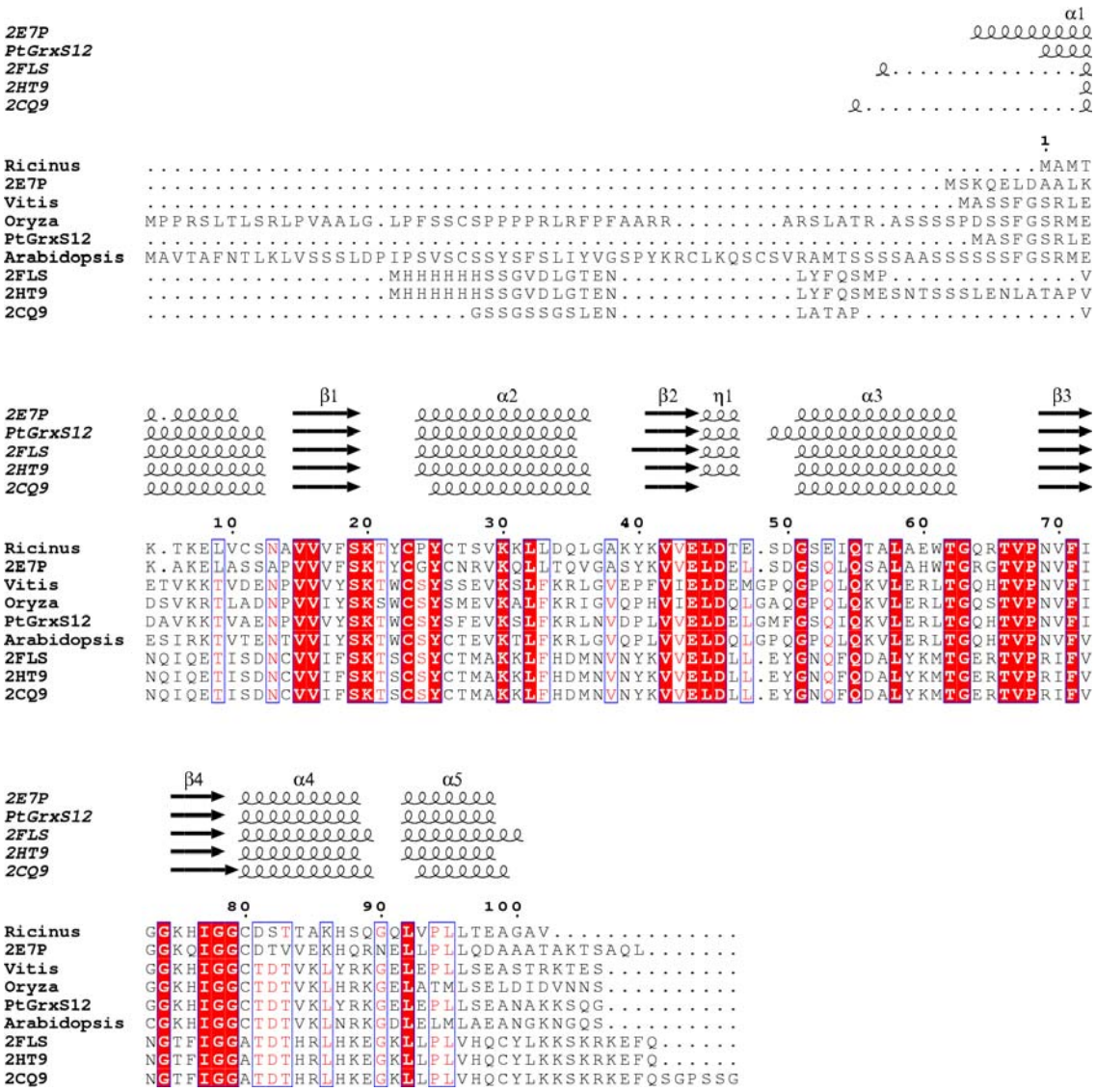


Figure 4.22 Multiple sequence alignment of 9 representative Grx proteins from mammals and plants (including the PtGrxS12) as compared to selected available secondary structure elements. References used are the holo form of Grx C1 from *Populus tremula x tremuloides* (PDB code 2E7P), the human Grx2 complexed with GSH (PDB code 2FLS), the dimeric human Grx 2 (PDB code 2HT9) and the human N-terminal domain of Grx2, Rsgi Ruh-044 (PDB code 2CQ9). The secondary structure of PtGrxS12 is shown in comparison to above-mentioned selected structures. Residues that are identical are boxed in red and displayed in white letters, while homolog residues are framed in blue and shown in red letters. For secondary structure representation, α and 3_{10} helices are displayed in coils with α and η labels, respectively, and β-strands are shown as arrows labeled β. The accession numbers for other plant Grx sequences (display highest sequences homology with PtGrxS12 based on the Blastp program) displayed in the final alignment are as follows: *Ricinus communis*, Ricinus: CAA89699; *Vitis vinifera*, Vitis: CAO40513; *Oryza sativa Japonica*, Oryza: Q0J3L4; *Arabidopsis thaliana*, Arabidopsis: NP_194602. The figure was prepared using ESPript.

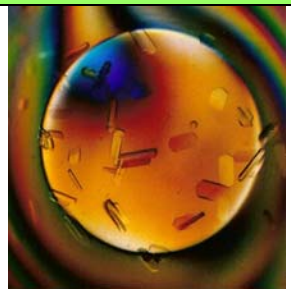
2.2 Crystal description and structure refinement statistics

Both WT-PtGrxS12$_{oxd[GSH]}$ and WT-PtGrxS12$_{oxd[GSH+HED]}$ are isomorphous and orthorhombic crystals. Based on the Matthew's coefficient calculation using the cell content analysis program in the CCP4 software package version 6.0.2, the presence of one subunit per asymmetric unit is assigned for both WT-PtGrxS12$_{oxd[GSH]}$ and WT-PtGrxS12$_{oxd[GSH+HED]}$ crystals (V_M value of 2.05 Å^3/Da, solvent content 40.15 %). Details of crystallographic data refinement statistics are summarized in Table 4.6.

2.3 Quality of models

PtGrxS12 is a protein containing 113 residues with a monomeric molecular mass of 12.49 kDa and exists as a monomer in solution. Both liganded forms of the enzyme (WT-PtGrxS12$_{oxd[GSH]}$ and WT-PtGrxS12$_{oxd[GSH+HED]}$) were crystallized as a monomer in the asymmetric unit. Final models of WT-PtGrxS12$_{oxd[GSH]}$ and WT-PtGrxS12$_{oxd[GSH+HED]}$ contain 106 amino acid residues, respectively, corresponding to residues from Gly5 to Lys110. Overall structures are well-defined in the $2F_o$-F_c electron density map. Very few residues have poorly defined side-chains and they are usually located on the protein surface. Residues with more than one conformations that could be modeled are found in both the WT-PtGrxS12$_{oxd[GSH]}$ (Asp10, Gln34, Arg41, Lys82, Lys92 and Glu105) and the WT-PtGrxS12$_{oxd[GSH+HED]}$ (Asp10, Gln34, Glu60, Arg67, Lys82, Lys92, Ser104 and Asn107) structures.

Table 4.6 Data refinement statistics for WT-PtGrxS12$_{oxd[GSH]}$ and WT-PtGrxS12$_{oxd[GSH+HED]}$ crystals.

Data Set	WT-PtGrxS12$_{oxd[GSH]}$	WT-PtGrxS12$_{oxd[GSH+HED]}$
		
Resolution range (Å)	36.00 - 1.70	30.00 - 1.80
Reflections used	11234	9301
Completeness (%)	99.26	99.61
R_{cryst} [a] (R_{free})[b]	19.28 (23.95)	16.78 (21.45)
Protein / waters / GSH / HED	106 residues / 186 / 1 / 0	106 residues / 207 / 1 / 1
Average *B*-value for peptide chain (Å^2)	17.91 (chain A)	13.37 (chain A)
Mean B factor		
Main-chain (Å^2)	16.53	12.04
Side-chain (Å^2)	23.33	14.79
Water (Å^2)	32.57	27.88
Ligand (Å^2)	15.31	17.00
All (Å^2)	20.52	16.24
Rms deviation from ideal geometry		
Bond lengths (Å)	0.012	0.011
Bond angles (°)	1.4	1.4
Dihedral angles (°)	23.5	23.7
Improper angles (°)	1.61	1.75
Ramachandran plot		
Residues in most favored regions (%)	96.7	95.6
Residues in additionally allowed regions (%)	3.3	4.4
Residues in generously allowed regions (%)	0	0

[a] $R_{cryst}=\Sigma|F_o-F_c|/\Sigma F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.
[b] The R_{free} value was calculated from 5 % of all data that were not used in the refinement.

In the model of WT-PtGrxS12$_{oxd[GSH]}$, a GSH molecule could be easily located at the active site covalently bound to Cys29 thanks to the well-defined electron density map at 1.7 Å resolution. Likewise for the WT-PtGrxS12$_{oxd[GSH+HED]}$ structure, an HED

molecule can also be modelled; covalently bound to the extra cysteine (Cys87) at the C-terminal of the glutathionylated enzyme and stabilized by residues Asp34 and Tyr31. All three residues are conserved in PtGrxS12 orthologs, suggesting their possible role for the recognition and/or fixation of ligands. The latter structure has equally well-defined electron density around the substrate/ligand binding sites (Figure 4.23).

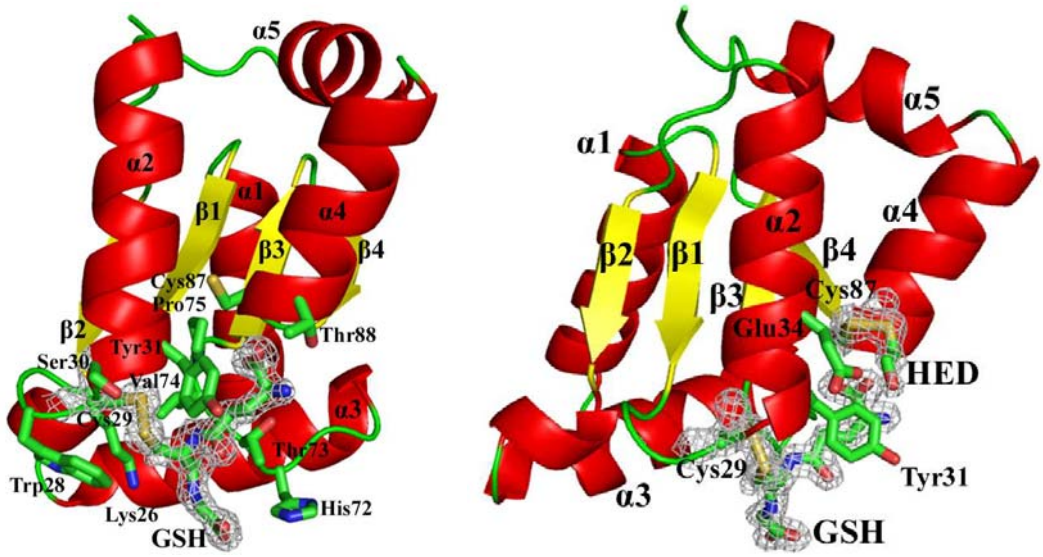


Figure 4.23 Cartoon representations of WT-PtGrxS12$_{oxd[GSH]}$ (left) and WT-PtGrxS12$_{oxd[GSH+HED]}$ (right) structures. All α-helices are shown in red while β-strands are in yellow and connecting loops are in green. Both Cys29 and Cys87, GSH and HED molecules are highlighted in sticks, with final $2F_o–F_c$ electron densities (1.2 σ level) covering chosen residues and ligands for clarity. Corresponding secondary structures are labeled.

2.4 Overall structures

The overall subunit structure of PtGrxS12 shows no significant difference to the known glutaredoxins which display a thioredoxin fold, consisting of a central four-stranded β-sheet, flanked by five α-helices. Regular secondary structure elements are conserved (Table 4.7, Figures 4.23 and 4.24). The overall structure of PtGrxS12 is very similar to

E. coli Grx2 (Xia et al. 2001) and can be described as follows: β1 strand is parallel with β2 and anti-parallel to β3 and β3 is anti-parallel to β4. The regularity in hydrogen bonding pattern is broken between β3 and β4 due to a classic *β* bulge (Nordstrand et al. 1999) at residues Val77 and $_{84}$IleGly$_{85}$ which is situated opposite to the *cis*-Pro75. The bulge accentuates the twist in the β-sheet. *β* Bulge is a common feature of a Trx-fold structure.

α-Helices 1 and 3 are located on the same side of the β-sheet and are oriented orthogonally to each other (Figures 4.23 and 4.24). α-Helices 2, 4, and 5 are located on the opposite side of the β-sheet, where α2 and α4 are essentially parallel. α-Helices 4 and 5 are almost continuous in sequence but are structurally distinct, tilted by 90°, most likely facilitated by the conserved Gly97 interspaced between the helices. In place of α4-α5 as in PtGrxS12, however, some of the oxidoreductases with Trx fold (*E. coli* Trx and T4 Grx) are lacking a C-terminal helix. In PtGrxS12, α-helix 5 is stabilized through a series of hydrophobic interactions (Leu99 and $_{102}$LeuLeu$_{103}$ of α5) with the core of the enzyme, involving residues from β1 (Val21), α2 (Leu38, Phe39, Ile42, Val44), β3 (Val77, Ile79), β4 (Ile84) and α4 (Thr90C$_{\gamma2}$, Leu93). A 3_{10} helix (residues 108 to 111) is located just before the C-terminus of the structure and it is also tilted by 90° in comparison to α5. This 3_{10} helix is connected to the C-terminus of α2 through a hydrogen bond involving Ser111N and Asn43O, which may be important in defining the orientation of α5 indirectly.

Table 4.7 Secondary structure elements in PtGrxS12.

Helices	α1	α2	α3	α4	α5
Residues	5 – 19	31 – 44	58 – 69	87 – 95	98 – 107
β-Strands	β1	β2	β3	β4	
Residues	21 – 25	47 – 51	76 – 80	82 – 85	

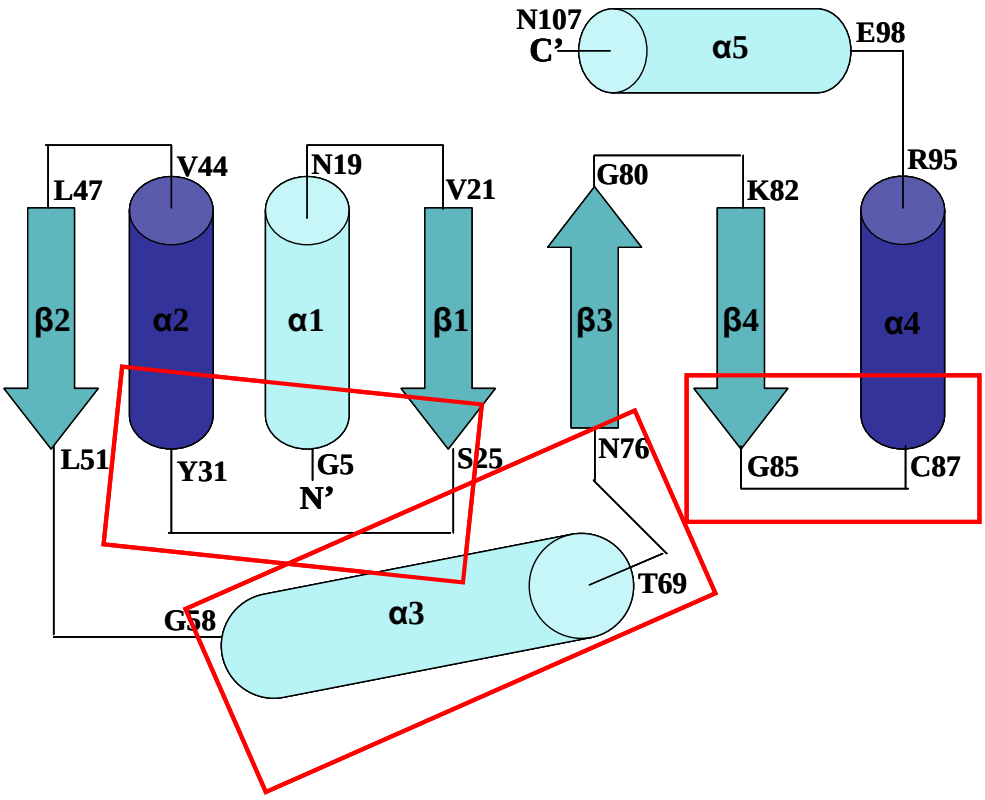


Figure 4.24 Topological diagram of PtGrxS12. In the diagram, the helices are represented as cylinders and the β-strands as arrows, and are coloured in degraded blues according to their localizations in the structure (with light blue represents the secondary structures situated at the front of the central β-sheet, medium blue represents the average plane of the central β-sheet and dark blue secondary structures situated at the back of the structure). The beginning and the ending residues with numbering of each secondary structural element are labeled. Regions involved in GSH recognition motif is boxed in red.

Superimposition between both liganded structures gives the r.m.s.d. value of 0.162 Å (based on alignments of 160 Cα positions). These low r.m.s.d. values suggest that different properties of substrates do not induce significant variations among these structures.

2.5 Active site

The active site of PtGrxS12 (Cys29–Ser32) (corresponding to the active site of *E. coli* monothiol Grx4, Cys30-Ser33; Fladvad et al. 2005) is located on the molecular surface of the enzyme, N-terminal to α-helix 2 and appears well-defined in the electron density maps of all structures solved in this study. Cys29 is exposed to the exterior of the protein (Figure 4.25) and is accessible to GSH (as seen in the liganded protein structures solved in this study) and other substrates. The conserved CxxS active-site sequence that is followed by an α helix is typically indicative of a redox function of the enzyme. The presence of one or two basic residues close to the active site is also characteristic of Grx particularly from subclass 1, and it has been suggested that it might be needed for the thioltransferase reaction due to the enhancement of the nucleophilicity of the reactive cysteine (Yang and Wells 1991). For the WT-PtGrxS12$_{oxd[GSH]}$ structure, the closest basic groups are the side-chains of Lys26 and His72, with their C^{α} atoms located approximately 6.1 and 11.1 Å away from the GSH-PtGrxS12 S-S mixed disulfide. The charged groups on their side chains lie further out than their respective C^{α} atoms in the liganded form. At the proximity of Ser32 at the active site of PtGrxS12, there is a conserved serine residue (Ser25) hydrogen-bonded to it (2.49 Å). This residue could play an important role in stabilizing the active site since this interaction is conserved in all Grxs with CxxS active site motif (see discussion).

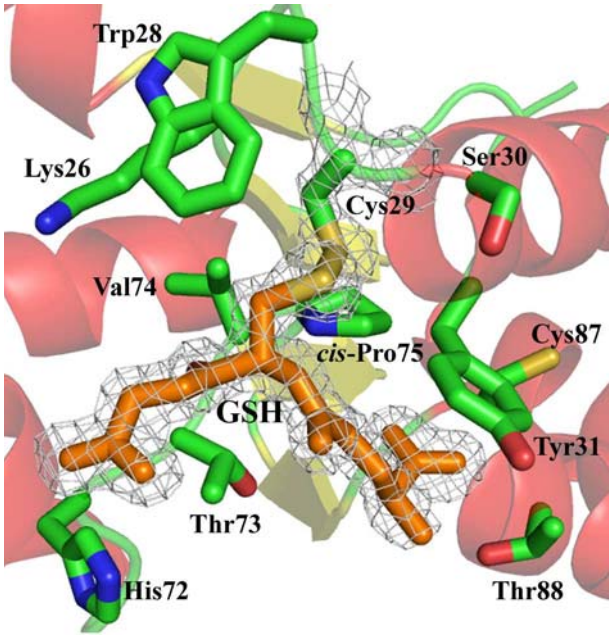


Figure 4.25 Cartoon representation of the PtGrxS12 active site. Relevant residues are labeled. All residues are shown here in sticks and colored according to their atom types. The catalytic Cys29 and the bound GSH molecule (coloured in orange) are highlighted with the display of the final $2F_o$-F_c electron density (1.2 σ level).

Parts of the active site region are rigid, as suggested by the *cis*-conformation adopted by Pro75, which is in close spatial proximity of GSH (Figure 4.25). A corresponding *cis*-Pro close to the active site is found in most Grxs/Trxs and is thought to play a role in folding and redox dynamics (Xia et al. 2001). GSH has also been reported to be stabilized by the hydrogen bonding between the substrate Cys backbone atoms and the protein backbone at the residue preceding the conserved *cis*-Pro (Bushweller et al. 1994, Nordstrand et al. 1999). In the case of PtGrxS12, a valine residue (Val74), found prior to the *cis*-Pro75, is interacting with GSH via mainchain-mainchain bonding with the cysteinyl part of GSH. Furthermore, a hydrogen bond between Cys29N and Lys26O, and a bifurcated hydrogen bonds linking Cys29O with Ser32N and Phe33N, could be important in determining the geometry of the CxxS tetrad and/or of the preceding loop (β2 – α2).

2.6 Glutathione binding site

The structure of the complex WT-PtGrxS12$_{oxd[GSH]}$ shows that GSH is recognised by PtGrxS12 primarily through residues Cys29$_{PtGrxS12}$ and Tyr31$_{PtGrxS12}$. These residues are part of the active site motif ($_{28}$WCSYS$_{32}$) of the enzyme. The whole GSH molecule is stabilized by a spatially defined motif on the PtGrxS12 surface composed of $_{28}$WCSY$_{31}$, $_{73}$TVP$_{75}$ and $_{86}$GCT$_{88}$ (referred to the GSH recognition motif in the following text) from three neighbouring loops between β1-α2, α3-β3 and β4-α4 respectively, all belonging to the core of the Trx-like fold (Figure 4.24). Similar substrate recognition loop motifs have been also described for Trx (Maeda et al. 2006) and GST (Ladner et al. 2004). In PtGrxS12, the GSH recognition motif constitutes a groove, mainly populated by polar residues, along which the backbone atoms of GSH are positioned (the part where PtGrxS12 accomodating the cysteinyl and glutamyl moieties of GSH; see Table 4.8).

The loop between β4 and α4, the backbone amino groups of Cys87 and Thr88 stabilizes the glutamyl group of GSH through hydrogen bonds. These interactions are strengthened by another hydrogen bond between the side-chain of Thr88 with the N-terminus of GSH (Table 4.8). Two immediately preceding residues of Cys87, $_{85}$GG$_{86}$ (GG kink), are strictly conserved in all PtGrxS12 and its orthologs and majority of the Grx subclass 1 members. This GG kink is at the proximity of the *cis*-Pro75$_{PtGrxS12}$. Similar to the role of the *cis*-Pro75$_{PtGrxS12}$ (see section 2.5 of this chapter), the GG kink may be important in determining the backbone geometry of the following residues which begins the α-helix 4.

On the other hand, no interaction is observed between the Gly$_{GSH}$ and residues of the PtGrxS12.

Table 4.8 Hydrogen bonding interactions at the PtGrxS12-GSH interface.

Interacted atoms		Distance (Å)	
WT-PtGrxS12	GSH	WT-PtGrxS12$_{oxd[GSH]}$	WT-PtGrxS12$_{oxd[GSH+HED]}$
	Cysteine residue		
Val74 N	O2	2.89	2.80
Val74 O	N2	2.83	2.81
	Glutamate residue		
Cys87 N	O11	2.83	2.83
Cys87 N	O12	3.21	3.21
Thr88 N	O12	2.80	2.87
Thr88 $O_{\gamma1}$	N1	3.09	3.44

2.7 Structural evaluation of conserved residues in the PtGrxS12 and orthologs

In addition to the conserved residues in all Grxs, detailed sequences comparison between PtGrxS12 and its orthologs (Figure 4.26) reveals several surface residues are nearly identical in sequence (numbering in PtGrxS12: G_5, $_{12}VK_{13}$, K_{26}, $_{28}WC_{29}$, $_{38}LF_{39}$, V_{44}, P_{46}, D_{52}, G_{55}, $_{64}VL_{55}$, G_{70}, $_{79}IG_{80}$, $_{89}DT_{90}$, L_{93}, K_{96}, L_{103}, E_{105}) throughout the PtGrxS12 orthologs. Among these residues, K_{26}, $_{28}WC_{29}$, $_{64}VL_{55}$, G_{70} and D_{89} are involved in the substrate recognition site based on the complex structure of WT-PtGrxS12$_{oxd[GSH]}$ or WT-PtGrxS12$_{oxd[GSH+HED]}$. The main cluster of strictly conserved residues (Figure 4.26) is located in direct contact with the suggested GSH recognition motif, implying a conserved substrate/ligand binding pocket in the Grxs family of subclass 1 particularly the orthologs of PtGrxS12. From the alignment, it is noteworthy to mention that residues which constitute β1, β3 and β4 are strictly conserved. There is an additional C-terminal cysteine (Cys87 in PtGrxS12) present in all the PtGrxS12 orthologs. However, this cysteine residue is replaced by an alanine in the human Grx2

(PDB codes 2FLS and 2HT9). Corresponding cysteine is also found in the *E. coli* monothiol Grx4 (Grx of subclass 2).

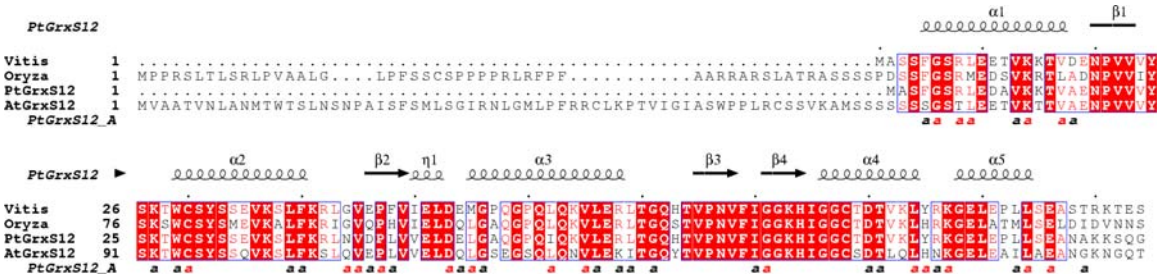


Figure 4.26 Multiple sequence alignment of PtGrxS12 and its strict homologs. The homolog representatives are plant GrxS12 from *Vitis vinifera* (CAO40513.1), *Oryza sativa* (Q0J3L4.2) and *Arabidopsis thaliana* (Q8LBS4). Identical residues are boxed in red and displayed in white letters, while homolog residues are framed in blue and shown in red letters. For secondary structure representation, α and 3_{10} helices are displayed in coils with α and η labels, respectively, and β-strands are shown as arrows labeled β. Surface residues are marked with a. The figure was prepared using ESPript.

B. DISCUSSION

1.1 PtGrxS12: an example of glutaredoxins subclass 1 of photosynthetic organisms

PtGrxS12 belongs to plant Grx of subclass 1 (Rouhier et al. 2004), with a more divergent active-site ($_{28}$WCSYS$_{32}$) and a single additional cysteine (Cys87). Crystal structure of liganded PtGrxS12 solved in this study is the first reported structure of a wild-type Grx that is covalently bound to GSH.

Despite having a "monothiol"-like active site motif (CxxS), preliminary enzymology work of this protein showed that PtGrxS12 was active in classical *in vitro* HED and dehydroascorbate reductase (DHAR) assays (N. Rouhier, *unpublished*

results), similar to other dithiol Grxs (Holmgren & Aslund 1995, Bushweller et al. 1992). Likewise, structural data from this study also reveal that PtGrxS12 possesses not only the biochemistry of Grx of subclass 1 (dithiol Grxs) but also shares minor yet could be important structural variation (see discussion below) that distinguish them from the classical monothiol Grxs with CGFS active sites. Notably, the last residue in β1, which is close to the active site, is strictly Met (for example, Met21 in the *E. coli* monothiol Grx4) in the "monothiol" Grxs, whereas smaller residues (Ser, Thr or Gly) are found in the corresponding positions of "dithiol" Grx structures (strictly conserved as a Ser in subclass 1, exemplified by Ser25 in PtGrxS12). In the Grxs of subclass 1 structures, this serine residue is hydrogen-bonded or in close proximity to the more C-terminal Cys/Ser residue of the CxxC/S active site motif. Indeed, in PtGrxS12, this conserved serine residue (Ser25) is hydrogen-bonded with Ser32 of the active site CxxS motif, *via* their hydroxyl side chains. This residue could play a role in maintaining the active site environment *in situ* and/or have a structural role in maintaining the active site architecture. Similar interaction is also present in monothiol enzymes. However, the analog of Ser25 (numbering in PtGrxS12) is commonly the serine residue preceding the catalytic cysteine at the active site (S_{29}CGFS, numbering in the *E. coli* monothiol Grx4) which is found at a different spatial location in the structure in monothiol Grx.

From the biochemistry point of view, it is noteworthy to know that dithiol Grxs use HED as their substrates *in vitro* (Holmgren & Aslund 1995, Bushweller et al. 1992). In addition, engineered dithiol Grxs in which the second cysteine of the active site has been substituted by a serine residue are also active in the HED assay (Bushweller et al. 1992) and this is also demonstrated for PtGrxS12 which have a natural CxxS active site. However, monothiol Grxs with the CGFS active site motif [for example, Grx4 of *E. coli* (Fernandes et al. 2005), Grx5 of *S. cerevisiae* (Tamarit et al. 2003) and Grx-like protein

(GLP-1) of *Plasmodium falciparum* (Rahlfs et al. 2001)] are shown not active in the HED assay. This characteristic remains true for poplar Grxs of subclass 2 (which also contain CGFS active sites) (N. Rouhier, *personal communication*). This means that monothiol Grxs with the CGFS motif are unable to deglutathionylate the small mixed disulfide between β-mercaptoethanol and a GSH moiety while the dithiol Grx (CxxS) mutants and PtGrxS12 can accomplish the task despite the fact that all Grxs share very similar tertiary structures. Distinct biochemical properties between dithiol and monothiol Grxs attributed to subtle structural differences can sometimes be difficult to underline.

1.2 Environment of the active site of PtGrxS12

Belonging to the common dithiol Grxs of subclass 1, however, PtGrxS12 is lacked of the C-terminal active site cysteine, displaying the $_{28}$WCSYS_{32} motif which is similar to a "monothiol-like" CxxS active site. From the complex structures of WT-PtGrxS12$_{oxd[GSH]}$ and WT-PtGrxS12$_{oxd[GSH+HED]}$, a substrate groove (constituting of the GSH recognition motif; see results section), rather polar, is observed at the active site accommodating a molecule of GSH.

In PtGrxS12, the central two residues in the active site are Ser30 and Tyr31, which are commonly a Pro and a Tyr in the consensus sequence among Grxs. Recent data on human mitochondrial Grx2 demonstrated that a Pro (residue immediately after the catalytic Cys) to Ser exchange in the active site dipeptide is the major determinant for the affinity toward glutathionylated substrates (Johansson et al. 2004). In this case, replacement of this Pro with Ser in the active site is most likely to affect the active site geometry, changing the pK_a of the thiol and thereby the redox properties of the protein

(Grauschopf et al. 1995, Mössner et al. 1998, Mössner et al. 2000). Indeed, human Grx1P23S and Grx2S38P mutants clearly demonstrate that a Pro to Ser exchange decreases the activity of the Grx but increases their affinity for glutathionylated substrates. Thus, PtGrxS12 with a serine residue following the catalytic Cys29 could share the same enzyme behavior as the human Grx1P23S mutant and wild type Grx2. It is interesting to note that we were not able to obtain a PtGrxS12 crystal structure that is without a GSH molecule binding to it, suggesting its high affinity towards GSH. This serine residue in the active site motif in Grx could allow the main chain to have more flexibility to accommodate the GSH binding.

In the GSH mixed disulfide observed in both WT-PtGrxS12$_{oxd[GSH]}$ and WT-PtGrxS12$_{oxd[GSH+HED]}$ structures, the side-chain of Tyr31 (located at the active site) packs near the γ-Glu residue of the GSH and at the same time in close proximity to the mixed-disulfide. Complementary with Tyr31, the side-chain of a Trp residue (Trp28) also piles against the mixed-disulfide. This enables a favourable interaction between the hydrophobic and highly polarizable sulphurs and the hydrophobic and polarizable phenol ring of Tyr and indole ring of Trp residues providing an additional determinant of peptide specificity towards GSH (particulary towards the γ-Glu$_{GSH}$ moiety) or glutathionylated-substrates. Tyr31 is highly conserved in all Grxs with several exceptions mostly of the replacement by a Phe residue (Yang et al. 1989, Ahn et al. 1992, Minakuch et al. 1994) which would also provide such a hydrophobic aromatic ring at the corresponding position in the active site. A mutagenesis study (Rouhier et al. 2001) on poplar Grx (with the $_{26}$YCPYC$_{30}$ active site) showed that replacement of Tyr29 with a Phe did not have any deleterious effect on catalysis, whereas replacement with a Pro resulted in a significant loss in reactivity, suggesting that it is the aromatic ring that is critical for the function of Grx. Tyrosine residue at this position has been

shown in *E. coli* Grx3 (Nordstand et al. 2000) to have a certain degree of flexibility from the reduced state to the GSH ligated form. If this is also the case for PtGrxS12, which means the corresponding Tyr31 is flexible, then most likely in its reduced form, the side-chain of this residue could swing to another position (more solvent accessible) in such that provides ample space for bulkier substrates and is stabilized upon GSH binding as observed in the both WT-PtGrxS12$_{oxd[GSH]}$ and WT-PtGrxS12$_{oxd[GSH+HED]}$ structures.

However, there are other factors apart from the two residues at the active site that could influence the redox potential when Grxs are considered. Apart from the influence of the dipole moment of the active site helix (corresponding to α-helix 2 of PtGrxS12) (Katti et al. 1990, Gane et al. 1995, Kortemme & Creighton 1995), the stabilization of the negative charge on the catalytic cysteine is also related to the electrostatic interactions between the thiolate anion and positively charged residues in the active site region (Lundstrom-Ljung et al. 1992, Kortemme et al. 1996). Of the two basic charged residues (Lys26 and His72) found at the proximity of the active site of PtGrxS12, Lys26 has a greater potential to swing its side chain so as to bring its positively charged ε-amino group closer to the active site without much steric hindrance. It could thereby stabilize the anionic thiolate group of Cys29 in the reduced form of the protein. Lys26, strictly conserved as a positive charge in all Grxs, may thus be critical for enhancing the reactivity of Cys29. Therefore, site-directed mutagenesis of Lys26 of PtGrxS12 could be interesting to investigate the effect of this residue towards the pK_a of Cys29.

1.3 GSH binding footprint and comparison with other Grxs

In PtGrxS12, the Grx-GSH complex is mainly stabilized by the heterodisulfide bond and five intermolecular backbone-backbone hydrogen bonds (Table 4.8 in Results chapter) involving a structural motif specific for GSH recognition, constitutes of $_{28}$WCSY$_{31}$, $_{73}$TVP$_{75}$ and $_{86}$GCT$_{88}$. This GSH recognition sequence motif is highly conserved in PtGrxS12 orthologs whereas to a lesser extend with the other Grxs. The GSH binding pocket architecture is very similar between PtGrxS12 and available Grxs with GSH bound; suggesting a common tertiary fold with a redox-active dithiol/monothiol centre is conserved throughout Grx family to enable these proteins to catalyze very efficiently deglutathionylation, reactions for which Trxs are poor catalysts (Rouhier et al. 2008a).

To date, very few biochemical and structural information are available for plant Grxs. Majority of the Grx structures solved to date are dithiol Grxs from various species, ranging from human to animal to bacteria and to T_4 phage. Only one natural "monothiol" Grx solution structure from *E. coli* (PDB code 1YKA) is available in the PDB while others are either "dithiol" Grxs. Of these accessible known Grx structures in the PDB, only "dithiol" Grxs with a GSH present can be retrieved (Figure 4.27): in three cases, the more C-terminal Cys of the active site motifs are mutated and the GSH is covalently bound [human Grx1 (1B4Q) (Yang et al. 1998), *E. coli* Grx1 (1GRX) (Sodano et al. 1991) and *E. coli* Grx3 (3GRX) (Aslund et al. 1996)] ; in one case, the GSH is not covalently bound to the wild type enzyme [human Grx2 (2FLS) (Johansson et al. 2007)] (Figure 4.27). Except the crystal structure of the human Grx2, the rest are NMR structures. The human GSH-liganded Grx2 shares the highest structural similarities with the PtGrxS12. The GSH recognition motif of PtGrxS12 ($_{28}$WCSY$_{31}$,

$_{73}$TVP$_{75}$ and $_{86}$GCT$_{88}$) is well-aligned with the motif comprised of residues $_{36}$SCSY$_{39}$, $_{80}$TVP$_{82}$ and $_{93}$GAT$_{95}$ from the crystal structure of human Grx2. The same manner of stabilizing the Cys$_{GSH}$ residue is also observed in a closely related GSH interacting oxidoreductase, GST (Ladner et al. 2004).

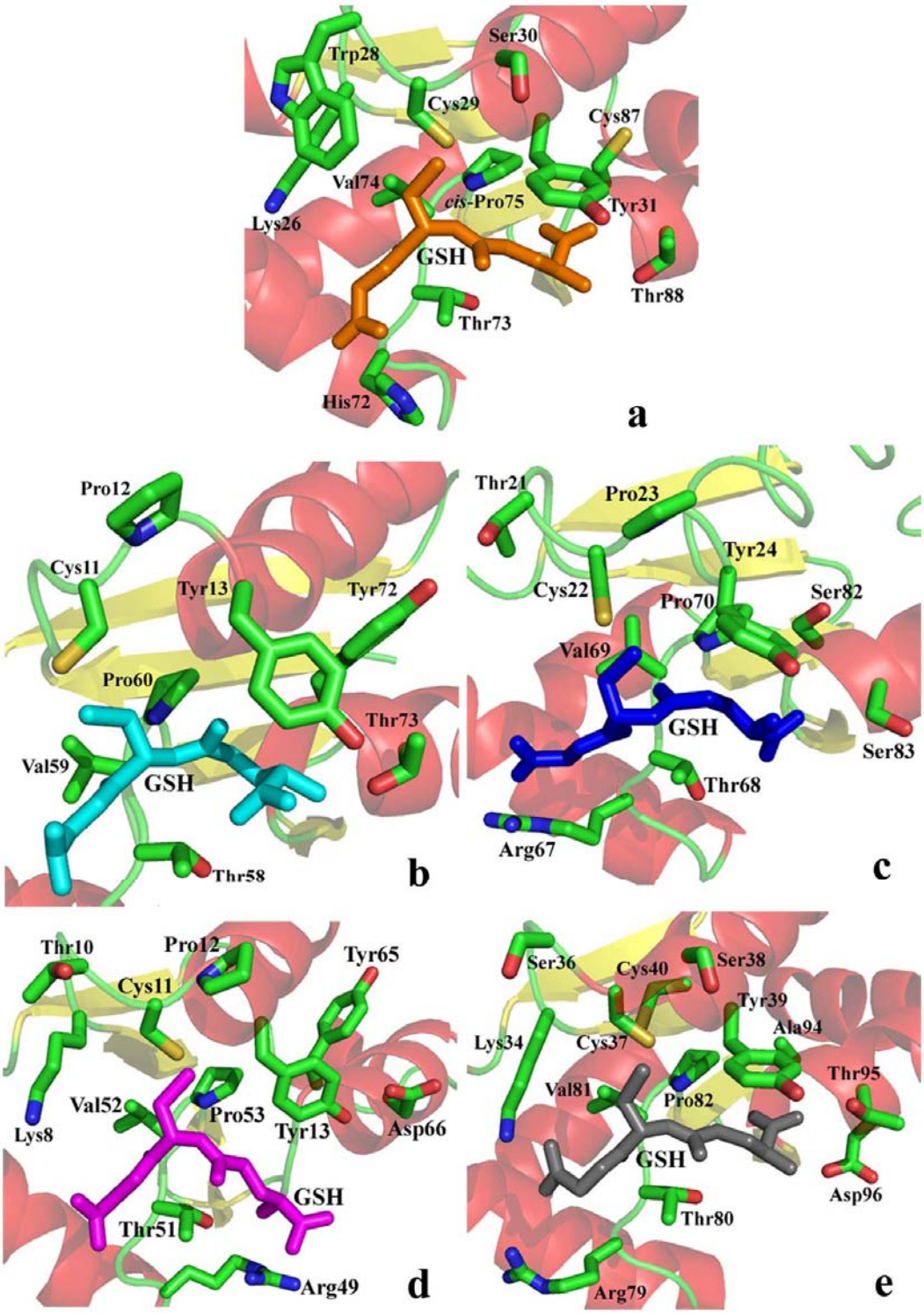


Figure 4.27 Cartoon representations of different glutathione binding sites of glutaredoxins. Active sites shown here are (a) PtGrxS12; (b) *E. coli* Grx1 (1GRX); (c) human Grx1 (1B4Q); (d) *E. coli* Grx3 (3GRX) and (e) human Grx2 (2FLS). Glutathione is highlighted and relevant residues are labeled for clarity. All residues are shown here in sticks and colored according to their atom types.

In all mentioned GSH-liganded Grx structures, GSH is in an anti-parallel orientation with respect to the main chains of the well-preserved -TVP- loop segment for mainchain-mainchain hydrogen bond formation. Favourable contact with GSH is enhanced with the presence the *cis*-Pro near the substrate binding site in which the geometry of the latter residue increases the stringency of the fold, ensures the correct conformation of the preceding valine and thus stabilizing favourable contacts with GSH.

Segment $_{86}$GCT$_{88}$ of PtGrxS12 constitutes a section of the GSH binding groove where the side-chain of Thr88 block the corresponding groove section and form hydrogen bonds with the γ-glutamate residue of GSH (Figure 4.27). This segment, with some variations in residues involved (can be Ser or Asp), is perfectly conserved in other Grx complexed with GSH structures. One can envisages that a polar residue present at the edge of a GSH binding groove that comes into contact with Glu_{GSH} is important to harbor the tripeptide molecule at the groove, making hydrogen bonds with its amide. This is indeed the case for all the liganded Grx structures mentioned above except for the human Grx1.

The position and orientation of the GSH diverge slightly at the glycine residue of GSH in all above-mentioned liganded structures. The Gly_{GSH} carboxylate is interacting with the side-chain of Arg67 in human Grx1, whereas no similar interaction is found for other Grx complex structures.

Here, I summarized those identified prerequisites of a GSH recognition site, taking into account of its homolog motifs in other Grxs. Proteins with Trx fold potentially recognize GSH moiety if they possess the following characteristics: (i) contain the CxxC/S active site motif localized at the N-terminal of α-helix 2 (in some Grxs, active site is located at the N-terminal of the first α-helix); (ii) the presence of a Tyr or or a Phe at close vicinity of the catalytic Cys; (iii) the presence of $_{73}$TVP$_{75}$ (numbering in PtGrxS12) loop motif; (iv) the presence of a *cis*-Pro (*cis*-Pro75$_{PtGrxS12}$) at the active site region; (v) the presence of a GG kink (for the β1-β2-β3-**GG**-β4 Grx scaffold) at the proximity of the active site; and (vi) the conservation of charged residues at both edges of the substrate binding groove (GSH binding pocket).

1.4 Is there a role for the additional active-site cysteine?

Most of the Trxs and Grxs have at least one additional Cys residue to those located in the active site. For example, mutagenesis of two additional cysteines (Cys78 and Cys82) in pig liver Grx (corresponding to Cys79 and Cys87 in human Grx) did not perturb the ability of the protein in reducing small-molecule GSH mixed disulfides or ascorbate to any significant extent (Yang & Wells 1991). The exact role of these additional cysteines in Grxs, if any, has not been elucidated because these mutants have not been assayed for their effect on protein disulfide reduction. However, their spatial location and local environment may provide hints as to possible functional importance.

There are more and more Grx examples containing CxxS active site and additional cysteine(s) that show the possibility of forming an intramolecular disulfide bond between the catalytic cysteine and the additional one. Indeed, yeast mitochondrial Grx5 (CGFS active site) has been reported to possess a disulfide bridge involving an

extra active site cysteine (Cys117) and the importance of the extra non-conserved Cys117 in the catalytic mechanism of the protein has been elucidated (Tamarit et al. 2003). More recently, a trypanosomal “monothiol” Grx has been reported to form an intramolecular disulfide bridge between the only active site (CAYS) cysteine and another single additional C-terminal cysteine (Cys181) (Filser et al. 2008). Besides that, additional cysteines (Cys79 and Cys83) have been also suggested to play a role in the oligomerization of Grx and in the binding of GSH to *E. coli* Grx1, respectively (Bushweller et al. 1994). Sun et al. (1998) have reported that these cysteines, distance of approximately 12 Å in the reduced protein, are able to form an intramolecular disulfide. It is suggested that formation of such a disulfide would substantially alter the structure of the protein.

Likewise, PtGrxS12 has an additional active-site cysteine (Cys87) at the N-terminal end of α4 (corresponding to the position of Cys83 in *E. coli* Grx1), which is also solvent-exposed as compared to Cys29. This cysteine is situated approximately 8.4 Å away from the catalytic Cys29, separating by the aromatic side chain of Tyr31. The corresponding conserved active site Tyr residue has been reported to undergo conformational rearrangement in *E. coli* Grx3 from the free to substrate bound form or vice versa (Nordstrand et al. 2000). In the oxidized form of *E. coli* Grx3, this tyrosine is solvent exposed while it adopts a less exposed conformation, stabilized by hydrogen bonds, in the mixed disulfide with GSH. The structures further suggest that the formation of a covalent linkage between GSH and Grx3 is necessary in order to induce these structural changes upon binding of the GSH peptide. Therefore based on this example, one can envisage that PtGrxS12 and its orthologs could undergo similar *in situ* redox-driven conformational changes.

The fact that this protein can efficiently catalyzing the reduction of small substrate like HED, comforted by the structure of WT-PtGrxS12$_{oxd[GSH+HED]}$, strongly suggests that Cys87 has a certain degree of reactivity despite the lacking of biochemical and structural data supporting the hypothesis of direct involvement of this additional cysteine in the catalytic mechanism of PtGrxS12. However, the possibility of the formation of an intramolecular disulfide between the Cys29 and Cys87 cannot be excluded since there is no experimental data available to date on the capability of these thiols in protein substrate reduction rather than the GSH-mixed disulfide reduction. From this crystallographic study, nevertheless, we can conclude from the WT-PtGrxS12$_{oxd[GSH]}$ structure that Cys87 is not able to reduce the Cys29-SG$_{GSH}$ mixed disulfide spontaneously since no fully oxidized (structure with intramolecular disulfide Cys29-Cys87) can be obtained from this structural study. Since Cys87 is fairly reactive therefore one can envisages that PtGrxS12 may have more functions apart from its regular GSH-specific oxidoreductase activity. In a promising manner, several preliminary tests (protein redox titrations) performed by the collaborator's group have shown that an intramolecular disulfide can be formed between Cys29 and Cys87 of WT-PtGrxS12, with a redox potential of approximately -305 mV at pH 7.9 (N. Rouhier, *personal communications*).

Conclusion and Perspectives

CHAPTER 5

CONCLUSION & PERSPECTIVES

Trx superfamily represents a wide range of enzymes that generally share limited sequence homology and yet they are structurally very similar, all having in common the catalytic residue Cys/SeCys and performing series of thiol/disulfide exchange. This manuscript is focusing on three members of the Trx superfamily of proteins known to be crucial for (i) scavenging ROS during oxidative stresses, GPX (PtGpx5); and (ii) maintaining a reduced intracellular redox state, Trx (PtTrx*h*4) and Grx (PtGrxS12).

5.1 Gpx

Oxidative stress in plants causes the induction of several enzymes. Contrast to well-documented animal Gpxs which are among the key enzymes responsible for scavenging vast of oxyradicals during oxidative stress, the existence of this enzyme in plants is reported only recently. Despite sharing high sequence homology with animal Gpxs, PtGpx5 (~19 kDa) and its orthologs share no similar biochemical behaviors as their animal counterparts. For the peroxide reduction, two cysteine residues are suggested to take part in the catalysis in PtGpx5 as compared to only a SeCys involved in animal Gpxs. Moreover, PtGpx5 depends on Trx as an electron donor *in vitro* instead of using GSH as their animal counterparts. No structural data on plant GPX was available prior to this current report. Here, I described two crystal structures of PtGpx5 in two different redox states that give insights for the denomination of this enzyme from a structural point of view. This structural study allowed us to identify the recycling cysteine

(Cys92) without ambiguity since there are three cysteines present in this enzyme. Taking into account previous studies (Navrot et al. 2006, Maiorino et al. 2007) and the present crystallographic evidence, we conclude that PtGpx5 is actually a thioredoxin peroxidase, structurally related to Gpxs but exhibits a catalytic and Trx-dependent recycling mechanism of Prx.

This study describes, for the first time, redox-driven drastic conformational changes in plant Gpx, much more remarkable than those described for atypical 2-Cys Prxs and prokaryotic thiol peroxidases so far. These structural rearrangements observed in PtGpx5, which involve the total unwinding of α-helix 2 and partial unwinding of one helix-turn of α-helix 1, are necessary for intramolecular disulfide bond formation and facilitate the recognition of its reducing partner, Trx. Three structural features, potential driving force responsible for causing helical instability thus leading to unwinding of helices, have been proposed in this study for these structural changes: a cluster of three same-charged residues distributed on the same side of an helix and spread over 3 consecutive helix-turns. The α-helix 2 of rPtGpx5 fulfills these prerequisites, comprising a negatively-charged cluster with residues Asp85, Asp89 and Cys92 (once the latter residue is in the thiolate form during catalysis) that is likely to facilitate the unwinding of the α-helix 2. The production of mutants lacking one of the two essential aspartate residues will definitely provides insights into the driving force of the unwinding of helices. In accordance with animal Gpxs, mutagenesis of the latter aspartate to a conserved lysine residue (D89K) is proposed. Mutant of this residue, if stable *in vitro*, can be subjected to structural studies if necessary. To understand how these helices could fold and unwind easily *in vivo*, it will be interesting to look at the process intermediates *in silico* (computational simulations to generate the reduced form enzyme from the oxidized form of X-ray crystal structure). The idea of this computer

simulation project is to qualitatively explain the likelihood of a loop to a well-defined secondary structure (in this case an α-helix) transition. Preliminary molecular simulation data show promising results and analysis of these computational models are still in progress.

