



HAL
open science

Etude structurale d'une Aldéhyde Déshydrogénase NADP dépendante de Streptococcus mutans

David Cobessi

► **To cite this version:**

David Cobessi. Etude structurale d'une Aldéhyde Déshydrogénase NADP dépendante de Streptococcus mutans. Biologie structurale [q-bio.BM]. Université Henri Poincaré - Nancy 1, 2000. Français. NNT : 2000NAN10035 . tel-01747582

HAL Id: tel-01747582

<https://hal.univ-lorraine.fr/tel-01747582>

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>

Thèse

Présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy I

**Mention Sciences
Spécialité Biologie Structurale**

Par

David COBESSI

**Etude structurale d'une Aldéhyde Déshydrogénase
NADP dépendante de *Streptococcus mutans***

Soutenue le 16 juin 2000 devant la commission d'examen:

Membres du jury:

Rapporteurs : M. Juan Carlos Fontecilla-Camps, Directeur de Recherches, IBS (CNRS-CEA), Grenoble
M. Richard Giége, Directeur de Recherches, IBMC (CNRS), Strasbourg

Examineurs: M. André Aubry, Directeur de Recherches, Université H. Poincaré, Nancy I
M. Guy Branlant, Professeur, Université H. Poincaré, Nancy I
M. Hans Eklund, Professeur, Université des Sciences Agricoles, Uppsala, Suède
M. Patrick Netter, Professeur, Université H. Poincaré, Nancy I

Le travail présenté dans ce manuscrit a été réalisé au Laboratoire de Cristallographie et Modélisation des Matériaux Minéraux et Biologiques dirigé par le Professeur Claude Lecomte. Je le remercie de m'avoir accueilli dans son laboratoire et de la bienveillance qu'il m'a témoignée.

J'exprime ma respectueuse reconnaissance à Monsieur le Professeur Hans Eklund, Messieurs les Directeurs de Recherches Juan Carlos Fontecilla-Camps et Richard Giegé pour l'honneur qu'ils me font de juger cette thèse.

Je remercie Monsieur le Professeur Patrick Netter d'avoir accepté de juger ce travail et Monsieur le Professeur Guy Branlant d'en être examinateur. Je le remercie également pour les nombreuses discussions scientifiques que nous avons eues et pour les critiques pertinentes qu'il a apportées à ce travail. Qu'il soit assuré de ma reconnaissance.

Mon inscription au DEA de Cristallographie et de RMN biologiques est une initiative d'André Aubry lorsque j'effectuais un stage au LCM3B. Je le remercie de m'avoir incité à suivre cette formation, pour la confiance qu'il m'a témoignée, et pour la simplicité de ses rapports humains. André Aubry a dirigé ce travail avec beaucoup d'optimisme en me laissant une très grande liberté d'initiative que j'ai appréciée. Je lui en suis reconnaissant.

Je tiens à remercier le Professeur Bernard Vitoux pour les discussions que nous avons eues et que les nombreuses charges assumées n'empêchent pas d'être à l'écoute des gens. Les cours de qualité en Biologie Structurale, qu'il m'a dispensés avec pédagogie et enthousiasme, en Licence et Maîtrise de Biochimie ont été déterminants dans le choix de mon orientation.

J'ai à cœur d'exprimer toute ma Gratitude et toute mon Amitié à Frédérique Tête Favier avec qui j'ai partagé de nombreuses journées au laboratoire et sur divers synchrotrons. Ses compétences, sa franchise, son soutien, sa bonne humeur et tous les moments et discussions stimulantes partagés (autour de publications, gâteaux, PC, croissants, éprouvettes, chocolats, fioles...) m'ont été d'une grande utilité. Je la remercie également d'avoir accepté de consacrer de son temps à la relecture de ce manuscrit.

Je remercie Jean-Paul Mangeot qui m'a permis de découvrir l'enseignement en me confiant des heures de travaux dirigés lors de ma deuxième année de thèse.

Merci également à Catherine Corbier d'avoir accepté de relire ce manuscrit et à Guillermo Mulliert pour sa disponibilité dans la maintenance de l'informatique du Groupe Biocristallographie.

Le soin avec lequel Claude Didierjean a entretenu le dispositif de rayons X m'a permis de réaliser l'acquisition des données de diffraction. Je l'en remercie.

Je remercie les techniciens et chercheurs du M. A. E. M. de m'avoir fourni la protéine, une matière première indispensable à la biocristallographie, en qualité et en quantité.

Je remercie également tous ceux qui de près ou de loin par leur travail, leurs connaissances, leur sympathie, leur bonne humeur ont contribué à la réalisation de ce travail. Qu'ils m'excusent de ne pas les nommer.

Sommaire	p.1
Abréviations	p. 7
Liste des figures	p. 9
Liste des tableaux et des graphiques	p. 11
Introduction	p. 13
Les aldéhyde déshydrogénases	p. 17
I- Aperçu des fonctions enzymatiques des ALDHs	p. 18
A- Les ALDHs chez les mammifères.....	p. 18
1- L'aldéhyde déshydrogénase mitochondriale ou ALDH2.....	p. 18
2- L'aldéhyde déshydrogénase cytosolique ou ALDH1.....	p. 19
3- L'aldéhyde déshydrogénase de la classe 3 ou ALDH3.....	p. 21
B- Les ALDHs chez les végétaux.....	p. 21
1- La bétaine aldéhyde déshydrogénase ou BADH.....	p. 21
2- La GAPDH NADP dépendante non phosphorylante.....	p. 22
C- Le domaine ALDH: cas de la FTHFDH.....	p. 24
D- Les ALDH structurales.....	p. 24
E- Les ALDH chez les bactéries.....	p. 25
II- Les structures de quelques ALDH	p. 25
A- L'ALDH de la classe 3 (Liu <i>et al.</i> , 1997).....	p. 25
B- L'ALDH de la classe 2 (Steinmetz <i>et al.</i> , 1997).....	p. 28
C- L'ALDH de la classe 9 (Johansson <i>et al.</i> , 1998).....	p. 30
D- L'ALDH de la classe 1 (Moore <i>et al.</i> , 1998).....	p. 31
III- Le mécanisme catalytique proposé pour les ALDH hors Sm-ALDH	p. 32
IV- Cas de l'ALDH de <i>S. mutans</i> (EC 1.2.1.9)	p. 33
Aspects méthodologiques	p. 37
I- Introduction	p. 37
II- Cristallisation de macromolécules	p. 38
III- Enregistrement des données de diffraction	p. 40
A- Les sources de rayons X.....	p. 41
B- La diffraction des rayons X.....	p. 42
C- Les détecteurs de rayons X.....	p. 44
D- Stratégie d'enregistrement.....	p. 45
IV- Le traitement des données de diffraction	p. 46
A- Recherche des paramètres de maille et indexation du cliché de diffraction.....	p. 46

1- Recherche des paramètres de maille.....	p. 46
2- Indexation du cliché de diffraction.....	p. 47
B- Intégration des données.....	p. 48
1- Méthode de la « shoe box ».....	p. 48
2- Méthode du « profile fitting ».....	p. 48
C- Mise à l'échelle des données.....	p. 48
V- Détermination de la structure.....	p. 49
A- Détermination de la symétrie non cristallographique.....	p. 50
1- La fonction de Patterson.....	p. 50
2- La fonction d'autocorrélation	p. 51
3- La fonction de Patterson native.....	p. 51
B- Le remplacement moléculaire.....	p. 52
1- La fonction de rotation croisée.....	p. 53
2- La fonction de translation.....	p. 53
C- Le remplacement isomorphe.....	p. 55
D- Modification de la densité électronique.....	p. 56
1- Nivellement de solvant.....	p. 56
2- Positivité.....	p. 57
3- Moyennation de la densité.....	p. 57
VI- Affinement de la structure.....	p. 57
A- Affinement par corps rigide.....	p. 59
B- Minimisation d'énergie.....	p. 60
C- La dynamique moléculaire.....	p. 60
D- Affinement des facteurs de température.....	p. 62
E- Introduction des molécules d'eau et des ions.....	p. 62
F- Validation de la structure.....	p. 63
G- Conclusion.....	p. 63
Cristallisations.....	p. 65
I- Apoenzyme de type 1 (Apo1) et holoenzyme.....	p. 65
II- L'apoenzyme de type 2 ou Apo2.....	p. 68
III- Etude du complexe ternaire.....	p. 68
A- Les différentes approches.....	p. 68
B- Le choix des mutations.....	p. 69
C- Les conditions de cristallisation.....	p. 69
Enregistrements et traitements des données de diffraction.....	p. 71
I- L'Apo1 et l'holoenzyme.....	p. 71
A- Recherche des conditions de cryogénie.....	p. 71
B- Collecte des données.....	p. 72
1- L'apoenzyme.....	p. 72
2- L'holoenzyme.....	p. 72
C- Analyse des jeux de données.....	p. 72

1- L'apoenzyme.....	p. 73
2- L'holoenzyme.....	p. 74
II- L'Apo2.....	p. 75
III- Le complexe ternaire.....	p. 76
A- Les essais de diffusion: cas du mutant E250A.....	p. 76
B- Les essais de diffusion: cas des cristaux orthorhombiques du mutant C284S....	p. 77
C- Les essais de diffusion: cas des cristaux trigonaux du mutant C284S.....	p. 78
Résolution des structures et affinements.....	p. 79
I- Résolution de la structure de l'Apo1 et de l'holoenzyme.....	p. 79
A- L'apo1.....	p. 79
1- Etudes préliminaires.....	p. 79
<i>La fonction d'autocorrélation.....</i>	p. 79
<i>La Patterson native.....</i>	p. 80
2- Remplacement moléculaire avec l'ALDH3.....	p. 82
3- Remplacement moléculaire avec l'ALDH2.....	p. 83
<i>Reconstruction et affinement.....</i>	p. 83
<i>Nivellement de solvant et moyennation.....</i>	p. 84
<i>Le logiciel O.....</i>	p. 84
<i>Conclusion.....</i>	p. 85
4- Remplacement moléculaire avec l'ALDH9.....	p. 88
5- Reconstruction et affinement.....	p. 88
<i>Stratégie.....</i>	p. 88
<i>Quelques remarques sur le suivi de l'affinement.....</i>	p. 89
B- L'holoenzyme.....	p. 91
1- Remplacement moléculaire.....	p. 91
2- Reconstruction et affinement.....	p. 91
C- Qualité des structures de l'Apo1 et de l'holoenzyme.....	p. 93
D- L'empilement cristallin.....	p. 95
II- Résolution de la structure de l'Apo2.....	p. 100
A- Remplacement moléculaire et affinement.....	p. 100
1- Remplacement moléculaire.....	p. 100
2- Affinement.....	p. 100
3- Qualité de la structure.....	p. 101
III- Etude du complexe ternaire.....	p. 102
A- L'holoenzyme C284S.....	p. 102
B- Le complexe ternaire.....	p. 102
Analyses structurales et discussion.....	p. 105
I- L'Apo1 et l'holoenzyme.....	p. 105
A- Description globale de la structure de Sm-ALDH.....	p. 105
B- Le domaine de Rossmann.....	p. 109
1- Description structurale et caractéristiques générales.....	p. 109

2- Cas de l'ALDH de <i>S. mutans</i>	p. 110
C- Le changement local de conformation.....	p. 112
D- La conformation du cofacteur.....	p. 113
E- Le domaine catalytique.....	p. 114
F- Relation entre la structure du monomère des ALDH et le degré d'oligomérisation.....	p. 116
II- L'Apo2.....	p. 118
III- Le mutant C284S et le complexe ternaire.....	p. 119
A- Structure du mutant C284S.....	p. 119
B- Le complexe ternaire.....	p. 120
IV- Les GAPDH NADP dépendantes non phosphorylantes.....	p. 121
A- Analyse des alignements de séquences.....	p. 122
1- Les séquences.....	p. 122
2- Analyse des alignements de séquences.....	p. 123
B- Stabilisation du 2'-phosphate du NADP chez les déshydrogénases à domaine de Rossmann.....	p. 129
1- Cas de l'ALDH de <i>S. mutans</i>	p. 129
2- Comparaison avec les déshydrogénases NADP dépendantes.....	p. 129
V- Structures cristallines des formes apo et holo d'une ALDH NADP dépendante de <i>Streptococcus mutans</i>.....	p.131
<i>Apo and holo crystal structures of an NADP-dependent aldehyde dehydrogenase from <i>Streptococcus mutans</i>.....</i>	p. 133
Le mécanisme catalytique des ALDH.....	p. 147
I- Les acteurs de la réaction.....	p. 147
II- Rappel des mécanismes proposés.....	p. 148
III- Les résultats de la cristallographie et de l'enzymologie: cas de l'ALDH de <i>S.</i> <i>mutans</i>.....	p. 148
A- Apoenzyme, holoenzyme, activation de la cystéine.....	p. 151
B- Le complexe ternaire et l'étape d'acylation.....	p. 152
C- La déacylation.....	p. 153
IV- Généralisation du mécanisme à l'ensemble des ALDH hors ALDH3.....	p. 154
V- Investigations cristallographiques et biochimiques du mécanisme catalytique d'une aldéhyde déshydrogénase NADP dépendante de <i>Streptococcus mutans</i>.....	p. 155
<i>Structural and biochemical investigations of the catalytic mechanism of an NADP- dependent aldehyde dehydrogenase from <i>Streptococcus mutans</i>.....</i>	p. 157
Conclusion et perspectives.....	p. 185

Bibliographie.....p. 189
Annexes.....p. 201

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Abréviations

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

1,3 dPG, 1,3 diphosphoglycérate
3-PGA, 3-phosphoglycérate
Å, angström
ADN, acide désoxyribonucléique
ADP, adénosine diphosphate
ALDH, aldéhyde déshydrogénase
Apo, apoenzyme
ATP, adénosine triphosphate
BADH, bétaine aldéhyde déshydrogénase
CCP4, Collaborative Computational Project, Number 4
DO, densité optique
ESRF, European Synchrotron Radiation Facility
 F_{calc} , F_c facteur de structure calculé
 F_{obs} , F_o facteur de structure observé
FTHFDH, 10-formyltétrahydrofolate déshydrogénase
GAPDH, glycéraldéhyde-3-phosphate déshydrogénase
G3P, glycéraldéhyde-3-phosphate
HEPES, acide 4-(2-hydroxyéthyl)-piperazine-1-éthanesulfonique
Holo, holoenzyme
 k_{cat} , constante catalytique
kDa, kilodalton
 K_M , constante de Michaelis
LURE, Laboratoire pour l'Utilisation du Rayonnement Electromagnétique
MAD, multiwavelength anomalous diffusion
MASC, multiwavelength anomalous with solvent contrast
 μl , microlitre
mg, milligramme
ml, millilitre
mm, millimètre
mM, millimole par litre
M, mole par litre
MIR, multiple isomorphous replacement
MIRAS, multiple isomorphous replacement with anomalous scattering
MMSDH, méthylmalonate semialdéhyde déshydrogénase
NAD, nicotinamide mononucléotide diphosphate
NADP, 2'-phosphate nicotinamide mononucléotide diphosphate
PEG, polyéthylène glycol
Pi, phosphate inorganique
PI, point isoélectrique
 R_{cryst} , facteur d'accord crystallographique
 R_{free} , facteur d'accord crystallographique sur n % des données omises dans l'affinement
Rms, root mean square deviation
RX, rayons X
s, seconde
SDXM, service commun de diffraction X sur monocristaux
SIR, single isomorphous replacement

SIRAS, single isomorphous replacement with anomalous scattering
SSDH, succénique semialdéhyde déshydrogénase

Code une lettre et trois lettres des acides aminés:

G, Gly, glycine
A, Ala, alanine
V, Val, valine
L, Leu, leucine
I, Ile, isoleucine
S, Ser, sérine
T, Thr, thréonine
C, Cys, cystéine
M, Met, méthionine
D, Asp, acide aspartique
N, Asn, asparagine
E, Glu, acide glutamique
Q, Gln, glutamine
K, Lys, lysine
R, Arg, arginine
H, His, histidine
F, Phe, phénylalanine
Y, Tyr, tyrosine
W, Trp, tryptophane
P, Pro, proline

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Liste des figures

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

1.1	Compartimentalisation et fonction biologique de différentes GAPDH.....	p. 23
1.2	Représentation en rubans et flèches du dimère de l'ALDH3.....	p. 26
1.3	Vue stéréoscopique du domaine de Rossmann de l'ALDH3 en interaction avec le NAD dans un nouveau mode de fixation.....	p. 27
1.4	Vue stéréoscopique du domaine de Rossmann de la GAPDH NAD dépendante de <i>Bacillus stearothermophilus</i> en interaction dans le mode classique avec le NAD.....	p. 27
1.5	Tétramère de l'ALDH2 représenté en rubans et flèches.....	p. 29
1.6	Vue stéréoscopique des cofacteurs de l'ALDH9 et de l'ALDH3 après superposition des domaines de Rossmann.....	p. 31
1.7	Cycle des pentoses phosphate.....	p. 34
2.1	Cristallisation par la méthode de la goutte suspendue.....	p. 40
2.2	Représentation schématique des étapes de la dynamique moléculaire.....	p. 61
3.1	Cristaux d'apoenzyme de l'ALDH de <i>S. mutans</i>	p. 66
3.2	Cristal d'holoenzyme obtenu en présence de sulfate d'ammonium.....	p. 67
5.1	Section $\kappa = 180^\circ$ de la fonction d'autocorrélation calculée entre 10 et 5 Å sur le jeu de données de l'apoenzyme.....	p. 80
5.2	Section de Harker $v = 0,5$	p. 81
5.3	Représentation schématique de la position des axes 2 non cristallographiques reliant les monomères des tétramères de l'ALDH dans la maille orthorhombique déduite de la fonction d'autocorrélation et de la Patterson native.....	p. 82
5.4	Tétramère de l'ALDH de <i>S. mutans</i> en association avec les ions sulfate.....	p. 90
5.5	Vue stéréoscopique d'un ion sulfate dans sa densité électronique.....	p. 91
5.6	Vue stéréoscopique de la densité électronique du NADP.....	p. 92
5.7	Carte de Ramachandran de l'Apo1.....	p. 93
5.8	Carte de Ramachandran de l'holoenzyme (forme orthorhombique)	p. 94

5.9 Vue stéréoscopique de l'empilement cristallin des tétramères d'ALDH dans la maille orthorhombique	p. 95
5.10 Vue stéréoscopique de l'empilement cristallin des tétramères d'ALDH dans la maille trigonale.....	p. 95
5.11 Diagramme de Ramachandran de l'Apo2.....	p. 101
5.12 Vue stéréoscopique de la carte F_o-F_c contourée à -3 et +3 σ après incorporation des ions sulfate dans la structure de l'Apo2.....	p. 102
6.1 Tétramère de l'holoenzyme de l'ALDH de <i>S. mutans</i> en interaction avec le NADP..	p. 106
6.2 Topologie du monomère de l'ALDH de <i>S. mutans</i>	p.106
6.3 Monomère de l'holoenzyme de Sm-ALDH et ses différents domaines.....	p. 108
6.4 Superposition de la première hélice α et du premier brin β du domaine de Rossmann de l'ALDH1, l'ALDH2, l'ALDH3, l'ALDH9 et de l'ALDH de <i>S. mutans</i>	p. 111
6.5 Représentation schématique du NADP	p. 113
6.6 Vue stéréoscopique des interactions entre un ion sulfate et les acides aminés du site catalytique de l'holoenzyme.....	p. 115
6.7 Représentation du G3P et du potentiel électrostatique du site catalytique de l'holoenzyme.....	p. 115
6.8 Superposition du dimère de l'ALDH3 et du dimère correspondant de l'ALDH de <i>S. mutans</i>	p. 116
6.9 Vue stéréoscopique de la superposition des deux ions sulfate de l'Apo2 et du G3P du complexe ternaire.....	p. 121
6.10 Alignement de sept séquences de GAPDH NADP dépendantes non phosphorylantes.....	p. 124-125
6.11 Alignement de 4 séquences de GAPDH NADP dépendantes non phosphorylantes p.	126
7.1 Schéma de la réaction catalytique catalysée par l'ALDH de <i>S. mutans</i>	p. 149-151
7.2 Vue stéréoscopique des cofacteurs de l'ALDH1, l'ALDH2, l'ALDH3, l'ALDH9 et de l'ALDH de <i>S. mutans</i> positionnés après superposition des domaines de Rossmann.	p. 154

Liste des tableaux et graphiques

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

1.1 Classification des ALDH humaines d'après Yoshida <i>et al.</i> (1998).....	p. 20
4.1 Statistiques des enregistrements de données sur un cristal d'apoenzyme de l'ALDH de <i>S. mutans</i>	p. 73
4.2 Statistiques d'un enregistrement de données à 3 Å sur un cristal de la forme trigonale de l'holoenzyme de l'ALDH de <i>S. mutans</i>	p. 74
4.3 Statistiques de l'enregistrement de données à 2,6 Å sur un cristal de la forme orthorhombique de l'holoenzyme de l'ALDH de <i>S. mutans</i>	p. 75
4.4 Statistiques des enregistrements de données de l'Apo2 dans deux conditions différentes de cristallisation.....	p. 76
4.5 Statistiques de l'enregistrement de données sur le mutant C284S (système orthorhombique).....	p. 77
4.6 Statistiques des enregistrements de données sur le mutant C284S et le complexe ternaire (système trigonal).....	p. 78
5.1 Evolution des facteurs d'accord au cours de l'affinement de la structure de l'apoenzyme (Apo1) à 1,8 Å de résolution.....	p. 89
5.2 Ecart rms sur différents paramètres structuraux pour les structures de l'Apo1 et de l'holoenzyme.....	p. 94
5.3 Interactions intermoléculaires entre les tétramères de l'holoenzyme dans le système orthorhombique.....	p. 96-98
5.4 Interactions intermoléculaires entre les tétramères de l'holoenzyme dans le système trigonal.....	p. 98-100
6.1 Topologie du monomère de l'ALDH de <i>S. mutans</i>	p. 107
6.2 Listes des interactions par liaisons hydrogène et ponts salins entre les différents monomères de l'ALDH de <i>S. mutans</i> en prenant comme référence le monomère A.....	p. 108-109
6.3 Groupements CH ₂ /CH ₃ présents dans une sphère de 5 Å autour de Glu250 pour le mutant C284S.....	p. 119

- 6.4 Interactions par liaisons hydrogène entre le substrat et la protéine pour le monomère A du complexe ternaire.....p. 120
- 6.5 Interactions entre le cofacteur et les acides aminés observées dans la structure de l'holoenzyme de Sm-ALDH.....p. 127
- 6.6 Interactions par liaisons hydrogène entre le 2'-phosphate du NADP et les acides aminés du domaine de Rossmann de plusieurs déshydrogénases NADP dépendantes.....p. 130

S.C.D. - U.N.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Introduction

Les aldéhyde déshydrogénases (ALDH) sont des enzymes ubiquitaires qui oxydent les aldéhydes en leurs acides carboxyliques correspondants avec réduction des cofacteurs à nucléotide tels que le NAD et le NADP (Croteau *et al.*, 1996). Les ALDH possèdent un très large spectre de substrats. L'étude de leur structure, nécessaire à l'interprétation des résultats enzymatiques, a débuté au début des années 90 (Rose *et al.*, 1990, Hurley & Weiner, 1992, Baker *et al.*, 1994). La première structure d'une ALDH ne sera résolue qu'en 1997 (Liu *et al.*, 1997). Actuellement les structures de huit ALDH sont disponibles dans la Protein Data Bank (Liu *et al.*, 1997; Steinmetz *et al.*, 1997; Johansson *et al.*, 1998; Moore *et al.*, 1998; Lamb & Newcomer, 1999; Cobessi *et al.*, 1999; Zhang *et al.*, 1999; Ni *et al.*, 1999). Parmi ces huit structures, deux sont dimériques et six sont tétramériques. Ces enzymes diffèrent par leur spécificité de substrat et de cofacteur, leur organisation oligomérique, ainsi que par certaines particularités structurales, telles que l'addition de boucles pour certaines ALDH. Des particularités cristallographiques sont également observées, telles que des densités électroniques mal définies au niveau du cofacteur ou au niveau de certains acides aminés essentiels à la catalyse.

L'investigation du mécanisme enzymatique des ALDH a débuté en 1949 (Racker, 1949). Il n'a été résolu que très récemment. Participent à celui-ci deux résidus essentiels: une cystéine et un glutamate (Védadi & Meighen, 1997). Comme la réaction catalytique se déroule en deux étapes, une acylation suivie d'une déacylation, il est nécessaire que la

cystéine soit accessible et activée afin de former l'espèce thiolate qui attaquera la fonction aldéhydique portée par le substrat. L'étape de déacylation est due à l'attaque d'une molécule d'eau sur le thioacylenzyme. La molécule d'eau est activée et positionnée par le glutamate (Marchal *et al.*, 2000). La variation des pKas de ces deux résidus peut être expliquée par l'étude des différentes structures de l'ALDH de *Streptococcus mutans*. En effet, des changements d'environnement protéique pour ces deux acides aminés sont observables pour l'holoenzyme de type sauvage en comparaison avec l'apoenzyme mais aussi au sein de différents mutants.

La structure de l'ALDH NADP dépendante de *S. mutans* constitue la première structure d'une ALDH tétramérique NADP dépendante (Cobessi *et al.*, 1999). Elle est également la première structure d'une GAPDH NADP dépendante non phosphorylante. Actuellement, les structures de la GAPDH NAD dépendante phosphorylante (eubactérie ou eucaryote) (Moras *et al.*, 1975; Skarzynsky *et al.*, 1988), de la GAPDH NADP dépendante phosphorylante d'archaebactérie (Isupov *et al.*, 1999, Charron *et al.*, 1999), et enfin de la GAPDH NADP dépendante non phosphorylante sont connues. Récemment, des cristaux de la GAPDH NADP dépendante phosphorylante du chloroplaste, ont été décrits (Sabatino *et al.*, 1999).

L'ensemble des résultats structuraux obtenus sur l'ALDH de *S. mutans* (Sm-ALDH) en association avec quelques résultats de l'enzymologie permettant la compréhension du mécanisme enzymatique des ALDH sont présentés dans ce manuscrit.

Le premier chapitre, constitué d'une étude bibliographique, décrit quelques activités aldéhyde déshydrogénase, puis quelques structures d'ALDH présentées selon l'ordre chronologique de leur résolution. Il résume également l'ensemble des mécanismes catalytiques proposés pour les ALDH. Cette première partie s'achève par la présentation de l'ALDH de *S. mutans*.

Les deuxième, troisième, quatrième et cinquième parties de ce mémoire présentent les méthodes cristallographiques utilisées, les essais de cristallisation, les enregistrements et les traitements de données ainsi que les résultats préliminaires obtenus sur Sm-ALDH. La sixième partie est consacrée à l'analyse structurale de l'ALDH de *S. mutans* et est constituée de résultats cristallographiques et de discussions non publiés dans l'article « *Apo and holo crystal structures of an NADP-dependent aldehyde dehydrogenase from Streptococcus*

mutans ». Ce premier article, publié dans *Journal of Molecular Biology* en 1999, est consacré à l'analyse cristallographique des structures de l'apo et de l'holoenzyme de Sm-ALDH. Il est inclus dans le manuscrit à la fin du sixième chapitre.

Une septième partie présente le mécanisme catalytique des ALDH en tenant compte des résultats de la cristallographie et de l'enzymologie. Elle inclut l'article « *Structural and biochemical investigations of the catalytic mechanism of an NADP-dependent aldehyde dehydrogenase from Streptococcus mutans* » accepté à *Journal of Molecular Biology*. Ce second article rassemble différents résultats enzymatiques et cristallographiques obtenus sur Sm-ALDH et montre de légers changements conformationnels pouvant intervenir au cours de l'acte catalytique. Il met également en évidence deux facteurs structuraux probablement responsables du nouveau mode de fixation du NAD(P) au domaine de Rossmann.

Enfin, une dernière partie intitulée conclusion et perspectives achève ce manuscrit.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

Les aldéhyde déshydrogénases

Les ALDH constituent une grande famille d'enzymes dont les premières études ont débuté en 1949 (Racker, 1949). L'évolution des techniques de génie génétique a permis de caractériser leur présence dans un grand nombre d'organismes et de séquencer de très nombreux gènes les exprimant (Hsu *et al.*, 1985; Hempel *et al.*, 1985; Johansson *et al.*, 1988; Jones *et al.*, 1988; Farrès *et al.*, 1989; Hsu & Chang, 1991; Miyauchi *et al.*, 1991; Hsu *et al.*, 1992; Kedishvili *et al.*, 1992; Kurys *et al.*, 1993). Les différentes études biochimiques ont montré que les ALDH catalysent l'oxydation des aldéhydes en acides carboxyliques avec réduction des cofacteurs NAD ou NADP associés à ces réactions. Récemment, un alignement de 145 séquences d'ALDH a été réalisé dans le cadre d'un projet intitulé : **Structure Function Relationships Among Aldehyde Dehydrogenases** (Perozich *et al.*, 1999). Une soixantaine d'autres séquences seront prochainement additionnées à cet alignement. La réduction de l'alignement à un alignement de 16 séquences en regroupant les 145 premières par classes d'enzymes a permis de mettre en évidence un certain nombre d'acides aminés plus ou moins conservés dont les acides aminés catalytiques (cystéine et glutamate) et d'autres importants dans la stabilisation du cofacteur (glutamate, lysine) ou supposés comme tels pour la stabilisation d'un intermédiaire catalytique (asparagine). Plusieurs de ces résidus seront mentionnés dans les paragraphes suivants. Enfin, certains résidus (proline et glycine) sont conservés en raison de leur participation à l'élaboration des structures secondaires. L'alignement fait apparaître l'existence d'au moins 14 classes d'ALDH (Perozich *et al.*, 1999). Compte tenu du nombre de séquences d'ALDH déterminées, la présentation suivante

ne se veut pas exhaustive. Elle est constituée de différents paragraphes résumant différents rôles des ALDH et décrivant quelques structures. Les exemples les illustrant ont été sélectionnés en rapport avec l'étude structurale de l'ALDH de *S. mutans*.

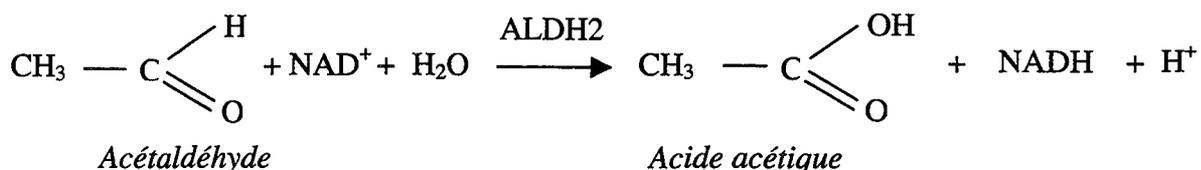
I- Aperçu des fonctions enzymatiques des ALDH

A- Les ALDH chez les mammifères

Les ALDH chez les mammifères sont réparties en différentes classes en fonction des pourcentages d'identité de séquences entre elles et en fonction de leur localisation cellulaire (Hempel *et al.*, 1989; Yoshida *et al.*, 1998). Le tableau 1.1 présente les douze classes d'ALDH répertoriées chez l'être humain, ainsi que leur localisation cellulaire et leur substrat principal (Yoshida *et al.*, 1998). Certaines de ces protéines font l'objet d'études enzymatiques très importantes en raison de leur implication dans certains processus et pathologies cellulaires tels que: le métabolisme de l'éthanol et l'alcoolisme, le syndrome de Sjogren-Larsson (Yoshida *et al.*, 1998), la différenciation cellulaire.

1- L'aldéhyde déshydrogénase mitochondriale ou ALDH2

L'ALDH2, enzyme tétramérique NAD dépendante d'environ 210 kDa, est responsable de l'oxydation de l'acétaldéhyde produit lors du métabolisme de l'éthanol, par l'alcool déshydrogénase (ADH), selon la réaction:

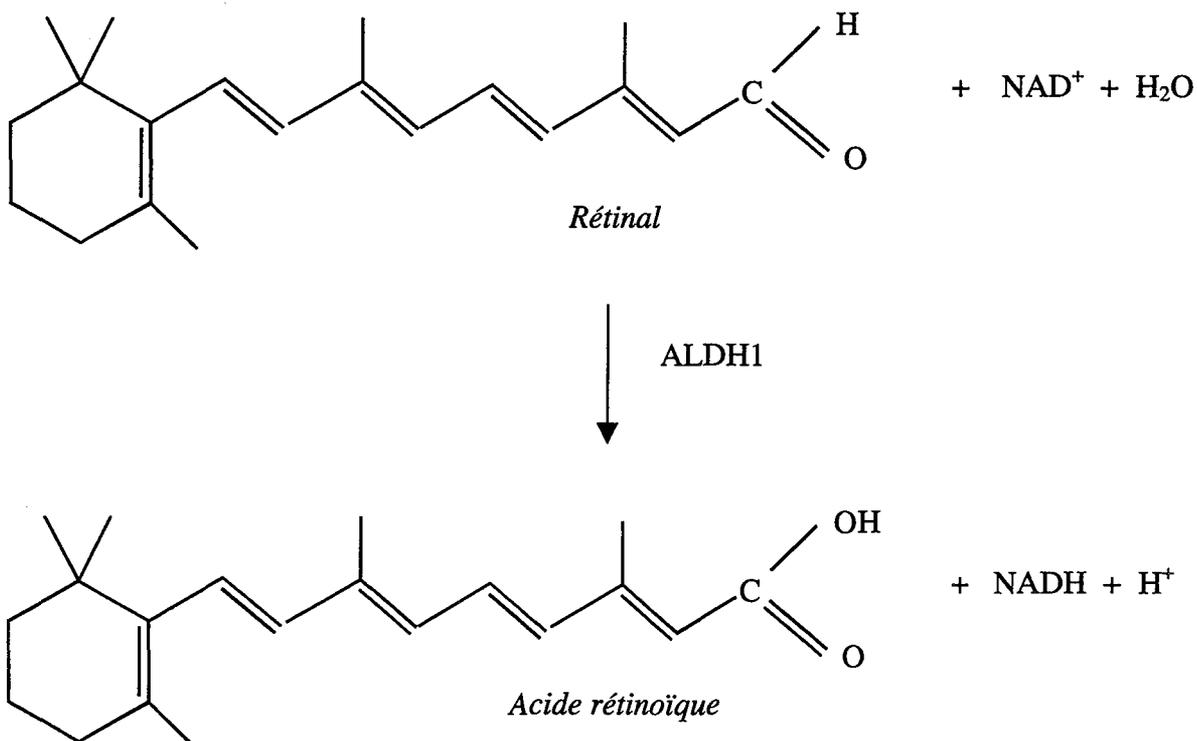


Son étude est importante en raison de son implication dans le métabolisme de l'éthanol. Sa caractérisation biochimique a permis de révéler l'importance d'une mutation du résidu Glu487 en lysine associée à une intolérance vis-à-vis de l'éthanol dans 50 % de la population asiatique (Steinmetz *et al.*, 1997). Cette mutation rendant inactive l'enzyme serait

à l'origine de l'accroissement du taux d'acétaldéhyde dans le sang entraînant alors une intolérance pour l'éthanol. La détermination récente de la structure tridimensionnelle de ses formes apo et holoenzyme a permis de mieux caractériser le mécanisme catalytique auquel participe une cystéine (Cys302), un glutamate (Glu268), et une molécule d'eau (Steinmetz *et al.*, 1997). Ces deux derniers points seront abordés dans un prochain paragraphe. Les ALDH2 partagent entre elles plus de 90 % d'identité de séquences et environ 70 % avec les ALDH cytosoliques ou ALDH1s (Hempel *et al.*, 1989). Ce type d'ALDH est codé par l'ADN nucléaire. Son transport dans la mitochondrie est assuré par un peptide signal que la mutation Leu2 → Val abolit (Hammen *et al.*, 1999). Ainsi, l'exportation de la protéine vers la mitochondrie serait essentiellement due au caractère hydrophobe de la séquence leader (Hammen *et al.*, 1999).

2- L'aldéhyde déshydrogénase cytosolique ou ALDH1

Tétramère de 210 kDa et NAD dépendante, elle participe au développement et à la différenciation cellulaire par l'oxydation du rétinol en acide rétinoïque (Moore *et al.*, 1998).



Gène (symbole)	Enzyme	Tissus et localisation cellulaire	Substrat principal	Séquence codante (acides aminés)	Taille de la sous unité (acides aminés)	Code de la GenBank
<i>aldh1</i>	ALDH1	foie, estomac... (cytosol)	rétinal	501	500	J04748
<i>aldh2</i>	ALDH2	foie, estomac... (mitochondrie)	acétaldéhyde	517	500	M20444
<i>aldh3</i>	ALDH3	estomac, poumons... (cytosol)	aldéhydes gras et aromatiques	453	452	M77477
<i>aldh4</i>	ALDH4	foie, reins (mitochondrie)	glutamate γ -semialdéhyde	563	539?	U24266
<i>aldh5</i>	ALDH5	testicules, foie (mitochondrie)	propionaldéhyde	517	500	M63967
<i>aldh6</i>	ALDH6	glandes salivaires, estomac, reins (cytosol)	aldéhyde aliphatique, rétinal	512	511	U07919
<i>aldh7</i>	ALDH7	reins, poumons (microsome)	aldéhydes aliphatiques et aromatiques	468	467	U10868
<i>aldh8</i>	ALDH8	parotides (microsome)	inconnu	451	450	U37519
<i>aldh9</i>	ALDH9	foie, reins, muscles (cytosol)	aldéhyde aminé	493	492	U34252
<i>aldh10</i>	ALDH10	foie, cœur, muscles (microsome)	aldéhydes gras et aliphatiques	485	484	U46689
<i>ssdh</i>	SSDH	cœur, foie (mitochondrie)	semialdéhyde succénique	inconnue	488	L34821
<i>mmsdh</i>	MMSDH	reins, foie, cœur (mitochondrie)	méthylmalonate semialdéhyde	535	503	M93405

Tableau 1.1: classification des ALDH humaines d'après Yoshida et al. (1998).

L'ALDH1 catalyse également l'oxydation de l'acétaldéhyde en acide acétique avec cependant un K_M plus élevé pour ce substrat par rapport à celui mesuré avec le *trans* ou le *cis* rétinol (Moore *et al.*, 1998).

Leur caractérisation montre les mêmes résidus que ceux cités précédemment impliqués dans la catalyse (Glu268 et Cys302). Les rétinolaldéhyde déshydrogénases peuvent être divisées en trois groupes en fonction de leur séquence et de leur pI (Moore *et al.*, 1998). La détermination récente de la structure tridimensionnelle de deux des trois types d'ALDH1 a permis d'émettre de nouvelles hypothèses sur un possible changement local de conformation lors de la catalyse (Moore *et al.*, 1998; Lamb & Newcommer, 1999).

3- L'aldéhyde déshydrogénase de la classe 3 ou ALDH3

L'ALDH3 est dimérique et possède moins de 30 % d'homologie de séquences avec les deux classes précédentes (Rose *et al.*, 1990). Ce fut la première structure tridimensionnelle d'une ALDH résolue. Un glutamate et une cystéine participent également à la réaction catalytique. L'ALDH3 est inductible par des agents carcinogènes (Jones *et al.*, 1988).

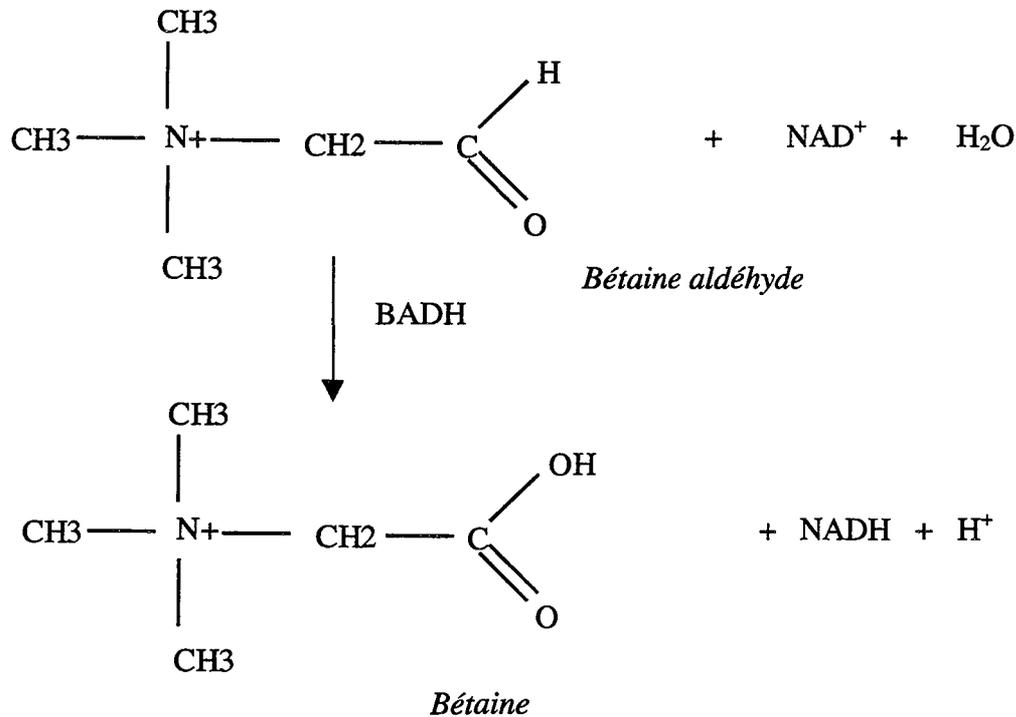
B- Les ALDH chez les végétaux

Deux exemples sont présentés dans ce paragraphe: la bétaine aldéhyde déshydrogénase et la glycéraldéhyde-3-phosphate déshydrogénase NADP dépendante non phosphorylante.

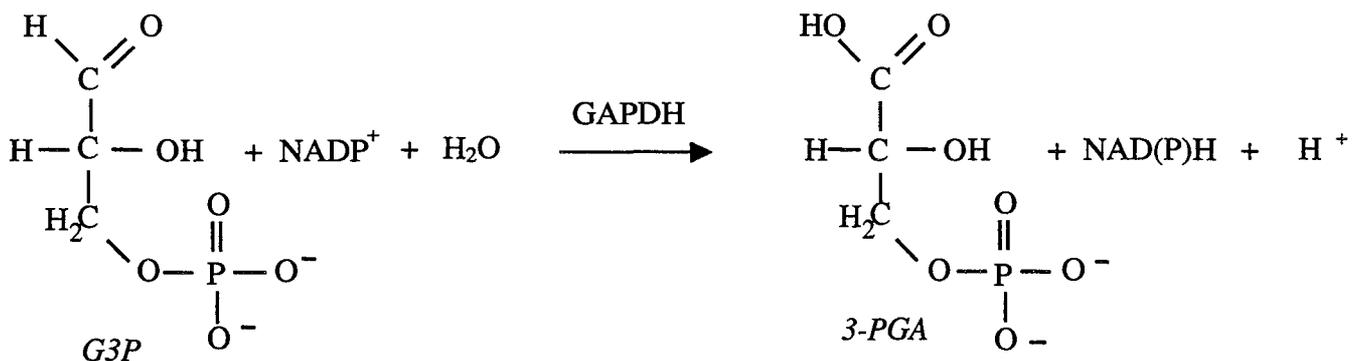
1- La bétaine aldéhyde déshydrogénase ou BADH

La BADH est un dimère de 60-63 kDa par monomère qui synthétise la bétaine par oxydation de la bétaine aldéhyde (Weretilnyk & Hanson, 1990). Elle est localisée dans le stroma du chloroplaste. Le gène l'exprimant est nucléaire ce qui suggère donc la présence d'un peptide signal permettant son exportation du cytosol vers le chloroplaste. La synthèse de bétaine est une réponse au stress hydrique chez les végétaux (Legaria *et al.*, 1998). La bétaine agit comme un osmolyte cytoplasmique permettant un fonctionnement normal du métabolisme cellulaire (Weretilnyk & Hanson, 1990). Ces protéines sont classées dans la

famille des ALDH sur la base des alignements de séquences. Ces alignements montrent non seulement la conservation de la cystéine essentielle à la catalyse (Cys302, ALDH2) mais aussi la conservation de dix résidus situés entre le domaine catalytique et celui du cofacteur dont le glutamate (Glu268, ALDH2) qui aurait un rôle de base durant la réaction catalytique (Weretilnyk & Hanson, 1990). Les pourcentages d'identité de séquences entre les BADHs et diverses ALDH s'étendent de 30 % à 40 %.