As structurally demonstrated in this study, PtGpx5 is a non-covalent dimeric enzyme. This oligomerization state is described for the first time in this family of enzymes, as all classical mammalian Gpxs characterized to date are tetrameric enzymes except for PHGpx (mammalian Gpx-4) which is monomeric. The dimerization interface of PtGpx5 is formed by a pseudo anti-parallel β-sheet which is stabilized mainly by orthologs-conserved hydrophobic interactions. Structures solved in this study reveal that the PtGpx5 dimer cannot be formed by other mammal Gpxs due to the structural differences that would cause steric hindrances for the same or similar dimer formation. Among the residues involved at the dimer interface of PtGpx5, Tyr151 is potentially a "hot-spot" residue that could strongly influent the dimerization property of this enzyme. The role of this residue is worth to be investigating further through mutagenesis, biochemical and structural experiments, in comparison to the properties of the wild-type enzyme. Other important interactions at the dimerization interface are the aromatic rings stacking of two tryptophan residues (Trp124 and Trp133) of two subunits. Mutant of Trp124 is probably sufficient to impair the dimerization of this enzyme since Trp133 is conserved in all Gpxs. PtGpx5 has been observed to accept larger and bulkier substrates (e.g. phospholipid peroxides) than peroxides *in vitro* (N. Navrot, *personal communications*). The presence of the two tryptophan residues at the edge of the dimerization interface and also at the proximity of the active site may serve as a substrate clamp that reinforces its affinity towards larger substrates. Therefore, it will certainly be interesting to further investigate these features, both biochemically and

structurally, to see the effect(s) of aromatic residues to the oligomerization state as well as to the catalytic capacity towards different kinds of substrates.

Comparison analysis of Gpxs active sites reveals that PtGpx5 shares high similarity, in terms of the active site cleft architecture, with those described for mammal SeCys-Gpx structures. However, the substitution of a glutamine residue at the catalytic triad of animal Gpx with a glutamate in plant PtGpx5 orthologs may be related to the catalytic properties of Gpxs with and without a SeCys at the active site. Thus, studying the E79Q mutant of PtGpx5 could provide some clues to the importance of this residue in the catalysis of Gpxs since this residue participates in the active site triad of Gpxs. It may act as a general acid or base, or have a role as a component of a salt-link, essential for the architecture of the active site. If this is the case, mutagenesis of glutamic acid to glutamine will probably abolish or strongly decrease the peroxidase activity of the protein.

Molecular dynamics simulations performed in this study also allow the exploration, for the first time, of the Gpx–Trx interface based on the model of complex oPtGpx5-Trx *h*1 and at the same time reveal the possibility of a transient intermolecular disulfide bond formation. A Trx recognition motif conserved in majority of plant Gpxs and identified Gpx recognition motifs in Trx were identified in this study. These computer-simulated representational models do provide insights into the protein-substrate interaction interfaces, notably pinpointing that the molecular surface of Trx interacting with its substrates, for example plant Gpxs, is different from the site of Trx interacting with its regenerating partner, NTR. To date, there is no complex structure between Gpx and Trx is available thus leaving a lot of ambiguities in the understanding of protein-protein recognition mechanisms. Therefore, it would be very interesting if described hypotheses in this study can be verified experimentally by obtaining the

heterodimer structure of Gpx-Trx (either by NMR and X-ray crystallography approach) in order to shed lights into the more "physiological-relevant" protein-substrate interactions (even though it is still an *in vitro* study) and at the same time verify certain computer simulation ambiguities.

Lastly, the fact that PtGpx5 could bind a considerable number of cadmium ions prompted us to investigate the possibility of this enzyme to act as a cadmium-sink, as also suggested for Trx. From all the available experimental and structural evidences, PtGpx5 and its other plant orthologs could likely play a broader protective role in the cellular compartments as a heavy-metal sink or as a biomarker of metal contamination/accumulation in plants than just as a commonly known peroxide scavenger. Nevertheless, this still preliminary proposition for this enzyme needs further investigations, e.g. testing its feasibility against various heavy metals in a more systematic field (with poplar trees and other available plant models) and laboratory (with recombinant proteins) studies, before proceeding to its possible biotechnology applications development.

5.2 Trx

Thioredoxins are a class of small 12 kDa redox proteins, capable of reducing protein disulfide bonds, known to be present in all eukaryotic and prokaryotic organisms. Among them, PtTrx*h*4 (~15 kDa) is the only characterized Trx to date that uses the GSH/Grx system *in vitro*. Belonging to the *h*-type Trx of subgroup 3, this enzyme possesses the usual Trx active site motif ($_{57}$WCGPC$_{61}$) and an additional cysteine (Cys4) at the N-terminal extension, with the latter being identified to also play a role in the catalytic mechanism very recently. Cys4 is strictly conserved at the same position in

the orthologs of the same subgroup of enzymes. In this study, I have solved two structures of the wild-type PtTrx*h*4 and its C61S mutant which provided structural clues on the missing part of the catalytic mechanism puzzle of this atypical type of Trx.

Structural data demonstrated in this study strongly suggest that the first reduction step of the catalytic mechanism shows no variation to other classical Trxs mechanism, as Cys58 is the catalytic cysteine and Cys61 is the first back-up cysteine. Analysis of the crystal structure of PtTrx*h*4 C61S shows that Cys4-Cys58 can be formed *in vitro* and therefore most likely *in vivo* for PtTrx*h*4 and its orthologs. Cys4 most likely acts as the second recycling cysteine by reducing the Cys58-Cys61 disulfide bond, leading to the formation of the second intramolecular disulfide bond. In this latter conformation, Cys61 is unlikely to attack the disulfide bond Cys4-Cys58 and the inefficiency of Cys61 at this step could due to the presence of Asn52 in this isoform instead of a conserved aspartate. Thus, future attempt to mutate this residue to an aspartate (N52D), mimicking the nature of common Trx, may justify this argument. It will be also interesting to check on the biochemical property (e.g. redox potential and catalytic activity) of the mutant. If the nature of this residue at the corresponding position is vital for the structural integrity of the active site, one can expect the site-directed mutagenesis of asparagine to aspartic acid could influent its nearby residues (Ser54 and Cys61).

From the biochemical data (Koh et al. 2008), it is quite convincing that the presence of the N-terminal extension (residues 1 - 24) is not the major deterrent of its interaction with typical NTR because the [SeMet]PtTrx*h*4 (truncated form, consists residues from Asn25 to Met134) was not reduced by the NTR of *E. coli* or the NTR of *A. thaliana*. Thus, structural reasons are proposed in this study to explain the incapability of PtTrx*h*4 to utilize NTR adequately. Among them, the conformation of

the side-chain of Trp57 potentially plays an important role in preventing a productive interaction between PtTrx*h*4 and NTR. The side-chain shows a high flexibility, depending probably on the redox state of the enzyme. An interesting motif that involves a conserved cis-proline residue, $_{79}$AMP$_{81}$ (numbering in poplar Trx *h*1), well-conserved in all plant Trxs and in most other Trxs from different organisms, is replaced by an $_{99}$ATP$_{101}$ motif in PtTrx*h*4 and its plant orthologs. The positioning of the substituted Thr100 in PtTrx*h*4 at the proximity of the disulfide bond in both structures certainly plays a role which needs to be identified by further mutagenesis and structural analyses. Mutagenesis of residue Thr100 to either Ile or Met (T100I and/or T100M) could be promising to see the effects of these residues on GSH binding capacity and the influences to the interaction with NTR (for the latter, mutagenesis should be done to the [SeMet]PtTrx*h*4 clone which is lacking of the N-terminal extension). However, one has to bear in mind that the NTR used in the biochemical experiments are from organisms other than poplar. Therefore, it is worth to conduct activity test in the near future utilizing the NTR of the same source as PtTrx*h*4. This suggestion is based on the observation that the suggested Trx recognition motif of NTR varies slightly in the corresponding motif in poplar NTR.

Summarizing all the biochemical (Koh et al. 2008) and structural (from this study) evidences, a four-step mechanism of reduced PtTrx*h*4 with an oxidized substrate is proposed. As for other Trxs, catalysis begins with the nucleophilic attack of the catalytic Cys58 to the oxidized target protein, proceeds through the formation of a heterodimer. The hetero-disulfide is then cleaved by Cys61, leading to the release of the reduced target protein and the formation of the first intramolecular disulfide bridge between Cys58 and Cys61. Then, the N-terminal Cys4 is involved in a nucleophilic attack on the Cys58-Cys61 disulfide bond, leading to the formation of a second

intramolecular disulphide bond (Cys4-Cys58). The oxidized protein can then be reduced with the Grx/GSH system. In this atypical Trx mechanism, the thiolate of Grx is proposed to initiate a nucleophilic attack on the Cys4-glutathionylated Trx (subgroup 3 of *h*-type Trx), leading to the release of fully reduced Trx.

In order to precisely characterize the protein-substrate interactions and biological functions of this unique isoform of Trx in plants, there is an emergence need to identify the putative targets *in vivo* of this enzyme by proteomics. There are three efficient methods used currently to fish out putative protein targets, which these techniques can be applied to PtTrx*h*4 and PtGrxS12 in this study: (i) the use of Cys-mutated Trx should allow the stabilization of the Trx-target complexes (structural studies of these complexes are of high interest in the field); (ii) the use of affinity columns with mutated monocysteinic Trxs as bait to isolate protein targets from whole plants or several subcellular fractions; and (iii) the use of a combination of several techniques (e.g. alkylation of free thiols, fluorescent or radioactive labelling of free thiols, 2D electrophoresis and mass spectroscopy) to identify Trx targets from plant extracts.

5.3 Grx

Similar to Trxs, Grxs are small proteins (~9 – 15 kDa) that exist abundantly in GSH-containing forms *in vivo*. These thiol-disulfide oxidoreductases possess high affinity and specificity towards the GSH moiety, a characteristic which is unique to Grxs in comparison to Trxs. Based on their active-site motifs and biochemical properties, plant Grxs can be divided in three major subclasses: subclass 1 with active site CxxC/S motif which is active in *in vitro* activity assays using HED or DHAR; subclass 2 with the specific active site CGFS motif and Grxs from this subclass is commonly known as

monothiol Grxs (none of them are active in the HED assay); and subclass 3 with a peculiar active site CCxx motif. PtGrxS12 belongs to the subclass 1 and contains a -WCSYS- active site motif, uncommon to typical CxxC active site motif. Nevertheless, this enzyme is active in the HED activity assay *in vitro*.

PtGrxS12 structure described in this study is the first Grx structure of subclass 1 solved from plant and this structure is also the first crystallographic structure of a GSH-protein mixed disulfide with an unusual but natural CxxS active site motif. Topology of PtGrxS12 shows no considerable structural differences with other Trx-fold containing thiol-disulfide oxidoreductases. This study enables the identification of a GSH recognition motif in PtGrxS12, very similar to that described before. The recognition motif ($_{28}$WCSY$_{31}$, $_{73}$TVP$_{75}$ and $_{86}$GCT$_{88}$) is highly conserved in PtGrxS12 orthologs. From this crystallographic study, the role of Cys87, if any, remains obscure and more biochemical and structural data are needed to tackle this ambiguity. Nevertheless, from the WT-PtGrxS12$_{oxd[GSH]}$ and WT-PtGrxS12$_{oxd[GSH+HED]}$ structures, Cys87 is spatially located at the vicinity of the catalytic Cys29, separated only by the active-site tyrosine residue (Tyr31).

Since from the literatures that Grx does undergo some conformational changes upon redox activation and the corresponding Tyr residue is shown to have certain degree of flexibility, hence it seems exigent to have a completely reduced form enzyme's structure (without the GSH or HED binding to the cysteines) and a totally oxidized PtGrxS12 structure (in order to see if the possibility of the formation of the intramolecular disulfide Cys29-Cys87). It would be equally interesting to assay the activity of the reduced PtGrxS12 with substrates like 5,5'Dithiobis-2-nitrobenzoic acid (DTNB), PrxIID and PtTrx*h*4 C61S mutant to see the effect of PtGrxS12 on protein disulfide reduction. If Cys87 does play a role as a recycling cysteine only under certain

circumstances for example substrate dependent (non-glutathionylated substrate) and oxidizing state, one should expect the formation of the intramolecular disulfide bond between Cys29-Cys87 if no reductant (e.g. GSH) is added to complete the enzyme regeneration *in vitro*.

References

REFERENCES

$\mathcal{A}$

Abrahams, J. P. & Ban, N. (2003). X-ray crystallographic structure determination of large asymmetric macromolecular assemblies. *Methods Enzymol*. **374**: 163-188.

Abrahams, J. P., Leslie, A. G., Lutter, R. & Walker, J. E. (1994). Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature*. **370**: 621-628.

Acharya, K. R. & Lloyd, M. D. (2005). The advantages and limitations of protein crystal structures. *Trends Pharm. Sci*. **26**: 10-14.

Adachi, M., Takenaka, Y., Gidamis, A. B., Mikami, B. & Utsumi, S. (2001). Crystal structure of soybean proglycinin A1aB1b homotrimer. *J. Mol. Biol.* **305**: 291-305.

Adachi, T., Pimental, D. R., Heibeck, T., Hou, X., Lee, Y. J., Jiang, B., Ido, Y. & Cohen, R. A. (2004). S-glutathionylation of Ras mediates redox-sensitive signaling by angiotensin II in vascular smooth muscle cells. *J. Biol. Chem.* **279**: 29857-29862.

Adams, P. D. (2006). PHENIX. http://phenix-online.org/.

Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1999). Extending the limits of molecular replacement through combined simulated annealing and maximum-likelihood refinement. *Acta Cryst.* ***D*****55**: 181-190.

Alexander P. A., He, Y., Chen, Y., Orban, J. & Bryan P. N. (2007). The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proc. Natl. Acad. Sci. USA*. **104**: 11963-11968.

Aloy, P., Ceulemans, H., Stark, A. & Russell, R. B. (2003). The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol*. **332**: 989-998.

Alphey, M. S., Bond, C. S., Tetaud, E., Fairlamb, A. H. & Hunter, W. N. (2000). The structure of reduced tryparedoxin peroxidase reveals a decamer and insight into reactivity of 2Cys-peroxiredoxins. *J. Mol. Biol.* **300**: 903–916.

Amunts, A., Drory, O. & Nelson, N. (2007). The structure of a plant photosystem I supercomplex at 3.4 Å resolution. *Nature*. **447**: 58-63.

Apic, G., Gough, J. & Teichmann, S. A. (2001). Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol*. **310**: 311-325.

Arnér E. S. & Holmgren, A. (2000). Physiological functions of thioredoxin and thioredoxin reductase. *Eur. J. Biochem*. **267**: 6102-6109.

Aslund, F., Nordstrand, K., Berndt, K. D., Nikkola, M., Bergman, T., Ponstingl, H., Jornvall, H., Otting, G. & Holmgren, A. (1996). Glutaredoxin-3 from *Escherichia coli*. Amino acid sequence, 1H and 15N NMR assignments, and structural analysis. *J. Biol. Chem*. **271**: 6736-6745.

Ataka, M. (1993). Protein crystal growth: an approach based on phase diagram determination. *Phase transitions*. **45**: 205-219.

B

Baker, M. L., Jiang, W., Bowmann, B. R., Zhou, Z. H., Quiocho, F. A., Rixon, F. J. & Chiu, W. (2003). Architecture of the herpes simplex virus major capsid protein derived from structural bioinformatics. *J. Mol. Biol*. **331**: 447-456.

Balatri, E., Banci, L., Bertini, I., Cantini, F. & Ciofi-Baffoni, S. (2003). Solution structure of Sco1: a thioredoxin-like protein involved in cytochrome c oxidase assembly. *Structure*. **11**: 1431-1443.

Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*. 289: 905-920.

Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P. B. & Steitz, T. A. (1999). Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature*. **400**: 841-847.

Bao, R., Chen, Y., Tang, Y-J., Janin, J. & Zhou, C-Z. (2007). Crystal structure of the yeast cytoplasmic thioredoxin Trx2. *Proteins*. **66**: 246-249.

Baumeister, W. (2002). Electron tomography: towards visualizing the molecular organization of the cytoplasm. *Curr. Opin. Struct. Biol*. **12**: 679-684.

Beckmann, R., Spahn, C. M., Eswar, N., Helmers, J., Penczek, P. A., Sali, A., Frank, J. & Blobel, G. (2001). Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell*. **107**: 361-372.

Ben-Shem, A., Frolow, F. & Nelson, N. (2003). Crystal structure of plant photosystem I. *Nature*. **426**: 630-635.

Berglund, G. I., Carlsson, G. H., Smith, A. T., Szoke, H., Henriksen, A. & Hajdu, J. (2002). The catalytic pathway of horseradish peroxidase at high resolution. *Nature*. **417**: 463-468.

Berk, C., Zhang, W., Pai, R. D. & Cate, J. H. D. (2006). Structural basis for mRNA and tRNA positioning on the ribosome. *Proc. Natl. Acad. Sci. USA*. **103**: 15830-15834.

Bernal, J. D. & Crowfoot, D. (1934). X-ray photographs of crystalline pepsin. *Nature*. **133**: 794-795.

Bernard, Y., Degoy, S., Lefaucheux, F. & Robert, M. C. (1994). A gel-mediated feeding technique for protein crystal growth from hanging drops. *Acta Cryst.* ***D*****50**: 504-507.

Berndt, C., Lillig, C. H. & Holmgren, A. (2007). Thiol-based mechanisms of the thioredoxin and glutaredoxin systems: implications for diseases in the cardiovascular system. *Am. J. Physiol. Heart Circ. Physiol.* **292**: 1227-1236.

Bijvoet J. M., Peerdeman A. F. & van Bommel A. J. (1951). Determination of the absolute configuration of optically active compounds by means of X-rays. *Nature*. **168**: 271–272.

Blundell, T. L. & Johnson, L. N. (1976). Protein crystallography. ed. 1. New York: Academic Press.

Bourgeois, D. & Royant, A. (2005). Advances in kinetic protein crystallography. *Curr. Opin. Struct. Biol.* **15**: 538-547.

Borissenko, L. & Groll, M. (2007). 20S Proteasome and its inhibitors: crystallographic knowledge for drug development. *Chem. Rev*. **107**: 687-717.

Branden, C. & Tooze, J. (1999). Introduction to protein structure. ed. 2. New York: Garland Publishing, Inc.

Bréhélin, C., Laloi, C., Setterdahl, A. T., Knaff, D. B. & Meyer, Y. (2004). Cytosolic, mitochondrial thioredoxins and thioredoxin reductases in *Arabidopsis thaliana*. *Photosynth. Res*. **79**: 295-304.

Broin, M., Cuiné, S., Peltier, G. & Rey, P. (2000). Involvement of CDSP 32, a drought-induced thioredoxin, in the response to oxidative stress in potato plants. *FEBS Lett*. **467**: 245-248.

Bushweller, J. H., Aslund, F., Wuthrich, K. & Holmgren, A. (1992) Structural and functional characterization of the mutant *Escherichia coli* glutaredoxin (C14-S) and its mixed disulfide with glutathione. *Biochemistry*. **31**: 9288-9293.

Bushweller, J. H., Billeter, M., Holmgren, A. & Wüthrich, K. (1994). The nuclear magnetic resonance solution structure of the mixed disulfide between *Escherichia coli* glutaredoxin (C14S) and glutathione. *J. Mol. Biol*. **235**: 1585-1597.

C

Capitani, G., Markovic-Housley, Z., DelVal, G., Morris, M., Jansonius, J. N. & Schurmann, P. (2000). Crystal structures of two functionally different thioredoxins in spinach chloroplasts. *J. Mol. Biol*. **302**: 135-154.

Carter Jr., C. W. (1997). Response surface methods for optimizing and improving reproducibility of crystal growth. *Methods Enzymol*. **276**: 74-99.

Cate, J . H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N. & Noller, H. F. (1999). X-ray crystal structures of 70S ribosome functional complexes. *Science*. **285**: 2095-2104.

Cazalis, R., Pulido, P., Aussenac, T., Pérez-Ruiz, J. M. & Cejudo, F. J. (2006). Cloning and characterization of three thioredoxin *h* isoforms from wheat showing differential expression in seeds. *J. Exp. Bot*. **57**: 2165-2172.

Chaudhry, C., Horwich, A. L., Brunger, A. T. & Adams, P. D. (2004). Exploring the structural dynamics of the *E. coli* chaperonin GroEL using translation-libration-screw crystallographic refinement of intermediate states. *J. Mol. Biol*. **342**: 229-245.

Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). An automated system for micro-batch protein crystallization and screening. *J. Appl. Cryst.* **23**: 297-302.

Chayen, N. E., Shaw Stewart, P. D. & Blow, D. M. (1992). Microbatch crystallization under oil – a technique allowing many small-volume crystallization trials. *J. Cryst.Growth* **122**: 176-180.

Chayen, N. E. (1997). A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *J. Appl. Cryst*. 30: 198-202.

Chayen, N. E. (1998). Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta Cryst*. ***D*54**: 8-15.

Chayen, N. E. (2004). Turning protein crystallization from an art into a science. *Curr. Opin. Struct. Biol*. **14**: 577-583.

Chayen, N. E. (2005). Methods for separating nucleation and growth in protein crystallization. *Prog. Biophys. Mol. Biol*. **88**: 329-337.

Chayen, N. E. & Saridakis, E. (2002). Protein crystallization for genomics: towards high-throughput optimization techniques – lead article. *Acta Cryst.* ***D*58**: 921-927.

Chayen, N. E., Saridakis, E., El Bahar, R. & Nemirvsky, Y. (2001). Porous silicon: an effective nucleation-inducing material for protein crystallization. *J. Mol. Biol*. **312**: 591-595.

Cheng, N. H. & Hirschi, K. D. (2003). Cloning and characterization of CXIP1, a novel PICOT domain-containing Arabidopsis protein that associates with CAX1. *J. Biol. Chem*. **278**: 6503-6509.

Cheng, N. H., Liu, J. Z., Brock, A., Nelson, R. S. & Hirschi, K. D. (2006). AtGRXcp, an Arabidopsis chloroplastic glutaredoxin, is critical for protection against protein oxidative damage. *J. Biol. Chem*. **281**: 26280-26288.

Chernov, A. A., García-Ruíz, J. M. & Thomas, B. R. (2001). Visualization of the impurity depletion zone surrounding apoferritin crystals growing in gel with holoferritin dimer impurity. *J. Cryst. Growth*. **232**: 184-187.

Chinenov, Y. V. (2000). Cytochrome c oxidase assembly factors with a thioredoxin fold are conserved among prokaryotes and eukaryotes. *J. Mol. Med*. **78**: 243-244.

Chivers, P. T., Prehoda, K. E. & Raines, R. T. (1997). The CXXC motif: a rheostat in the active site. *Biochemistry*. **36**: 4061-4066.

Choi, H. J., Kang, S. W., Yang, C. H., Rhee, S. G. & Ryu, S. E. (1998). Crystal structure of a novel human peroxidase enzyme at 2.0 Å resolution. *Nat. Struct. Biol*. **5**: 400-406.

Chrestensen, C. A., Starke, D. W. & Mieyal, J. J. (2000). Acute cadmium exposure inactivates thioltransferase (Glutaredoxin), inhibits intracellular reduction of protein-glutathionyl-mixed disulfides, and initiates apoptosis. *J. Biol. Chem*. **275**: 26556-26565.

Clemons, W. M. Jr., Brodersen, D. E., McCutcheon, J. P., May, J. L., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T. & Ramakrishnan, V. (2001). Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: purification, crystallization and structure determination. *J. Mol. Biol*. **310**: 827-843.

Clemons, W. M. Jr., May, J. L., Wimberly, B. T., McCutcheon, J. P., Capel, M. S. & Ramakrishnan, V. (1999). Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature*. **400**: 833-840.

Collin, V., Lamkemeyer, P., Miginiac-Maslow, M., Hirasawa, M., Knaff, D. B., Dietz, K. J. & Issakidis-Bourguet, E. (2004). Characterization of plastidial thioredoxins from *Arabidopsis* belonging to the new y-type. *Plant Physiol*. **136**: 4088-4095.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz Jr, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc*. **117**: 5179–5197.

Coudevylle, N., Thureau, A., Hemmerlin, C., Gelhaye, E., Jacquot, J. P. & Cung, M. T. (2005). Solution structure of a natural CPPC active site variant, the reduced form of thioredoxin *h*1 from poplar. *Biochemistry*. **44**: 2001-2008.

Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R. & Kornberg, R. D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*. **288**: 640-649.

D

Dai, S., Friemann, R., Glauser, D. A., Bourquin, F., Manieri, W., Schurmann, P. & Eklund, H. (2007). Structural snapshots along the reaction pathway of ferredoxin-thioredoxin reductase. *Nature*. **448**: 92-96.

Dai, S., Saarinen, M., Ramaswamy, S., Meyer, Y., Jacquot, J. P. & Eklund, H. (1996). Crystal structure of *Arabidopsis thaliana* NADPH dependent thioredoxin reductase at 2.5 Å resolution. *J. Mol. Biol*. **264**: 1044-1057.

D'Arcy, A., Mac Sweeney, A. & Haber, A. (2003). Using natural seeding material to generate nucleation in protein crystallization experiments. *Acta Cryst.* ***D*****59**: 1343-1346.

Dat, J., Vandenabeele, S., Vranova, E., Van Montagu, M., Inzé, D. & Van Breusegem, F. (2000). Dual action of the active oxygen species during plant stress responses. *Cell Mol. Life Sci.* **57**: 779–795.

Dauter, Z. (1999). Data-collection strategies. *Acta Cryst*. **D55**: 1703-1717.

Dauter, Z. (2005). Current state and prospects of macromolecular crystallography. *Acta Cryst*. **D62**: 1-11.

Dauter, Z. (2005). Efficient use of synchrotron radiation for macromolecular diffraction data collection. *Prog. Biophys. Mol. Biol*. **89**: 153-172.

Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). The benefits of atomic resolution. *Curr. Opin. Struct. Biol*. **7**: 681-688.

Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol*. **319**: 1097-1113.

Davis, A. M., Teague, S. J. & Kleywegt, G. J. (2003). Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem. Int. Ed. Engl*. **42**: 2718-2736.

Dawson, R. J. & Locher, K. P. (2006). Structure of a bacterial multidrug ABC transporter. *Nature*. **443**: 180-185.

Declercq, J. P., Evrard, C., Clippe, A., Stricht, D. V., Bernard, A. & Knoops, B. (2001). Crystal structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 Å resolution. *J. Mol. Biol*. **311**: 751-759.

Delaunay, A., Pflieger, D., Barrault, M. B., Vinh, J. & Toledano, M. B. (2002). A thiol peroxidase is an $HB_{2B}OB_{2B}$ receptor and redox-transducer in gene activation. *Cell*. **111**: 471–481.

Deponte, M., Becker, K. & Rahlfs, S. (2005). *Plasmodium falciparum* glutaredoxin-like proteins. *Biol. Chem*. **386**: 33-40.

Dixon, D. P., Fordham-Skelton, A. P. & Edwards, R. (2005). Redox regulation of a soybean tyrosine-specific protein phosphatase. *Biochemistry*. **44**: 7696-7703.

Doyle, D. A., Cabral, J. M., Pfuetzner R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T. & Mackinnon, R. (1998). The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science*. **280**: 69-77.

Drenth, J. (2007). Principles of protein X-ray crystallography. ed. 3. USA: Springer.

Dry, S., McCarthy, S. & Harris, T. (2000). Structural genomics in the biotechnology sector. *Nat. Struc. Biol. Struct.* **Genomics Suppl**.: 946-949.

Ducruix, A. & Giegé, R. (1999). Crystallization of nucleic acids and proteins. A practical approach, ed. 2. Oxford: Oxford University Press.

Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L. & McCafferty, J. (2004). Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC Biotechnol*. **4**: 32.

E

Ealick, S. E. (1997). Now we're cooking: new successes for shake-and-bake. *Structure*. **5**: 469-472.

Ealick, S. E. (2000). Advances in multiple wavelength anomalous diffraction crystallography. *Curr. Opin. Chem. Biol*. **4**: 495-499.

Eaton, P., Byers, H. L., Leeds, N., Ward, M. A. & Shattock, M. J. (2002). Detection, quantitation, purification, and identification of cardiac proteins S-thiolated during ischemia and reperfusion. *J. Biol. Chem*. **277**: 9806-9811.

Edwards, C. & Palmer, S. B. (1990). Thermal motion in protein crystals estimated using laser-generated ultrasound and young's modulus measurements. *Acta Cryst*. ***A*46**: 315-320.

Echols, N., Milburn, D. & Gerstein, M. (2003). MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res*. **31**: 478-482.

Eklund, H., Cambillau, C., Sjöberg, B. M., Holmgren, A., Jörnvall, H., Höög, J. O. & Brändén, C. I. (1984). Conformational and functional similarities between glutaredoxin and thioredoxins. *EMBO J*. **3**: 1443-1449.

Eklund, H., Gleason, F. K. & Holmgren, A. (1991). Structural and functional relations among thioredoxins of different species. *Proteins*. **11**: 13-28.

Epp, O., Ladenstein, R. & Wendel, A. (1983). The refined structure of the selenoenzyme glutathione peroxidase at 0.2-nm resolution. *Eur. J. Biochem*. **133**: 51–69.

Eshdat, Y., Holland, D., Faltin, Z. & Ben-Hayyim, G. (1997). Plant glutathione peroxidases. *Physiol. Plantarum*. **100**: 234–240.

Esposito, L., Vitagliano, L., Sica, F., Sorrentino, G., Zagari, A. & Mazzarella, L. (2000). The ultrahigh resolution crystal structure of ribonuclease A containing an isoaspartyl residue: hydration and stereochemical analysis. *J. Mol. Biol*. **297**: 713-732.

F

Farr, S. B. & Kogoma, T. (1991). Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium. Microbiol. Rev*. **55**: 561–585.

Feng, Y., Zhong, N., Rouhier, N., Hase, T., Kusunoki, M., Jacquot, J.-P., Jin, C. & Xia, B. (2006). Structural insight into poplar glutaredoxin C1 with a bridging iron-sulfur cluster at the active site. *Biochemistry*. **45**: 7998-8008.

Fersht, A. (1985). Enzyme structure and mechanism. ed. 2. New York: Freeman & Co.

Fermani, S., Falini, G., Minnucci, M. & Ripamonti, A. (2001). Protein crystallization on polymeric film surfaces. *J. Cryst. Growth*. **224**: 327-334.

Fernandes, A. P., Fladvad, M., Berndt, C., Andrésen, C., Lillig, C. H., Neubauer, P., Sunnerhagen, M., Holmgren, A. & Vlamis-Gardikas, A. (2005). A novel monothiol glutaredoxin (Grx4) from *Escherichia coli* can serve as a substrate for thioredoxin reductase. *J. Biol. Chem*. **280**: 24544-24552.

Fernandes, A. P. & Holmgren, A. (2004). Glutaredoxins: glutathione-dependent redox enzymes with functions far beyond a simple thioredoxin backup system. *Antiox. Redox Signal*. **6**: 63-74.

Filser, M., Comini, M. A., Molina-Navarro, M. M., Dirdjaja, N., Herrero, E. & Krauth-Siegel, R. L. (2008). Cloning, functional analysis, and mitochondrial localization of *Trypanosoma brucei* monothiol glutaredoxin-1. *J. Biol. Chem*. **389**: 21-32.

Fladvad, M., Bellanda, M., Fernandes, A. P., Mammi, S., Vlamis-Gardikas, A., Holmgren, A. & Sunnerhagen, M. (2005). Molecular mapping of functionalities in the solution structure of reduced Grx4, a monothiol glutaredoxin from *Escherichia coli. J. Biol. Chem.* **280**: 24553-24561.

Flohé, L., Loschen, G., Gunzler, W. A. & Eichele, E. (1972). Glutathione peroxidase, V. The kinetic mechanism. *Hoppe Seylers Z Physiol. Chem*. **353**: 987–999.

Fomenko, D. E. & Gladyshev, V. N. (2002). CxxS: fold-independent redox motif revealed by genome-wide searches for thiol/disulfide oxidoreductase function. *Protein Sci.* **11**: 2285-2296.

Foyer, C. H., Lescure, J. C., Lefebvre, C., Morot- Gaudry, J. F., Vincentz, M. & Vaucheret, H. (1994). Adaptations of photosynthetic electron transport, carbon assimilation, and carbon partitioning in transgenic *Nicotiana plumbaginifolia* plants to changes in nitrate reductase activity. *Plant Physiol*. **104**: 171–178.

Frey, M. (1994). Water structure associated with proteins and its role in crystallization. *Acta Cryst.* ***D*****50**: 663-666.

Friemann, R., Schmidt, H., Ramaswamy, S., Forstner, M., Krauth-Siegel, R. L. & Eklund, H. (2003). Structure of thioredoxin from *Trypanosoma brucei brucei*. *FEBS Lett*. **554**: 301-305.

Fu, J., Gnatt, A. L., Bushnell, D. A., Jensen, G. J., Thompson, N. E., Burgess, R. R., David, P. R. & Kornberg, R. D. (1999). Yeast RNA polymerase II at 5 Å resolution. *Cell*. **98**: 799-810.

Fujii, J., Iuchi, Y., Matsuki, S. & Ishii, T. (2003). Cooperative function of antioxidant and redox systems against oxidative stress in male reproductive tissues. *Asian J. Androl.* **5**: 231–242.

Funato, Y. & Miki, H. (2007). Nucleoredoxin, a novel thioredoxin family member involved in cell growth and differentiation. *Antioxid. Redox Signal*. **9**: 1035-1057.

G

Gallogly, M. M. & Mieyal, J. J. (2007). Mechanisms of reversible protein glutathionylation in redox signaling and oxidative stress. *Curr. Opin. Pharmacol*. **7**: 381-391.

García-Ruíz, J. M. (1991). Uses of crystal growth in gels and other diffusing-reacting systems. *Key Eng. Mater*. **88**: 87-106.

García-Ruíz, J. M., Drenth, J., Riès-Kautt, M. & Tardieu, A. (2001a). A world without gravity: Research in space for health and industrial processes. The Netherlands: ESA publication division.

García-Ruíz, J. M., Novella, M. L., Moreno, R. & Gavira, J. (2001b). Agarose as crystallization media for proteins. I. Transport processes. *J. Cryst. Growth* **232**: 165-172.

García-Ruíz, J. M., Gonzalez-Ramirez, L. A., Gavira, J. A. & Otalora, F. (2002). Granada crystallization box: a new device for protein crystallization by counter-diffusion techniques. *Acta Cryst.* ***D*****58**: 1638-1642.

García-Ruíz, J. M. & Moreno, A. (1994). Investigation on protein crystal growth by the gel acupuncture method. *Acta Cryst.* ***D*****50**: 484-490.

García-Ruíz, J. M., Novella, M. L., Moreno, R. & Gavira, J. A. (2002). Agarose as crystallization media for proteins; I: Transport processes. *J. Crystal Growth* **232**: 165-172.

Gavira, J. A., Toh, D., Lopéz-Jaramillo, J., García-Ruíz, J. M. & Ng, J. D. (2002). *Ab initio* crystallographic structure determination of insulin from protein to electron density without crystal handling. *Acta Cryst.* ***D*****58**: 1147-1154.

Gelhaye, E., Rouhier, N., Gerard, J., Jolivet, Y., Gualberto, J., Navrot, N., Ohlsson, P. I., Wingsle, G., Hirasawa, M., Knaff, D. B., Wang, H., Dizengremel, P., Meyer, Y. & Jacquot, J-P. (2004b). A specific form of thioredoxin *h* occurs in plant mitochondria and regulates the alternative oxidase. *Proc. Natl. Acad. Sci. USA.* **101**: 14545-14550.

Gelhaye, E., Rouhier, N. & Jacquot, J-P. (2003). Evidence for a subgroup of thioredoxin *h* that requires GSH/Grx for its reduction. *FEBS Lett*. **555**: 443-448.

Gelhaye, E., Rouhier, N. & Jacquot, J-P. (2004a). The thioredoxin *h* system of higher plants. *Plant Physiol. Biochem*. **42**: 265-271.

Gelhaye, E., Rouhier, N., Navrot, N. & Jacquot, J-P. (2005). The plant thioredoxin system. *Cell. Mol. Life Sci*. **62**: 24-35.

Gilliland, G. L., Tung, M. & Ladner, J. E. (2002). The biological macromolecule crystallization database: crystallization procedures and strategies. *Acta Cryst.* ***D*58**: 916-920.

Glusker, J. P., Lewis, M. & Rossi, M. (1994). Crystal structure analysis for chemist and biologist. ed. 1. New York: VCH Publishers, Inc.

Glusker, J. P. & Trueblood, K. N. (1985). Crystal structure analysis- A primer. *ed*. 2. Oxford: Oxford University Press.

Golas, M. M., Sander, B., Will, C. L., Luhrmann, R. & Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. *Science*. **300**: 980-984.

González, A. (2003). Optimizing data collection for structure determination. *Acta Cryst*. ***D*59**: 1935-1942.

Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript : multiple sequence alignments in PostScript. *Bioinformatics*. **15**: 305-308.

Grauschopf, U., Winther, J. R., Korber, P., Zander, T., Dallinger, P. & Bardwell, J. C. (1995). Why is DsbA such an oxidizing disulfide catalyst? *Cell*. **83**: 947-955.

Grignard, E., Morin, J., Vernet, P. & Drevet, J. R. (2005). GPX5 orthologs of the mouse epididymis-restricted and sperm-bound selenium-independent glutathione peroxidase are not expressed with the same quantitative and spatial characteristics in large domestic animals. *Theriogenology*. **64**: 1016–1033.

Groll, M. & Huber, R. (2005). Purification, crystallization and X-ray analysis of the yeast 20S proteosome. *Methods Enzymol*. **398**: 329-336.

Gustavsson, N., Kokke, B. P., Härndahl, U., Silow, M., Bechtold, U., Poghosyan, Z., Murphy, D., Boelens, W. C. & Sundby, C. (2002). A peptide methionine sulfoxide reductase highly expressed in photosynthetic tissue in *Arabidopsis thaliana* can protect the chaperone-like activity of a chloroplast-localized small heat shock protein. *Plant J.* **29**: 545-553.

Gutteridge, J. M. (1993). Free radicals in disease processes: a compilation of cause and consequence. *Free Radic. Res. Commun*. **19**: 141-158.

H

Haas, D. J. & Rossman, M. G. (1970). Crystallographic studies on lactate dehydrogenase at -75°C. *Acta Cryst.* ***B*****26**: 998-1004.

Halliwell, B. & Gutteridge, J. M. (1989). Free radicals in biology and medicine. Oxford: Clarendon Press.

Harata, K. & Kanai, R. (2002). Crystallographic dissection of the thermal motion of protein-sugar complex. *Proteins*. **48**: 53-62.

Hatahet, F. & Ruddock, L. W. (2007). Substrate recognition by the protein disulfide isomerases. *FEBS J.* **274**: 5223-5234.

Hauptman, H. (1986). The direct methods of X-ray crystallography. *Science*. **233**: 178-183.

Hauptman, H. (1997). Phasing methods for protein crystallography. *Curr. Opin. Struct. Biol*. **7**: 672-680.

Heikinheimo, P., Tuominen, V., Ahonen, A-K., Teplyakov, A., Cooperman, B. S., Baykov, A. A., Lahti, R. & Goldman, A. (2001). Toward a quantum-mechanical description of metal-assited phosphoryl transfer in pyrophosphatase. *Proc. Natl. Acad. Sci. USA*. **98**: 3121-3126.

Helliwell, J. R. (1992). Synchrotron X-ray crystallography techniques: time-resolved aspects of data collection. *Phil. Trans. R. Soc. Lond. A.* **340**: 221-232.

Helliwell, J. R. (2005). Protein crystal perfection and its application. *Acta Cryst.* ***D*61**: 793-798.

Hemming, S. A., Bochkarev, A., Darst, S. A., Kornberg, R. D., Ala, P., Yang, D. S. C. & Edwards, A. M. (1995). The mechanism of protein crystal growth from lipid layers. *J. Mol. Biol*. **246**: 308-316.

Hennecke, J., Spleiss, C., Gloockshuber, R. (1997). Influence of acidic residues and the kink in the active-site helix on the properties of the disulfide oxidoreductase DsbA. *J. Biol. Chem*. **272**: 189-195.

Herald, V. L., Heazlewood, J. L., Day, D. A. & Millar, A. H. (2003). Proteomic identification of divalent metal cation binding proteins in plant mitochondria. *FEBS Lett*. **537**: 96-100.

Heras, B. & Martin, J. L. (2005). Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Cryst.* ***D*61**: 1173-1180.

Herbette, S., Lenne, C., Leblanc, N., Julien, J.-L., Drevet, J. R. & Roeckel-Drevet, P. (2002). Two GPX-like proteins from *Lycopersion esculentum* and *Helianthus annuus* are antioxidant enzymes with phospholipid hydroperoxide glutathione peroxidase and thioredoxin peroxidase activities. *Eur. J. Biochem*. **269**: 2414–2420.

Herrero, E. & de la Torre-Ruiz, M. A. (2007). Monothiol glutaredoxins: a common domain for multiple functions. *Cell. Mol. Life Sci*. **64**: 1518-1530.

Herrero, E., Ros, J., Tamarit, J. & Bellí, G. (2006). Glutaredoxins in fungi. *Photosynth. Res*. **89**: 127-140.

Hirotsu, S., Abe, Y., Okada, K., Nagahara, N., Hori, H., Nishino, T. & Hakoshima, T. (1999). Crystal structure of a multifunctional 2-Cys peroxiredoxin heme-binding protein 23 kDa/proliferation-associated gene product. *Proc. Natl Acad. Sci. USA* **96**: 12333–12338.

Hodel, A., Kim, S. H., & Brunger, A. T. (1992). Model bias in macromolecular crystal structures. *Acta Cryst.* ***A*48**: 851-858.

Holland, D., Ben-Hayyim, G., Faltin, Z., Camoin, L., Strosberg, A. D. & Eshdat, Y. (1993). Molecular characterization of salt-stress-associated protein in citrus: protein and cDNA sequence homology to mammalian glutathione peroxidases. *Plant Mol. Biol*. **21**: 923–927.

Holmgren, A. (1976). Hydrogen donor system for *Escherichia coli* ribonucleoside-diphosphate reductase dependent upon glutathione. *Proc. Natl. Acad. Sci. USA.* **73**: 2275-2279.

Holmgren, A. (1979). Glutathione-dependent synthesis of deoxyribonucleotides. Characterization of the enzymatic mechanism of Escherichia coli glutaredoxin. *J. Biol. Chem.* **254**: 3672-3678.

Holmgren, A. (1981). Thioredoxin: structure and functions. *Trends in Biochem. Sci*. **6**: 26-29.

Holmgren, A. (1985). Thioredoxin. *Annu. Rev. Biochem*. **54**: 237-271.

Holmgren, A. (1989). Thioredoxin and glutaredoxin systems. *J. Biol. Chem*. **264**: 13963-13966.

Holmgren, A. & Aslund, F. (1995). Glutaredoxin. *Methods Enzymol*. **252**: 283-292.

Holmgren, A., Söderberg, B. O., Eklund, H. & Brändén, C. I. (1975). Three-dimensional structure of *Escherichia coli* thioredoxin-S2 to 2.8 Å resolution. *Proc. Natl. Acad. Sci. USA.* **72**: 2305-2309.

Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A. & Podjarny, A. (2004). Ultrahigh resolution drug design I: details of interactions in human aldose reductase-inhibitor complex at 0.66 Å. *Proteins*. **55**: 792-804.

Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph*. **14**: 33–38.

I

Isakov, N., Witte, S. & Altman, A. (2000). PICOT-HD: a highly conserved protein domain that is often associated with thioredoxin and glutaredoxin modules. *Trends Biochem. Sci.* **25**: 537-539.

J

Jacquot, J.-P., Gelhaye, E., Rouhier, N., Corbier, C., Didierjean, C. & Aubry, A. (2002). Thioredoxins and related proteins in photosynthetic organisms: molecular basis for thiol dependent regulation. *Biochem. Pharmacol*. **64**: 1065-1069.

Jacquot, J.-P., Rivera-Madrid, R., Marinho, P., Kollarova, M., Le Maréchal, P., Miginiac-Maslow, M. & Meyer, Y. (1994). *Arabidopsis thaliana* NAPHP thioredoxin reductase. cDNA characterization and expression of the recombinant protein in *Escherichia coli*. *J. Mol. Biol*. **235**: 1357-1363.

Jancarik, J., Kim, S.–H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr*. **24**: 409-411.

Jao, S. C., English Ospina, S. M., Berdis, A. J., Starke, D. W., Post, C. B. & Mieyal, J. J. (2006). Computational and mutational analysis of human glutaredoxin (thioltransferase): probing the molecular basis of the low pK_a of cysteine 22 and its role in catalysis. *Biochemistry*. **45**: 4785-4796.

Johansson, C., Kavanagh, K. L., Gileadi, O. & Opperman, U. (2007). Reversible sequestration of active site cysteines in a 2Fe-2S-bridged dimer provides a mechanism for glutaredoxin 2 regulation in human mitochondria. *J. Biol. Chem*. **282**: 3077-3082.

Johansson, C., Lillig, C. H. & Holmgren, A. (2004). Human mitochondrial glutaredoxin reduces S-glutathionylated proteins with high affinity accepting electrons from either glutathione or thioredoxin reductase. *J. Biol. Chem*. **279**: 7537-7543.

Jones, D. P., Go, Y.-M., Anderson, C. L., Ziegler, T. R., Kinkade, Jr. J. M. & Kirlin, W. G. (2004). Cysteine/ cystine couple is a newly recognized node in the circuitry for biologic redox signaling and control. *FASEB J*. **18**: 1246-1248.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys*. **79**: 926–935.

Jung, B. G., Lee, K. O., Lee, S. S., Chi, Y. H., Jang, H. H., Kang, S. S. et al. (2002). A chinese cabbage cDNAwith high sequence identity to phospholipids hydroperoxide glutathione peroxidases encodes a novel isoform of thioredoxin-dependent peroxidase. *J. Biol. Chem*. **277**: 12572–12578.

Jung, C. H. & Thomas, J. A. (1996). S-glutathionylated hepatocyte proteins and insulin disulfides as substrates for reduction by glutaredoxin, thioredoxin, protein disulfide isomerase, and glutathione. *Arch. Biochem. Biophys*. **335**: 61-72.

Juttner, J., Olde, D., Langridge, P. & Baumann, U. (2000). Cloning and expression of a distinct subclass of plant thioredoxins. *Eur. J. Biochem*. **267**: 7109-7117.

K

Kang, S.-G., Jeong, H. K. & Suh, H. S. (2004).Characterization of a new member of the glutathioneperoxidase gene family in *Oryza sativa*. *Mol. Cells*. **17**: 23–28.

Karle J. (1989). Direct methods in protein crystallography. *Acta Cryst*. **A45**, 765-781.

Katti, S. K., LeMaster, D. M. & Eklund, H. (1990). Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. *J. Mol. Biol*. **212**: 167-184.

Kauffmann, B., Weiss, M. S., Lamzin, V. S. & Schmidt, A. (2006). How to avoid premature decay of your macromolecular crystal: a quick soak for long life. *Structure*. **14**: 1099-1105.

Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M. & Creighton, T. E. (1996). Structure determination of the N-terminal thioredoxin-like domain of protein disulfide

isomerase using multidimensional heteronuclear 13C/15N NMR spectroscopy. *Biochemistry*. **35**: 7684-7691.

Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M. & Creighton, T. E. (1997). The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules. *Curr. Biol*. **7**: 239-245.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature*. **185**: 422-427.

Kettenberger, H., Armache, K. J. & Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell*. **114**: 347-357.

Kim, S. J., Woo, J. R., Hwang, Y. S., Jeong, D. G., Shin, D. H., Kim, K. & Ryu, S. E. (2003). The tetrameric structure of *Haemophilus influenza* hybrid Prx5 reveals interactions between electron donor and acceptor proteins. *J. Biol. Chem*. **278**: 10790-10798.

Kingston, R. L., Baker, H. M. & Baker, E. N. (1994). Search designs for protein crystallization based on orthogonal arrays. *Acta Cryst.* ***D*****50**: 429-440.

Kishan, K. V. (2007). Structural biology, protein conformations and drug designing. *Curr. Protein Pept*. Sci. **8**: 376-380.

Kleiger, G. & Eisenberg, D. (2002). GXXXG and GXXXA motifs stabilize FAD and NAD(P)-binding Rossmann folds through C(alpha)-H…. O hydrogen bonds and van der waals interactions. *J. Mol. Biol*. **323**: 69-76.

Kleywegt, G. J. (2000). Validation of protein crystal structures. *Acta Cryst.* ***D*****56**: 249-265.

Kleywegt, G. J. & Jones, T. A. (1997). Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Cryst.* ***D*****53**: 179-185.

Koh, C. S., Didierjean, C., Navrot, N., Panjikar, S., Mulliert, G., Rouhier, N., Jacquot, J.-P., Aubry, A., Shawkataly, O. & Corbier, C. (2007) Crystal structures of a poplar thioredoxin peroxidase that exhibits the structure of glutathione peroxidases: insights into redox-driven conformational changes. *J. Mol. Biol.* **370**: 512-529.

Koh, C. S., Navrot, N., Didierjean, C., Rouhier, N., Hirasawa, M., Knaff, D. B., Wingsle, G., Samian, R., Jacquot, J.-P., Corbier, C. & Gelhaye, E. (2008) An atypical catalytic mechanism involving three cysteines of thioredoxin. *J. Biol. Chem*. In press.

Kundrot, C. E. (1997). The preparation and crystallization of RNA: a sparse matrix approach. *Methods Enzymol*. **276**: 143-157.

$\mathcal{L}$

Ladner, J. E., Parsons, J. F., Rife, C. L., Gilliland, G. L. & Armstrong, R. N. (2004). Parallel evolutionary pathways for glutathione transferases: structure and mechanism of the mitochondrial class kappa enzyme rGSTK1-1. *Biochemistry.* **43**: 352-361.

Laloi, C., Rayapuram, N., Chartier, Y., Grienenberger, J. M., Bonnard, G. & Meyer, Y. (2001). Identification and characterization of a mitochondrial thioredoxin system in plants. *Proc. Natl. Acad. Sci. USA.* **98**: 14144-14149.

Lalonde, J. M., Bernlohr, D. A. & Banaszak, L. J. (1994). The up-and-down beta-barrel proteins. *FASEB J.* **8**: 1240-1247.

Lee, M. Y., Shin, K. H., Kim, Y. K., Suh, J. Y., Gu, Y. Y., Kim, M. R., Hur, Y. S., Son, O., Kim, J. S., Song, E., Lee, M. S., Nam, K. H., Hwang, K. H., Sung, M. K., Kim, H. J., Chun, J. Y., Park, M., Ahn, T. I., Hong, C. B., Lee, S. H., Park, H. J., Park, J. S., Verma, D. P. & Cheon, C. I. (2005). Induction of thioredoxin is required for nodule development to reduce reactive oxygen species levels in soybean roots. *Plant Physiol.* **139**: 1881-1889.

Leibundgut, M., Jenni, S., Frick, C. & Ban, N. (2007). Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science*. **316**: 288-290.

Lemaire, S. D. (2004). The glutaredoxin family in oxygenic photosynthetic organisms. *Photosynth. Res*. **79**: 305-318.

Lennon, B. W., Williams Jr, C. H. & Ludwig, M. L. (2000). Twists in catalysis: alternating conformations of *Escherichia coli* thioredoxin reductase. *Science*. **289**: 1190-1194.

Leslie, A. G. & Walker, J. E. (2000). Structural model of F1-ATPase and the implications for rotary catalysis. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. **355**: 465-471.

Lillig, C. H. & Holmgren, A. (2007). Thioredoxin and related molecules: from biology to health and disease. *Antioxid. Redox Signal*. **9**: 25-47.

Lillig, C. H., Prior, A., Schwenn, J. D., Aslund, F., Ritz, D., Vlamis-Gardikas, A. & Holmgren, A. (1999). New thioredoxins and glutaredoxins as electron donors of 3'-phosphoadenylylsulfate reductase. *J. Biol. Chem*. **274**: 7695-7698.

Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. (2001). Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Cryst.* ***D*****57**: 679-688.

Longhi, S., Czjzek, M. & Cambillau, C. (1998). Messages from ultrahigh resolution crystal structures. *Curr. Opin. Struct. Biol.* **8**: 730-737.

Longhi, S., Czjzek, M., Lamzin, V. S., Nicolas, A. & Cambillau, C. (1997). Atomic resolution (1.0 Å) crystal structure of *Fusarium solani* cutinase: stereochemical analysis. *J. Mol. Biol*. **268**: 779-799.

Lopez-Jaramillo, F. J., García-Ruíz, J. M., Gavira, J. A. & Otálora, F. (2001). Crystallization and cryocrystallography inside X-ray capillaries. *J. Appl. Cryst*. **34**: 365-370.

Lorber, B. & DeLucas, L. J. (1990). Large scale preparation of homogeneous bacteriorhodopsin. *FEBS Lett*. **261**: 14-18.

Lorber, B. & Giegé, R. (2001). Nucleation and growth of thaumatin crystals within a gel under microgravity on STS-95 mission vs. under Earth's gravity. *J. Cryst. Growth*. **231**: 252-261.

Lorber, B., Sauter, C., Ng, J. D., Zhu, D. W., Giegé, R., Vidal, O., Robert, M. C. & Capelle, B. (1999a). Characterization of protein and virus crystals by quasi-planar wave X-ray topography : a comparison between crystals grown in solution and in agarose. *J. Cryst. Growth*. **204**: 357-368.

Lorber, B., Sauter, C., Robert, M. C., Capelle, B. & Giegé, R. (1999b). Crystallization within agarose gel in microgravity improves the quality of thaumatin crystals. *Acta Cryst. **D***.**55**: 1491-1494.

M

Maeda, K., Hägglund, P., Finnie, C., Svensson, B. & Henriksen, A. (2006). Structural basis for target protein recognition by the protein disulfide reductase thioredoxin. *Structure*. **14**: 1701-1710.

Maiorino, M., Ursini, F., Bosello, V., Toppo, S., Tosatto, S. C., Mauri, P., Becker, K., Roveri, A., Bulato, C., Benazzi, L., De Palma, A. & Flohé, L. (2007). The thioredoxin specificity of Drosophila Gpx: a paradigm for a peroxiredoxin-like mechanism of many glutathione peroxidases. *J. Mol. Biol*. **365**: 1033-1046.

Martin, J. L. (1995). Thioredoxin – a fold for all reasons. *Structure*. **3**: 245-250.

Masutani, H., Bai, J., Kim, Y. C. & Yodoi, J. (2004). Thioredoxin as a neurotrophic cofactor and an important regulator of neuroprotection. *Mol. Neurobiol.* **29**: 229-242.

McCurbey, J. A. & Franklin, R. A. (2006). Reactive oxygen intermediates and signaling through kinase pathways. *Antioxid. Redox Signal*. **8**: 1745-1748.

McCurbey, J. A., LaHair, M. M. & Franklin, R. A. (2006). Reactive oxygen species-induced activation of the MAP kinase signaling pathways. *Antioxid. Redox Signal*. **8**: 1775-1789.

McPherson, A. (1982). Preparation and analysis of protein crystals. New York: John Wiley.

McPherson, A. (1990). Current approaches to macromolecular crystallization. *European J. Biochem*. **189**: 1-23.

McPherson, A. (1996). Macromolecular crystal growth in microgravity. *Crystallogr. Rev.* **6**: 157-308.

McPherson, A. (1999). Crystallization of biological macromolecules. New York: Cold spring harbor laboratory press.

McPherson, A. (2003). Macromolecular crystallization in the structural genomics era. *J. Struct. Biol.* **142**: 1-2.

McPherson, A. & Shlichta, P. J. (1987). Facilitation of the growth of protein crystals by heterogeneous/epitaxial nucleation. *J. Cryst. Growth*. **85**: 206-214.

McRee, D. E. (1999). XtalView/Xfit – A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol*. **125**: 156-165.

Menchise, V., Corbier, C., Didierjean, C., Jacquot, J. P., Benedetti, E., Saviano, M. & Aubry, A. (2000). Crystal structure of the W35A mutant thioredoxin *h* from *Chlamydomonas reinhardtii* : the substitution of the conserved active site Trp leads to modifications in the environment of the two catalytic cysteines. *Biopolymers*. **56**: 1-7.

Menchise, V., Corbier, C., Didierjean, C., Saviano, M., Benedetti, E., Jacquot, J. P. & Aubry, A. (2001). Crystal structure of the wild-type and D30A mutant thioredoxin *h* of *Chlamydomonas reinhardtii* and implications for the catalytic mechanism. *Biochem. J*. **359**: 65-75.

Messens, J. & Collet, J-F. (2005). Pathways of disulfide bond formation in *Escherichia coli. The International J. Biochem. Cell Biol.* **38**: 1050-1062.

Meyer, Y., Riondet, C., Constants, L., Abdelgawwad, M. R., Reichheld, J. P. & Vignols, F. (2006). Evolution of redoxin genes in the green lineage. *Photosynth. Res.* **89**: 179-192.

Meyer, Y., Siala, W., Bashandy, T., Riondet, C., Vignols, F. & Reichheld, J. P. (2007). Glutaredoxins and thioredoxins from plants. *Biochim. Biophys. Acta.* In press.

Miao, Y., Lv, D.,Wang, P.,Wang, X.-C., Chen, J., Miao, C. & Song, C.-P. (2006). An *Arabidopsis* glutathione peroxidase functions as both a redox tranducer and a scavenger in abscisic acid and drought stress responses. *Plant Cell*, **18**: 2749–2766.

Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *SnB*: crystal structure determination via *Shake-and-Bake*. *J. Appl. Cryst*. **27**: 613-621.

Milla, M. A. R., Maurer, A., Huete, A. R. & Gustafson, J. P. (2003). Glutathione peroxidase genes in Arabidopsis are ubiquitous and regulated by abiotic stresses through diverse signalling pathways. *The Plant J.* **36**: 602–615.

Minarik, P. (1997). Thioredoxin: a protein with a thousand faces. *Biologia (Bratislava).* **52**: 691-696.

Mittler, R., Vanderauwera, S., Gollery, M. & Van Breusegem, F. (2004). Reactive oxygen gene network of plants. *Trends Plant Sci.* **9**: 490-498.

Møller, I. M. & Kristensen, B. K. (2004). Protein oxidation in plant mitochondria as a stress indicator. *Photochem. Photobiol. Sci*. **3**: 730-735.

Møller, I. M. & Kristensen, B. K. (2006). Protein oxidation in plant mitochondria detected as oxidized tryptophan. *Free Radical Biol. Med*. **40**: 430-435.

Moreno, A., González-Ramírez, L. A., Hernández, M. d. l. A., Oliver-Salvador, C., Soriano-García, M. & Rodríguez-Romero, A. (1999). Investigations on gravity influence upon protein crystallization by the gel acupuncture technique. *J. Cryst. Growth*. **196**: 587-594.

Moreno, A., Saridakis, E. & Chayen, N. E. (2002). Combination of oils and gels for enhancing the growth of protein crystals. *J. Appl. Cryst*. **35**: 140-142.

Morris, R. J. & Brigcone, G. (2003). Sheldrick's 1.2 Å rule and beyond. *Acta Cryst.* ***D*59**: 615-617.

Mössner, E., Huber-Wunderlich, M. & Glockshuber, R. (1998). Characterization of *Escherichia coli* thioredoxin variants mimicking the active-sites of other thiol/disulfide oxidoreductases. *Protein Sci*. **7**: 1233-1244.

Mössner, E., Iwai, H. & Glockshuber, R. (2000). Influence of the pK(a) value of the buried, active-site cysteine on the redox properties of thioredoxin-like oxidoreductases. *FEBS Lett*. **477**: 21-26.

Mueller, U., Nyarsik, L., Horn, M., Rauth, H., Przewieslik, T., Saenger, W., Lehrach, H. & Eickhoff, H. (2007). Development of a technology for automation and miniaturization of protein crystallization. *J. Biotech*. **85**: 7-14.

Mugesh, G., du Mont,W. & Sies, H. (2001). Chemistry of biologically important synthetic organoselenium compounds. *Chem. Rev*. **101**: 2125–2179.

Muzet, N., Guillot, B., Jelsch, C., Howard, E. & Lecomte, C. (2003). Electrostatic complementarity in an aldose reductase complex from ultra-high-resolution crystallography and first-principles calculations. *Proc. Natl. Acad. Sci. USA*. **100**: 8742-8747.

N

Nakamura, H. (2005). Thioredoxin and its related molecules: update 2005. *Antioxid. Redox Signal.* **7**: 823-828.

Navrot, N., Collin, V., Gualberto, J., Gelhaye, E., Hirasawa, M., Rey, P., Knaff, D. B., Issakidis, E., Jacquot, J. P. & Rouhier, N. (2006). Plant glutathione peroxidases are functional peroxiredoxins distributed in several subcellular compartments and regulated during biotic and abiotic stresses. *Plant Physiol*. **142**: 1364–1379.

Newcomer M. E. & Ong, D. E. (2000). Plasma retinol binding protein: structure and function of the prototypic lipocalin. *Biochim. Biophys. Acta*. **1482**: 57-64.

Newkirk, K., Feng, W., Jiang, W., Tejero, R., Emerson, S. D., Inouye, M. & Montelione, G. T. (1994). Solution NMR structure of the major cold shock protein (CspA) from *Escherichia coli*: identification of a binding epitope for DNA. *Proc. Natl. Acad. Sci. USA*. **91**: 5114-5118.

Ng, J. D., Gavira, J. A. & García-Ruíz, J. M. (2003). Protein crystallization by capillary counterdiffusion for applied crystallographic structure determination. *J. Struct. Biol.* **142**: 218-231.

Ng, J. D., Lorber, B., Giegé, R., Koszelak, S., Day, J., Greenwood, A. & McPherson, A. (1997). Comparative analysis of thaumatin crystals grown on earth and in microgravity. *Acta Cryst.* ***D*****53**: 724-733.

Noctor, G. & Foyer, C. H. (1998). Ascorbate and glutathione: keeping active oxygen under control. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. **49**: 249–279.

Noguera, V., Walker, O., Rouhier, N., Jacquot, J.-P., Krimm, I. & Lancelin, J. M. (2005). NMR reveals a novel glutaredoxin-glutaredoxin interaction interface. *J. Mol. Biol*. **353**: 629-641.

Nordstrand, K., Åslund, F., Holmgren, A., Otting, G. & Berndt, K. D. (1999). NMR structure of *Escherichia coli* glutaredoxin 3-glutathione mixed disulfide complex: implications for the enzymatic mechanism. *J. Mol. Biol*. **286**: 541-552.

Nordstrand, K., Sandström, A., Åslund, F., Holmgren, A., Otting, G. & Berndt, K. D. (2000). NMR structure of oxidized glutaredoxin 3 from *Escherichia coli*. *J. Mol. Biol*. **303**: 423-432.

O

Oblong, J. E., Berggren, M. & Powis, G. (1994). Biochemical, structural and biological properties of human thioredoxin active site peptides. *FEBS Lett*. **343**: 81-84.

P

Pace, C., Shirley, B., McNutt, M. & Gajiwata, K. (1996). Forces contributing to the conformational stability of proteins. *FASEB J*. **10**: 75-83.

Pan, J. L. & Bardwell, J. C. (2006). The origami of thioredoxin-like folds. *Protein Sci.* **15**: 2217-2227.

Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Auto-Rickshaw* – An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Cryst.* ***D*****61**: 449-457.

Pauling, L., Corey, R. B. & Branson, H. R. (1951). The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA*. **37**: 235-240.

Pedersen, B. P., Buch-Pedersen, M. J., Morth, J. P., Palmgren, M. G. & Nissen, P. (2007). Crystal structure of the plasma membrane proton pump. *Nature*. **450**: 1111-1114.

Pellecchia, M., Sebbel, P., Hermanns, U., Wuthrich, K. & Glockshuber, R. (1999). Pilus chaperone FimC-adhesin FimH interactions mapped by TROSY-NMR. *Nat. Struct. Biol.* **6**: 336-339.

Pérez-Ruiz, J. M., Spínola, M. C., Kirchsteiger, K., Moreno, J., Sahrawy, M. & Cejudo, F. J. (2006). Rice NTRC is a high-efficiency redox system for chloroplast protection against oxidative damage. *Plant Cell*. **18**: 2356-2368.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol*. **6**: 458-463.

Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. C. T. (1960). Structure of myoglobin: A three-dimensional Fourier synthesis at 5.5 Angstrom resolution, obtained by X-ray analysis. *Nature*. **185**: 416-422.

Peterson, F. C., Lytle, B. L., Sampath, S., Vinarov, D., Tyler, E., Shahan, M., Markley, J. L. & Volkman, B. F. (2005). Solution structure of thioredoxin *h*1 from *Arabidopsis thaliana*. *Protein Sci*. **14**: 2195-2200.

Petry, S., Brodersen, D. E., Murphy, F. V. 4th., Dunham, C. M., Selmer, M., Tarry, M. J., Kelley, A. C. & Ramakrishnan, V. (2005). Crystal structures of the ribosome

in complex with release factors RF1 and RF2 bound to a cognate stop codon. *Cell*. **123**: 1255-1266.

Pitzcheke, A., Forzani, C. & Hirt, H. (2006). Reactive oxygen species signaling in plants. *Antioxid. Redox Signal*. **8**: 1757-1764.

Powis, G. & Kirkpatrick, D. L. (2007). Thioredoxin signaling as a target for cancer therapy. *Curr. Opin. Pharmacol*. **7**: 392-397.

Powis, G. & Monfort, W. R. (2001). Properties and biological activities of thioredoxins. *Annu. Rev. Biophys. Biomol. Struct*. **30**: 421-455.

Prabhakar, R., Morokuma, K. & Musaev, D. G. (2006). Peroxynitrite reductase activity of selenoprotein glutathione peroxidase: a computational study. *Biochemistry*. **45**: 6967–6977.

Prabhakar, R., Vreven, T., Morokuma, K. & Musaev, D. G. (2005). Elucidation of the mechanism of selenoprotein glutathione peroxidase (Gpx)-catalyzed hydrogen peroxide reduction by two glutathione molecules: a density functional study. *Biochemistry*. **44**: 11864–11871.

Przbylska, M. (1988). A novel method of mounting a protein crystal on a surface perpendicular to the X-ray capillary. *J. Appl. Cryst*. **21**: 272-273.

Pusey, M. L., Liu, Z-J., Tempel, W., Praissman, J., Lin, D., Wang, B-C., Gavira, J. A. & Ng, J. D. (2005). Life in the fast lane for protein crystallization and X-ray crystallography. *Prog. Biophys. Mol. Biol*. **88**: 359-386.

R

Rahlfs, S., Fischer, M. & Becker, K. (2001). *Plasmodium falciparum* possesses a classical glutaredoxin and a second, glutaredoxin-like protein with a PICOT homology domain. *J. Biol. Chem*. **276**: 37133-37140.

Ravelli, R. B., Nanao, M. H., Lovering, A., White, S. & McSweeney, S. (2005). Phasing in the presence of radiation damage. *J. Synchrotron Radiat.* **12**: 276-284.

Reichheld, J.-P., Mestres-Ortega, D., Laloi, C. & Meyer, Y. (2002). The multigenic family of thioredoxin *h* in *Arabidopsis thaliana*: specific expression and stress response. *Plant Physiol. Biochem*. **40**: 685-690.

Ren, B., Huang, W., Akesson, B. & Ladenstein, R. (1997). The crystal structure of seleno-glutathione peroxidase from human plasma at 2.9 Å resolution. *J. Mol. Biol*. **268**: 869-885.

Ren, X., Yang, L., Liu, J., Su, D., You, D., Liu, C., Zhang, K., Luo, G., Mu, Y., Yan, G. & Shen, J. (2001). A novel glutathione peroxidase mimic with antioxidant activity. *Arch. Biochem. Biophys*. **387**: 250–256.

Reynaert, N. L., Ckless, K., Guala, A. S., Wouters, E. F., van der Vliet, A. & Janssen-Heininger, Y. M. (2006). *In situ* detection of S-glutathionylated proteins following glutaredoxin-1 catalyzed cysteine derivatization. *Biochim. Biophys. Acta*. **1760**: 380-387.

Rey, P., Pruvot, G., Becuwe, N., Eymery, F., Rumeau, D. & Peltier, G. (1998). A novel thioredoxin-like protein located in the chloroplast is induced by water deficit in *Solanum tuberosum* L. plants. *Plant J*. **13**: 97-107.

Rhodes, G. (2006). Crystallography made crystal clear. A guide for users of macromolecular models. ed. 2. USA: Academic Press Publications.

Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). Single-wavelength anomalous diffraction phasing revisited. *Acta Cryst.* ***D*****56**: 1413-1420.

Ridder, I. S., Rozeboom, H. J. & Dijkstra, B. W. (1999). Haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10 refined at 1.15 Å resolution. *Acta Cryst.* ***D*****55**: 1273-1290.

Rodgers, D. W. (1997). Practical cryocrystallography. *Methods in Enzymol*. **276**: 183-203.

Rodríguez-Manzaneque, M. T., Ros, J., Cabiscol, E., Sorribas, A. & Herrero, E. (1999). Grx5 glutaredoxin plays a central role in protection against protein oxidative damage in *Saccharomyces cerevisiae*. *Mol.Cell.Biol.* **19**: 8180-8190.

Rodríguez-Manzaneque, M. T., Tamarit, J., Belli, G., Ros, J. & Herrero, E. (2002). Grx5 is a mitochondrial glutaredoxin required for the activity of iron/sulfur enzymes. *Mol. Biol. Cell*. **13**: 1109-1121.

Rose G., Fleming P., Banavar J. & Maritan A. (2006). A backbone-based theory of protein folding. *Proc. Natl. Acad. Sci. USA.* **103**: 16623-16633.

Rosenbaum, G., Holmes, K. C. & Witz, J. (1971). Synchrotron radiation as a source for X-ray diffraction. *Nature*. **230**: 434-437.

Rossmann, M. G. (1972). The molecular replacement method. New York: Gordon & Breach.

Rouhier, N., Couturier, J. & Jacquot, J.-P. (2006). Genome-wide analysis of plant glutaredoxin systems. *J. Exp. Bot*. **57**: 1685–1696.

Rouhier, N., Gelhaye, E. & Jacquot, J.-P. (2004). Plant glutaredoxins: still mysterious reducing systems. *Cell. Mol. Life Sci*. **61**: 1266–1277.

Rouhier, N. & Jacquot, J.-P. (2002). Plant peroxiredoxins: alternative hydroperoxide scavenging enzymes. *Photosynth. Res*. **74**: 259–268.

Rouhier, N. & Jacquot, J.-P. (2005). The plant multigenic family of thiol peroxidases. *Free Radic. Biol. Med*. **38**: 1413–1421.

Rouhier, N., Koh, C. S., Gelhaye, E., Corbier, C., Favier, F., Didierjean, C. & Jacquot, J.-P. (2008a). Redox based anti-oxidant systems in plants: Biochemical and structural analyses. *Biochim. Biophys. Acta*. In press.

Rouhier, N., Lemaire, S. D. & Jacquot, J.-P. (2008b). The role of glutathione in photosynthetic organisms: emerging functions for glutaredoxin and glutathionylation. *Annu. Rev. Plant Biol*. In press.

Rouhier, N., Unno, H., Bandyopadhay, S., Masip, L., Kim, S. K., Hirasawa, M., Gualberto, J., Lattard, V., Kusunoki, M., Knaff, D. B., Geogiou, G., Hase, T., Johnson, M. K. & Jacquot, J. P. (2007). Functional, structural and spectroscopic characterization of a glutathione-ligated [2Fe-2S] cluster in poplar glutaredoxin C1. *Proc. Natl. Acad. Sci. USA.* **104**: 7379-7384.

Roussel, A. & Cambillau, C. (1989). Silicon graphics geometry partner directory. Pp. 77-78. Silicon Graphics, Mountain View, CA.

Roy, G., Sarma, B. K., Phadnis, P. P. & Mugesh, G. (2005). Selenium-containing enzymes in mammals: chemical perspectives. *J. Chem. Sci*. **117**: 287–303.

Ruiz, F., Hazemann, I., Mitschler, A., Joachimiak, A., Schneider, T., Karplus, M. & Podjarny, A. (2004). The crystallographic structure of the aldose reductase-IDD552 complex shows direct proton donation from tyrosine 48. *Acta Cryst.* ***D*60**: 1347-1354.

Russell R. B., Alber F., Aloy, P., Davis, F. P., Korkin, D., Pichaud, M., Topf, M. & Sali, A. (2004). A structural perspective on protein-protein interactions. *Curr. Opin. Struct. Biol.* **14**: 313-324.

Russell R. B., Sasieni, P. D. & Sternberg, M. J. (1998). Supersites within superfolds. Binding site similarity in the absence of homology. *J. Mol. Biol.* **282**: 903-918.

S

Saarinen, M., Gleason, F. K. & Eklund, H. (1995). Crystal structure of thioredoxin-2 from Anabaena. *Structure*. **3**: 1097-1108.

Sali, A. (2003). NIH workshop on structural proteomics of biological complexes. *Structure*. **11**: 1043-1047.

Sali, A., Glaeser, R., Earnest, T. & Baumeister, W. (2003). From words to literature in structural proteomics. *Nature*. **422**: 216-225.

Sali, A. & Kuriyan, J. (1999). Challenges at the frontiers of structural biology. *Trends Cell Biol*. **9**: M20-M24.

Sandalova, T., Zhong, L., Lindqvist, Y., Holmgren, A. & Schneider, G. (2001). Three-dimensional structure of a mammalian thioredoxin reductase: implications for mechanism and evolution of a selenocysteine-dependent enzyme. *Proc. Natl. Acad. Sci. USA*. **98**: 9533-9538.

Sanjoh, A., Tsukihara, T. & Gorti, S. (2001). Surface-potential controlled Si-microarray devices for heterogeneous protein crystallization screening. *J. Cryst. Growth*. **232**: 618-628.

Santos, V. D. C. & Rey, P. (2006). Plant thioredoxins are key actors in the oxidative stress response. *Trends Plant Sci*. **11**: 329-334.

Sauter, C., Lorber, B. & Giegé, R. (2002). Towards atomic resolution with crystals grown in gel: the case of thaumatin seen at room temperature. *Proteins.* **48**: 146-150.

Scapin, G. (2006). Structural biology and drug discovery. *Curr. Pharm. Des*. **12**: 2087-2097.

Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*. **436**: 138-141.

Scheerer, P., Borchert, A., Krauss, N., Wessner, H., Gerth, C., Höhne, W. & Kuhn, H. (2007). Structural basis for catalytic activity and enzyme polymerization of phospholipid hydroxyperoxide glutathione peroxidase-4 (Gpx4). *Biochemistry.* **46**: 9041-9049.

Schmidt, A., Gonzales, A., Morris, T., Costabel, M., Alzari, P. M. & Lamzin, V. S. (2002). Advantages of high resolution phasing: MAD to atomic resolution. *Acta Cryst*. ***D*58**: 1433-1441.

Schmidt, A. & Lamzin, V. S. (2002). Veni, vidi, vici: atomic resolution unraveling the mysteries of protein function. *Curr. Opin. Struct. Biol*. **12**: 698-703.

Schmidt, A. & Lamzin, V. S. (2007). From atoms to proteins. *Cell. Mol. Life Sci*. **64**: 1959-1969.

Schröder, E., Littlechild, J. A., Lebedev, A. A., Errington, N., Vagin, A. A. & Isupov, M. N. (2000). Crystal structure of decameric 2-Cys peroxiredoxin from human erythrocytes at 1.7 Å resolution. *Structure*. **8**: 605–615.

Schürmann, P. (2003). Redox signaling in the chloroplast: the ferredoxin/thioredoxin system. *Antioxid. Redox Signal*. **5**: 69-78.

Segelke, B. W. (2001). Efficiency analysis of sampling protocols used in protein crystallization screening. *J. Cryst. Growth*. **232**: 553-562.

Sheldrick, G. M. (1990). Phase annealing in SHELX-90: direct methods for larger structures. *Acta Cryst*. ***A*46**: 467-473.

Sheldrick, G. M. (2006). SHELX. http://shelx.uni-ac.gwdg.de/SHELX/index.html.

Shelton, M. D., Chock, P. B. & Mieyal, J. J. (2005). Glutaredoxin: role in reversible protein S-glutathionylation and regulation of redox signal transduction and protein translocation. *Antioxid. Redox Signal*. **7**: 348-366.

Shelton, M. D., Kern, T. S. & Mieyal, J. J. (2007). Glutaredoxin regulates nuclear factor kappa-B and intercellular adhesion molecule in Müller cells: model of diabetic retinopathy. *J. Biol. Chem*. **282**: 12467-12474.

Sica, F., Demasi, D., Mazzarella, D. L., Zagari, A., Capasso, S., Pearl, L. H., D'Auria, S., Raia, C. A. & Rossi, M. (1994). Elimination of twinning in crystals of *Sulfolobus sofataricus* alcohol dehydrogenase holo-enzyme by growth in agarose gels. *Acta Cryst.* ***D*****50**: 508-511.

Simeoni, F. & Divita, G. (2007). The Dim protein family: from structure to splicing. *Cell. Mol. Life Sci*. **64**: 2079-2089.

Smeets, A., Evrard, C., Landtmeters, M., Marchand, C., Knoops, B. & Declercq, J. P. (2005). Crystal structures of oxidized and reduced forms of human mitochondrial thioredoxin 2. *Protein Sci*. **14**: 2610-2621.

Snell, E. H., Weisgerber, S., Helliwell, J. R., Weckert, E., Holzer, K. & Schroer, K. (1995). Improvements in lysozyme protein crystal perfection through microgravity growth. *Acta Cryst.* ***D*****51**: 1099-1102.

Sodano, P., Xia, T. H., Bushweller, J. H., Bjornberg, O., Holmgren, A., Billeter, M. & Wuthrich, K. (1991). Sequence-specific 1H NMR assignments and determination of the three-dimensional structure of reduced *Escherichia coli* glutaredoxin. *J. Mol. Biol*. **221**: 1311-1324.

Southworth-Davies, R. J. & Garman, E. F. (2007). Radioprotectant screening for cryocrystallography. *J. Synchrotron Radiat*. **14**: 73-83.

Spahn, C. M., Beckmann, R., Eswar, N., Penczek, P. A., Sali, A., Blobel, G. & Frank, J. (2001). Structure of the 80S ribosome from *Saccharomyces cerevisiae*-tRNA-ribosome and subunit-subunit interactions. *Cell*. **107**: 373-386.

Stefanková, P., Kollárová, M. & Barák, I. (2005). Thioredoxin – structural and functional complexity. *Gen. Physiol. Biophys*. **24**: 3-11.

Stevens, R. C. (2004). Long live structural biology. *Nat. Struct. Mol. Biol.* **11**: 293-295.

Stout, G. H. & Jensen, L. H. (1989). X-ray structure determination. ed. 2. Londres: MacMillan.

Srinivasan, J., Miller, J., Kollman, P. A. & Case, D. A. (1998). Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn*. **16**: 671–682.

Srinivasan, U., Mieyal, P. A. & Mieyal, J. J. (1997). pH profiles indicative of rate-limiting nucleophilic displacement in thioltransferase catalysis. *Biochemistry*. **36**: 3199-3206.

Sugimoto, M. & Sakamoto, W. (1997). Putative phospholipids hydroperoxide glutathione peroxidase gene from *Arabidopsis thaliana* induced by oxidative stress. *Genes Genet. Syst*. **72**: 311–316.

Sun, C., Berardi, M. J. & Bushweller, J. H. (1998). The NMR solution structure of human glutaredoxin in the fully reduced form. *J. Mol. Biol*. **280**: 687-701.

Sztajer, H., Gamain, B., Aumann, K. D., Slomianny, C., Becker, K., Brigelius-Flohe, R. & Flohe, L. (2001). The putative glutathione peroxidase gene of *Plasmodium falciparum* codes for a thioredoxin peroxidase. *J. Biol. Chem*. **276**: 7397–7403.

$\mathcal{T}$

Tamarit, J., Bellí, G., Cabiscol, E., Herrero, E. & Ros, J. (2003). Biochemical characterization of yeast mitochondrial Grx5 monothiol glutaredoxin. *J. Biol. Chem*. **278**: 25745-25751.

Tanaka, T., Hosoi, F., Yamaguchi-Iwai, Y., Nakamura, H., Masutani, H., Ueda, S., Nishiyama, A., Takeda, S., Wada, H., Spyrou, G. & Yodoi, J. (2002). Thioredoxin-2 (TRX-2) is an essential gene regulating mitochondria-dependent apoptosis. *EMBO J*. **21**: 1695–1703.

Tanaka, T., Izawa, S. & Inoue, Y. (2005). GPX2, encoding a phospholipid hydroperoxide glutathione peroxidase homologue, codes for an atypical 2-Cys peroxiredoxin in *Saccharomyces cerevisiae*. *J. Biol. Chem*. **276**: 7397–7403.

Teng, T. Y. (1990). Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J. Appl. Cryst*. **23**: 387-391.

Terwilliger, T. C. (2001). Map-likelihood phasing. *Acta Cryst.* ***D*****57**: 1763-1775.

Terwilliger, T. C. (2004). Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Cryst.* ***D*****60**: 2144-2149.

Thakur, A. S., Robin, G., Guncar, G., Saunders, N. F. W., Newman, J. Martin, J. L. & Kobe, B. (2007). Improved success of sparse matrix protein crystallization screening with heterogeneous nucleating agents. *PLoS ONE*. **2**: e1091.

Thiessen, K. J. (1994). The use of two novel methods to grow protein crystals by microdialysis and vapor diffusion in an agarose gel. *Acta Cryst.* ***D*****50**: 491-495.

Timasheff, S. N. (1995). Solvent stabilization of protein structure. *Methods Mol. Biol.* **40**: 253-269.

Torres, M. A., Dangl, J. L. & Jones, J. D. (2002). Arabidopsis gp91^{phox} homologues AtrbohD and AtrbohF are required for accumulation of reactive oxygen intermediates in the plant defense response. *Proc. Natl. Acad. Sci. USA*. **99**: 517-522.

Toth, E. A. (2007). Molecular replacement. *Methods Mol. Biol*. **364**: 121-148.

Tsukamoto, S., Morita, S., Hirano, E., Yokoi, H., Masumura, T. & Tanaka, K. (2005). A novel *cis*-element that is responsive to oxidative stress regulates three antioxidant defense genes in rice. *Plant Physiol*. **137**: 317-327.

U

Uson, I. & Sheldrick, G. M. (1999). Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol*. **9**: 643-648.

V

Vaughn, D. E. & Bjorkman, P. J. (1996). The (Greek) key to structures of neural adhesion molecules. *Neuron*. **16**: 261-273.

Vidal, O., Robert, M. C., Arnoux, B. & Capelle, B. (1999). Crystalline quality of lysozyme crystals grown in agarose and silica gels studied by X-ray diffraction techniques. *J. Cryst. Growth*. **196**: 559-571.

Vlamis-Gardikas, A. & Holmgren, A. (2002). Thioredoxin and glutaredoxin isoforms. *Methods Enzymol*. **347**: 286-296.

W

Wahl, M. C., Irmler, A., Hecker, B., Schirmer, R. H. & Becker, K. (2005). Comparative structural analysis of oxidized and reduced thioredoxin from *Drosophila melanogaster*. *J. Mol. Biol*. **345**: 1119-1130.

Wang, Z., Luecke, H., Yao, N. & Quiocho, F. A. (1997). A low energy short hydrogen bond in very high resolution structures of protein receptor-phosphate complexes. *Nat. Struct. Biol*. **4**: 519-522.

Wang, Z.-X. (1996). How many fold types of protein are there in nature? *Prot. Struc. Func. Gen.* **26**: 186-191.

Weeks, C. M., Blessing, R. H., Miller, R., Mungee, R., Potter, S. A., Rappleye, J., Smith, G. D., Xu, H. & Furey, W. (2002). Towards automated protein structure determination: *BnP*, the *SnB*-PHASES interface. *Z. Kristallogr*. **217**: 686-693.

Weeks, C. M. & Miller, R. (1999). The design and implementation of SnB v2.0. *J. Appl. Cryst*. **32**: 120-124.

Weichsel, A., Gasdaska, J. R., Powis, G. & Montfort, W. R. (1996). Crystal structures of reduced, oxidized, and mutated human thioredoxins: evidence for a regulatory homodimer. *Structure*. **4**: 735-751.

Witte, S., Villalba, M., Bi, K., Liu, Y., Isakov, N. & Altman, A. (2000). Inhibition of the c-Jun N-terminal kinase/AP-1 and NF-kappaB pathways by PICOT, a novel protein kinase C-interacting protein with a thioredoxin homology domain. *J. Biol. Chem*. **275**: 1902-1909.

Wood, Z. A., Schroder, E., Harris, J. R. & Poole, L. B. (2003). Structure, mechanism, and regulation of peroxiredoxins. *Trends Biochem. Sci.* **28**: 32-40.

Wood, Z. A., Poole, L. B., Hantgan, R. R. & Karplus, P. A. (2002). Dimers to doughnuts: redox-sensitive oligomerization of 2-cysteine peroxiredoxins. *Biochemistry*. **41**: 5493–5504.

X

Xia, B., Vlamis-Gardikas, A., Holmgren, A., Wright, P. E. & Dyson, H. J. (2001). Solution structure of *Escherichia coli* glutaredoxin-2 shows similarity to mammalian glutathione-S-transferases. *J. Mol. Biol*. **310**: 907-918.

Xu, H. & Hauptman, H. A. (2006). Recent advances in direct phasing methods for heavy-atom substructure determination. *Acta Cryst.* ***D*62**: 897-900.

Xu, Q. S., Roberts, R. J. & Guo, H. C. (2005). Two crystal forms of the restriction enzyme MspI-DNA complex show the same novel structure. *Protein Sci*. **14**: 2590-2600.

Xu, Z., Horwich, A. L. & Sigler, P. B. (1997). The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature*. **388**: 741-750.

Y

Yang, Y., Jao, S., Nanduri, S., Starke, D. W., Mieyal, J. J. & Qin, J. (1998). Reactivity of the human thioltransferase (glutaredoxin) C7S, C25S, C78S, C82S mutant and NMR solution structure of its glutathionyl mixed disulfide intermediate reflect catalytic specificity. *Biochemistry*. **37**: 17145-17156.

Yang, Y. & Wells, W. W. (1991). Identification and characterization of the functional amino acids at the active site center of pig liver thioltransferase by site-directed mutagenesis. *J. Biol. Chem*. **266**: 12759-12765.

Yoshida, T., Oka, S., Masutani, H., Nakamura, H. & Yodoi, J. (2003). The role of thioredoxin in the aging process: involvement of oxidative stress. *Antioxid. Redox Signal.* **5**: 563-570.

Yu, B. P. (1994). Cellular defenses against damage from reactive oxygen species. *Physiol. Rev*. **74**: 139-162.

Z

Zanotti, G. & Berni, R. (2004). Plasma retinol-binding protein: structure and interactions with retinol, retinoids and transthyretin. *Vitam. Horm*. **69**: 271-295.

Zeppezauer, M., Eklund, H. & Zeppezauer, E. S. (1968). Micro diffusion cells for the growth of single protein crystals by means of equilibrium dialysis. *Arch. Biochem. Biophys*. **126**: 564-573.

Zhou, Z. H., Baker, M. L., Jiang, W., Dougherty, M., Jakana, J., Dong, G., Lu, G. & Chiu, W. (2001). Electron cryomicroscopy and bioinformatics suggest protein fold models for rice dwarf virus. *Nat. Struct. Biol*. **8**: 868-873.

Publications

PUBLICATIONS

International Journals:

Koh, C. S., Navrot, N., Didierjean, C., Rouhier, N., Hirasawa, M., Knaff, D. B., Wingsle, G., Samian, R., Jacquot, J. P., Corbier, C. & Gelhaye, E. (2008). An atypical catalytic mechanism involving three cysteines of thioredoxin. *J. Biol. Chem.* In press.

Rouhier, N., **Koh, C. S.**, Gelhaye, E., Corbier, C., Favier, F., Didierjean, C. & Jacquot, J. P. (2008). Redox based anti-oxidant systems in plants: Biochemical and structural analyses. *Biochim. Biophys. Acta.* In press.

Koh, C. S., Didierjean, C., Navrot, N., Panjikar, S., Mulliert, G., Rouhier, N., Jacquot, J. P., Aubry, A., Shawkataly, O., Corbier, C. (2007). Crystal structures of a poplar thioredoxin peroxidase that exhibits the structure of glutathione peroxidases: insights into redox-driven conformational changes. *J. Mol. Biol*. **370**: 512-529.

Conference proceedings:

Koh, C. S., Didierjean, C., Jacquot, J. P., Corbier, C. & Gelhaye, E. (2007). Elucidation du mécanisme catalytique d'une thiorédoxine de peuplier glutarédoxine-dépendante par cristallographie. *Proceedings of the GTBio colloquium 2007***,** 8-11 October, Lille, France.

Koh, C. S., Didierjean, C. & Corbier, C. (2007). Structure cristalline de la glutathion peroxydase 5 de peuplier, une enzyme thiorédoxine-dépendante: mise en évidence de changements conformationnels rédox-dépendents. *Proceedings of the 15th Réunion du Groupe Français des Peptides et des Protéines (GFPP) 2007*, 20-25 May, Dinard, France.

Koh, C. S., Didierjean, C. & Corbier, C. (2006). The first plant glutathione peroxidase crystal structure: insights to the redox-driven conformational changes and a new mode of dimerization. *Proceedings of the 20th Regio Meeting on Biocrystallography 2006 (International conference between Germany, Switzerland and France)*, 4-6 October, Bildungszentrum Kloster Hegne, Germany.

doi:10.1016/j.jmb.2007.04.031 *J. Mol. Biol.* (2007) **370**, 512–529

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structures of a Poplar Thioredoxin Peroxidase that Exhibits the Structure of Glutathione Peroxidases: Insights into Redox-driven Conformational Changes

Cha San Koh[1], Claude Didierjean[1]*, Nicolas Navrot[2], Santosh Panjikar[3] Guillermo Mulliert[1], Nicolas Rouhier[2], Jean-Pierre Jacquot[2] André Aubry[4], Omar Shawkataly[5] and Catherine Corbier[6]*

[1]*LCM3B, Equipe Biocristallographie, UMR 7036 CNRS-UHP, Faculté des Sciences et Techniques, Nancy Université, BP 239, 54506 Vandoeuvre-lès-Nancy France*

[2]*Unité Mixte de Recherche INRA-UHP 1136, Interactions Arbres/Micro-organismes IFR 110 GEEF, Faculté des Sciences et Techniques Nancy Université, BP 239 54506 Vandoeuvre-lès-Nancy France*

[3]*EMBL Hamburg Outstation c/o DESY, Notkestr 85 D-22603 Hamburg Germany*

[4]*Laboratoire de Chimie Physique Macromoléculaire UMR CNRS-INPL 1,rue Grandville, BP 451 54001 Nancy, France*

[5]*Chemical Sciences Programme School of Distance Education Universiti Sains Malaysia Minden, 11800, USM, Penang Malaysia*

[6]*URAFPA, Equipe PB2P Faculté des Sciences et Techniques, Nancy Université BP 239, 54506 Vandoeuvre-lès-Nancy, France*

**Corresponding authors*

Glutathione peroxidases (GPXs) are a group of enzymes that regulate the levels of reactive oxygen species in cells and tissues, and protect them against oxidative damage. Contrary to most of their counterparts in animal cells, the higher plant GPX homologues identified so far possess cysteine instead of selenocysteine in their active site. Interestingly, the plant GPXs are not dependent on glutathione but rather on thioredoxin as their *in vitro* electron donor. We have determined the crystal structures of the reduced and oxidized form of *Populus trichocarpa*×*deltoides* GPX5 (PtGPX5), using a selenomethionine derivative. PtGPX5 exhibits an overall structure similar to that of the known animal GPXs. PtGPX5 crystallized in the assumed physiological dimeric form, displaying a pseudo ten-stranded β sheet core. Comparison of both redox structures indicates that a drastic conformational change is necessary to bring the two distant cysteine residues together to form an intramolecular disulfide bond. In addition, a computer model of a complex of PtGPX5 and its *in vitro* recycling partner thioredoxin *h*1 is proposed on the basis of the crystal packing of the oxidized form enzyme. A possible role of PtGPX5 as a heavy-metal sink is also discussed.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: glutathione peroxidase; X-ray structure; thioredoxin peroxidase; redox state; Molecular Dynamics

Abbreviations used: ROS, reactive oxygen species; GSH, reduced glutathione; PRX, peroxiredoxin; GPX, glutathione peroxidase; MAD, multiple anomalous dispersion.

E-mail addresses of the corresponding authors: Claude.Didierjean@lcm3b.uhp-nancy.fr; Catherine.Corbier@scbiol.uhp-nancy.fr

0022-2836/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.

Introduction

In aerobic organisms, reactive oxygen species (ROS) such as superoxide radicals, hydroxyl radicals or hydrogen peroxide are generated during the incomplete reduction of molecular oxygen to water,[1] following exposure to environmental factors[2] or during several metabolic processes (e.g. photosynthesis, respiration).[3] The accumulation of ROS can give rise to oxidative stress and lead to cell damage, mutation, or even death. The ROS challenge to land plants is made difficult because they are immobile in a constantly changing environment and, besides consuming O_2 during respiration, they also generate it during photosynthesis.[4] Plants have developed several non-enzymatic and enzymatic systems to withstand the oxidative damage caused by these ROS. While the major non-enzymatic antioxidant compounds include carotenoids, tocopherols, reduced glutathione (GSH) and ascorbate,[5] the enzymatic systems rely on superoxide dismutases, catalases, ascorbate peroxidases, peroxiredoxins (PRXs) and glutathione peroxidases (GPXs). In mammals, GPXs that form a group of enzymes with an important role in protecting cells against ROS, using GSH as the reducing substrate, have been studied extensively.[6] In plants, different isoforms of GPX are found to be expressed in various subcellular compartments. Recently, one of these GPXs was proposed to play a role in signal transduction during stress conditions.[7,8] *Saccharomyces cerevisiae* possesses three plant-type GPXs, GPX3 functioning as a redox sensor involved in gene activation.[9]

Typical plant enzymes display sequences similar to those of animal GPX enzymes, except that the SeCys in the catalytic site of most animal GPX is replaced by Cys in plant GPX, which in general show reduced catalytic efficiency.[10–15] Interestingly, recent studies have demonstrated that some plants[10,11,16] and yeast[17] GPXs can reduce peroxides, much more efficiently or sometimes exclusively, using TRX instead of GSH as an electron donor. In the unicellular parasite *Plasmodium falciparum*, a GPX-like protein closely related to plant GPXs is also specific for TRX as a reductant.[18] Most animal GPXs catalyse the reduction of hydrogen peroxide by consuming two GSH molecules to reduce the oxidized form of SeCys (selenenic acid) that is generated upon catalysis ($H_2O_2 + 2GSH \rightarrow GS\text{-}SG + 2H_2O$).[19–22] On the contrary, plant GPXs possess a second Cys residue to complete the reduction of ROS.[10] Navrot *et al.* (2006) have recently proposed a three-step reaction mechanism for plant GPXs similar to that used by peroxiredoxin Q (PRX Q),[10,23] with two Cys forming an intramolecular disulfide bridge in the oxidized state. These steps are: (i) a nucleophilic attack of the peroxidatic cysteine (Cys107, numbering in PtGPX3.2[10]) on the peroxide with the release of an alcohol and the concomitant formation of a sulfenic acid; (ii) an attack of the sulfenic acid by the resolving Cys (Cys155, numbering in PtGPX3.2[10]) and formation of an intramolecular disulfide bridge between Cys107 and Cys155; and (iii) a reduction of the disulfide bridge by TRX.[10] Plant GPXs have low substrate specificity, being capable to reduce a wide spectrum of peroxides, including more complex hydroperoxides.[10] On the basis of biochemical evidence, rather than on phylogenetic classification, Rouhier and Jacquot (2005) suggested that plant GPXs should be included in the large thioredoxin-dependent peroxidase family, also known as PRXs and constitute a fifth group of this family.[23] The levels of sequence identity between GPXs and PRXs are relatively low[23] (ranging from 1% with cytosolic peroxiredoxin type IIB to 16% with the PRX Q; see Rouhier and Jacquot[23] for the nomenclature of peroxiredoxin).

To date, there are six GPX crystal structures available, all of them from mammals. These crystal structures include one bovine GPX-1 (PDB code 1GP1)[24] structure and five human GPX structures (GPX-1, PDB code 2F8A; GPX-2, PDB code 2HE3; GPX-3, from Professor Rudolf Ladenstein;[25] GPX-4, PDB code 2GS3 and GPX-5, PDB code 2IY3). These GPXs are homotetrameric enzymes, except that human GPX-4 (an isoform that specifically detoxifies phospholipid hydroperoxides) is reported as a monomeric enzyme. Depending on the active site amino acid, these enzymes are generally classified into two groups: one contains SeCys at its active center (at least four out of five isozymes belong to selenium-containing GPX in mammals), the other contains a regular cysteine (the mammalian epididymis-specific isoform, GPX-5).[26] The non-selenium type GPX-5 is speculated to serve as the backup enzyme for the selenium-containing GPX to counteract ROS-mediated damage of sperm cells, since selenium deficiency causes male infertility and selenium-containing GPX is suspected to be a candidate for the defective molecule.[26,27] The activity of selenium-independent GPX is low and, hence, its contribution as a GSH-dependent peroxide scavenger is ambiguous.[27] Nevertheless, the overall structures of all mammal GPX monomers are similar, except that the environment around their active site catalytic residues is quite different.[19,24,25]

Black cottonwood (poplar) has been chosen as the model plant in this study. Very recently, an exhaustive overview was done on poplar GPXs and the annotation of the first release of the poplar genome indicates the presence of six complete GPX genes.[10] In this study, we describe the crystal structures of the GPX5 of *Populus trichocarpa* × *deltoides* (PtGPX5) in its oxidized (oPtGPX5) and reduced (rPtGPX5) states, revealing that PtGPX5 is a GPX-type thioredoxin-dependent peroxidase with an unique dimerization pattern mainly depending on hydrophobic contacts. The differences between the structures are discussed, with emphasis on the transition between the oxidized and the reduced forms, and on the formation of dimers†.

† The peroxidatic Cys44 and the resolving Cys92 in this study correspond to Cys107 and Cys155 of poplar GPX3 (PtGPX3.2) in the previous study.[10] The denomination of plant GPXs does not correspond to the denomination of mammalian GPXs, for example, PtGPX5 in current study does not correspond to the human epididymis GPX5.