2- La GAPDH NADP dépendante non phosphorylante



Elle catalyse l'oxydation irréversible du glycéraldéhyde-3-phosphate (G3P) en 3-phosphoglycérate (3-PGA) en réduisant son cofacteur le NADP en NADPH (Crow & Wittenberger, 1979).

Le produit de la réaction est transféré vers le stroma du chloroplaste où il est utilisé dans le cycle de Calvin-Benson pour la production de 1,3-diphosphoglycérate (1,3-dPG) qui sera utilisé pour produire du glycéraldéhyde-3-phosphate (G3P) dont une partie est mobilisée pour la synthèse de polysaccharides.

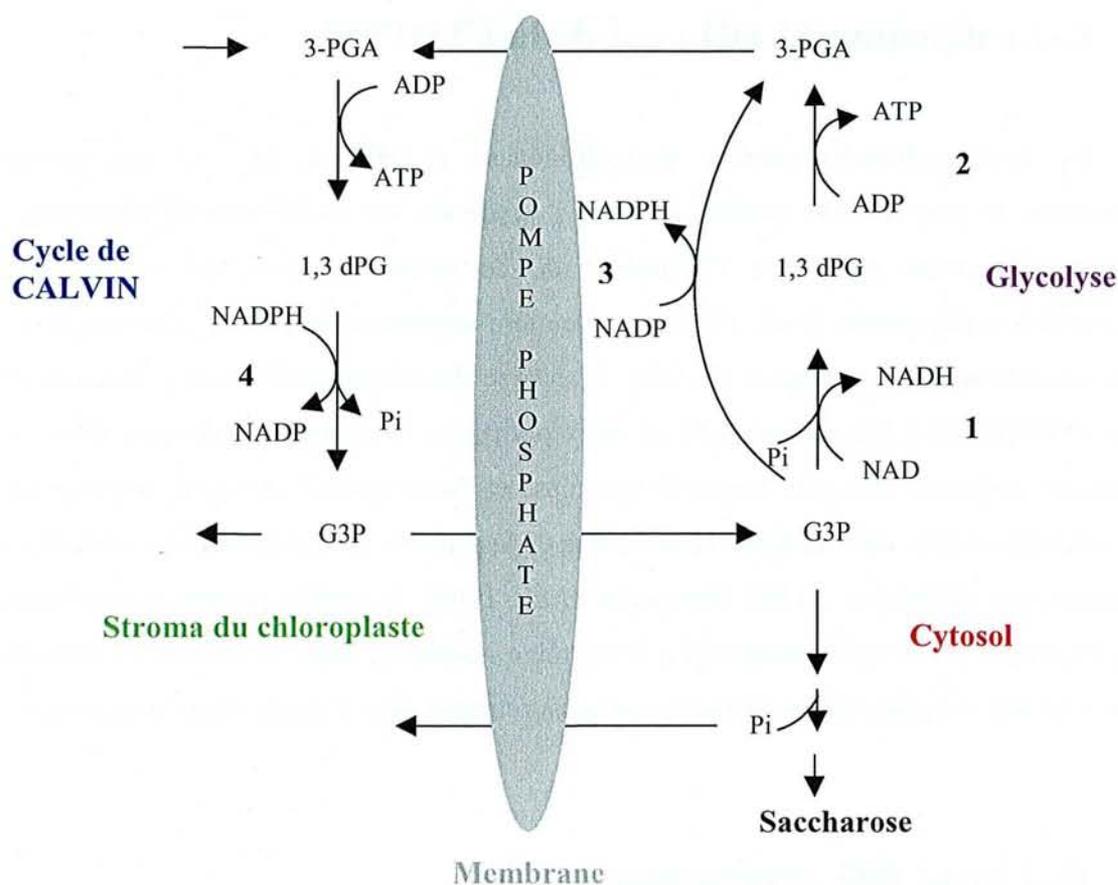


Figure 1.1: Compartimentalisation et fonction biologique de différentes GAPDH (Habenicht et al., 1994). Seules les réactions principales *in vivo* sont indiquées. 1, GAPDH NAD dépendante phosphorylante (GAPC); 2, Phosphoglycérate kinase (PGK); 3, GAPDH NADP dépendante non phosphorylante (GAPN); 4, GAPDH NADP dépendante phosphorylante (GAPA/GAPB).

Cette enzyme d'environ 200 kDa a été classée dans la famille des ALDH sur la base des alignements de séquences (Habenicht et al., 1994). Elle ne possède aucune similarité de séquences avec les GAPDH NAD(P) dépendantes phosphorylantes d'environ 145 kDa qui

convertissent de façon réversible le G3P, non pas en 3-PGA, mais en 1,3-dPG lors de la glycolyse par phosphorylation oxydative. La GAPDH NADP dépendante non phosphorylante est également présente chez *Streptococcus mutans* (Boyd *et al.*, 1995), streptocoque déficient dans une voie de biosynthèse du NADPH. Mais elle a surtout été étudiée chez de nombreux organismes photosynthétiques tels que l'épinard (Iglesias *et al.*, 1988), *Chlorella fusca* (Serrano *et al.*, 1991), *Chlamydomonas reinhardtii* (Iglesias *et al.*, 1987)...

C-Le domaine ALDH : cas de la FTHFDH

La 10-formyltétrahydrofolate déshydrogénase (FTHFDH, EC 1.5.1.6), protéine tétramérique de plus de 900 résidus, catalyse l'oxydation du 10-formyltétrahydrofolate en tétrahydrofolate mais également l'hydrolyse de ce même composé en formate et en tétrahydrofolate (Krupenko *et al.*, 1995). La première oxydation est NADP dépendante alors que la seconde requiert également du CO₂. L'analyse des alignements de séquences montre que la FTHFDH peut être décomposée en deux domaines. Cette enzyme possède deux sites catalytiques distincts: l'un pour l'activité hydrolase et l'autre pour l'activité déshydrogénase. Le second site se situe dans la partie C-terminale de la protéine qui possède 46 % d'identité de séquences avec différentes ALDH (Krupenko *et al.*, 1995). L'activité correspondante met en jeu la cystéine Cys707 qui correspond à la cystéine impliquée dans la catalyse et conservée chez les ALDH. Le glutamate catalytique est aussi retrouvé dans l'alignement de séquences.

D- Les ALDH structurales

La présence d'aldéhyde déshydrogénases dans l'œil des vertébrés et des invertébrés a également été rapportée (Tomarev *et al.*, 1991; Zinovieva *et al.*, 1993; Graham *et al.*, 1996). Dans ce cas, les ALDH n'ont plus de rôle catalytique mais un rôle structural ce qui laisse penser à un type de régulation particulier ou bien à la présence de mutations empêchant cette activité. Par exemple, dans le cas de l' Ω -cristalline, rapportée comme étant une ALDH1, la cystéine catalytique est mutée en arginine.

E- Les ALDH chez les bactéries

Diverses ALDH bactériennes sont étudiées: la lactaldéhyde déshydrogénase, la benzaldéhyde déshydrogénase (Perozich *et al.*, 1999), la glycéraldéhyde-3-phosphate déshydrogénase NADP dépendante non phosphorylante, une ALDH dimérique de *Vibrio harveyi* spécifique des aldéhydes à longue chaîne carbonée (Zhang *et al.*, 1999)...

Actuellement, quelques séquences d'ALDH d'archae sont également connues comme celles des GAPDH NAD(P) dépendantes non phosphorylantes de *Thermoproteus tenax* (Brunner *et al.*, 1998), *Methanobacterium thermoautotrophicum* (Smith *et al.*, 1997) et de *Methanococcus jannaschii* (Bult *et al.*, 1996).

II- Les structures de quelques ALDH

A- L'ALDH de la classe 3 (Liu *et al.*, 1997)

Cette première structure d'ALDH (ALDH3 du rat) a été résolue, sous sa forme holoenzyme en complexe avec le NAD, en 1997 par la méthode SIRAS (single isomorphous replacement with anomalous scattering) (Liu *et al.*, 1997). La structure de l'ALDH3 peut se décomposer en deux monomères reliés par une symétrie non cristallographique d'ordre 2 (Figure 1.2). Son monomère peut être divisé en trois domaines:

- le domaine de fixation du cofacteur
- le domaine d'oligomérisation
- le domaine catalytique

Le monomère débute par un groupement de 4 hélices α suivi d'un domaine, composé de 5 brins β parallèles ceinturés de part et d'autre de 2 hélices α : le repliement de Rossmann (Rossmann *et al.*, 1974). Bien que le groupement N-terminal de 4 hélices α soit proche du domaine de Rossmann dans l'espace, il en est séparé dans la séquence par le premier brin β et la longue boucle du domaine d'oligomérisation, le second brin de ce domaine étant composé des acides aminés de la partie C-terminale. En vis-à-vis du domaine de Rossmann se trouve le domaine catalytique composé de brins β et d'hélices α .

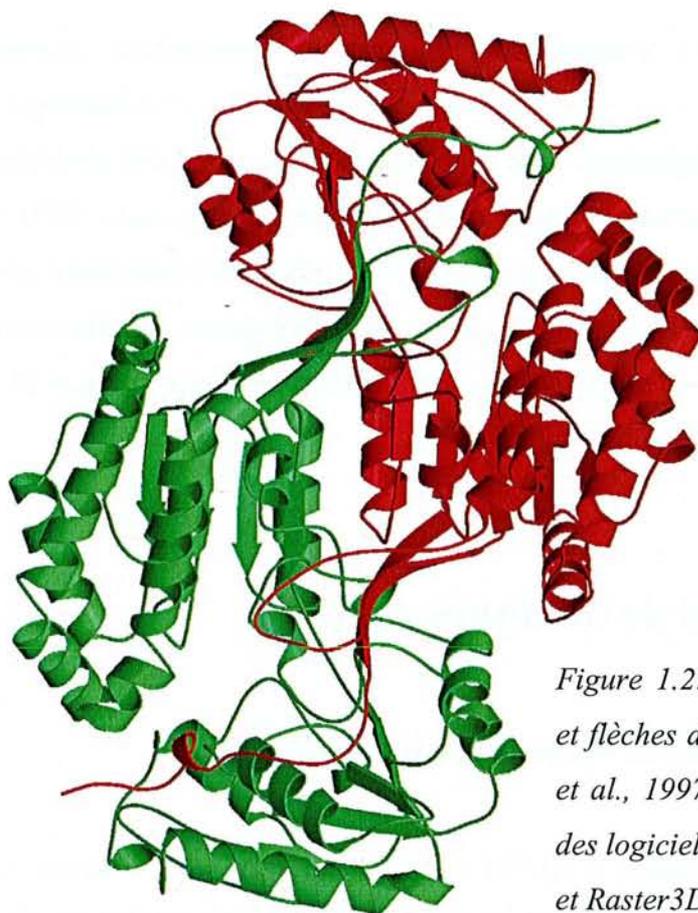


Figure 1.2: représentation en rubans et flèches du dimère de l'ALDH3 (Liu *et al.*, 1997). Dessin effectué à l'aide des logiciels Molscript (Kraulis, 1991) et Raster3D (Meritt & Murphy, 1994).

L'analyse de la structure a permis de montrer un nouveau mode de fixation d'un cofacteur à nucléotide au domaine de Rossmann (Figure 1.3). L'interaction entre le cofacteur et le domaine de Rossmann est différente de celle observée couramment chez les déshydrogénases possédant un repliement de Rossmann (Figure 1.4). Chez ces déshydrogénases, les nucléosides du cofacteur sont situés de part et d'autre du feuillet β du domaine de Rossmann et le groupement pyrophosphate du cofacteur est en général en interaction avec la boucle GXGXXG, empreinte du domaine de Rossmann, qui se situe entre la première hélice α et le premier brin β de ce domaine (Wierenga *et al.*, 1986). La stabilisation du groupement pyrophosphate est assurée par le dipôle de la première hélice α du domaine de Rossmann. Chez les ALDH, le cofacteur enjambe le domaine de Rossmann dont l'empreinte GX₄G se situe dans une boucle reliant le quatrième brin β à la quatrième et dernière hélice α . Il n'y a aucune stabilisation du groupement pyrophosphate par le dipôle d'une hélice α ou par des résidus chargés positivement.

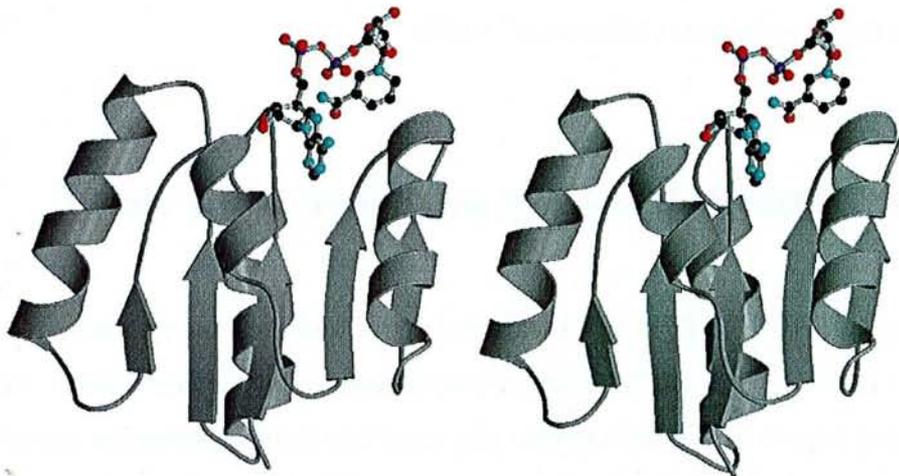


Figure 1.3: vue stéréoscopique du domaine de Rossmann de l'ALDH3 en interaction avec le NAD dans un nouveau mode de fixation.

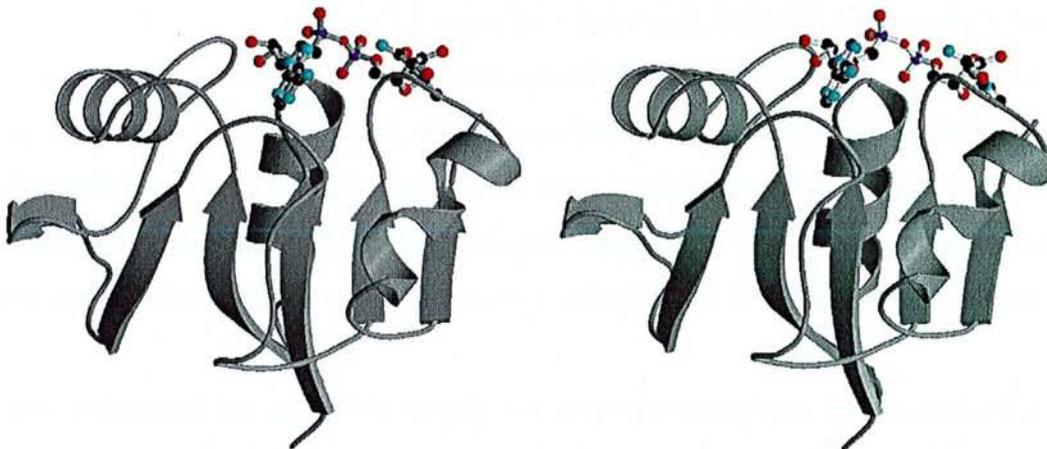


Figure 1.4: vue stéréoscopique du domaine de Rossmann de la GAPDH NAD dépendante de *Bacillus stearothermophilus* en interaction dans le mode classique avec le NAD.

L'analyse de la structure du monomère révèle les positions de la cystéine et du glutamate, acides aminés impliqués dans la réaction catalytique. Le glutamate (Glu209) se situe à 7 Å de la cystéine (Cys243). Cependant, cette analyse n'a pas conduit les auteurs à émettre des hypothèses quant à la réaction catalytique.

L'étude de cette structure montre une orientation *syn* adoptée par le nicotinamide en relation avec un transfert d'hydrure de type pro-*S* alors que le résultat des études RMN optait

pour un transfert de type *pro-R* et donc laissait envisager une conformation de type *anti* du nicotinamide mononucléoside (Jones *et al.*, 1987).

B- L'ALDH de la classe 2 (Steinmetz *et al.*, 1997)

La structure de l'ALDH mitochondriale bovine a été résolue en même temps que celle de l'ALDH3 mais par la méthode MIR (multiple isomorphous replacement). Elle a été tout d'abord résolue sous la forme holoenzyme puis sous une forme apoenzyme par élimination du NAD présent dans les cristaux (Steinmetz *et al.*, 1997). Le mode de fixation du NAD est semblable à celui décrit précédemment. Sa stabilisation diffère par la présence du dipôle de la dernière hélice α du domaine de Rossmann qui stabilise le groupement pyrophosphate. Par contre, la conformation du nicotinamide mononucléoside (*anti*) est différente de celle observée chez l'ALDH3. Comme la soustraction du NAD au cristal se produit sans perturbation du réseau cristallin, il est possible qu'aucun changement conformationnel important n'intervienne lors de la fixation du cofacteur à l'enzyme.

La première structure d'une ALDH tétramérique est décrite comme une association de dimères présentant une symétrie non cristallographique 222 (Figure 1.5). Le monomère a été divisé en trois domaines comparables à ceux décrits précédemment. Une différence existe cependant entre les monomères de ces deux enzymes: l'ALDH2 possède en N-terminal quatre brins β de plus que l'ALDH3 et son domaine d'oligomérisation est un feuillet β antiparallèle à 3 brins.

L'analyse du site catalytique a permis aux auteurs d'émettre une hypothèse concernant le schéma réactionnel. Dans la structure de l'holoenzyme de l'ALDH2, une molécule d'eau est située entre le glutamate Glu268 et la cystéine catalytique Cys302. Le groupement carboxylate du glutamate est situé à 6,8 Å de la cystéine et ne peut donc activer la cystéine à une telle distance. Par contre, un relais par l'intermédiaire de la molécule d'eau, couplé à un mouvement du NAD, est postulé. Le glutamate activerait la molécule d'eau et celle-ci déprotonerait la cystéine afin de permettre l'étape d'acylation. Cependant, il est également proposé que l'activation de la cystéine soit due à la présence des groupements NH de la chaîne principale appartenant à la cystéine et à l'acide aminé suivant non conservé chez les ALDH, et à la charge positive portée par le nicotinamide car l'orientation de la chaîne latérale de la cystéine ne varie pas sous l'effet de la soustraction du cofacteur aux cristaux. Lors de la

réaction, les intermédiaires tétraédriques chargés négativement et thioacylenzyme pourraient être stabilisés par des interactions, entre autres, avec un site oxyanion composé du groupement NH de Cys302 et du groupement NH₂ de la fonction amide de l'asparagine conservée Asn169. La déacylation de l'enzyme est assurée par une seconde molécule d'eau puisqu'il y a oxydation du substrat aldéhydique en acide carboxylique. La seconde molécule d'eau serait activée par le glutamate. En raison de la présence du cofacteur, les auteurs proposent également l'existence de mouvements du nicotinamide afin que les étapes d'acylation et de déacylation se produisent. Cette hypothèse sera renforcée par l'étude structurale de l'ALDH1.

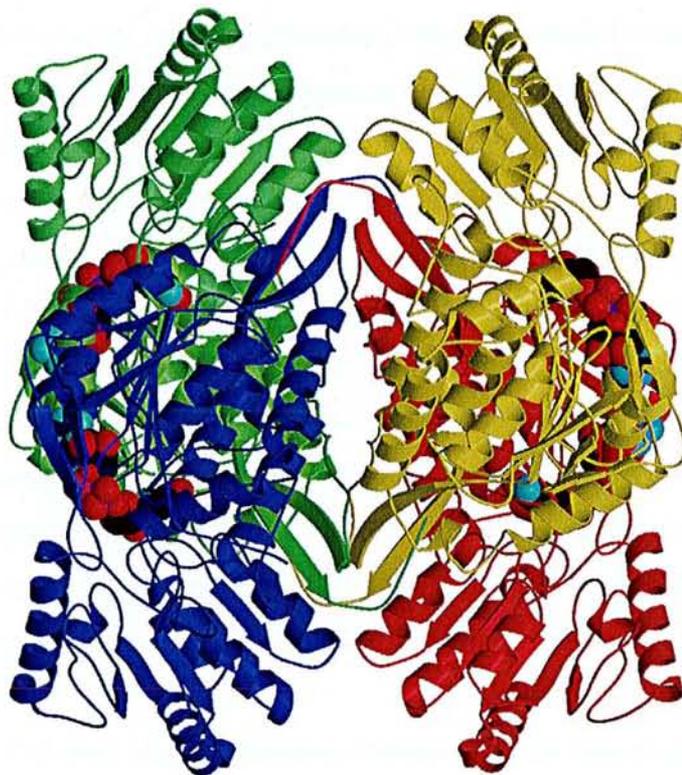


Figure 1.5: tétramère de l'ALDH2 représenté en rubans et flèches (Steinmetz et al., 1997). Le cofacteur est représenté en sphère de Van der Waals.

Les études enzymatiques récentes (Mann & Weiner, 1999) montrent que l'étape limitante de la réaction serait différente entre les ALDH de la classe 2 et les ALDH de la classe 3. Pour un grand nombre d'ALDH2, l'étape limitante est l'étape de déacylation, et plus

précisément l'hydrolyse du thioacylenzyme (Wang & Weiner, 1995). Pour l'ALDH3 humaine, le transfert d'hydrure est l'étape limitante (Mann & Weiner, 1999). Les études de mutagenèse dirigée montrent également que le glutamate catalytique serait Glu333 pour l'ALDH3 humaine. La mutation du glutamate correspondant de l'ALDH2 (Glu399), qui interagit avec les groupements hydroxyle du ribose du nicotinamide, entraîne un changement de l'étape limitante qui devient le transfert d'hydrure. Ainsi, des résidus conservés dans les alignements de séquences et dans leur position dans la structure tridimensionnelle auraient des rôles différents dans l'acte catalytique (Mann & Weiner, 1999).

C- L'ALDH de la classe 9 (Johansson *et al.*, 1998)

La structure de la bétaine aldéhyde déshydrogénase du foie de morue a été résolue peu de temps après celle de la classe 2 par remplacement moléculaire à l'aide du modèle de l'ALDH mitochondriale. Son analyse sous ces formes apo et holoenzyme a permis de soulever une différence importante au niveau de la conformation du cofacteur entre les ALDH2, 9, et l'ALDH3 (Johansson *et al.*, 1998). Comme dans le cas de l'ALDH2, aucun changement conformationnel global n'intervient lors de la fixation du cofacteur à la protéine (Steinmetz *et al.*, 1997). Selon les auteurs, l'absence de changement de conformation résulte de la présence de très nombreux contacts entre le domaine catalytique et le domaine de fixation du cofacteur. Cette protéine présente 37 % d'identité de séquences avec l'ALDH2 et 21 % avec l'ALDH3 (Johansson *et al.*, 1998). Les faibles valeurs rms calculées entre les différentes structures d'ALDH montrent que celles-ci sont très similaires dans le repliement, dans l'association des dimères, et dans l'organisation dimérique (Johansson *et al.*, 1998).

La conformation *anti* du nicotinamide mononucléoside dans la structure de l'ALDH9 a été discutée dans cet article en comparaison avec l'orientation *syn* adoptée par le nicotinamide dans la structure de l'ALDH3. Dans la structure de l'ALDH2, l'orientation *anti* du nicotinamide est également observée. Si des conformations différentes du cofacteur existaient chez les ALDH (Figure 1.6), alors le transfert d'hydrure serait différent en fonction de la classe à laquelle chaque enzyme appartient. Il serait Pro-*R* chez les ALDH des classes 2 et 9 et Pro-*S* pour les ALDH de la classe 3. Or, le transfert d'hydrure est de type pro-*R* pour

plusieurs ALDH (Jones *et al.*, 1987). Donc, un changement de conformation du cofacteur doit se produire avant le transfert d'hydrure dans le cas des ALDH de la classe 3.

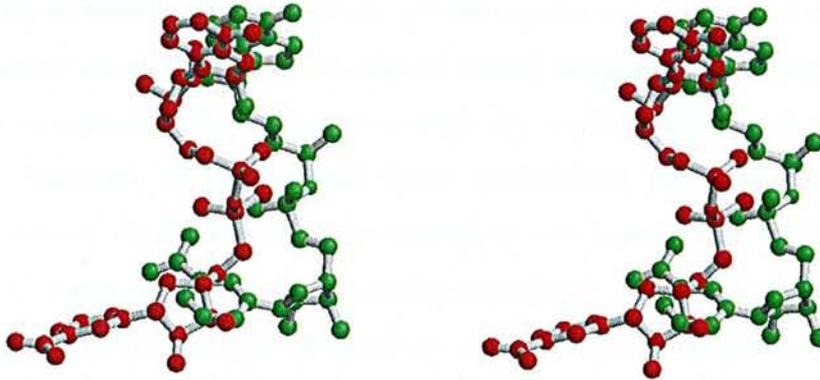


Figure 1.6 : vue stéréoscopique des cofacteurs de l'ALDH9 (rouge) et de l'ALDH3 (vert) après superposition des domaines de Rossmann.

D- L'ALDH de la classe 1 (Moore *et al.*, 1998)

La structure de l'ALDH cytosolique du foie de mouton a été résolue par remplacement moléculaire à l'aide de la structure de l'ALDH2. Son analyse apporte des informations sur un probable réarrangement conformationnel de la structure au cours de la réaction catalytique (Moore *et al.*, 1998). Ce point avait déjà été évoqué lors de l'étude structurale de l'ALDH2. Il était alors proposé une possible différence de conformation entre le NADH et le NAD conformément aux résultats de mutagenèse puisque certaines mutations affectent seulement la fixation du NAD et non celle du NADH (Steinmetz *et al.*, 1997). L'analyse de la structure de l'ALDH1 fournit également des éléments de réponse supplémentaires sur la spécificité de substrat. Pour ce type d'ALDH, l'étape limitante de la réaction est associée au relargage du cofacteur réduit donc à l'étape de déacylation (Wang & Weiner, 1995).

Dans cette structure, la partie nicotinamide mononucléotide du NAD présente deux conformations différentes déduites de l'observation de la carte de densité électronique. La partie nicotinamide mononucléotide est donc affectée de taux d'occupation différents en fonction de sa conformation. Le taux d'occupation le plus faible correspond à la position que

le cofacteur occupe dans les structures de l'ALDH2 et de l'ALDH9, c'est-à-dire une position où le carbone C4 acceptant l'hydrure se trouve à environ 3 Å de la cystéine. Le taux d'occupation le plus élevé correspond à une position du cofacteur où le C4 du nicotinamide se situe à 7,3 Å de la cystéine. Les valeurs élevées des facteurs d'agitation thermique des atomes de la chaîne latérale du glutamate jouant le rôle de base générale au cours de la réaction, associées à une densité électronique très faible sur la chaîne, suggèrent sa possible flexibilité. L'association de ces deux phénomènes serait nécessaire à la fonction enzymatique. La flexibilité du NAD serait due à son mode particulier de fixation au domaine de Rossmann. L'absence d'interaction entre le pyrophosphate, le dipôle d'une hélice α et des charges positives, comme observées dans le mode classique de fixation du NAD à ce domaine, et la présence d'un potentiel négatif à proximité du pyrophosphate faciliteraient cette flexibilité.

L'analyse de la structure de l'ALDH1 révèle la spécificité de substrat pour le rétinol en comparaison avec celle pour l'acétaldéhyde, substrat de l'ALDH2. La spécificité de substrat serait due, selon les auteurs, à la taille et la forme des sites catalytiques plutôt qu'à leur caractère plus ou moins hydrophobe. Ainsi, l'entrée du site catalytique de l'ALDH1 est plus large que celui de l'ALDH2 ce qui lui permettrait de mieux fixer le rétinol (Moore *et al.*, 1998).

L'analyse de la structure de l'ALDH1 a aussi conduit les auteurs à imaginer une possible interaction entre l'ALDH1 et une protéine (CRBP II) fixant le rétinol ou des protéines fixant l'acide rétinoïque en raison des nombreuses boucles, structures secondaires flexibles, entourant le site catalytique de l'ALDH1. Ces différentes interactions permettraient un passage du rétinol de CRBP II à l'ALDH1 et ensuite de l'ALDH1 aux protéines fixant l'acide rétinoïque.

III- Le mécanisme catalytique proposé pour les ALDH hors Sm-ALDH

La résolution des structures d'ALDH et les différentes études d'enzymologie ont permis de proposer un mécanisme catalytique général des ALDH pour lesquelles la déacylation est l'étape limitante de la réaction. Il était déjà acquis que les principaux acteurs de la réaction étaient un glutamate, une cystéine, le cofacteur et le substrat (Wang & Weiner,

1995; Farrès *et al.*, 1995). Le glutamate (Glu268, ALDH1/2) jouerait le rôle de base générale permettant la déacylation de l'enzyme. L'étape d'acylation et donc d'activation de la cystéine pourrait être due à différents mécanismes: implication d'une molécule d'eau et de Glu268, ou effet des dipôles de la chaîne principale et de la charge positive du nicotinamide sur le pKa de la cystéine, les études de RMN montrant que le transfert d'hydrure est de type pro-*R* chez les ALDH (Jones *et al.*, 1987). Ce schéma a été complété par les observations de la structure de l'ALDH1. Lors de la réaction, le mouvement du cofacteur serait couplé à celui du glutamate. Ainsi, au cours de la seconde étape, le cofacteur serait partiellement relargué de l'enzyme, permettant à une molécule d'eau qui serait activée par le glutamate de prendre place pour l'étape de déacylation (Moore *et al.*, 1998). Au cours de la réaction, la charge négative portée par l'intermédiaire tétraédrique pourrait être stabilisée par un site oxyanion composé du groupement amide d'une asparagine hautement conservée chez les ALDH et du groupement NH de la cystéine catalytique.

En plus de leur implication dans l'oxydation des aldéhydes, les ALDH, comme les protéases, et la glycéraldéhyde-3-phosphate déshydrogénase, possèdent une seconde activité: l'activité estérase.

IV- Cas de l'ALDH de *S. mutans* (EC 1.2.1.9)

L'ALDH présentée dans ce manuscrit est une enzyme tétramérique d'environ 200 kDa qui catalyse l'oxydation irréversible du glycéraldéhyde-3-phosphate (G3P) en 3-phosphoglycérate (3-PGA). Au cours de cette réaction, elle réduit son cofacteur, le NADP en NADPH¹. Elle permet la production de NADPH nécessaire pour une croissance correcte de la bactérie. Cependant, une mutation dans le gène *gapN* codant cette protéine n'empêche pas la croissance de la bactérie mais la ralentit (Boyd *et al.*, 1995). Ainsi les streptocoques déficients dans les enzymes de connexion de la voie des hexoses monophosphate à celle des pentoses phosphate (glucose-6-phosphate déshydrogénase et 6-phosphogluconate déshydrogénase) (Figure 1.7) et dans le gène codant cette ALDH possèdent d'autres enzymes qui permettent de pallier ces déficiences (Crow & Wittenberger, 1979).

¹ Les formules développées du substrat et du cofacteur de l'ALDH de *S. mutans* sont présentées dans la partie annexe.

Comme cette ALDH oxyde de façon irréversible son substrat le G3P en réduisant le cofacteur NADP, elle est appelée glycéraldéhyde-3-phosphate déshydrogénase NADP dépendante non phosphorylante (EC 1.2.1.9). Sur la base des alignements de séquences, elle est classée dans la famille des ALDH (Habenicht *et al.*, 1994). Elle ne possède aucune similarité de séquences avec les GAPDH NAD(P) dépendantes phosphorylantes qui catalysent l'oxydation réversible du même substrat mais en 1,3-diphosphoglycérate (1,3 dPG) (EC 1.2.1.12). Ce type de réaction requiert la présence d'un phosphate inorganique.

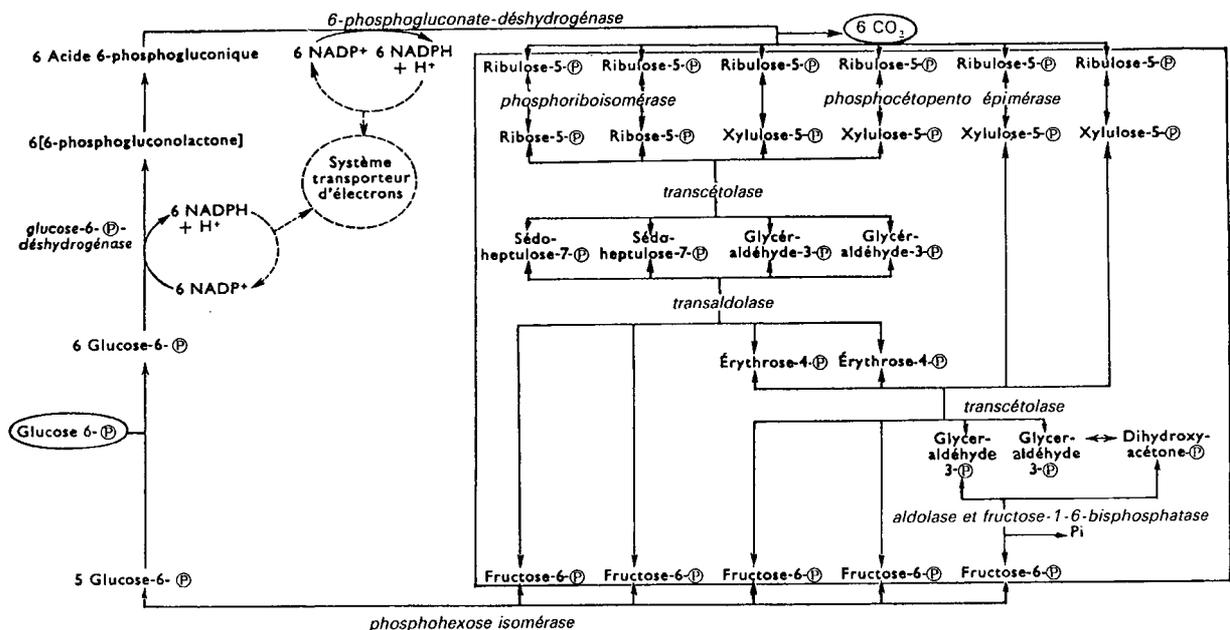


Figure 1.7: cycle des pentoses phosphate (Weil, 4^e édition, 1983).

Ainsi, les GAPDH et les ALDH constituent deux familles de protéines distinctes sur la base des alignements de séquences, de la structure tridimensionnelle et du mécanisme réactionnel. Chez les GAPDH, il fait également intervenir une cystéine, comme la plupart des protéases à cystéine, et une base, une histidine, comme la plupart des protéases à sérine et à cystéine. Le transfert d'hydrure est pro-*S* contrairement aux ALDH. Enfin, l'alignement réalisé entre la séquence de l'ALDH de *S. mutans* et celle de la GAPDH de *B. stearothermophilus* montre une vingtaine d'acides aminés similaires entre les parties C-terminale et N-terminale respectivement. Cependant, le substrat des deux enzymes est un aldéhyde et le produit un acide, reflétant ainsi pour ces deux types d'enzymes une caractéristique commune, l'oxydation des aldéhydes, l'une étant de nature réversible et phosphorylante, l'autre étant irréversible et non phosphorylante.

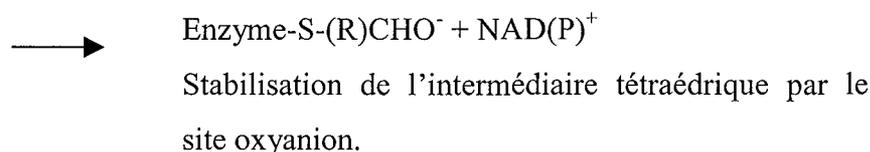
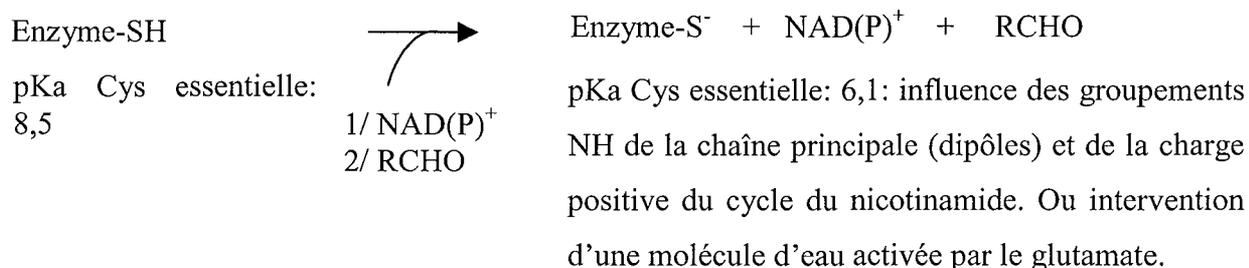
Une cystéine, Cys284 (Cys302, ALDH1/2), et un glutamate, Glu250 (Glu268, ALDH1/2), participent au mécanisme réactionnel de l'ALDH de *S. mutans*. Cette enzyme présentent environ 30 % d'identité de séquences avec les autres ALDH et 50 % d'identité de séquences avec les GAPDH NADP dépendantes non phosphorylantes du pois et du maïs (Boyd *et al.*, 1995). Si les ALDH possèdent des repliements similaires, il est fort probable que la structure de cette ALDH soit plus proche de celle des ALDH NAD(P) dépendantes oxydant de façon irréversible le G3P chez les organismes photosynthétiques et chez certaines bactéries. Un alignement de 74 séquences d'ALDH réalisé avec le programme Clustal W (Thompson *et al.*, 1994) est présenté en annexe.

La résolution de plusieurs structures de la GAPDH NADP dépendante non phosphorylante de *Streptococcus mutans*, décrite dans ce manuscrit, permet de retrouver des caractéristiques mentionnées dans les autres structures, mais aussi d'observer une distance plus faible entre Glu250 et Cys284 ainsi que des changements locaux de conformation de la protéine dus à la fixation du NADP (mouvement de la chaîne principale et rotation de la chaîne latérale de Cys284). Ces deux points, en relation avec des résultats de l'enzymologie, ont permis de compléter le schéma de la réaction catalysée par les ALDH (proposé paragraphe III). Les analyses structurales mettent également en évidence toute une série de résidus conservés chez les GAPDH NADP dépendantes non phosphorylantes pouvant être impliqués dans la fixation du substrat. La comparaison des différentes structures d'ALDH révèle le rôle de certains résidus, conservés dans les alignements de séquences, dans la détermination de la spécificité de cofacteur. Elle révèle également le rôle d'autres acides aminés dans la détermination de l'état oligomérique. L'analyse du site catalytique de cette ALDH permet de modéliser la fixation du substrat et de montrer l'importance de certains résidus dans sa fixation.

Cette dernière observation est confirmée par l'étude structurale du complexe ternaire G3P-NADP-enzyme (mutant C284S). L'analyse des différentes structures résolues, en association avec certains résultats de l'enzymologie, permet de proposer un modèle de la réaction catalysée par les ALDH qui ne nécessite pas l'intervention d'une molécule d'eau lors de l'étape d'acylation et s'accompagne de réarrangements locaux de la structure et du cofacteur.

Le schéma suivant résume les différentes étapes du mécanisme réactionnel décrites dans les paragraphes précédents. Les valeurs de pKa indiquées sont celles mesurées pour l'ALDH de *S. mutans* (Marchal & Branlant, 1999; Marchal *et al.*, 2000). L'étape limitante de la réaction pour Sm-ALDH est l'hydrolyse du thioacylenzyme comme dans le cas des ALDH de la classe 2.

- Activation de la cystéine lors de l'étape d'acylation



- Activation de la molécule d'eau par un résidu glutamate, étape de déacylation



Thioacylenzyme stabilisé entre autres par des interactions avec le site oxyanion.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
 Rue du Jardin Botanique
 54600 VILLERS LES NANCY

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Aspects méthodologiques

I- Introduction

La détermination de la structure tridimensionnelle d'une macromolécule par diffraction des rayons X se déroule en plusieurs étapes allant de l'obtention des cristaux à l'affinement de la structure de la molécule en passant par l'analyse de la diffraction du rayonnement X.

Toute détermination d'une structure de protéine nécessite une certaine quantité de matériel biologique avec un haut degré de pureté. Le point de départ de la résolution d'une structure consiste à réaliser des études de cristallogenèse en vue de l'obtention de cristaux de qualité à partir d'une solution de macromolécule ayant un haut degré de pureté. La cristallisation utilise différents composés très purs permettant l'insolubilisation de la molécule étudiée.

Après analyses préliminaires des caractéristiques cristallines, des jeux de données de diffraction X sont enregistrés sur un ou plusieurs cristaux. L'analyse des intensités diffractées I_{hkl} permet d'obtenir la structure de l'objet diffractant, et donc la densité électronique de celui-ci qui s'exprime par la relation:

$$\rho(x, y, z) = (1/V) \sum_h \sum_k \sum_l F_{hkl} e^{-2i\pi(hx + ky + lz)}$$

Dans cette relation, V représente le volume de la maille, hkl sont les indices des nœuds du réseau réciproque et donc des réflexions, et le vecteur F est le facteur de structure. L'accès à cette quantité n'est possible que si la valeur de la phase de l'onde qu'il représente est connue, l'expérience ne permettant d'obtenir que le module du vecteur par la relation:

$$\|F_{hkl}\| = I_{hkl}^{1/2} \text{ et } F_{hkl} = \|F_{hkl}\| e^{i\phi_{hkl}}$$

La solution du problème de la phase ne sera obtenue qu'à l'instant où un ou plusieurs atomes seront positionnés dans la maille. Ces atomes peuvent être des atomes lourds facilement localisables en raison de leur facteur de diffusion élevé ou les atomes d'une molécule dont la structure est proche de celle recherchée. Ces méthodes de phasage sont le phasage par le remplacement isomorphe ou le remplacement moléculaire.

Une fois le modèle moléculaire construit dans la carte de densité électronique calculée à partir de la position des atomes lourds ou de la position du modèle de remplacement moléculaire, la dernière étape de la résolution de la structure débute. Il s'agit de l'affinement, qui consiste à corriger les erreurs introduites dans la structure au cours de sa construction. Cette procédure permet d'affiner les paramètres structuraux par rapport à des paramètres standards afin d'obtenir une structure avec une faible énergie libre.

Ensuite, la structure validée sur le plan géométrique à l'aide de différents programmes est étudiée sur le plan structural, biochimique et biologique (évolution par exemple) et ses coordonnées sont déposées dans la Protein Data Bank (Bernstein *et al.*, 1977).

II- Cristallisation de macromolécules

La détermination de la structure tridimensionnelle d'une protéine par radiocristallographie nécessite l'obtention de cristaux permettant l'enregistrement des données de diffraction. La cristallisation des macromolécules biologiques est l'une des étapes

limitantes à l'étude de leur structure et les processus régissant la cristallisation de telles molécules sont très étudiés.

Un cristal est constitué par la répétition périodique du même motif à travers l'espace. L'association de ces motifs est assurée par l'existence de forces intermoléculaires diverses (liaisons hydrogène, interactions électrostatiques, de Van der Waals et hydrophobes). Le pourcentage de solvant à l'intérieur du cristal peut atteindre des valeurs proches de 80 %, ce contenu ayant une moyenne de l'ordre de 50 %.

La cristallisation d'une protéine ou d'autres macromolécules (ADN, ARN) est obtenue en augmentant progressivement la concentration d'une solution protéique diluée en présence d'une solution précipitante amenant la protéine dans un état de sursaturation. Ceci conduit à l'apparition de deux phases en équilibre: une phase liquide saturée en protéine et une phase solide représentée par les cristaux ou/et un précipité amorphe. L'état de sursaturation est créé par la présence de sels qui entraîne une diminution de l'activité de l'eau autour de la protéine. Les molécules d'eau hydratent les électrolytes ce qui entraîne une insolubilisation de la protéine et donc l'apparition d'agrégats. Dans l'état de sursaturation, la concentration de la protéine en solution est supérieure à sa solubilité. La solubilité représente la concentration résiduelle de macromolécules en solution en équilibre avec la phase cristalline. L'état de sursaturation permet d'initier l'étape de la nucléation nécessaire à l'apparition de germes cristallins, qui grandiront ensuite jusqu'à ce que l'équilibre soit atteint.

Ce phénomène dépend d'un grand nombre de paramètres tels que le pH, la température, la nature des sels, du tampon et des agents précipitants utilisés, la présence d'additifs et dépend bien sûr de l'état de la protéine et de sa nature. L'intégrité structurale et fonctionnelle de la protéine associée à une grande pureté sont des éléments déterminants pour la cristallisation de telles molécules dans la plupart des cas. De même, la concentration des solutions protéiques et salines a un rôle sur la qualité des cristaux, qualité qui peut également être affectée par leur vitesse de croissance.

Afin d'obtenir les premières conditions de cristallisation, une étude de cristallogénèse est réalisée. Compte tenu de la faible quantité de matériel disponible et du nombre important de conditions préalables pouvant être testées, différentes microtechniques de cristallisation ont été développées, dont la méthode de la diffusion de vapeur en goutte suspendue (McPherson, 1990) (Figure 2.1).

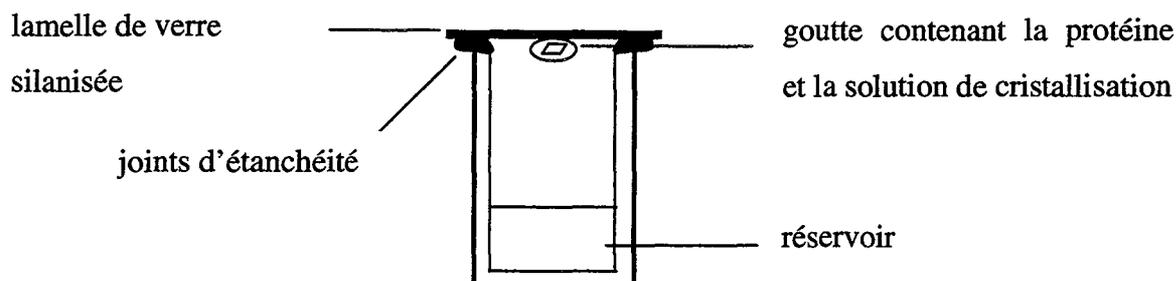


Figure 2.1: cristallisation par la méthode de la goutte suspendue.

Quelques microlitres de solution protéique sont mélangés à quelques microlitres de solution tamponnée, en présence d'agents précipitants et/ou d'additifs sur une lamelle en verre préalablement silanisée. La lamelle est alors retournée en vis-à-vis d'un réservoir dont la concentration en agents précipitants et/ou en additifs est supérieure à celle contenue dans la goutte. Par évaporation du solvant de la goutte vers le puits, la concentration de la protéine dans la goutte augmentera progressivement au cours du temps. Si les conditions de cristallisation sont correctement ajustées, quelques cristaux de qualité suffisante pour une étude cristallographique pourront être obtenus. Cette technique a été utilisée pour la cristallisation de l'ALDH de *S. mutans*.

La première caractérisation des cristaux de protéine peut être réalisée grâce à la diffusion d'un colorant à l'intérieur du cristal en utilisant les canaux de solvant qui le traversent. Contrairement aux cristaux de molécules inorganiques, les cristaux de protéine contiennent une grande proportion de solvant. Ce fort pourcentage de solvant peut être utilisé pour faire diffuser différents composés tels que des atomes lourds, les cofacteurs, les substrats ou encore des colorants. La coloration des cristaux permet dans un premier temps de vérifier que ceux-ci sont probablement des cristaux de protéine et non de sels.

Cependant, les meilleures caractérisations des cristaux de protéine demeurent les caractérisations biochimiques et physiques.

III- Enregistrement des données de diffraction

Après avoir obtenu des cristaux de protéine, il est nécessaire de les caractériser par différentes méthodes, biochimiques, par exemple dissolution du cristal et dépôt sur gel, ou physiques telle que la diffraction des rayons X. La diffraction des rayons X par des objets macroscopiques permet de trouver la structure atomique présente à l'intérieur du cristal grâce à l'intensité diffractée. Le jeu de données est composé par l'ensemble des intensités diffractées appelées I_{hkl} .

A-Les sources de rayons X

Différentes sources de rayons X peuvent être utilisées pour des études de diffraction par des cristaux de protéine. Les principales sources utilisées sont les anodes tournantes et le rayonnement synchrotron. L'utilisation de tubes scellés devient marginale en raison de la faible intensité du rayonnement émis par rapport aux sources précédentes.

La production de rayons X par les anodes tournantes est obtenue par application d'une différence de potentiel entre une anode et une cathode constituée par un filament. L'anode la plus fréquemment utilisée en biocristallographie est une anode en cuivre, le filament étant en tungstène. La différence de potentiel existant entre l'anode et le filament arrache des électrons au filament qui vont, sous cette accélération, entrer en collision avec l'anode. Le freinage brutal des électrons permet la conversion de leur énergie cinétique en rayonnement continu. Le spectre est constitué d'un fond continu auquel s'ajoutent principalement les raies K_{α} et K_{β} caractéristiques de l'élément constituant l'anode. L'utilisation de monochromateurs en graphite ou de miroirs associés à un filtre de nickel permet de sélectionner la longueur d'onde désirée.

L'autre source de rayonnement utilisée est le rayonnement synchrotron. Des électrons ou des positrons, soumis à une accélération dans un anneau appelé « booster » sont injectés dans un anneau de stockage. Sous l'action de champs magnétiques produits par des aimants de courbure courbant leur trajectoire, ces particules subissent une décélération qui se traduit par l'émission du rayonnement synchrotron capté en différentes parties de l'anneau appelées lignes de lumière. L'intensité du rayonnement émis est beaucoup plus élevée que celle fournie

par des anodes tournantes, ce qui permet d'obtenir des taches de diffraction d'intensités plus élevées et donc des jeux de données à plus haute résolution.

B- La diffraction des rayons X

Un cristal est caractérisé par la répétition périodique du même motif à travers l'espace. Ainsi lorsqu'il y a interaction entre le rayonnement incident et le cristal avec une orientation des plans réticulaires du cristal satisfaisant la loi de Bragg, le rayonnement est diffracté dans une certaine direction de l'espace. La loi de Bragg est donnée par la relation :

$$2 \times d_{hkl} \times \sin\theta = n \times \lambda \quad d_{hkl} : \text{distance interréticulaire}$$

La diffraction se caractérise par l'apparition de taches de diffraction ou réflexions sur le détecteur. Elles résultent de la traversée de la sphère d'Ewald par les nœuds hkl du réseau réciproque, chaque nœud correspondant à une famille de plans hkl du réseau direct en position de diffraction. Chaque réflexion est caractérisée par ses indices hkl et par son intensité I_{hkl} . L'intensité diffractée par un cristal est décrite par la relation suivante (relation de Darwin):

$$I_{hkl} = Vc/(Vm)^2 \times 1/\omega \times I_0 \times A_{hkl} \times P_{hkl} \times L_{hkl} \times ||F_{hkl}||^2 \times \lambda^3$$

L'intensité diffractée I_{hkl} dépend donc de:

- Vc le volume du cristal irradié
- Vm le volume de la maille
- ω la vitesse angulaire, c'est-à-dire la vitesse de rotation du cristal
- I_0 l'intensité incidente
- A_{hkl} le facteur d'absorption
- P_{hkl} le facteur de polarisation qui dépend de l'angle θ
- L_{hkl} le facteur de Lorentz, caractéristique du temps de traversée de la sphère d'Ewald par un nœud du réseau réciproque
- λ la longueur d'onde
- $||F_{hkl}||$ module du facteur de structure

La mesure précise de I_{hkl} permettra de déterminer $\|F_{hkl}\|$, le module du facteur de structure. L'enregistrement de la totalité de l'espace réciproque indépendant est nécessaire pour la détermination complète de l'unité asymétrique. La connaissance de cette unité en association avec celle du groupe d'espace permet de définir la maille. Le nombre de réflexions à enregistrer dans le cas des cristaux de macromolécules est très élevé en raison de la dimension des paramètres de maille qui peuvent atteindre plus de 500 Å, par exemple pour les cristaux de virus ou de ribosomes. Le nombre théorique de réflexions enregistrées à la résolution d est donné par la relation :

$$N_t = 4/3 \times \pi \times 1/d^3 \times V_m \times 1/n$$

- N_t : nombre théorique de taches de diffraction enregistrées à la résolution d
- d : résolution désirée en Å
- n : valeur dépendant du mode de réseau (1: réseau primitif)

Le nombre théorique de réflexions indépendantes est alors donné par la relation:

$$N_i = N_t/m$$

où m représente la multiplicité de la classe de Laue.

Le nombre de taches de diffraction indépendantes pour les cristaux d'apoenzyme de l'ALDH de *S. mutans* s'élève à plus de 215000 à une résolution de 1,8 Å. Le nombre de réflexions enregistrées avant la moyennation des données s'élève à 676000. Ainsi, il est possible d'introduire la notion de redondance qui représente le nombre de fois qu'une réflexion ou une de ses équivalentes par la symétrie de la classe de Laue a été mesurée.

La loi de Friedel montre qu'à toute réflexion hkl correspond une réflexion équivalente $-h-k-l$. L'égalité par la loi de Friedel n'existe plus dès lors qu'un atome ayant une forte contribution anormale à la diffraction est présent dans l'unité asymétrique. Cette propriété est très utile pour la détermination des structures de protéine par les méthodes MAD (multiwavelength anomalous diffusion) ou SIRAS (single isomorphous replacement with anomalous scattering). Notons également que l'introduction de tels atomes dans le solvant a permis le développement d'une nouvelle méthode de résolution des structures appelée méthode MASC (multiwavelength anomalous with solvent contrast).

Ainsi, si le nombre de réflexions à enregistrer est élevé, il faut que le matériel soit adapté et permette à la fois d'enregistrer tout l'espace réciproque indépendamment tout en optimisant le temps de collecte.

C- Les détecteurs de rayons X

Si Perutz avait à déterminer aujourd'hui la structure tridimensionnelle de l'hémoglobine, l'acquisition des données représenterait une étape banale de la résolution de la structure. En effet, l'utilisation des films photographiques a été remplacée par l'utilisation de détecteurs bidimensionnels type image plate ou CCD (charge-coupled device).

Les détecteurs ponctuels, c'est-à-dire mesurant les réflexions les unes après les autres, ne sont quasiment plus utilisés en biocristallographie. En effet, l'utilisation de ce type de détecteur rend la collecte de données excessivement longue.

L'apparition des plaques images et des détecteurs à fils a permis de diminuer le temps de collecte des données. En effet, elles permettent d'enregistrer simultanément de nombreuses réflexions dans deux directions de l'espace (plus de 2000 par plaque dans le cas de l'ALDH de *S. mutans* jusqu'à 1,8 Å de résolution). Leur association avec les anodes tournantes permet d'enregistrer l'espace indépendamment en quelques jours voire une journée en fonction du pouvoir diffractant du cristal et de sa classe de Laue. Si l'intensité du rayonnement incident est augmentée, comme dans le cas du rayonnement synchrotron, alors une collecte de données peut être ramenée à quelques heures.

Un nouveau type de détecteur est apparu récemment: le CCD (charge-coupled device). Ce type de détecteur permet en quelques secondes, contrairement aux plaques images, d'avoir accès au cliché de diffraction. De plus sa sensibilité est supérieure à celle des plaques images. Cependant, le faible diamètre de ce type de détecteur, dû à son coût de fabrication et à divers problèmes technologiques, oblige à diviser les collectes en plusieurs parties notamment lorsque les cristaux utilisés diffractent à haute résolution.

D- Stratégie d'enregistrement

Avant de débiter tout enregistrement, il est nécessaire de déterminer les paramètres de maille du cristal ainsi que son réseau de Bravais. Un certain nombre de réflexions doivent être mesurées pour y parvenir. Une oscillation du cristal autour d'un axe perpendiculaire aux rayons X incidents permet d'obtenir ce nombre suffisant.

La rotation du cristal dans le flux de rayons X amène plusieurs plans du réseau direct en position de diffraction et donc autant de nœuds du réseau réciproque à traverser la sphère d'Ewald. Ainsi, un grand nombre de taches de diffraction peut être collecté sur une plaque image. Le nombre de réflexions enregistrées dépendra des paramètres de maille, de l'orientation du détecteur, de la longueur d'onde et du pouvoir diffusant du cristal. Pour des cristaux possédant de petits paramètres de maille comme les cristaux de molécules inorganiques, une rotation de plusieurs degrés peut être nécessaire pour cette étape. En ce qui concerne les cristaux de protéine, une rotation de 1 degré permet d'obtenir suffisamment de réflexions. Si cette rotation est trop importante, alors les taches de diffraction seront en recouvrement sur la plaque image. Or, il est important pour la qualité des mesures de bien séparer les taches de diffraction sur le détecteur dont le recouvrement dépend

Le recouvrement des taches de diffraction existe dans un autre cas. La distance dans l'espace réciproque est l'inverse de celle existant dans l'espace direct. Ce recouvrement des taches de diffraction dépend des paramètres de maille, de la longueur d'onde utilisée, de la distance cristal détecteur et de la mosaïcité.

Une fois les paramètres de maille déterminés, une stratégie d'enregistrement doit être définie. La distance cristal détecteur doit rendre possible l'enregistrement d'un espace complet de diffraction à une certaine résolution tout en ne conduisant à aucun recouvrement.

Il est également nécessaire de tenir compte d'un autre facteur qui est la résistance du cristal à l'irradiation. En effet, l'absorption du rayonnement par les cristaux de protéine conduit à des dégradations par réactions radicalaires et par effet thermique (Burmeister, 2000; Ravelli & McSweeney, 2000). Ce type de dégradations nécessite de changer de cristal pour la collecte des données dans de nombreux cas et le recours à la cryogénie devient alors une nécessité (Rodgers, 1994). De cette façon, il sera possible d'exposer plus longtemps le cristal

aux rayons X. De ce fait, une meilleure résolution pourra être atteinte. Cependant, si la faible intensité des taches de diffraction à haute résolution est augmentée en prolongeant l'exposition du cristal, le bruit accumulé à la surface du détecteur le sera aussi. Ainsi, le temps d'exposition doit être optimisé pour obtenir le meilleur rapport entre le signal et le bruit. Malgré l'utilisation de la cryogénie, des dégradations peuvent encore se produire (Burmeister, 2000; Ravelli & McSweeney, 2000)

La limite de résolution de l'enregistrement ne dépend pas seulement de l'intensité incidente, du temps d'exposition du cristal, de sa taille ou de la longueur d'onde utilisée. Cette limite dépendra également de l'angle qu'il est possible de choisir entre le cristal et le détecteur, compte tenu du dispositif expérimental.

IV- Le traitement des données de diffraction

Comme la formule de Darwin le décrit (page 40), l'intensité enregistrée dépend de plusieurs facteurs. Traiter des données revient donc à estimer la contribution de chacun de ces facteurs à l'intensité diffractée. Le bruit collecté à la surface des détecteurs sera retranché à l'intensité de chaque réflexion *hkl*.

Les deux logiciels utilisés pour le traitement des données enregistrées sur les cristaux de l'ALDH de *S. mutans* sont DENZO (Otwinowski & Minor, 1997), et XDS (Kabsch, 1988). Comme le logiciel DENZO a majoritairement été utilisé lors du traitement des données des différents enregistrements de cristaux de l'ALDH de *S. mutans*, les points suivants traiteront de son utilisation.

A- Recherche des paramètres de maille et indexation du cliché de diffraction

1- Recherche des paramètres de maille

A partir des réflexions de plus forte intensité sur le cliché, le calcul des paramètres de maille débute. Le logiciel recherche le plus petit trièdre irréductible permettant de passer d'une réflexion à l'autre, donc les plus petits vecteurs linéairement indépendants. Lors de cette étape, le logiciel ne prend pas en compte la rotation du cristal. Il considère que tous les nœuds du réseau réciproque ont traversé la sphère d'Ewald au même instant, donc à une valeur de rotation du cristal correspondant à la moitié de l'angle d'oscillation. Afin que cette approximation reste valable, l'amplitude de l'oscillation doit demeurer faible mais être suffisante afin qu'un nombre suffisant de réflexions soit mesuré.

La liste des 14 réseaux de Bravais est alors proposée avec les paramètres de maille correspondant à chacun, associé à un facteur d'accord. Le plus faible facteur d'accord est toujours obtenu pour le système triclinique puisque ce dernier n'impose aucune contrainte sur les paramètres de maille. Le choix du système cristallin et donc des paramètres de maille étant réalisé, il est nécessaire de procéder à sa vérification et donc de passer à l'étape suivante qui est l'indexation du cliché de diffraction.

2- Indexation du cliché de diffraction

L'indexation du cliché de diffraction amène le logiciel à prédire la position des taches de diffraction sur le cliché. Des facteurs d'accord sont calculés entre les positions des taches prédites et celles observées. Les prédictions se rapprochent de l'observation si les différents paramètres de l'enregistrement (longueur d'onde, distance cristal-détecteur, centre de la plaque...) ainsi que les paramètres liés au cristal (maille, réseau de Bravais et mosaïcité) ont été correctement définis. Il est donc nécessaire au début de chaque expérience de préciser au mieux ces paramètres afin d'obtenir les meilleurs paramètres cristallins pour l'affinement des paramètres expérimentaux.

Le terme de mosaïcité utilisé précédemment concerne les petits blocs désorientés les uns par rapport aux autres qui constituent le cristal. Cette désorientation entraîne alors un élargissement des profils de raies puisque le nœud correspondant proviendra de plans équivalents en position de diffraction dans des directions très peu différentes. Cette désorientation entre les blocs est appelée mosaïcité. Cependant, son estimation reste délicate, en raison de la divergence du faisceau de rayons X et de la taille du cristal, mais nécessaire

pour estimer au mieux la position prédite des taches de diffraction pour calculer l'intensité de chaque réflexion.

B- Intégration des données

Pour chaque tache de diffraction, le calcul de l'intensité sera décomposé en deux étapes.

1- Méthode de la « shoe box »

Le bruit est calculé à l'intérieur d'une boîte entourant la tache de diffraction. La surface de la boîte doit être suffisamment large afin d'estimer au mieux le bruit. Généralement la surface de cette boîte correspond au double de la surface de la tache. Le bruit est soustrait à l'intensité mesurée.

2- Méthode du « profile fitting »

Des réflexions faibles ou fortes, proches sur un cliché de diffraction, possèdent en général le même profil de raie (Diamond, 1969). Il est donc possible de calculer des profils moyens dans différentes parties du cliché et de les appliquer alors aux taches présentes dans ces régions en centrant ce pic sur la tache, d'où la nécessité d'une estimation correcte de la mosaïcité afin de déterminer au mieux ce centre. Ainsi, le profil de chaque tache de diffraction est obtenu.

C- Mise à l'échelle des données

Le jeu de données est constitué de plusieurs clichés de diffraction. Etant donné que les deux plaques images utilisées au service commun de diffraction X sur monocristaux (SCDX) pour la collecte peuvent être légèrement différentes, que l'intensité incidente peut diminuer au cours de la collecte, que le cristal peut se dégrader au cours de l'enregistrement, que l'absorption est différente en fonction de l'orientation du cristal, il est nécessaire de mettre à

l'échelle les différents clichés avant de moyenner les intensités afin de ne conserver qu'un ensemble de réflexions indépendantes.

Le jeu de données est caractérisé par différents paramètres:

- la redondance, qui caractérise le nombre de fois qu'une même réflexion ou l'une de ses équivalentes a été mesurée.
- le R_{sym} : ce facteur permet de juger de la déviation des différentes intensités équivalentes par rapport à leur moyenne. Le R_{sym} est donné par la formule:

$$R_{\text{sym}} = \frac{\sum |I_j - \langle I \rangle|}{\sum \langle I \rangle}$$
 où I_j représente la $j^{\text{ème}}$ mesure de la réflexion hkl et $\langle I \rangle$ la valeur moyenne.
- la complétude ou « completeness »: cette valeur représente la portion de l'espace indépendant mesuré. La complétude est le rapport N_m/N_i où N_m représente le nombre de réflexions indépendantes mesurées et N_i le nombre théoriquement mesurable de réflexions indépendantes à la résolution atteinte lors de l'enregistrement.
- le rapport $I/\sigma(I)$: ce facteur représente le rapport entre l'intensité mesurée et l'erreur estimée sur cette mesure.

- les extinctions: pour certaines combinaisons des indices hkl , le facteur de structure F_{hkl} et donc l'intensité associée sont égales à zéro. Ces combinaisons de hkl sont spécifiques du groupe d'espace du cristal, c'est-à-dire de la présence d'éléments de symétrie hélicoïdaux et/ou de la présence d'un mode de réseau ou de miroirs avec translation (n'existant pas dans les cristaux de protéine). Compte tenu du caractère asymétrique des protéines dû à leur stricte composition en acides aminés L, seuls les deux premiers opérateurs de symétrie peuvent se rencontrer dans les cristaux de protéine. De ce fait, l'analyse des extinctions permet de déterminer le groupe d'espace du cristal. Cependant des ambiguïtés demeurent quant à la translation le long de certains axes hélicoïdaux. Elles ne pourront être levées que par la détermination de la structure, donc par le phasage.

V- Détermination de la structure

La détermination de la structure tridimensionnelle d'une protéine consiste tout d'abord à résoudre le problème des phases. Trois méthodes de résolution des structures de protéine sont majoritairement utilisées:

- le remplacement moléculaire

- le remplacement isomorphe multiple
- l'utilisation de la diffusion anormale

Cependant, des études préliminaires visant à déterminer le nombre de molécules dans l'unité asymétrique et les symétries les reliant doivent être réalisées.

A-Détermination de la symétrie non cristallographique

La symétrie cristallographique permet de générer le cristal par son application à l'unité asymétrique. Pour les macromolécules biologiques, elle peut être un axe de rotation, une translation due au mode de réseau, ou un axe hélicoïdal associant la rotation à une translation, ou toute autre combinaison de plusieurs de ces éléments.

Contrairement à la symétrie cristallographique, la symétrie non cristallographique ne peut s'appliquer qu'à l'intérieur de l'unité asymétrique. Elle permet de relier les molécules les unes aux autres dans cette unité ou bien de relier entre elles les sous unités d'une macromolécule (monomères d'un oligomère). La symétrie non cristallographique peut être une rotation ou une translation. Elle est décrite comme propre s'il s'agit d'une rotation pure ou impropre lorsque la rotation est associée à une translation.

La connaissance exacte de cette symétrie est importante pour l'amélioration des phases lorsque les méthodes de modification de densité sont utilisées et pour faciliter la reconstruction du modèle dans la densité électronique (Kleywegt & Read, 1997). La recherche et l'orientation des éléments de symétrie non cristallographique s'effectuent par le calcul des fonctions d'autocorrélation (Rossmann & Blow, 1962) et éventuellement des fonctions de Patterson natives. Leur interprétation sera facilitée si au préalable le nombre de molécules dans l'unité asymétrique a été évaluée par le calcul du taux moyen de solvant (Matthews, 1968).

1- La fonction de Patterson

La fonction de Patterson correspond au produit de convolution de la densité électronique. Cette fonction est centrosymétrique et donne naissance aux 24 groupes de Patterson. Elle est donnée par la formule:

$$P(U) = \int_V \rho(X)\rho(X+U)dX, \text{ avec } X = (x, y, z) \text{ et } U = (u, v, w)$$

Cette fonction contient des pics situés à l'extrémité des vecteurs interatomiques et dont l'intensité (la hauteur) correspond au produit des poids atomiques de chaque atome, que ces vecteurs soient intra ou intermoléculaires. Calculée à partir du cliché de diffraction donc à partir des modules des facteurs de structure, la fonction de Patterson correspond à la transformée de Fourier du module du facteur de structure au carré.

2- La fonction d'autocorrélation

La fonction d'autocorrélation permet de superposer les vecteurs interatomiques d'une molécule à ceux de la molécule avec laquelle elle est reliée par la symétrie non cristallographique en appliquant une rotation à l'un des deux jeux de vecteurs. Ces vecteurs sont représentés cristallographiquement par la fonction de Patterson. Donc, une fonction d'autocorrélation superpose les vecteurs de Patterson d'une macromolécule avec ceux tournés de la molécule qui lui est symétrique.

Le calcul de la fonction d'autocorrélation est effectué dans une sphère dont le rayon est choisi de sorte à ne prendre en compte que les vecteurs interatomiques intramoléculaires ou au sein de la même sous-unité. Au niveau des données, une coupure sur la haute résolution est effectuée car l'identité entre les molécules ou leurs sous-unités n'est qu'une approximation. En effet, des disparités peuvent exister à haute résolution. L'orientation de l'axe de rotation ainsi que la nature de la rotation sont données par 3 angles, par exemple les angles polaires φ, ψ, κ . Dans ce cas, les deux premiers angles définissent l'orientation de l'axe de rotation dans l'espace, le dernier la valeur de la rotation autour de l'axe.

3- La fonction de Patterson native

Cette fonction permet de trouver la position d'un axe de rotation non cristallographique orienté particulièrement par rapport aux axes de la maille. Les pics de la

fonction de Patterson ont leur maximum d'intensité lorsqu'ils relient des atomes en positions équivalentes. Ces pics apparaissent alors dans les sections remarquables de Harker. Ainsi, si un axe de symétrie non cristallographique parallèle à un axe de symétrie cristallographique existe, un pic intense apparaîtra dans une section perpendiculaire à l'axe de symétrie non cristallographique.

B- Le remplacement moléculaire

La recherche des phases permet de convertir le module du facteur de structure $\|F_{hkl}\|$, racine carrée de l'intensité diffractée, en facteur de structure F_{hkl} . Le remplacement moléculaire consiste à orienter puis à positionner un modèle moléculaire proche de celui étudié dans la maille de la structure inconnue. Le positionnement de ce modèle permet alors le calcul d'un premier jeu de phases à partir de la position des atomes, et celui des premières cartes de densité électronique.

L'utilisation du remplacement moléculaire suppose qu'une structure proche de celle étudiée soit disponible. La recherche de cette structure peut être entreprise par l'étude des alignements de séquences. Un pourcentage d'identité de séquences supérieur à 30 % ou un pourcentage élevé d'homologie de séquences doit attirer l'attention du cristallographe, d'autant plus que le nombre de structures disponibles dans la PDB est en constante augmentation. Ainsi cette méthode est de plus en plus utilisée. Même lorsque le pourcentage d'identité de séquences ou d'homologie est élevé, il est nécessaire de déterminer une certaine gamme de résolution dans laquelle la recherche des phases sera effectuée. Une troncature du jeu de données à très basse et vers la haute résolution doit être réalisée. En effet, d'une part le modèle utilisé ne contient pas de molécules d'eau alors que leur contribution est importante dans les réflexions à basse résolution, et d'autre part deux modèles proches en première approximation peuvent présenter à haute résolution de grandes différences qui masqueraient alors une similitude apparente à moyenne résolution. La connaissance de l'état oligomérique de la protéine ainsi que le nombre de molécules dans l'unité asymétrique sont impératifs afin d'utiliser le nombre exact de modèles à positionner dans l'unité asymétrique. La connaissance de la symétrie non cristallographique pourra aider à déterminer ce nombre.

1- La fonction de rotation croisée

Le calcul de cette fonction est sensiblement le même que celui employé pour le calcul de la fonction d'autocorrélation. La différence provient des fonctions de Patterson. La fonction d'autocorrélation consiste à superposer la fonction de Patterson de la macromolécule avec elle-même. La fonction de rotation croisée superpose la fonction de Patterson du modèle à la fonction de Patterson de la molécule dont la structure est inconnue. Un maximum de cette fonction est obtenue lorsque les deux fonctions de Patterson se superposent au mieux. La fonction de rotation croisée est donnée par la formule :

$$R_c(\theta_1, \theta_2, \theta_3) = \int_{|U| < d} P(U) \times P_1 C(U) dU$$

où $U=(u,v,w)$ et $P(U)$ représente la fonction de Patterson du cristal, $P_1 C(U)$, celle du modèle tournée par l'opérateur de rotation C défini par les angles d'Euler $(\theta_1, \theta_2, \theta_3)$.

La corrélation est calculée sur un volume, généralement une sphère, choisi de façon à sélectionner les vecteurs U de module inférieur au diamètre de la protéine. En effet, si la rotation de la molécule n'affecte pas la distribution des vecteurs intramoléculaires, il n'en est rien en ce qui concerne la distribution des vecteurs intermoléculaires. Le rayon de la sphère doit donc être choisi de façon à exclure les vecteurs interatomiques intermoléculaires. Ce rayon ne sera qu'approximatif en raison de l'absence d'information structurale concernant la protéine étudiée et donc proche de celui de la molécule servant de modèle.

2- La fonction de translation

Le modèle orienté est translaté dans la maille du cristal. Pour une position correcte du modèle, il apparaît un maximum de corrélation entre la fonction de Patterson du cristal et celle du modèle déplacé dans la maille. A ce maximum, il correspondra un vecteur intermoléculaire entre deux molécules du modèle superposable à un vecteur intermoléculaire entre deux molécules du cristal. La fonction de translation est donnée par la relation:

$$T(t) = \int_v P_{01}(U,t) \times P(U) dU$$

où $P(U)$ représente la fonction de Patterson observée pour le cristal, et $P_{01}(U,t)$ la fonction de Patterson calculée à partir de la molécule référence, située à une distance t d'un

opérateur de symétrie (cristallographique ou non), et de la molécule qui lui est équivalente par cet axe de symétrie. Le calcul est effectué sur le volume v de la maille.

Cette fonction prend en compte les vecteurs interatomiques entre deux molécules reliées par la symétrie non cristallographique mais également entre deux molécules reliées par la symétrie cristallographique. Dans ce dernier cas, étant donné que seule la translation du modèle perpendiculairement à l'axe de symétrie cristallographique affecte la répartition des vecteurs intermoléculaires, une recherche du maximum de corrélation entre ces deux fonctions est effectuée dans les sections de Harker. En rejetant les vecteurs intramoléculaires, le rapport signal sur bruit peut être amélioré.

La recherche de la solution correcte de cette seconde étape peut également être trouvée en suivant l'évolution du facteur d'accord R donné par la relation:

$$R = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

Ce facteur correspond à une comparaison entre les facteurs de structure observés et les facteurs de structure calculés à partir de la position du modèle.

Enfin, la visualisation graphique des différentes solutions issues de l'étape de la translation, permet de sélectionner la ou les solutions possibles du remplacement moléculaire en rejetant les positions conduisant à des recouvrements entre les différentes molécules dans la maille.

Ainsi, il est possible de calculer un premier jeu de phases à l'aide du modèle orienté et positionné dans la maille puis une première carte de densité électronique à partir de ces phases et des facteurs de structure calculés à partir du jeu de données. Alors la construction de la macromolécule peut débuter. Cependant, il est possible qu'elle soit rapidement stoppée en raison de la piètre qualité des phases calculées, due à un modèle trop éloigné de la structure recherchée. Alors, il est nécessaire de procéder à une amélioration des phases par modification de la densité et, le cas échéant, à la recherche d'un ou de plusieurs atomes lourds.

C- Le remplacement isomorphe

Si aucun modèle suffisamment proche de la structure étudiée n'est disponible pour résoudre le problème des phases, alors il faut recourir à des méthodes faisant intervenir des diffuseurs en général absents naturellement des protéines: les atomes lourds (Hg, Pt, Eu, Sm...). Les méthodes utilisables sont les méthodes de phasage SIR, MIR, SIRAS, MAD, MIRAS, ou MASC. L'utilisation de la méthode MAD connaît actuellement un essor notamment grâce à l'incorporation par génie génétique de séléno-méthionines dans les protéines (Yang *et al.*, 1990; Smith & Thompson, 1998).

La localisation d'un ou de plusieurs site(s) d'atomes lourds dans le cristal permet de déterminer les phases et ainsi de résoudre la structure de la protéine étudiée. L'incorporation des atomes lourds au sein de la protéine par interactions électrostatiques ou covalentes, doit s'effectuer sans distorsion des paramètres de maille. Elle est réalisée par diffusion de ces composés dans le cristal ou obtenue par cocristallisation de la protéine en leur présence (Blundell & Johnson, 1976).

Si l'enregistrement des données sur un cristal natif (sans atome lourd) permet d'accéder au module du facteur de structure $\|F_P\|$, celui effectué sur un dérivé (en présence d'atomes lourds) donne accès au module du facteur de structure $\|F_{PH}\|$. La fonction de Patterson différence, permettant d'interpréter les pics correspondant à des vecteurs interatomiques entre atomes lourds, est donnée par la relation:

$$P(x,y,z) = (1/V) \sum_h \sum_k \sum_l |\Delta F|_{hkl}^2 \exp -2\pi i(hx+ky+lz)$$

$$\text{et } |\Delta F|_{hkl}^2 = (\|F_{PH}\| - \|F_P\|)^2$$

L'étude de ces pics conduira à la position des atomes lourds et ainsi à deux valeurs possibles de la phase pour les facteurs de structure de la protéine. L'utilisation d'un second dérivé permettra de lever cette ambiguïté.

L'interprétation de la densité électronique est souvent difficile en raison de la piètre qualité des phases, parfois même après affinement de la position des atomes lourds. Il est alors nécessaire d'améliorer la qualité des phases par modification de la densité électronique.

Cette étape peut également être réalisée après un phasage par remplacement moléculaire si l'interprétation de la densité électronique est délicate.

D- Modification de la densité électronique

La densité électronique calculée à partir des paramètres expérimentaux (F_{obs} , φ_{calc} ou φ_{MIR}) doit être en accord avec les caractéristiques théoriques de la densité électronique, à savoir: sa positivité, sa constance dans les zones de solvant, son identité entre les molécules reliées par la symétrie non cristallographique. Les différences existant entre l'expérience et la théorie proviennent d'erreurs sur le calcul des phases (donc sur la position des atomes lourds ou du modèle utilisé lors du remplacement moléculaire) et sur l'estimation du facteur de structure, l'erreur la plus importante étant réalisée sur le calcul de la phase.

Par modification de la densité électronique, il est possible d'améliorer progressivement les phases. Cette amélioration conduit alors à une meilleure interprétation de la densité électronique permettant ainsi la construction ou un positionnement plus précis du modèle dans la densité électronique, à partir duquel sont calculées les phases. Ce processus est répété jusqu'à convergence de différents critères comme le coefficient de corrélation entre la densité électronique calculée à un cycle de modification et la densité au cycle suivant, ou encore le facteur R entre les facteurs de structures observés et calculés.

1- Nivellement du solvant

Cette méthode suppose que l'enveloppe moléculaire soit déterminée afin de différencier la densité électronique de la protéine de celle du solvant. Elle repose sur le fait qu'à l'intérieur de la protéine, de fortes fluctuations de la densité existent contrairement aux zones de solvant où la densité est estimée constante du fait du mouvement statistique important des molécules. Ainsi, il est possible de délimiter zone de solvant et de protéine. Une carte de densité lissée est calculée en associant à chaque point la densité électronique moyenne calculée sur les valeurs de densité en ce point, et en ceux l'environnant situés dans une sphère de rayon donné. La connaissance de la proportion approximative de solvant permet alors de définir sur cette carte un contour séparant zone de protéine et de solvant. Une fois ce masque connu, il est alors possible d'aplanir la densité électronique dans les zones de

solvant sans modifier celle de la protéine. Les facteurs de structure sont recalculés sur cette carte idéale et les phases correspondantes sont combinées aux phases expérimentales. Ce processus cyclique est répété jusqu'à sa convergence.

2- Positivité

Il est possible que la densité électronique en certains points de la carte soit négative, ce qui n'a aucun sens physique. Ce phénomène est dû à une erreur sur le calcul de la phase. Simultanément au nivellement de solvant et/ou à la moyennation de la densité, ces valeurs sont éliminées.

3- Moyennation de la densité

Si la molécule étudiée présente des sous-unités identiques ou plusieurs copies dans l'unité asymétrique, alors la densité électronique entre les différentes sous-unités ou copies est supposée identique jusqu'à une certaine résolution. Cette situation permet d'imposer une contrainte à la densité électronique: la valeur de la densité électronique pour les points reliés par la symétrie non cristallographique sera calculée comme étant leur moyenne. Ainsi, la symétrie non cristallographique doit être connue avec précision afin de moyenner des points réellement équivalents. Si une telle symétrie existe, différentes enveloppes moléculaires ou masques doivent être déterminés pour la procédure de moyennation généralement réalisée dans le même temps que le nivellement de solvant.

Après un ou plusieurs cycles de modification de densité, l'amélioration de la densité permet la reconstruction du modèle, ou son meilleur positionnement dans la carte de densité, et donc le calcul de phases plus précises. Alors peut débuter la dernière partie de la résolution d'une structure: son affinement.

VI- Affinement de la structure

Cette dernière étape permet de corriger les erreurs introduites dans la structure pendant la reconstruction. En effet, l'étape de reconstruction graphique introduit des erreurs

géométriques dans la structure tel que: la longueur des liaisons, les angles de valence (ϕ, ψ, χ). D'autres erreurs comme la connectivité entre les structures secondaires, l'orientation d'une chaîne latérale, le non respect de la séquence, les distances aberrantes entre atomes non liés... sont également corrigées automatiquement ou visualisées et corrigées manuellement lors de l'examen de la carte de densité électronique. En fait, les deux étapes d'affinement et de reconstruction sont menées de façon alternative, jusqu'à obtenir les plus faibles valeurs possibles pour les facteurs d'accord appelés R_{crist} et R_{free} (Brünger, 1992). L'affinement de la structure est la recherche d'un minimum énergétique obtenu en minimisant l'écart entre les paramètres de la structure et les valeurs moyennes d'angles, de longueurs de liaisons, de distances interatomiques... connues grâce aux structures de petites molécules. La recherche du minimum énergétique se fait par minimisation d'énergie ou bien, en faisant varier ces paramètres par dynamique moléculaire autour de la valeur provenant d'une minimisation d'énergie ou d'une étape de reconstruction. Cette combinaison affinement - reconstruction permet une amélioration des phases très perceptible par l'amélioration de la carte de densité. En effet, l'information structurale augmentant et s'affinant au cours de ces différents cycles, les phases peuvent être calculées plus précisément.

L'évolution de l'affinement est suivi par le calcul de deux facteurs d'accord: le R_{crist} et le R_{free} . L'expression du R_{crist} consiste à comparer les facteurs de structure observés F_{obs} et ceux calculés F_{calc} . Le R_{crist} est donné par la relation:

$$R_{\text{crist}} = \frac{\sum_{\text{hkl}} ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_{\text{hkl}} |F_{\text{obs}}|}$$

Le R_{free} s'exprime par la même relation mais son calcul est réalisé sur 10 % des réflexions préalablement exclues de l'affinement (Brünger, 1992).

D'un point de vue énergétique, la fonction à minimiser est:

$$E_{\text{pot}} = E_{\text{chim}} + w_A E_{\text{RX}}$$

Cette fonction correspond à l'addition de l'énergie chimique du système aux contraintes cristallographiques pondérées par le facteur w_A représentant le poids qui leur est attribué. E_{chim} est donnée par la relation:

$$\begin{aligned}
E_{\text{chim}} = & \sum_{\text{liaisons}} 1/2 \times K_b \times (r_i - r_0)^2 && \text{longueurs des liaisons} \\
& + \sum_{\text{angles}} 1/2 \times K_h \times (\theta_i - \theta_0)^2 && \text{angles de valence} \\
& + \sum_{\text{torsion}} 1/2 \times K_\zeta \times (\zeta_i - \zeta_0)^2 && \text{dièdres impropres} \\
& + \sum_{\text{dièdres}} 1/2 K_\theta \times (1 + \cos(n\phi_i - \delta)) && \text{dièdres propres} \\
& + \sum_i \sum_j \text{électrostatiques } q_i q_j / 4\pi\epsilon_0\epsilon_r r_{ij} && \text{interactions électrostatiques} \\
& + \sum_i \sum_j \text{Lennard-Jones } (A_{ij} r_{ij}^{-12} - B_{ij} r_{ij}^{-6}) && \text{atomes non liés}
\end{aligned}$$

D'autres termes peuvent être ajoutés au calcul de l'énergie chimique comme: la symétrie non cristallographique, les interactions par la symétrie cristallographique, les interactions par liaisons hydrogène.

E_{RX} est donnée par:

$$E_{\text{RX}} = \sum_{\text{hkl}} (|F_{\text{obs}}| - |F_{\text{calc}}|)^2$$

Les paragraphes suivants traiteront des procédures utilisées pour affiner différentes structures de l'ALDH de *S. mutans*. Une différence dans ces procédures d'affinement est l'absence d'une contrainte de symétrie non cristallographique lors de la résolution de la structure de l'apoenzyme appelée Apo1. Pour toutes les autres structures décrites dans ce manuscrit, une contrainte sur la symétrie non cristallographique a été appliquée lors de l'affinement. Cette contrainte restreint l'amplitude du déplacement des atomes reliés par la symétrie non cristallographique lors de l'affinement.

A- Affinement par corps rigide

Cet affinement consiste à considérer la molécule comme une entité rigide. Il est appliqué après le remplacement moléculaire pour ajuster la position du tétramère d'ALDH dans l'unité asymétrique par rotations et translations. Une fois positionné, le tétramère est divisé en quatre monomères et la procédure est répétée tout en augmentant progressivement la résolution. Cette procédure a été appliquée jusqu'à une résolution de 4 Å pour l'apoenzyme résolue à 1,8 Å (Apo1), et 3 Å pour les autres structures avec un affinement par corps rigide uniquement effectué sur le tétramère et les monomères.