Results

Quality of the model

PtGPX5 is a 170 residue protein with a monomeric molecular mass of 19.36 kDa and exists as a dimer in solution.[10] The reduced form of the enzyme (rPtGPX5) was crystallized as two dimers (dimer 1, subunits A and B; dimer 2, subunits C and D) in the asymmetric unit and this structure was solved from rhombohedric selenium-substituted protein crystal dataset (2.7 Å) using the single wavelength anomalous diffraction method. The structure was completed using the high-resolution data (1.95 Å) collected from the native rPtGPX5. The final model of rPtGPX5 structure, with R and R_{free} factors of 17.85% and 20.04%, respectively, contains 160 or 161 amino acid residues corresponding to residues from Glu10 to Ile169 in monomers B, C, and D and from Glu10 to Ser170 in monomer A. The overall structure is well defined in the electron density map except for the missing portions mentioned above. Very few residues have poorly defined side-chains and they are usually located in flexible loops on the protein surface. In addition, there are 32 cadmium atoms, eight acetate anions and 668 water molecules found in the asymmetric unit of the reduced form crystal. The rms overlaps of all the subunits fall in the range from 0.15 Å–0.33 Å based on alignments of C^{α} positions, thus suggesting that all subunits are equivalent. The most ordered with well-defined electron density subunit, as indicated by the average B-value (Table 1), is subunit A that participates most highly in the subunit–subunit interactions. Hence, unless indicated otherwise, our discussion will focus on this subunit.

The oxidized form of the enzyme (oPtGPX5) was crystallized as a dimer in the asymmetric unit. The final model of oPtGPX5 contains residues from Ser6 to Ile169 in each subunit with an R-value of 21.08% (R_{free} = 28.03%) at 2.45 Å resolution. Compared to

Table 1. Data collection, phasing and refinement statistics for the [SeMet] rPtGPX5, the wild-type native rPtGPX5 and the wild-type native oPtGPX5 crystals

Data Set	[SeMet] reduced (peak)	Reduced native	Oxidized native
A. *Data collection and processing statistics*			
Data collection site	BM30A ESRF-Grenoble	X11 DESY-Hamburg	BM30A ESRF-Grenoble
Wavelength (Å)	0.98004	0.81560	0.80630
Space group	*R*3	*R*3	$P3_121$
Unit cell dimensions			
a (Å)	222.19	221.67	71.58
b (Å)	222.19	221.67	71.58
c (Å)	48.36	48.14	117.82
Content of asymmetric subunit	Four subunits	Four subunits	Two subunits
Resolution range (Å)[a]	50.00–2.70 (2.85–2.70)	50.00–1.95 (2.02–1.95)	50.00–2.45 (2.49–2.45)
Redundancy	6.50 (6.50)	5.04 (4.52)	7.60 (7.79)
Completeness (%)	100.0 (100.0)	99.8 (99.0)	99.8 (100.0)
I/σ (I)[a]	21.30 (7.70)	11.48 (2.20)	17.24 (2.91)
R_{merge}[a,b]	0.079 (0.300)	0.066 (0.409)	0.062 (0.335)
Figure of merit (MLPHARE/DM)	0.342/0.665		
B. *Refinement statistics*			
Resolution range (Å)		30.00–2.00	50.00–2.45
Reflections used		59452	12907
R_{cryst}[c]		17.85	21.08
R_{free}[d]		20.04	28.03
Number of atoms			
(Protein/water/Cd^{2+}/Ca^{2+}/acetate ions)		5187/668/32/0/8	2650/224/0/5/0
Average B-value for individual chain (Å^2)		16.32/25.41/21.08/29.68 (subunit A/B/C/D)	39.15/53.58 (subunit A/B)
Mean B-factor			
Main chain (Å^2)		22.10	46.24
Side-chain (Å^2)		24.12	46.49
Water (Å^2)		33.19	42.08
All (Å^2)		24.40	46.07
rms deviation from ideal geometry			
Bond lengths (Å)		0.005	0.006
Bond angles (°)		1.3	1.3
Dihedral angles (°)		22.9	23.9
Improper angles (°)		0.67	0.72
Ramachandran plot			
In most favored regions (%)		90.5	81.4
In additionally allowed regions (%)		9.3	16.2
In generously allowed regions (%)		0.2	2.4

[a] The values in parentheses are for the highest resolution bin.
[b] $R_{merge} = \sum_i | I_i - \langle I \rangle | / \sum | \langle I \rangle |$, where I is the intensity for the ith measurement of an equivalent reflection with the indices h, k, l.
[c] $R_{cryst} = \sum | F_o - F_c | / \sum F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.
[d] The R_{free} value was calculated from 10% of all data that were not used in the refinement.

the structure of the reduced form, oPtGPX5 structure has more poorly defined regions. Most of the residues located in a long flexible loop (residues 73–100) are found in the generously allowed regions of the Ramachandran plot (Table 1). The electron density of the region between residues 77 to 84 can hardly be seen. Among the subunits in the asymmetric unit, subunit A shows the clearest densities for the main-chain atoms. Thus, we have adopted the conformation of the flexible loop of subunit A for the loop positioning of the subunit B as this monomer has more poorly defined regions. In the model of the oxidized form, we have assigned five calcium atoms per asymmetric unit as calcium chloride was present in the crystallization condition.

Overall fold and secondary structure of PtGPX5 subunit

The subunit structure shows a thioredoxin fold,[28] which consists mainly of a central twisted β-sheet and several flanking α-helices, with few additional secondary structure elements, similar to that described for the GPX[24,25] and PRX[29–32] families. The reduced form (rPtGPX5) subunit structure consists of four α helices and seven β strands (Figure 1). Helices α1, α2 and α4 are located on one side of the 45° twisted central β sheet (β3, β2, β1, β4 and β5) and α3 is on the other side. Helices α1 and α4 are nearly parallel and their axes are roughly parallel with the central β strands. The two other α helices, α2 and α3, are oriented perpendicular to each other and to the helices α1 and α4 (Figure 1(b)). The N-terminal end exhibits an extra chain folding into a β-hairpin designated as βN1 and βN2, respectively.

Superimposition of both reduced and oxidized form of PtGPX5 subunit structures shows a rms deviation of 4.02 Å (based on alignments of 160 C^{α} positions) (Figure 2). This high rms value is due to the large local differences observed in α1 and α2 helices between both structures. The beginning of α1 helix in the oPtGPX5, where the peroxidatic Cys44 is

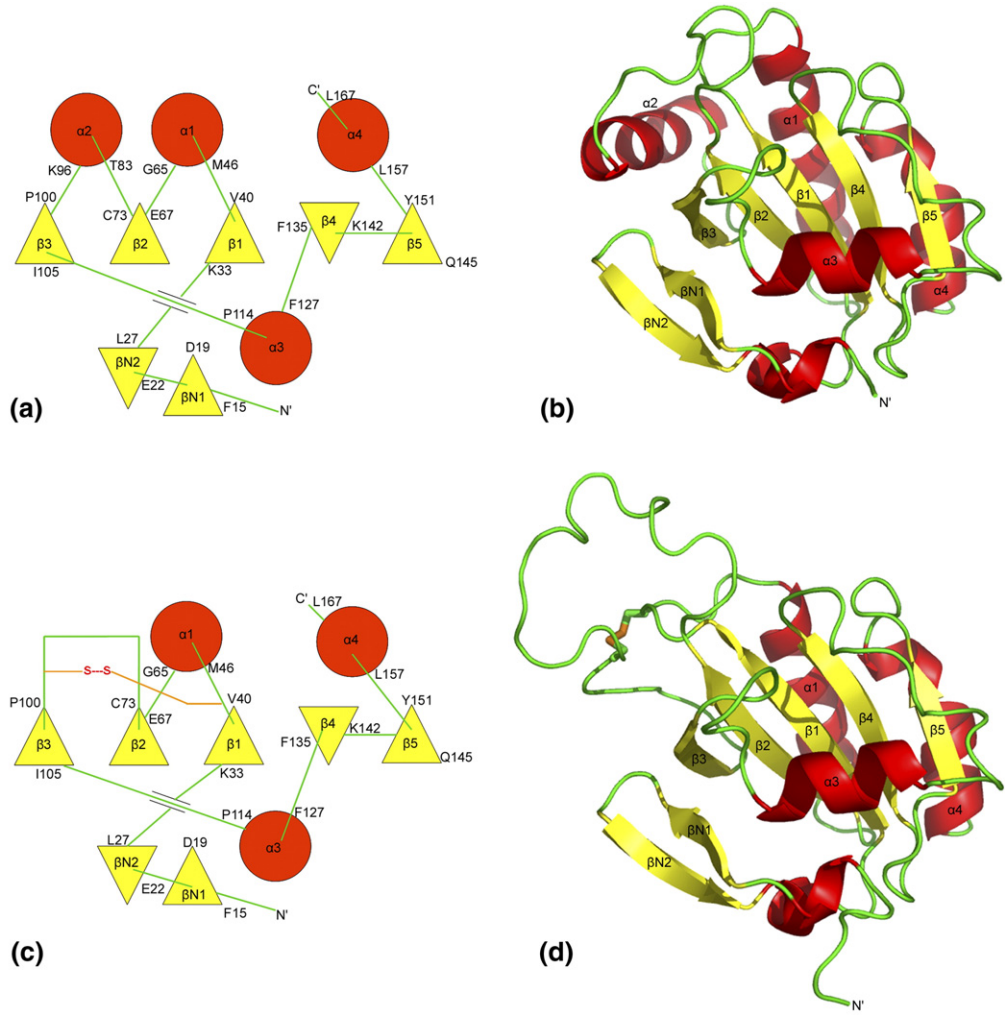


Figure 1. (a) and (c) Topological diagrams and (b) and (d) cartoon representation showing the overall structures of the rPtGPX5 (top two panels) and the oPtGPX5 (bottom two panels). All α helices are shown in red while β strands are in yellow. In the topological diagrams, the helices are represented as circles and the β strands as triangles. The beginning and the ending residues with numbering of each secondary structural element are labelled. The intramolecular disulfide bond between Cys44 and Cys92 is highlighted.

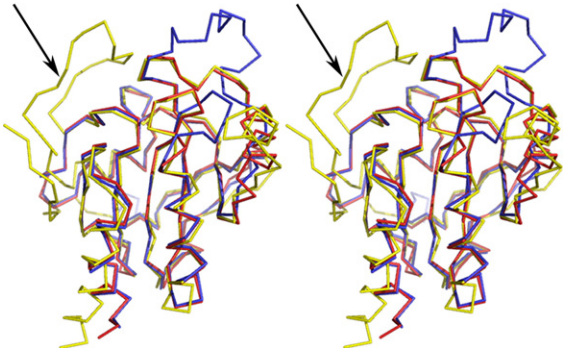


Figure 2. Stereoscopic view of superimpositions between the C^{α} traces of the rPtGPX5 (red), the oPtGPX5 (blue) and a typical mammalian GPX, the human plasma GPX-3 (yellow).[25] The major difference between PtGPX5 and the classical mammalian GPX is the absence of the oligomerization loop (indicated by the arrow) in the PtGPX5.

located, undergoes a local unwinding and rearrangement. The local flipping of this small loop towards the direction of helix α2, however, is rather small as compared to the complete unwinding of the α2 helix, where the resolving Cys92 is situated. The latter motion results in a flexible extra long loop (residues 73–100), connecting β2 and β3 strands. These conformational changes lead to the loss of 58 interactions, which were involved in the stabilization of the reduced form enzyme, whereas 28 interactions are newly created to stabilize the oxidized form. If the two structures are superimposed, excluding the conformational changed regions, the rms deviation is only 0.525 Å (based on alignments of only 135 C^{α} positions) (Figure 2). Indeed, the rest of both rPtGPX5 and oPtGPX5 structures superimpose well.

Oligomerization state of PtGPX5

One of the inter-subunit contact types observed within the crystal lattice of PtGPX5 likely represents the subunit interaction that stabilizes the dimeric enzyme in solution. Indeed, from two different experiments, using different crystallization conditions, we observed the same dimerization pattern in the crystal structures regardless of their redox states. The dimer interface areas in both reduced and oxidized structures range from 822 $Å^2$ to 847 $Å^2$, involving 60% of non-polar atoms and 40% of polar atoms. In this study, we highlight a new mode of dimerization pattern that differs from previously reported GPXs and PRXs. The dimerization interface localizes at the C-terminal region of PtGPX5. The main dimer building block consists of two antiparallel β sheet ($\beta5_A$-$\beta5_B$) which are positioned side-by-side in a head-to-tail manner, related by local 2-fold symmetry axis. Formation of the PtGPX5 dimer results in a central pseudo-10-pleated β sheet core through the dimer (Figure 3). However, although the two β5 strands are adjacent to each other, the -CO groups of one $\beta5_A$ strand and the -NH groups on the adjacent $\beta5_B$ strand are 6 Å apart. In fact, the dimer interface is stabilized by several hydrogen bonds involving polar side-chains of residues located in the loop that connects α3 to β4, β5 and α4 (summarized

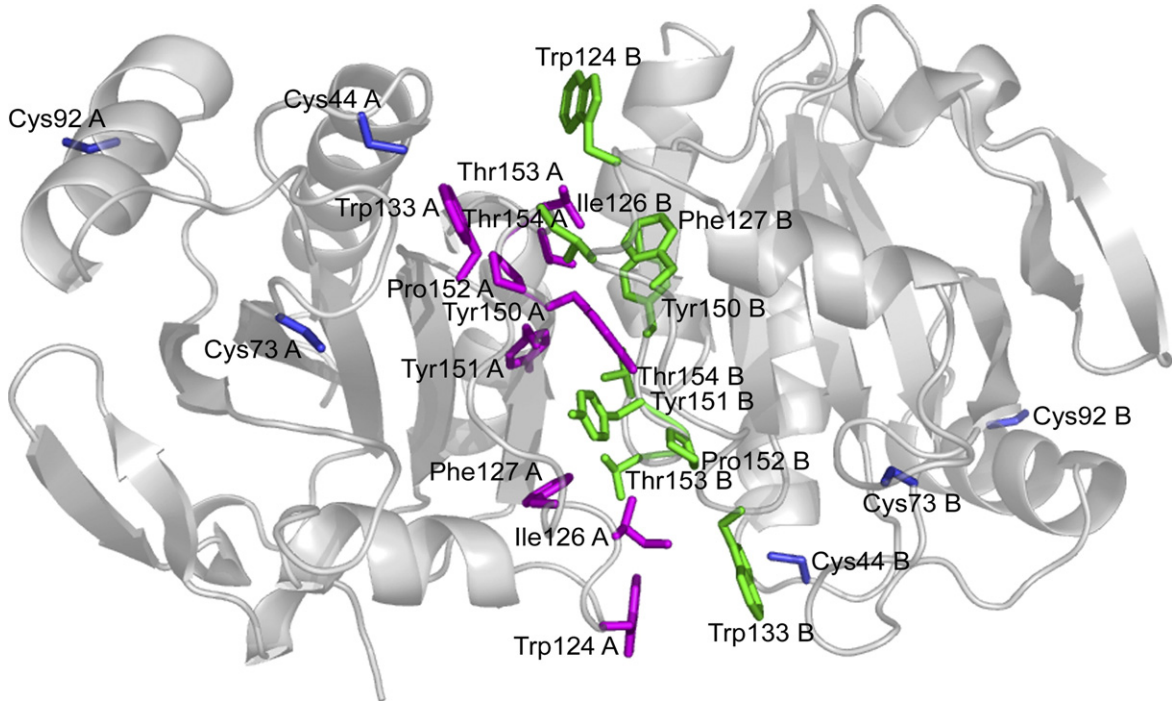


Figure 3. Cartoon representation of the dimerization interface of the rPtGPX5 (monomers A and B). This Figure highlights the dimerization interface involving the hydrophobic and aromatic residues (shown as sticks). The side-chains of monomer A and monomer B, which are involved in the dimerization, are coloured magenta and green, respectively. The locations of the peroxidatic Cys44, the highly conserved Cys73 and the resolving Cys92 are coloured blue.

Table 2. Hydrogen bonding interactions at the dimerization interface

Monomer A	Monomer B	Distance (Å) rPtGPX5	Distance (Å) oPtGPX5
Loop between α3 and β4 (124–133)			
Asp 129 $O_{\delta1}$	Gln 132 $N_{\varepsilon2}$	2.67	2.44
Asp 129 $O_{\delta1}$	Tyr 151 OH	2.82	3.00
Gln 132 $N_{\varepsilon2}$	Asp 129 $O_{\delta2}$	2.71	2.82
Gln 132 $N_{\varepsilon2}$	Tyr 151 OH	2.84	2.82
β5 (147–155)			
Arg 149 NH_1	Tyr 151 OH	2.94	(3.73)
Arg 149 NH_2	Tyr 151 OH	3.17	(3.55)
Tyr 151 OH	Asp 129 $O_{\delta2}$	2.65	3.05
Tyr 151 OH	Arg 149 NH_1	3.14	2.98
Tyr 151 OH	Gln 132 $N_{\varepsilon2}$	2.89	2.85
Ser 155 O_γ	Arg 161 NH_2	3.28	3.37
α4 (156–166)			
Ser 158 O_γ	Asp 162 $O_{\delta2}$	2.81	3.05
Ser 158 O_γ	Asp 162 $O_{\delta1}$	2.93	2.41
Ser 158 O_γ	Arg 161 NH_1	2.73	3.03
Arg 161 NH_1	Ser 155 O_γ	2.66	3.21
Arg 161 NH_1	Ser 158 O_γ	2.92	2.72
Asp 162 $O_{\delta1}$	Ser 158 O_γ	2.83	2.52
Asp 162 $O_{\delta2}$	Ser 158 O_γ	2.87	3.19

Distance >3.5 Å is not considered as a hydrogen bonding interaction and is indicated for comparison purposes (marked in parentheses).

in Table 2) and van der Waals interactions which involve the participation of a hydrophobic and aromatic cluster (Trp124, Ile126, Phe127, Trp133, Tyr150, Tyr151, Pro152, C^γ of Thr153 and C^γ of Thr154).

Environment of the peroxidatic Cys in the reduced form of the enzyme (rPtGPX5)

The active site residues in the neighbourhood of the peroxidatic Cys44 are well-defined in ($3F_o$–$2F_c$) electron density map. The active site pocket is formed by residues in the long loop between α3 and β4 (Phe127–Phe135), the loops between β2 and α2 (Cys73–Thr83), β5 and α4 (Tyr151–Leu157) and β1 and α1 (Val40–Met46), while the assumed peroxidatic Cys44 of PtGPX5[10,23] is positioned on the seven residue-long loop just before the α1 helix (near the interface of a dimer of the asymmetric unit). The side-chain of Cys44 is pointing towards the interior of the protein, surrounded mainly by non-polar groups (Met46, Phe76, Phe135, Trp133, Pro152 and the aliphatic chain of Lys43), while the remaining residues are hydrophilic (Glu79, Asn134) (Figure 4(a)). The peroxidatic Cys44 remains accessible to solvent because of its strategic location near the surface of the subunit.

The active-site cleft appears to be occupied by a cadmium ion (Cd^{2+}) that originated from the crystallization solution, ligated by three residues and well coordinated with three water molecules. The coordination of this cadmium ion displays a distorted octahedron, ligated with Cys44-S^γ, Glu79-$O^{\varepsilon1}$, Trp133-$N^{\varepsilon1}$ and three water molecules (see Table 3 and Figure 4(a)). The positions of Cd^{2+}-surrounding ligands show no significant geometry changes compared with the active sites described for mammalian GPXs (which do not possess any cadmium ion).[24,25]

Further, the active site of rPtGPX5 is shielded by the adjacent subunit of the dimer (residues Gly122–Ile126 at the C terminus of the α3), which suggests that the dimer in PtGPX5 and probably in all plant GPXs is of catalytic significance. In the structure of the reduced form, the peroxidatic cysteine (Cys44) of a dimer are ~38 Å apart and occupy opposite sites (Figure 3). In addition to the catalytic site Cys44, there are two other cysteine residues present in each PtGPX5 subunit. Cys73 is located on the loop that connects β2 and α2, about 11 Å apart from Cys44, while Cys92 is located on α2 and is 21 Å away from the peroxidatic residue.

Environment of the peroxidatic Cys in the oxidized form of the enzyme (oPtGPX5)

In the structure of the oxidized form (oPtGPX5), an intramolecular disulfide bond between the peroxidatic Cys44 and the resolving Cys92 is clearly visible. The formation of the disulfide bond causes two regions in oPtGPX5 to rotate toward each other, decreasing the C^α–C^α distance between residues 44 and 92 by 12.1 Å relative to the reduced molecule. The torsion angles for the bonds constituting the disulfide bond between Cys44 and Cys92 are as follows: Cys44, $x_1 = -58°$, $x_2 = -60°$, $x_3 = +96°$, $x_{2'} = +84°$, $x_{1'} = -165°$, Cys92.

Examining the surroundings of the disulfide bond in the oPtGPX5 reveals that one side is freely exposed to the solvent while the other side faces an aromatic residue (Phe95). Close to the disulfide bond, there are several other residues, including Met46, Thr93 and Arg94 (Figure 4(b)). In addition, a spoon-shaped, surface-exposed loop connecting β2 and β3 (residues 77–100) is observed facing the oxidized active site. This long loop results from the unwinding of α2 helix mentioned earlier. The conformation of this loop is an obtuse angle between protein core and the surface loop (~120°).

Proposed model of the oPtGPX5-TRX *h*1 complex

As a result of crystal packing, a Trp residue (Trp124′) from a crystallographically symmetric neighbouring molecule is observed in the vicinity of the intramolecular disulfide bond (Figure 4(b)). As mentioned before, oPtGPX5 is regenerated by TRX,[10] and the Trp residue at the catalytic site of TRX plays an important role in TRX-substrate recognition (see Discussion). It is tempting to speculate that a similar situation at the active site may occur when the oPtGPX5 binds the reducing redox partner, TRX. Therefore, we were prompted to model the complex that enables us to visualize the possible interaction between the ROS scavenger (oPtGPX5) and its *in vitro* recycling partner (TRX *h*1,[33] PDB code 1TI3). Several molecular dynamics

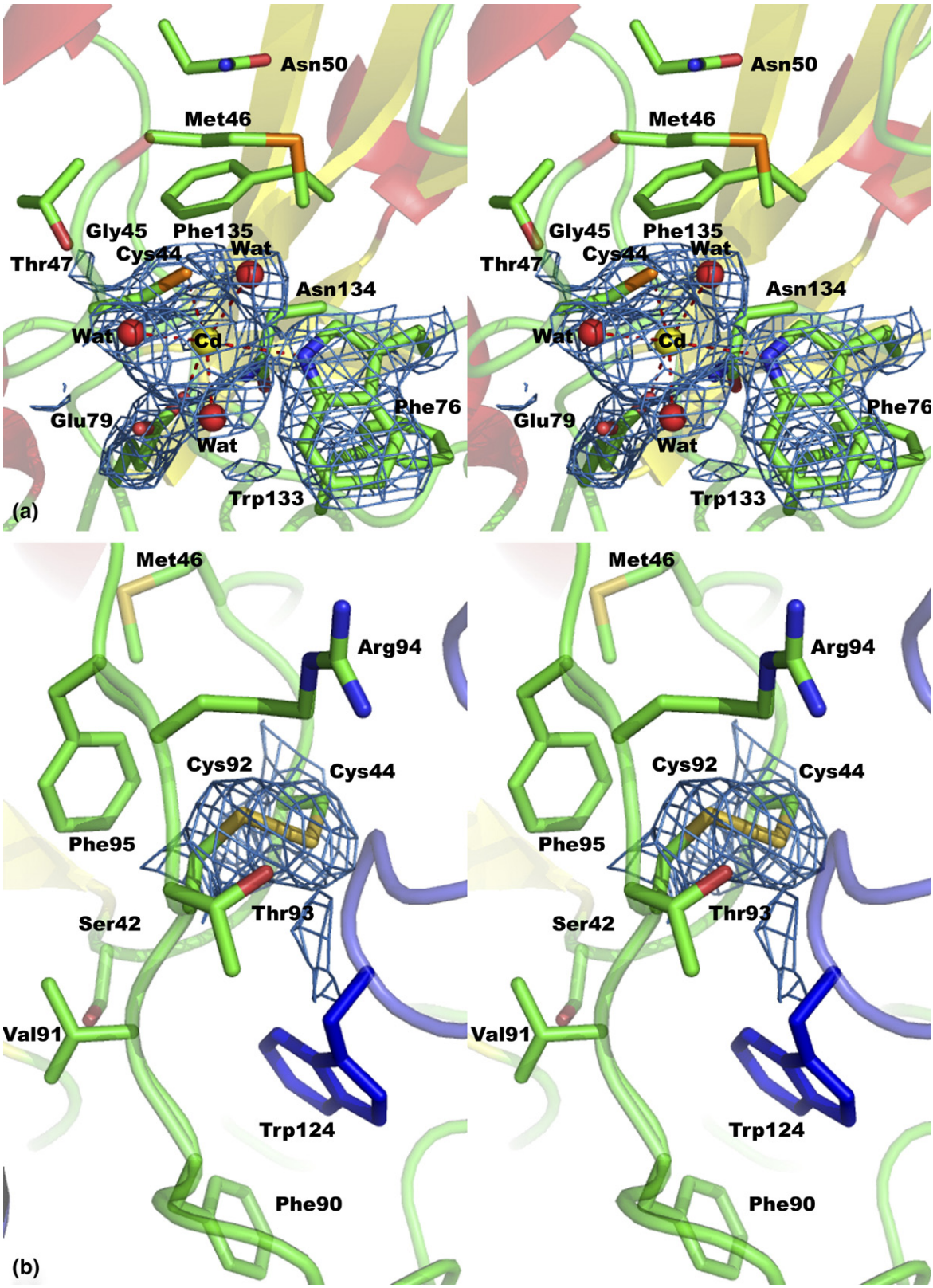


Figure 4. Stereoview details of the active site cleft of both (a) rPtGPX5 and (b) oPtGPX5 with final $3F_o$–$2F_c$ electron densities (1.2 σ level) covering chosen residues for clarity. Relevant residues are labelled. In the rPtGPX5 active site, a cadmium ion is coordinated to the peroxidatic Cys44, Trp133, Glu79 and three water molecules, forming an octahedral geometry. In the oPtGPX5 active site, an intramolecular disulfide bond is formed between Cys44 and Cys92. All the residues shown here (in sticks, coloured according to atom type) are from monomer A. (b) The Trp124 residue, coloured blue, stacks from a symmetry-related subunit.

Table 3. Coordination geometry of the active-site cadmium ion

Bond lengths (Å)		Bond angles (°)	
Cys44 S^{γ} - Cd	2.59	Cys44 S^{γ} - Cd - Glu79 $O^{\varepsilon 1}$	90
Glu79 $O^{\varepsilon 1}$ - Cd	2.44	Cys44 S^{γ} - Cd - Trp133 $N^{\varepsilon 1}$	105
Trp133 $N^{\varepsilon 1}$ - Cd	2.75	Cys44 S^{γ} - Cd - Wat318	167
Wat318 - Cd	2.65	Cys44 S^{γ} - Cd - Wat482	96
Wat482 - Cd	2.71	Cys44 S^{γ} - Cd - Wat626	103
Wat626 - Cd	2.81	Glu79 $O^{\varepsilon 1}$ - Cd - Trp133 $N^{\varepsilon 1}$	92
		Glu79 $O^{\varepsilon 1}$ - Cd - Wat318	93
		Glu79 $O^{\varepsilon 1}$ - Cd - Wat482	171
		Glu79 $O^{\varepsilon 1}$ - Cd - Wat626	111
		Trp133 $N_{\varepsilon 1}$ - Cd - Wat318	88
		Trp133 $N^{\varepsilon 1}$ - Cd - Wat482	79
		Trp133 $N^{\varepsilon 1}$ - Cd - Wat626	143
		Wat318 - Cd - Wat482	84
		Wat318 - Cd - Wat626	64
		Wat482 - Cd - Wat626	75

(MD) processes were performed by placing the TRX molecule next to oPtGPX5, in such a position that Trp124′ of the symmetry-related oPtGPX5 is replaced by Trp37 of TRX *h*1 (see Materials and Methods). Two complex models can be built with a transient intermolecular disulfide bond formed between the catalytic Cys38 of poplar TRX *h*1 and either the peroxidatic Cys44 (complex 1) or the resolving Cys92 (complex 2) of oPtGPX5. The complex 1 exhibits a plus left-handed hook disulfide bond, while the complex 2 possesses a minus right-handed hook disulfide bond. In terms of calculated free energy, however, complex 2 reveals a more stable complex as compared to complex 1.

In both complex models, several edge-to-face aromatic interactions were essential at the protein–protein interfaces involving especially the Trp37 of TRX *h*1 (Figure 5). The $N^{\varepsilon 1}$ atom of this residue is hydrogen bonded with Gln86-O in complex 1 or with Asn84-O in complex 2. Otherwise, this residue is always located between a Phe residue and a Pro residue. In both cases, the Phe ring interacts perpendicularly, with its edge pointing towards the Trp face, while Pro tends to make CH–π interaction on the Trp face through major contributions from its C^{γ} and C^{δ}. In complex 1, both of these Phe and Pro residues are from oPtGPX5 (Phe90 and Pro81), whereas in complex 2, Phe90 of oPtGPX5 and Pro39 of TRX *h*1 are involved. Nevertheless, all these residues are conserved in plant GPX homologues and TRX *h*1 subgroup. Details of the hydrogen bonding interactions involved in both complexes are summarized in Table 4.

Comparison to homologous structures

PtGPX5 shares a high level of identity with various plant GPXs, ranging from 72% to 91%. However, none of these plant GPXs has any structural reference. To date, six structures of mammalian GPXs have been solved. They are the classical bovine erythrocyte GPX-1 (PDB code 1GP1),[24] the SeCys to glycine mutant of human GPX-1 (PDB code 2F8A), the human GPX-2 (PDB code 2HE3), the human plasma GPX-3 (from Professor Rudolf Ladenstein),[25] the SeCys to glycine mutant of human GPX-4 (PDB code 2GS3) and the human GPX-5 (PDB code 2IY3). All the classical mammalian GPX structures known so far are in the tetrameric form, except for human GPX-4, which is a monomeric enzyme (PDB code 2GS3). GPX-5 from poplar presents as a dimer with interactions that differ from those existing in other known GPX or 2-Cys PRX structures (see Discussion).

The superimposition of rPtGPX5 subunit with the other mammal GPXs leads to rms deviations varying from 0.72 Å with human GPX-4 (based on alignments of 152 C^{α} positions) to 1.08 Å with human plasma GPX-3 (based on alignments of 144

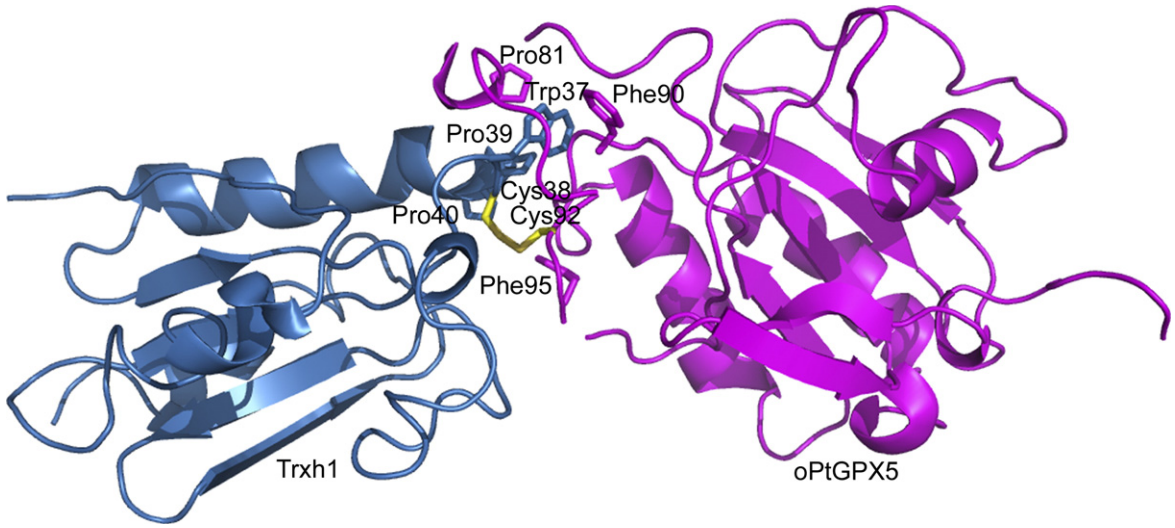


Figure 5. Molecular model of the oPtGPX5-TRX *h*1 complex. The modelled complex 2 (see Results) structure shown in cartoon representation displays an intermolecular disulfide bond (yellow) between the catalytic Cys38 of TRX *h*1 and the resolving Cys92 of oPtGPX5. Protein–protein interactions between the oPtGPX5 (magenta) and the TRX *h*1 (blue) involving several aromatic residues at the interface are shown as sticks and are coloured according to respective molecules.

Table 4. Hydrogen bonding interactions between oPtGPX5 and TRX *h*1 in both complex 1 and complex 2 models

Molecule	PtGPX5	TRX h1	*d* (Å)
A. *Complex 1*	Cys44 N	Trp37 O	3.24
	Gln86 $N^{\varepsilon 2}$	Asp64 $O^{\delta 2}$	3.25
	Gln86 $N^{\varepsilon 2}$	Asp66 $O^{\delta 2}$	2.88
	Gln86 O	Trp37 $N^{\varepsilon 1}$	3.22
	Thr88 $O^{\gamma 1}$	Ser36 O^{γ}	3.29
	Thr88 O	Lys67 N^{ζ}	2.81
B. *Complex 2*	Glu80 O	Lys42 N^{ζ}	2.89
	Thr83 $O^{\gamma 1}$	Ser36 O	3.00
	Asn84 O	Trp37 $N^{\varepsilon 1}$	3.03
	Gln86 $N^{\varepsilon 2}$	Ser36 O^{γ}	2.76
	Arg94 NH_1	Val77 O	2.71

Complex 1; disulfide bond between Cys44 of oPtGPX5 and Cys38 of TRX *h*1. Complex 2; disulfide bond between Cys92 of oPtGPX5 and Cys38 of TRX *h*1.

C^{α} positions). Structure-based alignments demonstrated that the main differences between PtGPX5 and other mammalian GPXs are the absence of an extended N terminus, the deletion of the region coding for the classical oligomerization loop, and the shortening of helix α2 (by five residues) (Figure 6). In terms of the N terminus, PtGPX5 does not possess the extra N-terminal residues (the region before the N-terminal β-hairpin) observed in the human GPX-3 enzyme structure. Likewise, the other mammalian GPX-1, GPX-2, GPX-4 and GPX-5 also lack this extra feature. In addition, a large part of the classical mammalian sequence (positions 125–142 in bovine GPX-1 structure numbering) is deleted in PtGPX5 and in human phospholipid hydroperoxide GPX-4 (PDB code 2GS3). This missing part corresponds precisely to the subunit interaction sites of the tetrameric mammalian GPXs.

Another important difference is that PtGPX5 displays an overall negatively charged protein surface, while the rest of the mammalian GPXs show more evenly distributed charged protein surfaces. Sequence analysis also demonstrated that PtGPX5 contains 15% of negatively charged residues (Asp and Glu), while the mammalian GPXs contain 11%. Remarkably, the α2 helix in PtGPX5, where the resolving Cys92 is situated, exhibits a highly negatively charged region (Figure 7(d)) as compared to the corresponding regions in all mammalian GPXs. Further, active site residues of rPtGPX5 and their geometries are similar to those described for the most classical GPX enzymes. However, a peroxidatic Cys (Cys44) and a Glu (Glu79) are found in the active site pocket of the present structure instead of SeCys and Gln (Gln83; numbering in human plasma GPX-3 structure) residues in classical mammalian GPX structures. These differences may suggest enzyme adaptation to substrate binding and specificity or to the adapted redox partner (see Discussion).

Cadmium-binding sites

There are 32 cadmium ions present in the asymmetric unit of PtGPX5 in its reduced form (Figure 8(a)). We have distinguished four potential cadmium-binding sites per monomer (16 sites per asymmetric unit) on the basis of the following three prerequisites: (i) the cadmium-binding site is present in the four independent subunits; (ii) at least two non-water ligands are involved in ion-binding; and (iii) the ion-binding site is independent of the crystal packing. The first cadmium-binding site is situated at the active site cavity, displaying the coordination geometry of an octahedron as described before (see the previous section). The second binding site located at the dimer interface also shows the typical octahedron geometry by binding to six ligands, including Asp148-$O^{\delta 1}$ of one subunit, Thr153-O of the adjacent subunit and four water molecules (Figure 8(b)). Anchored to the subunit interface, the third cadmium-binding site (Figure 8(c)) is hepta-coordinated by two bidentate residues (Asp85 and Asp89) and water molecules. A similar coordination geometry is found for the fourth cadmium-binding site (Figure 8(d)) at the β-hairpin loop of each subunit, except that the two bidentate residues involved are Glu22 and Asp103.

Discussion

Re-evaluation of PtGPX5 denomination

Sequence comparison revealed that PtGPX5 and other plant homologues share relatively high levels

Figure 6. Multiple sequence alignment of 14 representative glutathione peroxidase proteins from mammals and plants (including the PtGPX5) as compared to the available secondary structure elements from the bovine erythrocyte GPX-1 (PDB code 1GP1), the human GPX-1 (PDB code 2F8A), the human GPX-2 (PDB code 2HE3), the human plasma GPX-3 (from Professor Rudolf Ladenstein), the human GPX-4 (PDB code 2GS3) and the human GPX-5 (PDB code 2IY3) as references. Both reduced and oxidized states secondary structures of PtGPX5 are shown in comparison to the mammalian GPX structures. Residues that are identical are boxed in red and displayed in white letters, while homolog residues are framed in blue and shown in red letters. For secondary structure representation, α and 3_{10} helices are displayed in coils with α and η labels, respectively, β strands are shown as arrows labelled β, and strict α and β turns are marked TT. Contacts between protein residues and supported hetero-compounds are shown as black stars. Sequence comparison between plant GPXs and mammalian GPXs (except mammal GPX-4) show a large gap in plant GPX sequences, while this region (framed in red) codes for an oligomerization loop in animal GPXs. The accession numbers for other plant GPX sequences displayed in the final alignment are as follows: *Arabidopsis thaliana*, ATGPX8: NP_564813; *Medicago truncatula*, Medicago: ABE92132; *Solanum lycopersicum*, LE2: AI898013; *Vitis vinifera*, VV: CB978870; *Helianthus annuus*, Ha1: CAA74775; *Mesembryanthemum crystallinum*, Mesembryanthemum: CAC83045; *Citrus sinensis*, CIT-SAP: CAA47018. A motif from residue Glu80 to Lys96 (numbering in PtGPX5) in plant GPXs is proposed (framed in dark blue) to be the recognition site for thioredoxin. The Figure was prepared using ESPript.[54]

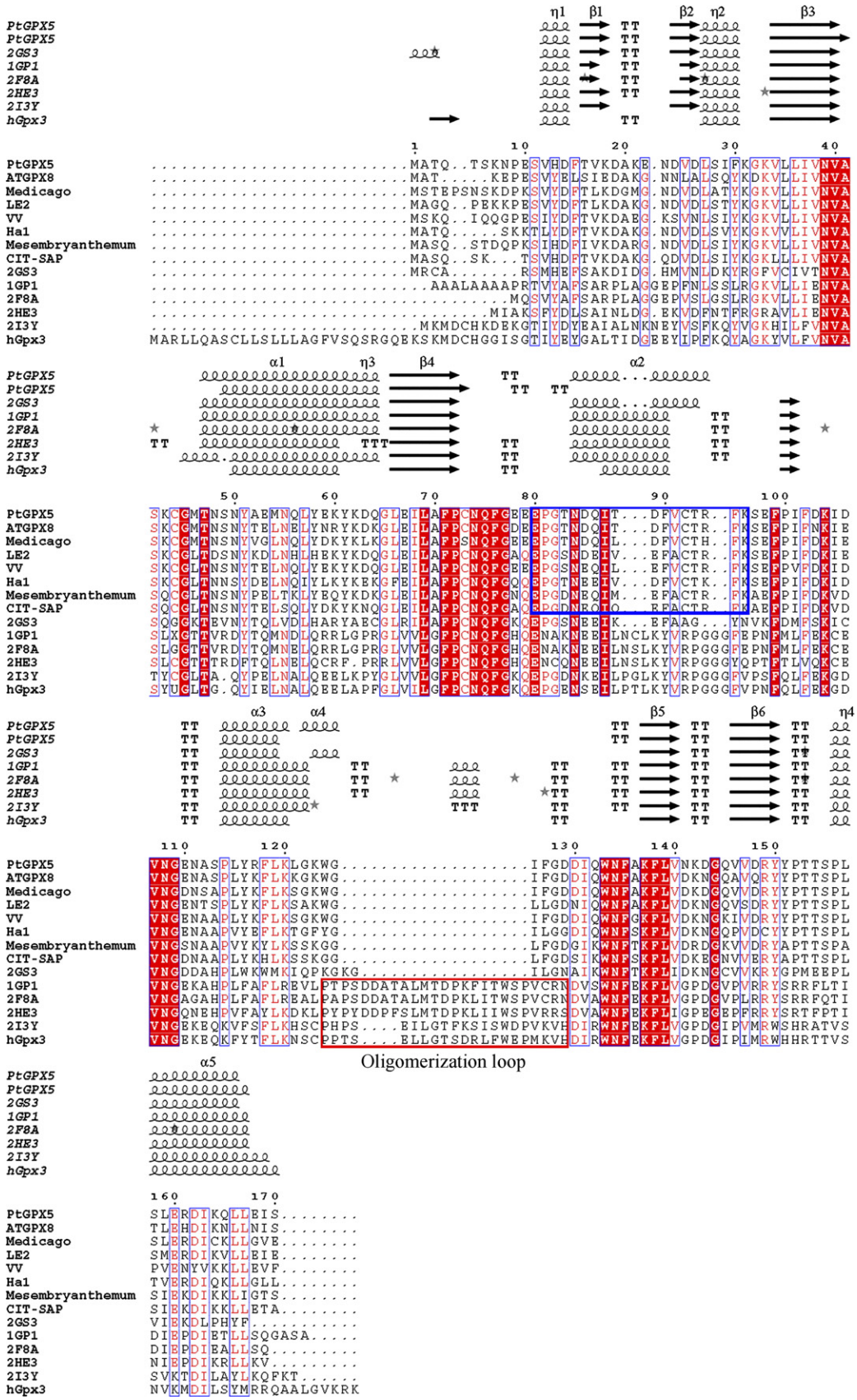


Figure 6 (*legend on previous page*)

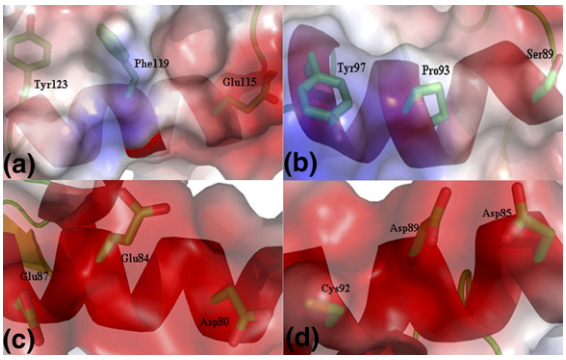


Figure 7. Local surface charges of α2 helix of (d) rPtGPX5 and its corresponding helices of other mammalian GPX structures ((a) the monomeric human GPX-4, PDB code 2GS3; (b) the classical tetrameric human GPX-3, human plasma GPX-3). The corresponding α2 helix in calmodulin (PDB code 1CLL) that exhibits a helix–coil transition is shown in (c). Negatively and positively charged regions are shown in red and blue, respectively, while the rest are shown in grey. Three residues, situated at each helix turn and facing the same side of the helix, are displayed as sticks.

of sequence similarities with mammal GPXs. Contrary to the majority of mammalian GPXs, which involve only a SeCys residue in the catalytic mechanism, the higher plant GPX homologues identified so far possess two cysteine residues (peroxidatic Cys, C_P and resolving Cys, C_R) that take part in the reduction of peroxide. These mammalian GPXs use GSH for their regeneration of the SeCys catalytic residue. However, it has been suggested that GPX-type enzymes are not completely specific for GSH but may react also with CxxC motifs present in TRX or tryparedoxins.[34] Indeed, most of the plant GPXs probably use reduced TRX for their regeneration,[10–12,16,17] and do not react with GSH or glutaredoxin, as demonstrated recently for PtGPX5.[10] The reaction mechanism of PtGPX5 was shown to be homologous to that of atypical 2-Cys PRXs, in which (i) the reduction of the peroxide is accompanied by the formation of a sulfenic acid on the catalytic cysteine, (ii) a second Cys residue forms an intramolecular disulfide bridge with the C_P and acts as a C_R in being indispensable for the reduction by TRX.[10] Thus, in a classification based on existing biochemical evidence rather than on phylogenetic linkage, the plant GPXs constitute a fifth group of plant PRXs.[23] On the other hand, a reclassification of the group of enzymes in which SeCys is replaced by Cys in GPX homologs is being suggested on the basis of biochemical properties, as GPX-type TRX peroxidases.[35] We describe here for the first time in plants, two structures of PtGPX5 in two different redox states that give insights for the denomination of this enzyme from a structural point of view.

GPX and PRX present the same overall fold, namely the TRX fold, which consists of a central β sheet and flanking α helices. However, with respect to the nature of their redox centers, the SeCys residue in mammal GPXs is at hydrogen bonding distance from a Trp and a Gln residue[25] (active site

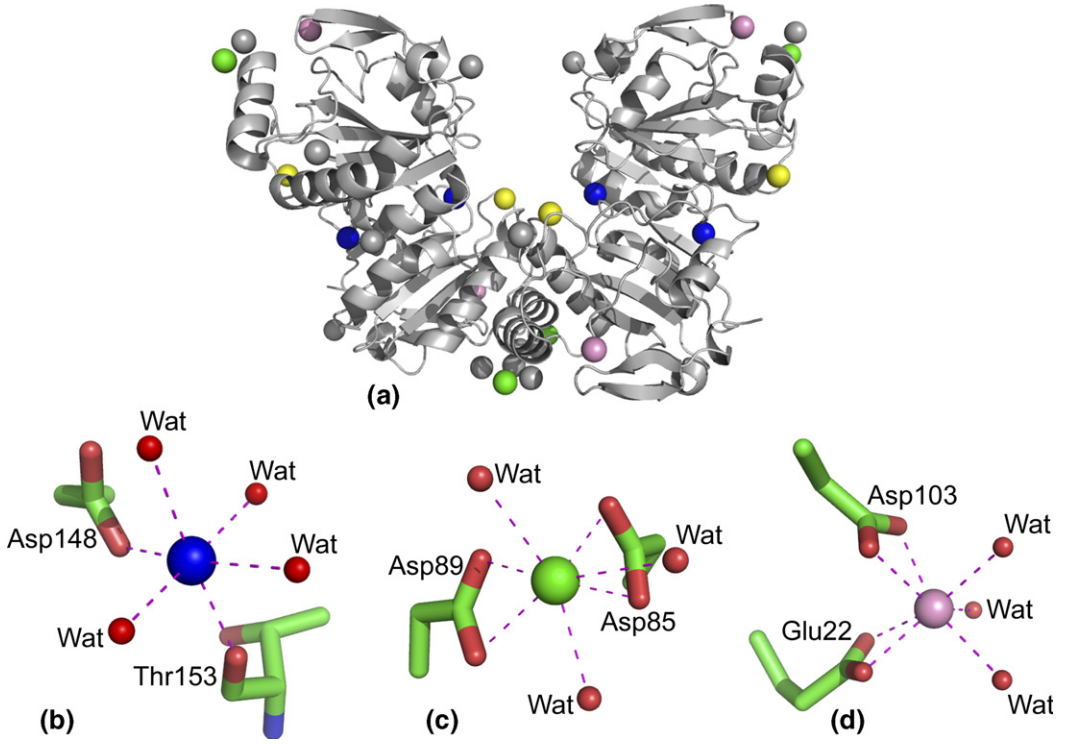


Figure 8. The 32 independent cadmium atoms in the rPtGPX5 crystal structure are displayed. (a) The two dimers of rPtGPX5 are shown in cartoon representation and Cd^{2+} involved in potential cadmium-binding sites are highlighted (yellow, blue, green and pink for sites 1, 2, 3 and 4, respectively). The sites 1 (Figures 4(a)), 2(b), 3(c) and 4(d) are shown in detail with Cd^{2+} in the same colour code as in (a).

signature of mammalian GPXs, which is highly conserved), while the immediate neighbouring residues of the catalytic Cys at the active site pocket of PRXs are Thr and Arg.[36] As expected, the rPtGPX5 displays the same overall fold as classical mammal GPXs and PRXs. Interestingly, the rPtGPX5 possesses an active site cleft architecture (consisting of Cys, Glu and Trp) very similar to that reported for mammal SeCys-GPX structures, which clearly distinguishes rPtGPX5 from PRXs. On the other hand, when rPtGPX5 is superimposed onto the thiol peroxidase structures (which belong to 2-Cys PRXs), the locations of both C_P and C_R are similar in structure, although the proteins have a very low level of sequence identity. This observation suggests the importance of the positions of these Cys residues in order to facilitate the peroxide reduction and the regeneration process. Flohé *et al*. (2003) suggested that thiol-dependent peroxidases care little about homology-based assignments of function.[34] Taking into account previous studies[10,35] and the present crystallographic evidence, we conclude that PtGPX5 is actually a thioredoxin peroxidase, structurally related to glutathione peroxidases but exhibiting catalytic and Trx-dependent recycling mechanisms of peroxiredoxin. The historical term of GPX can be used accurately only to describe a subgroup of GPX family[35] that uses GSH as electron donor despite the high level of similarity between the sequences among the family members.