Pour l'apoenzyme (Apo1), une dernière division a été réalisée en accord avec l'alignement de séquences par rapport au modèle utilisé. Les monomères ont été divisés en autant de domaines qu'il existait de lacunes dans l'alignement de séquences.

B- Minimisation d'énergie

Durant cette étape, les atomes sont déplacés de façon à diminuer l'énergie potentielle E_{pot} . Le déplacement des atomes s'effectue dans le sens opposé du gradient d'énergie dont le calcul est donné par la relation:

$$\text{grad}(E_{\text{pot}}) = - \delta E_{\text{pot}}(r_1, r_2, \dots, r_i, \dots, r_n) / \delta r_i$$

où $r_1, r_2, \dots, r_i, \dots, r_n$ représentent les positions des atomes.

L'inconvénient de ce mode d'affinement est l'existence de minima locaux auxquels il est possible d'aboutir alors que le minimum énergétique recherché est le minimum profond. La minimisation d'énergie ne permettant pas de sortir des minima locaux, des étapes de reconstruction graphique associées à des cycles de minimisation sont nécessaires pour en sortir. Ce long processus itératif de reconstruction - affinement peut être accéléré par l'utilisation de la dynamique moléculaire.

C- La dynamique moléculaire

Une autre méthode utilisée pour sortir des minima locaux consiste à fournir au système l'énergie nécessaire au franchissement des barrières énergétiques (Figure 2.2). Une vitesse aléatoire est attribuée à chaque atome. L'énergie cinétique fournie permet alors au système de franchir des barrières d'énergie de l'ordre de $k_B T$ (k_B : constante de Boltzmann).

Pendant l'évolution du système, chaque atome est soumis à une force $f_i(t)$ correspondant à l'opposé du gradient de l'énergie potentielle. Il est donc soumis à une accélération donnée par la loi de Newton:

$$m_i \delta^2 r_i(t) / \delta t^2 = -\delta E_{\text{pot}}(r_1(t), r_2(t) \dots, r_i(t) \dots, r_n(t)) / \delta r_i = f_i(t)$$

où r_i représente les coordonnées de l'atome i à un instant t et m_i sa masse.

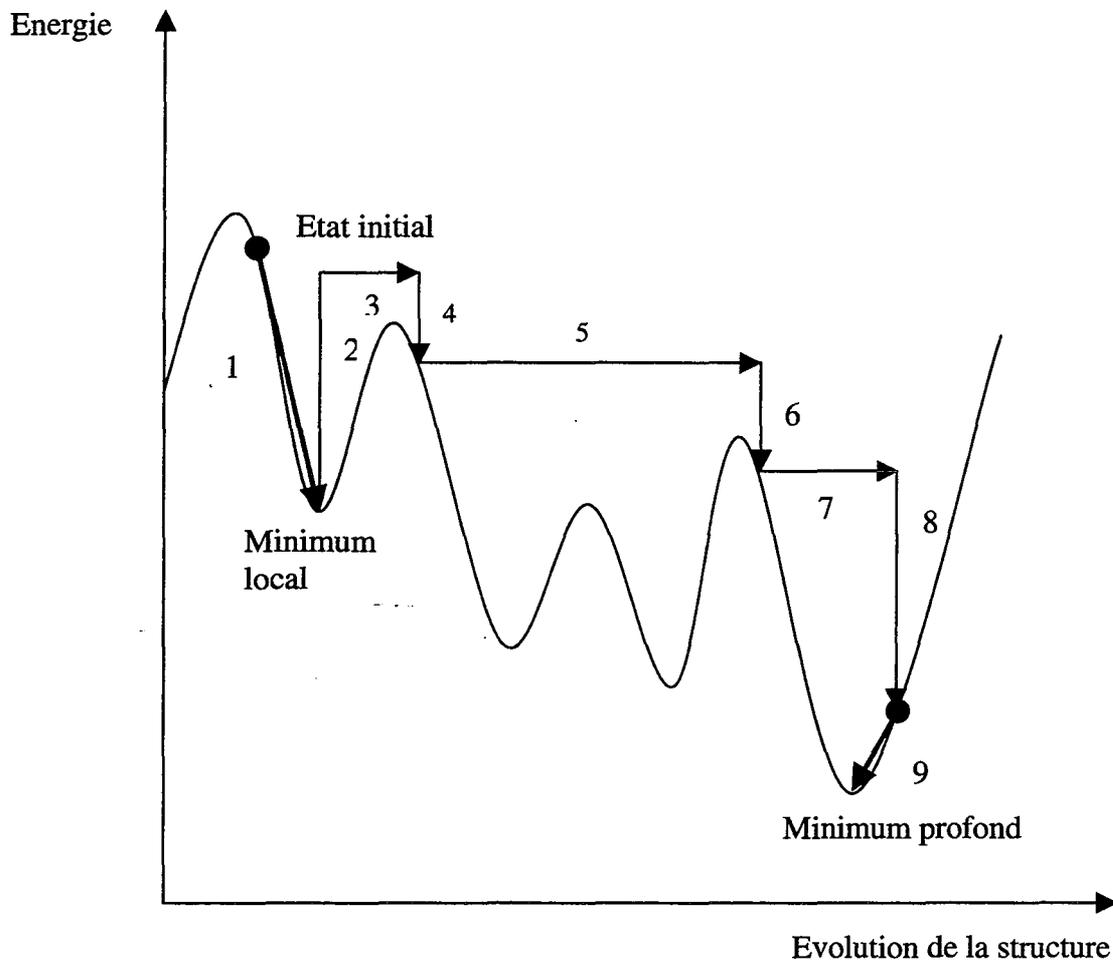


Figure 2.2: représentation schématique des étapes de la dynamique moléculaire. Les étapes 1 et 9 représentent la minimisation d'énergie permettant d'atteindre des minima locaux et profond. Le recuit simulé commence en 2 par l'augmentation de la température suivie d'une diminution de celle-ci (étapes 4, 6 et 8) et de l'évolution du système entre ces différentes baisses (étapes 3, 5 et 7).

La procédure de dynamique moléculaire implantée dans X-PLOR (Brünger, 1990) utilise le recuit simulé qui consiste à fournir de l'énergie au système par augmentation de la température jusqu'à des valeurs supérieures à 3000 K. Cette augmentation se traduit par une augmentation de l'énergie cinétique nécessaire au franchissement des barrières énergétiques. Partant d'une valeur de 3000 K, la température est abaissée progressivement à 300 K durant la

simulation, avec une diminution de 25 K de la température toutes les 25 fs ce qui amène cette simulation à couvrir environ 3 picosecondes. Cette étape s'achève par une minimisation d'énergie.

Si cette procédure permet un gain considérable en temps lors de la reconstruction, un examen graphique de la structure est nécessaire après chaque étape utilisant la dynamique moléculaire. En effet, cette méthode puissante peut déplacer un fragment de la chaîne polypeptidique dans un modèle incomplet vers une zone de densité inoccupée. La puissance de la dynamique permet aussi par exemple de déplacer des chaînes latérales de quelques angströms ou de changer la stéréochimie des liaisons peptidiques.

D- Affinement des facteurs de température

Le désordre structural existant dans une molécule peut être modélisé par le facteur d'agitation thermique B_i de chaque atome qui représente son déplacement autour d'une position moyenne. L'ajustement des paramètres B_i est effectué en fin d'affinement afin que les erreurs de positionnement des atomes dans la structure durant les premières phases de l'affinement ne soient pas cumulées dans la valeur des paramètres d'agitation thermique. Au début de l'affinement, ces facteurs sont affectés d'une valeur de 15 \AA^2 représentant une valeur moyenne des facteurs d'agitation rencontrés dans les protéines. Des contraintes peuvent être introduites durant leur affinement. Par exemple, deux atomes liés doivent avoir des facteurs d'agitation thermique voisins. L'énergie de restriction est donnée par la relation:

$$E_{B_i} = \sum_{\text{liaisons-}i,j} 1/2 k_B (B_i - B_j)^2$$

De même, l'application de la symétrie non cristallographique lors de cette étape introduit une restriction sur le calcul des facteurs d'agitation, en limitant leur variation autour d'une valeur moyenne calculée entre les atomes reliés par la symétrie non cristallographique.

E- Introduction des molécules d'eau et des ions

Les protéines contiennent dans leur structure et à leur surface des molécules d'eau plus ou moins ordonnées et donc plus ou moins visibles dans des cartes de densité électronique. Ces différentes molécules sont introduites de façon automatique ou manuellement dans la structure en fin d'affinement, donc à des valeurs raisonnables du R_{cris} et du R_{free} , du moment où les densités sont interprétables. L'introduction des molécules d'eau se traduit par une diminution des facteurs d'accord. Les ions présents dans les protéines, qu'ils soient catalytiques, structuraux, ou en interaction avec celles-ci en raison des conditions de purification ou de cristallisation, sont également introduits en fin d'affinement, lorsque la carte de densité permet clairement de les identifier.

Dans le cas de l'ALDH de *S. mutans*, les facteurs d'agitation thermique de ces molécules sont affinés sans restriction sur leur calcul.

F- Validation de la structure

Bien que les facteurs d'accord soient faibles en fin d'affinement, des erreurs peuvent encore être présentes dans la structure, notamment des erreurs de géométrie du modèle et/ou de positionnement des molécules d'eau. Toute molécule d'eau doit être reliée par liaison hydrogène à au moins un groupement donneur ou accepteur de la protéine ou à une autre molécule d'eau en interaction avec la protéine. En ce qui concerne la géométrie, les valeurs des angles et des longueurs de liaisons covalentes doivent appartenir à des intervalles définis par la connaissance de la structure de certaines protéines à haute résolution. La recherche de telles erreurs est facilitée par l'utilisation des logiciels X-PLOR, PROCHECK (Laskowsky *et al.*, 1993) et WHAT_CHECK (Hoof *et al.*, 1996). Une fois corrigées des dernières erreurs, les coordonnées de la structure étudiée sont déposées dans la Protein Data Bank.

G- Conclusion

Ainsi, s'il fallait plusieurs années de travail pour déterminer la structure d'une macromolécule dans les années 60, jusqu'au début des années 80 dans la majeure partie des cas, il n'en est rien aujourd'hui en raison de l'évolution rapide des moyens informatiques, des

appareils de diffraction X, et des connaissances acquises tout au long de ces différentes années. Actuellement, les étapes limitantes de la résolution d'une structure de macromolécule restent la mise au point des conditions de purification, de cristallisation et la recherche de dérivés lourds. Ces différentes étapes font que la détermination tridimensionnelle de telles molécules peut encore parfois prendre plusieurs années.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES NANCY

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

Cristallisations

L'ALDH de *S. mutans* a été produite par génie génétique et purifiée au laboratoire de Maturation des ARN et d'Enzymologie Moléculaire (M.A.E.M.) dirigé par le Docteur C. Branlant dans le groupe du Professeur G. Branlant.

I- Apoenzyme de type 1 (Apo1)¹ et holoenzyme

La protéine, précipitée au sulfate d'ammonium en tampon phosphate de potassium 50 mM, pH 6,4, 5 mM β -mercaptoéthanol, est dialysée contre ce même tampon en présence de DTT à une concentration de 1 mM. En fin de dialyse, la mesure de la DO à 280 nm permet d'estimer la concentration de la protéine (ϵ_{280} : 204000 M⁻¹cm⁻¹).

La recherche des conditions de cristallisation de l'ALDH de *S. mutans* a été abordée de différentes façons, toujours en diffusion de vapeur en goutte suspendue (McPherson, 1990). Tout d'abord, étant donné que l'enzyme est précipitée au sulfate d'ammonium en fin de purification, les premières études de cristallogénèse ont débuté en utilisant ce sel comme

¹ Le terme d'apoenzyme de type 1 est utilisé pour désigner la première structure résolue d'apoenzyme de Sm-ALDH. Elle se différencie d'une seconde structure par la présence d'un seul ion sulfate dans le site catalytique. Pour la seconde structure d'apoenzyme de type sauvage, le terme Apo2 sera utilisé.

agent de cristallisation en présence d'un tampon phosphate de potassium 50 mM, pH 6,4. En parallèle, une matrice d'échantillonnage contenant 50 solutions ayant permis de cristalliser plusieurs protéines a été utilisée (Jancarick & Kim, 1991). L'ensemble de ces essais a conduit à l'obtention de plusieurs conditions de cristallisation. Tous les essais de cristallisation ont été réalisés à 20°C.

Une condition de cristallisation a été retenue pour l'étude cristallographique de l'ApoI et de l'holoenzyme.

Composition du milieu de cristallisation (puits):

- sulfate d'ammonium: 1,9 à 2,0 M
- tampon HEPES: 0,1 M, pH 7,5
- PEG 400: 2 %

L'ajout du sulfate d'ammonium à une concentration de 1,9 M ou 2,0 M à la solution d'HEPES tamponnée par NaOH pour atteindre un pH de 7,5, s'accompagne d'une augmentation du pH de la solution qui atteint des valeurs proches de 8,0 en raison de la formation d'ammoniaque.

Composition des gouttes:

- 2 à 3 μ l de protéine à 40 mg/ml
- 2 à 6 μ l de solution précipitante

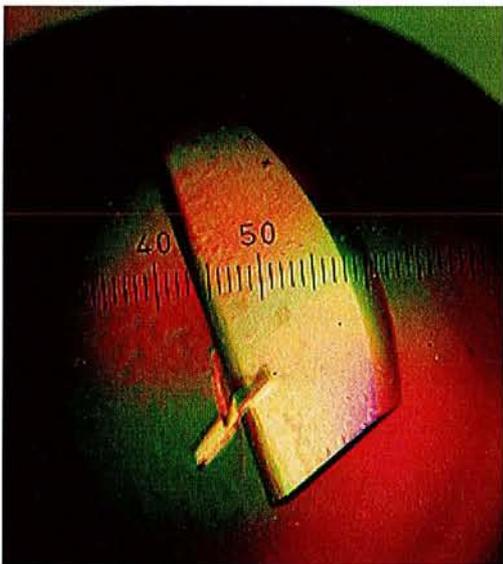


Figure 3.1: cristaux d'apoenzyme de l'ALDH de S. mutans. 1 graduation représente 0.02 mm.

Ce mélange porte la concentration initiale de la protéine dans la goutte à 10 mg/ml pour un mélange de 2 μ l de protéine et 6 μ l de solution précipitante. Lors de la contraction de

la goutte, la concentration de la protéine va progressivement augmenter pour atteindre 20 mg/ml. A l'état final, cette forte concentration de protéine associée aux agents précipitants permet l'obtention de cristaux de taille satisfaisante pour une étude cristallographique, atteignant parfois plus d'un millimètre de côté dans une direction de l'espace avec des dimensions de l'ordre de quelques centaines de microns pour les deux autres directions.

Les essais préalables de cristallisation ont été réalisés en mélangeant 4 μ l de protéine à une concentration estimée de 5 mg/ml et 1 μ l de solution précipitante, ce qui amenait la concentration finale de la protéine à 20 mg/ml avec un nombre plus important de cristaux dans la goutte. Ainsi, l'affinement de la condition de cristallisation, par variation de la cinétique d'équilibre, a permis d'arriver au même état final avec une cinétique plus lente et un taux de nucléation diminué. L'augmentation de la taille des cristaux est due à l'augmentation de la quantité de matériel protéique disponible. En effet celle-ci devient plus importante du fait de la diminution du nombre de cristaux en croissance dans la goutte.

Grâce à l'obtention de cristaux de taille satisfaisante, des données cristallographiques à haute résolution pour la forme Apo1 de l'ALDH de *S. mutans* ont été enregistrées.

L'holoenzyme a été préparée par addition d'un large excès de NADP (environ 16 mM final) à la solution protéique dans les mêmes conditions de tampon que l'Apo1. La solution de cristallisation et la concentration en protéine demeurent identiques. Des cristaux de la même forme que ceux d'Apo1 ont été obtenus.

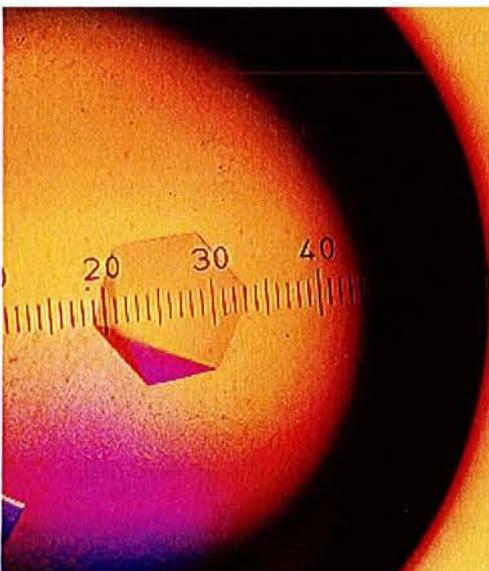


Figure 3.2 : cristal d'holoenzyme obtenu en présence de sulfate d'ammonium. 1 graduation représente 0,02 mm.

Une seconde forme cristalline de l'holoenzyme est également obtenue en présence uniquement de sulfate d'ammonium à une concentration de 2,0 M ou en association avec un tampon phosphate de potassium 50mM, pH 6,4 (Figure 3.2). La concentration initiale en protéine est estimée à 40 mg/ml. Ces cristaux apparaissent après précipitation de la protéine. Les gouttes contiennent 2 µl de protéine et 6 µl de solution précipitante.

II- L'apoenzyme de type 2 ou Apo2

Deux conditions de cristallisation différentes ont été utilisées pour l'étude de cette forme enzymatique. Les essais sont effectués à 20°C par la méthode de la goutte suspendue en phase vapeur (McPherson, 1990). La première condition de cristallisation est:

- sulfate d'ammonium 2,0 M
- tampon HEPES 0,1 M pH 7,5
- PEG 400, 2 %
- concentration en protéine: 5 mg/ml
- composition de la goutte: 2-6 µl de protéine et 1-2 µl de solution précipitante.

Comme les séquences des GAPDH NADP dépendantes non phosphorylantes du pois et du maïs présentent 50 % d'identité avec celle de *S. mutans* (Boyd *et al.*, 1995) et que leur purification utilise, en plus du sulfate d'ammonium, l'acétone (Habenicht *et al.*, 1994), alors le PEG est substitué par de l'acétone dans les essais de cristallisation. Cependant cette condition s'est avérée peu reproductible car sensible au lot d'acétone.

III- Etude du complexe ternaire

A- Les différentes approches

Afin d'obtenir un cristal dans lequel la densité électronique du substrat fixé à la protéine puisse être visualisée, il est nécessaire de disposer soit:

- d'un mutant inactif ou faiblement actif
- d'un substrat synthétique et d'une réaction dont le déclenchement peut se faire sous l'action d'une impulsion lumineuse, impliquant l'utilisation de la technique de Laue (Johnson, 1992).

L'étude du complexe ternaire a été réalisée en utilisant des mutants faiblement actifs ou inactifs.

B- Le choix des mutations

Trois mutants peuvent raisonnablement être utilisés pour cette étude: E250A, C284A et C284S.

Il est probable que travailler avec le mutant E250A permette de mimer beaucoup mieux la réalité puisque dans ce cas, la cystéine catalytique est présente. Cependant, ce mutant possède encore une activité catalytique. En effet, la constante de vitesse de la réaction k_{cat} est de $0,01 \text{ s}^{-1}$. Le choix des deux autres mutants se révélera par la suite plus judicieux d'un point de vue structural pour l'étude du mécanisme réactionnel.

L'utilisation des deux autres mutants, C284S et C284A, semble plus appropriée pour cette étude. En particulier, le mutant C284S (k_{cat} de $0,001 \text{ s}^{-1}$) mime bien le groupement sulfhydryle de la cystéine par la présence du groupement hydroxyle de la sérine. Quant au mutant C284A totalement inactif, il ne nécessite aucune précaution particulière dans le contrôle du temps de diffusion du substrat dans le cristal.

C- Les conditions de cristallisation

Trois conditions de cristallisation ont été retenues. Les deux premières correspondent à celles citées pour cristalliser l'Apo1 et l'holoenzyme et sont utilisées pour la cristallisation des mutants C284A et C284S:

Puits : - sulfate d'ammonium: 1,9 à 2,0 M

- tampon HEPES: 0,1 M, pH 7,5

- PEG 400: 2 %

ou 2,0 M sulfate d'ammonium

Goutte : - 2 à 3 µl de protéine à 40 mg/ml

- 2 à 6 µl de solution précipitante

Les conditions de cristallisation et de dialyse pour le mutant E250A sont différentes de celles de l'Apo1 par la valeur du pH (pH 6,4), en raison de la réactivité et de la déacylation de la cystéine favorisées à un pH basique. Un pH de 7 est imposé aux solutions de dialyse et de cristallisation. De cette façon, l'étape de déacylation est ralentie. Ainsi, il serait possible d'obtenir le thioacylenzyme. Dans tous les cas, les mutants sont cristallisés sous leur forme holoenzyme à une concentration estimée de 40 mg/ml.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Enregistrements et traitements des données de diffraction

I- L'Apo1 et l'holoenzyme

A- Recherche des conditions de cryogénie

Les données de diffraction ne pouvant être enregistrées à température ambiante en raison de la dégradation des cristaux sous l'influence du rayonnement, des conditions de cryogénie ont été recherchées (Rodgers, 1994). Il s'agit d'empêcher la formation d'un réseau de glace autour et dans le cristal ce qui perturberait la diffraction. La présence de glace se caractérise par un aspect opaque de la solution de cryogénie lors de la congélation et/ou, si ce n'est pas le cas, par l'apparition d'anneaux sur le cliché de diffraction du fait de la diffraction des RX par la glace. Ces anneaux résultent de l'orientation multiple des cristaux de glace à l'intérieur et à l'extérieur du cristal de protéine et sont identiques à ceux observés sur les clichés de poudre où une sphère et non plus un nœud traverse la sphère d'Ewald.

Pour pallier ce problème, des tests sont réalisés sur des solutions cryogéniques contenant non seulement le cryoprotecteur mais aussi la solution saline dans laquelle le cristal s'est développé. Une condition utilisant de l'éthylène glycol à une concentration de 17,5 % a été retenue. Cependant le trempage direct des cristaux dans une solution à haute concentration

en éthylène glycol entraîne leur dégradation. Leur transfert dans une solution concentrée n'est possible que s'ils sont au préalable trempés dans des solutions de concentrations en éthylène glycol croissantes par pas de 5 %.

B- Collecte des données

Ce sont en tout quatre jeux de données qui ont été enregistrés pour déterminer la structure de ces deux formes enzymatiques.

1- L'apoenzyme

Un premier jeu de données a été enregistré à 2,2 Å au laboratoire sur un cristal d'une taille d'environ 0,2 x 0,2 x 1,0 mm³ à une température de 100 K. Les données sur ce même cristal ont été enregistrées à une résolution de 1,8 Å au LURE sur la ligne W32. Pour cela, le cristal est récupéré sur le diffractomètre dans de l'éthane liquide. L'ensemble est ensuite plongé dans l'azote liquide.

2- L'holoenzyme

Le premier jeu de données a été collecté à 2,8 Å sur la ligne de lumière D2AM à l'ESRF sur un cristal obtenu en tampon phosphate de potassium 50 mM, pH 6,4, 2,0 M sulfate d'ammonium. L'analyse des données montrait la présence d'un paramètre de maille de 280,9 Å obligeant à éloigner le détecteur pour éviter le recouvrement des taches de diffraction.

Le second jeu de données de cette forme protéique a été enregistré au laboratoire à 2,6 Å à 100 K sur des cristaux différents (groupe d'espace P2₁2₁2) des premiers (groupe d'espace P3₂21), obtenus dans les mêmes conditions de cristallisation que ceux d'apoenzyme.

C- Analyse des jeux de données

Les données ont été traitées en utilisant le logiciel DENZO (Otwinowsky & Minor, 1997) pour déterminer et affiner les paramètres cristallins, et le programme SCALEPACK (Otwinowsky & Minor, 1997) pour la mise à l'échelle des données lors de leur réunion.

1- L'apoenzyme

Les statistiques de l'enregistrement de données au laboratoire et au LURE sont résumées dans le tableau 4.1:

Jeu de données	W32 LURE	Laboratoire (SDXM)
Résolution (Å)	1,8 ($\lambda= 0,96$ Å)	2,2 ($\lambda= 1,5418$ Å)
Température (K)	100	100
Groupe d'espace	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Paramètres de maille (Å)	a= 144,4; b= 154,7; c= 114,0	a= 144,3; b= 154,6; c= 114,3
Réflexions uniques	215812	127097
R _{sym} (%)	6,1	8,3
Complétude (%)	94,9	97,9
Multiplicité	3,5	3,6
Moyenne I/ σ (I)	11,7	-

Tableau 4.1: statistiques des enregistrements de données sur un cristal d'apoenzyme de l'ALDH de S. mutans.

La valeur élevée de la redondance, comprise entre 3 et 4 pour les deux jeux de données, associée à une complétude supérieure ou égale à 95 % et à une valeur raisonnable du R_{sym} permettent d'affirmer que ces jeux de données sont de bonne qualité. La différence de R_{sym} entre ces deux jeux de données provenant du même cristal ne peut résulter que de différences sur les paramètres expérimentaux utilisés durant le traitement des données. Le dispositif influence en effet l'intensité des taches qui dépend de la puissance du flux de RX (les taches de plus faible intensité mesurées sur une anode tournante voient leurs intensités augmentées au synchrotron), la valeur donnée à la mosaïcité (augmentée de la divergence du faisceau) et de fait, la taille des boîtes à utiliser durant le traitement des données....

2- L'holoenzyme

Le premier jeu de données enregistré à l'ESRF, traité avec XDS (Kabsch, 1988), fait apparaître une valeur de R_{sym} légèrement supérieure à 10 % et une faible redondance (comprise entre 1 et 2). En raison de ces statistiques médiocres, ce jeu ne sera jamais utilisé pour étudier la structure de l'holoenzyme. Les données de cette forme cristalline (obtenue en sulfate d'ammonium seul ou en présence d'un tampon phosphate) seront enregistrées au LURE sur la ligne D41 à 2,9 Å de résolution. Les données sont traitées jusqu'à 3,0 Å en raison de la valeur élevée du R_{sym} au-delà de cette résolution. Les statistiques de l'enregistrement sont résumées dans le tableau 4.2.

Jeu de données	D41 LURE
Résolution (Å)	3 Å ($\lambda = 1,375$ Å)
Température (K)	100
Groupe d'espace	P3 ₂ 21
Paramètres de maille (Å)	a= 156,3; b= 156,3; c= 280,9
Réflexions uniques	76372
R_{sym} (%)	8,3
Complétude (%)	95,6
Multiplicité	3-4
Moyenne $I/\sigma(I)$	10,9

Tableau 4.2: statistiques d'un enregistrement de données à 3 Å sur un cristal de la forme trigonale de l'holoenzyme de l'ALDH de S. mutans.

Un autre jeu de données enregistré au laboratoire sur une forme cristalline orthorhombique (cristaux obtenu dans les mêmes conditions que ceux de l'Apo1) ne montre pas des valeurs plus encourageantes de ces statistiques pour une étude structurale (Tableau 4.3). Le R_{sym} de 8,3 % associé à une faible complétude (80 %) et à une redondance comprise entre 1 et 2 caractérise un jeu de données de piètre qualité. La valeur élevée du R_{sym} provient d'une mauvaise qualité cristalline observée lors de l'étape de trempage des cristaux dans la solution de cryogénie. Durant cette étape, les cristaux avaient tendance à se détériorer.

Jeu de données (SDXM)	Holoenzyme
Résolution (Å)	2,6 ($\lambda = 1,5418$ Å)
Température (K)	100
Groupe d'espace	P2 ₁ 2 ₁ 2
Paramètres de maille (Å)	a= 144,8; b= 154,8; c= 114,3
Réflexions uniques	65491
R _{sym} (%)	8,7
Complétude (%)	82,6
Multiplicité	1,5
Moyenne I/ σ (I)	7

Tableau 4.3: statistiques de l'enregistrement de données à 2,6 Å sur un cristal de la forme orthorhombique de l'holoenzyme de l'ALDH de S. mutans.

Les faibles valeurs de la complétude et de la redondance résultent du nombre de données exclues lors du traitement en raison de très médiocres facteurs d'accord. Compte tenu de la meilleure résolution, la forme cristalline orthorhombique a été utilisée pour résoudre la structure de l'holoenzyme.

II- L'Apo2

Les enregistrements des données ont également été réalisés à 100 K mais avec des cryoprotecteurs différents de celui utilisé pour l'Apo1 et l'holoenzyme (Tableau 4.4). La valeur élevée du R_{sym} pour les cristaux obtenus en présence d'acétone peut être expliquée par leur détérioration lors du trempage dans la solution cryoprotectrice. Une solution de glucose à 20 % a été utilisée alors qu'une solution de glycérol à 15 % a été employée pour les cristaux obtenus en présence de PEG. La différence de paramètres de maille observée entre les cristaux d'apoenzyme de types 1 et 2 résulte probablement de la différence des cryoprotecteurs utilisés. En effet, en fonction du type de cryoprotecteur, de sa concentration, et de la durée du trempage, l'effet sera différent sur le cristal selon qu'il y pénètre plus ou moins.

	Apo2 (cristallisée en présence de PEG400) (SDXM)	Apo2 (cristallisée en présence d'acétone) (SDXM)
Résolution (Å)	2,5 ($\lambda= 1,5418 \text{ \AA}$)	3,0 ($\lambda= 1,5418 \text{ \AA}$)
Groupe d'espace	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Paramètres de maille (Å)	a= 140,8; b= 157,5; c= 110,3	a= 139,6; b= 154,1; c= 109,3
Réflexions uniques	81795	46364
R _{sym} (%)	5,9	8,2
Complétude (%)	96	98,4
Redondance	4,5	5,5

Tableau 4.4: statistiques des enregistrements de données de l'Apo2 dans deux conditions différentes de cristallisation.

III- Le complexe ternaire

L'obtention d'un complexe ternaire nécessite la présence simultanée dans la protéine du substrat et du cofacteur par cocrystallisation ou par diffusion du substrat dans des cristaux d'holoenzyme.

A- Les essais de diffusion: cas du mutant E250A

Compte tenu de la constante de vitesse de la réaction k_{cat} ($0,01 \text{ s}^{-1}$), la diffusion du substrat dans le cristal s'effectue au moment où le cristal est plongé dans la dernière solution de cryogénie. Ceci permet à la fois de faire diffuser le cryoprotecteur, ici l'éthylène glycol, et le G3P. Le dernier trempage dans la solution d'éthylène glycol permet d'amener sa concentration à 17,5 % et d'introduire le G3P. Après 80 à 90 secondes, le cristal est extrait et soumis à un flux d'azote à 100K. Différentes concentrations de G3P ont été testées en raison de la diffraction anisotrope observée lors des essais. La diffusion du substrat dans les cristaux orthorhombiques du mutant E250A entraîne un désordre tel que la diffraction s'en trouve perturbée. Aucun jeu de données ne sera enregistré sur ces cristaux. Pour cette raison, le

mutant E250A a été abandonné au profit du mutant C284S cristallisé dans deux conditions différentes.

B- Les essais de diffusion: cas des cristaux orthorhombiques du mutant C284S

En raison de la faible activité catalytique de ce mutant, le trempage est réalisé dans la goutte où les cristaux ont grandi. Diverses concentrations de G3P sont testées jusqu'à une concentration avoisinant le K_M (46 μM ; Marchal *et al.*, 2000) en raison de la diffusion anisotrope persistante comparable à celle observée dans le cas du mutant E250A. Le temps maximal de trempage réalisé est de 16 minutes pour le mutant C284S. L'absence de diffraction anisotrope permet alors d'enregistrer les données. Les statistiques de l'enregistrement à 2,2 Å de résolution sont résumées dans le tableau 4.5. La résolution de cette structure montrera que le G3P n'est pas présent dans le site catalytique de l'enzyme.

	C284S (SDXM)
Résolution (Å)	2,2 ($\lambda= 1,5418$ Å)
Groupe d'espace	P2 ₁ 2 ₁ 2
Paramètres de maille (Å)	a= 144,8; b= 154,8; c= 114,6
Réflexions uniques	124052
R _{sym} (%)	7,9
Complétude (%)	93,8
Redondance	5,3
I/ σ (I)	9,9

Tableau 4.5: statistiques de l'enregistrement de données sur le mutant C284S (système orthorhombique).

La réunion des résultats expérimentaux permet d'avancer l'hypothèse d'un problème d'empilement cristallin pour les cristaux de la forme orthorhombique. En effet, plus celui-ci est compact, et plus les molécules diffusent difficilement dans le cristal en entraînant

probablement des ruptures au niveau des contacts intermoléculaires qui en assurent la cohésion.

En raison de ces problèmes de diffraction, la forme cristalline trigonale sera utilisée. Dans ce cas, les cristaux obtenus en sulfate d'ammonium 2,0 M possèdent un très grand pourcentage de solvant comparé à la forme orthorhombique (75 % contre 60 %).

C- Les essais de diffusion: cas des cristaux trigonaux du mutant C284S

Ce type de cristaux a déjà été obtenu pour l'enzyme sauvage en présence de cofacteur. Leur taux de solvant est d'environ 75 %. Dans ce cas, un trempage d'une minute environ dans 17,5 % d'éthylène glycol suffit pour enregistrer à 100 K les données de diffraction alors que les cristaux orthorhombiques nécessitent un trempage plus doux avec des concentrations croissantes en éthylène glycol. Le D-G3P est ajouté à la goutte contenant le cristal de façon à obtenir une concentration finale de 2,5 mM. Au préalable, les données ont été enregistrées sur un cristal du mutant sous sa forme holoenzyme obtenu dans les mêmes conditions. Les statistiques des jeux de données sont consignées dans le tableau 4.6.

	C284S (SDXM)	C284S ternaire (SDXM)
Résolution (Å)	3,15 ($\lambda= 1,5418$ Å)	3 ($\lambda= 1,5418$ Å)
Groupe d'espace	P3 ₂ 21	P3 ₂ 21
Paramètres de maille (Å)	a= 157,7; b= 157,7; c= 280,9	a= 158,5; b= 158,5; c= 282,8
Réflexions uniques	63894	65980
R _{sym} (%)	12,7	9,7
Complétude (%)	91,5	80,3
Redondance	1,6	1-2
I/ σ (I)	5,4	7,3

Tableau 4.6: statistiques des enregistrements de données sur le mutant C284S et le complexe ternaire (système trigonal).

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

Résolution des structures et affinements

I- Résolution de la structure de l'Apo1 et de l'holoenzyme

A- L'Apo1

Les premiers essais de remplacement moléculaire sont réalisés à l'aide de la structure dimérique de l'ALDH3 (Liu *et al.*, 1997), première structure disponible, puis par l'utilisation de la structure tétramérique de l'ALDH2 (Steinmetz *et al.*, 1997) dès sa parution. Dans ces deux cas, les chaînes latérales des acides aminés non conservés dans l'alignement de séquences sont remplacées par des alanines, puisque le remplacement moléculaire utilise la similitude de repliement ainsi que le poids électronique de chaque atome lors de la comparaison des fonctions de Patterson. Donc le remplacement des résidus non conservés par des alanines devrait globalement diminuer tous les signaux des différentes fonctions utilisées en remplacement moléculaire, le signal de la solution étant néanmoins plus facilement décelable en raison d'un niveau de bruit nettement abaissé.

1- Etudes préliminaires

- *La fonction d'autocorrélation*

Son calcul à l'aide du programme GLRF (Tong & Rossmann, 1990) permet de confirmer la présence d'un seul tétramère dans l'unité asymétrique, comme le suggérait le résultat du calcul de la proportion de solvant (60 %) (Matthews, 1968).

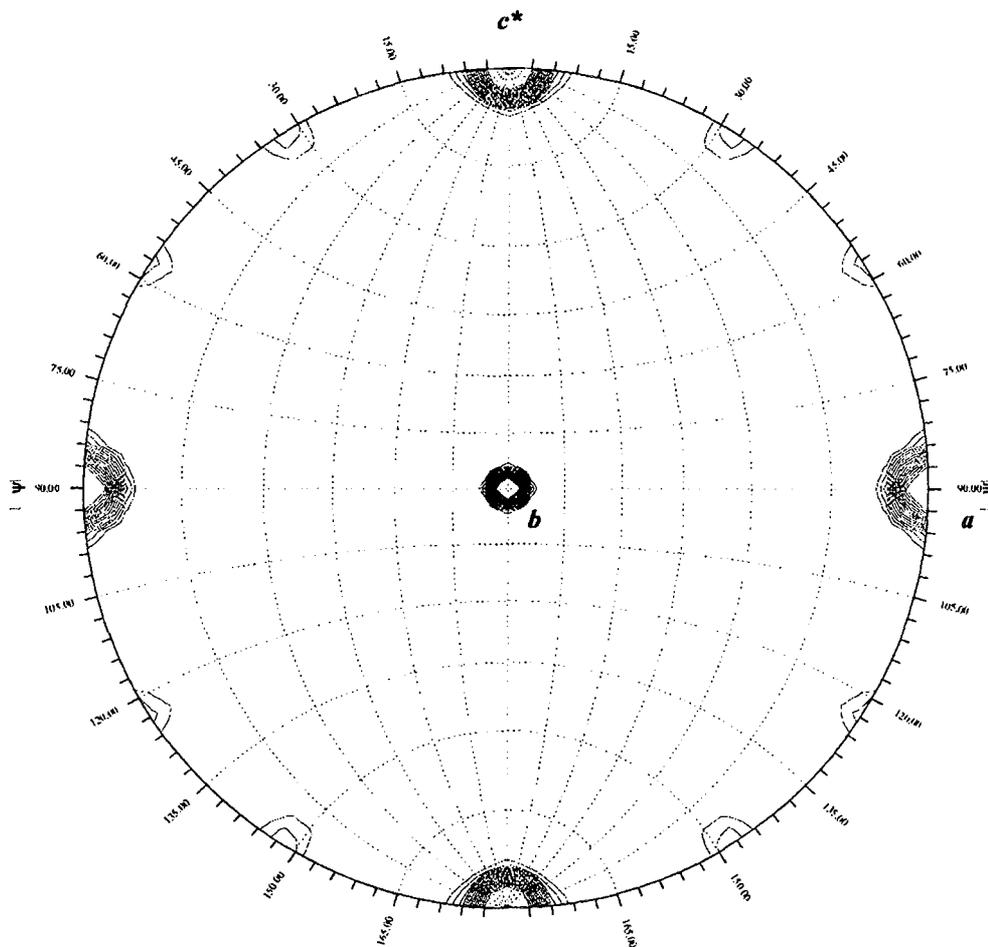


Figure 5.1: section $\kappa=180^\circ$ de la fonction d'autocorrélation calculée entre 10 et 5 Å sur le jeu de données de l'apoenzyme.

La fonction d'autocorrélation révèle l'orientation des deux axes 2 reliant les monomères dans le plan (a,c) de la maille. Ils se situent l'un et l'autre à 33° et 123° (ou à 57° et 147°) de l'axe a (Figure 5.1). Si deux axes sont perpendiculaires l'un à l'autre, il existe un troisième axe d'ordre 2 qui leur est perpendiculaire. Le troisième axe 2 est alors dirigé selon b dans ce cas.

- La Patterson native

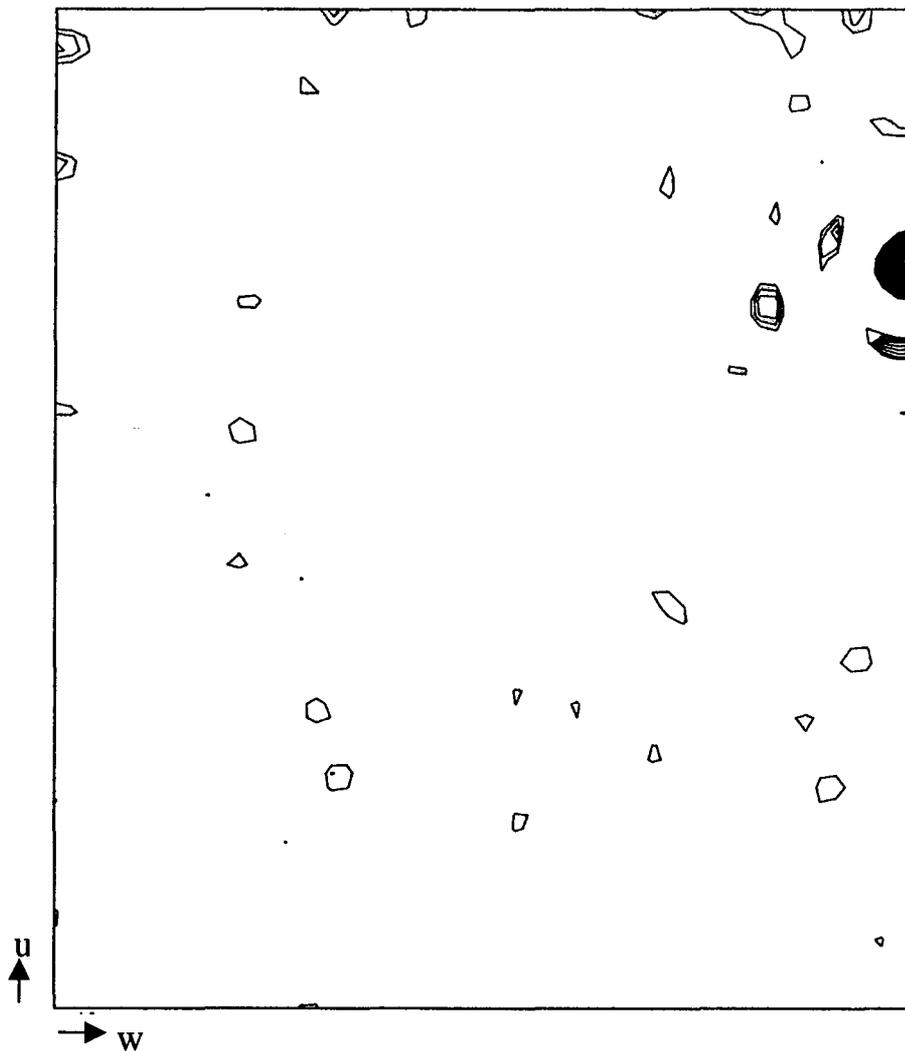


Figure 5.2: section de Harker $v=0,5$. La fonction de Patterson native a été calculée entre 10 et 5 Å sur le jeu de données de l'apoenzyme.

La fonction de Patterson native est calculée en utilisant le programme FFT (Eyck, 1973) de la chaîne de programmes CCP4 (Collaborative Computational Project, Number 4, 1994). La recherche de l'axe 2 non cristallographique parallèle à **b** est effectuée dans la section de Harker $v=0,5$. Dans cette section, $u=2x-1/2$ et $w=2z$. Le pic dû à la symétrie non cristallographique se situe en $u=0,37$ et $w=0,5$ et donc l'axe 2 non cristallographique parallèle à **b** est sécant au plan (**a,c**) en $x=0,9365$ ou $0,4365$ et $z=0,25$ ou $0,75$ (Figure 5.2).

Le nombre de molécules présentes dans l'unité asymétrique, quatre, implique comme modèle pour le remplacement moléculaire l'association de deux dimères d'ALDH3 ou l'utilisation du tétramère de l'ALDH2 ou de l'ALDH9 (Johansson *et al.*, 1998).