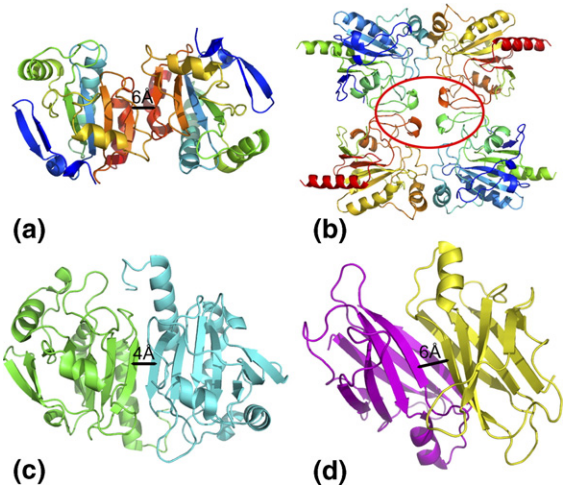


Figure 9. Cartoon representation of different oligomerization states of enzymes, including (a) the dimeric plant PtGPX5 in its reduced state (PDB code 2P5Q), (b) the classical tetrameric human plasma GPX-3 (from Professor Rudolf Ladenstein), (c) the mammalian 2-Cys peroxiredoxin heme-binding protein, HBP23 (PDB code 1QQ2) and (d) the human transthyretin or prealbumin (PDB code 2PAB). Note that the presence of oligomerization loops, circled in red, in the human plasma GPX-3 structure in (b) is likely to cause steric hindrance and obstruct the dimer formation if this feature exists in PtGPX5. In the1QQ2 structure in (c), a true ten-stranded β-sheet is observed (the intermolecular β-sheet distance is indicated) while in (a) the rPtGPX5 dimer a pseudo ten-stranded β sheet similar to that observed (d) for the 2PAB structure is seen (see the text).

Oligomerization mode of PtGPX5

Recent biochemical studies[10] showed that poplar GPXs are non-covalent dimeric enzymes in solution. This dimeric arrangement, which is found in both the reduced and oxidized form, exhibits a pseudo antiparallel β sheet that is stabilized mostly by hydrophobic clusters. Most of the residues involved in the dimerization are conserved in plant GPXs, suggesting that similar form of enzyme could be observed for all plant GPXs. Among the residues involved at the dimer interface of PtGPX5, Tyr151 is structurally an important residue (Table 2). It makes several critical contacts with residues (particularly Asp129, Gln132 and Arg149) from another subunit and is also a part of an aromatic cluster (see Results).

To date, all classical mammalian GPXs characterized are tetrameric enzymes (Figure 9(b)) except PHGPX, a mammal GPX4 that is monomeric. The sequence alignment (Figure 6) indicates that the region coding for the oligomerization loop (located in the core of the tetramer, Figure 9(b)) in typical tetrameric mammal GPXs (positions 125–142 in the bovine GPX-1 sequence) is absent from all plant GPXs as well as mammal GPX4, suggesting that these enzymes might not be tetrameric. Our structures reveal that the PtGPX5 dimer cannot be formed by other mammal GPXs due to the structural differences that would cause steric hindrances at the dimerization interface. For example in mammal GPX-1, GPX-2, GPX-3 and GPX-5, part of the oligomerization loop covers the C-terminal part of the β5 strand that is the dimerization interface in PtGPX5. In addition, mammal GPX-3 possesses a longer N-terminal chain that completely covers the β5 strand, making this enzyme unable to form the same dimer as PtGPX5. However, in the mammal monomeric GPX4[37] there is neither the oligomerization loop nor the extra-long N-terminal chain, yet this enzyme cannot form the dimeric pattern of PtGPX5. Further structural comparison reveals that a loop (residues Lys111–Ala126) in mammal GPX4 obstructs the formation of a β-sheet interaction between two subunits and hinders the dimerization.

Similar dimer formation by antiparallel association of two β-strands has been reported for typical 2-Cys and 1-Cys PRXs.[32,36] However instead of a central true ten-stranded β sheet as observed in PRX (Figure 9(c)), dimerization of PtGPX5 results in a central pseudo ten-stranded β sheet (Figure 9(a)) due to the relatively large distance between two β strands (~6 Å apart) at the dimer interface that does not allow backbone hydrogen bond formation. Therefore, the hydrophobic and aromatic networks play an important role in forming the dimer of PtGPX5 independently of its redox state. This rather unusual kind of β-sheet interactions has already been described in an amyloid sheet,[38–40] and in the prealbumin protein.[41] Sharing similar dimer properties with prealbumin, PtGPX5 dimer also produces an apparently more open quaternary structure due

to the greater separation of the β sheet at the dimerization site (Figure 9(d)).

Comparison between reduced and oxidized form PtGPX5: redox-driven conformational changes

Resolution of the reduced and oxidized structures of PtGPX5 allowed us to examine how the intramolecular disulfide bond can be formed in plant GPXs. Upon scavenging peroxide, the resulting peroxidatic Cys44-SOH intermediate will react rapidly with the resolving Cys92 to generate a stable disulfide bond at the active site of oPtGPX5. This reaction involves two conformational changes (Figure 2): (i) first, the bowing of the loop that connects β1 and α1 (where peroxidatic Cys44 is located) by about 90° towards Cys92 (in α2 helix). The flipping of this helical turn causes the peroxidatic Cys44 in rPtGPX5 to move ~10 Å before the formation of an intramolecular disulfide bond. (ii) Second, the unwinding of the whole α2 helix bringing Cys92 closer to Cys44. The absence of α2 helix in the oPtGPX5 structure (Figure 1) results in a long flexible loop that now connects β2 directly to β3. Due to its flexibility, a loop motion seems possible, leading to a "closed" conformation where the active site of oPtGPX5 is covered. We speculate that this motion is needed to cover/uncover the active site cavity of oPtGPX5 during the recycling process of this enzyme by reduced TRX.

Similar extensive conformational changes could occur for the other plant GPXs, since they share closely similar amino acid sequences and biochemical properties. Comparatively, the conformational changes observed in the present study are much more drastic than those described for atypical 2-Cys PRXs and prokaryotic thiol peroxidases so far. Indeed, in atypical 2-Cys PRXs, the change of redox state is accompanied by limited conformational changes, a helical turn of peroxidatic-Cys- (C_P-) and resolving-Cys- (C_R-) loops.[29]

The α2 helix of rPtGPX5 shows a noticeable analogy with ribosomal protein L20[42] and calmodulin (CaM),[43] which exhibit a helix–coil transition. The flexible regions of L20 and CaM share a similar sequence motif that confers helical instability, a cluster of charged amino acid residues distributed on the same helical face and spanning three helical turns. In the α2 helix of rPtGPX5, Cys92 and two negatively charged residues, Asp85 and Asp89, are located at each helix-turn, with all side-chains oriented facing the same side of the α2 helix. Once Cys92 is present in its thiolate form during catalysis, the emerging negatively charged cluster of α2 helix (Figure 7(d)) tends to make it unstable. This may lead to the complete unwinding of the α2 helix, as observed from the differences between the reduced and the oxidized form of PtGPX5. Interestingly, these two negatively charged residues are conserved or replaced by homologue residues in plant GPXs, suggesting that similar conformational changes could happen for the other GPX homologues.

In all three cases (PtGPX5, L20 and CaM), side-chain electrostatic repulsion could be responsible for helical instability, and therefore constitute a hot spot for α helix unwinding. This helix–coil transition was suggested to confer to CaM the required plasticity for binding to an exceptional variety of distinct partners.[44] We suggest here that the drastic structural rearrangements observed in PtGPX5 are needed to allow disulfide bond formation and facilitate the recognition of its reducing partner, as also proposed for methionine sulfoxide reductases,[45] enzymes that repair the oxidized methionine sulfoxide derivatives generated from ROS.[45,46]

Possible interactions between PtGPX5 and thioredoxin *h*1 from poplar

We have already mentioned that oPtGPX5 is regenerated by TRX[10] and the Trp residue adjacent to the catalytic Cys of TRX plays an important role in TRX-substrate recognition.[47] Indeed, Menchise *et al*. (2001) suggested that the Trp residue at the active site of TRX has a dual function both to force the active site in the bioactive conformation and to mediate protein–protein recognition.[47] We therefore simulated two model complexes of oPtGPX5-TRX *h*1. The poplar TRX *h*1 enzyme exhibits an unusual active site with the sequence WCPPC rather than the typical WCGPC motif. The TRX *h*1 structure (PDB code 1TI3), however, adopts the classical TRX fold, in general, although it appears to be more rigid.[33] The major interaction force in GPX5-TRX *h*1 complex is derived from two charge interactions: negative on the oPtGPX5 domain and positive on the TRX *h*1 domain.

Our molecular dynamics simulations suggest that the complex that contains an intermolecular disulfide bond between the resolving Cys92 of oPtGPX5 and catalytic Cys38 of TRX *h*1 (complex 2) is more favourable in terms of calculated binding free energy. This result corroborates recent evidence that a disulfide can be formed between the resolving Cys (Cys91) of the oxidized *Drosophila melanogaster* (*Dm*)GPX and the catalytic Cys (Cys32) of a reduced TRX C35S mutant.[35] In this study, we provide structural information describing for the first time the GPX–TRX interface and the possibility of a transient intermolecular disulfide bond formation. Taking both complexes into account, we propose a TRX recognition motif found in the majority of plant GPX sequences, the -^{80}EPGxx(D/E)xIxx(F/M)(V/A)CT(R/K)FK^{96}- motif (where x represents any residue, Figure 6), whereas the identified GPX recognition motifs in TRX, -^{36}SWC(P/G)PC(K/R)^{42}- and -^{64}D(V/T)DELK^{69}- (numbering based on the 1TI3 structure), is conserved also in the TRX *h*1 subgroup.

The protein–protein interaction illustrated in the model of oPtGPX5 and TRX *h*1 complex is in agreement with the suggestion that a cluster of charged groups on one side of the S-S bond of TRX participates in the mechanism of the redox reactions or to substrate recognition.[48] In contrast, there is also

a flat and hydrophobic surface on the other side of the S-S bond in TRX, which is suggested to be the interaction area with thioredoxin reductases[49,50] *via* ping-pong mechanisms.[51,52] Hence, based on the model of complex oPtGPX5-TRX *h*1 in this study, we conclude that the molecular surface of TRX interacting with its substrates, for example plant GPXs, is different from the site of TRX interacting with its regenerating partner, TRX reductase.

Protective role of PtGPX5 against Cd^{2+}

Cadmium is non-redox-reactive but strongly hazardous to metal-sensitive enzymes, and its toxicity is believed to be due to the depletion of glutathione and sulfhydryl groups in proteins,[52–54] contributing indirectly to oxidative stress by affecting the cellular thiol redox balance. Vido *et al*.[55] showed that the expression of proteins involved in protection of the cell against peroxides and superoxides is increased upon exposure to Cd^{2+}. In response to Cd^{2+} stress, higher plants utilise a number of defence systems, such as retention in cell walls, binding by organic molecules in the cytosol and sequestration in the vacuole.[56] Thiol-rich compounds such as GSH are also proposed as Cd^{2+}-detoxifying compounds in plants.[57,58] However GSH depletion in response to Cd^{2+} has been reported in several plant species.[59] Alternatively, thioredoxin from the thiol-disulfide oxidoreductase family has been proposed to be an efficient Cd^{2+} chelator upon acute Cd^{2+} exposure.[60,61] In a more relevant context, yeast GPX-3 over-expression was shown to be effective in raising the threshold of cadmium tolerance.[62] Moreover, expression of GPX-3^{C82S} did not complement the cadmium sensitivity of the gpx3Δ strain, corroborating the notion that a peroxidase activity of GPX-3 is required for cadmium resistance. The present reduced form structure (rPtGPX5) displays the ability of two dimers to interact with 32 atoms of Cd^{2+} in the asymmetric unit (Figure 8(a)) in which eight of them are trapped independently by a dimer, suggesting that this enzyme could play an additional role as a Cd^{2+}-sink in defence against heavy metals, particularly Cd^{2+}. Four potential cadmium-binding sites are identified in each subunit (see Results). It is worth mentioning that the first three cadmium-binding sites, including the one at the active site, involve residues that are conserved in other plant GPX homologues. The cadmium ion captured at the active site also has an important role in locking the reduced conformation of PtGPX5 (Figure 4(a)) since this enzyme oxidizes spontaneously in solution. The fourth cadmium-binding site, however, involves non-conserved residues in plant GPXs, thus suggesting that PtGPX5 may be an efficient Cd^{2+}-sink as compared to other isoforms. Recently, Navrot *et al*. (2006) have demonstrated that when poplar leaves are subjected to metal stresses such as the presence of high concentrations of cadmium or copper, the levels of GPX are modified.[10] Likewise, plant GPX5 homologues could play a broader role as a heavy-metal sink than just as a peroxide scavenger.

Materials and Methods

Protein purification and crystallization

The PtGPX5 gene was amplified from a *P. trichocarpa*×*deltoides* root cDNA library. The gene was inserted in the pET-3d expression plasmid, between NcoI and BamHI sites. Recombinant plasmids carrying the gene of interest were electroporated into methionine auxotrophic strain of *Escherichia coli* BL21 (DE3) pSBET.[63] Bacteria were cultured at 37 °C in M9 medium supplemented with the usual amino acid residues (100 mg/l) except methionine. A final culture volume of 2 l was reached by two successive subculturing processes. Expression of recombinant PtGPX5 was induced at exponential phase with 100 μM isopropyl-β-d-thiogalactoside (IPTG) and cell growth continued for another 4 h before harvesting. Simultaneous with the addition of IPTG, selenomethionine was added to the culture medium at a final concentration of 100 mg/l. Purification of PtGPX5 was completed by a two-step chromatography as described,[10] and the purified protein was kept at −20 °C in storage buffer (30 mM Tris–HCl (pH 8), 1 mM EDTA). Mass spectrometry was performed to assess purity and to confirm the full incorporation of selenomethionine. Electrospray mass analysis of the native reduced form of PtGPX5 (19279(±1.61) Da) showed that the N-terminal methionine residue was not present in the purified protein sample. Therefore, two methionine residues were expected to be substituted in one monomer of [SeMet]PtGPX5. Indeed, the observed molecular mass of 19374(±1.96) Da indicates that the two methionine residues were fully substituted by selenomethionine in each monomer of [SeMet]PtGPX5. As for the native PtGPX5 used in the present crystallographic work, the protein was over-expressed and purified as described.[10]

Crystallization conditions were screened extensively at 20 °C by batch methods, using microliter amounts in microdroplets under oil. Drops used for the initial crystallization trials consisted of 1.5 μl of the protein solution concentrated at 20 mg/ml mixed with 1.5 μl of various crystallization solutions. The most promising screen conditions from Jena Bioscience GmbH (Jena, Germany) and Hampton Research (CA, USA) were optimized to obtain large crystals suitable for X-ray analysis. Both the native and the SeMet reduced form of PtGPX5 crystals were obtained after 7 days in Hampton crystal screen 2 solution 34 (0.1 M Hepes (pH 7.5), 0.05 M cadmium sulphate hydrate, 1.0 M sodium acetate). On the other hand, native oxidized PtGPX5 crystals were obtained within 2 days in JBS 2–C4 crystallization condition containing 0.1 M Tris–HCl (pH 8.5), 25% (w/v) PEG 4000, 0.2 M calcium chloride. All crystals were cryoprotected in a mother liquor supplemented with 20% (v/v) glycerol before flash-cooling in liquid ethane and liquid nitrogen.

Data collection and processing

X-ray data for cryopreserved crystals of both selenomethionine-substituted (reduced form) and native (oxidized form) enzymes were collected at the selenium absorption edge (0.980043 Å) and at wavelength

0.80630 Å, respectively. Both sets of diffraction data were collected by using the Mar Research CCD detector and the radiation produced by the ESRF beamline BM30A (Grenoble, France). As for the native reduced PtGPX5 crystal, a dataset was collected at the DORIS storage ring, EMBL X11 beamline (DESY, Hamburg, Germany) at wavelength 0.81560 Å. All images were indexed, integrated and scaled using either the HKL program[64] or the XDS program package.[65] While both the [SeMet] and the native reduced form of PtGPX5 crystals are rhombohedral (designated form I), the native oxidized PtGPX5crystal is trigonal (designated form II).

Form I crystals belong to the rhombohedral space group $R3$ with cell dimensions of $a=222.7$ Å and $c=48.1$ Å. Assuming four subunits per asymmetric unit, the V_M value of the crystal is 2.93 Å^3/Da.[66] Form II exhibits space group $P3_121$ with cell dimensions of $a=71.6$ Å and $c=117.8$ Å. The presence of two subunits per asymmetric unit in form II crystal gives the V_M value of 2.24 Å^3/Da. For details of data collection statistics, see Table 1.

Phasing and structure refinements

Multiple anomalous dispersion (MAD) data were collected from PtGPX5 crystal to 2.2 Å resolution at the selenium edge. The 3W-MAD, 2W-MAD and single wavelength anomalous diffraction approaches were tested using Auto-Rickshaw, an automated crystal structure determination platform‡ at webserver, However only the single wavelength anomalous diffraction approach using the Beamline Version of Auto-Rickshaw was successful and resulted in an interpretable map with a partial model. The Beamline Version uses the data to 3.0 Å resolution whereas the Advanced Version of Auto-Rickshaw includes maximum resolution of the data. The version was not successful with any phasing protocol; this observation indicated that the crystal might have suffered severely with radiation damage (see below). High-resolution reflections are more sensitive to radiation damage and such damage causes problems for scaling and merging the datasets, and can prevent measurement of anomalous and dispersive differences that consequently affects the experimental phasing. Therefore, the peak data were reprocessed and scaled. Data quality was found to be good to 2.7 Å for phasing (Table 1).

Overall processes that have taken place in the pipeline included: SHELXD,[67] which located eight expected selenium atoms present in the asymmetric unit; MLPHARE[68] for initial phasing; DM[69] for density modification and twofold NCS averaging; and ESSENS[70] to position α helices in the structure. The initial phases were extended to 2.7 Å resolution using RESOLVE version 2.02.[71]

Initially, molecular replacement using the closest structure (bovine erythrocyte GPX, PDB code 1GP1) was unsuccessful against the native dataset; however, the model was placed successfully in electron density using phased molecular replacement. The resultant model was checked on graphics for the crystal packing, which clearly showed that a long loop (19 residues) clashes with the neighbouring residues; therefore, the loop was removed and the phased molecular replacement was repeated. This gave the best fit to the experimental electron density. The model was used as a starting point for further manual model building using the graphics program COOT.[72]

Once the model was 90% complete, positional and B-factor refinements were performed using the protocols implemented in CNS, version 1.1.[73] The partially refined [SeMet]PtGPX5 model was used as a search model in a molecular replacement procedure using MOLREP[74] against the native data, and the resultant model was used for automatic model building using the program ARP/wARP:[75] 99% of the model was built automatically. Further, the model was improved by manual building with Turbo-Frodo§ interspersed with refinement using CNS. At this stage, it became evident that the PtGPX5 crystal contained additional heavy atoms. The calculation of anomalous difference Fourier map indicated 32 cadmium atoms bound to the protein in one asymmetric unit. Some cadmium was expected to bind the protein, as the cadmium sulphate hydrate was used during crystallization; however such a large number of cadmium ions bound to the protein is generally not common. Therefore, the radiation damage to the [SeMet] crystal was significantly high and prohibited phasing information using MAD. In addition, eight acetate anions, which was also an component of the crystallization solution, were added to the final native reduced form structure of PtGPX5 before several refinement steps using both REFMAC5[68] and CNS.

A model of the reduced form of native enzyme was used for molecular replacement to solve the structure of the oxidized form of enzyme. The solution was further refined using CNS, and iteratively combined with manual building with the program described above. Throughout the model building and refinement process, the qualities of the models were assessed using the program PROCHECK.[76] Refinement statistics are summarised in Table 1. The Figures were prepared with PyMol‖.

Molecular modelling

The structure of thioredoxin in complex with a protein substrate is not available. Thus, the initial model for the oxidized PtGPX5-thioredoxin (oPtGPX5-1TI3) complex was constructed as follow. The initial model of the complex was generated using the program Turbo-Frodo by simply superimposing the minimized average structure of 1TI3 (as the structure of TRX $h1$ was solved by a nuclear magnetic resonance approach and there is actually an ensemble of 20 structures) to the nearest symmetry mate of the oPtGPX5 molecule (monomer A), paying special attention to Trp124 of the symmetry mate (Trp124′). The complex was built to calculate the most energy-favourable complex in terms of its stability. All minimizations and MD simulations of the complex were performed using the AMBER8.0 simulation package¶ and the AMBER parm99 force field[77] with a TIP3P water model.[78] The complex molecules were explicitly solvated in a truncated octahedral box and all counterions were excluded from all the calculations. The calculations were performed using two processors (Intel P4 Xeon) on the local in-house cluster with 3 GHz CPUs. The non-bonded cutoff was set to 12.0 Å in all calculations unless mentioned otherwise. Before the MD simulations, all structures were energy-minimized using the SANDER module in AMBER 8.0, first by holding the complex fixed while minimizing the positions of water molecules using the conjugate gradient method for 1000 cycles and then minimizing the entire molecular system for 2500 cycles. MD simulations were carried out thereafter. The temperature of the system was raised gradually from 0 K to 300 K

‡ http://www.embl-hamburg.de/Auto-Rickshaw

§ http://www.afmb.univ-mrs.fr/-TURBO-
‖ http://pymol.sourceforge.net
¶ http://www.amber.scripps.edu

(with weak restraint on the complex) in 20 ps followed by 100 ps equilibration (without any restraint) at 300 K. All MD simulations were performed with a time-step of 2.0 fs. Various parameters (density, temperature, pressure, kinetic energy and potential energy) were monitored during the MD simulations, and generally were stable during the 100 ps equilibration.

Using the molecular mechanics Poisson–Boltzmann/surface area (MM-PB/SA) method,[79] snapshots for energy analyses were obtained from the MD trajectories of both complexes. The water molecules were deleted for energy analyses. We have selected 100 snapshots from each equilibrated trajectory in 100 ps intervals. Snapshots from the trajectories were visually examined and illustrated with visual molecular dynamics (VMD) program.[80]

Sequence analysis and structure alignments

Deduced amino acids of PtGPX5 (170 residues), deposited in the Genbank database with the accession number ABN59534 was analysed using the programs BLASTP[81] to search for its homologous protein sequences and ClustalW[82] to align selected sequences. As for the tertiary structure, superimpositions of the present structures with their homologous structures obtained from PDB were performed using LSQMAN program from the DEJAVU package[83] and Lsqkab (superpose) program of the CCP4 program suite 5.0.2[68] with default parameters proposed by the author. The protein–protein interface was analysed using the Protein-Protein Interaction Server (V1.5).[84]

Protein Data Bank accession codes

All coordinates and structure factors of the native reduced and oxidized forms of PtGPX5 have been deposited in the RCSB Protein Data Bank, with the accession code 2P5Q and 2P5R, respectively.

Acknowledgements

We are very grateful to the FIP-team in ESRF, Grenoble, France and the DESY-team in EMBL-Hamburg Outstation, Germany, for providing access to beamlines BM30A and X11, respectively. This work was supported by financial aid from the CNRS and the French government. C.S.K. is a recipient of the ASTS fellowship from the Universiti Sains Malaysia.

References

1. Farr, S. B. & Kogoma, T. (1991). Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium*. *Microbiol. Rev.* **55**, 561–585.
2. Dat, J., Vandenabeele, S., Vranova, E., Van Montagu, M., Inzé, D. & Van Breusegem, F. (2000). Dual action of the active oxygen species during plant stress responses. *Cell Mol. Life Sci.* **57**, 779–795.
3. Milla, M. A. R., Maurer, A., Huete, A. R. & Gustafson, J. P. (2003). Glutathione peroxidase genes in Arabidopsis are ubiquitous and regulated by abiotic stresses through diverse signalling pathways. *The Plant J.* **36**, 602–615.
4. Foyer, C. H., Lescure, J. C., Lefebvre, C., Morot-Gaudry, J. F., Vincentz, M. & Vaucheret, H. (1994). Adaptations of photosynthetic electron transport, carbon assimilation, and carbon partitioning in transgenic *Nicotiana plumbaginifolia* plants to changes in nitrate reductase activity. *Plant Physiol.* **104**, 171–178.
5. Noctor, G. & Foyer, C. H. (1998). Ascorbate and glutathione: keeping active oxygen under control. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 249–279.
6. Roy, G., Sarma, B. K., Phadnis, P. P. & Mugesh, G. (2005). Selenium-containing enzymes in mammals: chemical perspectives. *J. Chem. Sci.* **117**, 287–303.
7. Miao, Y., Lv, D., Wang, P., Wang, X.-C., Chen, J., Miao, C. & Song, C.-P. (2006). An Arabidopsis glutathione peroxidase functions as both a redox tranducer and a scavenger in abscisic acid and drought stress responses. *Plant Cell*, **18**, 2749–2766.
8. Eshdat, Y., Holland, D., Faltin, Z. & Ben-Hayyim, G. (1997). Plant glutathione peroxidases. *Physiol. Plantarum*, **100**, 234–240.
9. Delaunay, A., Pflieger, D., Barrault, M. B., Vinh, J. & Toledano, M. B. (2002). A thiol peroxidase is an $HB_{2B}OB_{2B}$ receptor and redox-transducer in gene activation. *Cell*, **111**, 471–481.
10. Navrot, N., Collin, V., Gualberto, J., Gelhaye, E., Hirasawa, M., Rey, P. *et al.* (2006). Plant glutathione peroxidases are functional peroxiredoxins distributed in several subcellular compartments and regulated during biotic and abiotic stresses. *Plant Physiol.* **142**, 1–17.
11. Jung, B. G., Lee, K. O., Lee, S. S., Chi, Y. H., Jang, H. H., Kang, S. S. *et al.* (2002). A chinese cabbage cDNA with high sequence identity to phospholipids hydroperoxide glutathione peroxidases encodes a novel isoform of thioredoxin-dependent peroxidase. *J. Biol. Chem.* **277**, 12572–12578.
12. Kang, S.-G., Jeong, H. K. & Suh, H. S. (2004). Characterization of a new member of the glutathione peroxidase gene family in *Oryza sativa*. *Mol. Cells*, **17**, 23–28.
13. Holland, D., Ben-Hayyim, G., Faltin, Z., Camoin, L., Strosberg, A. D. & Eshdat, Y. (1993). Molecular characterization of salt-stress-associated protein in citrus: protein and cDNA sequence homology to mammalian glutathione peroxidases. *Plant Mol. Biol.* **21**, 923–927.
14. Sugimoto, M. & Sakamoto, W. (1997). Putative phospholipids hydroperoxide glutathione peroxidase gene from *Arabidopsis thaliana* induced by oxidative stress. *Genes Genet. Syst.* **72**, 311–316.
15. Ren, X., Yang, L., Liu, J., Su, D., You, D., Liu, C. *et al.* (2001). A novel glutathione peroxidase mimic with antioxidant activity. *Arch. Biochem. Biophys.* **387**, 250–256.
16. Herbette, S., Lenne, C., Leblanc, N., Julien, J.-L., Drevet, J. R. & Roeckel-Drevet, P. (2002). Two GPX-like proteins from *Lycopersion esculentum* and *Helianthus annuus* are antioxidant enzymes with phospholipid hydroperoxide glutathione peroxidase and thioredoxin peroxidase activities. *Eur. J. Biochem.* **269**, 2414–2420.
17. Tanaka, T., Izawa, S. & Inoue, Y. (2005). GPX2, encoding a phospholipid hydroperoxide glutathione peroxidase homologue, codes for an atypical 2-Cys peroxiredoxin in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**, 7397–7403.
18. Sztajer, H., Gamain, B., Aumann, K. D., Slomianny, C., Becker, K., Brigelius-Flohe, R. & Flohe, L. (2001). The putative glutathione peroxidase gene of *Plasmodium falciparum* codes for a thioredoxin peroxidase. *J. Biol. Chem.* **276**, 7397–7403.

19. Prabhakar, R., Vreven, T., Morokuma, K. & Musaev, D. G. (2005). Elucidation of the mechanism of selenoprotein glutathione peroxidase (GPx)-catalyzed hydrogen peroxide reduction by two glutathione molecules: a density functional study. *Biochemistry*, **44**, 11864–11871.
20. Prabhakar, R., Morokuma, K. & Musaev, D. G. (2006). Peroxynitrite reductase activity of selenoprotein glutathione peroxidase: a computational study. *Biochemistry*, **45**, 6967–6977.
21. Flohé, L., Loschen, G., Gunzler, W. A. & Eichele, E. (1972). Glutathione peroxidase, V. The kinetic mechanism. *Hoppe Seylers Z Physiol. Chem.* **353**, 987–999.
22. Mugesh, G., du Mont, W. & Sies, H. (2001). Chemistry of biologically important synthetic organoselenium compounds. *Chem. Rev.* **101**, 2125–2179.
23. Rouhier, N. & Jacquot, J.-P. (2005). The plant multigenic family of thiol peroxidases. *Free Radic. Biol. Med.* **38**, 1413–1421.
24. Epp, O., Ladenstein, R. & Wendel, A. (1983). The refined structure of the selenoenzyme glutathione peroxidase at 0.2-nm resolution. *Eur. J. Biochem.* **133**, 51–69.
25. Ren, B., Huang, W., Åkesson, B. & Ladenstein, R. (1997). The crystal structure of seleno-glutathione peroxidase from human plasma at 2.9 Å resolution. *J. Mol. Biol.* **268**, 869–885.
26. Grignard, E., Morin, J., Vernet, P. & Drevet, J. R. (2005). GPX5 orthologs of the mouse epididymis-restricted and sperm-bound selenium-independent glutathione peroxidase are not expressed with the same quantitative and spatial characteristics in large domestic animals. *Theriogenology*, **64**, 1016–1033.
27. Fujii, J., Iuchi, Y., Matsuki, S. & Ishii, T. (2003). Cooperative function of antioxidant and redox systems against oxidative stress in male reproductive tissues. *Asian J. Androl.* **5**, 231–242.
28. Martin, J. F. (1995). Thioredoxin – a fold for all reasons. *Structure*, **3**, 245–250.
29. Choi, J., Choi, S., Choi, J., Cha, M. K., Kim, I. H. & Shin, W. (2003). Crystal structure of *Escherichia coli* thiol peroxidase in the oxidized state: insights into intramolecular disulfide formation and substrate binding in atypical 2-Cys peroxiredoxins. *J. Biol. Chem.* **278**, 49478–49486.
30. Hirotsu, S., Abe, Y., Okada, K., Nagahara, N., Hori, H., Nishino, T. S. & Hakoshima, T. (1999). Cystal structure of a multifunctional 2-Cys peroxiredoxin heme-binding protein 23 kDa/proliferation-associated gene product. *Proc. Natl Acad. Sci. USA*, **96**, 12333–12338.
31. Schröder, E., Littlechild, J. A., Lebedev, A. A., Errington, N., Vagin, A. A. & Isupov, M. N. (2000). Crystal structure of decameric 2-Cys peroxiredoxin from human erythrocytes at 1.7 Å resolution. *Structure*, **8**, 605–615.
32. Declercq, J. P., Evrard, C., Clippe, A., Stricht, D. V., Bernard, A. & Knoops, B. (2001). Crystal structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 Å resolution. *J. Mol. Biol.* **311**, 751–759.
33. Coudevylle, N., Thureau, A., Hemmerlin, C., Gelhaye, E., Jacquot, J. P. & Cung, M. T. (2005). Solution structure of a natural CPPC active site variant, the reduced form of thioredoxin h1 from poplar. *Biochemistry*, **44**, 2001–2008.
34. Flohé, L., Jaeger, T., Pilawa, S. & Sztajer, H. (2003). Thiol-dependent peroxidases care little about homology-based assignments of function. *Redox Rep.* **8**, 256–264.
35. Maiorino, M., Ursini, F., Bosello, V., Toppo, S., Tosatto, S. C., Mauri, P. *et al.* (2006). The thioredoxin specificity of Drosophila GPx: a paradigm for a peroxiredoxin-like mechanism of many glutathione peroxidases. *J. Mol. Biol.* **365**, 1033–1046.
36. Wood, Z. A., Schröder, E., Harris, J. R. & Poole, L. B. (2003). Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem. Sci.* **28**, 32–40.
37. Ursini, F., Maiorino, M., Brigelius-Flohé, R., Aumann, K. D., Roveri, A., Schomburg, D. & Flohé, L. (1995). Diversity of glutathione peroxidases. *Methods Enzymol.* **252**, 38–53.
38. Serag, A. A., Altenbach, C., Gingery, M., Hubbell, W. L. & Yeates, T. O. (2002). Arrangement of subunits and ordering of beta-strands in an amyloid sheet. *Nature Struct. Biol.* **9**, 734–739.
39. Hwang, W., Zhang, S., Kamm, R. D. & Karplus, M. (2004). Kinetic control of dimer structure formation in amyloid fibrillogenesis. *Proc. Natl Acad. Sci. USA*, **101**, 12916–12921.
40. Lei, H., Wu, C., Wang, Z. & Duan, Y. (2006). Molecular dynamics simulations and free energy analyses on the dimer formation of an amyloidogenic heptapeptide from human beta2-microglobulin: implication for the protofibril structure. *J. Mol. Biol.* **356**, 1049–1063.
41. Blake, C. C., Geisow, M. J., Oatley, S. J., Rerat, B. & Rerat, C. (1978). Structure of prealbumin: secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å. *J. Mol. Biol.* **121**, 339–356.
42. Timsit, Y., Allemand, F., Chiaruttini, C. & Springer, M. (2006). Coexistence of two protein folding states in the crystal structure of ribosomal protein L20. *EMBO Rep.* **7**, 1013–1018.
43. Babu, A., Orr, G. & Gulati, J. (1988). Calmodulin supports the force-generating function in desensitized muscle fibers. *J. Biol. Chem.* **263**, 15485–15491.
44. Hoeflich, K. P. & Ikura, M. (2002). Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell.* **108**, 739–742.
45. Boschi-Muller, S., Olry, A., Antoine, M. & Branlant, G. (2005). The enzymology and biochemistry of methionine sulfoxide reductases. *Biochim. Biophys. Acta*, **1703**, 231–238.
46. Moskovitz, J., Bar-Noy, S., Williams, W. M., Requena, J., Berlett, B. S. & Stadtman, E. R. (2001). Methionine sulfoxide reductase (MsrA) is a regulator of antioxidant defense and lifespan in mammals. *Proc. Natl Acad. Sci. USA*, **98**, 12920–12925.
47. Menchise, V., Corbier, C., Didierjean, C., Jacquot, J. P., Benedetti, E., Saviano, M. & Aubry, A. (2000). Crystal structure of the W35A mutant thioredoxin h from *Chlamydomonas reinhardtii*: the substitution of the conserved active site Trp leads to modifications in the environment of the two catalytic cysteines. *Biopolymers*, **56**, 1–7.
48. Eklund, H., Cambillau, C., Sjoberg, B. M., Holmgren, A., Jornvall, H., Hoog, J. O. & Branden, C. I. (1984). Conformational and functional similarities between glutaredoxin and thioredoxins. *EMBO J.* **3**, 1443–1449.
49. Slaby, I. & Holmgren, A. (1979). Structure and enzymatic functions of thioredoxin refolded by complementation of two tryptic peptide fragments. *Biochemistry*, **18**, 5584–5591.
50. Thelander, L. (1974). Reaction mechanism of ribonucleoside diphosphate reductase from Escherichia coli. Oxidation-reduction-active disulfides in the B1 subunit. *J. Biol. Chem.* **249**, 4858–4862.
51. Holmgren, A. (1979). Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. *J. Biol. Chem.* **254**, 9627–9632.
52. Fortuniak, A., Zadzinski, R., Bilinski, T. & Bartosz, G.

(1996). Glutathione depletion in the yeast *Saccharomyces cerevisiae*. *Biochem. Mol. Biol. Int.* **38**, 901–910.
53. Zenk, M. H. (1996). Heavy metal detoxification in higher plants – a review. *Gene*, **179**, 21–30.
54. Stohs, S. J., Bagchi, D., Hassoun, E. & Bagchi, M. (2001). Oxidative mechanisms in the toxicity of chromium and cadmium ions. *J. Environ. Pathol. Toxicol. Oncol.* **20**, 77–88.
55. Vido, K., Spector, D., Lagniel, G., Lopez, S., Toledano, M. B. & Labarre, J. (2001). A proteome analysis of the cadmium response in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**, 8469–8474.
56. Vázquez, S., Goldsbrough, P. & Carpena, R. O. (2006). Assessing the relative contributions of phytochelatins and the cell wall to cadmium resistance in white lupin. *Physiol. Plant.* **128**, 487–495.
57. Howden, R., Goldsbrough, P. B., Andersen, C. R. & Cobbett, C. S. (1995). Cadmium-sensitive, cad1 mutants of *Arabidopsis thaliana* are phytochelatin deficient. *Plant Physiol.* **107**, 1059–1066.
58. Vogeli-Lange, R. & Wagner, G. J. (1990). Subcellular localization of cadmium and cadmium-binding peptides in tobacco leaves: implication of a transport function for cadmium-binding peptides. *Plant Physiol.* **92**, 1086–1093.
59. Steffens, J. C. (1990). The heavy metal-binding peptides of plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **41**, 553–575.
60. Rollin-Genetet, F., Berthomieu, C., Davin, A. H. & Quemeneur, E. (2004). *Escherichia coli* thioredoxin inhibition by cadmium: two mutually exclusive binding sites involving Cys32 and Asp26. *Eur. J. Biochem.* **271**, 1299–1309.
61. Lemaire, S. D., Stein, M., Issakidis-Bourguet, E., Keryer, E., Benoit, V. V., Pineau, B. *et al.* (1999). The complex regulation of ferredoxin/thioredoxin-related genes by light and the circadian clock. *Planta*, **209**, 221–229.
62. Avery, A. M., Willetts, S. A. & Avery, S. V. (2004). Genetic dissection of the phospholipid hydroperoxidase activity of yeast gpx3 reveals its functional importance. *J. Biol. Chem.* **279**, 46652–46658.
63. Schenk, P. M., Baumann, S., Mattes, R. & Steinbiss, H. H. (1995). Improved high-level expression system for eukaryotic genes in *Escherichia coli* using T7 RNA polymerase and rare Arg tRNAs. *Biotechniques*, **19**, 196–198.
64. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
65. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallog.* **26**, 795–800.
66. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
67. Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallog. sect. D*, **58**, 1772–1779.
68. CCP4: Collaborative Computational Project. Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
69. Cowtan, K. (1994). Dm: an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34–38.
70. Kleywegt, G. J. & Jones, T. A. (1997). Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Crystallog. sect. D*, **53**, 179–185.
71. Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallog. sect. D*, **56**, 965–972.
72. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallog. sect. D*, **60**, 2126–2132.
73. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
74. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallog.* **30**, 1022–1025.
75. Perrakis, A., Morris, R. J. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
76. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
77. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Jr *et al.* (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
78. Jorgensen, W. L., Chandrasekhar, J., Madura, J. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
79. Srinivasan, J., Miller, J., Kollman, P. A. & Case, D. A. (1998). Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. DynAM.* **16**, 671–682.
80. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38.
81. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
82. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
83. Kleywegt, G. T. & Jones, T. A. (1997). Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* **277**, 525–545.
84. Jones, S. & Thornton, J. M. (1996). Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.

Edited by M. Guss

(Received 5 February 2007; received in revised form 6 April 2007; accepted 9 April 2007)
Available online 19 April 2007

Available online at www.sciencedirect.com
ScienceDirect
Biochimica et Biophysica Acta xx (2008) xxx – xxx
Biochimica et Biophysica Acta BBA
www.elsevier.com/locate/bbagen

Review

Redox based anti-oxidant systems in plants: Biochemical and structural analyses

Nicolas Rouhier [a], Cha San Koh [b], Eric Gelhaye [a], Catherine Corbier [b], Frédérique Favier [b], Claude Didierjean [b], Jean-Pierre Jacquot [a,*]

[a] *Unité Mixte de Recherches 1136 Interaction Arbres Microorganismes, IFR 110 GEEF, Nancy University, Faculté des Sciences 54506 Vandoeuvre-lès-Nancy, Cedex, France*
[b] *LCM3B, Equipe Biocristallographie, UMR 7036 CNRS-UHP, Faculté des Sciences et Techniques, Nancy Université, BP 239, 54506 Vandoeuvre-lès-Nancy, France*

Received 9 November 2007; received in revised form 11 December 2007; accepted 17 December 2007

Abstract

We provide in this paper a comparative biochemical and structural analysis of the major thiol oxidoreductases (thioredoxin and glutaredoxin) of photosynthetic organisms in relation with their reductases and with target proteins, especially those involved either in the detoxication of peroxides such as hydrogen peroxide (thiol-peroxidases) or in the repair of oxidized methionines in proteins (methionine sulfoxide reductases). Particular emphasis will be given to the catalytic and regeneration mechanisms used by these enzymes. In addition, the protein–protein interactions of these systems will be discussed, leading to an integrated view of the functioning of these systems in various plant sub-cellular compartments.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Glutaredoxin; Methionine sulfoxide reductase; Protein–protein interaction; Structure; Thiol-peroxidase; Thioredoxin

1. The need for anti-oxidant systems in plants

In photosynthetic and non-photosynthetic organisms, the functioning of the mitochondrial transport chain generates ROS (reactive oxygen species) which can at high concentration be damaging to macromolecules [1]. They include superoxide ions, peroxides and hydroxyl radicals. Because of their photosynthetic capacities, plants have to deal with a second source of oxygen-related molecules generated at the level of the photosynthetic electron chain. In these organisms, the generation of these compounds (together with singlet oxygen) is greatly enhanced in reactions occurring mostly at the level of photosystem I and photosystem II [2]. No wonder then that the anti-oxidant/repair enzymatic equipment of plants is considerably enhanced and diversified compared to the ones of bacterial or animal systems. A third reason that explains the exaltation of the redox defence in plants is that the overwhelming majority of plant species are essentially immobile and have to deal with environmental adversity and deprivation that can include stresses linked to water deficit, temperature fluctuation, heavy metal pollution, etc. Most of these stresses are accompanied by an oxidative burst that needs to be controlled. While animal models can escape those unfavourable conditions by simply moving away, land plants cannot do that. A last explanation for the multiplicity of anti-oxidant redox catalysts in plants is that they are widely distributed in the organs (roots, leaves, stems, flowers, fruits etc…) or in the sub-cellular compartments of higher plants.

2. The thiol-dependent peroxidases: peroxiredoxins and "glutathione" peroxidases

Hydrogen peroxide (H_2O_2) is one of the peroxides generated in plants subjected to stress, but other more complex peroxides known as alkyl hydroperoxides can also be generated at the level of fatty acids and lipids. H_2O_2 is generated from superoxide ions via the catalytic action of superoxide dismutases, and it is both an oxidant and a signal transmitter [3]. Although it is a

* Corresponding author.
E-mail address: j2p@scbiol.uhp-nancy.fr (J.-P. Jacquot).

0304-4165/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbagen.2007.12.007

moderate oxidant, it is necessary to control the levels of H_2O_2 in order to prevent damages to macromolecules and subsequent cellular degradation and also to prevent its transformation into hydroxyl radicals through the Fenton reaction. Among the enzymes responsible for the reduction of H_2O_2 are catalases in peroxisomes and ascorbate peroxidases in the cytosol and in the chloroplast. Both are haem-containing enzymes with very high specificity and very high turnover with H_2O_2 but unable to reduce alkyl hydroperoxides. More recently, additional peroxide-degrading enzymes which can reduce a broader range of peroxide substrates have been detected and analyzed in plants. They constitute a superfamily of thiol-peroxidases which includes peroxiredoxins (Prxs) and so-called "glutathione" peroxidases (Gpxs) [4].

2.1. Catalytic and regeneration mechanisms

All thiol-peroxidases function in a similar way: the peroxide molecule is attacked by a peroxidatic cysteine resulting into the formation of a sulfenic acid together with the release of the corresponding alcohol. The resulting sulfenic acid can then be further oxidized to sulfinic acid, a redox state that can be reversed in some Prxs by sulfiredoxin [5], or sulfonic acid, an irreversible oxidation state associated with the loss of activity of the enzyme (Fig. 1). Alternatively, in the normal catalytic cycle of those proteins, the sulfenic acid is reduced either directly via glutathione or glutaredoxin (type II Prx) or indirectly through a resolving cysteine that itself belongs to the protein (Fig. 2) [6]. This results in an intra-(Prx Q, Gpxs) or sometimes inter-subunit disulfide (typical 2-Cys Prx) which is in turn generally reduced to the dithiol form via thioredoxin. The oxidized thioredoxin generated upon regeneration of the Prx or Gpx protein is then converted into the reduced state via NADPH and NADPH

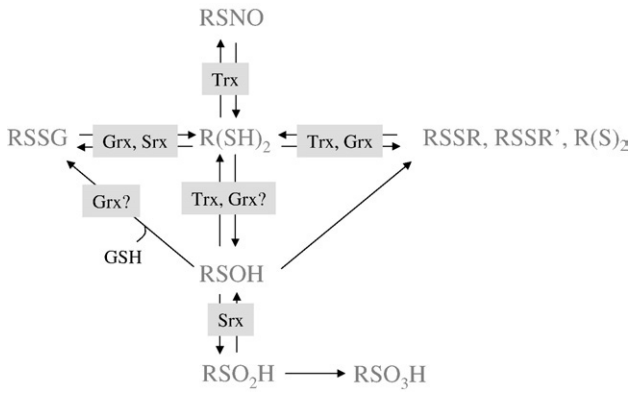


Fig. 1. Cysteine oxidation states and their regulation via proteins of the redoxin family. From a protein with two thiol groups, intra-$R(S)_2$ or inter-molecular (RSSR and RSSR′) disulfide bridges can be formed. These post-translational modifications can be reversed either by Grxs or by Trxs. It has been postulated several times that both protein families are also able to reduce directly sulfenic acids. For proteins using the sulfenic acid chemistry, a drawback is the formation of overoxidized forms such as sulfinic and sulfonic acids. Only in the first case, have sulfiredoxins been shown to be able to reduce some Prxs. More specific functions could be attributed to Trx and Grxs. The latter are efficient enzymes for deglutathionylation reaction and maybe for the glutathionylation reaction. Srxs have also been proposed to possess deglutathionylation activity [87]. Trxs could be implicated in the nitrosylation [98].

thioredoxin reductase (NTR) or photoreduced ferredoxin and ferredoxin-thioredoxin reductase (FTR) [7].

In plants, while the 2-Cys Prx, Prx Q and Gpx are reduced and regenerated only via thioredoxin or cyclophilin [8–11], the type II Prx is unique as it can be reduced either by glutaredoxin or thioredoxin and sometimes even by glutathione [12–16]. The active site of a type II Prx from poplar was investigated in detail and three amino acids form a catalytic triad. Besides the catalytic cysteine, it comprises a conserved threonine and a conserved arginine, both of which are required for activity [17]. The type II peroxiredoxin–glutaredoxin/thioredoxin interactions have been studied using biochemical and structural approaches. The sulfenic acid generated on the catalytic cysteine of poplar Prx IIB can be attacked by the catalytic cysteine of either glutaredoxin or thioredoxin and the use of cysteine mutants together with the addition of oxidants has permitted to demonstrate the formation of a heterodisulfide between the proteins [13]. On the other hand, upon glutathionylation of the protein, the non-covalent homodimer of Prx IIB readily dissociates into monomers [18]. Thus, when the GSH/Grx regenerating system operates, it is likely that the sulfenic acid of Prx IIB is first attacked by glutathione and then, in a second step, Grx is able to enzymatically deglutathionylate it, leading to the active reduced form of Prx IIB [18]. To date, plant type II Prx is the only peroxiredoxin which is regenerated by an external glutaredoxin, but this initial discovery was followed and confirmed shortly after by the detection in several pathogenic bacteria and cyanobacteria of natural hybrid proteins made of a Prx II module in the N-terminus fused by a linker peptide to a Grx module situated in the C-terminus [19,20]. The determination of the 3D crystallographic structure of the hybrid protein from *Haemophilus influenzae* has revealed the molecular nature of the Prx–Grx interface [21], and provided some clues to explain why the poplar Prx IIB can accommodate both glutaredoxin and thioredoxin as electron donors. Using genetic engineering and based on the linker peptide sequence present in the *Neisseria meningitidis* fusion, Rouhier et al. have engineered artificial fusion proteins between Prx IIB and either Grx or Trx [22]. All fusion proteins were correctly folded and catalytically active suggesting that they can be good candidates for establishing the 3D structure of the complex between the two proteins.