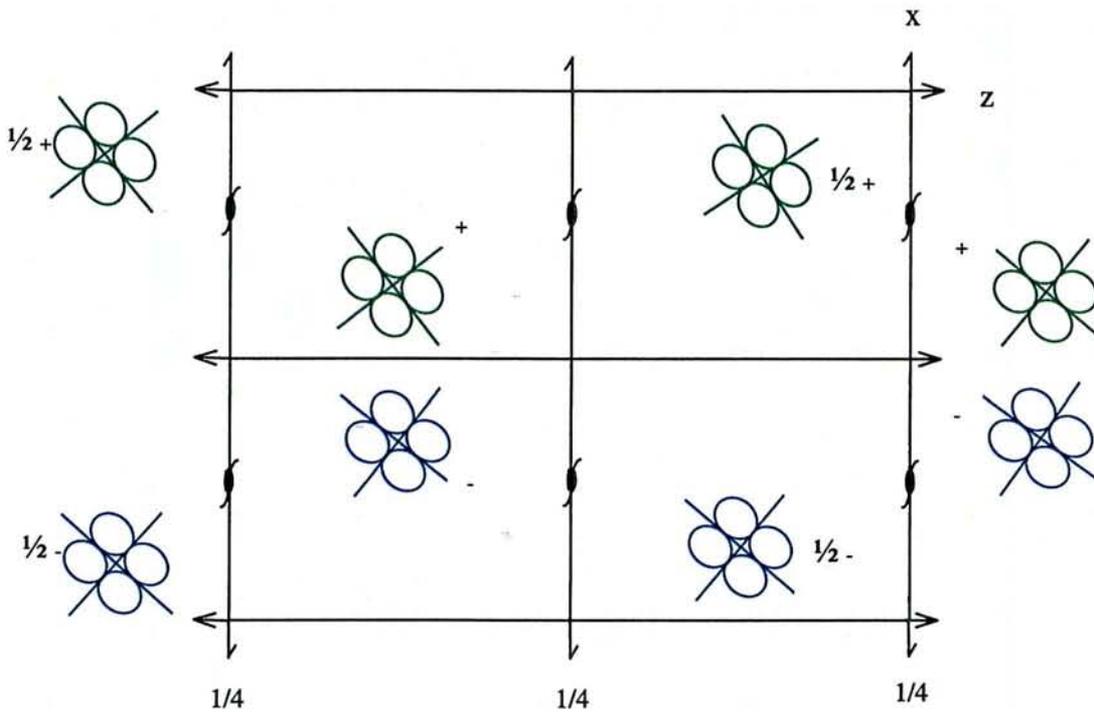


Figure 5.3: représentation schématique de la position des axes 2 non cristallographiques reliant les monomères des tétramères de l'ALDH dans la maille orthorhombique déduite de la fonction d'autocorrélation et de la fonction de Patterson native.

2- Remplacement moléculaire avec l'ALDH3

Tous les essais de remplacement moléculaire, quels que soient les modèles utilisés, sont réalisés en utilisant le logiciel AMoRe (Navaza, 1994). La première structure utilisée comme modèle est celle de l'ALDH3 (dimérique) résolue en 1997. Sa séquence présente 27,3 % d'identité avec celle de l'ALDH de *S. mutans*. La partie N-terminale de la structure comporte une quarantaine de résidus de moins que celle de l'ALDH de *S. mutans*. Toutes les tentatives de remplacement moléculaire avec ce modèle en faisant varier différents paramètres comme la résolution, les facteurs d'agitation thermique, la division des dimères en monomères, ont échoué. Les essais de remplacement moléculaire à l'aide de ce modèle furent stoppés dès que les coordonnées de l'ALDH2, ALDH tétramérique (Steinmetz *et al.*, 1997), furent disponibles dans la Protein Data Bank.

L'échec de la résolution de la structure par remplacement moléculaire à l'aide du modèle de l'ALDH3 est probablement dû à la faible identité de séquences entre les deux protéines (27,3 %) et au nombre d'acides aminés en N-terminal manquant dans cette

structure, associés à l'agencement des monomères en dimère (voir analyses structurales et discussion, paragraphe F) (Cobessi *et al.*, 1999).

3- Remplacement moléculaire avec l'ALDH2

Le remplacement moléculaire est réalisé à partir d'un tétramère en utilisant les coordonnées atomiques de l'ALDH mitochondriale ou ALDH2. La séquence du monomère de l'ALDH2 possède 30,7 % d'identité avec celle du monomère de l'ALDH de *S. mutans*. De nouveau, les résidus non conservés sont remplacés par des alanines. Une solution de la fonction de rotation très contrastée par rapport aux autres proposées est obtenue pour une gamme de résolution comprise entre 9 et 6 Å. Après l'étape de translation, les valeurs du coefficient de corrélation et du facteur d'accord cristallographique sont respectivement de 47,6 % et 47,8 %. Les matrices de rotation et de translation sont appliquées aux coordonnées du modèle pour débiter les étapes d'affinement et de reconstruction de la structure.

- *Reconstruction et affinement*

Le modèle orienté et positionné dans la maille orthorhombique est soumis aux différentes étapes de l'affinement par corps rigide. Le tétramère puis les monomères et enfin les différents domaines créés en fonction des lacunes de l'alignement de séquences sont translétés et tournés dans l'unité asymétrique avec une extension de la gamme de résolution au cours des différentes étapes. Ceci se traduit par une baisse des facteurs d'accord R_{cryst} et R_{free} .

A des valeurs stables du R_{cryst} et du R_{free} , un affinement par minimisation d'énergie est réalisé en utilisant le logiciel X-PLOR (Brünger, 1990). Puis une carte de densité électronique est calculée. Le modèle d'un monomère est reconstruit en utilisant TURBO-FRODO (Roussel & Cambillau, 1991) très partiellement en raison du faible nombre de chaînes latérales apparaissant dans les cartes de densité électronique. Le tétramère est régénéré par application de la symétrie non cristallographique. Les matrices de rotation et de translation la décrivant sont calculées par le programme Lsq (Kleywegt & Jones, 1994). Elle est déterminée à partir du modèle dont la position et l'orientation ont été affinées par corps rigide et minimisation d'énergie dans la maille après le remplacement moléculaire.

Cette seconde étape de la reconstruction a été interrompue en raison d'une divergence des facteurs d'accord révélant des calculs de phase trop imprécis. Le modèle utilisé comportait moins de 50 % des atomes de la structure finale. De plus, de nombreuses parties du modèle étaient dépourvues de densité électronique sur la chaîne principale. Le manque d'information structurale était sans doute responsable de ces erreurs de calcul des phases et donc d'une qualité médiocre de la carte de densité. Il a donc été nécessaire de modifier la densité électronique afin de mieux repositionner le modèle pour améliorer le calcul des phases.

- Nivellement de solvant et moyennation

Cette procédure est menée en utilisant la chaîne de programmes CCP4. Le calcul des facteurs de structure, à partir du modèle repositionné après affinement par corps rigide sur les domaines créés en fonction de l'alignement de séquences, est réalisé par le programme SFALL (Argawal, 1978). Une mise à l'échelle en utilisant RSTATS est réalisée entre les facteurs de structure observés F_{obs} et les facteurs de structure calculés F_{calc} , puis les figures de mérite pour la phase de chaque réflexion sont calculées par le programme SIGMAA (Read, 1986). Le nivellement de solvant et la moyennation de densité sont effectués par le programme DM (Cowtan, 1994). Les masques correspondant à chaque monomère sont calculés par les programmes MAMA (Kleygwet & Jones, 1993) et NCSMASK de CCP4. Après le nivellement de solvant, une carte de densité est calculée, et examinée pour replacer des chaînes latérales et repositionner la chaîne principale. Cette procédure cyclique a été de nouveau interrompue puisqu'aucune densité nouvelle n'apparaissait.

- Le logiciel O (Jones & Kjeldgaard, 1990)

La carte de densité électronique est différente selon les monomères. Or, si une symétrie non cristallographique est présente, les densités électroniques entre les différents monomères devraient être identiques jusqu'à une certaine résolution. Ainsi se justifie la superposition des cartes de densité électronique après amélioration des phases jusqu'à une résolution de 4 Å.

La procédure de nivellement et moyennation est couplée à ce nouveau type de reconstruction et les phases sont calculées sur un modèle en polyalanine. Le modèle polyalanine est reconstruit dans une carte de densité électronique correspondant à la

superposition de 4 cartes. Les matrices correspondant à la symétrie non cristallographique sont calculées par le logiciel O à l'aide duquel s'effectue cette reconstruction. Après chaque étape de reconstruction - modification de la densité, le facteur d'accord R calculé par le programme DM (Cowtan, 1994) diminue. Les corrélations entre les monomères ainsi que les figures de mérite augmentent également. La gamme de résolution comprise entre 17 et 4 Å pour ces calculs est étendue jusqu'à 3,5 Å.

Malgré cette amélioration, le positionnement des chaînes latérales demeure impossible et un nouveau passage dans X-PLOR (Brünger, 1990) pour un ultime affinement par corps rigide et minimisation d'énergie montre une évolution divergente des facteurs d'accord R_{crist} et R_{free} .

- Conclusion

Il est fort probable que l'utilisation d'un modèle très incomplet, en raison des lacunes de l'alignement de séquences et de la structure de la polyalanine utilisée, bien que très proche en repliement de la protéine étudiée, explique ce problème de phasage. L'utilisation d'un modèle élaboré sur la base d'une nouvelle matrice de score (blosum62 au lieu de pileuppep) aurait probablement permis de résoudre la structure. Ainsi, un mauvais alignement de séquences entre les structures primaires de l'ALDH de *S. mutans* et de l'ALDH2 a entraîné l'utilisation d'un modèle très incomplet qui a été responsable de l'échec du phasage. Les alignements de séquences réalisés avec deux matrices de scores par le programme GCG (Wisconsin Package, 1996) sont représentés pages 84 et 85. Le premier correspond à l'utilisation de la matrice pileuppep et le second à blosum62. Ceci montre qu'un alignement de séquences est fortement sensible à la matrice utilisée; dans la mesure du possible, il doit être optimisé sur la base des superpositions de structures.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

	1	50
SM-ALDH MTK QYKNYVNGEW	
ALDH2	MLRAVALAAA RLGPRQGRRL LSAATQAVPT PNQQPEVLYN QI..FINNEW	
	51	100
SM-ALDH	K..LSENEIK IYEPASGAEL GSVFAMSTEE VDYVYASAKK A...QPAWRA	
ALDH2	HDAVSKKTFP TVNPSTGDVI CHVAEGDKAD VDRAVKAARA AFQLGSPWRR	
	101	150
SM-ALDH	LSYIERAAYL HKVADILMRD KEKIGAILSK EVAKGYKSAV SEVVRTAETII	
ALDH2	MDASERGRLL NRLADLIERD RTYLAALETL DNGKPY.... ..II	
	151	200
SM-ALDH	NYAAEEGLRM EGEVLEGGSF EAASKKKIAVVRR EPVGLVLAIS	
ALDH2	SYLVLDLDMVL KCLRYAGWA DKYHGKTIPI DGDYFSYTRH EPVGVCGQII	
	201	250
SM-ALDH	PFNYFVNLAG SKIAPALIAG NVIAFKPPTQ GSISGLLLAE AFAEAGLPAG	
ALDH2	PWNFPLMQA WKLGPALATG NVVVMKVAEQ TPLTALYVAN LIKEAGFPPG	
	251	300
SM-ALDH	VFNTITGRGS EIGDYIVEHQ AVNFINFYGS TGIGERIGKM AG...MRPIM	
ALDH2	VVNVIPIGFGP TAGAAIASHE DVDKVAFTGS TEVGHLIQVA AGKSNLKRVT	
	301	350
SM-ALDH	LELGKDSAI VLEDADLELT AKNIIAGAFG YSGQRCTAVK RVLVMESVAD	
ALDH2	LEIGGKSPNI IMSDADMDWA VEQAHFALFF NQGQCCAGS RTFVQEDIYA	
	351	400
SM-ALDH	ELVEKIREKV LALTIGNP.E DDADITPLID TKSADYVEGL INDANDKGAT	
ALDH2	EFVERSVARA KSRVVGPNPF SRTEQGPQVD ETQFKKVLGY IKSGKEEGLK	
	401	450
SM-ALDH	AL...TEIKR EGNLICPILF DKVTTDMRLA WEEFPGPVLV IIRVTSVEEA	
ALDH2	LLCGGAAAAD RGYFIQPTVF GDLQDGMTIA KEEIFGPVMQ ILKFKSMEEV	
	451	500
SM-ALDH	IEISNKSEYG LQASIFTNDF PRAFGIAEQL EVGTVHINNK TQRGTDNFPF	
ALDH2	VGRANNSKYG LAAAVFTKDL DKANYLSQAL QAGTVWVNCY DVFGAQS.PF	
	501	535
SM-ALDH	LGAKKSGAGI QGVKYSIEAM TTVKSVVFDI K....	
ALDH2	GGYKLSGSGR ELGEYGLQAY TEVKTVTVRV PQKNS	

Alignement des séquences de l'ALDH2 et de Sm-ALDH en utilisant la matrice pileup et le programme GCG (Wisconsin Package, 1996).

```

1                               50
SM-ALDH .....MTKQY KNYVNGEW..
ALDH2  MLRAVALAAA RLGPRQGRRL LSAATQAVPT PNQQPEVLYN QIFINNEWHD

51                               100
SM-ALDH KLSENEIKIY EPASGAELGS VPAMSTEEVD YVYASAKKA. ..QPAWRALS
ALDH2  AVSKKTFPTV NPSTGDVICH VAE GDKADVD RAVKAARAAF QLGSPWRRMD

101                              150
SM-ALDH YIERAAYLHK VADILMRDKE KIGAILSKEV AKGY.KSAVS EVVRTAEIIN
ALDH2  ASERGRLLNR LADLIERDRT YLAALETLDN GKPYIISYLV DLDMVLKCLR

151                              200
SM-ALDH YAAEGLRME GEVLEGGSF E AASKKKIAVV RREPVGVLVA ISPFNYPVNL
ALDH2  YYAGWADKYH GKT..... PIDGDYFSYT RHEPVGVCQG IIPWNFPLLM

201                              250
SM-ALDH AGSKIAPALI AGNVIAFKPP TQGSISGLLL AEAFAEAGLP AGVFNTITGR
ALDH2  QAWKLGPALA TGNVVMKVA EQTPLTALYV ANLIKEAGFP PGVVNVIPGF

251                              300
SM-ALDH GSEIGDYIVE HQAVNFINFT GSTGIGERIG KMAG...MRP IMLELGGKDS
ALDH2  GPTAGAAIAS HEDVDKVAFT GSTEVGHLIQ VAAGKSNLKR VTLEIGGKSP

301                              350
SM-ALDH AIVLEDADLE LTAKNIIAGA FGYSQRCTA VKRVLVMSV ADELVEKIRE
ALDH2  NIIMSDADM WAVEQAHFAL FPNQGQCCCA GSRTFVQEDI YAEFVERSVA

351                              400
SM-ALDH KVLALTIGNP EDD.ADITPL IDTKSADYVE GLINDANDKG ATALT...EI
ALDH2  RAKSRVVGPN FDSRTEQGPQ VDETQFKKVL GYIKSGKEEG LKLLCGGGAA

401                              450
SM-ALDH KREGNLICPI LFDKVTDMR LAWEPPFGPV LPIIRVTSVE EAIEISNKSE
ALDH2  ADRGYFIQPT VFGDLQDGMT IAKEEIFGPV MQILKFKSME EVVGRANNSK

451                              500
SM-ALDH YGLQASIFTN DFPRAFGIAE QLEVGTVHIN NKTQRGTDNF PFLGAKKSGA
ALDH2  YGLAAAVFTK DLDKANYLSQ ALQAGTVVWN CYDVFGAQS. PFGGYKLSGS

501                              527
SM-ALDH GIQGVKYSIE AMTTVKSUVF DIK....
ALDH2  GRELGEYGLQ AYTEVKTVTV RVPQKNS
    
```

S.C.D. - U.H.P. NANCY 1
 BIBLIOTHÈQUE DES SCIENCES
 Rue du Jardin Botanique
 54600 VILLERS LES NANCY

Alignement des séquences de l'ALDH2 et de Sm-ALDH en utilisant la matrice blosum62 et le programme GCG (Wisconsin Package, 1996).

4- Remplacement moléculaire avec l'ALDH9

L'alignement des séquences de Sm-ALDH et de l'ALDH9 du foie de morue (30 % d'identité) par le programme GCG (Wisconsin Package, 1996) en utilisant la matrice de score blosum62 permet de créer un tétramère modèle dans lequel plus de 330 résidus sont remplacés par des alanines. Le remplacement moléculaire fait apparaître une seule solution de la fonction de rotation et de translation. Le coefficient de corrélation s'élève à 53,7 % et le factor d'accord R à 46,5 % au lieu respectivement de 47,6 % et 47,8 % dans le cas du modèle issu de l'ALDH2. La procédure de reconstruction - affinement est alors amorcée sur le modèle positionné dans la maille par application des matrices de rotation et translation issues du remplacement moléculaire.

5- Reconstruction et affinement

Cette étape de la résolution d'une structure alterne les reconstructions du modèle dans la densité électronique et donc dans l'espace direct avec des affinements de la structure dans l'espace réciproque. Les étapes de reconstruction sont effectuées à l'aide des logiciels O (Jones & Kjeldgaard, 1990) pour la plupart et TURBO-FRODO (Roussel & Cambillau, 1991) pour quelques unes. Les monomères sont décomposés en 6 domaines conformément à l'alignement de séquences. Le modèle est reconstruit dans des cartes de densité électronique $F_{\text{obs}}-F_{\text{calc}}$ et $3F_{\text{obs}}-2F_{\text{calc}}$. Le premier type de carte permet de faire apparaître les parties fausses du modèle ou bien des zones incomplètes du modèle, car le calcul revient à retrancher la densité calculée ρ_{calc} de la densité observée ρ_{obs} . Les pics de densité correspondant à des parties manquantes dans le modèle (densité positive) seront encore plus prononcés par le calcul de synthèses de Fourier $2F_{\text{obs}}-F_{\text{calc}}$ ou $3F_{\text{obs}}-2F_{\text{calc}}$. Les affinements sont réalisés en utilisant le logiciel X-PLOR (Brünger, 1990).

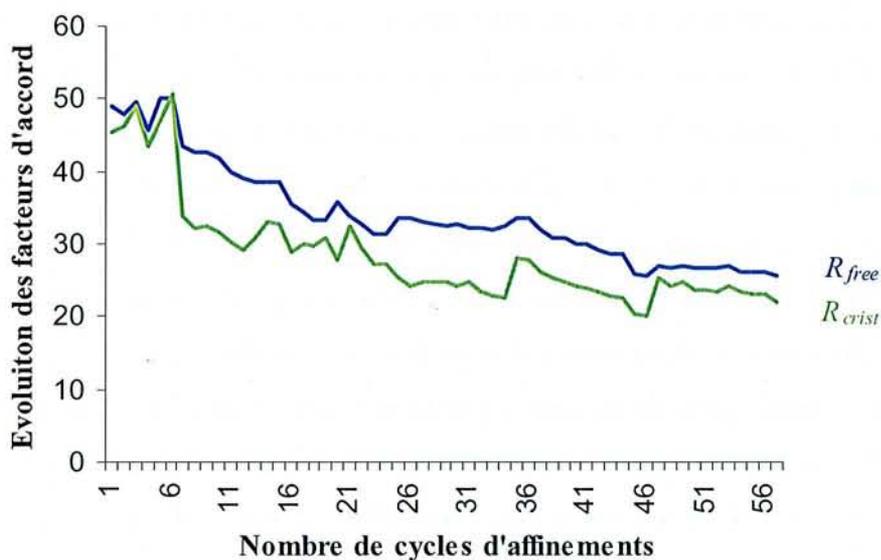
- Stratégie

La structure est reconstruite en utilisant le jeu de données enregistré au laboratoire à 2,2 Å puisque les facteurs d'accord calculés lors du remplacement moléculaire et lors des premières étapes de l'affinement étaient meilleurs avec ce jeu qu'avec celui enregistré au LURE à 1,8 Å de résolution. La symétrie non cristallographique est maintenue durant la

totalité de la reconstruction de la chaîne polypeptidique, c'est-à-dire qu'un seul monomère sur les quatre est reconstruit, les 3 autres étant générés par le calcul. Cette contrainte sera relâchée lorsque les facteurs d'accord se stabiliseront. Après affinement à 2,2 Å de résolution avec une troncature sur les données à 2σ et introduction des premières molécules d'eau, l'ensemble du jeu de données du LURE est utilisé pour mener l'affinement jusqu'à 1,8 Å de résolution sans troncature en σ . Les dernières molécules d'eau et les ions sont placés. La reconstruction a nécessité plusieurs étapes intermédiaires de minimisation d'énergie, en raison du nombre important de chaînes latérales à positionner, entrecoupées de quelques étapes d'affinement des facteurs d'agitation thermique et de dynamique moléculaire à 3000 K.

- Quelques remarques sur le suivi de l'affinement.

La qualité de l'affinement et du positionnement des acides aminés et de leurs chaînes latérales dans la carte de densité est évaluée par l'analyse des cartes de densité électronique et des facteurs d'accord R_{crist} et R_{free} dont les variations au cours de l'affinement sont représentées sur le graphique 5.1.



Graphique 5.1: évolution des facteurs d'accord au cours de l'affinement de la structure de l'apoenzyme (Apo1) à 1,8 Å de résolution.

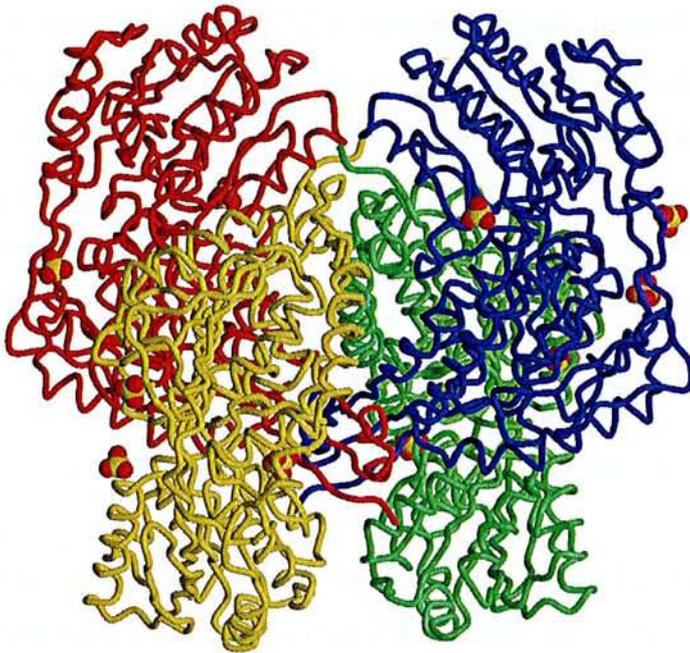


Figure 5.4: tétramère de l'ALDH de S. mutans en association avec les ions sulfate.

Au cours de l'affinement, un repositionnement du modèle dans la maille est effectué au moyen d'un affinement par corps rigide. Les chaînes latérales, pour lesquelles aucune densité n'apparaissait, sont orientées par rapport à leur environnement. Les erreurs sur les angles φ et ψ de la structure sont corrigées après analyse des cartes de Ramachandran. Ces légères corrections ont permis une diminution des facteurs d'accord. Puis, les molécules d'eau sont introduites, tout d'abord par l'utilisation du programme Arp (Lamzin & Wilson, 1993) puis par analyse de la carte de densité F_o-F_c à 2,2 Å puis à 1,8 Å. Dix ions sulfate sont introduits après examen de l'environnement électrostatique des sites présentant une densité tétraédrique pouvant être interprétée soit comme un ion sulfate, soit comme un ion phosphate (Figure 5.5). Compte tenu de la concentration des ions sulfate présents dans le milieu de cristallisation (2,0 M), ces ions sont choisis au lieu d'ions phosphate présents dans la solution de dialyse (50 mM). L'introduction des ions sulfate et des molécules d'eau a permis une amélioration des cartes de densité électronique et une diminution sensible des facteurs R_{crist} et R_{free} qui en fin de reconstruction ont des valeurs respectives de 22,1 % et 25,6 % sur l'ensemble des données comprises entre 8 Å et 1,8 Å. La structure a été déposée dans la Protein Data Bank sous le code 1EUH.

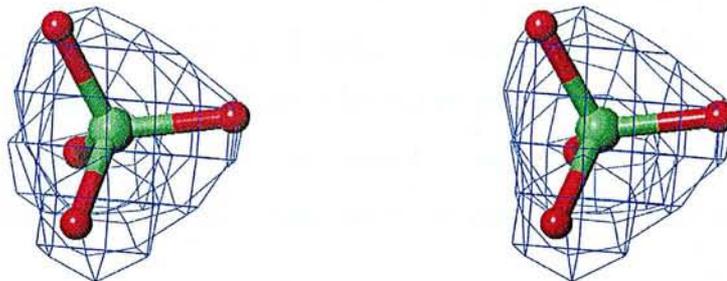


Figure 5.5: vue stéréoscopique d'un ion sulfate dans sa densité électronique. Carte $3F_o-2F_c$ contournée à $1,2 \sigma$ et calculée à $1,8 \text{ \AA}$ de résolution.

B- L'holoenzyme

1- Remplacement moléculaire

Le modèle utilisé pour résoudre la structure de l'holoenzyme est celui de l'apoenzyme sans molécule d'eau, ni ion sulfate. Les calculs sont effectués entre 9 et 6 Å. Le programme AMoRe (Navaza, 1994) trouve une solution unique présentant un très fort coefficient de corrélation et un faible facteur d'accord puisque les deux structures possèdent des repliements similaires.

2- Reconstruction et affinement

La stratégie de l'affinement et de la reconstruction de la structure dans sa densité électronique suit la procédure décrite précédemment.

L'analyse de la densité électronique montre qu'aucun changement conformationnel global n'intervient lors de la fixation du NADP et révèle dans tous les monomères la présence d'une densité inoccupée dans le domaine de fixation du cofacteur correspondant au NADP (Figure 5.6). Le cofacteur y est placé et les premiers cycles de reconstruction - affinement débutent avec l'application de la symétrie non cristallographique. Si aucun changement global

de la protéine n'est observé, il n'en est pas de même en ce qui concerne des réarrangements locaux intervenant lors de la fixation du cofacteur. Le plus important concerne une zone de la chaîne principale du domaine de Rossmann qui connecte le troisième brin β à la troisième hélice α . La réorientation de cette zone se traduit par une amplification de la courbure de l'axe d'une l'hélice α pour la stabilisation du 2'-phosphate du ribose de l'adénosine du NADP (Cobessi *et al.*, 1999).

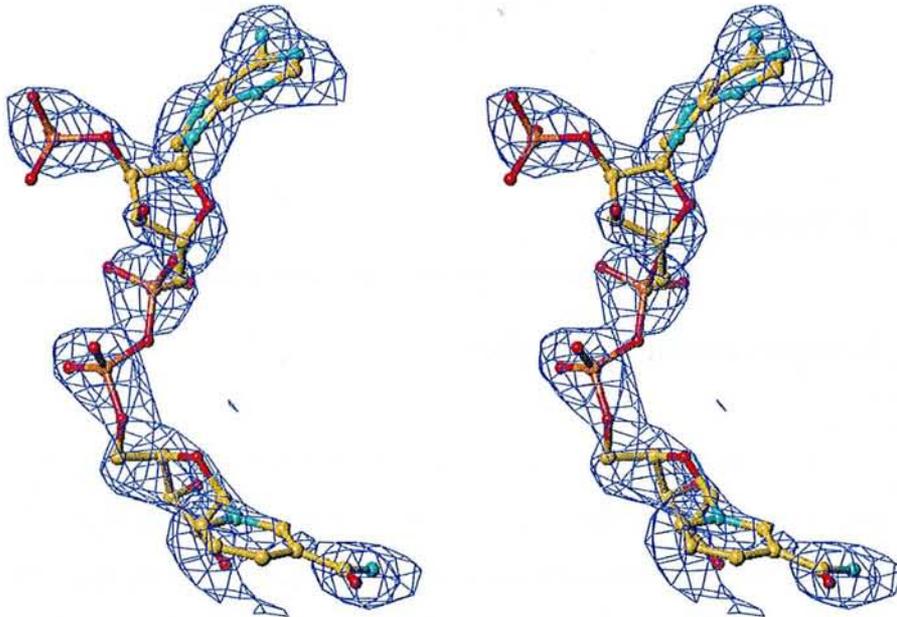


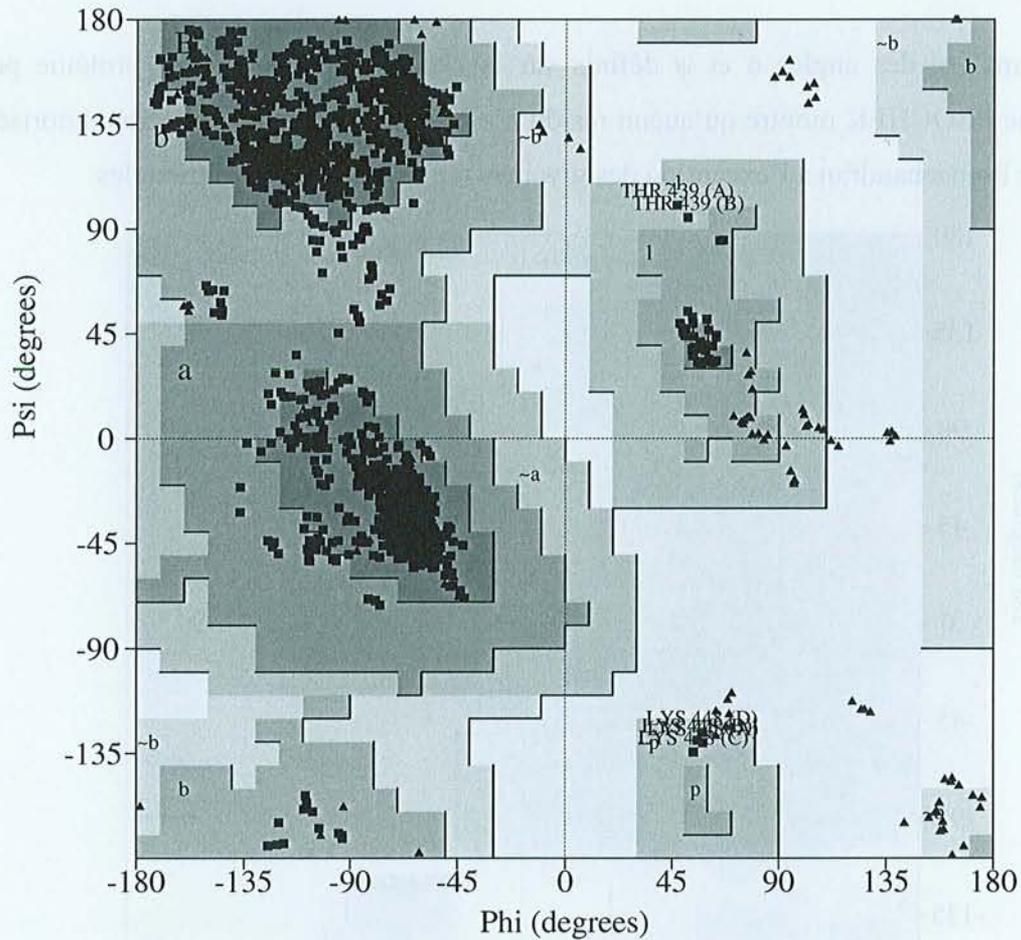
Figure 5.6: vue stéréoscopique de la densité électronique du NADP.

Carte $3Fo-2Fc$ calculée à 2,6 Å et contournée à 1,2 σ .

Au cours de la reconstruction, la symétrie non cristallographique n'a pas été relâchée. Les molécules d'eau et cinq ions sulfate sont introduits en fin d'affinement, le monomère A ne possédant aucun ion sulfate. Les valeurs respectives du R_{crist} et du R_{free} sont 23,5 % et 28,8 % sur toutes les données entre 8 et 2,6 Å de résolution. Cette structure a été déposée sous le code 2EUH dans la Protein Data Bank. Les valeurs du R_{crist} et du R_{free} sont respectivement de 25 % et 28,2 % pour la forme trigonale de l'holoenzyme dont les données ont été enregistrées à 3 Å de résolution au LURE.

C- Qualité des structures de l'Apo1 et de l'holoenzyme

La qualité géométrique des structures est évaluée par comparaison des paramètres structuraux de la molécule avec les paramètres standards des protéines. Cette évaluation est réalisée en utilisant les programmes GEOMANAL de X-PLOR et PROCHEK (Laskowsky *et al.*, 1993) de la chaîne de programmes CCP4.



Residues in most favoured regions [A,B,L]	1514	92.1%
Residues in additional allowed regions [a,b,l,p]	124	7.5%
Residues in generously allowed regions [-a,-b,-l,-p]	6	0.4%
Residues in disallowed regions	0	0.0%

Number of non-glycine and non-proline residues	1644	100.0%
Number of end-residues (excl. Gly and Pro)	8	
Number of glycine residues (shown as triangles)	168	
Number of proline residues	76	

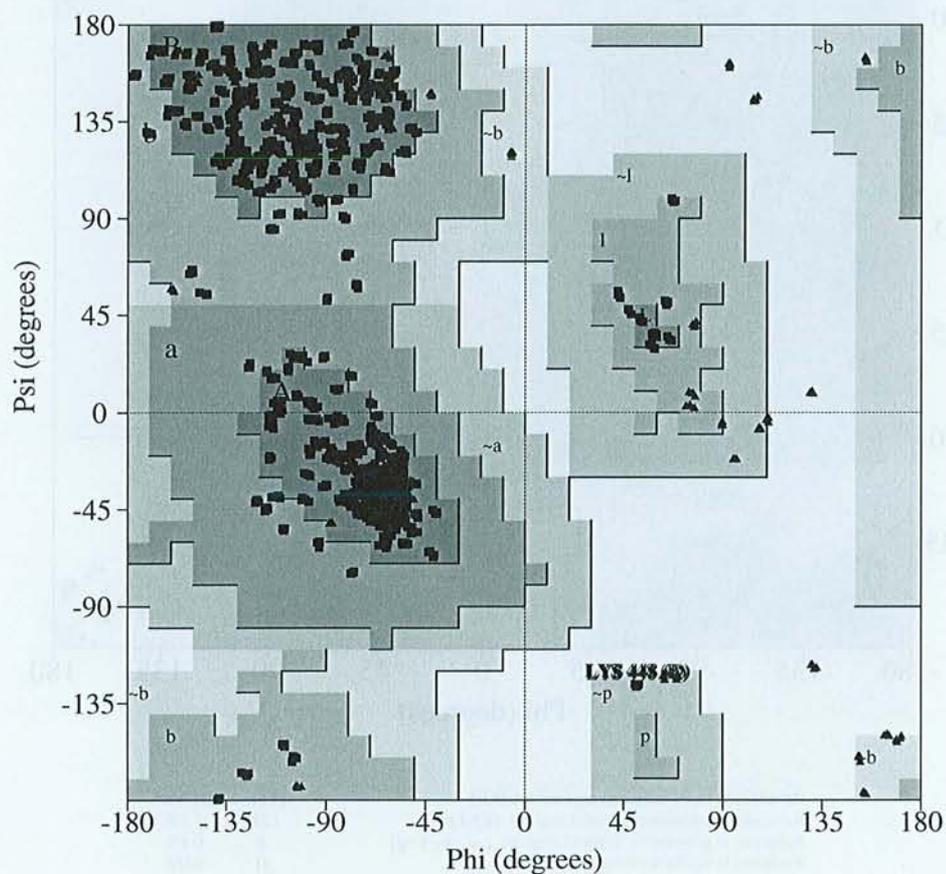
Total number of residues	1896	

Figure 5.7 : carte de Ramachandran de l'Apo1.

	Apo1	Holoenzyme
Écarts rms sur la longueur des liaisons (Å)	0,006	0,008
Écarts rms sur la valeur des angles de liaisons (°)	1,31	2,44
Angles dièdres impropres (°)	1,23	0,95
Angles dièdres (°)	23,30	25,40

Tableau 5.2: écarts rms sur différents paramètres structuraux pour les structures de l'Apo1 et de l'holoenzyme.

L'analyse des angles ϕ et ψ définis sur la chaîne principale de la protéine par le programme PROCHEK montre qu'aucun résidu n'est présent dans une zone non autorisée de la carte de Ramachandran à l'exception des glycines représentées par des triangles.



Residues in most favoured regions [A,B,L]	1515	92.2%
Residues in additional allowed regions [a,b,l,p]	125	7.6%
Residues in generously allowed regions [-a,-b,-l,-p]	4	0.2%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	1644	100.0%
Number of end-residues (excl. Gly and Pro)	8	
Number of glycine residues (shown as triangles)	168	
Number of proline residues	76	
Total number of residues	1896	

Figure 5.8: carte de Ramachandran de l'holoenzyme (forme orthorhombique).

D- L'empilement cristallin

Comme la structure de l'holoenzyme est résolue dans deux systèmes cristallins différents, l'empilement des tétramères dans les deux mailles est analysé. Les contacts intermoléculaires calculés à l'aide du programme GEOMANAL de X-PLOR sont présentés dans le tableau 5.3 pour la forme orthorhombique et 5.4 pour la forme trigonale. Les figures 5.9 et 5.10 montrent des vues stéréoscopiques de l'empilement cristallin de l'holoenzyme respectivement dans les systèmes orthorhombique et trigonal.

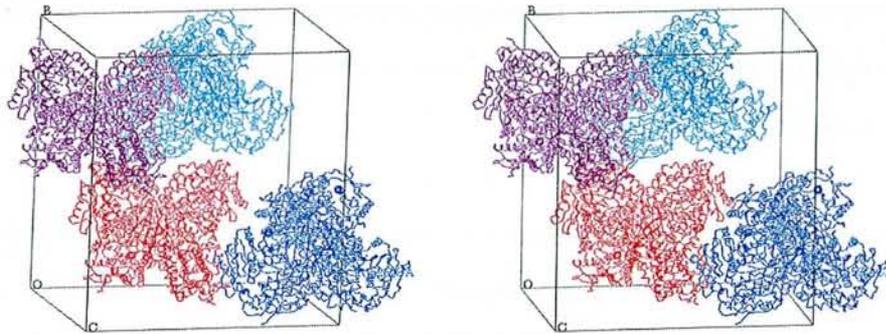


Figure 5.9: vue stéréoscopique de l'empilement cristallin des tétramères d'ALDH dans la maille orthorhombique.

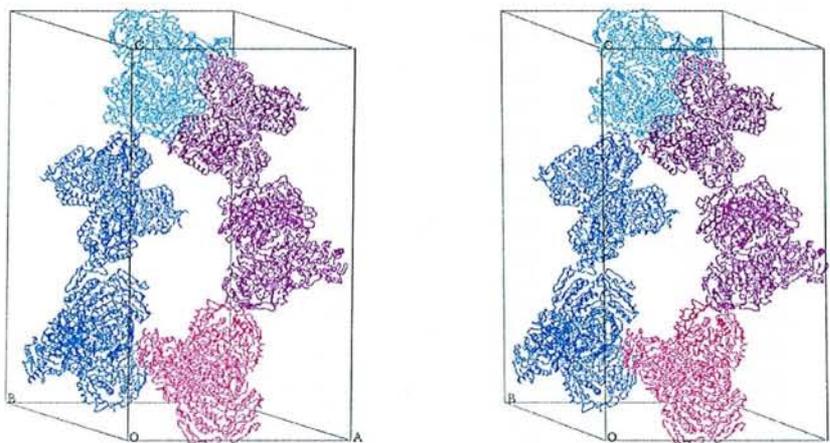


Figure 5.10: vue stéréoscopique de l'empilement cristallin des tétramères d'ALDH dans la maille trigonale.

Le nombre des contacts intermoléculaires assurant la cohésion cristalline est très inférieur dans la forme trigonale, ce qui est en accord avec le taux de solvant (75 %) qui y est plus élevé que dans la forme orthorhombique (60 %). De plus, la taille des canaux de solvant est plus importante dans le cas du système trigonal que dans le cas du système orthorhombique. L'analyse des interactions et du taux de solvant permet alors d'expliquer la stratégie des expériences de cryogénie lors de l'étape de trempage des cristaux dans la solution de cryoprotecteur. Un trempage doux pour la forme orthorhombique permet à l'éthylène glycol de diffuser doucement dans le cristal sans perturber, ou peu, le réseau cristallin. Dans le cas du système trigonal, où les canaux de solvant sont plus volumineux par rapport au système orthorhombique, un trempage brusque dans une solution d'éthylène glycol à 17,5 % n'entraîne aucune perturbation du réseau cristallin puisque le cryoprotecteur peut y diffuser très facilement.

Les interactions intermoléculaires assurant la cohésion cristalline sont résumées dans les tableaux suivants. Le premier répertorie les interactions intermoléculaires entre tétramères de l'holoenzyme dans le système orthorhombique, et le second dans le système trigonal.

Monomères et résidus	Monomères et résidus	Distances (Å)
MONA 77 -ARG -NE	MONC 317 -GLU -CB	3.57
MONA 77 -ARG -NE	MONC 317 -GLU -CG	3.86
MONA 77 -ARG -CZ	MONC 317 -GLU -CB	3.80
MONA 77 -ARG -CZ	MONC 317 -GLU -CG	3.94
MONA 77 -ARG -NH2	MONC 317 -GLU -CG	3.66
MONA 77 -ARG -O	MONC 88 -LYS -NZ	3.55
MONA 78 -ASP -CG	MONC 317 -GLU -OE2	3.79
MONA 78 -ASP -OD1	MONC 317 -GLU -CG	3.51
MONA 78 -ASP -OD1	MONC 317 -GLU -CD	3.57
MONA 78 -ASP -OD1	MONC 317 -GLU -OE2	3.31
MONA 80 -GLU -CB	MONC 27 -SER -OG	3.95
MONA 80 -GLU -CB	MONC 29 -ALA -CB	3.66
MONA 80 -GLU -CG	MONC 29 -ALA -CB	2.89
MONA 80 -GLU -CD	MONC 29 -ALA -CB	3.55
MONA 80 -GLU -OE2	MONC 29 -ALA -CB	3.79
MONA 81 -LYS -CD	MONC 315 -ASN -OD1	3.55
MONA 81 -LYS -CD	MONC 317 -GLU -OE2	3.45
MONC 77 -ARG -CG	MONC 77 -ARG -CG	3.51
MONC 77 -ARG -CG	MONC 77 -ARG -CD	3.57
MONC 77 -ARG -CD	MONC 77 -ARG -CD	3.18
MONC 77 -ARG -CD	MONC 77 -ARG -NE	3.22
MONC 77 -ARG -CD	MONC 77 -ARG -CZ	3.98
MONC 77 -ARG -CD	MONC 78 -ASP -OD1	3.99
MONC 77 -ARG -NE	MONC 77 -ARG -NE	3.30
MONC 77 -ARG -NE	MONC 77 -ARG -CZ	3.64
MONC 77 -ARG -NE	MONC 77 -ARG -NH1	3.86
MONC 77 -ARG -CZ	MONC 77 -ARG -CZ	3.53
MONC 77 -ARG -CZ	MONC 77 -ARG -NH1	3.68

MONC 77 -ARG -CZ	MONC 77 -ARG -NH2	3.99
MONC 77 -ARG -NH1	MONC 77 -ARG -NH2	3.71
MONC 80 -GLU -CB	MOND 81 -LYS -CG	3.82
MOND 80 -GLU -OE1	MOND 81 -LYS -CD	3.72
MOND 317 -GLU -OE2	MOND 317 -GLU -OE2	3.75
MONA 302 -GLU -CG	MOND 211 -SER -O	3.46
MONA 302 -GLU -CD	MOND 211 -SER -CB	3.73
MONA 302 -GLU -CD	MOND 211 -SER -OG	3.58
MONA 302 -GLU -CD	MOND 211 -SER -O	3.66
MONA 302 -GLU -OE1	MOND 211 -SER -CB	3.85
MONA 302 -GLU -OE2	MOND 211 -SER -CB	3.52
MONA 302 -GLU -OE2	MOND 211 -SER -OG	2.90
MONA 302 -GLU -OE2	MOND 211 -SER -O	3.56
MONA 305 -ARG -NH1	MOND 211 -SER -OG	3.38
MONA 306 -ARG -CG	MOND 212 -GLU -OE2	3.91
MONA 306 -GLU -CD	MOND 212 -GLU -CG	3.92
MONA 306 -GLU -CD	MOND 212 -GLU -CD	3.75
MONA 306 -GLU -CD	MOND 212 -GLU -OE2	3.31
MONA 306 -GLU -OE1	MOND 212 -GLU -CD	3.80
MONA 306 -GLU -OE1	MOND 212 -GLU -OE2	3.46
MONA 306 -GLU -OE2	MOND 212 -GLU -CG	3.20
MONA 306 -GLU -OE2	MOND 212 -GLU -CD	3.52
MONA 306 -GLU -OE2	MOND 212 -GLU -OE2	3.39
MONA 342 -ASN -OD1	MOND 332 -ASP -O	3.63
MONA 342 -ASN -O	MOND 372 -ARG -CZ	3.85
MONA 342 -ASN -O	MOND 372 -ARG -NH1	3.13
MONA 343 -ASP -CA	MOND 372 -ARG -NH1	3.71
MONA 343 -ASP -CG	MOND 339 -ASN -CB	3.97
MONA 343 -ASP -OD1	MOND 336 -GLY -CA	3.72
MONA 343 -ASP -OD1	MOND 336 -GLY -O	3.84
MONA 343 -ASP -OD1	MOND 339 -ASN -CB	3.50
MONA 343 -ASP -C	MOND 372 -ARG -NH1	3.62
MONA 343 -ASP -O	MOND 340 -ASP -CB	3.93
MONA 343 -ASP -O	MOND 340 -ASP -OD2	3.83
MONA 343 -ASP -O	MOND 372 -ARG -CD	3.32
MONA 343 -ASP -O	MOND 372 -ARG -NH1	3.28
MONA 348 -ALA -O	MOND 329 -LYS -CE	3.89
MONA 349 -LEU -CA	MOND 329 -LYS -CE	3.93
MONA 349 -LEU -CA	MOND 329 -LYS -NZ	3.89
MONA 349 -LEU -C	MOND 329 -LYS -CE	3.73
MONA 349 -LEU -C	MOND 329 -LYS -NZ	3.70
MONA 349 -LEU -O	MOND 329 -LYS -CD	3.83
MONA 349 -LEU -O	MOND 329 -LYS -CE	3.55
MONA 349 -LEU -O	MOND 329 -LYS -NZ	3.15
MONA 351 -GLU -CA	MOND 332 -ASP -OD2	3.90
MONA 351 -GLU -CB	MOND 328 -THR -CG2	3.65
MONA 351 -GLU -CG	MOND 328 -THR -CG2	3.44
MONA 351 -GLU -CD	MOND 328 -THR -CG2	3.57
MONA 351 -GLU -OE2	MOND 328 -THR -CB	3.88
MONA 351 -GLU -OE2	MOND 328 -THR -CG2	3.17
MONA 351 -GLU -C	MOND 332 -ASP -OD2	3.90
MONA 352 -ILE -N	MOND 332 -ASP -OD2	2.96
MONA 352 -ILE -CA	MOND 332 -ASP -OD2	3.73
MONA 352 -ILE -CB	MOND 332 -ASP -OD2	3.57
MONA 352 -ILE -CG1	MOND 332 -ASP -OD2	3.94
MONA 352 -ILE -O	MOND 332 -ASP -OD2	3.71
MONB 211 -SER -CB	MONC 302 -GLU -OE2	3.87
MONB 211 -SER -CB	MONC 305 -ARG -NH1	3.83
MONB 211 -SER -OG	MONC 302 -GLU -OE2	3.50
MONB 211 -SER -OG	MONC 305 -ARG -CZ	3.84
MONB 211 -SER -OG	MONC 305 -ARG -NH1	3.27

MONB 211 -SER -C	MONC 302 -GLU -OE2	3.91
MONB 211 -SER -O	MONC 302 -GLU -CG	3.61
MONB 211 -SER -O	MONC 302 -GLU -CD	3.38
MONB 211 -SER -O	MONC 302 -GLU -OE2	3.11
MONB 212 -GLU -CG	MONC 306 -GLU -OE2	3.73
MONB 328 -THR -CG2	MONC 351 -GLU -OE2	3.80
MONB 329 -LYS -CE	MONC 348 -ALA -O	3.80
MONB 329 -LYS -NZ	MONC 349 -LEU -C	3.81
MONB 329 -LYS -NZ	MONC 349 -LEU -O	3.39
MONB 332 -ASP -OD2	MONC 352 -ILE -CB	3.73
MONB 332 -ASP -OD2	MONC 352 -ILE -O	3.82
MONB 332 -ASP -O	MONC 342 -ASN -OD1	3.63
MONB 336 -GLY -CA	MONC 343 -ASP -OD1	3.42
MONB 336 -GLY -O	MONC 343 -ASP -OD1	3.82
MONB 339 -ASN -CB	MONC 343 -ASP -OD1	3.76
MONB 340 -ASP -CB	MONC 343 -ASP -O	3.94
MONB 372 -ARG -CD	MONC 343 -ASP -O	3.07
MONB 372 -ARG -NE	MONC 343 -ASP -O	3.69
MONB 372 -ARG -NE	MONC 345 -GLY -CA	3.98
MONB 372 -ARG -CZ	MONC 342 -ASN -O	3.76
MONB 372 -ARG -CZ	MONC 343 -ASP -O	3.56
MONB 372 -ARG -CZ	MONC 345 -GLY -N	3.89
MONB 372 -ARG -CZ	MONC 345 -GLY -CA	3.94
MONB 372 -ARG -NH1	MONC 342 -ASN -O	3.17
MONB 372 -ARG -NH1	MONC 343 -ASP -CA	3.43
MONB 372 -ARG -NH1	MONC 343 -ASP -C	3.23
MONB 372 -ARG -NH1	MONC 343 -ASP -O	2.80
MONB 372 -ARG -NH1	MONC 345 -GLY -N	3.99

Tableau 5.3: interactions intermoléculaires entre les tétramères de l'holoenzyme dans le système orthorhombique.

Monomères et résidus	Monomères et résidus	Distances (Å)
MONB 26 -ALA -O	MONC 30 -GLU -N	3.48
MONB 26 -ALA -O	MONC 30 -GLU -CB	3.48
MONB 27 -SER -C	MONC 29 -ALA -CB	3.77
MONB 27 -SER -C	MONC 30 -GLU -O	3.92
MONB 27 -SER -O	MONC 29 -ALA -CB	3.51
MONB 27 -SER -O	MONC 30 -GLU -O	3.08
MONB 28 -GLY -N	MONC 29 -ALA -CB	3.93
MONB 28 -GLY -CA	MONC 29 -ALA -CB	3.83
MONB 315 -ASN -ND2	MONC 21 -LYS -NZ	3.61
MONB 328 -THR -CB	MONC 27 -SER -O	3.97
MONB 328 -THR -OG1	MONC 27 -SER -CA	3.67
MONB 328 -THR -OG1	MONC 27 -SER -C	3.26
MONB 328 -THR -OG1	MONC 27 -SER -O	2.54
MONB 332 -ASP -CG	MONC 356 -GLY -CA	3.71
MONB 332 -ASP -OD2	MONC 356 -GLY -CA	3.24
MONB 355 -GLU -CB	MONC 23 -TYR -CD1	3.82
MONB 355 -GLU -CB	MONC 23 -TYR -CE1	3.28
MONB 355 -GLU -CB	MONC 23 -TYR -CZ	3.83
MONB 355 -GLU -CG	MONC 23 -TYR -CE1	3.90
MONB 355 -GLU -C	MONC 23 -TYR -CD1	3.98
MONB 356 -GLY -N	MONC 23 -TYR -CD1	3.93
MONB 356 -GLY -CA	MONC 23 -TYR -CB	3.89

MONB 356 -GLY -CA	MONC 28 -GLY -CA	3.34
MONB 356 -GLY -CA	MONC 28 -GLY -C	3.91
MONB 356 -GLY -CA	MONC 28 -GLY -O	3.55
MONB 356 -GLY -C	MONC 28 -GLY -CA	3.82
MONB 357 -ASN -N	MONC 28 -GLY -CA	3.93
MONB 357 -ASN -CG	MONC 27 -SER -O	3.48
MONB 357 -ASN -OD1	MONC 27 -SER -O	3.48
MONB 357 -ASN -OD1	MONC 28 -GLY -CA	3.98
MONB 357 -ASN -ND2	MONC 27 -SER -O	3.04
MONA 343 -ASP -CB	MOND 51 -LYS -NZ	3.52
MONA 343 -ASP -CG	MOND 51 -LYS -CE	3.80
MONA 343 -ASP -CG	MOND 51 -LYS -NZ	2.58
MONA 343 -ASP -OD1	MOND 51 -LYS -CE	3.12
MONA 343 -ASP -OD1	MOND 51 -LYS -NZ	2.27
MONA 343 -ASP -OD2	MOND 51 -LYS -NZ	2.95
MONB 305 -ARG -CZ	MOND 261 -GLU -O	3.42
MONB 305 -ARG -NH2	MOND 261 -GLU -C	3.59
MONB 305 -ARG -NH2	MOND 261 -GLU -O	2.51
MONB 305 -ARG -NH2	MOND 262 -ASP -CA	3.84
MONB 309 -LEU -CD1	MOND 261 -GLU -CB	3.39
MONB 309 -LEU -CD1	MOND 261 -GLU -C	3.64
MONB 309 -LEU -CD1	MOND 261 -GLU -O	3.70
MONB 309 -LEU -CD1	MOND 262 -ASP -N	3.81
MONB 309 -LEU -CD2	MOND 261 -GLU -CB	3.97
MONB 309 -LEU -CD2	MOND 261 -GLU -CD	3.85
MONB 309 -LEU -CD2	MOND 261 -GLU -OE2	3.67
MONB 349 -LEU -O	MOND 295 -SER -OG	3.36
MONB 351 -GLU -N	MOND 295 -SER -CB	3.85
MONB 351 -GLU -OE1	MOND 389 -THR -CB	3.94
MONB 351 -GLU -OE1	MOND 389 -THR -O	3.81
MONC 10 -ASN -CG	MONC 44 -TYR -OH	3.80
MONC 10 -ASN -ND2	MONC 44 -TYR -OH	3.30
MONC 14 -LYS -CD	MONC 51 -LYS -NZ	3.22
MONC 14 -LYS -CE	MONC 51 -LYS -NZ	3.34
MONC 14 -LYS -NZ	MONC 51 -LYS -CE	3.82
MONC 14 -LYS -NZ	MONC 51 -LYS -NZ	3.17
MONC 44 -TYR -CA	MONC 47 -ALA -CB	3.94
MONC 44 -TYR -CB	MONC 51 -LYS -CD	3.51
MONC 44 -TYR -CG	MONC 51 -LYS -CD	3.55
MONC 44 -TYR -CG	MONC 51 -LYS -CE	3.78
MONC 44 -TYR -CG	MONC 51 -LYS -NZ	3.86
MONC 44 -TYR -CD1	MONC 44 -TYR -CE1	3.98
MONC 44 -TYR -CD1	MONC 48 -SER -N	3.90
MONC 44 -TYR -CD1	MONC 48 -SER -CA	3.81
MONC 44 -TYR -CD1	MONC 48 -SER -OG	3.59
MONC 44 -TYR -CE1	MONC 44 -TYR -CE1	3.25
MONC 44 -TYR -CE1	MONC 44 -TYR -CZ	3.69
MONC 44 -TYR -CE1	MONC 44 -TYR -OH	3.79
MONC 44 -TYR -CE1	MONC 48 -SER -CA	3.91
MONC 44 -TYR -CE1	MONC 48 -SER -CB	3.95
MONC 44 -TYR -CE1	MONC 48 -SER -OG	2.93
MONC 44 -TYR -CD2	MONC 51 -LYS -CD	3.76
MONC 44 -TYR -CD2	MONC 51 -LYS -CE	3.54
MONC 44 -TYR -CD2	MONC 51 -LYS -NZ	3.14
MONC 44 -TYR -CE2	MONC 51 -LYS -CE	3.82
MONC 44 -TYR -CE2	MONC 51 -LYS -NZ	3.37
MONC 44 -TYR -CZ	MONC 44 -TYR -CZ	3.97
MONC 44 -TYR -CZ	MONC 44 -TYR -OH	3.62
MONC 44 -TYR -CZ	MONC 48 -SER -OG	3.98
MONC 44 -TYR -OH	MONC 44 -TYR -OH	2.83
MONC 47 -ALA -CB	MONC 47 -ALA -CB	3.62

MONC 237 -ARG -NH2	MOND 370 -ASP -OD1	3.65
MONC 237 -ARG -NH2	MOND 370 -ASP -OD2	3.84

Tableau 5.4: interactions intermoléculaires entre les tétramères de l'holoenzyme dans le système trigonal.

II- Résolution de la structure de l'Apo2

A- Remplacement moléculaire et affinement

1- Remplacement moléculaire

Le modèle utilisé est la structure de l'Apo1 sans molécule d'eau, ni ion sulfate. Le remplacement moléculaire est réalisé sur les deux jeux de données, enregistrés respectivement à 3 et 2,5 Å de résolution, en utilisant AMoRe. Le calcul des fonctions de rotation croisée et de translation est réalisé entre 6 et 9 Å. Seule la structure résolue à 2,5 Å sera affinée.

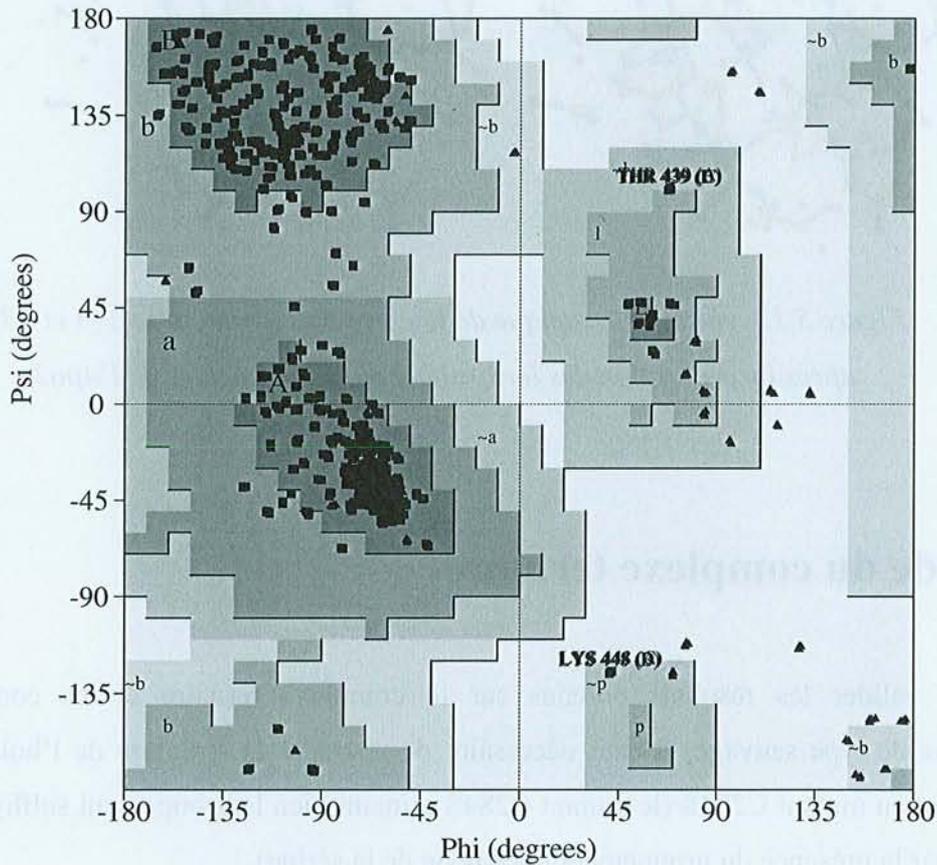
2- Affinement

D'abord effectué en corps rigide puis par minimisations d'énergie entrecoupées de reconstructions graphiques et de dynamique moléculaire, l'affinement s'achève à des valeurs de R_{free} de 28,0 % et de R_{crist} de 23,2 % sur toutes les réflexions comprises entre 8 et 2,5 Å. Durant tout l'affinement, la symétrie non cristallographique a été maintenue et 603 molécules d'eau ont été incluses dans le modèle.

L'analyse de la carte de densité électronique de type $F_{\text{obs}}-F_{\text{calc}}$ montre la présence de densités positives de plus de 3 σ supérieures à la moyenne dans le site catalytique et à proximité du site de fixation du 2'-phosphate du NADP. Ces pics sont interprétés comme des ions sulfate en raison de la présence du sulfate d'ammonium dans la solution de cristallisation. L'affinement de leurs positions couplé au calcul et à l'analyse de ces mêmes cartes révèlent des taux d'occupation différents en fonction des monomères allant de 0,5 pour 8 ions à 1 pour les 4 autres.

3- Qualité de la structure

Elle est évaluée par l'utilisation des programmes X-PLOR et PROCHECK. Le diagramme de Ramachandran de la structure est présenté ci-dessous.



Residues in most favoured regions [A,B,L]	1512	92.0%
Residues in additional allowed regions [a,b,l,p]	124	7.5%
Residues in generously allowed regions [-a,-b,-l,-p]	8	0.5%
Residues in disallowed regions	0	0.0%

Number of non-glycine and non-proline residues	1644	100.0%
Number of end-residues (excl. Gly and Pro)	8	
Number of glycine residues (shown as triangles)	168	
Number of proline residues	76	

Total number of residues	1896	

Figure 5.11: diagramme de Ramachandran de l'Apo2.

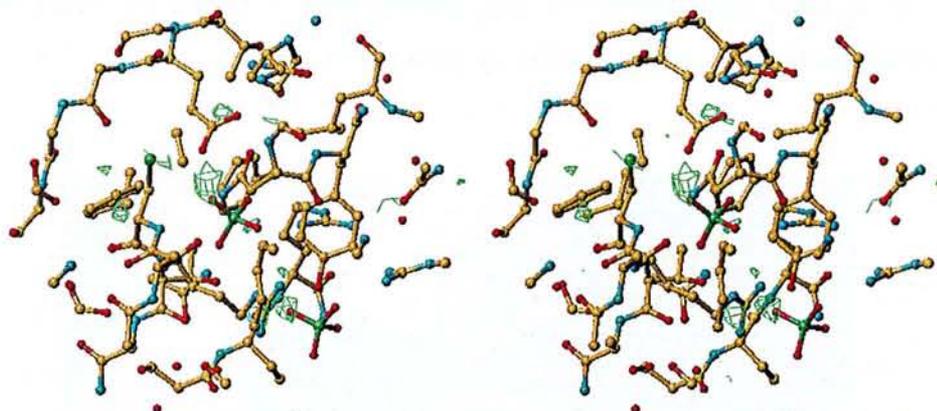


Figure 5.12: vue stéréoscopique de la carte F_o-F_c contourée à -3 et $+3 \sigma$ après incorporation des ions sulfate dans la structure de l'Apo2.

III- Etude du complexe ternaire

Pour valider les résultats obtenus sur le complexe ternaire et les comparer à l'holoenzyme de type sauvage, il était nécessaire de résoudre la structure de l'holoenzyme correspondant au mutant C284S (le mutant C284S mimant bien le groupement sulfhydryle de la cystéine par la présence du groupement hydroxyle de la sérine).

A- L'holoenzyme C284S

Sa structure a été résolue à 3,15 Å de résolution. Les facteurs d'accord sont respectivement de 22,8 % pour le R_{crist} et 27,3 % pour le R_{free} . L'analyse de cette structure montre la sérine 284 en densité et le cofacteur dans une densité électronique de qualité médiocre sur toute la partie nicotinamide mononucléoside.

B- Le complexe ternaire

Les phases sont calculées à partir du modèle de l'holoenzyme de type sauvage dont la structure a été résolue dans le groupe d'espace $P3_221$. Le modèle est d'abord replacé dans sa maille trigonale par remplacement moléculaire puis sa position est affinée par des cycles d'affinement par corps rigide suivis d'une minimisation d'énergie et d'une dynamique moléculaire à 3 Å. Des cartes de densité électronique de type $3F_o-2F_c$ et F_o-F_c sont calculées. Il est possible d'y observer la densité électronique du cofacteur (de qualité médiocre au niveau de la partie nicotinamide mononucléoside), du substrat et des chaînes latérales du site catalytique qui sont déplacées par rapport à l'holoenzyme de type sauvage. Il s'agit notamment des chaînes latérales du glutamate Glu250 et de l'arginine Arg437. Un cycle de minimisation d'énergie est réalisé après le repositionnement des chaînes latérales dans la densité électronique. Le substrat est placé dans sa densité. Une nouvelle minimisation d'énergie suivie d'un calcul de cartes de densité électronique permet d'attribuer un taux d'occupation de 0,5 au G3P. Il en sera de même pour le NADP dont la conformation est différente de celle observée dans la structure du mutant C284S. Après positionnement de ces deux molécules, un cycle de minimisation d'énergie est réalisé, suivi d'un affinement des facteurs d'agitation thermique en les contraignant par chaînes, puis d'une dynamique moléculaire. Ces affinements permettent d'amener les facteurs d'accord cristallographique à 24,2 % pour le R_{cris} et 28,3 % pour le R_{free} à 3 Å de résolution.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

Analyses structurales et discussion

I- L'Apo1 et l'holoenzyme

Une comparaison de la structure de Sm-ALDH avec les premières structures d'ALDH résolues montre que malgré un faible pourcentage d'identité de séquences, ces protéines adoptent un repliement similaire avec de très faibles valeurs rms entre leurs coordonnées atomiques (Cobessi *et al.*, 1999).

A-Description globale de la structure de Sm-ALDH

Le tétramère de l'ALDH de *S. mutans* (Figure 6.1) peut être décrit comme une association de deux dimères, les monomères étant reliés les uns aux autres par trois axes 2 non cristallographiques. Le monomère peut être divisé en trois domaines: un premier domaine s'étendant des acides aminés Thr2 à Gly118 puis de Gly145 à Gly252 et de Ser450 à Met464 représentant le domaine de fixation du cofacteur. Dans ce premier domaine se trouve le repliement de Rossmann partiellement ceinturé par 4 brins β et 5 hélices α . Le second domaine, constitué d'hélices α et de brins β , s'étend de Gly253 à Lys449 et constitue le domaine catalytique. Enfin un dernier domaine composé d'un feuillet β antiparallèle constitue le domaine d'oligomérisation. Il comprend les résidus Glu119 à Val144 et Thr465 à Lys475.

Concernant les acides aminés impliqués dans la catalyse, la distance entre le carboxylate du glutamate Glu250 et l'atome de soufre de la cystéine essentielle Cys284 est de 6,9 Å. L'ensemble de la topologie du monomère est décrite dans le tableau 6.1

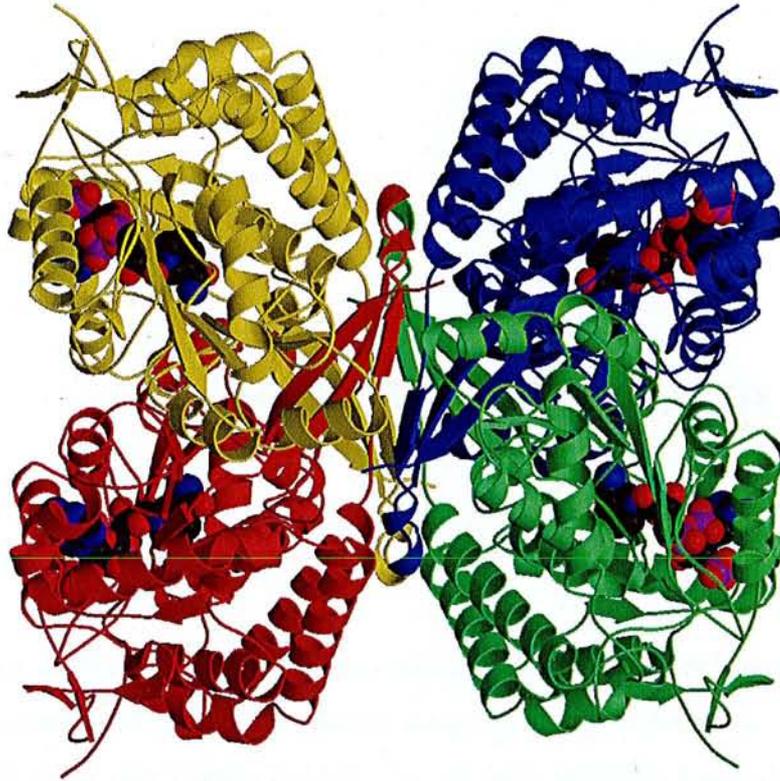


Figure 6.1: tétramère de l'holoenzyme de l'ALDH de *S. mutans* en interaction avec le NADP.

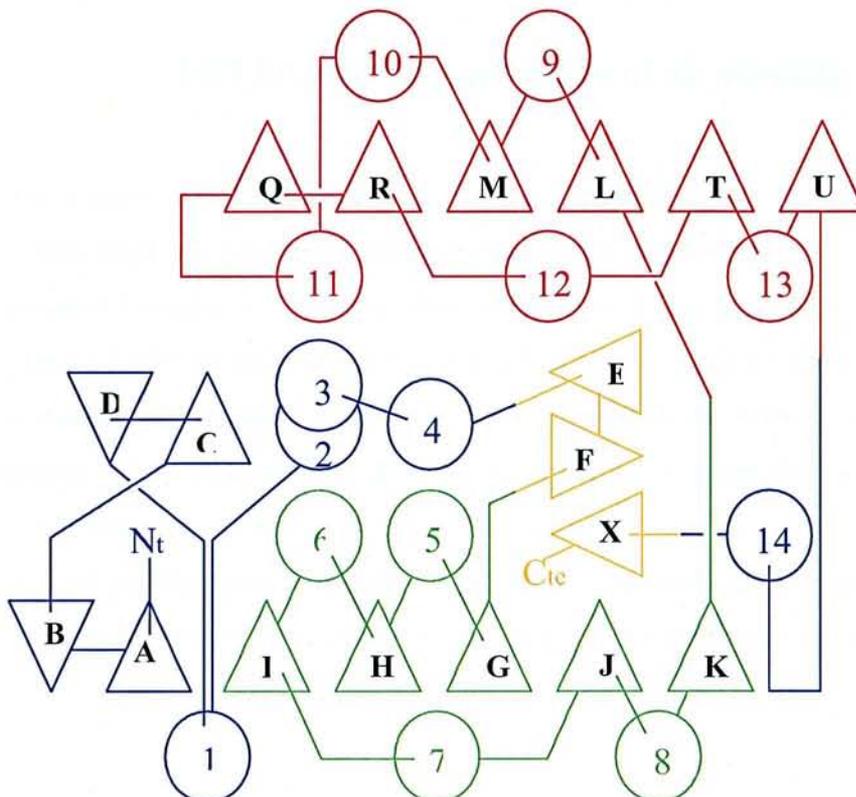


Figure 6.2: topologie du monomère de l'ALDH de *S. mutans*.

Eléments de structure secondaire	Résidus	Eléments de structure secondaire	Résidus
Brin β_A	Asn7 - Val9	Brin β_J	Phe225 - Thr229
Brin β_B	Glu12 - Lys14	Hélice α_8	Thr232 - Met241
Brin β_C	Glu19 - Tyr23	Brin β_K	Ile247 - Glu250
Brin β_D	Gly32 - Pro35	Brin β_L	Asp255 - Val259
Hélice α_1	Thr39 - Ala58	Hélice α_9	Leu265 - Gly281
Hélice α_2	Tyr61 - Arg77	Brin β_M	Arg289 - Leu291
Hélice α_3	Lys79 - Val90	Hélice α_{10}	Glu294 - Leu309
Hélice α_4	Gly115 - Glu119	Hélice α_{11}	Thr328 - Asp343
Brin β_E	Glu121 - Glu124	Brin β_Q	Ile362 - Asp365
Brin β_F	Lys135 - Pro143	Brin β_R	Val382 - Thr388
Brin β_G	Leu146 - Glu150	Hélice α_{12}	Val391 - Lys400
Hélice α_5	Asn158 - Ile169	Brin β_T	Gln406 - Phe410
Brin β_H	Ile173 - Lys177	Hélice α_{13}	Phe414 - Gln423
Hélice α_6	Ser183 - Ala196	Brin β_U	Thr428 - Ile431
Brin β_I	Phe203 - Thr205	Hélice α_{14}	Val457 - Ala463
Hélice α_7	Gly210 - Glu219	Brin β_X	Thr465 - Asp473

Tableau 6.1: topologie du monomère de l'ALDH de *S. mutans*.

Au sein du tétramère, les interactions entre les monomères impliquent principalement les domaines d'oligomérisation qui interagissent non seulement entre eux mais aussi avec d'autres régions du domaine catalytique et du domaine du cofacteur. Ces interactions sont des liaisons hydrogène, des ponts salins ou des contacts de Van der Waals. Les liaisons hydrogène entre les résidus se font de façon directe, ou bien *via* une ou plusieurs molécules d'eau. La liste des liaisons hydrogène et des ponts salins intermoléculaires (Tableau 6.2) a été obtenue en utilisant le programme CONTACT (Skarzynski, 1988) de la chaîne de programmes CCP4.

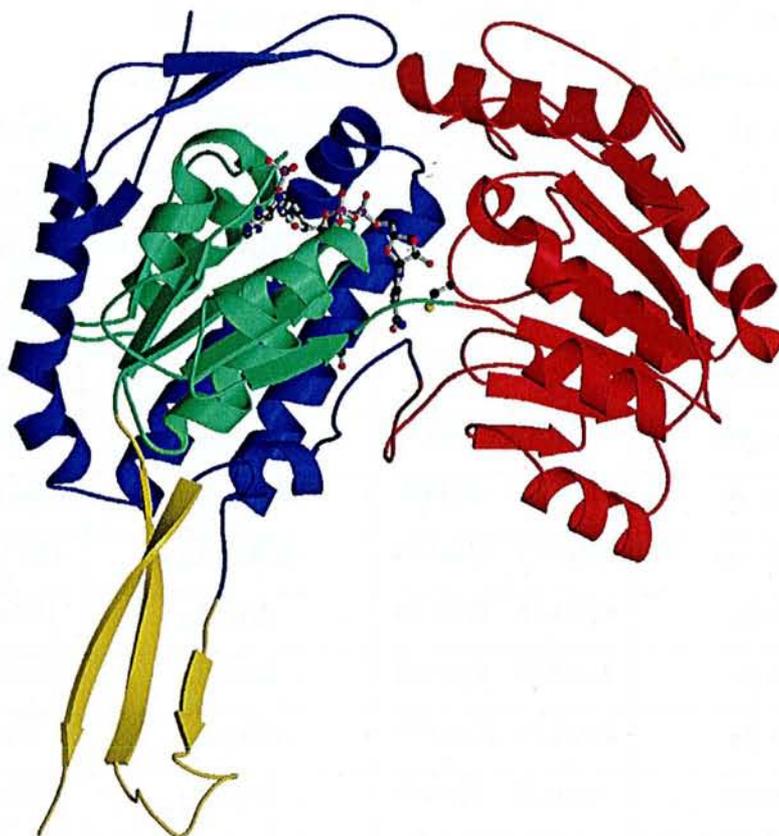


Figure 6.3: monomère de l'holoenzyme de Sm-ALDH et ses différents domaines: bleu et vert, domaine de fixation du cofacteur (le domaine de Rossmann est coloré en vert); rouge, domaine catalytique; jaune, domaine d'oligomérisation. Les résidus impliqués dans l'acte catalytique (Glu250 et Cys284) et le NADP sont représentés en mode sphères et bâtons.

Résidus monomère A			Résidus en contact			distance	angle
ARG	A57	NH1	GLU	B422	O	3.27	136.2
ARG	A57	NH2	GLU	B422	O	2.98	123.0
ALA	A58	O	LYS	D133	NZ	3.07	0.0
TYR	A61	N	GLY	D126	O	3.50	115.7
ILE	A62	N	GLY	D126	O	3.08	158.9
TYR	A110	O	ARG	C117	NH2	2.97	123.5
GLU	A114	OE1	ARG	C117	NE	2.79	140.6
LEU	A116	O	SER	D127	OG	2.64	99.1
ARG	A117	NE	GLU	C114	OE1	2.76	120.7
ARG	A117	NH2	TYR	C110	O	2.90	121.7
GLU	A119	OE1	LYS	C458	NZ	2.92	0.0
			GLU	D121	OE1	2.73	126.2
GLU	A119	OE2	LYS	C458	NZ	2.75	0.0
GLY	A120	N	VAL	D122	O	2.94	146.3
GLY	A120	O	VAL	D122	N	3.00	167.8
GLU	A121	OE1	LYS	B458	NZ	3.02	0.0
			GLU	D119	OE1	3.29	124.8
GLU	A121	OE2	TYR	B459	OH	2.73	116.9
VAL	A122	N	GLY	D120	O	3.02	97.6
VAL	A122	O	GLY	D120	N	2.82	164.1
GLU	A124	OE1	ARG	D140	NH2	3.07	162.1
GLU	A124	OE2	ARG	D140	NH2	3.08	125.1

GLU	A124	O	ASN	B441	ND2	3.01	136.5
GLY	A126	O	ILE	D62	N	3.14	154.2
SER	A127	OG	LEU	D116	O	2.73	154.3
LYS	A133	NZ	ALA	D58	O	2.89	139.2
LYS	A134	NZ	GLU	C422	OE1	2.94	122.5
ARG	A140	NE	GLU	D124	OE1	2.78	132.3
ARG	A140	NH1	GLU	B422	OE2	2.83	132.9
ARG	A140	NH2	GLU	B422	OE2	3.26	166.7
			GLU	D124	OE1	3.24	94.4
			GLU	D124	OE2	3.07	102.6
PRO	A415	O	LYS	C475	N	3.09	159.1
ALA	A421	O	LYS	B468	NZ	2.70	0.0
GLU	A422	OE2	ARG	B140	NH1	2.95	151.6
			ARG	B140	NH2	3.49	130.5
GLU	A422	O	ARG	B57	NH1	3.37	133.8
			ARG	B57	NH2	3.01	148.4
			LYS	B468	NZ	3.29	0.0
LEU	A424	O	LYS	B468	NZ	2.84	0.0
GLY	A427	O	SER	B469	N	2.97	176.9
VAL	A429	N	SER	B469	O	2.72	168.6
VAL	A429	O	VAL	B471	N	2.83	163.2
ILE	A431	N	VAL	B471	O	2.88	173.2
ASN	A441	ND2	GLU	B124	O	2.95	131.1
LYS	A458	NZ	GLU	B121	OE2	3.38	140.3
			GLU	C119	OE2	3.05	134.1
TYR	A459	OH	GLU	B121	OE2	2.85	141.1
LYS	A468	NZ	LEU	B424	O	2.77	153.0
			ALA	B421	O	2.75	141.5
			GLU	B422	O	3.17	99.1
SER	A469	N	GLY	B427	O	2.93	161.3
SER	A469	O	VAL	B429	N	2.76	167.1
VAL	A471	N	VAL	B429	O	2.92	160.3
VAL	A471	O	ILE	B431	N	3.00	170.9
ASP	A473	OD2	ASN	B433	ND2	3.00	152.3
LYS	A475	N	PRO	C415	O	3.07	107.9

Tab 6.2: liste des interactions par liaisons hydrogène et ponts salins entre les différents monomères de l'ALDH de *S. mutans* en prenant comme référence le monomère A.

B- Le domaine de Rossmann

Le domaine de Rossmann (Figures 1.3; 1.4) est constitué de brins β et d'hélices α . C'est un domaine commun à de nombreuses déshydrogénases chez lesquelles il interagit avec les cofacteurs à nucléotide. Son rôle a été mis en évidence par Rossmann *et al.* (1974) à travers l'étude structurale de 4 déshydrogénases: la malate déshydrogénase, la lactate déshydrogénase, la glycéraldéhyde-3-phosphate déshydrogénase et l'alcool déshydrogénase.

1- Description structurale et caractéristiques générales

Le domaine de Rossmann est constitué d'un feuillet β parallèle central clôturé de part et d'autre par des hélices α . Le grand nombre de structures de déshydrogénases montre que le nombre d'hélices α et de brins β est très variable. Les alignements de séquences à travers les différentes familles de déshydrogénases possédant ce repliement révèlent une empreinte générale caractéristique. Elle se caractérise par la succession d'acides aminés GXGXXG (Wierenga *et al.*, 1986). Structuralement, elle se situe dans la boucle reliant le premier brin β à la première hélice α . Cette boucle interagit en général avec le groupement pyrophosphate des cofacteurs NAD ou NADP se fixant au domaine de Rossmann. Chez les déshydrogénases autres que les ALDH, et possédant un domaine de Rossmann, le cofacteur se fixe du côté C-terminal des brins β , et ses nucléosides se situent de part et d'autre du feuillet β . La stabilisation du pyrophosphate est favorisée par son interaction avec le dipôle de la première hélice α et dans certains cas par la présence de résidus chargés positivement. Ces cofacteurs adoptent une conformation étendue lors de leur fixation au domaine de Rossmann.

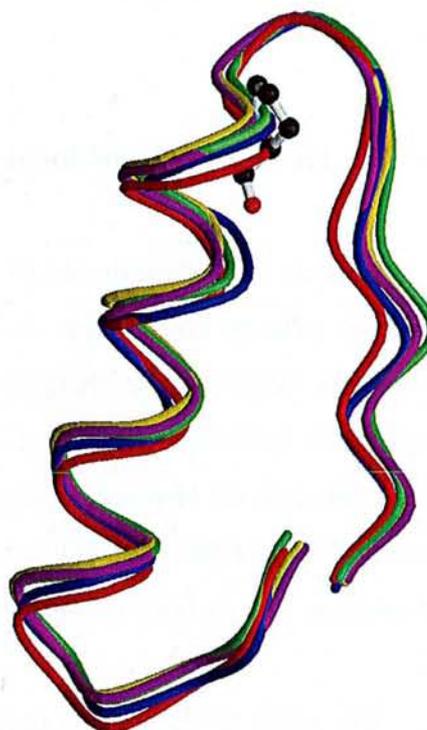
2- Cas de l'ALDH de *S. mutans*

La fixation du cofacteur NADP au domaine de Rossmann de l'ALDH de *S. mutans* est totalement différente de celle observée chez les autres déshydrogénases n'appartenant pas à la famille des ALDH. Le cofacteur enjambe le feuillet β du domaine de Rossmann. Ce nouveau mode de fixation des cofacteurs à nucléotide trouve probablement son origine dans l'orientation et dans la longueur de la première boucle du domaine de Rossmann (Cobessi *et al.*, 2000). La superposition du domaine à celui de plusieurs déshydrogénases NAD(P) dépendantes montre que la boucle située entre le premier brin β et la première hélice α est plus longue dans le cas des ALDH. Sa longueur ainsi qu'une proline, Pro156 chez *S. mutans*, sont conservées chez les ALDH (Figure 6.4).

La modélisation de la fixation du NAD au domaine de Rossmann de Sm-ALDH tel qu'il interagit avec le domaine de Rossmann de la GAPDH NAD dépendante de *Bacillus stearothermophilus* révèle un obstacle à sa fixation en l'état. Il s'agit de contraintes stériques apportées par la longueur de la boucle dont l'orientation semble influencée par la présence de la proline conservée Pro156 dans la séquence de Sm-ALDH. Ainsi, pour se fixer à l'enzyme, le NAD et le NADP n'interagissent pas avec le domaine de Rossmann dans un mode classique mais en adoptant un nouveau mode de fixation (Liu *et al.*, 1997; Steinmetz *et al.*,

1997; Johansson *et al.*, 1998; Moore *et al.*, 1998; Lamb & Newcomer, 1999; Cobessi *et al.*, 1999).

Figure 6.4: superposition de la première hélice α et du premier brin β du domaine de Rossmann de l'ALDH1 (violet), l'ALDH2 (jaune) l'ALDH3 (bleu), l'ALDH9 (vert) et de l'ALDH de *S. mutans* (rouge). La proline conservée chez les ALDH est représentée en mode sphères et bâtons.



Chez les ALDH, le domaine de Rossmann est composé de 5 brins β et de 4 hélices α et l'empreinte GXGXXG n'est pas présente. Par contre, une succession d'acides aminés GX₄G ou plus précisément GstXXG est observée à travers les alignements de séquences. Ces acides aminés constituent une boucle entre le brin β_1 et l'hélice α_8 du domaine de Rossmann. La présence de résidus peu volumineux (Gly, Ser, Thr) dans cette boucle est certainement nécessaire à un positionnement correct du groupement pyrophosphate du cofacteur dont la stabilisation n'est assurée par aucun acide aminé chargé positivement ni par le dipôle d'une hélice α , excepté dans le cas de l'ALDH2 (Steinmetz *et al.*, 1997). Le cycle adénine du cofacteur interagit avec la protéine par une liaison hydrogène (Asp215 OD2...NH2 adénosine) et par contacts de Van der Waals. Des absences de densité électronique, également décrites dans les structures de l'ALDH2, l'ALDH9 et de la RalDH2 (Lamb & Newcomer, 1999), sont observées au niveau du cofacteur. Elles concernent essentiellement la moitié nicotinamide mononucléoside qui interagit avec la protéine par une liaison hydrogène entre son groupement carboxamide et le carbonyle de la leucine Leu251, par des liaisons hydrogène entre Glu377 et

les hydroxyles du ribose et par interaction électrostatique entre la charge positive du cycle nicotinamide et la charge négative du thiolate de la cystéine Cys284. Dans le cas de l'apoenzyme, un ion sulfate est observé à une position similaire à celle occupée par le groupement 2'-phosphate du ribose de l'adénosine. Le glutamate Glu250 se situe à la fin du dernier brin β du domaine de Rossmann.