The regeneration system used by glutathione peroxidases is also a complex story. In animal cells, there are several glutathione peroxidases and interestingly some of them contain a selenocysteine (SeCys) at the active site rather than a regular cysteine [23,24]. The SeCys-containing enzymes are all reduced by glutathione and they possess catalytic activities three orders of magnitude superior to those of glutathione peroxidases containing cysteines at the active site [24]. While SeCys-containing Gpx are present in green algae, the SeCys machinery has apparently been lost in land plants where all Gpxs characterized so far are of the cysteine type [25]. Accordingly, land plant Gpxs display lower catalytic activities than the selenoenzymes. Moreover, they are not reduced via glutathione but through thioredoxin [10,26]. The thioredoxin-linked regeneration of poplar Gpxs has been studied in detail by Navrot et al. who have shown that following oxidation of the catalytic cysteine into a sulfenic acid, an intra-subunit

Please cite this article as: N. Rouhier, et al., Redox based anti-oxidant systems in plants: biochemical and structural analyses, Biochim. Biophys. Acta (2008), doi:10.1016/j.bbagen.2007.12.007

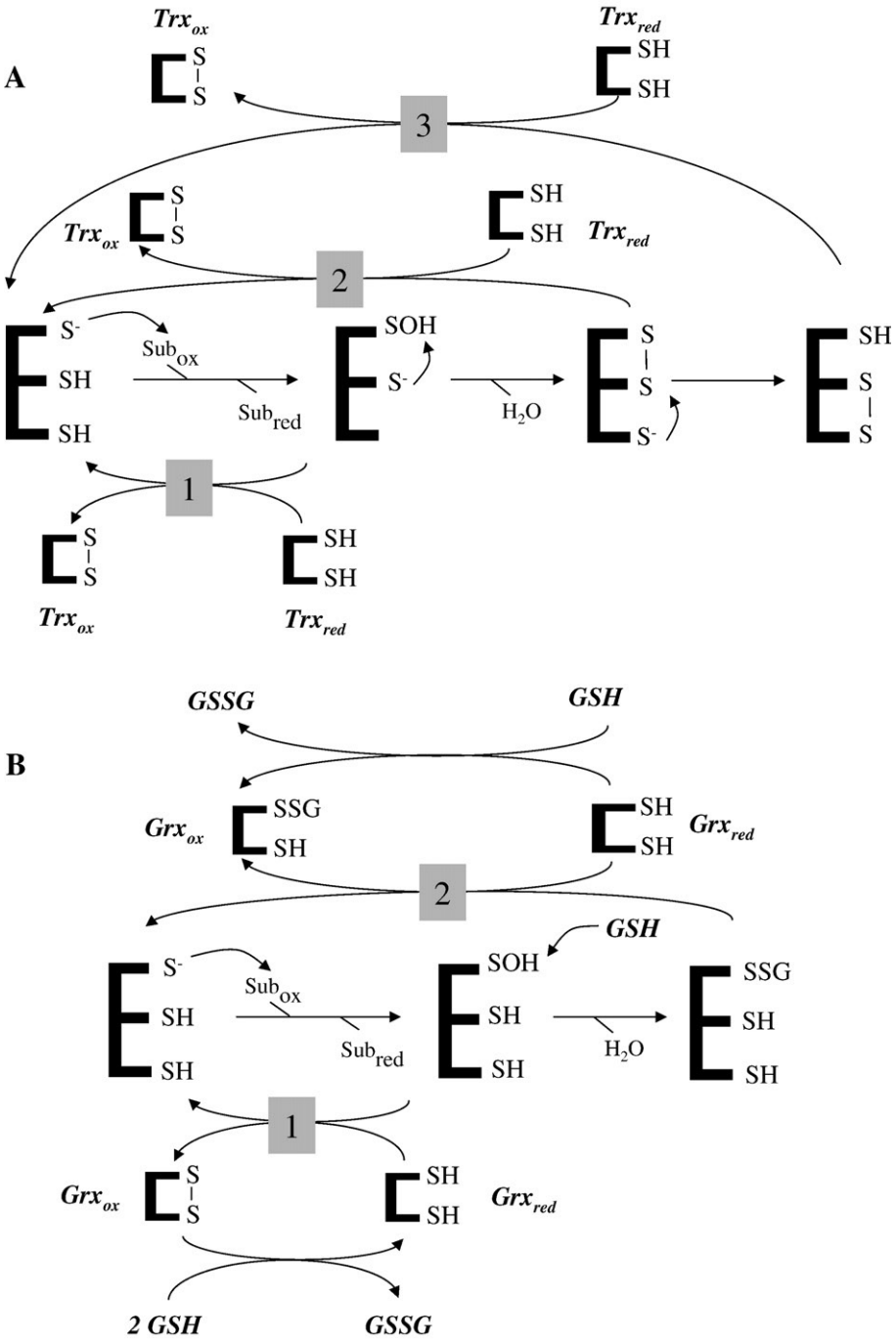


Fig. 2. Catalytic and regeneration mechanisms of Prxs and Msrs. A. Trx-mediated recycling of Msrs or Prxs. The reduction of the oxidized substrate (either ROOH or MetSO) leads to the formation of a sulfenic acid on the catalytic cysteine. It has been shown or proposed that the sulfenic acid can directly be attacked by Trx, especially when the Prx or Msr do not possess any recycling cysteine (pathway 1). The alternative regeneration pathways are linked to the presence of one (pathway 2) or two recycling cysteines (pathway 3) in Prxs or Msrs. B. Grx-mediated recycling of Msrs or Prxs. As in A, once the sulfenic acid formed on the enzyme, it can be regenerated either via direct reduction by Grxs, which most likely leads to the formation of an intra-molecular disulfide bridge on the Grx which is then reduced using two glutathione molecules (pathway 1). An alternative pathway more accepted is the attack of the sulfenic acid by glutathione forming a glutathionylated protein which is reduced either by another glutathione molecule (not represented here) or by Grxs (pathway 2). In this case, there is no intra-molecular disulfide formed onto the Grx and the newly glutathionylated Grx is reduced using one glutathione molecule. Finally, other options not presented here, but shown in the pathways 2 and 3 of the part A, is the reduction of a disulfide bridge in a manner similar to Trxs.

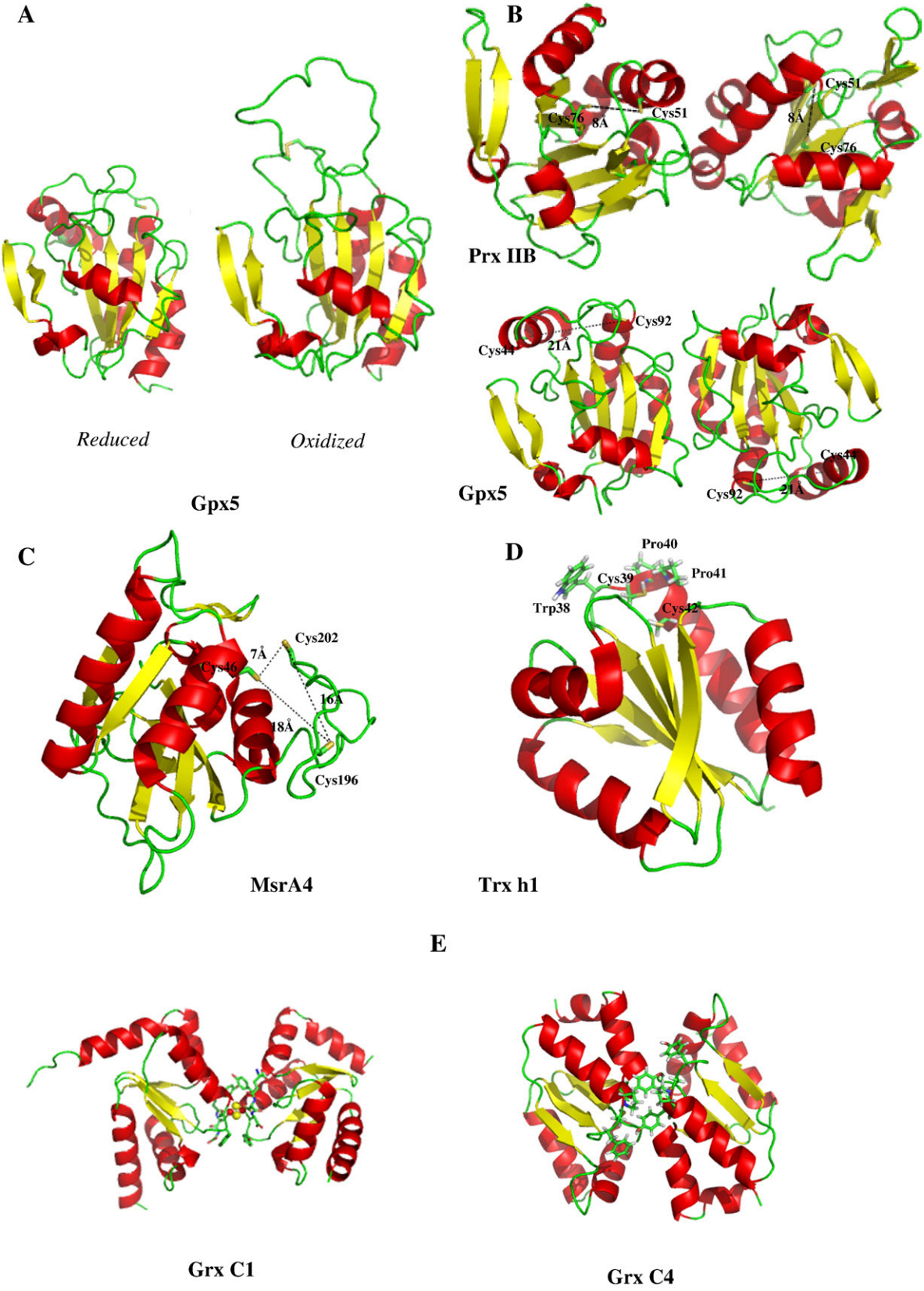
disulfide bond is formed between the catalytic and the resolving cysteines (Cys44 and Cys92, poplar Gpx5 numbering) prior to reduction by thioredoxin. In cyanobacteria, two NADPH-dependent Gpxs have also been described [27].

2.2. Structural considerations

The general Prx topology is β_1, β_2, α_1, β_3, α_2, β_4, α_3, β_5, α_4, β_6, β_7, α_5, with some N-terminal or C-terminal extensions or

Please cite this article as: N. Rouhier, et al., Redox based anti-oxidant systems in plants: biochemical and structural analyses, Biochim. Biophys. Acta (2008), doi:10.1016/j.bbagen.2007.12.007

insertions in some Prx or Gpx families [28]. Within a monomer, the β−strands organize in a central β−sheet surrounded by α helices much as in the canonical structural model of thioredoxin. The quaternary structure within the thiol-peroxidase family is dependent on the redox and oligomerisation state. Except for Prx Q or homologous proteins, which are monomeric



Please cite this article as: N. Rouhier, et al., Redox based anti-oxidant systems in plants: biochemical and structural analyses, Biochim. Biophys. Acta (2008), doi:10.1016/j.bbagen.2007.12.007

enzymes, most other are homodimeric proteins, which can in some conditions form higher order oligomers such as octamers, decamers or even dodecamers [for a review see [28]]. As far as the dimer is concerned, two types of interfaces have been described, either parallel or perpendicular to the central β−sheets of molecules.

Site-directed mutagenesis, biochemical approaches and crystallographic study by our group shed light on the catalytic mechanism of plant Gpxs, by identifying both the peroxidatic cysteine (Cys44) and the resolving cysteine (Cys92) [10,29]. Resolutions of both the reduced (PDB code 2P5Q) (Fig. 3A) and the oxidized form (PDB code 2P5R) structures of poplar Gpx5 revealed unambiguously the presence of the intra-molecular disulfide, the oligomerization nature of the enzyme and the possibility of this enzyme as metal sink [29]. In the same study, the structure of the reduced form was also obtained (PDB code 2P5Q) indicating that the reduced to oxidized redox state transition is accompanied by a striking rearrangement in the protein itself, which involves the total unwinding of the helix α_2 and the unwinding of a helix turn of α_1. These extensive reorganizations of the secondary structure are necessary to bring the two far away cysteines to form an intra-molecular disulfide. Poplar Gpx subunits are arranged as a dimer with a peculiar anti-parallel arrangement of the 5 β-strands at the dimerization interface [29]. In addition to the thioredoxin fold, this enzyme possesses two supplementary β-strands (namely β_{N1} and β_{N2} respectively) at the most N-terminal end of the polypeptide. Fig. 3B shows a comparison of the poplar Gpx homodimer and of the poplar Prx IIB homodimer (PDB code 1TP9) [30]. The interface between the two dimers is parallel to the central β-sheet plane in Gpx and perpendicular in Prx IIB. The monomer-monomer contacts between subunits in both cases are mainly hydrophobic. In the reduced form, the distance between the Sγ atoms of cysteines, both in Prx IIB or Gpx is too large (around 8 Å in Prx IIB and 21 Å in Gpx) to form a disulfide without a large structural rearrangement (Fig. 3B) and this has been observed in many other Prx structures. However, these conformational changes are subtle as compared to the one reported for poplar Gpx5. While such a drastic conformational change has been shown for Gpxs, it is presently not clear whether an intra-molecular disulfide can be formed in Prx IIB.

2.3. Functionality in cells: peroxidase, chaperone or signal transducer activity

The low catalytic turnover of these thiol-containing peroxidases has suggested that they might play additional roles rather than merely as peroxidases and especially that they could be involved in signal transduction [31,32]. Although this argument is certainly valid, the low catalytic efficiency generally measured for Prxs (from 10^3 to 10^5 M^{-1} S^{-1}) could be due to the method of measurement used, i.e. a spectrophotometric test linked to NADPH oxidation, which actually measures the rate of regeneration by thioredoxins or glutaredoxins. It has been demonstrated for Msr for example that this step is by far the rate-limiting one [33]. As Prxs and Msrs use similar mechanisms (see below and Fig. 2), it is likely that the activity measured is underestimated and does not really reflect the intrinsic Prx activity. Some recent studies, using an alternative activity test, have demonstrated that the catalytic efficiency of Prxs should be in the range of 10^7 M^{-1} s^{-1} [34]. With this high catalytic efficiency and their high abundance in the cells, thiol-peroxidases could thus be much more efficient than initially envisaged, and moreover they are likely to degrade a broad variety of substrates, including peroxidized lipids and peroxynitrite.

Prxs have been recently suggested to be involved in signal transduction and this concept was initially derived from the description of the overoxidation of the catalytic cysteine of eukaryotic 2-Cys Prxs into sulfinic or sulfonic acids and the slow reduction of sulfinic acids into sulfenic acids by sulfiredoxin [5]. The “floodgate” theory basically proposed that the accumulation of overoxidized inactive Prxs at high peroxide concentration allows signalling via H_2O_2 [35]. A structural analysis has shown that eukaryotic 2-Cys Prxs possess two additional structural motifs as compared to prokaryotic enzymes, a GGLG insertion in the loop between α_3-β_4 and a YF extension at the C-terminal arm, which prevent a fast recycling of the peroxidatic cysteine by the resolving cysteine and hence allow the subsequent reaction with a second peroxide [35]. Another well described example is the description in fungi of the interaction of thiol-peroxidases such as *Saccharomyces cerevisiae* Gpx3 or *Schizosaccharomyces pombe* thioredoxin peroxidase with stress responsive transcription factors of the AP1 family [31,36]. Yap1 is activated by oxidation through Gpx3 when hydroperoxide levels increase and thioredoxin turns off the pathway by reducing both the sensor and regulator.

Finally, based on the identification of a switch from low to high molecular weight Prx complexes under stress situations (oxidative stress and heat shock), a chaperone function has been described for two cytosolic yeast Prxs [37].

3. Protein oxidation: repair mechanism by methionine sulfoxide reductases (Msr)

Methionine sulfoxide reductases are enzymes that reduce methionine sulfoxide (MetSO) back to methionine using a thiol-

Fig. 3. Selected 3D structures of components of the poplar redox-based antioxidant systems. A: Monomer of poplar “glutathione” peroxidase 5, left reduced (2P5Q), right oxidized (2P5R). The transition from the reduced to the oxidized state is accompanied by the total unwinding of helix α_2. B: Comparison of the dimeric structures of poplar Prx IIB (1TP9) (top) and Gpx5 (2P5Q) (bottom) in the reduced forms. The distance between the sulphur atoms of cysteines 44 and 92 in the Gpx reduced form and between Cys 51 and 76 in Prx IIB are around 21 Å and 8 Å respectively and are shown in dashed line. Note the differential arrangement of the β-sheets in both dimers and the additional two β-hairpins at the N-terminal parts of the two enzymes. C: Methionine sulfoxide reductase A4 (2J89) (reduced form). The three catalytic cysteines are located, together with the distance between their sulphur atoms. The catalytic cysteine is Cys46 and the recycling cysteines Cys196 and Cys202. D: Mean NMR structure of poplar thioredoxin h1 (1TI3) in the reduced form. The protein has an atypical WCPPC active site. The positions of the Trp, Cys and Pro residues at the top of helix α_2 have been indicated. E: Comparison of the dimeric structures of Grx C1 (2E7P) (active site CGYC) left, and Grx C4 (active site CPYC) right. Grx C1 bridges an iron sulphur center between the two subunits. In Grx C4, the side chain of the proline residue at the active site prevents this incorporation.

Please cite this article as: N. Rouhier, et al., Redox based anti-oxidant systems in plants: biochemical and structural analyses, Biochim. Biophys. Acta (2008), doi:10.1016/j.bbagen.2007.12.007

based regeneration mechanism [33,38]. The existence of two different enantiomers of the sulfoxide function of MetSO coincides with the existence of two different subgroups of enzymes, called MsrA and MsrB, MsrA being able to reduce the S-MetSO form and MsrB the R-MetSO. These two classes are very divergent in terms of primary and tertiary structures. As methionine oxidation by ROS does not always lead to protein inactivation, cyclic oxidation/reduction of methionine has been proposed to be part of the antioxidant system by eliminating excess ROS [39].

3.1. Catalytic and regeneration mechanisms

In several aspects, the catalytic and regeneration mechanisms used by Msrs are similar to those described for Prxs (Fig. 2A and B). Indeed, the first step, consisting of a nucleophilic attack of the substrate by a reactive catalytic cysteine followed by the rearrangement of the resulting intermediate, leads to the formation of a sulfenic acid on that cysteine. Its regeneration mechanism is generally dependent on the presence of recycling cysteines, leading to the formation of a disulfide subsequently reduced by an external reductant. In fact, there are some MsrA or MsrB which do not possess additional cysteines, and the mechanism associated with the sulfenic acid reduction in these enzymes remains obscure. Nevertheless, it has been demonstrated recently that a plant plastidial MsrB1 is regenerated by an atypical chloroplastic bimodular Trx, called CDSP32 (chloroplast drought induced protein of 32 kDa) or by glutaredoxin but not by traditional monomodular Trx [40]. The reason is yet unclear but it could be related to the compatibility of molecular surface charges in CDSP32 or glutaredoxin but absent in thioredoxin which may be essential for the recognition of MsrB1. Based on those observations, glutaredoxin could be a possible reductant for Msrs lacking recycling cysteines. Alternative reductants such as selenocompounds or thionein have also been proposed based on experiments with some human MsrB which are poorly reduced by Trx [41,42]. Regardless of the Msr family considered, when a single recycling cysteine is present, besides the catalytic cysteine, the sulfenic acid formed on the catalytic cysteine is attacked, leading to the formation of an intra-molecular disulfide. This latter disulfide is reduced by thioredoxin [43–45]. To date, unlike some Prx subgroups, there is no description of an inter-molecular disulfide in Msrs. Finally when two recycling cysteines are present, two successive intra-molecular disulfide bonds are formed. This is the case for some MsrAs studied (*E. coli*, *Bos taurus*) and in particular for the plant MsrAs characterized so far [43,46]. Nevertheless, the position and the order in which the recycling cysteines are involved are different between plants and the other organisms [46]. In addition, plant MsrAs are equally well regenerated using a glutathione/glutaredoxin reducing system than using a thioredoxin system, which may not be the case for other organisms [Rouhier, unpublished observations]. This observation is supported by the existence of an active protein in *Gracilaria gracilis* in which two N-terminal Grx modules are linked to a MsrA domain present in the C-terminus. The recombinant protein can reduce MetSO at the expense of glutathione [Rouhier, unpublished observations].

3.2. Structural informations

As the *E. coli* enzyme, plant MsrA is a monomeric protein and its topology is β_1, α_1, β_2, $\alpha_{1'}$, β_3, α_2, $\beta_{3'}$, $\beta_{3''}$, β_4, α_3, β_5, β_6, while the MsrB from *Neisseira gonorrhoeae*, also a monomeric protein, has a topology of the form α_1, α_2, 3_{10}, β_1, β_2, β_3, β_8, β_9, α_3, 3_{10} [47,48]. The catalytic cysteine is just before the α_1 helix in MsrA while in MsrB, it is included in the β_8 strand. In MsrA, the position of the recycling cysteine(s) is variable and it is very often located in the unstructured C-terminal region, far from the active site. Therefore, the formation of the disulfide bridge(s) necessary to regenerate MsrA should require a large structural rearrangement. All those observations are illustrated in the reduced poplar MsrA4 structure (PDB code 2J89), which is the only plant Msr the structure of which has been solved to date (Fig. 3C) [46].

When compared to other known non-plant MsrA structures, some features seem specific to plant enzymes. The C-terminal end is much more structured as compared to other known MsrA. In all MsrAs using two recycling cysteines characterized so far but plant MsrA, the two cysteines, located in the C-terminal part of the protein, are surrounded by many glycinyl residues [48]. In addition, a GYC motif (or sometimes GYxC in MsrA using only one recycling cysteine), is present before the first C-terminal cysteine and acts as the first recycling cysteine in the regeneration mechanism. This conserved tyrosine residue is very important for substrate recognition [49]. In poplar MsrA and in most plant MsrAs, there are no such motifs but instead the GYC motif is reversed and includes the most C-terminal cysteine in a highly conserved sequence (C_{196}NDPIRC_{202}YG), supporting the proposal that Cys202 actually acts as the first recycling cysteine [46]. Nevertheless, although the biochemical analyses using mutated proteins support the involvement of Cys202 as the first recycling cysteine, its Sγ atom is found 7.13Å away from the catalytic cysteine Cys46 (Fig. 3C). However, structural changes required to displace Cys202 to the proximity of Cys46, as compared to those needed to move Cys 196, are somehow less important. The resolution of an oxidized MsrA structure would thus be crucial to understand the transition from the reduced active form to the oxidized inactive form.

4. The thioredoxin-dependent regenerating systems

Thioredoxins are small ubiquitous disulfide oxidoreductases considered to be the best reductants in the cell. Its small size, compact globular structure and high stability have prompted its widespread utilization in biotechnology, especially as a fusion module that enhances both the solubility and stability of recombinant proteins. In animal and bacterial cells there are multiple thioredoxins which are reduced via NADPH and a thioredoxin reductase [50,51]. In plants, the thioredoxin system is particularly developed in comparison to non-photosynthetic organisms. For instance, in *Arabidopsis thaliana*, at least 20 genes encoding “true thioredoxins” with a conserved catalytic

Please cite this article as: N. Rouhier, et al., Redox based anti-oxidant systems in plants: biochemical and structural analyses, Biochim. Biophys. Acta (2008), doi:10.1016/j.bbagen.2007.12.007

site WC[G/P]PC, have been detected in the whole sequenced genome [52–54]. These include the *f*, *m*, *x* and *y* types in the chloroplasts, the o type in mitochondria and the *h* type in the cytosol and mitochondria. In chloroplasts, besides their function in the ROS detoxification system, thioredoxins are involved in the light regulation of carbon metabolism through fine-tuning of the reducing pentose phosphate pathway and also of the C_4 pathway of photosynthesis [6]. In this regulatory pathway, the reducing power provided by light is transferred through ferredoxin and FTR to thioredoxin. In addition, a fusion protein between a thioredoxin reductase and a thioredoxin module called NTRc has been shown to be expressed in plastids, which increases even more the complexity of the plastidial Trx system [55]. In contrast, in cytosolic and mitochondrial systems, the reducing power from NADPH is mainly transferred to thioredoxins via a specific NTR [56,57].

4.1. Catalytic mechanism

Trxs perform dithiol–disulfide exchange reactions in which the dithiol form of a Trx attacks the disulfide of a target protein. In classical Trx, the two cysteines of the active site are non equivalent, the one next to the conserved Trp residue and closer to the N-terminal part of the protein is the catalytic one, with the pKa of its thiol group being lowered in part by the presence of a conserved buried Asp residue [58]. This thiol group is thus easily deprotonated at neutral pH (in thiolate form) and prone to attack a target disulfide, creating an intermediate heterodisulfide between Trx and its target. The half-life/lifespan of this disulfide can be extended indefinitely by mutating the second cysteine (back-up cysteine) of the active site into a serine residue for example. This has been the basis of a number of papers where Trx targets have been identified by proteomics [59]. With a dithiol Trx, the heterodisulfide is reduced by the back-up cysteine, leaving the target in the dithiol form and Trx in the oxidized state with a disulfide bond. The reduced state of Trx is then regenerated from NADPH or photoreduced ferredoxin in the plastids.

4.2. Structures and interactions with reductases and targets

Structures of thioredoxins from photosynthetic organisms have been solved using either NMR or X-ray crystallography (spinach thioredoxin m 1FB6 and 1FB0, *Chlamydomonas* thioredoxin m 1DBY, spinach thioredoxin *f* 1F9M, poplar thioredoxin h1 1TI3, *Chamydomonas* thioredoxin *h* 1EP7, 1EP8 and 1TOF, barley thioredoxin h2 2IWT, rice thioredoxin *h* 1WMJ, *Arabidopsis* thioredoxin *h* 1XFL) [60–67]. In all cases, the following succession of secondary structures is conserved: β_1, α_1, β_2, α_2, β_3, α_3, β_4, β_5 and α_4. In the oxidized form of the proteins, the two catalytic cysteines are exposed and localized in a protruding loop between β_2 and α_2 including the first turn of α_2. The catalytic site of thioredoxin also contains a conserved tryptophan, which has been shown to be important for protein–protein interactions [68]. Another conserved buried Asp residue, in which the carboxyl group exhibits a particularly high pKa has been shown to be important for the catalytic efficiency of plant thioredoxins similarly to what was observed in other organisms [65]. This latter conserved residue could be involved together with a water molecule in the protonation/deprotonation mechanism of the back-up cysteine.

In the thioredoxin of class *h*, several isoforms differ from the canonical WCGPC thioredoxins, exhibiting the variant active site WCPPC [69]. Biochemical studies have shown that these plant variants display similar catalytic efficiency both with NTR and with potential physiological targets such as peroxiredoxins or "glutathione" peroxidases. Structural data obtained from the poplar enzyme PtTrx*h*1 suggest that the G to P mutation does not give rise to important conformational changes but leads to more rigid structural characteristics [67]. Nevertheless, using yeast complementation experiments, it has been shown that mutating the active site of an *A. thaliana* Trx *h*3 from WCPPC to WCGPC, allowed cells to grow on a medium devoid of organic sulphur source (one of the phenotype associated to the deletion of the yeast thioredoxins), suggesting that in some cases, minor structural changes can be associated to a loss or gain of function [70]. On the contrary, both AtTrx *h*2 with WCGPC and WCPPC active sites are able to restore growth on sulphate.

One area where structural progress related to chloroplastic thioredoxins has been remarkable is the elucidation of the 3D structures of the two regulatory enzymes fructose-1,6-bisphosphatase and NADP malate dehydrogenase [71,72]. However, the molecular contacts between chloroplastic thioredoxins and these enzymes still need to be refined by the elucidation of heterostructures. In this respect, the isolation of a covalent heterostructure containing both thioredoxin *f* and fructose-1,6-bisphosphatase is a technological advance of considerable interest [73]. In contrast, the functioning, the structure and the interactions of thioredoxin with FTR are now fully characterized at the molecular level, since the structure of the ternary complex involving ferredoxin, FTR and thioredoxin *f* has been recently solved (PDB code 2PVO) [74]. FTR, found only in oxygenic photosynthetic organisms, is a heterodimeric protein, exhibiting a variable and a catalytic subunit. The catalytic subunit containing the [4Fe-4S] cluster interacts with reduced ferredoxin leading to an intermediate where an iron atom is penta-coordinated and a cysteinyl residue free to further reduce Trx disulfide bridge [74]. After a second electron transfer from reduced ferredoxin, the second FTR catalytic cysteinyl residue involved in the iron coordination becomes reduced and allows the release of reduced thioredoxin. FTR catalyses the reduction of the chloroplastic thioredoxins *f* and *m* involved in the regulation of key carbon-fixation enzymes. As indicated above, chloroplasts contains also the thioredoxins *x* and *y* which are rather involved in responses of oxidative stress, since they are particularly efficient in the reduction of some Prxs [11,75], but their 3D structures and molecular interactions with FTR or with targets are presently unknown.

The interactions of plant thioredoxins with NTRs remain poorly understood at the molecular level in comparison to their bacterial and mammal counterparts [76,77], but nevertheless a crystallographic structure of an *A. thaliana* NTR has been described (PDB code 1VDC) [78]. Interestingly, FTR and NTR although they are extremely different in their amino acid sequences and subunit composition, have apparently converged towards a similar 3D structure if one compares the heterodimer

Please cite this article as: N. Rouhier, et al., Redox based anti-oxidant systems in plants: biochemical and structural analyses, Biochim. Biophys. Acta (2008), doi:10.1016/j.bbagen.2007.12.007

of FTR to the monomer of NTR (NTR is a homodimer). In NTR, the subunit contains two domains which organize in a similar way to the two subunits of FTR with the FAD and disulfide at the junction of the domains, similar to what is observed in FTR where the iron sulphur center and the disulfide are in a similar position facing each other at the junction of the two subunits of the heterodimer.

4.3. Cross-talk and glutathionylation

The cytosolic and mitochondrial thioredoxins are mainly reduced by NTR. However, thioredoxins belonging to the third group of the class *h* have be shown to interact with glutaredoxins suggesting a cross-talk between the different networks involved in the cellular redox regulation [79]. Likewise, the study of thioredoxin reductase knock-out lines could indicate that some Trx *h* are likely to be reduced in this context by glutathione and/or glutaredoxins [80]. Besides this interaction, it has been shown *in vivo* that a human thioredoxin [81] and *in vitro* that some plant thioredoxins can be glutathionylated on cysteines not present at the active site. This post-translational modification has been shown to reduce the catalytic efficiency of Trx *f* during its interaction with several protein targets such as GAPDH or NADP-MDH [82]. In a similar way, the glutathionylation of mitochondrial Trx*h*2 from poplar leads to an alteration of its interaction with NTR [83].

5. The glutaredoxin-dependent regenerating systems

Glutaredoxins are also small ubiquitous oxidoreductases belonging to the large Trx superfamily. The global active site sequence of glutaredoxins can be defined as CxxC or CxxS. Nevertheless, a careful examination of the plant genomes available and phylogenetic analyses led to the classification of around 30 Grx isoforms into three distinct subgroups with respect to the active site sequences, either CxxC/S (subgroup I) or CGFS (subgroup II) or CCxC/S (subgroup III) [84,85].

5.1. Catalytic and mechanistic properties

Compared to Trxs, Grxs are considered to be less efficient reductants, because they possess a higher redox potential (around −170 mV compared to −300 mV for Trxs) [86]. This is probably true for many dithiol–disulfide exchange reactions, but on the other hand, the interaction between redoxins and their targets is also an important aspect to consider. In addition, Grxs possess the ability to catalyze very efficiently deglutathionylation reactions, i.e. the reduction of a protein-glutathione mixed disulfide, reactions for which Trxs are poor catalysts (Fig. 2). Other putative candidates are sulfiredoxins but more convincing experimental proofs are still required [87]. Although protein glutathionylation may proceed spontaneously, the cells might also need catalysts that are able to do the reverse reaction, but experimental evidence is still lacking. The human Grx1 was shown to catalyze this reaction via glutathione thiyl radical [88]. Other candidates for this step are glutathione-*S*-transferases or sulfhydryl oxidases [88].

Considering these two different reactions, three different mechanisms seem to operate for plant Grxs. Concerning dithiol containing Grxs (essentially in subgroup I), the dithiol–disulfide exchange reactions require the involvement of two cysteines, most probably the two active site cysteines. However for some other reactions such as the reduction of hydroxyethyl disulfide or of dehydroascorbate, only the most N-terminus localized active site cysteine is required [89]. For Grxs with a monothiol active site, it cannot be excluded that extra active site cysteines can sometimes be involved. The deglutathionylation reaction necessitates one or two cysteines. The well-established process is the removal of the glutathione moiety from the target protein via the first Grx active site cysteine, which becomes itself glutathionylated, before another glutathione molecule reduces this disulfide. This possibility is described in Fig. 2B. An alternative has recently been proposed for a plant Grx with a monothiol CGFS active site, but with an additionally conserved C-terminal cysteine [90]. In this case, the C-terminal cysteine attacks the active site glutathionylated Grx, leading to the formation of an intra-molecular disulfide, which is then reduced by FTR.

5.2. Structural informations

From a structural point of view, only two structures of plant Grxs of subgroup I (Grx C1 and C4) have been investigated both by NMR and X-ray crystallography (PDB codes 1Z7P, 1Z7R, 2E7P) [86,91,92]. The arrangement of secondary structures is quite similar to those of Trx, with an α_1, β_1, α_2, β_2, α_3, β_3, β_4, α_4, α_5 organisation and the four β−strands forming a β−sheet flanked by five helices [91,92]. Compared to most prokaryotic Grxs, the poplar Grxs and other eukaryotic Grxs possess two supplementary α−helices in the N and C-terminal regions. As in thioredoxin, the first active site cysteine is found at the beginning of the helix α2 and exposed to the solvent while the second cysteine, when present, is more buried. The Grx C4 can specifically self-associate into a monomer-dimer equilibrium, with the auto-association surface comprising the active site and the GSH binding site [92]. This dimer interface is different from the only one described so far in the *H. influenzae* Prx–Grx hybrid protein, essentially because of the presence of this extra N-terminal helix [21,92]. The Prx–Grx interface of the HiPrxGrx protein is stabilized by electrostatic interactions involving negatively charged residues in the Prx domain and positively charged residues in the Grx domain. Those amino acid residues are conserved both in the poplar Grx C4 and Prx IIB, but the two important residues in Grx C4, a lysine in position 24 and an arginine in position 69, are located at the auto association surface, probably preventing the interaction with Prx when and if the protein is organized as a dimer. Although the formation of a homodimer is possibly not favored in the cell, the presence of this newly described interface supports other results obtained later with Grx C1, which bridged an iron sulphur cluster into a homodimer via two catalytic cysteines and two external glutathione molecules [86,91]. Nevertheless, the quaternary structure is different since in Grx C4, the monomers are arranged in a head-to-tail orientation, while in Grx C1 they are in a mirrored conformation (Fig. 3E). Among the poplar

Please cite this article as: N. Rouhier, et al., Redox based anti-oxidant systems in plants: biochemical and structural analyses, Biochim. Biophys. Acta (2008), doi:10.1016/j.bbagen.2007.12.007

subgroup I Grxs, only Grx C1 forms an iron–sulfur cluster. From mutagenesis results and analysis of the structures, it appeared that the side chain of the active site proline residue in Grx C2, C3, or C4 (CPFC or CPYC active sites) likely prevents the incorporation of an iron–sulfur cluster [86]. Indeed, the replacement of this proline by a glycine in Grx C2, C3 and C4, allows incorporation of the cluster as in Grx C1 [86]. In light of these results, the presence of a small residue, and especially a glycine and the role played by the two tyrosine residues surrounding the active site (illustrated in Fig. 3E), allow to delimitate a small pocket essential for Fe–S cluster incorporation.

6. Integrated functioning of the redox-dependent reactions in sub-cellular compartments

Based on the biochemical and functional analyses described above and elsewhere, a tentative comprehensive scheme can be proposed to describe the complexity of the protein–protein interactions and molecular functioning of redox-dependent antioxidant systems in a given sub-cellular compartment (Fig. 4).

The complexity and redundancy of those systems seem to be exacerbated in the chloroplast. Indeed, there are at least eleven conventional thioredoxins [93], the specificity of which vs Prxs and Msrs has been studied quite extensively. Among the thiol-peroxidase family, the 2-Cys Prx are most likely reduced by NTRc, Trx *x* and CDSP32; the Prx Q would rather be reduced by CDSP32 or Trxs y1 or y2; Gpxs almost exclusively by Trx *y*1 or *y*2 and Prx IIE by Grxs [[10,11,55,75,94,95], Rouhier unpublished results]. Among the Msr family, the regeneration of MsrB2 would occur mainly through Trx *y*1 or *y*2 or Trxs *m*, while MsrB1 is only reduced by CDSP32 and Grx S12 [40]. MsrA4 is reduced *in vitro* by Grxs and cytosolic Trxs *h*, but chloroplastic Trxs could also probably catalyse this reaction although their specificity is unknown for the moment. In general, the preferential electron donor for a given enzyme is not always clearly established and even though there is often more than a single candidate for reduction. In addition, the physiological electron donors might be different depending on the organs or developmental stages considered.

In mitochondria, the system is less developed with only two thiol-peroxidases identified so far and no Msrs. The Prx IIF was shown to accept either Trx *o* or glutathione, alone or together with Grx S10, as electron donors [15,16]. On the other hand, the reductant for Gpx 3.2 is not known as the Trx *o* and Trx *h*2 do not support its regeneration [83].

In the cytosol, poplar Gpx3.1 only accepts electrons from the Trx *h*1 or *h*3, while poplar Prx IIB can be reduced by those two

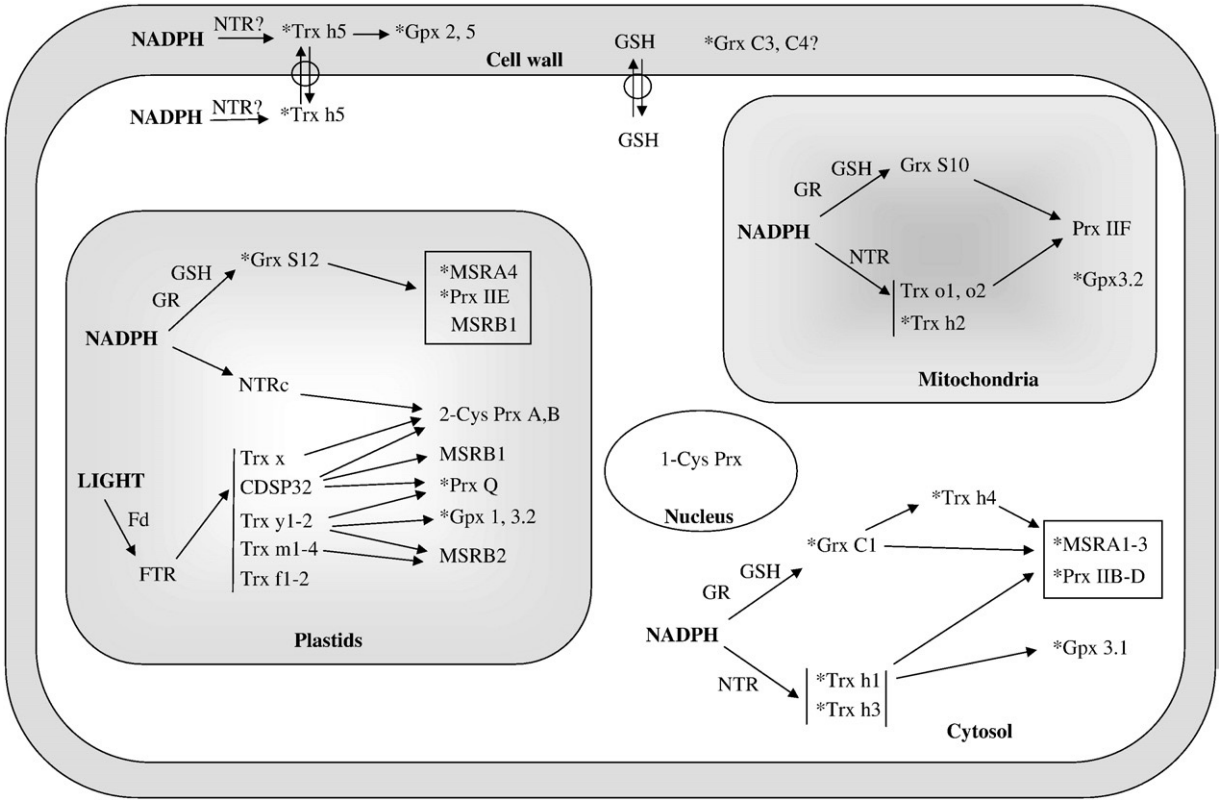


Fig. 4. Present knowledge about the sub-cellular compartmentalization and interactions between reducing and antioxidant systems. The references are indicated in the text. In some cases, only the preferred reductants have been indicated. Most data come from studies with *A. thaliana* or poplar enzymes and sometimes a combination of enzymes from both origins. As the results are sometimes not transposable from one species to another and as the names are sometimes different between these species, we have added an asterisk when poplar nomenclature and proteins have been used. This discrepancy between the *A. thaliana* and poplar models is especially true in the case of the Trx *h* family, while the nomenclature used for Grxs, Msrs and Prxs is similar. Indeed, in *A. thaliana*, there are 8 Trx *h* while in poplar there are only 7 Trx *h* which have been initially numbered in the order of their discovery and not based on the *A. thaliana* nomenclature. The only example represented in this scheme, which is not coming from studies with poplar or *Arabidopsis* proteins is the secretion of a *Nicotiana alata* Trx *h*, homologous to poplar Trx *h*5. We have thus hypothesized that poplar Trx *h*5 could be secreted too and located in the apoplast.

Please cite this article as: N. Rouhier, et al., Redox based anti-oxidant systems in plants: biochemical and structural analyses, Biochim. Biophys. Acta (2008), doi:10.1016/j.bbagen.2007.12.007

Trxs *h*, but also by Grxs and especially Grx C1 [12]. On the other hand, *Arabidopsis* Prx IIB is only reduced by Grxs but not Trxs [14]. Poplar Grxs C2, C3 and C4 are also able to support Prx IIB activity, but localisation predictions indicate that those proteins may be exported by the secretory pathway and thus routed to the apoplast [Rouhier, unpublished results]. The cytosolic MsrA1 to 3, which are very similar to MsrA4, are reduced both by Trxs *h* and Grxs [[46], Rouhier unpublished results]. Finally, another Trx *h* from poplar, called Trx *h*4, can also reduce cytosolic MsrA but at the expense of the GSH/Grx system and not through the NTR system [79].

The reductant for 1-Cys Prx, that is present in the nucleus, is currently unknown. Nucleoredoxins, proteins composed of three Trx domains, are also localized in the same sub-cellular compartment [96]. Whether they can interact with 1-Cys Prx still remains an open question.

Finally, from the literature, a Trx *h* from *Nicotiana alata,* belonging to the subgroup II of Trxs *h* and homologous to the poplar Trx *h*5, has been shown to be secreted, while Gpx 2 and 5 are also predicted to be secreted [97]. There is thus a possibility that these proteins could interact, especially in the apoplast, although to date the presence of NTR in this compartment remains to be established.

References

[1] N. Navrot, N. Rouhier, E. Gelhaye, J.P. Jacquot, ROS generation and antioxidant systems in plant mitochondria, Physiol. Plant. 129 (2007) 185–195.
[2] C.H. Foyer, G. Noctor, Redox sensing and signalling associated with reactive oxygen in chloroplasts, peroxisomes and mitochondria, Physiol. Plant. 119 (2003) 355–364.
[3] C. Laloi, K. Apel, A. Danon, Reactive oxygen signalling: the latest news, Curr. Opin. Plant Biol. 7 (2004) 323–328.
[4] N. Rouhier, J.P. Jacquot, Plant peroxiredoxins: a multigenic family of thiol peroxidases, Free Radic. Biol. Med. 38 (2005) 1413–1421.
[5] B. Biteau, J. Labarre, M.B. Toledano, ATP-dependent reduction of cysteine-sulphinic acid by *S. cerevisiae* sulphiredoxin, Nature 425 (2003) 980–984.
[6] N. Rouhier, J.P. Jacquot, Plant peroxiredoxins: alternative hydroperoxide scavenging enzymes, Photosynth. Res. 74 (2002) 259–268.
[7] P. Schürmann, J.P. Jacquot, Plant thioredoxin systems revisited, Annu. Rev. Plant Physiol. Plant Mol. Biol. 51 (2000) 371–400.
[8] J. Konig, K. Lotte, R. Plessow, A. Brockhinke, M. Baier, K.J. Dietz, Reaction mechanism of plant 2-Cys peroxiredoxin. Role of the C-terminus and the quaternary structure, J. Biol. Chem. 278 (2003) 24409–24420.
[9] M. Laxa, J. König, K.J. Dietz, A. Kandlbinder, Role of the cysteine residues in *Arabidopsis thaliana* cyclophilin CYP20-3 in peptidyl-prolyl *cis-trans* isomerase and redox-related functions, Biochem. J. 401 (2007) 287–297.
[10] N. Navrot, V. Collin, J. Gualberto, E. Gelhaye, M. Hirasawa, P. Rey, D.B. Knaff, E. Issakidis, J.P. Jacquot, N. Rouhier, Plant glutathione peroxidases are functional peroxiredoxins distributed in several subcellular compartments and regulated during biotic and abiotic stresses, Plant Physiol. 142 (2006) 1364–1379.
[11] N. Rouhier, E. Gelhaye, J.M. Gualberto, M.N. Jordy, E. De Fay, M. Hirasawa, S. Duplessis, S.D. Lemaire, P. Frey, F. Martin, W. Manieri, D.B. Knaff, J.P. Jacquot, Poplar peroxiredoxin Q. A thioredoxin-linked chloroplast antioxidant functional in pathogen defense, Plant Physiol. 134 (2004) 1027–1038.
[12] N. Rouhier, E. Gelhaye, P.E. Sautiere, A. Brun, P. Laurent, D. Tagu, J. Gerard, E. de Fay, Y. Meyer, J.P. Jacquot, Isolation and characterization of a new peroxiredoxin from poplar sieve tubes that uses either glutaredoxin or thioredoxin as a proton donor, Plant Physiol. 127 (2001) 1299–1309.
[13] N. Rouhier, E. Gelhaye, J.P. Jacquot, Glutaredoxin dependent peroxiredoxin from poplar: protein–protein interaction and catalytic mechanism, J. Biol. Chem. 277 (2002) 13609–13614.
[14] C. Brehelin, E.H. Meyer, J.P. de Souris, G. Bonnard, Y. Meyer, Resemblance and dissemblance of *Arabidopsis* type II peroxiredoxins: similar sequences for divergent gene expression, protein localization, and activity, Plant Physiol. 132 (2003) 2045–2057.
[15] I. Finkemeier, M. Goodman, P. Lamkemeyer, A. Kandlbinder, L.J. Sweetlove, K.J. Dietz, The mitochondrial type II peroxiredoxin F is essential for redox homeostasis and root growth of *Arabidopsis thaliana* under stress, J. Biol. Chem. 280 (2005) 12168–12180.
[16] F. Gama, O. Keech, F. Eymery, I. Finkemeier, E. Gelhaye, P. Gardestrom, K.J. Dietz, P. Rey, J.P. Jacquot, N. Rouhier, The mitochondrial type II peroxiredoxin from poplar, Physiol. Plant. 129 (2007) 196–206.
[17] N. Rouhier, E. Gelhaye, C. Corbier, J.P. Jacquot, Active site mutagenesis and phospholipid hydroperoxide reductase activity of poplar type II peroxiredoxin, Physiol. Plant. 120 (2004) 57–62.
[18] V. Noguera-Mazon, J. Lemoine, O. Walker, N. Rouhier, A. Salvador, J.P. Jacquot, J.M. Lancelin, I. Krimm, Glutathionylation induces the dissociation of 1-Cys D-peroxiredoxin non-covalent homodimer, J. Biol. Chem. 281 (2006) 31736–31742.
[19] F. Pauwels, B. Vergauwen, F. Vanrobaeys, B. Devreese, J.J. Van Beeumen, Purification and characterization of a chimeric enzyme from *Haemophilus influenzae* Rd that exhibits glutathione-dependent peroxidase activity, J. Biol. Chem. 278 (2003) 16658–16666.
[20] N. Rouhier, J.P. Jacquot, Molecular and catalytic properties of a peroxiredoxin–glutaredoxin hybrid from *Neisseria meningitidis*, FEBS Lett. 554 (2003) 149–153.
[21] S.J. Kim, J.R. Woo, Y.S. Hwang, D.G. Jeong, D.H. Shin, K. Kim, S.E. Ryu, The tetrameric structure of *Haemophilus influenza* hybrid Prx5 reveals interactions between electron donor and acceptor proteins, J. Biol. Chem. 278 (2003) 10790–10798.
[22] N. Rouhier, F. Gama, G. Wingsle, E. Gelhaye, P. Gans, J.P. Jacquot, Engineering functional artificial hybrid proteins between poplar peroxiredoxin II and glutaredoxin or thioredoxin, Biochem. Biophys. Res. Commun. 341 (2006) 1300–1308.
[23] B. Ren, W. Huang, B. Akesson, R. Ladenstein, The crystal structure of seleno-glutathione peroxidase from human plasma at 2.9 Å resolution, J. Mol. Biol. 268 (1997) 869–885.
[24] J.R. Arthur, The glutathione peroxidases, Cell Mol. Life Sci. 57 (2000) 1825–1835.
[25] S.V. Novoselov, M. Rao, N.V. Onoshko, H. Zhi, G.V. Kryukov, Y. Xiang, D.P. Weeks, D.L. Hatfield, V.N. Gladyshev, Selenoproteins and selenocysteine insertion system in the model plant cell system, *Chlamydomonas reinhardtii*, EMBO J. 21 (2002) 3681–3693.
[26] S. Herbette, C. Lenne, N. Leblanc, J.L. Julien, J.R. Drevet, P. Roeckel-Drevet, Two GPX-like proteins from *Lycopersicon esculentum* and *Helianthus annuus* are antioxidant enzymes with phospholipid hydroperoxide glutathione peroxidase and thioredoxin peroxidase activities, Eur. J. Biochem. 269 (2002) 2414–2420.
[27] A. Gaber, A.M. Tamoi, T. Takeda, Y. Nakano, S. Shigeoka, NADPH-dependent glutathione peroxidase-like proteins (Gpx-1, Gpx-2) reduce unsaturated fatty acid hydroperoxides in *Synechocystis* PCC 6803, FEBS Lett. 499 (2001) 32–36.
[28] V. Noguera-Mazon, I. Krimm, O. Walker, J.M. Lancelin, Protein–protein interactions within peroxiredoxin systems, Photosynth. Res. 89 (2006) 277–290.
[29] C.S. Koh, C. Didierjean, N. Navrot, S. Panjikar, G. Mulliert, J.P. Jacquot, N. Rouhier, A. Aubry, O. Shawakataly, C. Corbier, Crystal structures of poplar thioredoxin peroxidase which exhibits the structure of glutathione peroxidases: Insights into redox driven conformational changes, J. Mol. Biol. 370 (2007) 512–529.
[30] A. Echalier, X. Trivelli, C. Corbier, N. Rouhier, O. Walker, P. Tsan, J.P. Jacquot, A. Aubry, I. Krimm, J.M. Lancelin, X-ray and NMR studies of a homodimeric D (type II) peroxiredoxin reduced by both glutaredoxin and thioredoxin: Prx–Prx interfaces in the Prx family, Biochemistry 44 (2005) 1755–1767.