C- Le changement local de conformation

Lors de la fixation du NADP, la zone Gly208-Asp215, constituée d'une boucle et de l'hélice α_7 subit un changement de conformation qui se traduit par une amplification de la courbure de l'hélice α . Le changement de conformation permet de stabiliser le groupe 2'-phosphate du NADP par l'addition du dipôle de l'hélice α . De plus, ce changement porte la chaîne principale de la protéine au contact du cofacteur par interactions de Van der Waals (Cobessi *et al.*, 1999). Ser211 est l'acide aminé le plus affecté par le changement de conformation, son carbone C_α se déplaçant de 4,1 Å.

Une autre réorganisation conformationnelle intervient lors de la fixation du cofacteur. Sous l'effet de sa fixation, la chaîne latérale de la cystéine Cys284 s'oriente vers le site catalytique. Cette réorientation est nécessaire pour la suite de la réaction où le groupement thiolate attaque le carbone C1 du G3P pour former l'intermédiaire covalent thiohémicétal. Le groupement sulfhydryle de Cys284 est alors positionné dans un nouvel environnement, à proximité des groupements NH de la chaîne principale de Arg283 et Thr285.

Steinmetz *et al.* (1997) proposent qu'une molécule d'eau pourrait être activée par le glutamate afin de déprotonner la cystéine en vue de son activation puisque, pour l'ALDH2, aucun changement de conformation n'a été observé entre les formes apo et holo de l'enzyme. Ce rôle du glutamate est invalidé par les études de mutagenèse dirigée (Marchal *et al.*, 2000). Ainsi, le changement d'environnement du groupement sulfhydryle de la cystéine et la charge positive portée par le cycle nicotinamide du NADP entraînent une baisse du pKa de Cys284 qui passe de 8,5 à 6,1 (Marchal & Branlant, 1999). Cette baisse permet à Cys284 d'exister sous une forme thiolate à un pH physiologique (Cobessi *et al.*, 2000).

D- La conformation du cofacteur

Elle constitue un point de désaccord entre les deux premières structures d'ALDH résolues, où le nicotinamide mononucléoside est observé dans une conformation *anti* dans l'ALDH2, et *syn* dans l'ALDH3. Ces dénominations pour les cofacteurs NAD ou NADP signifient que le nicotinamide adopte une orientation *syn* ou *anti* par rotation autour de la liaison NN1–NC1'.

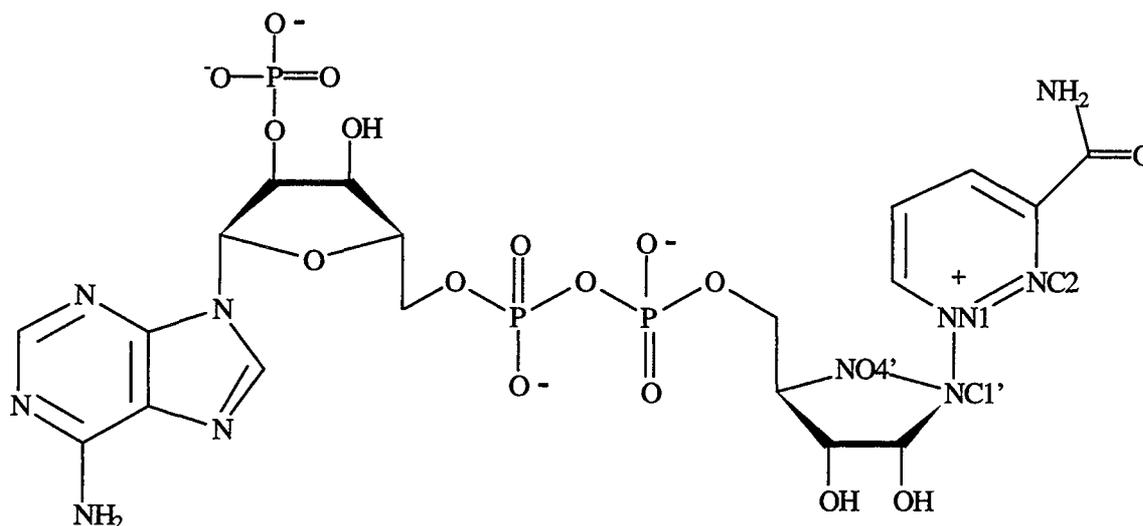


Figure 6.5: représentation schématique du NADP.

L'angle NC2–NN1–NC1'–NO4' a une valeur proche de 120° dans le cas de la conformation *anti* et voisine de -60° pour la conformation *syn* (Carugo & Argos, 1997). Cette conformation est à relier avec le transfert d'hydrure pro-*R* ou pro-*S*.

L'analyse de la structure de l'ALDH9 a montré une orientation *anti* du nicotinamide confirmant ainsi les observations précédentes sur l'ALDH2 (Johansson *et al.*, 1998). L'orientation *anti* du nicotinamide est également observée pour le NADP lié à l'ALDH de *S. mutans*, et le NAD lié à l'ALDH1. Est-il possible qu'une conformation pour le cofacteur soit privilégiée chez les ALDH tétramériques par rapport à celle adoptée chez l'ALDH3 (ALDH dimérique)? Il est également à signaler que la fixation du NAD dans l'ALDH3 est différente de celle observée dans les autres structures d'ALDH. En effet, le cofacteur est moins enfoui dans la protéine. De ce fait, le carbone C4 du nicotinamide ne se serait pas en position optimale pour le transfert d'hydrure, et il n'est observé aucune interaction correspondant à

celle existant entre Glu377 (chez l'ALDH de *S. mutans*) et les hydroxyles du ribose du nicotinamide, conservée chez les ALDH tétramériques.

La résolution de la structure de l'ALDH1 (Moore *et al.*, 1998) a permis de mettre en évidence deux conformations adoptées par le cofacteur NAD. Une première où le C4 du nicotinamide se situe à 7,3 Å de la cystéine catalytique Cys302 et une seconde, minoritaire, où la distance est semblable à celle observée dans l'ALDH de *S. mutans*. En association avec ces taux d'occupation différents, la densité électronique de la chaîne latérale du glutamate Glu268 est mal définie suggérant sa flexibilité. Ces observations ont conduit les auteurs à émettre l'hypothèse de l'existence de deux conformations du cofacteur au cours de la réaction, l'une d'entre elle permettant le transfert d'hydrure et l'autre la déacylation qui serait due à une molécule d'eau observée dans la structure de l'ALDH1 en interaction avec Glu268. Les deux positions du NAD seraient favorisées par le nouveau mode de fixation du cofacteur et par le faible nombre d'interactions entre le NAD et l'enzyme (il n'y a aucun dipôle d'hélice α et aucune charge positive en interaction avec les pyrophosphates du NAD(P)).

E- Le domaine catalytique

Il s'étend du résidu Gly253 au résidu Lys449. Il est constitué d'un feuillet β parallèle et d'hélices α . Deux ions sulfate sont fixés au domaine catalytique des monomères B et D de l'apo et de l'holoenzyme. Un ion est présent dans les monomères A et C de l'Apo1, alors que le monomère A de l'holoenzyme n'en possède pas. Les ions sulfate communs aux monomères B, C et D se situent dans le site catalytique de la protéine représentant probablement le groupement phosphate du G3P. Ils interagissent avec les groupements guanidinium des arginines Arg103 et Arg283, le groupement hydroxyle de la thréonine Thr285, et avec les groupements NH de la chaîne principale de Arg437 et Gly438 (Figure 6.6). Le second ion sulfate, situé à la surface du monomère, interagit avec le groupement NH de Thr328, les groupements NH et ammonium de Lys329 et des molécules d'eau.

Moore *et al.* (1998) rapportent que la spécificité de substrat pour les aldéhydes résulte plutôt de la forme et de la taille du site catalytique que de son caractère hydrophobe. L'analyse du site catalytique de l'ALDH de *S. mutans* révèle son très grand caractère hydrophile de part la nature des résidus qui le composent. En effet, les chaînes latérales de

deux arginines (R103, R283), d'une thréonine (T285), d'une tyrosine (Y151) et d'un aspartate (D440) sont présentes. L'ensemble de ces chaînes lui confère également un potentiel positif lui permettant d'interagir avec les ions sulfate et *a priori* avec les charges négatives du phosphate du G3P.

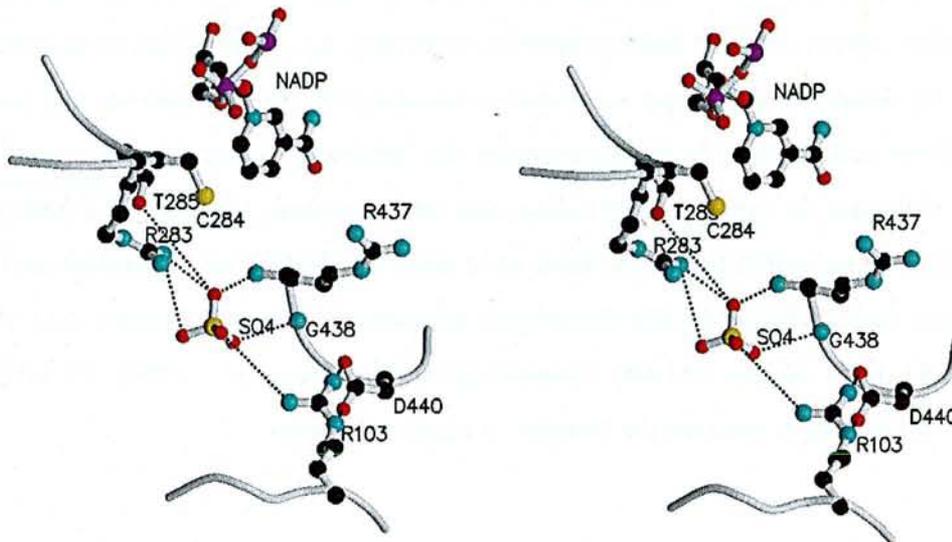


Figure 6.6: vue stéréoscopique des interactions entre un ion sulfate et les acides aminés du site catalytique de l'holoenzyme.

L'hydrophilie et l'environnement positif vont de pair avec la nature du substrat oxydé durant l'acte catalytique. Le G3P porte à la fois des charges négatives sur son phosphate et des fonctions susceptibles d'échanger des liaisons hydrogène: les atomes d'oxygène du phosphate, la fonction hydroxyle en C2 et la fonction aldéhydique en C1. Le potentiel positif du site catalytique superposé au G3P (telle que sa modélisation de sa fixation sur la base de la présence des ions sulfate a été décrite (Cobessi *et al.*, 1999)) est représenté en figure 6.7. Le potentiel électrostatique a été calculé en utilisant le logiciel GRASP (Nicholls *et al.*, 1991).

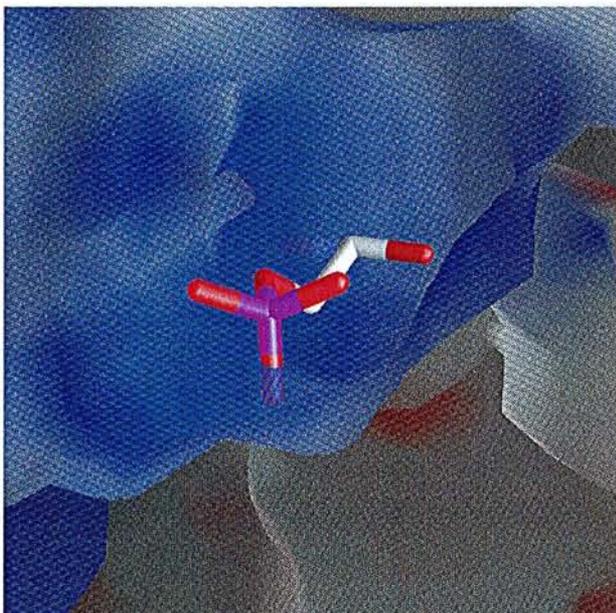


Figure 6.7: représentation du G3P et du potentiel électrostatique du site catalytique de l'holoenzyme.

F- Relation entre la structure du monomère des ALDH et le degré d'oligomérisation

La superposition des structures d'holoenzyme de trois ALDH (ALDH2, ALDH3, ALDH9) avec celle de *S. mutans*, en utilisant le logiciel O (Jones & Kjeldgaard, 1990), montre des valeurs rms très faibles entre les structures des monomères et souvent plus petites lorsque les domaines de chaque monomère sont comparés séparément les uns par rapport aux autres. Pour cette raison, le positionnement des différents domaines a été étudié en prenant comme référence de superposition le domaine de Rossmann. Cette étude a permis de montrer que l'angle entre le domaine catalytique et le domaine d'oligomérisation est plus petit pour le monomère de l'ALDH3 (ALDH dimérique) que pour les monomères de l'ALDH9, l'ALDH2 et l'ALDH de *S. mutans* (ALDH tétramériques) (Cobessi *et al.*, 1999). La fermeture de cet angle résulterait de la rotation du domaine d'oligomérisation.

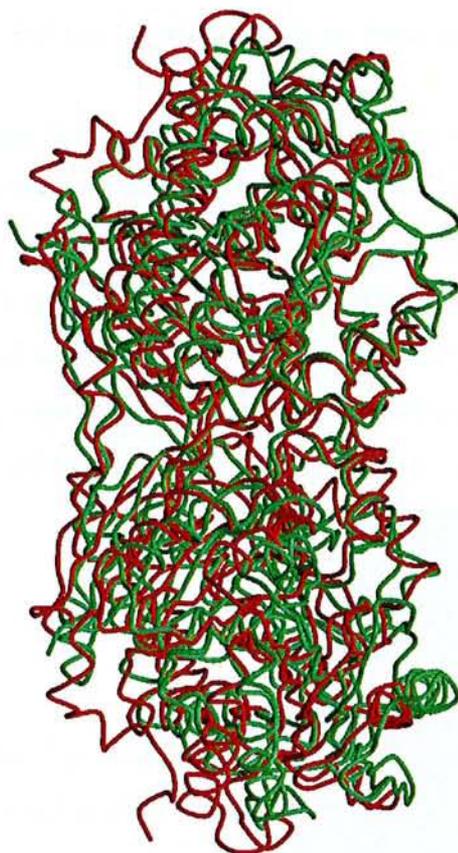


Figure 6.8: superposition du dimère de l'ALDH3 (rouge) et du dimère correspondant de l'ALDH de *S. mutans* (vert).

L'étude des alignements de séquences révèle la conservation de quatre résidus pouvant avoir un rôle clef dans la détermination de l'état oligomérique. Chez l'ALDH3, une proline (Pro78) et un tryptophane (Trp80) sont conservés et situés à l'entrée du domaine d'oligomérisation. La proline est également présente chez l'ALDH9 où l'angle entre le domaine d'oligomérisation et le domaine catalytique est le plus petit des ALDH tétramériques retenues pour cette étude. Une glycine située dans la séquence de Sm-ALDH à la position 120, après la proline chez l'ALDH9, est principalement conservée chez les ALDH tétramériques. Elle est mutée en des résidus variables chez l'ALDH3. Son rôle serait d'apporter de la flexibilité à la chaîne principale alors que celui de la proline serait de contraindre l'orientation de la chaîne principale. Enfin, les rôles de Trp80 et de Phe418, quatrième résidu, conservés chez l'ALDH3, seraient d'empêcher tout mouvement du domaine d'oligomérisation. Ces résidus volumineux apportent des contraintes stériques en raison de leur proximité spatiale. Ces acides aminés pourraient donc être les principaux déterminants de l'état oligomérique des ALDH. Cependant, ce ne sont peut-être pas les seuls. Il se pourrait que des résidus situés aux interfaces des dimères soient également impliqués (Cobessi *et al.*, 1999).

Suite à cette étude et en raison de l'échec du remplacement moléculaire avec la structure de l'ALDH3, les dimères de l'ALDH3 et de l'ALDH de *S. mutans* ont été superposés en utilisant les programmes Lsqman (Keleywegt & Jones, 1994) et Moleman2 (Kleywegt, 1997) (Figure 6.8) et l'angle entre les monomères de chaque dimère calculé. L'angle entre les monomères du dimère de l'ALDH3 est plus petit que celui calculé pour l'ALDH de *S. mutans*. Si le tétramère était généré à partir du dimère de l'ALDH3 par application de la symétrie non cristallographique observée chez l'ALDH de *S. mutans*, cela conduirait à la formation d'un oligomère dont les interfaces entre dimères présenteraient des contacts non autorisés entre atomes en vis-à-vis. Ces contacts sont tellement nombreux et surtout défavorables qu'il est impossible qu'un tel tétramère puissent se former, ce qui pourrait ajouter une nouvelle explication à l'échec du remplacement moléculaire avec le modèle de l'ALDH3. Une telle association ne pourrait se produire qu'à la suite d'un changement de conformation des deux partenaires.

A partir de ces analyses, des mutations de la séquence de Sm-ALDH ont été proposées et réalisées au M.A.E.M afin de générer un dimère. La triple mutation L116P, G120E, M118F a permis l'obtention d'une solution protéique polydisperse contenant entre autre, une forme

dimérique. Afin d'obtenir uniquement une forme dimérique, une nouvelle série de mutations de résidus situés aux interfaces du tétramère a été proposée (Arg117, Glu119, Glu124, Lys133 et Lys134). Ces expériences sont en cours de réalisation.

II- L'Apo2

Si le repliement de la protéine est identique à celui décrit pour l'Apo1, il existe des différences non négligeables dans le positionnement de deux chaînes latérales du site catalytique et du domaine du cofacteur: celles de Glu250 et Arg437.

Le groupement carboxylate de la chaîne latérale du glutamate Glu250 est positionné à 3,7 Å de la cystéine essentielle au lieu de 6,9 Å observés dans la structure de l'Apo1. La cystéine conserve la même orientation que celle observée dans la structure de l'Apo1. Aucune molécule d'eau n'est visible entre Glu250 et Cys284. Les chaînes latérales de trois autres résidus, Arg283, Arg437, et Gln455, subissent également des déplacements au sein des deux sites. Arg437 est en interaction avec Glu250 dans les structures de l'Apo1 et de l'holoenzyme. Compte tenu de l'orientation de la chaîne latérale de Glu250, le pont salin qui existait dans l'Apo1 et l'holoenzyme est rompu. Le groupement guanidinium de Arg437 est en interaction avec un second ion sulfate présent dans le site catalytique qui possède un taux d'occupation de 0,5. Ce second ion sulfate (SO_{4b}) partage également une liaison hydrogène avec le groupement guanidinium de l'arginine Arg283. L'orientation de la chaîne latérale de l'arginine Arg103 est quant à elle peu affectée par la présence de SO_{4b} dans le site catalytique. SO_{4b} interagit également par liaisons hydrogène avec le groupement amide de Asn154 et avec le groupement NH de Cys284. Ces dernières interactions sont celles suggérées par Steinmetz *et al.* (1997) dans le cadre de la stabilisation de l'oxyanion de l'intermédiaire tétraédrique lors de la réaction catalytique. Ainsi, l'association de ce nouvel ion et de celui précédemment observé (SO_{4a}) correspondrait à la trace du substrat dans le site catalytique. L'étude structurale d'un complexe ternaire confirmera cette hypothèse.

Le changement d'environnement du glutamate vers un environnement plus hydrophobe associé à la perte de l'interaction avec Arg437 dans l'Apo2 pourrait être responsable d'une augmentation du pKa de Glu250 vers une valeur de 7,6 mesurée dans le complexe ternaire. Cet environnement est fourni par des groupements CH₂ et CH₃ provenant

des acides aminés suivants: Thr229, Leu251, Leu405, Phe444 qui sont fortement conservés, et Leu159 qui est souvent remplacée par une méthionine ou une isoleucine dans différentes séquences d'ALDH (Perozich *et al.*, 1999). L'environnement hydrophobe serait donc préservé dans les différentes structures d'ALDH.

Afin de confirmer les hypothèses sur les rôles de Arg103 et Arg283 pour la stabilisation du phosphate du G3P, sur les rôles du site oxyanion et de Asn154, le complexe ternaire ALDH–NADP–G3P et le mutant C284S sont étudiés.

III- Le mutant C284S et le complexe ternaire

A-Structure du mutant C284S

Glu250	CH ₂ /CH ₃	Distance (Å)
OE1	Leu159 CD1	4.6
OE1	Leu159 CD2	4.5
OE1	Thr229 CG2	3.2
OE2	Thr229 CG2	4.4
OE2	Thr285 CG2	4.8
OE2	Leu405 OE2	4.5
OE1	Phe444 CZ	4.8
OE1	Phe444 CE2	4.8
OE1	Phe444 CD2	4.6
OE2	Phe444 CZ	3.7
OE2	Phe444 CE2	3.4
OE2	Phe444 CD2	4.7

Tableau 6.3: groupements CH₂/CH₃ présents dans un rayon de 5 Å autour de Glu250 pour le mutant C284S.

Le repliement global est similaire à celui observé pour l'holoenzyme. Il n'existe plus d'interaction entre Glu250 et Arg437. Glu250 se situe à 3 Å de Ser284. Le glutamate se trouve alors dans un environnement hydrophobe. Les acides aminés cités précédemment comme pouvant influencer le pKa de Glu250 par leur groupement CH₂ et CH₃ sont consignés dans le tableau 6.3. Le changement d'orientation du glutamate Glu250 par rapport à sa position dans l'holoenzyme de type sauvage est favorisé par les mouvements du cofacteur pour lequel la densité électronique au niveau de la partie nicotinamide mononucléoside est mal définie. Ce désordre structural est probablement favorisé par la perte de l'interaction électrostatique entre la charge positive du cycle nicotinamide et la charge négative que portait la cystéine. Afin de positionner au mieux le NADP dans sa densité électronique, il a été nécessaire de modifier la conformation de la partie nicotinamide mononucléotide par rapport à celle observée dans l'holoenzyme.

B- Le complexe ternaire

Comme dans le cas du mutant C284S, le groupement carboxylate du glutamate Glu250 se situe à 3 Å de la sérine au lieu de 6,9 Å observés dans les structures de l'Apo1 et de l'holoenzyme, ou encore au lieu des 3,7 Å observés dans la structure de l'Apo2. De ce fait, l'interaction entre le groupement carboxylate de Glu250 et le groupement guanidinium de Arg437, observée dans les structures de l'holoenzyme et de l'Apo1, est perdue.

Protéine	Substrat	Distance (Å)
Arg103 NH1	OC3	3.2
Asn154 ND2	O1	3.2
Arg283 NH2	OC2	2.7
Ser284 N	O1	3.2
Thr285 OG1	OPC	3.0
Arg437 N	OC2	3.1
Tyr155 OH	OPC	3.1

Tableau 6.4: interactions par liaisons hydrogène entre le substrat et la protéine pour le monomère A du complexe ternaire.

Pour la thèse de David COBESSI,

Il n'y pas Page 121-126.

Le tableau 6.5 présente les interactions par liaisons hydrogène entre les acides aminés et le NADP observées dans la structure de l'holoenzyme de Sm-ALDH. Les acides aminés conservés dans l'alignement de séquences 1 sont indiqués en gras.

Donneurs	Accepteurs
Phe153 N	NO2
Lys177NZ	AO2*
Thr180 N	AOP2
Thr180 OG1	AOP3
Gly210 N	AOP3
Ser231 N	AO1
AO3*	Ser151 O
AN6	Asp215 OD2
NN7	Leu251 O
NO2*	Glu377 OE2
NO3*	Glu377 OE1

Tableau 6.5: interactions entre le cofacteur et les acides aminés observées dans la structure de l'holoenzyme de Sm-ALDH.

Un acide aminé parmi les neuf répertoriés pourrait être responsable de la spécificité de cofacteur pour le NADP chez les ALDH. Il s'agit de la thréonine Thr180 qui est remplacée par un glutamate, une proline ou une isoleucine chez les ALDH NAD dépendantes. La GAPDH NAD dépendante non phosphorylante de *T. tenax* possède dans sa séquence une isoleucine. La thréonine interagit avec deux des atomes d'oxygène portés par le 2'-phosphate du NADP. Ainsi la spécificité de cofacteur pour le NADP serait due à la perte de charge et à l'absence de contraintes stériques apportées par les résidus Glu, Pro ou Ile mutés en Thr ou Ser. D'ailleurs, la mutation de la thréonine en glutamate ou en aspartate pour l'ALDH NADP dépendante de *Vibrio harveyi* se traduit par un changement de spécificité de cofacteur (Zhang *et al.*, 1999). Cependant, il est probable que ce ne soit pas le seul déterminant de la spécificité de cofacteur pour le NADP puisque la GAPDH NADP dépendante non phosphorylante de *M. thermoautotrophicum* possède à cette position une leucine.

Concernant les autres résidus, et notamment ceux qui interagissent avec le NADP par leur chaîne latérale, ils sont mutés en des acides aminés possédant la même fonction sur leur chaîne latérale. C'est le cas de Lys177, mutée en histidine, de Thr180 mutée en sérine et de Asp215 muté en glutamate. Pour les résidus interagissant par la chaîne principale avec le NADP, ils sont mutés en des résidus présentant des propriétés différentes au niveau de leur chaîne latérale.

L'analyse de l'alignement de séquences révèle également la conservation de résidus impliqués dans l'interaction avec le substrat. Les acides aminés les plus conservés sont Asn154, Tyr155 et Arg103 qui est en interaction dans la structure étudiée avec l'aspartate Asp440 conservé dans six séquences sur sept. Arg103 est mutée en alanine chez *T. tenax*. Cependant, une lysine située avant cette alanine pourrait avoir le même rôle que Arg103 alors que Asp440 est muté en glycine. L'arginine Arg283 est conservée dans cinq séquences sur sept et est mutée en valine chez *M. thermoautotrophicum* et *M. jannaschii*. Quant à Arg437, elle est mutée en histidine chez *T. tenax* et en phénylalanine chez *M. thermoautotrophicum* et *M. jannaschii*. Pour les deux dernières séquences, une arginine est présente après la phénylalanine. Ainsi, l'histidine et cette arginine pourraient avoir le même rôle que Arg437 chez Sm-ALDH.

Le second alignement de séquences (Figure 6.11), réalisé entre les GAPDH NADP dépendantes non phosphorylantes du pois, du maïs, du tabac et de *S. mutans* permet d'observer un plus grand pourcentage d'identité de séquences (supérieur à 45 %) ainsi qu'une très forte homologie de séquences. De plus, les acides aminés interagissant avec le substrat, et listés précédemment, sont conservés sans exception. Ceci permet d'affirmer sans ambiguïté que ces quatre ALDH possèdent la même structure quaternaire.

Afin d'évaluer si les sept protéines du premier alignement possèdent la même structure tridimensionnelle, les séquences des ALDH de *Methanococcus jannaschii*, *Thermoproteus tenax*, *Methanobacterium thermoautotrophicum* sont comparées une après l'autre à celle de *S. mutans*. Des identités de séquences de plus de 35 % apparaissent. Ce résultat, couplé aux observations précédentes permet d'affirmer sans ambiguïté que toutes les GAPDH NAD(P) dépendantes non phosphorylantes adoptent un repliement similaire quel que soit l'organisme dont elles sont isolées, eucaryote, eubactérie ou archae. La résolution de la structure de la GAPDH NAD dépendante non phosphorylante de *T. tenax*, actuellement en cours (Brunner *et al.*, 2000), devrait apporter des renseignements supplémentaires sur ces observations mais

également sur les éléments structuraux pouvant avoir un rôle dans la stabilisation des structures secondaires des protéines isolées d'organismes hyperthermophyles.

B- Stabilisation du 2'-phosphate du NADP chez les déshydrogénases à domaine de Rossmann

1- Cas de l'ALDH de *S. mutans*

La fixation du cofacteur au domaine de Rossmann entraîne un réarrangement local de la zone Gly208-Asp215 constituée d'une boucle et de l'extrémité N-terminale de l'hélice α_7 . Le déplacement maximal entre les C α de Ser211 dans l'apo et l'holoenzyme atteint 4,1 Å. Le changement local de conformation est lié à la stabilisation du 2'-phosphate du cofacteur par le dipôle de l'hélice α qui, dans l'apoenzyme, interagit avec les charges négatives d'un ion sulfate. Le phosphate interagit également par liaisons hydrogène avec l'hydroxyle de Thr180 et les groupements NH de Thr180 et Gly210 et avec l'amine de Lys177 par l'intermédiaire d'un pont salin. Dans l'apoenzyme, l'ion sulfate crée des liaisons hydrogène avec l'hydroxyle de Thr180 et le groupement NH de Gly210. La distance entre la position du sulfate (qui pourrait correspondre au 2'-phosphate avec la conformation de la boucle de l'apoenzyme) et le 2'-phosphate du NADP est de 4,3 Å. Bien qu'une arginine soit présente en position 209, elle n'est impliquée dans aucune interaction avec ces groupements chargés négativement.

2- Comparaison avec des déshydrogénases NADP dépendantes

Pour réaliser cette comparaison, 16 déshydrogénases possédant un domaine de Rossmann ont été sélectionnées dans la PDB. Les interactions par liaisons hydrogène entre le 2'-phosphate du NADP et les acides aminés du domaine de Rossmann sont présentées dans le tableau 6.6. Dans 14 structures observées (Sm-ALDH compris), au moins un résidu basique interagit par sa chaîne latérale avec le groupement 2'-phosphate du NADP. Dans la structure de Sm-ALDH, il s'agit d'une lysine (Lys177). Dans les structures de la formate déshydrogénase et de la 17- β -hydroxystéroïde déshydrogénase, aucune interaction entre le 2'-phosphate et la chaîne latérale d'un résidu basique n'est observée. Dans 12 structures, des interactions entre la chaîne principale et le groupement phosphate se produisent.

Protéine (code PDB)	Enzyme	Organisme	Acide aminé en interaction avec le 2'-phosphate du NADP par liaison hydrogène	Distance O...D D: groupement donneur) en Å
1CYD	Carbonyl réductase (Tanaka <i>et al.</i> , 1996)	<i>Mus musculus</i>	K17 NZ R39 NE R39 NH2 T40 OG	2,90 2,72 2,81 2,54
1FDT	17- β -hydroxystéroïde déshydrogénase (Breton <i>et al.</i> , 1997)	<i>Homo sapiens</i>	C10 N S11 N R37 N R37 N	2,83 3,19 2,73 2,69
1GET	Glutathione réductase (Mittl <i>et al.</i> , 1994)	<i>Escherichia coli</i>	R198 N R198 NH2 R204 NH1 R204 NH1	2,89 2,92 2,85 3,17
1TDF	Thiorédoxine réductase (Waksman <i>et al.</i> , 1994)	<i>Escherichia coli</i>	R176 NE R176 NH2 R176 NH2	3,00 3,16 2,80
1TYP	Trypanothione réductase (Hunter <i>et al.</i> , 1992)	<i>Crithidia fasciculata</i>	Y 221 OH R222 NH2 R222 NH2	2,99 2,79 3,20
1YKF	Alcool déshydrogénase (Korkhin <i>et al.</i> , 1998)	<i>Thermoanaerobium brockii</i>	S199 N S199 OG R200 N R200 NE Y218 OH	2,99 2,98 3,01 3,18 2,68
1HFQ	Dihydrofolate réductase (Cody <i>et al.</i> , 1997)	<i>Homo sapiens</i>	K54 NZ S76 OG R77 N	2,76 2,64 2,70
1PGO	6-phosphogluconate déshydrogénase (Adams <i>et al.</i> , 1994)	<i>Ovis orientalis aries</i>	R33 NE R33 NH2 T34 N T34 OG	2,92 2,93 2,94 2,99
1QOR	Quinone oxydo-réductase (Thorn <i>et al.</i> , 1995)	<i>Escherichia coli</i>	R137 NH2 G173 N K177 NZ	2,75 2,86 2,78
2DPG	Glucose-6-phosphate déshydrogénase (Cosgrove <i>et al.</i> , 1998)	<i>Leuconostoc mesenteroides</i>	R46 N R46 NE R46 NH2	3,20 2,94 3,03
1QP8	Formate déshydrogénase (Peat <i>et al.</i> , non publié)	<i>Pyrobaculum aerophilum</i>	T156 N T156 OG T156 OG	3,02 2,62 3,24
1AE1	Tropinone réductase (Nakajima <i>et al.</i> , 1998)	<i>Dratura stramonium</i>	K31 NZ R53 N R53 NH2	2,70 3,30 2,87
1DAP	Acide diaminopimélique déshydrogénase (Scapin <i>et al.</i> , 1996)	<i>Corynebacterium glutamicum</i>	Y11 N S35 OG R36 N R37 NH1	2,78 2,88 3,02 2,95
1NAS	Sépiaptérine réductase (Auerbach <i>et al.</i> , 1997)	<i>Mus musculus</i>	S17 OG R43 N R43 NE R43 NH2 S44 N R48 NH2	2,77 2,87 2,79 3,25 3,13 3,27
1CF2	Glyceraldéhyde-3- phosphate déshydrogénase (Charron <i>et al.</i> , 2000)	<i>Methonotermus fervidus</i>	K33 NZ K33NZ R35 N T34 N T34 OG	2,96 2,80 2,87 2,73 2,88

Tableau 6.6: interactions par liaisons hydrogène entre le 2'-phosphate du NADP et les acides aminés du domaine de Rossmann de plusieurs déshydrogénases NADP dépendantes.

Les liaisons hydrogène créées avec d'autres chaînes latérales dont celles de la thréonine ou de la sérine, résidus présents avec une fréquence de plus 50 %, s'associent aux interactions précédentes. Dans toutes les structures, des interactions se produisent avec les acides aminés de la seconde boucle du domaine de Rossmann dont la longueur et la composition en acides aminés diffèrent entre les structures.

V- Structures cristallines des formes apo et holo d'une ALDH NADP dépendante de *Streptococcus mutans*.

Apo and holo crystal structures of an NADP-dependent aldehyde dehydrogenase from Streptococcus mutans (Cobessi et al., 1999).

Ce premier article publié dans *Journal of Molecular Biology* est consacré à l'analyse structurale de Sm-ALDH. Après une brève description de la structure de l'holoenzyme, un premier paragraphe présente la comparaison des monomères de l'ALDH2, l'ALDH3, l'ALDH9 et Sm-ALDH. Cette comparaison structurale révèle une différence considérable entre les monomères au niveau de l'orientation du domaine d'oligomérisation. Il apparaît que l'angle entre celui-ci et le domaine catalytique est plus fermé dans le cas de l'ALDH3 (dimérique) que dans les autres ALDH. Sur la base de cette analyse, les acides aminés semblant impliqués dans ce changement d'orientation ont été proposés comme déterminants structuraux du degré d'oligomérisation des ALDH.

La comparaison entre les formes apo et holo montre un changement structural se produisant au niveau de la zone 208-215 nécessaire à la stabilisation du 2'-phosphate du NADP. Dans l'apoenzyme, un ion sulfate est situé dans la poche qui fixe le 2'-phosphate du NADP. La comparaison des structures des ALDH NAD dépendantes avec la structure de Sm-ALDH montre l'importance du résidu thréonine 180 (Thr180) dans la détermination de la spécificité de cofacteur. Thr180 est remplacée par Ile, Glu ou Pro chez les ALDH NAD dépendantes. Ainsi la spécificité de cofacteur pour le NADP serait due à l'absence de charge et de contraintes stériques à proximité du 2'-phosphate du NADP.

Le dernier paragraphe de l'article est consacré à l'analyse du site catalytique dans lequel est fixé un ion sulfate stabilisé par les charges positives de Arg103 et Arg283. L'ion sulfate du site catalytique correspondrait au groupement phosphate du G3P. Sur la base de cette hypothèse, la modélisation de la fixation du G3P a été réalisée. Une analyse structurale

avec la GAPDH NAD dépendante phosphorylante de *Bacillus stearothermophilus* est également présentée. Elle révèle des similitudes dans le recrutement des acides aminés impliqués dans la fixation du substrat.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

JMB



S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

**Apo and Holo Crystal Structures of an
NADP-dependent Aldehyde Dehydrogenase
from *Streptococcus mutans***

David Cobessi, Frédérique Tête-Favier, Stéphane Marchal Saïd Azza,
Guy Branlant and André Aubry

Article No. jmbi.1999.2853 available online at <http://www.idealibrary.com> on IDEAL® *J. Mol. Biol.* (1999) 290, 161–173

JMB

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY



Apo and Holo Crystal Structures of an NADP-dependent Aldehyde Dehydrogenase from *Streptococcus mutans*

David Cobessi¹, Frédérique Tête-Favier¹, Stéphane Marchal²
Saïd Azza², Guy Branlant² and André Aubry^{1*}

¹Laboratoire de Cristallographie et Modélisation des Matériaux Minéraux et Biologiques, Groupe Biocristallographie, ESA 7036 Faculté des Sciences, BP 239 54506 Vandoeuvre-lès-Nancy France

²Laboratoire de Maturation des ARNs et Enzymologie Moléculaire, UMR 7567 CNRS-UHP, Faculté des Sciences, BP 239 54506 Vandoeuvre-lès-Nancy France

The aldehyde dehydrogenases (ALDHs) are a superfamily of multimeric enzymes which catalyse the oxidation of a broad range of aldehydes into their corresponding carboxylic acids with the reduction of their cofactor, NAD or NADP, into NADH or NADPH. At present, the only known structures concern NAD-dependent ALDHs. Three structures are available in the Protein Data Bank: two are tetrameric and the other is a dimer. We solved by molecular replacement the first structure of an NADP-dependent ALDH isolated from *Streptococcus mutans*, in its apo form and holo form in complex with NADP, at 1.8 and 2.6 Å resolution, respectively. Although the protein sequence shares only approximately 30% identity with the other solved tetrameric ALDHs, the structures are very similar. However, a large local conformational change in the region surrounding the 2' phosphate group of the adenosine moiety is observed when the enzyme binds NADP, in contrast to the NAD-dependent ALDHs.

Structure and sequence analyses reveal several properties. A small number of residues seem to determine the oligomeric state. Likewise, the nature (charge and volume) of the residue at position 180 (Thr in ALDH from *S. mutans*) determines the cofactor specificity in comparison with the structures of NAD-dependent ALDHs. The presence of a hydrogen bond network around the cofactor not only allows it to bind to the enzyme but also directs the side-chains in a correct orientation for the catalytic reaction to take place. Moreover, a specific part of this network appears to be important in substrate binding. Since the enzyme oxidises the same substrate, glyceraldehyde-3-phosphate (G3P), as NAD-dependent phosphorylating glyceraldehyde-3-phosphate dehydrogenases (GAPDH), the active site of GAPDH was compared with that of the *S. mutans* ALDH. It was found that Arg103, Arg283 and Asp440 might be key residues for substrate binding.

© 1999 Academic Press

Keywords: aldehyde dehydrogenase; NADP; cofactor specificity; X-ray structure; substrate specificity

*Corresponding author

Abbreviations used: 3-PGA, 3-phosphoglycerate; G3P, glyceraldehyde-3-phosphate; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; ALDH, aldehyde dehydrogenase; Sm-ALDH, *Streptococcus mutans* aldehyde dehydrogenase; rmsd, root mean square deviation.

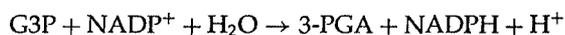
E-mail address of the corresponding author: aubry@lcm3b.u-nancy.fr

Introduction

NAD(P)-dependent aldehyde dehydrogenases catalyse the irreversible oxidation of aldehydes to their corresponding acids. Many such enzymes are involved in cellular detoxification and cellular differentiation, the most widely studied being ALDH1 (cytosolic, EC 1.2.1.3), and ALDH2 (mitochondrial, EC 1.2.1.2), which oxidise the acetaldehyde produced by alcohol dehydrogenase during the conversion of ethanol (Steinmetz *et al.*, 1997). ALDHs are known to exist in dimeric or

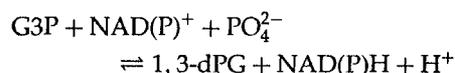
tetrameric state. A total of 12 classes have been described in humans (Yoshida *et al.*, 1998). At present, four structures belonging to four different classes have been solved: one is a dimeric enzyme (ALDH3, EC 1.2.1.5) (Liu *et al.*, 1997) and the others are tetrameric (ALDH1; ALDH2; ALDH9, EC 1.2.1.8) (Moore *et al.*, 1998; Steinmetz *et al.*, 1997; Johansson *et al.*, 1998). These structures were determined in their holoform in complex with NAD. Structural analyses revealed a similar assembly of the catalytic domain and Rossmann fold and showed a new type of interaction between NAD and the Rossmann fold (Liu *et al.*, 1997). A catalytic mechanism was proposed in which two residues, a cysteine and a glutamate, respectively, participate as the nucleophile and general base for activation of the catalytic cysteine and deacylation of the enzyme (Sheikh *et al.*, 1997).

NADP-dependent ALDH from *Streptococcus mutans* (Sm-ALDH, EC 1.2.1.9) is a tetrameric enzyme of 474 residues per monomer which catalyses the irreversible oxidation of glyceraldehyde-3-phosphate (G3P) into 3-phosphoglycerate (3-PGA) with reduction of its cofactor NADP into NADPH according to the reaction:



This reaction produces the NADPH necessary for the correct growth of some *S. mutans* strains which lack both glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase. Both these enzymes belong to the oxidative part of the hexose monophosphate pathway where they generate the NADPH required for other biosynthesis pathways (Crow & Wittenberger, 1979; Boyd *et al.*, 1995).

On the basis of this reaction and due to the nature of the substrate, this enzyme is an NADP-dependent non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Yet, on the basis of the sequence alignment, it has been classified in the ALDH superfamily together with NADP-dependent non-phosphorylating GAPDHs from pea and maize (Habenicht *et al.*, 1994). Indeed, these enzymes do not display sequence homology with other phosphorylating GAPDHs (EC 1.2.1.12 or EC 1.2.1.13), for which more than ten structures are available in the Protein Data Bank. Phosphorylating GAPDHs catalyse the reversible oxidation of G3P into 1,3-diphosphoglycerate (1,3-dPG) with reduction of NAD(P) to NAD(P)H by phosphorylation according to the reaction:



Sm-ALDH shares 50% sequence identity with both maize and pea ALDHs and approximately 30% with ALDH1, ALDH2, ALDH3 and ALDH9. Sequence alignment of ALDHs reveals that the catalytic cysteine and glutamate (Cys284 and Glu250

in Sm-ALDH), together with a large number of glycine and proline residues, have been conserved.

We solved the structures of the apo- and holoenzyme of NADP-dependent ALDH from *S. mutans* by molecular replacement using the structure of ALDH from cod liver (Johansson *et al.*, 1998) as a template. The structure of the apoenzyme was solved at 1.8 Å resolution and that of the holoenzyme, in complex with NADP, at 2.6 Å resolution. Here we describe these structures and compare them with the structures of three NAD-dependent ALDHs and *Bacillus stearothermophilus* GAPDH.

Results and Discussion

Overall structure of the holo form and comparison with other ALDHs

Holo ALDH from *S. mutans* exists as a tetramer (Figure 1) with four identical subunits related by non-crystallographic 222 symmetry. The tetramer can be considered as a dimer of dimers, as described for ALDH2 (Steinmetz *et al.*, 1997). The root-mean-square deviation (rmsd) values for all possible comparisons of the C α trace of monomers vary from 0.23 Å to 0.36 Å. Each monomer can be divided into three domains (Figure 2): a cofactor binding domain composed of a core which resembles the Rossmann fold (Rossmann *et al.*, 1974) (residues 145-252) flanked by five helices and four strands (residues 2-118, 450-464), a catalytic α/β domain (residues 253-449), and a small protruding domain which enables the oligomerisation (residues 119-144 and 465-475).