Please cite this article as: N. Rouhier, et al., Redox based anti-oxidant systems in plants: biochemical and structural analyses, Biochim. Biophys. Acta (2008), doi:10.1016/j.bbagen.2007.12.007

[31] A. Delaunay, D. Pflieger, M.B. Barrault, J. Vinh, M.B. Toledano, A thiol peroxidase is an H_2O_2 receptor and redox-transducer in gene activation, Cell 111 (2002) 471–481.
[32] S. Herbette, P. Roeckel-Drevet, J.R. Drevet, Seleno-independent glutathione peroxidases. More than simple antioxidant scavengers, FEBS J. 274 (2007) 2163–2180.
[33] S. Boschi-Muller, A. Olry, M. Antoine, G. Branlant, The enzymology and biochemistry of methionine sulfoxide reductases, Biochim. Biophys. Acta 1703 (2005) 231–238.
[34] D. Parsonage, D.S. Youngblood, G.N. Sarma, Z.A. Wood, P.A. Karplus, L.B. Poole, Analysis of the link between enzymatic activity and oligomeric state in AhpC, a bacterial peroxiredoxin, Biochemistry 44 (2005) 10583–10592.
[35] Z.A. Wood, L.B. Poole, P.A. Karplus, Peroxiredoxin evolution and the regulation of hydrogen peroxide signalling, Science 300 (2003) 650–653.
[36] A.P. Vivancos, E.A. Castillo, B. Biteau, C. Nicot, J. Ayte, M.B. Toledano, E. Hidalgo, A cysteine-sulfinic acid in peroxiredoxin regulates H_2O_2-sensing by the antioxidant Pap1 pathway, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 8875–8880.
[37] H.H. Jang, K.O. Lee, Y.H. Chi, B.G. Jung, S.K. Park, J.H. Park, J.R. Lee, S.S. Lee, J.C. Moon, J.W. Yun, Y.O. Choi, W.Y. Kim, J.S. Kang, G.W. Cheong, D.J. Yun, S.G. Rhee, M.J. Cho, S.Y. Lee, Two enzymes in one; two yeast peroxiredoxins display oxidative stress-dependent switching from a peroxidase to a molecular chaperone function, Cell 117 (2004) 625–635.
[38] N. Rouhier, C. Vieira Dos Santos, L. Tarrago, P. Rey, Plant methionine sulfoxide reductase A and B multigenic families, Photosynth. Res. 89 (2006) 247–262.
[39] E.R. Stadtman, J. Moskovitz, B.S. Berlett, R.L. Levine, Cyclic oxidation and reduction of protein methionine residues is an important antioxidant mechanism, Mol. Cell. Biochem. 234–235 (2000) 3–9.
[40] C. Vieira Dos Santos, E. Laugier, L. Tarrago, V. Massot, E. Issakidis-Bourguet, N. Rouhier, P. Rey, Specificity of thioredoxins and glutaredoxins as electron donors to two distinct classes of *Arabidopsis* plastidial methionine sulfoxide reductases B, FEBS Lett. 581 (2007) 4371–4376.
[41] D. Sagher, D. Brunell, J.F. Hejtmancik, M. Kantorow, N. Brot, H. Weissbach, Thionein can serve as a reducing agent for the methionine sulfoxide reductases, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 8656–8661.
[42] D. Sagher, D. Brunell, N. Brot, B.L. Vallee, H. Weissbach, Selenocompounds can serve as oxidoreductants with the methionine sulfoxide reductase enzymes, J. Biol. Chem. 281 (2006) 31184–31187.
[43] S. Boschi-Muller, S. Azza, S. Sanglier-Cianferani, F. Talfournier, A. Van Dorsselear, G. Branlant, A sulfenic acid enzyme intermediate is involved in the catalytic mechanism of peptide methionine sulfoxide reductase from *Escherichia coli*, J. Biol. Chem. 275 (2000) 35908–35913.
[44] A. Olry, S. Boschi-Muller, M. Marraud, S. Sanglier-Cianferani, A. Van Dorsselear, G. Branlant, Characterization of the methionine sulfoxide reductase activities of PILB, a probable virulence factor from *Neisseria meningitidis*, J. Biol. Chem. 277 (2002) 12016–12022.
[45] C. Vieira Dos Santos, S. Cuine, N. Rouhier, P. Rey, The *Arabidopsis* plastidic methionine sulfoxide reductase B proteins. Sequence and activity characteristics, comparison of the expression with plastidic methionine sulfoxide reductase A, and induction by photooxidative stress, Plant Physiol. 138 (2005) 909–922.
[46] N. Rouhier, B. Kauffmann, F. Tete-Favier, P. Paladino, P. Gans, G. Branlant, J.P. Jacquot, S. Boschi-Muller, Functional and structural aspects of poplar cytosolic and plastidial type A methionine sulfoxide reductases, J. Biol. Chem. 282 (2007) 3367–3378.
[47] W.T. Lowther, H. Weissbach, F. Etienne, N. Brot, B.W. Matthews, The mirrored methionine sulfoxide reductases of *Neisseria gonorrhoeae* pilB, Nat. Struct. Biol. 9 (2002) 348–352.
[48] B. Kauffmann, A. Aubry, F. Favier, The three-dimensional structures of peptide methionine sulfoxide reductases: current knowledge and open questions, Biochim. Biophys. Acta 1703 (2005) 249–260.
[49] A. Gand, M. Antoine, S. Boschi-Muller, G. Branlant, Characterization of the amino acids involved in substrate specificity of methionine sulfoxide reductase A, J. Biol. Chem. 282 (2007) 20484–20491.
[50] A. Holmgren, Thioredoxin and glutaredoxin systems, J. Biol. Chem. 264 (1989) 13963–13966.
[51] C.H. Lillig, A. Holmgren, Thioredoxin and related molecules—from biology to health and disease, Antioxid. Redox Signal. 9 (2007) 25–47.
[52] E. Gelhaye, N. Rouhier, N. Navrot, J.P. Jacquot, The plant thioredoxin system, Cell Mol. Life Sci. 62 (2005) 24–35.
[53] Y. Meyer, C. Riondet, L. Constans, M.R. Abdelgawwad, J.P. Reichheld, F. Vignols, Evolution of redoxin genes in the green lineage, Photosynth. Res. 89 (2006) 179–192.
[54] J.P. Jacquot, E. Gelhaye, N. Rouhier, C. Corbier, C. Didierjean, A. Aubry, Thioredoxins and related proteins in photosynthetic organisms: molecular basis for thiol dependent regulation, Biochem. Pharmacol. 64 (2002) 1065–1069.
[55] J.M. Perez-Ruiz, M.C. Spinola, K. Kirchsteiger, J. Moreno, M. Sahrawy, F.J. Cejudo, Rice NTRC is a high-efficiency redox system for chloroplast protection against oxidative damage, Plant Cell. 18 (2006) 2356–2368.
[56] J.P. Jacquot, R. Rivera-Madrid, P. Marinho, M. Kollarova, P. Le Marechal, M. Miginiac-Maslow, Y. Meyer, *Arabidopsis thaliana* NAPH thioredoxin reductase. cDNA characterization and expression of the recombinant protein in *Escherichia coli*, J. Mol. Biol. 235 (1994) 1357–1363.
[57] C. Laloi, N. Rayapuram, Y. Chartier, J.M. Grienenberger, G. Bonnard, Y. Meyer, Identification and characterization of a mitochondrial thioredoxin system in plants, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 14144–14149.
[58] H.J. Dyson, M.F. Jeng, L.L. Tennant, I. Slaby, M. Lindell, D.S. Cui, S. Kuprin, A. Holmgren, Effects of buried charged groups on cysteine thiol ionization and reactivity in *Escherichia coli* thioredoxin: structural and functional characterization of mutants of Asp 26 and Lys 57, Biochemistry 36 (1997) 2622–2636.
[59] B.B. Buchanan, Y. Balmer, Redox regulation: a broadening horizon, Annu. Rev. Plant Biol. 56 (2005) 187–220.
[60] G. Capitani, Z. Markovic-Housley, G. DelVal, M. Morris, J.N. Jansonius, P. Schurmann, Crystal structures of two functionally different thioredoxins in spinach chloroplasts, J. Mol. Biol. 302 (2000) 135–154.
[61] N. Coudevylle, A. Thureau, C. Hemmerlin, E. Gelhaye, J.P. Jacquot, M.T. Cung, Solution structure of a natural CPPC active site variant, thioredoxin h1 from poplar, Biochemistry 44 (2005) 2001–2008.
[62] F.C. Peterson, B.L. Lytle, S. Sampath, D. Vinarov, E. Tyler, M. Shahan, J.M. Markley, B.F. Volkman, Solution structure of thioredoxin H1 From *Arabidopsis thaliana*, Protein Sci. 14 (2005) 2195–2200.
[63] I. Krimm, S. Lemaire, E. Ruelland, M. Miginiac-Maslow, J.P. Jacquot, M. Hirasawa, D.B. Knaff, J.M. Lancelin, The single mutation W35A in the 35–40 redox site of *Chlamydomonas reinhardtii* thioredoxin *h* affects the biochemical activity and the pH dependence of C36–C39 1H-NMR but not the redox potential, Eur. J. Biochem. 255 (1998) 185–195.
[64] J.M. Lancelin, L. Guilhaudis, I. Krimm, M.J. Blackledge, D. Marion, J.P. Jacquot, NMR structures of thioredoxin m from the green alga *Chlamydomonas reinhardtii*, Proteins 41 (2000) 334–349.
[65] V. Menchise, C. Corbier, C. Didierjean, M. Saviano, E. Benedetti, J.P. Jacquot, A. Aubry, Crystal structure of the wild-type and D30A mutant thioredoxin *h* of *Chlamydomonas reinhardtii* and implications for the catalytic mechanism, Biochem. J. 359 (2001) 65–75.
[66] V. Mittard, M.J. Blackledge, M. Stein, J.P. Jacquot, D. Marion, J.M. Lancelin, NMR solution structure of an oxidised thioredoxin *h* form the eukaryotic alga *Chlamydomonas reinhardtii*, Eur. J. Biochem. 243 (1997) 374–378.
[67] K. Maeda, P. Hagglund, C. Finnie, B. Svensson, A. Henriksen, Structural basis for target protein recognition by the protein disulfide reductase thioredoxin, Structure 14 (2006) 1701–1710.
[68] V. Menchise, C. Corbier, C. Didierjean, J.P. Jacquot, E. Benedetti, M. Saviano, A. Aubry, Crystal structure of the W35A mutant thioredoxin *h* from *Chlamydomonas reinhardtii*: the substitution of the conserved active site Trp leads to modifications in the environment of the two catalytic cysteines, Biopolymers 56 (2000–2001) 1–7.
[69] E. Gelhaye, N. Rouhier, J.P. Jacquot, The thioredoxin *h* system of higher plants, Plant Physiol. Biochem. 42 (2004) 265–271.
[70] C. Bréhelin, N. Mouaheb, L. Verdoucq, J.M. Lancelin, Y. Meyer, Characterization of determinants for the specificity of *Arabidopsis* thioredoxins *h* in yeast complementation, J. Biol. Chem. 275 (2000) 31641–31647.

Please cite this article as: N. Rouhier, et al., Redox based anti-oxidant systems in plants: biochemical and structural analyses, Biochim. Biophys. Acta (2008), doi:10.1016/j.bbagen.2007.12.007

[71] M. Chiadmi, A. Navaza, M. Miginiac-Maslow, J.P. Jacquot, J. Cherfils, Redox signalling in the chloroplast: structure of oxidized pea fructose-1,6-bisphosphatase, EMBO J. 18 (1999) 6809–6815.
[72] K. Johansson, S. Ramaswamy, M. Saarinen, M. Lemaire-Chamley, E. Issakidis-Bourguet, M. Miginiac-Maslow, H. Eklund, Structural basis for light activation of a chloroplast enzyme: the structure of sorghum NADP-malate dehydrogenase in its oxidized form, Biochemistry 38 (1999) 4319–4326.
[73] Y. Balmer, P. Schurmann, Heterodimer formation between thioredoxin *f* and fructose 1,6-bisphosphatase from spinach chloroplasts, FEBS Lett. 492 (2001) 58–61.
[74] S. Dai, R. Friemann, D.A. Glauser, F. Bourquin, W. Manieri, P. Schurmann, H. Eklund, Structural snapshots along the reaction pathway of ferredoxin-thioredoxin reductase, Nature 448 (2007) 92–96.
[75] V. Collin, E. Issakidis-Bourguet, C. Marchand, M. Hirasawa, J.M. Lancelin, D.B. Knaff, M. Miginiac-Maslow, The *Arabidopsis* plastidial thioredoxins: new functions and new insights into specificity, J. Biol. Chem. 278 (2003) 23747–23752.
[76] T. Sandalova, L. Zhong, Y. Lindqvist, A. Holmgren, G. Schneider, Three-dimensional structure of a mammalian thioredoxin reductase: implications for mechanism and evolution of a selenocysteine-dependent enzyme, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 9533–9538.
[77] B.W. Lennon, C.H. Williams Jr, M.L. Ludwig, Twists in catalysis: alternating conformations of *Escherichia coli* thioredoxin reductase, Science 289 (2000) 1190–1194.
[78] S. Dai, M. Saarinen, S. Ramaswamy, Y. Meyer, J.P. Jacquot, H. Eklund, Crystal structure of *Arabidopsis thaliana* NADPH dependent thioredoxin reductase at 2.5 Å resolution, J. Mol. Biol. 264 (1996) 1044–1057.
[79] E. Gelhaye, N. Rouhier, J.P. Jacquot, Evidence for a subgroup of thioredoxin *h* that requires GSH/Grx for its reduction, FEBS Lett. 555 (2003) 443–448.
[80] J.P. Reichheld, M. Khafif, C. Riondet, M. Droux, G. Bonnard, Y. Meyer, Inactivation of thioredoxin reductases reveals a complex interplay between thioredoxin and glutathione pathways in *Arabidopsis* development, Plant Cell 19 (2007) 1851–1865.
[81] S. Casagrande, V. Bonetto, M. Fratelli, E. Gianazza, I. Eberini, T. Massignan, M. Salmona, G. Chang, A. Holmgren, P. Ghezzi, Glutathionylation of human thioredoxin: a possible crosstalk between the glutathione and thioredoxin systems, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 9745–9749.
[82] L. Michelet, M. Zaffagnini, C. Marchand, V. Collin, P. Decottignies, P. Tsan, J.M. Lancelin, P. Trost, M. Miginiac-Maslow, G. Noctor, S.D. Lemaire, Glutathionylation of chloroplast thioredoxin *f* is a redox signaling mechanism in plants, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 16478–16483.
[83] E. Gelhaye, N. Rouhier, J. Gerard, Y. Jolivet, J. Gualberto, N. Navrot, P.I. Ohlsson, G. Wingsle, M. Hirasawa, D.B. Knaff, H. Wang, P. Dizengremel, Y. Meyer, J.P. Jacquot, A specific form of thioredoxin *h* occurs in plant mitochondria and regulates the alternative oxidase, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 14545–14550.
[84] N. Rouhier, E. Gelhaye, J.P. Jacquot, Plant glutaredoxins: still mysterious reducing systems, Cell. Mol. Life Sci. 61 (2004) 1266–1277.
[85] N. Rouhier, J. Couturier, J.P. Jacquot, Genome-wide analysis of plant glutaredoxin systems, J. Exp. Bot. 57 (2006) 1685–1696.
[86] N. Rouhier, H. Unno, S. Bandyopadhay, L. Masip, S.K. Kim, M. Hirasawa, J. Gualberto, V. Lattard, M. Kusunoki, D.B. Knaff, G. Georgiou, T. Hase, M.K. Johnson, J.P. Jacquot, Functional, structural and spectroscopic characterization of a glutathione-ligated [2Fe–2S] cluster in poplar glutaredoxin C1, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 7379–7384.
[87] V. J Findlay, D.M. Townsend, T.E. Morris, J.P. Fraser, L. He, K.D. Tew, A novel role for human sulfiredoxin in the reversal of glutathionylation, Cancer Res. 66 (2006) 6800–6806.
[88] M.M. Gallogly, J.J. Mieyal, Mechanisms of reversible protein glutathionylation in redox signaling and oxidative stress, Curr. Opin. Pharmacol. 7 (2007) 381–391.
[89] N. Rouhier, E. Gelhaye, J.P. Jacquot, Exploring the active site of plant glutaredoxin by site-directed mutagenesis, FEBS Lett. 511 (2002) 145–149.
[90] N. Rouhier, S.D. Lemaire, J.P. Jacquot, The role of glutathione in photosynthetic organisms: emerging functions for glutaredoxin and glutathionylation, Annu. Rev. Plant Biol. 59 (2008) 143–166.
[91] Y. Feng, N. Shong, N. Rouhier, T. Hase, M. Kusunoki, J.P. Jacquot, C. Jin, B. Xia, Structural insight into poplar glutaredoxin C1 with a bridging iron sulfur center near the active site, Biochemistry 45 (2006) 7998–8008.
[92] V. Noguera, O. Walker, N. Rouhier, J.P. Jacquot, I. Krimm, J.M. Lancelin, NMR reveals a novel glutaredoxin–glutaredoxin interaction interface, J. Mol. Biol. 353 (2005) 629–641.
[93] S.D. Lemaire, L. Michelet, M. Zaffagnini, V. Massot, E. Issakidis-Bourguet, Thioredoxins in chloroplasts, Curr. Genet. 51 (2007) 343–365.
[94] P. Lamkemeyer, M. Laxa, V. Collin, W. Li, I. Finkemeier, M.A. Schottler, V. Holtkamp, V.B. Tognetti, E. Issakidis-Bourguet, A. Kandlbinder, E. Weis, M. Miginiac-Maslow, K.J. Dietz, Peroxiredoxin Q of *Arabidopsis thaliana* is attached to the thylakoids and functions in context of photosynthesis, Plant J. 45 (2006) 968–981.
[95] P. Rey, S. Cuine, F. Eymery, J. Garin, M. Court, J.P. Jacquot, N. Rouhier, M. Broin, Analysis of the proteins targeted by CDSP32, a plastidic thioredoxin participating in oxidative stress responses, Plant J. 41 (2005) 31–42.
[96] B.J. Laughner, P.C. Sehnke, R.J. Ferl, A novel nuclear member of the thioredoxin superfamily, Plant Physiol. 118 (1998) 987–996.
[97] J.A. Juarez-Diaz, B. McClure, S. Vazquez-Santana, A. Guevara-Garcia, P. Leon-Mejia, J. Marquez-Guzman, F. Cruz-Garcia, A novel thioredoxin *h* is secreted in *Nicotiana alata* and reduces S-RNase *in vitro*, J. Biol. Chem. 281 (2006) 3418–3424.
[98] D.A. Mitchell, M.A. Marletta, Thioredoxin catalyzes the *S*-nitrosation of the caspase-3 active site cysteine, Nat. Chem. Biol. 1 (2005) 154–158.

Please cite this article as: N. Rouhier, et al., Redox based anti-oxidant systems in plants: biochemical and structural analyses, Biochim. Biophys. Acta (2008), doi:10.1016/j.bbagen.2007.12.007

JBC Papers in Press. Published on June 14, 2008 as Manuscript M802093200
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M802093200

AN ATYPICAL CATALYTIC MECHANISM INVOLVING THREE CYSTEINES OF THIOREDOXIN

Cha San Koh[¥‖], Nicolas Navrot*[§], Claude Didierjean[¥], Nicolas Rouhier*, Masakazu Hirasawa[#], David B. Knaff[#], Gunnar Wingsle[§], Razip Samian[‖], Jean-Pierre Jacquot*, Catherine Corbier[¤], and Eric Gelhaye*[ж]

From LCM3B, Equipe Biocristallographie, UMR 7036 CNRS-Université Henri Poincaré[¥], Unité Mixte de recherche INRA-UHP 1136, Interactions Arbres/Micro-organismes*, URAFPA, Equipe PB2P[¤], Faculté des Sciences et Techniques, Nancy Université, BP 239, 54506 Vandoeuvre Cedex France; [#]Department of Chemistry and Biochemistry, and Center for Biotechnology and Genomics, Texas Tech University, Lubbock, Texas, 79409-1061, USA; [‖]School of Biology, Universiti Sains Malaysia, Minden, 11800 USM, Penang, Malaysia and the [§]Department of Forest Genetics and Plant Physiology, Faculty of Forestry, Swedish University of Agricultural Sciences, SE-901 83 Umea, Sweden.

Running head: Glutaredoxin-dependent thioredoxin

Address correspondence to : Eric Gelhaye, Unité Mixte de recherche INRA-UHP 1136, Interactions Arbres/Micro-organismes, Faculté des Sciences et Techniques, Nancy Université, BP 239, 54506 Vandoeuvre Cedex France. Tel : +33 (0)3 83 68 42 28; E-mail: gelhaye@lcb.uhp-nancy.fr

Unlike other thioredoxins *h* characterized so far, a poplar thioredoxin of the *h* type, PtTrx*h*4, is reduced by glutathione and glutaredoxin (Grx) but not NADPH:thioredoxin reductase (NTR). PtTrx*h*4 contains three cysteines; one localized in an N-terminal extension (Cys4) and two (Cys58 and Cys61) in the classical thioredoxin active site ($_{57}$WCGPC$_{61}$). The property of a mutant in which Cys58 was replaced by serine demonstrates that it is responsible for the initial nucleophilic attack during the catalytic cycle. The observation that the C4S mutant is inactive in presence of Grx but fully active when dithiothreitol is used as a reductant indicates that Cys4 is required for the regeneration of PtTrx*h*4 by Grx. Biochemical and X-ray crystallographic studies indicate that two intramolecular disulfide bonds involving Cys58 can be formed, linking it to either Cys61 or Cys4. We propose thus a four-step disulfide cascade mechanism involving the transient glutathionylation of Cys4 to convert this atypical thioredoxin *h* back to its active reduced form.

Thioredoxins (Trxs) are small molecular weight proteins found in all organisms from prokaryotes to higher eukaryotes. They are involved in many cellular processes, dealing primarily with cell redox regulation. In plants, numerous isoforms have been reported. For example, at least 20 genes coding for Trxs are present in the completely sequenced genome of *Arabidopsis thaliana* (1). The Trxs *f*, *m*, *x* and *y* are present in chloroplasts (2, 3), whereas the Trxs *o* are localized in mitochondria (4). The Trxs *h* constitute a large group which includes cytosolic and mitochondrial isoforms (1, 5, 6, 7). Trxs *h* have been divided in three distinct subgroups in a classification based on their primary structure (5). Members of the first and second groups are reduced by NADPH in a reaction catalyzed by NTR. Members of the first and second Trxs *h* subgroups contain a conserved WC[G/P]PC catalytic site. The first cysteine is the one involved in the nucleophilic attack on disulfide bonds present in target proteins, leading to the formation of a disulfide bond between the target protein and Trx. This intermolecular disulfide is then reduced by the second cysteine, leading to the release of reduced target protein and oxidized Trx.

It is only recently that members of plant Trx *h* subgroup 3 have been detected and characterized especially in poplar (8, 9). The poplar thioredoxin PtTrx*h*4, which belongs to this subgroup, contains a regular WCGPC catalytic site but differs from previously characterized Trxs *h* in being reduced *in vitro* by glutaredoxins but not by NTR (8). This unique feature raised several questions about the reaction mechanism of PtTrx*h*4. For example, as most characterized Trxs have a redox midpoint potential of ca -290 mV while Grxs are more electropositive (ca -200 mV), questions about the thermodynamic favourability of reduction of a Trx-like molecule by Grx naturally arise.

Copyright 2008 by The American Society for Biochemistry and Molecular Biology, Inc.

Downloaded from www.jbc.org by on June 21, 2008

The structural and redox properties of animal, bacterial and some plant Trxs have been studied extensively, but little structural information about higher plant Trx *h* is available (10, 11). In addition, there is no subgroup 3 Trx *h* structure solved to date. Since members of this subgroup exhibit an N-terminal extension containing a conserved cysteine in the fourth position that is absent in other subgroups, questions concerning the role of this extension and its additional cysteine arise. It is especially of interest to know how this extension is positioned with respect to the conservation of the Trx fold and more importantly to understand why this protein does not react with its traditional reducing partner, NTR. We show here that, in contrast to other Trxs, three cysteines rather than two are involved in the catalytic mechanism of PtTrx*h*4. Two of these cysteines are present in the classical Trx catalytic site ($WC_{58}GPC_{61}$), whereas the third one is localized in the N-terminal extension (Cys4). From the kinetic and structural data, a new catalytic mechanism is proposed for this Trx isoform.

Experimental procedures

Cloning and mutations of PtTrxh4- The procedures for cDNA isolation of PtTrx*h*4 and its subsequent cloning are described in (8). The PtTrx*h*4 mutants C4S, C58S, and C61S mutants were generated by PCR using cloning and mutagenic oligonucleotides shown below (*Nco*I and *BamH*I sites are underlined and mutagenic bases are in bold): PtTrx*h*4 direct, 5'- CCCCCCATGGGACTTTGCTTGGAT -3'; and PtTrx*h*4 reverse, 5' CCCCGGATCCTCATTTGTCACTAGGGGG CAA 3'; PtTrx*h*4C4S direct, 5'-CCCCCCATGGGACTT**AGC**TTGGATAAG CAT-3' ; PtTrx*h*4C58S direct, 5'-TTCAGTGCAACATGGAGTGG**TCC**TTGT AGACAG-3'; PtTrx*h*4C58S reverse, 5'-CTGTCTACAA**GGA**CCACTCCATGTTGC ACTGAA-3'; PtTrx*h*4C61S direct, 5'-ACATGGTGTGGTCCT**AGT**AGACAGATT GCACCG-3'; PtTrx*h*4C61S reverse, 5'-CGGTGCAATCTGTCT**ACT**AGGACCACA CCATGT-3'.

The mutated PCR products that contained the restriction sites have been cloned into the expression plasmid pET-3d, yielding the constructions pET PtTrx*h*4C4S, pET PtTrx*h*4C58S and pET PtTrx*h*4C61S. The mutations of the recombinant plasmids were verified by DNA sequencing.

Expression and purification of the recombinant proteins- All procedures for the expression and purification of *Arabidopsis thaliana* NTR B (AtNTRB), poplar PrxQ (PtPrxQ), and Grx (WT and mutants) are described elsewhere (12, 13, 14, 15). All the PtTrx*h*4 proteins have been expressed in the *Escherichia coli* strain BL21(DE3), which was also co-transformed with the plasmid helper pSBET as described in (16). The [SeMet]PtTrx*h*4 gene was amplified from a *P. tremula* × *tremuloides* cDNA library. The gene was inserted in the pET-3d expression plasmid, between *Nco*I and B*amH*I sites. Recombinant plasmids carrying the gene of interest were electroporated into methionine auxotrophic strain of *Escherichia coli* BL21(DE3) pSBET. Bacteria were cultured in M9 medium supplemented with selenomethionine (SeMet) and protein overexpression performed as previously described (17). Mass spectrometry was performed to assess purity and to confirm the full incorporation of SeMet, all purification steps of the mutated PtTrx*h*4 proteins being similar to those described for PtTrx*h*4 in (8).

Thiol content titration- The thiol content of each protein preparation was measured using the DTNB procedure as described in (14). All thiol titrations were performed in the presence of SDS on enzymes either as purified or after DTT reduction and dialysis. Consequently, all thiols are titrated regardless of whether or not they were accessible in the protein.

Glutathionylation experiments- The reaction mixture (50 µL) containing 30 mM Tris-HCl, pH 8.0, 1 mM 1,4-dithiothreitol (DTT) and 50 µg poplar Trx (concentration *ca* 80 µM) was incubated for 10 min before adding 5 mM oxidized glutathione.

Electrospray mass spectrometry- A Micromass Q-Tof Ultima (Waters Micromass MS Technologies) hybrid tandem mass spectrometer was used for the acquisition of the electrospray ionization (ESI) mass spectra. This instrument is equipped with a nanoflow electrospray source. The samples were infused into the mass spectrometer using nanoflow capillaries (Proxeon Biosystems, Denmark). The needle voltage was ~1800 V, and the collision energy was 10 eV for the MS analyses. Samples for flow injection analyses were diluted 1:20 with a solution of 50:50 acetonitrile:0.1% formic acid. Data analysis

Downloaded from www.jbc.org by on June 21, 2008

was accomplished with a MassLynx data system and Transform deconvolution software supplied by the manufacturer (Waters Micromass MS Technologies).

Redox potential determination- Oxidation-reduction titrations were carried out as described previously using the fluorescence of the monobromobimane (mBBr)-modified form of the reduced protein to monitor the extent of the protein's reduction (18, 19). Ambient potentials (E_h) were established using mixtures of oxidized glutathione (GSSG) and reduced glutathione (GSH) and the PtTRX*h*4 samples were incubated at these defined E_h values for 2 hours to reach redox equilibrium. The oxidation-reduction midpoint potential (E_m) value was shown to be independent of the total concentration [GSSG and GSH] present in the redox equilibration buffer over the range from 2 to 5 mM. The E_m was calculated by fitting the data to the Nernst Equation for a two-electron process as described previously (20).

PrxQ activity measurement- The reduction of H_2O_2 by poplar PrxQ in the presence of PtTrx*h*4 was followed spectrophotometrically, using a Cary 50 spectrophotometer, by monitoring the decrease in absorbance arising from NADPH oxidation in a coupled enzyme assay system. The reaction mixture (500 µL) contained 30 mM Tris-HCl pH 8.0, 1 mM EDTA, 200 µM NADPH, 0.5 IU glutathione reductase, 1 mM GSH, 6 µM PtGrx C4, 16 µM PtTrx*h*4 WT and mutants, 500 µM H_2O_2 and 2 µM PtPrxQ.

Alternatively, H_2O_2 disappearance was followed directly. The reaction mixture (100 µL) contained 30 mM Tris-HCl pH 7.0, 500 µM DTT, 4 µM PtPrxQ and 36 µM PtTrx*h*4. The reaction was started by adding 500 µM H_2O_2. After given incubation times, 5 µL were mixed with 495 µL of FOX1 (ferrous oxidation in xylenol orange) reagent (21). The absorbance was then read at 560 nm after one-hour incubation.

Crystallization- Three different samples were crystallized, one with Se-Met (WT PtTrx*h*4) and two with regular methionines (WT and PtTrx*h*4C61). Crystallization conditions were screened extensively at 20°C with the microbatch method. Drops used for the initial crystallization trials consisted of 2 µl of the protein solution (20 mg/ml) mixed with 2 µl of various crystallization solutions. The [SeMet]PtTrx*h*4 crystals were grown in 1.0 M sodium/potassium phosphate buffer (pH 6.9) (Hampton SaltRx Screen 2 – solution 54) while the WT PtTrx*h*4 was obtained by using the JBS screen 2 solution D2 (30% PEG 4000, 0.1 M NaHEPES pH 7.5, 0.2 M $CaCl_2$).

For PtTrx*h*4C61S protein, orthorhombic crystals were obtained by using the Hampton SaltRx Screen condition 55 buffer (1.0 M sodium/potassium phosphate pH 8.2). The drop was formed by mixing 25 mg/ml protein with crystallization solution in 1:1 ratio. Crystals were cryoprotected and flash-cooled in liquid ethane at 100 K.

Data collection and processing, structure solutions and refinements- Information and statistics of data collection and processing of the three crystals are presented in Table 1. The diffraction images of the different crystals were indexed, integrated and scaled using either the HKL program (22) or the XDS program package (23) then analyzed using the CCP4 software package version 6.0.2 (24).

A SeMet-labeled protein was prepared and a SAD dataset was collected. Although three selenium atoms per protein were expected (including the one associated to the first methionine residue), only two were found ($SeMet_{79}$ and $SeMet_{134}$) using *SHELXL*97 (25). Subsequent mass spectroscopy analysis allowed us to identify a SeMet-labeled protein which is presumed to be truncated (113 instead of 139 amino acids; absence of the N-terminal part) during its production in *E. coli* system. The solution with the highest correlation coefficient in heavy atom position determination was fed into SHARP (26) for further refinement of SeMet-sulfur positions and phasing. After refinement, SHARP reported that occupancies of heavy atoms were 1.0 respectively, and calculated experimental phase had an overall figure of merit (FOM) of 0.35 (acentric reflections) and 0.06 (centric reflections) for 35 – 2.5 Å diffraction data. The phase improvement was made using SOLOMON (27) where the FOM gradually increased to 0.79 and the electron density map became interpretable. This map was submitted to the automatic model-building program ARP/wARP, version 6.1 (28) and 95 residues were built of the expected 113. The initial model was further improved by manual building with Turbo-Frodo (29) or Coot (30) interspersed with refinements using both CNS, version 1.1 (31) and REFMAC5 (24).

Models of the wild type and the PtTrx*h*4C61S mutant were solved by

Downloaded from www.jbc.org by on June 21, 2008

molecular replacement (MOLREP, 32) using the [SeMet]PtTrx*h*4 as search model. The final structure of the WT PtTrx*h*4 was obtained after manual rebuilding of many parts of the structure (mainly near the active site and the α-helix 3 regions) and refinements using CNS and REFMAC5. Concerning the PtTrx*h*4C61S mutant, 82% of the model was built automatically using ARP/wARP. Final mutant model was obtained using the same procedure as done for WT. Throughout the model building and refinement processes, qualities of all models were assessed using the program PROCHECK (33). Refinement statistics are summarized in Table 1. Figures were prepared with PyMOL (34).

Structure superimpositions- Superimpositions of the present structures with their homologous structures obtained from PDB were performed using LSQMAN program from the DEJAVU package (35), Lsqkab (superpose) program of the CCP4 suite with default parameters proposed by the authors.

Protein Data Bank accession codes- Atomic coordinates and structure factors have been deposited in the Protein Data Bank. Accession codes are 3D21 (WT PtTrx*h*4) and 3D22 (PtTrx*h*4C61S).

RESULTS

PtTrx*h*4 belongs to the third Trx *h* subgroup and exhibits the classical WCGPC active site. In addition to these two cysteinyl residues, members of this subgroup contain one additional conserved cysteine found in the fourth position in an N-terminal extension (comprising the 24 most N-terminus residues) (Figure 1). Sequence analysis using several localization prediction programs suggests that the N-terminal extension does not correspond to a signal peptide and that this protein is likely to be located in the cytosol.

Thiol content determination. In order to investigate the putative role of the three conserved cysteines in the catalytic mechanism of PtTrx*h*4, site-directed mutagenesis has been used in order to produce the full-length recombinant proteins in *E. coli*: PtTrx*h*4C4S, PtTrx*h*4C58S and PtTrx*h*4C61S. Thiol content of these proteins has been determined both in reducing and non-reducing conditions. A summary of these data is shown in Table 2. Nearly three thiols per protein are titrated for the reduced WT protein, while approximately one SH group is present in the non-reduced WT protein. Under reducing conditions, thiol contents of the PtTrx*h*4C4S, PtTrx*h*4C58S and PtTrx*h*4C61S mutants were all close to 2 SH/mol in good agreement with the expected theoretical values. In the absence of reductant, The PtTrx*h*4C61S mutant is fully oxidized (no titrated thiol group) while PtTrx*h*4C4S and PtTrx*h*4C58S mutants are only partially oxidized (i.e., thiol content values lower than 1 SH/mol were measured).

SDS-PAGE was performed under reducing and non-reducing conditions (see supplementary data). In the absence of DTT, both monomers and dimers were detected for PtTrx*h*4 WT and C61S with monomers being the dominant species for both proteins. Although thiol titration results are most consistent with the formation of an intramolecular disulfide bond, the presence of reducible dimers for the C61S mutant indicates that Cys58 and/or Cys4 could be also involved in intermolecular disulfide bonds. Only monomers were observed for the C4S mutant in the absence of DTT, suggesting that an intramolecular disulfide bond is formed between Cys58 and Cys61. In the case of the C58S mutant, oligomeric forms (disappearing under reducing conditions) were present in high abundance, which is consistent with the presence of intermolecular disulfide bonds involving Cys4 and/or Cys61.

Redox potential determination. The redox titrations of PtTrx*h*4 and of its three C/S variants have been carried out at pH 7.0 over the potential range of -50 to -250 mV. Each result represents the average of at least two determinations and the average deviations suggest that the experimental uncertainty in E_m is between 5 and 10 mV. The data obtained for the PtTrx*h*4 and PtTrx*h*4C61S give a good fit to the curve expected for a single two-electron process with an E_m value of -165 ± 10 mV and −178 ± 10 mV, respectively. Data from redox titrations of PtTrx*h*4C4S and PtTrx*h*4C58S could not be fitted to the Nernst Equation for a single two-electron process but did give good fits for the sum of the Nernst Equation for two separate two-electron processes. In the case of PtTrx*h*4C4S, the E_m values for the two components are −140 ± 10 mV and -200 ± 10 mV and for PtTrx*h*4C58S the E_m values of the two components are -130 ± 10 mV and -180 ± 10 mV. The two-electron nature of these redox couples, the E_h range over which they titrate

Downloaded from www.jbc.org by on June 21, 2008

and the fact that the experiments rely on a thiol-specific reagent, mBBr, to monitor the course of the titrations leave little doubt that the components being titrated are dithiol / disulfide couples.

As the redox titration were carried out using the GSH/GSSG couple for redox buffering, it may be possible that some of the components of the redox titration arise from protein-GSH adducts. To investigate the susceptibility of PtTrx*h*4 to glutathionylation, the protein was reduced by DTT, incubated with large excess of GSSG, and analyzed by quadrupole time-of-flight mass spectrometry. It showed the formation of a low-amplitude peak with a mass of 15,761.00 Da (around 35 % of the total preparation) in addition to the peak corresponding to the oxidized recombinant PtTrx*h*4 (15,455.125 Da). This additional peak is compatible with the addition of one GSH molecule (305.5 Da). Concerning the mutants, a large-amplitude additional peak (around 90% of the total preparation) corresponding to a GSH adduct was detected with PtTrx*h*4C58S; PtTrx*h*4C61S was glutathionylated to a lesser extent (around 35 % of the total preparation); while PtTrx*h*4C4S showed no detectable glutathionylation. To identify the glutathionylation site(s) of the WT protein, we performed tryptic digestion of fully reduced or GSSG-oxidized protein and analyzed the tryptic fragments by mass spectrometry. The data indicated that Cys4 is the glutathionylation site on PtTrx*h*4, a result consistent to the absence of GSH adduct on PtTrx*h*4C4S.

Due to the possibilities of extensive glutathionylation and the multiplicity of possible disulfides (both intra- and intermolecular), the potential redox values obtained are actually the average values of a mixture of disulfides. Hence, it is not yet possible to provide a unique assignment of each E_m component in these two-component titrations. Nevertheless, we can conclude that the redox midpoint potential values for all of the couples involved are more positive than -200 mV, a value that is very much more electropositive than the redox potential of typical Trxs (ca. -300 mV).

PrxQ activity. The activity of the recombinant proteins was tested using a non-physiological PrxQ-based system involving PtPrxQ, GSH, and PtGrx (type C4 with an active site CPYC) (8). Both PtTrx*h*4C4S and PtTrx*h*4C58S are totally inactive in this system whereas the mutant PtTrx*h*4C61S retained some of the activity exhibited by WT PtTrx*h*4 (Figure 2a). Taking into account the data reported from previously characterized Trxs, Cys58 is most likely the catalytic residue. As sequence comparisons suggest that Cys58 and Cys61 are the active-site residues (with Cys58 making the initial nucleophilic attack characteristic of Trxs), it is not surprising that converting either of these cysteine residues to serine results in loss of activity. The most interesting feature of PtTrx*h*4, based on the observed total loss of activity for the PtTrx*h*4C4S mutant, is that three cysteines seem to be involved in the catalytic mechanism.

The interactions between PtTrx*h*4 and several PtGrx mutants have also been investigated using this assay. The PtGrx C27S mutant, in which the catalytic cysteine of PtGrx C4 has been replaced to serine, is not active. In this system, mutations of residues surrounding the Grx catalytic cysteines (Y26A and Y29F) altered the reduction efficiency of Grx (36) and thus the PtTrx*h*4 reduction by these mutants. In contrast, Grx C30S (in which the second cysteine of the active site is removed) was still able to reduce PtTrx*h*4, suggesting that monothiol Grx may be able to interact with this kind of Trx.

The interactions between Trxs and PrxQ were also investigated by measuring H_2O_2 disappearance in presence of DTT (Figure 2b). Again, the mutant C58S is totally inactive in this system, consistent with the hypothesis that Cys58 is the catalytic cysteine. The PtTrx*h*4C61S mutant exhibited a much lower activity than PtTrx*h*4, showing the importance of Cys61 in target protein reduction by PtTrx*h*4. In contrast to the results obtained when glutaredoxin was used as the reductant, the PtTrx*h*4C4S mutant is fully active when DTT is the donor.

Overall description and structure comparison of wild-type PtTrxh4 and PtTrxh4C61S. The two PtTrx*h*4 models presented in this study display intramolecular disulfide bonds involving the catalytic Cys58 and either Cys61 (WT PtTrx*h*4, despite the crystallization in the presence of DTT) or Cys4 (PtTrx*h*4C61S mutant). Here, unless indicated otherwise, structure A refers to the WT PtTrx*h*4 (Figure 3a) and structure B to the PtTrx*h*4C61S mutant (Figure 3b). Generally,

Downloaded from www.jbc.org by on June 21, 2008

both forms of PtTrx*h*4 structures retain the overall thioredoxin fold, which is shared by all members of the thiol-disulfide oxidoreductase family, with a five-stranded β sheet surrounded by four α helices in a βα βα βα ββα topology (Figure 3a and 3b).

In structure A, the electron density is well-defined along the main-chains and for most of the side-chains except that the 23 most N-terminal and the 5 most C-terminal amino acids of the WT enzyme could not be located due to weak electron densities, thereby suggesting a high degree of structural flexibility of these parts of the molecule. Hence, this structural model comprises residues 24 – 134 of PtTrx*h*4. Nevertheless, the full length of the protein was verified by mass spectroscopy. The oxidized form A is present as two monomers in the asymmetric unit with active site regions facing each other. The active site architecture is similar to other known Trxs (37 - 40). One side of the redox-active disulfide bond is flat and rather hydrophobic (formed by residues Ala55, $_{59}$GP$_{60}$, $_{99}$ATP$_{101}$, $_{116}$VGA$_{118}$; named as hydrophobic site 1; Figure 3c), while the other side is more shielded with a pitcher-like cleft predominantly hydrophobic (this cleft is formed by Val26, Leu28, Asn52, Ser54, $_{64}$IAP$_{66}$, $_{68}$YI$_{69}$, Val82, Phe103, Leu115, Leu123; named as hydrophobic site 2; Figure 3d). At the edge of the pitcher-like cleft lies a positively-charged patch, $_{56}$RW$_{57}$ and Lys62. The hydrophobic site 2 is accommodating hydrophobic residues from a symmetry-related molecule.

On the other hand, the B structure consists of 129 amino acids (residues 2 – 139, excluding residues from 7 to 15) and the protein crystallizes as a monomer in the asymmetric unit. Upon formation of Cys4-Cys58 disulfide bond, the active site of form B is covered by the N-terminal extension. The electron density between residues 7 and 15 was barely visible, so no residue was assigned to this assumed highly flexible loop region. Nevertheless, the disulfide bond is undoubtedly present (Figure 3b). The N-terminal is kinked from the rest of the molecule at the double-glycine region ($_{23}$GG$_{24}$) and this extension is further stabilized by making extensive hydrophobic interactions which involve residues Ile19 and $_{21}$LAGG$_{24}$ with the residues located at hydrophobic site 2. Apart from the hydrophobic interactions, the N-terminal extension is also stabilized by hydrogen bonding (Val26N-Gly23O, Ile19N-Pro17O, Thr100N-Cys4O, Cys4N-Thr100O, Ala118N-Gly2O, Ile19O-Arg56N$_{\eta 1}$). Upon closure of the N-terminal region towards the core of the protein, the molecular surface property of hydrophobic site 1 remains unchanged (just as described in structure A, see Figure 3c and 3e for comparison). However, the hydrophobic site 2 pocket is shielded by residues from the N-terminal extension (Pro17 – Gly24) and therefore its depth has been greatly reduced (see Figure 3d and 3f for comparison). Indeed, half of the pitcher-like cleft wall (hydrophobic site 2) has now become more hydrophilic in the B structure (Figure 3f).

If the first 23 amino acids are excluded, superimposition of form A and form B structures yield root-mean-square deviation (r.m.s.d.) values of 0.6 Å (for 110 common C$^{\alpha}$ atoms). The positions of the active site cysteines (or Ser61 in form B structure) remain unchanged in both structures. The A and B structures exhibit different conformations in their active site regions with the most obvious difference observed at residue Trp57. In the B structure, the position of Trp57 (χ_1 value = +50.01°, residue stabilized by making a hydrogen bond with the O$_{\delta 2}$ atom of a conserved residue Asp86) is the same as in most of other classical Trxs (χ_1 values between +40° and +55°) and its side-chain covers an important part of the active site (Cys4-Cys58 disulfide bond) surface. This residue has been shown to be important for the Trx-protein interaction or recognition (41, 42). In contrast, the plane of the indole ring of Trp57 (χ_1 value = -81.87°) in the A structure has to "flip out" by about 90° with respect to its usual position as described in the B structure and orients towards the surface of the protein. In the A structure, the Trp57 residues of the two independent monomers are facing each other. This residue possesses an elevated average B factor (72.55 Å^2) and the indole side-chain lacks electron density, suggesting the flexibility of the side-chain. Figure 4 displays a superimposition of the active-site region close-up view of form A PtTrx*h*4, form B PtTrx*h*4 and other Trx structures, highlighting the position and orientation of the active site Trp residue. The atypical position of Trp57 in

Downloaded from www.jbc.org by on June 21, 2008

structure A is not retained in structure B mainly due to the constraint of the N-terminal.

Apart from the flexibility of Trp57, superimposition of monomers of the A and B structures also revealed that its neighboring residue, Arg56, also possesses a certain degree of mobility. This latter side-chain has moved in towards the molecule by about 40° when comparing the A and B structures. The movements of these two residues actually make way for the N-terminal extension of PtTrx*h*4 to gain access to the active site of the enzyme. Another discrepancy between the two forms is the direction of the side-chain of Ser54. Contrary to the A structure where it is more solvent-exposed, the O_{γ} atom of Ser54 is hydrogen bonded to the $O_{\delta 1}$ atom of Asn52 (~2.7 Å) in the B structure. In this structure, the O_{γ} atom of Ser54 is also hydrogen bonded to the S_{γ} atom (~3 Å) of Ser61 (this residue is originally a cysteine residue). Ser54 is highly conserved in PtTrx*h*4 homologues, but usually replaced by a threonine in other plant *h*-type Trxs or an aromatic residue (Phe or Trp) in other non-*h*-type Trxs.

Comparison with other Trxs. Numerous Trx structures (from different organisms) are known, and all display a high degree of structural homology. Among the plant *h*-type Trxs, there are 4 structures of the plant subgroups 1 and 2 available in the PDB to date (they comprise both NMR and X-ray structures), which show very little differences with other known Trx structures. Therefore, we limit the structure comparison here with *h*-type Trxs unless mentioned otherwise.

Structures alignment with 100-109 C^{α} showed the r.m.s.d. values ranging from 1.0 to 1.6 Å, with Trx *h* of *Chlamydomonas reinhardtii* (green alga, PDB code 1EP7, 40), Trx *h*2 of *Hordeum vulgare* (barley, PDB code 2IWT, 43), Trx *h*1 of *A. thaliana* (mouse-ear cress, PDB code 1XFL, 44) and Trx *h*1 of *P. tremula* (poplar tree, PDB code 1TI3, 10). Sequence identity among plant *h*-type Trxs mentioned above ranged between 30% and 35% (50% to 57% of sequence homology). The main differences reside on the extended N-terminal sequence of PtTrx*h*4 which is absent in *h*-type Trxs of subgroups 1 and 2. Nevertheless the canonical active site motif of Trx, WCxxC (x can be any amino acid), is in view. Interestingly, another motif that involves a conserved cis-proline residue, $_{79}$AMP$_{81}$ (numbering in poplar Trx *h*1), is well-conserved in all plant Trxs mentioned above and also in most other Trxs from different organisms, but is replaced by an $_{99}$ATP$_{101}$ motif in PtTrx*h*4 and its plant orthologs. The positioning of Thr100 in PtTrx*h*4 near the disulfide bond (the $O_{\gamma 1}$ of Thr100 is less than 4 Å away from the S_{γ} of the catalytic Cys58) and also near hydrophobic site 2, also suggests an important role for this residue.

A noticeable difference between PtTrx*h*4 and other Trx structures is that Val82 of PtTrx*h*4 which is situated in the hydrophobic site 2 mentioned earlier (see previous section) is replaced by a lysine residue in all other Trxs (corresponding to Lys57 in *E. coli* Trx). In all plant Trxs *h* of subgroup 3, a hydrophobic and relatively small residue is always present. This substitution underlines the hydrophobicity of site 2 whereas in other Trxs a positively charged residue (Lys57) changes the properties of the cavity. In classical Trxs, the corresponding cavity is more solvent-exposed whereas in PtTrx*h*4, Ser54 is the most hydrophilic and polar residue embedded at the bottom of the hydrophobic site 2. All these results suggest that the "hydrophobic cavity" at the hydrophobic site 2 may be involved in regulating the flipping of the N-terminal extension of PtTrx*h*4. Indeed, the site 2 is complementarily covered by the N-terminal extension in form B structure.

Another peculiarity of the active site of PtTrx*h*4 is the absence of a buried carboxylate behind the active-site cysteines. Instead of the common aspartate residue, this corresponding position is occupied by Asn52 in PtTrx*h*4 with equivalent conformation. The buried Asp26 carboxyl (using the *E. coli* Trx numbering) and Lys57 ε–amino groups mentioned, which are in equivalent positions of residues Asn52 and Val82 in PtTrx*h*4, are suggested to have significantly modulating the pK_as of the active site thiols, thereby enhancing the rates of thiol-disulfide reactions at physiological pH (45). In other words, the substitution of these two highly conserved residues of Trxs in PtTrx*h*4 and its orthologs may be responsible for the deviation observed in the more electropositive redox potentials. The substitution of the conserved Asp (but above-mentioned Lys remains conserved) by another residue in PtTrx*h*4, has also been reported for *Anabaena*

Downloaded from www.jbc.org by on June 21, 2008

Trx-2 (replaced by Tyr26) (46) and for *Trypanosoma* Trx (replaced by Trp24) (47). However, no pK_a values of the active-site thiols of these proteins have been determined experimentally.

DISCUSSION

Uniqueness of PtTrxh4. All plant Trxs characterized up to now are reduced by either NTR or ferredoxin/thioredoxin reductase except for poplar thioredoxin PtTrx*h*4 (6). To date, PtTrx*h*4 is the only characterized Trx that is reduced by the GSH/Grx system *in vitro*. It was thus of interest to examine the reasons for the original properties of this Trx. Three conserved cysteines are present in PtTrx*h*4, two found in the classical Trx active site (WCGPC) (Cys58 and Cys61 respectively) and one localized in the N-terminal part of the protein (Cys4). Cys4, shown to play a role in the catalytic mechanism in this study, is a well-conserved residue found at the same position in all other Trxs of subgroup 3 to which PtTrx*h*4 belongs (8).

Proposed catalytic mechanism for an atypical thioredoxin. Based on the biochemical and structural data described here, we propose a four-step mechanism for the reaction of reduced PtTrx*h*4 with an oxidized target protein (e.g., PrxQ) and its regeneration. As for other Trxs, catalysis proceeds through the formation of a heterodimer between the target protein and PtTrx*h*4 involving the catalytic Cys58. The hetero-disulfide is then cleaved by Cys61, leading to the release of the reduced target protein and the formation of a disulfide bond between Cys58 and Cys61 visible in the WT PtTrx*h*4 crystal structure (this step is identical to the one that has been documented for other Trxs). Then, the N-terminal Cys4 is involved in a nucleophilic attack on the Cys58-Cys61 disulfide bond, leading to the formation of a second intramolecular disulphide bond between Cys4 and Cys58. The oxidized protein can then be reduced by the Grx/GSH system. In this atypical Trx mechanism, the thiolate of Grx is proposed to initiate a nucleophilic attack on the Cys4-glutathionylated Trx, leading to the release of fully reduced Trx. The latter steps of this reaction have not been reported in other Trxs and are discussed further below.

Cys4 of PtTrxh4 acts as the second resolving cysteine. From the activity assay results (Figure 2a), we know that PtTrx*h*4C4S is nearly as inactive as the PtTrx*h*4C58S when the Grx/GSH system is used as a reductant. In structure B, we observed that i) PtTrx*h*4 can form a disulfide bond between Cys4 and Cys58, ii) this disulfide bond is non-artifactual since the N-terminal extension fits perfectly in the hydrophobic site 2 pocket by complementary matching involving invariant residues of subgroup 3 Trxs *h*, iii) Cys61 is unlikely to attack the disulfide bond since it is totally buried in the enzyme, and iv) Cys4 is the only solvent-accessible cysteine, suggesting that it is able to interact with the electron donor. This study suggests that Cys4 acts as the second recycling cysteine by reducing the Cys58-Cys61 disulfide bond, leading to the formation of a second intramolecular disulfide bond, i.e., Cys4-Cys58. The catalytic mechanism proposed for PtTrx*h*4 is related to the one described for some peptide methionine sulfoxide reductases A (MsrA), where upon catalysis three cysteines form two consecutive intramolecular disulfide bonds before the regeneration of the enzyme by Trxs (48).

PtTrxh4 is regenerated by GSH/Grx reducing system but not by the typical NTR reducing system. In terms of redox potential, WT PtTrx*h*4 (E_m average value of -165 mV, see results) should be reducible by the NTR reducing system as demonstrated for the vast majority of the Trxs (E_m values in the -280 mV to -310 mV range observed for other Trxs (6, 7, 49)).Thus, there must be reasons other than thermodynamic driving force for the inability of NTR to regenerate PtTrx*h*4. The superimposition with the structure of the Trx-NTR complex (PDB code 1F6M, 50), reveals that the N-terminal extension of PtTrx*h*4 is preventing the interaction with NTR. In addition, structure B exhibits a highly negatively charged molecular surface around the active site which may disfavor its interaction with NTR which also possesses a negatively charged active site surface. Clearly, the N-terminal part of PtTrx*h*4 is not the only determinant responsible for the lack of reactivity with NTR because the [SeMet]PtTrx*h*4 (truncated form, consisting of residues Asn25 to Met134) is also not reduced either by *A. thaliana* or *E. coli* NTR.

Downloaded from www.jbc.org by on June 21, 2008

The conformation of the side-chain of Trp57 at the active site (WCxxC) may also play an important role in preventing a productive interaction between PtTrx*h*4 and NTR. Although the major role of this residue is related to substrate/Trx recognition and interaction (41, 42, 51), it is also involved in the Trx/TrxR interaction (50). In both A (WT and [SeMet]PtTrx*h*4) and B (PtTrx*h*4C61S) structures, the side-chain of Trp57 shows a high flexibility, so that it can either be close to the disulfide bond (structure B) or somewhat distant from the disulfide bond (structure A). A similar flexibility of the corresponding tryptophan has been reported for other Trxs which do not use the common NTR reducing system: the spinach chloroplastic Trx *f* (short form) which uses the ferredoxin/thioredoxin reductase system for its reduction (52) and the Trx of *T. brucei* (Figure 4) which is spontaneously reduced by the trypanothione/ trypanothione reductase system of this parasite (53). These interesting reports hint that Trxs with increased flexibility for this conserved Trp most likely use regeneration system(s) other than the NTR reducing system.

One point that is still unclear is whether the Cys4-Cys58 disulfide bond, once formed, is reduced by GSH or by Grx. In terms of redox potential, an E_m value of -165 mV for PtTrx*h*4 is fully compatible with its thermodynamically favourable complete reduction by either glutathione (GSH/GSSG, E_m value of -241 mV) (54) or glutaredoxins (E_m values of -233 mV, -198 mV and -170 mV ± 10 mV for *E. coli* Grx1, *E. coli* Grx3 and plant Grx C4, respectively) (55-57). We show here that neither GSH nor Grx alone can complete the regeneration of the enzyme but both together are essential for the catalytic mechanism. The fact that we have not been able to obtain a complex between a mutated PtGrx C30S and the PtTrx*h*4C61S suggests that the possibility of direct reduction by Grx is unlikely. It is interesting to note that PtTrx*h*4 can be easily glutathionylated *in vitro* (using mass spectrometry to characterize tryptic digests of glutathionylated WT PtTrx*h*4). Hence, we can postulate that the reaction with a GSH molecule occurs prior to the reaction with Grx in the catalytic cycle.

Physiological considerations of PtTrxh4. Orthologs of PtTrx*h*4 are also present in different plants. Based on amino acid sequence comparisons, it seems quite reasonable to assume that these proteins should, like PtTrx*h*4, be able to use Grx as an electron donor. This assumption is consistent with the report that the *Phalaris cerelescens* ortholog is not reduced by NTR (9). It has been shown that ribonucleotide reductase cannot use monothiol Grx as donor whereas other reactions catalyzed by Grx require either monothiol or dithiol Grxs (55, 58, 59). Based on these considerations, it is tempting to speculate that both kinds of Grx may be able to promote the activity of PtTrx*h*4 and its orthologs. Moreover, there are many naturally occurring monothiol Grx in plants. It will be of considerable interest to compare the relative efficiencies of the corresponding proteins for PtTrx*h*4 catalysis.

The physiological role of this particular kind of GSH/Grx reducible Trx remains to be elucidated. Although PtTrx*h*4 is unable to reduce several isoforms of poplar glutathione peroxidases (Navrot, *unpublished results*), the enzyme is capable to reduce a cytosolic MsrA and a Prx II from poplar (8). In a recent study, it has been shown that *A. thaliana* NTR knockout mutants are viable and fertile and that the inactivation of NTR could be compensated by pathways linked to GSH (60). It is tempting to speculate that this particular subgroup of Trx reduced by Grxs could be involved in this compensatory mechanism. Preliminary experiments suggest that PtTrx*h*4 is expressed both at transcriptional and protein levels in different poplar tissues (Rouhier, *unpublished results*). Further studies are required to identify other possible targets, keeping in mind that the E_m value of PtTrx*h*4 is significantly more positive than that of most Trxs and thus thermodynamic considerations would limit the number of protein disulfides that PtTrx*h*4 is capable of reducing.

ACKNOWLEDGEMENTS

We thank F. Favier for careful reading of the manuscript. We are very grateful to the FIP-team in ESRF, Grenoble, France and the DESY-team in EMBL-Hamburg Outstation, Germany, for providing

Downloaded from www.jbc.org by on June 21, 2008

access to beamlines BM30A, X11 and BW7A. This research was supported by a grant (D-0710 to DBK) from the Rovert A. Welch Foundation. C.S.K. was a recipient of the ASTS fellowship from the Universiti Sains Malaysia.

REFERENCES

1. Meyer, Y., Vignols, F., and Reichheld, J.P. (2002) *Methods Enzymol*. **347**, 394-402
2. Schürmann, P. (2003) *Antiox. Redox Signal*. **5**, 69-78
3. Gelhaye, E ., Rouhier, N., Navrot, N., and Jacquot, J.P. (2004) *Cell Mol. Life Sci.* **62**, 24-35
4. Laloi, C., Rayapuram, N., Chartier., Y., Grienenberger J.M., Bonnard G., and Meyer Y. (2001) *Proc. Natl. Acad. Sci. U S A*. **98**, 14144-14149
5. Gelhaye, E ., Rouhier, N., Navrot, N., and Jacquot, J.P. (2004) *Plant Physiol. Biochem.* **42**, 265-271
6. Bréhélin, C., Laloi, C., Setterdahl, A.T., Knaff, D.B., and Meyer, Y. (2004) *Photosynth. Res.* **79**, 295-304
7. Gelhaye, E., Rouhier, N., Gerard, J., Jolivet, Y., Gualberto, J., Navrot, N., Ohlsson, P.I., Wingsle, G., Hirasawa, M., Knaff, D.B., Wang, H., Dizengremel, P., Meyer, Y., and Jacquot, J.P. (2004) *Proc. Natl. Acad. Sci. U S A*. **101**, 14545-14550
8. Gelhaye, E., Rouhier, N., and Jacquot J.P. (2003) *FEBS Lett.* **555**, 443-448
9. Juttner, J., Olde, D., Langridge, P., and Baumann, U. (2000) *Eur. J. Biochem.* **267**, 7109-7117
10. Coudevylle, N., Thureau, A., Hemmerlin, C., Gelhaye, E., Jacquot, J.P. & Cung, M.T. (2005) *Biochemistry*. **44**, 2001-2008
11. Peterson, F.C., Lytle, B.L., Sampath, S., Vinarov, D., Tyler, E., Shahan, M., Markley, J.L. & Volkman, B.F. (2005) *Protein Sci*. **14**, 2195-2200
12. Jacquot, J.P., Rivera-Madrid, R., Marinho, P., Kollarova, M., Le Marechal, P., Miginiac-Maslow, M., and Meyer, Y. (1994) *J. Mol. Biol.* **235**, 1357-1363
13. Rouhier, N., Gelhaye, E., Gualberto, J.M., Jordy, M.N., De Fay, E., Hirasawa, M., Duplessis, S., Lemaire, S.D., Frey, P., Martin, F., Manieri, W., Knaff, D.B., and Jacquot, J.P. (2004) *Plant Physiol*. **134**, 1027-1038
14. Rouhier, N., Gelhaye, E., and Jacquot, J.P. (2002) *J. Biol. Chem.* **277**, 13609-13614
15. Rouhier, N., Gelhaye, E., Sautiere, P.E., and Jacquot, J.P. (2002) *Protein Expr. Purif.* **24**, 234-241
16. Schenk, P.M., Baumann, S., Mattes, R., and Steinbiss, H.H. (1995) *Biotechniques*. **19**, 196-200
17. Navrot, N., Collin, V., Gualberto, J., Gelhaye, E., Hirasawa, M., Rey, P., Knaff, D. B., Issakidis, E., Jacquot, J. P. and Rouhier, N. (2006) *Plant Physiol.* **142**, 1364-1379
18. Krimm, I., Lemaire, S., Ruelland, E., Miginiac-Maslow, M., Jacquot, J.P., Hirasawa, M., Knaff, D.B., and Lancelin, J.M. (1998) *Eur. J. Biochem.* **255**, 185-195
19. Hirasawa, M., Ruelland, E., Schepens, I., Issakidis-Bourguet, E., Miginiac-Maslow, M., and Knaff, D.B. (2000) *Biochemistry* **39**, 3344-3350
20. Masuda, S., Dong, C., Swem, D., Knaff, D.B., and Bauer, C.E. (2002) *Proc. Natl. Acad. Sci. USA*. **99**, 7078-7083
21. Deiana, L., Carru, C., Pes, G., and Tadolini, B. (1999) *Free Radic. Res.* **31**, 237-244
22. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol*. **276**, 307-326
23. Kabsch, W. (1993) *J. Appl. Crystallog.* **26**, 795-800
24. CCP4: Collaborative Computational Project Number 4 (1994) *Acta Crystallog. sect. D*. **50**, 760-763
25. Sheldrick, G.M. & Schneider, T.R. (1997) *Methods Enzymol*. **277**, 319-343
26. de La Fortelle, E. & Bricogne, G. (1997) *Methods Enzymol*. **276**, 472-494
27. Abrahams, J.P. & Leslie, A .G. (1996) *Acta Crystallog. sect. D*. **52**, 30-42
28. Perrakis, A., Morris, R. & Lamzin, V.S. (1999) *Nat. Struct. Biol*. **6**, 458-463
29. Roussel, A. & Cambillau, C. (1989) *Silicon Graphics Geometry Partners Directory*, pp.77-78 ; Mountain View, CA : Silicon Graphics
30. Vagin, A. & Teplyakov, A. (1997) *J. Appl. Crystallog.* **30**, 1022-1025

Downloaded from www.jbc.org by on June 21, 2008

31. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. & Warren, G.L. (1998) *Acta Crystallog. sect. D*. **54**, 905-921
32. Emsley, P. & Cowtan, K. (2004) *Acta Crystallog. sect. D*. **60**, 2126-2132
33. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallog.* **26**, 283-291
34. DeLano, W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, Palo Alto, CA, USA. http://www.pymol.org
35. Kleywegt, G.J. & Jones, T.A. (1997) *Acta Crystallog. sect. D*. **53**, 179-185
36. Rouhier, N., Gelhaye, E. & Jacquot, J.P. (2002) *FEBS Lett*. **511**, 145-149
37. Holmgren, A., Söderberg, B. O., Eklund, H. & Brändén, C. I. (1975) *Proc. Natl. Acad. Sci. USA*. **72**, 2305-2309
38. Katti, S. K., LeMaster, D. M. & Eklund, H. (1990) *J. Mol. Biol*. **212**, 167-184
39. Weichsel, A., Gasdaska, J. R., Powis, G. & Monfort, W. R. (1996) *Structure*. **4**, 735-751
40. Menchise, V., Corbier, C., Didierjean, C., Saviano, M., Benedetti, E., Jacquot, J. P. & Aubry, A. (2001) *Biochem. J*. **359**, 65-75
41. Menchise, V., Corbier, C., Didierjean, C., Jacquot, J.P., Benedetti, E., Saviano, M. & Aubry, A. (2000-2001) *Biopolymers*. **56**, 1-7
42. Krause, G. & Holmgren, A. (1991) *J. Biol. Chem*. **266**, 4056-4066
43. Maeda, K., Hägglund, P., Finnie, C., Svensson, B. & Henriksen, A. (2006) *Structure*. **14**, 1701-1710
44. Peterson, F. C., Lytle, B. L., Sampath, S., Vinarov, D., Tyler, E., Shahan, M., Markley, J. L. & Volkman, B. F. (2005) *Protein Sci*. **14**, 2195-2200
45. Dyson, H. J., Jeng, M.-F., Tennant, L. L., Slaby, I., Lindell, M., Cui, D.-S., Kuprin, S. & Holmgren, A. (1997) *Biochemistry*. **36**, 2622-2636
46. Saarinen, M., Gleason, F.K. & Eklund, H. (1995) *Structure*. **3**, 1097-1108
47. Friemann, R., Schmidt, H., Ramaswamy, S., Forstner, M., Krauth-Siegel, R.L. & Eklund, H. (2003) *FEBS Lett*. **554**, 301-305
48. Boschi-Muller, S., Azza, S., Sanglier-Cianferani, S., Talfournier, F., Van Dorsselear, A., and Branlant, G. (2000) *J. Biol. Chem.* **275**, 35908-35913
49. Wangensteen, O.S., Chueca, A., Hirasawa, M., Sahrawy, M., and Knaff D.B., Lopez Gorge, J. (2001) *Biochim. Biophys. Acta.* **1547**,156-166
50. Lennon, B.W., Williams, C.H. Jr. & Ludwig, M.L. (2000) *Science*. **289**, 1190-1194
51. Koh, C. S., Didierjean, C., Navrot, N., Panjikar, S., Mulliert, G., Rouhier, N., Jacquot, J. P., Aubry, A., Shawkataly, O. & Corbier, C. (2007) *J. Mol. Biol*. **370**, 512-529
52. Crawford, N.A., Droux, M., Kosower, N.S. & Buchanan, B.B. (1989) *Arch. Biochem. Biophys*. **271**, 223-239
53. Schmidt, H. & Krauth-Siegel, R.L. (2003) *J. Biol. Chem*. **278**, 46329-46336
54. Cheng, Z., Arscott, L.D., Ballou, D.P. & Williams, C.H. Jr. (2007) *Biochemistry*. **46**, 7875-7885
55. Rouhier, N., Gelhaye, E., and Jacquot, J.P. (2004) *Cell Mol. Life Sci.* **61**,1266-1277
56. Aslund, F., Berndt, K.D. & Holmgren, A. (1997) *J. Biol. Chem*. **272**, 30780-30786
57. Rouhier, N., Unno, H., Bandyopadhyay, S., Masip, L., Kim, S.K., Hirasawa, M., Gualberto, J.M., Lattard, V., Kusunoki, M., Knaff, D.B., Georgiou, G., Hase, T., Johnson, M.K. & Jacquot, J.P. (2007) *Proc. Natl. Acad. Sci. USA*. **104**, 7379-7384
62. Fernandes, A. P. & Holmgren, A. (2004) *Antioxid. Redox Signal*. **6**, 63-74
63. Bushweller, J. H., Åslund, F., Wüthrich, K & Holmgren, A. (1992) *Biochemistry*. **31**, 9288-9293
64. Reichheld, J. P., Khafif, M., Riondet, C., Droux, M., Bonnard, G. & Meyer, Y. (2007) *Plant Cell*. **19**, 1851-1865

FOOTNOTES

Downloaded from www.jbc.org by on June 21, 2008

The abbreviations used are: Grx, glutaredoxin; NTR, NADPH thioredoxin reductase; Prx, peroxiredoxin; Trx, thioredoxin; GSH, glutathione; DTNB, dithionitrobenzoate; DTT, dithiothreitol; SDS, sodium dodecyl sulfate; SAD, single-wavelength anomalous dispersion.

Figure 1. Multiple alignment of poplar Trx*h*4 related sequences. Amino acid sequences were aligned using CLUSTALW software. Accession numbers are as follows : Pt1 : *Populus tremula x Populus tremuloides* (AAL99941); Sec : *Secale cereale* (AF159386); Ta : *Triticum aestivum* (AF438359); Hv : *Hordeum vulgare* (AF435815) ; Hb : *Hordeum bulbosum* (AF159385) ; Lp : *Lolium perenne* (159387); Pc : *Phalaris coerulescens* (AF159388) ; Os : *Oryza sativa* (AF435817) ; Zm : *Zea mays* (AF435816); Ls : *Lactuca sativa* (TC9851); Vv : *Vitis vinifera* (CB004453); Gm : *Glycine max* (CA799351); Mc : *Mesembryanthemum crystallinum* (CA838461); Pt4 : *Populus tremula x P. trichocarpa* (P85801); At9: *Arabidopsis thaliana* (At3g08710). The asterisk corresponds to strict identity, the colon corresponds to functional homology, and the period corresponds to structural homology. Amino acids that compose the hydrophobic site 1 and 2 are highlighted in dark grey and light grey, respectively. Three cysteines proposed to be involved in the atypical catalytic mechanism of PtTrx*h*4 (see text for details) are well-conserved in its orthologs (shown in black).

Figure 2. Activity of PtPrxQ in the presence of PtTrx*h*4 with (a) PtGrx C4 or with (b) DTT. For assay in (a), PtPrxQ (2 µM) was incubated with either PtTrx*h*4WT or the different mutants (16 µM) in presence of the GSH/Grx system (200 µM NADPH, 1 mM GSH, 0.5 IU glutathione reductase, 6 µM PtGrx C4, 500 µM H_2O_2). NADPH oxidation was measured at 340 nm as absorbance change per min. For assay in (b), the reduction of H_2O_2 by PtPrxQ (5 µM) was measured by following the disappearance of 500 µM H_2O_2 using 1 mM DTT in presence of either PtTrxh4WT or the different mutants (30 µM). Depletion of H_2O_2 was measured at 560 nm as absorbance change per min.

Figure 3. Overall structures of the WT PtTrx*h*4 (blue) and the PtTrx*h*4C61S (orange) in cartoon representation (a, b) and the electrostatic potential surface charges of the WT (c and d) and PtTrx*h*4C61S mutant (e and f), viewing from both sides of the molecule (rotated by 180°). Molecules in (c), (d), (e) and (f) correspond to the orthogonal view of the molecule in (a). For (a) and (b), the intramolecular disulfide bond in each structure is highlighted, with final $2F_o$-F_c electron density (1.2 σ level) covering chosen residues for clarity. The N-terminal extension (residues 1 - 23) is not modeled in the WT PtTrx*h*4 structure (a) due to the lacking of electron density, whereas in the PtTrx*h*4C61S partial of the N-terminal extension (residues 2 – 6 and 16 – 24, colored in green) can be assigned to the structure (b, see text for details). In the latter structure, lacking residues (7 – 15) are shown in an imaginary gray loop for a better overview of the structure. Relevant residues and secondary structures are labeled. Two hydrophobic sites, site 1 (c) and site 2 (d), are circled for clarity. Corresponding sites in the mutant can be found in the bottom panel (e and f). Notice that that site 1 cleft is shallower than the pitcher-like hydrophobic site 2 (for WT PtTrx*h*4). The figure was prepared using PyMOL.

Figure 4. Superimpostion of the active site cysteines and tryptophan from poplar WT PtTrx*h*4 (blue) (PDB code 3D21), poplar PtTrx*h*4C61S (orange) (PDB code 3D22), *C. reinhardtii* (green) (PDB code 1EP7), *T. brucei* (yellow) (PDB code 1R26) and spinach Trx*f* short form (pink) (PDB code 1F9M). For PtTrx*h*4, Cys58 is the catalytic cysteine (corresponding to $Cys36_{1EP7}$, $Cys46_{1F9M}$ and $Cys30_{1R26}$), Cys61 is the first resolving cysteine (corresponding to $Cys39_{1EP7}$, $Cys49_{1F9M}$ and $Cys33_{1R26}$) while Cys4 is the additional N-terminal cysteine or the second resolving cysteine in the proposed atypical mechanism of PtTrx*h*4 (see text for details). In contrast to the typical position in most Trxs, exemplified by Trp35 of green alga Trx *h* (green) and Trp57 of poplar WT PtTrx*h*4 (blue), the Trp57

Downloaded from www.jbc.org by on June 21, 2008

of the PtTrx*h*4C61S mutant has a flipped-out position. The unusual Trp side-chain position seen in PtTrx*h*4 has also been found in two other Trx structures mentioned in the text ($Trp45_{1F9M}$ and $Trp29_{1R26}$). The figure was prepared using PyMOL.

Figure 5. A proposed four-step disulfide cascade mechanism for Grx-dependent PtTrx*h*4 catalysis. 1) Cys58 attacks initially the S_γ atom of an oxidized target protein, 2) the heterodisulfide is then reduced by the active site recycling cysteine (Cys61), 3) the resulting disulfide bond (Cys58-Cys61) is further reduced by the second recycling cysteine, Cys4, and 4i) a glutathione molecule (GSH) reduces the Cys4-Cys58 disulfide bond prior to 4ii) the final nucleophilic attack of the Trx-SG adduct by the catalytic Cys of monothiol or dithiol Grx.

Table 1. Data collection, phasing and refinement statistics for the [SeMet]PtTrx*h*4 , the WT PtTrx*h*4 and the PtTrx*h*4C61S crystals.

Data Set	**[SeMet]PtTrx*h*4**	**WT PtTrx*h*4**	**PtTrx*h*4 C61S**
Data Collection and Processing Statistics			
Data collection site	BM30A ESRF-Grenoble	X11 DESY-Hamburg	BW7A DESY-Hamburg
Wavelength (Å)	0.9805	0.8123	1.2400
Space group	$P4_12_12$	$P6_1$	$P2_12_12_1$
Unit cell dimensions (Å) (a,b,c)	44.89 44.89 131.74	47.35 47.35 196.00	31.78 44.10 85.68
Asymmetric unit	1 subunit	2 subunits	1 subunit
Resolution range (Å)[a]	32.94 - 2.46 (2.60 - 2.46)	50.00 - 2.15 (2.23 - 2.15)	30.00 - 1.60 (1.66 - 1.60)
Redundancy [a]	7.0 (3.6)	28.3 (37.6)	5.2 (5.3)
Completeness (%)[a]	93.4 (65.7)	99.8 (100.0)	98.5 (98.4)
I/sigma I [a]	17.85 (3.10)	22.98 (2.60)	16.03 (6.20)
R_{merge} [a, b]	0.089 (0.588)	0.053 (0.392)	0.092 (0.266)
Phasing power (acentric/centric)	1.283 / 0		
R_{cullis} (isomorphous/anomalous)	0 / 0.754		
Figure of merit(acentric/centric)	0.34824 / 0.06270		
Refinement statistics			
Resolution range (Å)		41.0 - 2.15	30.0 - 1.6
Reflections used		12725	15470
R_{cryst} [c] (R_{free})[d]		20.97 (28.53)	19.06 (19.63)
Protein / waters / PO_4		2 x (111 residues) / 63 / 0	129 residues / 115 / 1
Mean B factor (Å^2)			
Main-chain		55.20	19.40
Side-chain		57.67	21.52
Water		59.29	32.47
All		56.53	21.69
Rms deviation from ideal geometry			
Bond lengths (Å)		0.030	0.010
Bond angles (°)		2.5	1.4
Dihedral angles (°)		26.3	24.1

Downloaded from www.jbc.org by on June 21, 2008

Improper angles (°)	5.59	1.46
Ramachandran plot		
Residues in most favored regions (%)	91.5	92.8
Residues in additionally allowed regions (%)	7.5	7.2
Residues in generously allowed regions (%)	1.0	0.0

[a] The values in parentheses are for the highest resolution bin.

[b] $R_{merge}=\Sigma i|Ii-bIN|/\Sigma|bIN|$, where I is the intensity for the ith measurement of an equivalent reflection with the indices *h*, *k*, *l*.

[c] $R_{cryst}=\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.

[d] The R_{free} value was calculated from 5% of all data that were not used in the refinement.

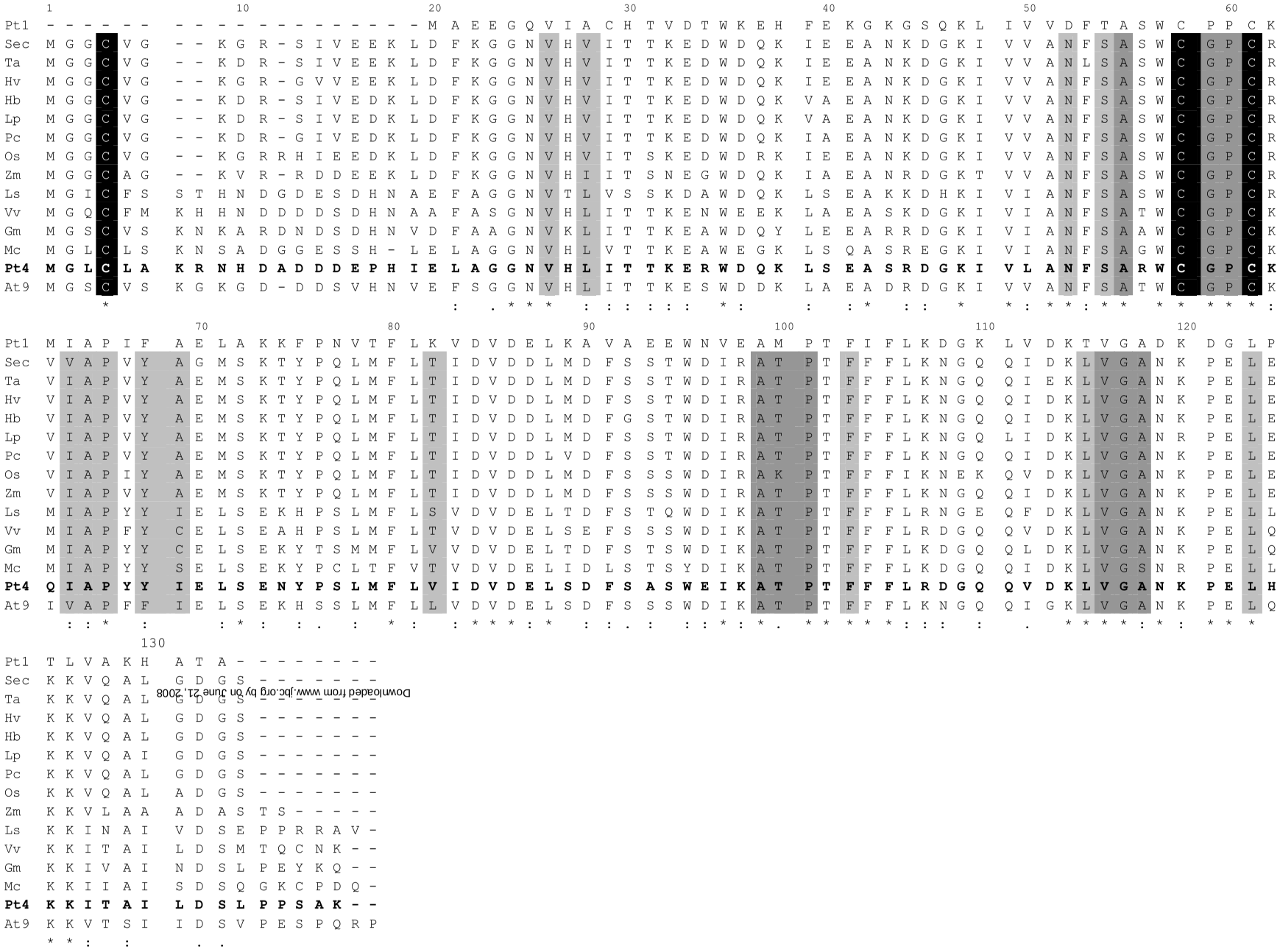
Table 2. Thiol content of PtTrx*h*4 under nonreducing or reducing conditions. Thiols were titrated using DTNB as described under "Experimental procedures." Data are expressed in mol SH/mol enzyme. The S.D. is typically ± 0.2 SH/ mol SH.

	Nonreducing conditions	Reducing conditions
WT PtTrx*h*4	0.75	2.88
PtTrx*h*4C4S	0.93	1.64
PtTrx*h*4C58S	0.64	1.93
PtTrx*h*4C61S	0.05	1.55

Supplementary data

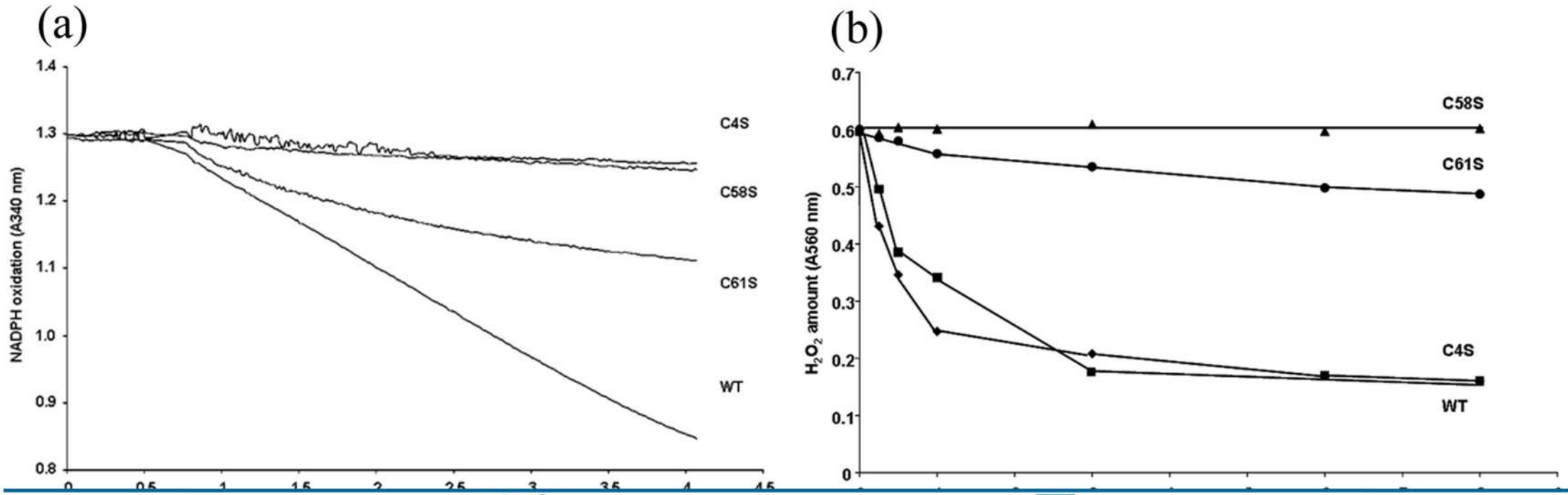
Figure 1. Analysis of purified recombinant proteins by SDS-PAGE. 14% SDS-PAGE under both reducing and nonreducing conditions showing PtTrx*h*4WT, PtTrx*h*4C4S, PtTrx*h*4C58S and PtTrx*h*4C61S.

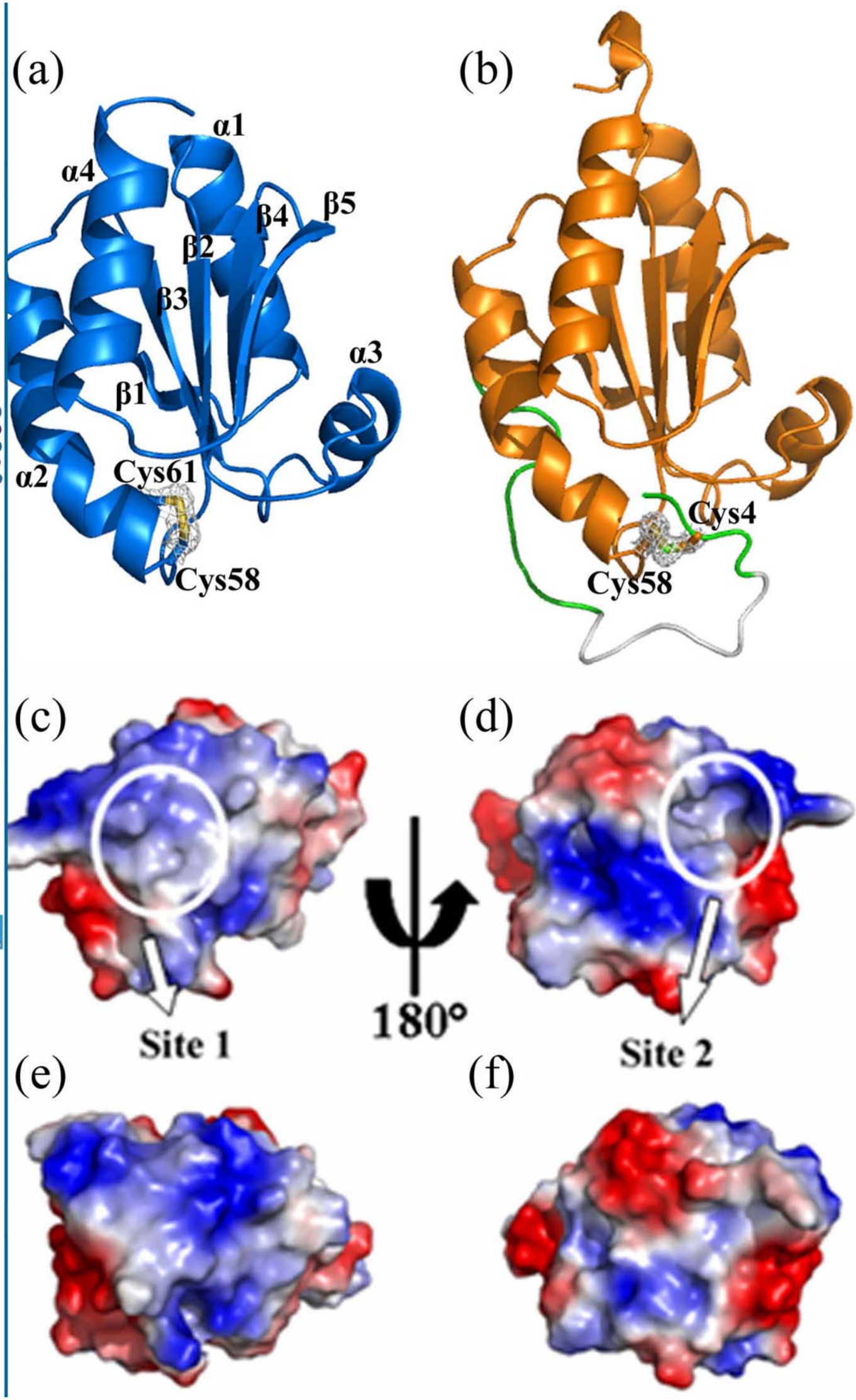
Downloaded from www.jbc.org by on June 21, 2008



Downloaded from www.jbc.org by on June 21, 2008

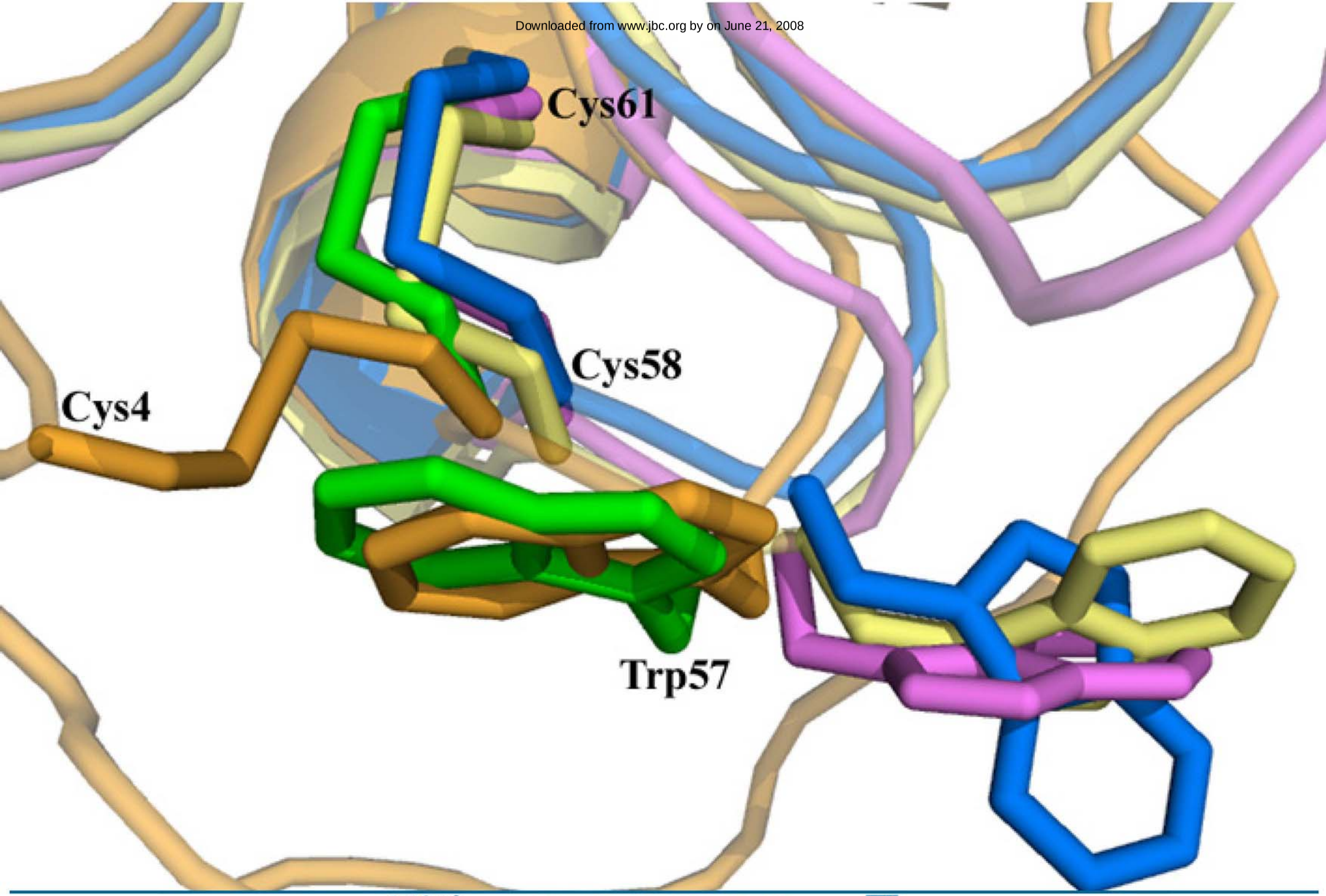
Downloaded from www.jbc.org by on June 21, 2008

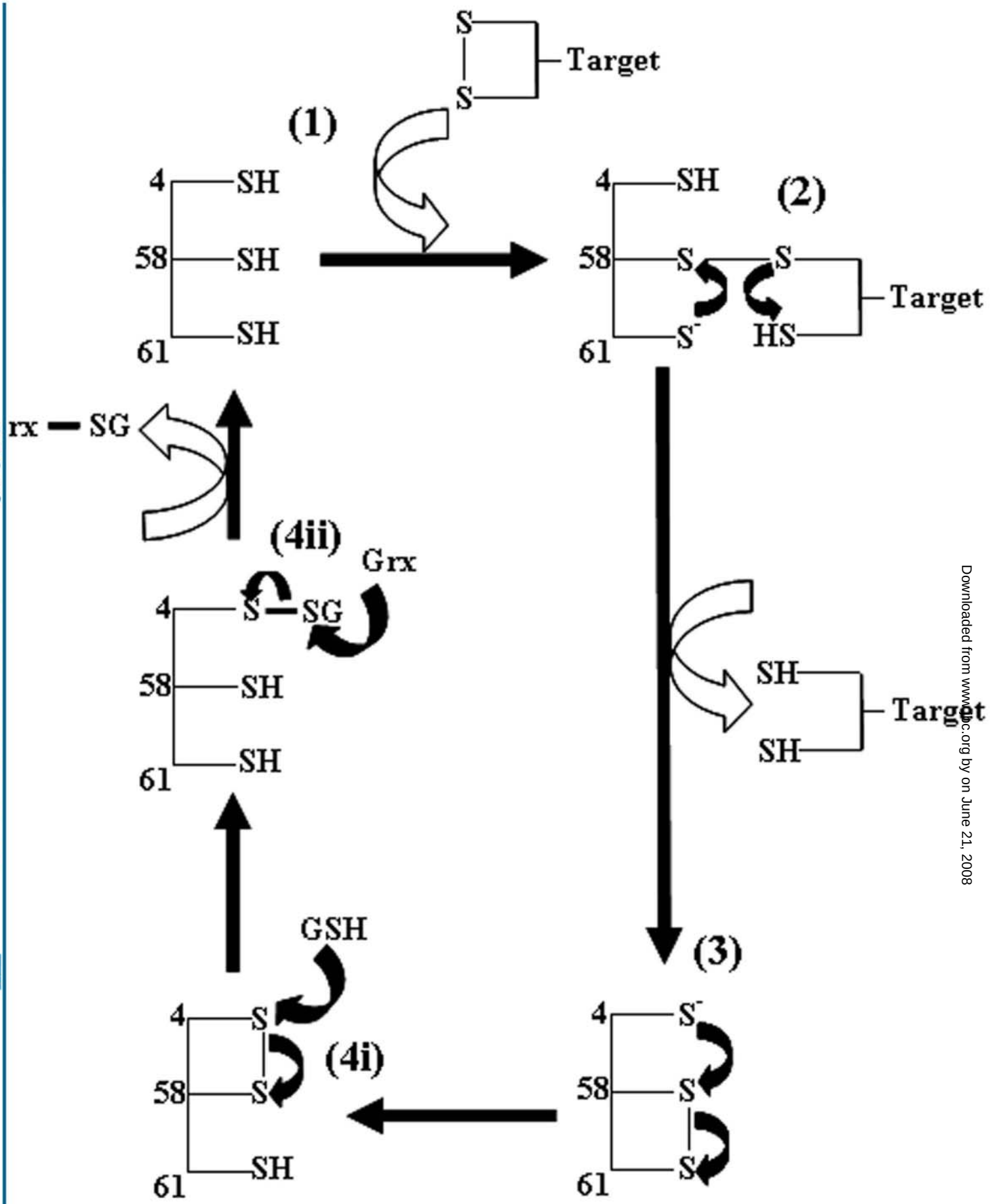




The Journal of Biological Chemistry
jbc
Downloaded from www.jbc.org by on June 21, 2008

Downloaded from www.jbc.org by on June 21, 2008





Downloaded from www.jbc.org by on June 21, 2008

Nancy-Université

Université Henri Poincaré

Mademoiselle **KOH CHA SAN**

DOCTORAT DE L'UNIVERSITÉ HENRI POINCARÉ, NANCY 1

en CHIMIE & PHYSICO-CHIMIE MOLECULAIRES

VU, APPROUVÉ ET PERMIS D'IMPRIMER

Nancy, le 16/10/08

Le Président de l'Université

JP. FINANCE

Université Henri Poincaré, Nancy I
24-30 rue Lionnois - B.P. 3069 - 54013 NANCY Cédex
Tél. : 03 83 68 20 00 - Fax : 03 83 68 21 00

ÉTUDE CRISTALLOGRAPHIQUE DES OXYDORÉDUCTASES IMPLIQUÉES DANS LA RÉPONSE AU STRESS OXYDATIF CHEZ LE PEUPLIER POUR UNE MEILLEURE COMPRÉHENSION DES MÉCANISMES CATALYTIQUES

RÉSUMÉ

La production des espèces réactives de l'oxygène (ROS) est une conséquence inévitable de la vie dans un milieu aérobie. Une fois produites, ces espèces peuvent être éliminées par différents systèmes enzymatiques. Les structure de trois oxydoréductases (la glutathion peroxydase (Gpx), la thiorédoxine (Trx) et la glutarédoxine (Grx)) de *Populus trichocarpa* × *deltoides* (le peuplier) ont été résolues par diffraction des rayons X. Les Gpxs forment un groupe d'enzymes qui régulent la concentration des ROS dans les cellules, et qui les protègent des effets d'un stress oxydant. Contrairement à leurs homologues d'origine animale, les Gpxs végétales ne dépendent pas du glutathion (GSH) mais des Trx pour leur fonctionnement. De plus, elles possèdent une cystéine dans leur site actif au lieu d'une sélénocystéine. Dans cette étude, j'ai résolu les structures cristallines des formes réduite et oxydée de la Gpx5 de peuplier (PtGpx5) et montré que des changements conformationnels drastiques sont nécessaires pour permettre la formation d'un pont disulfure intramoléculaire. Les Trxs constituent une famille de petites protéines ubiquitaires, régulant diverses protéines cibles par la réduction de leur pont disulfure. Par cette étude structurale, mon objectif était de comprendre précisément le mécanisme catalytique d'une nouvelle isoforme caractérisée chez le peuplier, la PtTrx*h*4, dont la capacité à accepter des électrons de la Grx a été récemment démontrée. A la différence des autres Trxs *h* caractérisées jusqu'ici, la PtTrx*h*4 ne peut pas être réduite par la voie typique qui implique la NADPH-Trx réductase. Cette PtTrx*h*4 contient trois cystéines, la première localisée dans une extension en position N-terminale (Cys4) et deux (Cys58 et Cys61) situées dans le site actif classique (WC_1GPC_2) de la Trx. Les résolutions des structures de l'enzyme sauvage et du mutant C4S sous forme oxydée m'ont permis de proposer un mécanisme catalytique en quatre étapes en accord avec les études enzymatiques. Les Grxs sont des protéines qui utilisent des électrons de GSH en particulier pour catalyser des réactions d'échange de thiol-disulfure. Ici, je présente la structure de la PtGrxS12 (en complexe avec le GSH), la première structure de la Grx végétale sous-classe 1 avec un motif du site actif atypique $_{28}WCSYS_{32}$. Cette protéine possède une autre cystéine (Cys87), dont le rôle reste obscur. Le site de fixation du GSH est détaillé ultérieurement dans ce manuscrit. Ces structures éclaircissent notre compréhension du mécanisme d'oxydoréduction chez les végétaux et nous permettent de mieux comprendre les interactions enzyme-substrat.