Among the other known ALDH structures, ALDH2 and ALDH9 are tetrameric and were solved in both apo and holo forms (Steinmetz *et al.*, 1997; Johansson *et al.*, 1998). ALDH3 is dimeric and was solved in its holo form (Liu *et al.*, 1997). When optimising the sequence alignment (Figure 3) and the superimposition of the three-dimensional structures of holo forms, the tetrameric structures display the same fold (Figure 4). On the other hand, the dimeric enzyme lacks the N-terminal sequence corresponding to the first four β -strands (labelled A, B, C and D in Sm-ALDH). Except for this difference, the superimposition of the backbones is good, as illustrated by the low global rmsd values of Sm-ALDH compared with ALDH2, ALDH3 or ALDH9 (Table 1). However, the rmsd values calculated for the three separate domains are often smaller than the global values. Therefore, as only the positioning between domains can be responsible for such a discrepancy, another comparative approach was used.

Since the Rossmann fold (residues 145-252) constitutes the core of the central domain, it can be chosen as the reference for measurement of the rotation of the other two domains (Table 1 and Figure 4). For the catalytic domain, only slight differences are found. The observed low values of the rotation angles might be explained by the strength of the interactions between the cofactor

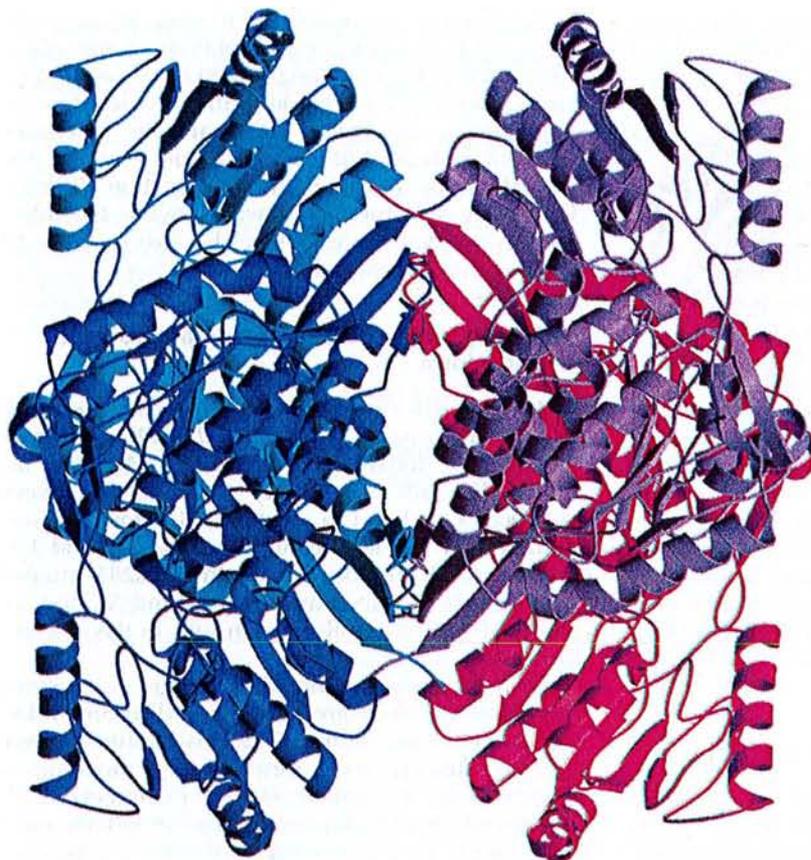


Figure 1. Ribbon diagram of the Sm-ALDH tetramer using MOLSCRIPT (Kraulis, 1991) and Raster3D (Meritt & Murphy, 1994). The monomers of each dimer are similarly coloured.

and the catalytic domains. Indeed, they are maintained by the main chain which passes twice from one domain to the other, and by numerous hydrogen bonds, hydrophobic contacts, and salt bridges. These strong interactions prevent flexibility, resulting in similar positioning of the catalytic and the cofactor domains. The largest value of the rotation angle is observed for ALDH3, i.e. the dimeric enzyme, which differs by 6.4° . This observation may be explained by the absence of the N terminus, thereby reducing the number of possible non-covalent interactions, as well as by the level of oligomerisation.

Looking at the oligomerisation domains, the observed rotations are much more significant. They correspond to different opening angles of the L-shaped monomer (Figure 4). Again, ALDH3 shows the highest deviation (17.4°) from Sm-ALDH, which results in a reduced angle between the catalytic and oligomerisation domains, leading to a sharper bend in the monomer. Likewise, in order to superimpose the ALDH3 dimer on the corresponding Sm-ALDH dimer, the angle between both monomers of ALDH3 must be increased by 8.5° . Thus the rotation of the oligomerisation domain appears to lead to the formation of differently orientated dimers. Similarly, this rotation could prevent the dimers from

assembling to form the tetramer. Indeed, generation of the tetramer from the dimer of ALDH3 using the 2-fold axis observed in Sm-ALDH encounters an insurmountable obstacle: numerous bad contacts and clashes of side-chains are observed. These occur mainly in the region of the first strand of the oligomerisation domain, resulting from the misorientation of the entire antiparallel β -sheet (βE , βF , βX) because of structural constraints.

Analysis of the sequences around the hinge region, which is where the oligomerisation domain emerges from the core of the monomer, revealed a conserved proline, Pro78, in ALDH3, which is dimeric. Proline residues are known to stiffen the main chain and might therefore maintain the concerned domain in a correct geometry, as often observed in oligomerisation performed by domain swapping or arm exchange (Bergdoll *et al.*, 1997). However, an equivalent proline, Pro134, is present in the same position in ALDH9. Remarkably, its monomer also shows a strong bend. Thus, the proline may be responsible for the rotated oligomerisation domain, but it is not the only determining factor for the oligomeric state, since ALDH9 is tetrameric. Concerning the flexibility of the main chain, we found that glycine Gly120 in Sm-ALDH is mainly conserved in tetrameric enzymes, while

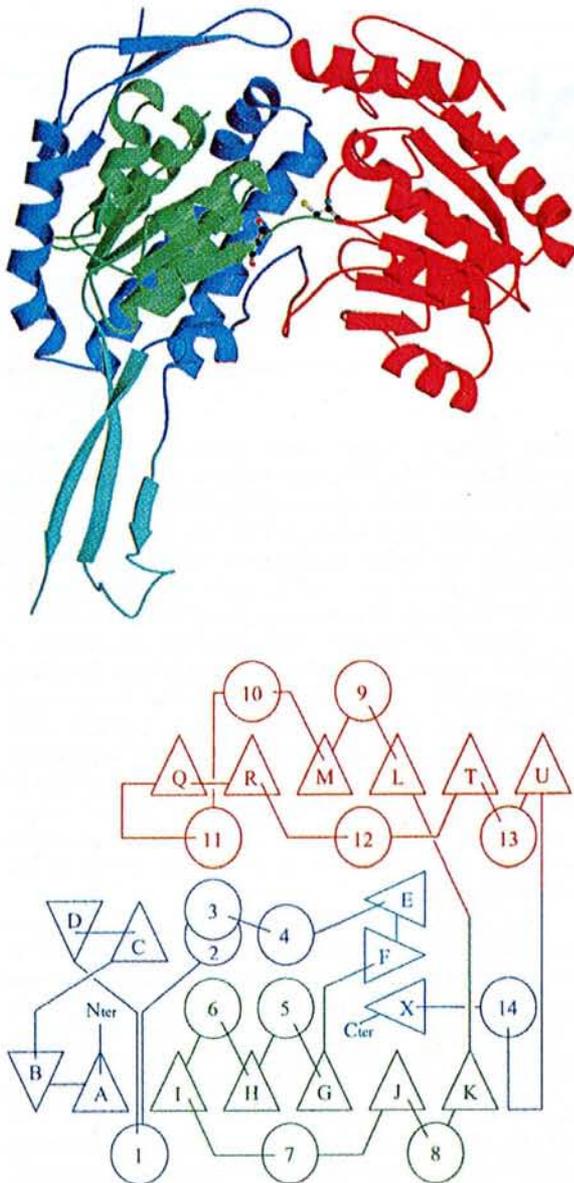


Figure 2. (a) Domain localisation in the Sm-ALDH monomer using MOISCRIP (Kraulis, 1991) and Raster3D (Meritt & Murphy, 1994). The cofactor binding domain is coloured in blue and Rossmann fold in green. The catalytic domain is coloured in red and oligomerisation domain in light blue. The catalytic residues, Glu250 and Cys284, are represented in ball-and-stick. (b) Topology diagram of the Sm-ALDH monomer.

dimeric enzymes have variable non-glycine amino acids at this position. Together with the absence of the glycine, neighbouring aromatic residues Trp80 and Phe418 are found in ALDH3. Their simultaneous presence appears necessary to decrease the L angle in ALDH3 (Table 1) in order to avoid steric hindrance, leading to the dimeric state. These structural characteristics may

be the main determinants of the oligomeric state. However, the origin of the rotation of the oligomerisation domain might involve additional amino acids. Moreover, key amino acids for the oligomeric state may exist at the interfaces between dimers, but their variability within the ALDH family makes their identification difficult. Only new structural data will provide the information needed to establish the precise role of these different amino acids.

Holoenzyme structure and comparison with the apo form

As shown for other ALDHs, the cofactor straddles the Rossmann fold (Liu *et al.*, 1997; Steinmetz *et al.*, 1997; Johansson *et al.*, 1998; Moore *et al.*, 1998). The adenine moiety is situated between α -helices 7 and 8. The nicotinamide moiety is surrounded by the loop connecting the β -strand J to the α -helix 8, the loop containing Cys284 situated between the α -helix 9 and the β -strand M, and the loop connecting the Rossmann fold to the catalytic domain.

Overall conformational changes involving domain movement are not observed in Sm-ALDH upon coenzyme binding. The rmsd value between the apo and the holo form for one monomer is 0.4 Å. However, unlike ALDH2 (Steinmetz *et al.*, 1997) and ALDH9 (Johansson *et al.*, 1998), the main chain at the end of the loop connecting β -strand I to α -helix 7 undergoes a large shift up to 4 Å (Figure 5). The adenine moiety slot is therefore larger in the apo form of Sm-ALDH than in its corresponding holo form and in other ALDHs. This change brings residues 208 to 216 closer to the cofactor, leading to a state quite similar to other ALDHs.

Cofactor stabilisation

The adenine is stabilised by Van der Waals contacts with CH₂ groups of Gly210 and Gly214 which add up to a few hydrogen bonds fixing NADP (Figure 6 and Table 2). The 3' hydroxyl group of the adenosine ribose is bound to the carbonyl group of Ser151.

The shift of the main chain that occurs upon coenzyme binding spreads to the N terminus of helix 7, amplifying its curvature. Therefore, the helix dipole is capable of stabilising the 2' phosphate group of NADP. This ability of the helix to bind anions (phosphate or sulphate) is illustrated by the apoenzyme structure, where a sulphate ion is found at the beginning of the helix (Figure 5). However, the positions of both anions differ by 4.3 Å due to the change in helix curvature when the cofactor binds. In addition to the α -helix effect, the absence of steric hindrance from neighbouring Gly210 creates sufficient space to receive the phosphate group, and to bind it *via* hydrogen bonding with its NH group. The other amino acids involved in the phosphate stabilisation are Thr180 and

Table 1. rmsd values between the Sm-ALDH structure and structures of ALDH2, ALDH3, and ALDH9, and rotation angles between their domains

	Overall structure	Rossmann domain	Oligomerisation domain	Catalytic domain
ALDH2 rmsd (Å)	1.40 (450 residues)	1.12 (108 residues)	1.46 (29 residues)	1.27 (193 residues)
ALDH2 rotation (°)			6.1	3.1
ALDH3 rmsd (Å)	1.69 (372 residues)	1.25 (90 residues)	1.90 (32 residues)	1.48 (175 residues)
ALDH3 rotation (°)			17.4	6.4
ALDH9 rmsd (Å)	1.47 (450 residues)	1.13 (107 residues)	1.39 (30 residues)	1.35 (193 residues)
ALDH9 rotation (°)			14.7	3.7

The structures were superimposed by Modeller (Sali & Blundell, 1993). The superimposition was optimised by Lsq (Jones & Kjeldgaard, 1990). Sm-ALDH structure is taken as the reference for the overall rmsd value calculations by O. The other rmsd value calculations for the catalytic and oligomerisation domains are performed by Lsq taking the Sm-ALDH Rossmann fold as the reference.

Lys177. This association of the dipole of one "phosphate binding helix", one hydrogen bond with a glycine, and one salt bridge (between Lys177 and the phosphate group), has previously been described as the mechanism by which enzymes stabilise the pyrophosphate group of cofactors (Hol, 1985).

In Sm-ALDH, the pyrophosphate group interacts with the main chains of Phe153 (one hydrogen bond with the nicotinamide pyrophosphate) and Ser231 (one hydrogen bond with the adenine pyrophosphate). Ser231 is located in the loop connecting β -strand J with α -helix 8. Like Thr232, it is one of the highly conserved residues of the fingerprint of the dinucleotide binding domain, G_{st}X₂G (described as GX₄G by Liu *et al.*, 1997), characteristic of ALDHs.

As observed in ALDH2 and ALDH9 (Steinmetz *et al.*, 1997; Johansson *et al.*, 1998), the electronic density is weaker around the nicotinamide group (Figure 6(b)) while two conformations are observed in ALDH1 (Moore *et al.*, 1998). In Sm-ALDH, the nicotinamide group is sandwiched between Cys284 and Gly230 through Van der Waals interactions. Moreover, its carboxamide group is maintained by a hydrogen bond with the main-chain carbonyl group of Leu251. Both hydroxyl groups of the nicotinamide mononucleotide ribose are bound to the carboxyl group of the conserved Glu377.

Hydrogen bond network around the nicotinamide moiety

Independently of the above mentioned residues, a large number of hydrogen bonds are observed around the nicotinamide moiety (Figure 7) involving the side-chains and/or the main chain of residues Lys163, Thr229, Thr232, Glu250, Leu251, Gly252, Gly253, Arg283, Cys284, Lys288, Glu377, Arg437 and Ser450.

As noted above, the carbonyl group of Leu251 directly interacts with the carboxamide group of the cofactor. The correct positioning of the main chain in this region is thus critical. However, Leu251 is situated in the Glu250 to Gly253 loop connecting the cofactor binding domain to the cata-

lytic domain, where both successive conserved Gly252 and Gly253 could impart flexibility to the main chain. Since these residues have to be held lightly, two hydrogen bonds are observed, one between the hydroxyl group of residue Thr232 and the NH group of Gly252, and one between the hydroxyl group of conserved Ser450 and the carbonyl group of Gly252. Although the side-chain of Leu251 cannot be involved in hydrogen bonding, its surprising conservation may be due to the steric volume that must be maintained as Leu251 interacts with the conserved Gly235 of the fingerprint GX₄G. Concerning the ribose, the NH group of Gly253 binds to the carboxyl group of the strictly conserved Glu377 and orients its side-chain towards the hydroxyl groups of the ribose of the nicotinamide moiety.

The network continues through the residues Lys163, Thr229, Arg437 and the catalytic Glu250. In holo ALDH2, it has been reported that one water molecule is bound to the side-chain of the catalytic glutamate (Glu268), and that this water molecule can deprotonate the catalytic cysteine, Cys302 (Sheikh *et al.*, 1997). Thus, Glu268 might act as the general base. In the apo form of Sm-ALDH, the same interaction is observed between Glu250 and a water molecule. Moreover, a remarkable interaction is observed between the side-chains of residues Thr229, Glu250 and Arg437. Thr229 appears necessary to maintain Glu250 in such a direction that it can interact with Arg437, which is not conserved in ALDHs, such interactions between two charged residues leading to variations in the pK_a of the partners.

The last part of the network concerns the residues Arg283, Cys284, Lys288 and Glu377. In addition to the interactions mentioned above, Glu377 is bound to the side-chain of Lys288, which interacts with the carbonyl groups of Cys284 and Arg283. Among the three aligned sequences, Arg283 and Lys288 are conserved in NADP-dependent non-phosphorylating GAPDHs. In the structures of the apo and holoenzyme, a sulphate ion interacts with Arg283 (Figure 8). During catalysis, Sm-ALDH oxidises G3P into 3-PGA. Thus, this sulphate ion may correspond to the phosphate group of G3P. It should be noted that another highly

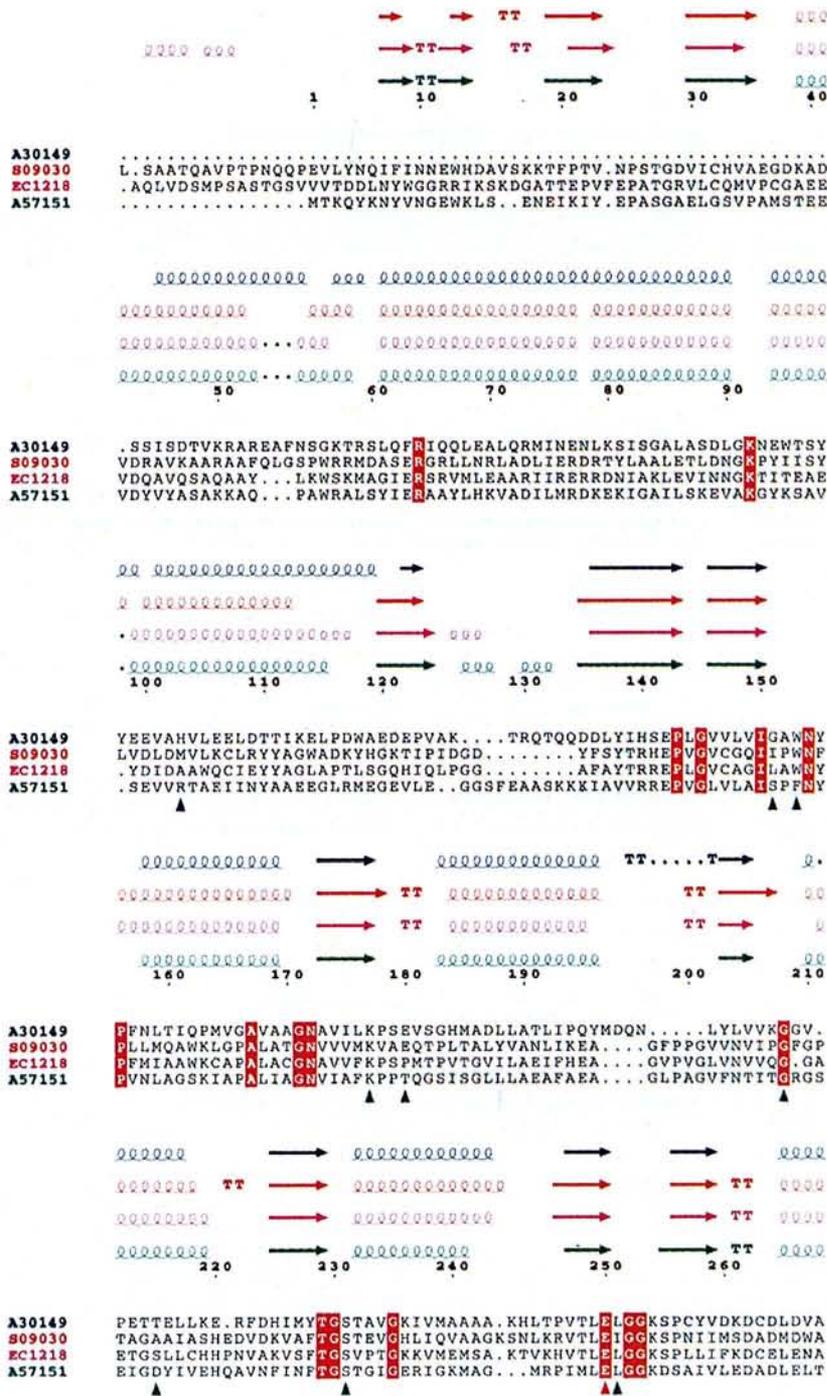


Figure 3 (Legend shown opposite)

conserved arginine in NADP-dependent non-phosphorylating GAPDHs, Arg103, participates in the fixation of the sulphate ion, and therefore in the putative binding of the substrate. The last part of the network may therefore be specific to this class of ALDHs.

Cofactor specificity

At present, three ALDH structures in complex with NAD are available in the Protein Data Bank. As the cofactor reduced during the reaction in Sm-ALDH is NADP, we searched for the structural

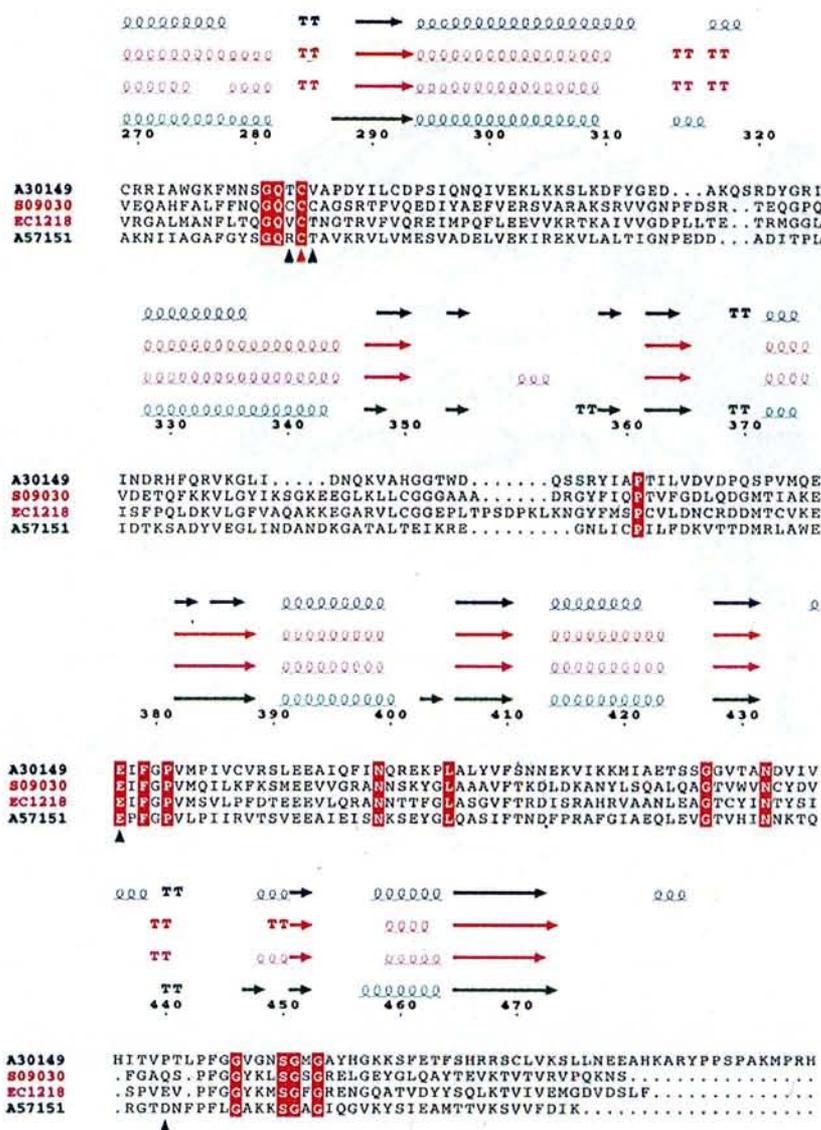


Figure 3. Sequence alignment and secondary structure alignment. The four structures are colour-coded as follows: ALDH2 (red), ALDH3 (blue), ALDH9 (pink), Sm-ALDH (green). β -Strands are represented by arrows, α -helices by curls. The alignment representation of the four sequences and corresponding secondary structures was done with ESPript (Gouet *et al.*, 1998) on the basis of the alignment of the 15 sequences. The residues shown in red are conserved throughout the 15 aligned sequences. The catalytic residues Glu250 and Cys284 are indicated by red triangles. Blue triangles indicate the residues of the anionic pocket situated in the catalytic site and probably bound to the phosphate group of G3P. Green triangles indicate the residues involved in the cofactor binding by hydrogen bonds.

factors responsible for NADP *versus* NAD specificity in ALDHs.

In ALDH9, a conserved lysine (Lys189) has been reported to interact with the 2' and 3' oxygen atoms of the ribose of the adenine moiety (Johansson *et al.*, 1998). In ALDH3, 2' and 3' hydroxyl groups bind to Glu140 (Liu *et al.*, 1997) instead of the conserved lysine, as the cofactor is less buried. However, even in the structures where the lysine is involved, the position of this conserved glutamate remains constant. One exception is observed in

NAD-dependent ALDHs: this residue is mutated to proline in cod liver ALDH.

In Sm-ALDH, the corresponding conserved lysine, Lys177, interacts with two oxygen atoms of the NADP 2' phosphate group including the 2' oxygen of the phosphomonoester bond. Thr180 is found in place of Glu or Pro in NAD-dependent ALDHs. This provides a pocket designed to receive and to bind the phosphate group of NADP, as also illustrated by the binding of a sulphate ion in the apoenzyme structure. The volume required to accept the phosphate group is not sufficient in

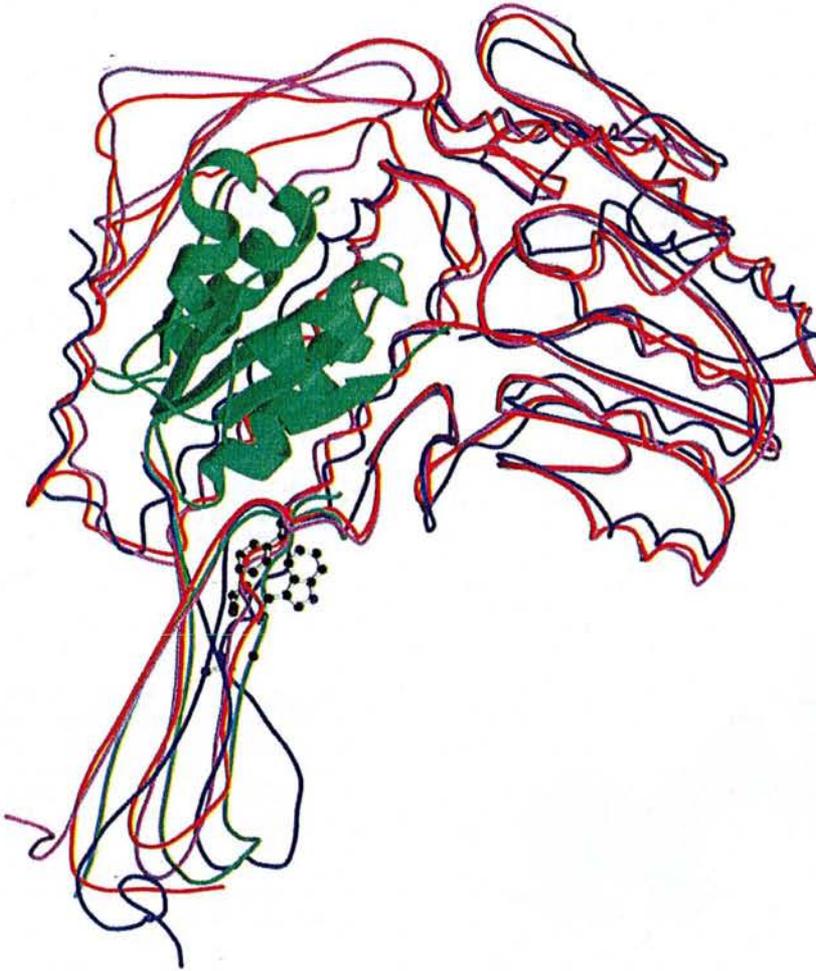


Figure 4. Superimposition of the C^{α} trace of each monomer of the solved ALDH structures, using MOLSCRIPT (Kraulis, 1991) and Raster3D (Meritt & Murphy, 1994). The superimposition of the monomer structures was done with Modeller (Sali & Blundell, 1993) and Lsq from O (Jones & Kjeldgaard, 1990). The Rossmann fold is drawn in ribbon. The monomers of the four structures are colour-coded as follows: ALDH2 (red); ALDH3 (blue); ALDH9 (pink); Sm-ALDH (green). The residues mentioned in the text are represented in ball-and-stick.

NAD-dependent ALDHs due to the presence of bulky residues, as seen in Figure 9. In addition to steric hindrance, the negative charge of a glutamate would prevent the stabilisation of negative charges of the phosphate group.

In Sm-ALDH, Thr180 interacts with the phosphate group through its hydroxyl group and its main-chain NH group. Among the four aligned NADP-dependent non-phosphorylating GAPDHs, the steric constraints and the ability to interact by hydrogen bonds with the phosphate group are

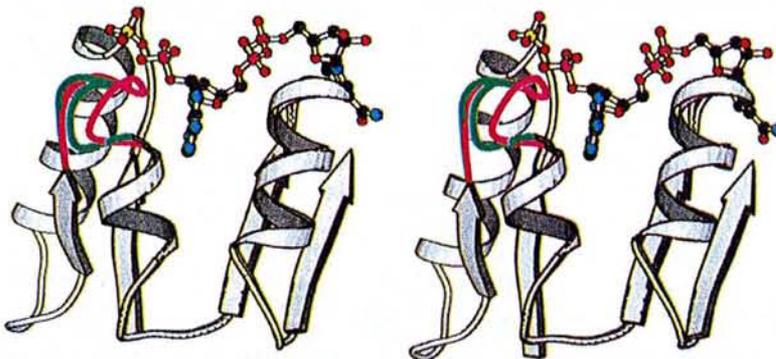


Figure 5. Stereo view of the local conformational change upon coenzyme binding. The Rossmann fold is drawn in ribbon. The green trace corresponds to the apoenzyme and the purple trace to the holoenzyme. The sulphate ion observed in the apoenzyme, and NADP of the holoenzyme are represented in ball-and-stick. The Figure was drawn using MOLSCRIPT (Kraulis, 1991).

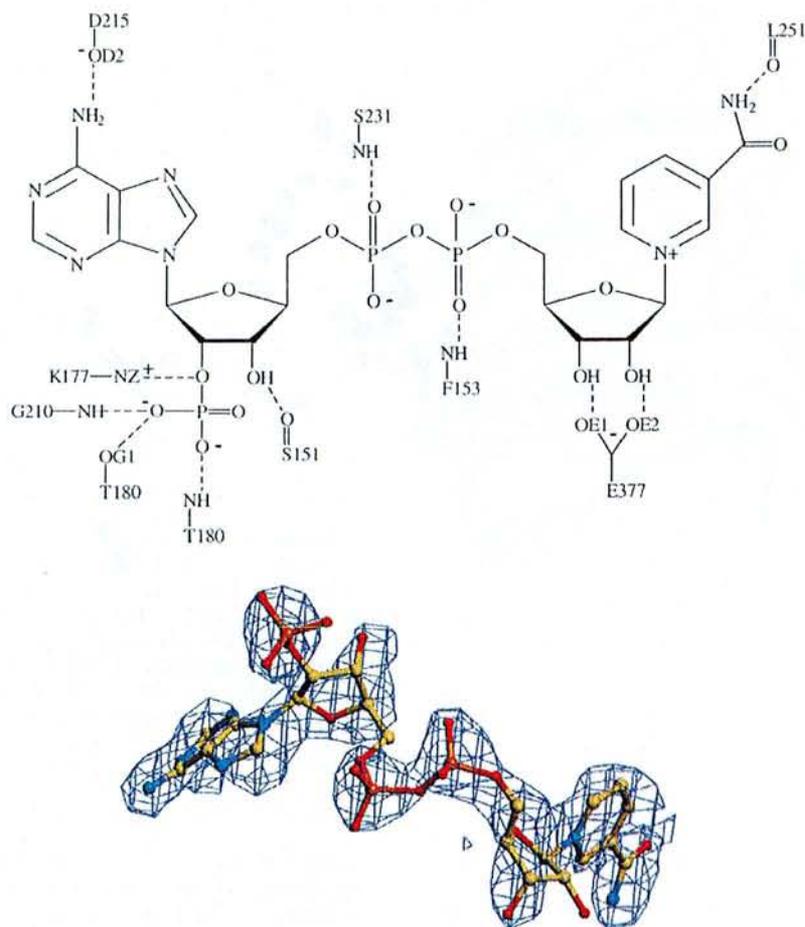


Figure 6. (a) Schematic drawing of the amino acids interacting with NADP by hydrogen bonds. Note that a salt bridge is present between Lys177 and the phosphate group. (b) Representation of NADP in a $3F_o - 2F_c$ electronic density map contoured at 1.2σ generated using TURBO-FRODO (Roussel & Cambillau, 1991).

conserved. Indeed, a threonine or serine is found at this position. Note that the sequence of an NAD-dependent non-phosphorylating GAPDH from *Thermoproteus tenax* (Brunner *et al.*, 1998) possesses an isoleucine at this position, thus displaying the same structural characteristics as ALDH9, which is NAD-dependent. Thus, a threonine or a serine at this position appears to be the structural determinant of cofactor specificity for NADP in ALDHs.

Substrate site, substrate specificity and comparison with phosphorylating GAPDHs

The localisation of the active site in this structure is facilitated by the presence of a sulphate ion at the surface of the protein at 7.9 Å from the thiolate group of the catalytic cysteine. This sulphate ion may correspond to the phosphate group of G3P (Figure 8). Moreover, the entrance of the site is situated near the interfaces inside the tetramer, as seen in ALDH2 (Steinmetz *et al.*, 1997).

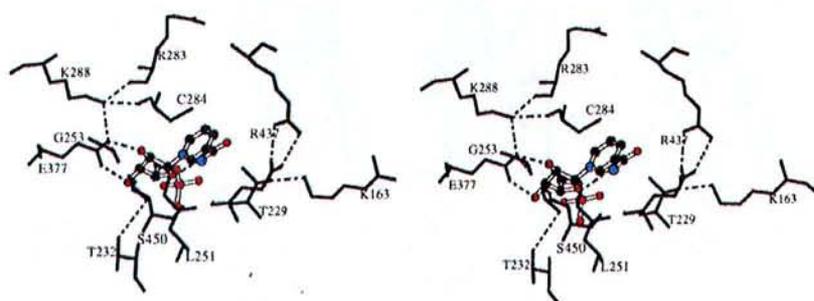


Figure 7. Stereo view of the network of hydrogen bonds involving amino acids around the cofactor and the catalytic site, using MOLSCRIPT (Kraulis, 1991). For clarity, the bonds between the residues are not drawn, and only the nicotinamide moiety is displayed and only a few residues are labelled.

Table 2. Hydrogen bonds between the cofactor and neighbouring residues

Donors	Acceptors
Phe153 N	NO2
Lys177NZ	AO2*
Thr180 N	AOP2
Thr180 OG1	AOP3
Gly210 N	AOP3
Ser231 N	AO1
AO3*	Ser151 O
AN6	Asp215 OD2
NN7	Leu251 O
NO2*	Glu377 OE2
NO3*	Glu377 OE1

The residues conserved in the sequence alignment are shown in bold faced letters.

In contrast to the solved NAD-dependent ALDH structures, the site is essentially composed of hydrophilic residues interacting with the substrate, which has been modelled in the holoenzyme to form a ternary complex (Figure 8). In NADP-dependent non-phosphorylating GAPDHs, three residues are conserved: Thr285, Arg283 and Arg103, the latter being highly conserved. As noted above for Sm-ALDH, Arg103 and Arg283 interact with a sulphate ion in the holo and apoenzyme. In this hypothetical ternary complex, these positively charged residues constitute an anionic pocket and interact with the phosphate group of the substrate for its stabilisation. Two hydrogen bonds occur with Arg283 and one with Arg103. In the apo- and holoenzyme, the latter arginine interacts by hydrogen bonding with Asp440, which is highly conserved in this ALDH class. This interaction directs the arginine toward the sulphate ion and probably toward the phosphate group of G3P. Note that the phosphate group of the substrate is slightly shifted in relation to the observed sulphate ion. The carbonyl group of carbon 1 of G3P may be stabilised by the Cys284 nitrogen group.

As the substrate oxidised by Sm-ALDH during the catalytic reaction, G3P, is the same as that used by NAD-dependent GAPDHs and is transformed

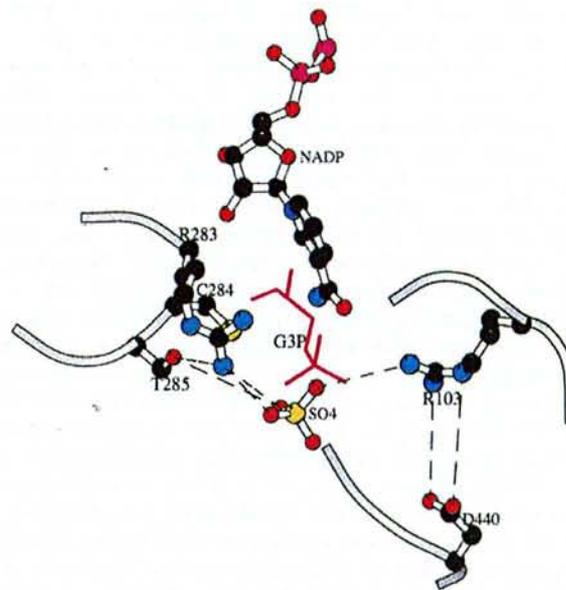


Figure 8. Modelling of G3P in the catalytic site using MOLSCRIPT (Kraulis, 1991). The stabilising residues, the sulphate ion, and the nicotinamide moiety of NADP are represented in ball-and-stick, and G3P by its trace. Hydrogen bonds are drawn in broken lines.

into the same thioacyl intermediate, we compared the catalytic sites of both enzymes, although the folds of the catalytic domains and the structure of their catalytic sites are different. In the structures of NAD-dependent phosphorylating GAPDHs, two arginine residues (Arg195 and the conserved Arg231) stabilise a sulphate ion present in a site called P_s, located 6.9 Å from the catalytic cysteine Cys149 in *Bacillus stearothermophilus* GAPDH. In the catalytic site of Sm-ALDH, both arginine residues Arg103 and Arg283 also define an anionic pocket and interact with a sulphate ion located 7.9 Å from Cys284. Moreover, a conserved aspartic acid, Asp440, is in close contact with Arg103. In agreement with the sequence alignment, Arg103,

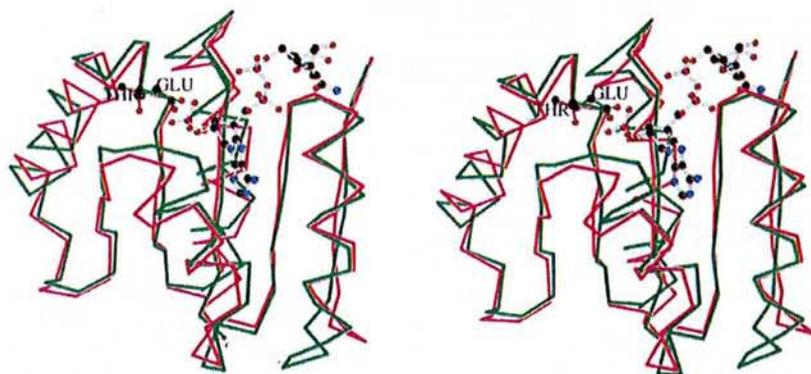


Figure 9. Stereo view of cofactor specificity. The ALDH2 Rossmann fold coloured in purple is superimposed onto the Sm-ALDH Rossmann fold coloured in green, interacting with NADP represented in ball-and-stick. Glu195 of ALDH2 and Thr180 of Sm-ALDH are represented in ball-and-stick. The Figure was prepared using MOLSCRIPT (Kraulis, 1991).

Arg283 and Asp440 may be specific to NADP-dependent ALDHs, which oxidise G3P to 3-PGA. In *B. stearothermophilus* GAPDH, Asp181 interacts with Arg195, the arginine most distant from Cys149, in the same way that Asp440 in Sm-ALDH interacts with Arg103, the arginine most distant from Cys284. The residue Asp181 might stabilise Arg195 in GAPDH to direct the guanidinium group toward the sulphate ion of the P_s site, and so toward the phosphate group of G3P. Therefore, these three residues may play a similar role in NAD-dependent phosphorylating GAPDHs and in NADP-dependent non-phosphorylating GAPDHs although these enzymes possess no sequence identity. Another site named P_i is also observed in GAPDHs. This site may bind the inorganic phosphate group in NAD-dependent phosphorylating GAPDHs (Moras *et al.*, 1975). It has no equivalent in Sm-ALDH, since Sm-ALDH does not phosphorylate G3P. Thus, the binding of G3P in Sm-ALDH may be similar to that proposed by Moras *et al.* (1975), who postulate that the fixation of G3P involves the anionic P_s site in NAD-dependent phosphorylating GAPDHs. More recently, Skarzynsky *et al.* (1987) proposed as an alternative hypothesis that the initial fixation of G3P involves the P_i site for stabilisation of its phosphate group in the first step of the reaction.

Conclusion

The analysis of the apo and holo structures of an ALDH isolated from *S. mutans* has provided the first crystallographic data for an NADP-dependent ALDH. Several residues were found to be critical determinants of the oligomeric state, although other residues may also play a role. Identification of these latter amino acids should be possible through analysis of additional structural data, for example retinal dehydrogenase and ALDH1 from sheep liver deposited in the Protein Data Bank but not available at the time of writing.

As shown in the previously determined ALDH structures, there is no overall conformational change between the holo- and apoenzyme structures of the *S. mutans* enzyme. The absence of conformational change upon coenzyme binding may be related to the experimental conditions used. A conformational change might occur under certain conditions of pH and ionic strength, leading to movement of the main chain, or involving only a few side-chains of the enzyme, the holoenzyme keeping the same global structure.

A large local movement of the main chain, not reported in the structures of NAD-dependent ALDHs, stabilises the 2' phosphate group of NADP when the enzyme binds the cofactor. The specificity for NADP in relation to the sequence alignment defines a class of ALDHs containing NADP-dependent non-phosphorylating GAPDHs. NAD-dependent non-phosphorylating GAPDHs may form another class in the ALDH family.

Table 3. Summary of crystallographic data

Data sets	Apoenzyme	Holoenzyme
	1.8 ($\lambda = 0.9 \text{ \AA}$; W32 LURE)	
Resolution (\AA)	100	2.6 ($\lambda = 1.5418 \text{ \AA}$)
Temperature (K)	100	100
Space group	$P2_12_12$	$P2_12_12$
Unit cell (\AA)	144.4 154.7 114.0	144.8 154.8 114.3
Unique reflections	215,812	65,491
R_{sym} (%)	6.1	8.7
(last shell)	20.6	31.5
Completeness (%)	94.9	82.6
Multiplicity	3.5	1.5
Mean $I/\sigma(I)$	11.7	7

$R_{\text{sym}} = \sum |I_i - \langle I \rangle| / \sum (I_i)$, where i is the i th measurement and $\langle I \rangle$ is the weighted mean of I .

However, the sequence of only one NAD-dependent non-phosphorylating GAPDH from *Thermoproteus tenax* has so far been determined (Brunner *et al.*, 1998).

A hydrogen bond network surrounding the cofactor was also found. It comprises two parts, one probably common to all ALDHs, and the second specific to this particular ALDH due to the nature of the substrate. The substrate specificity reveals the existence of an anionic pocket in this class of ALDH, as in GAPDHs. Modelling of G3P binding in the catalytic site of Sm-ALDH is similar to the binding mode of G3P during the first step of the catalytic reaction in GAPDHs proposed by Moras *et al.* (1975), with the participation of an anionic pocket. Attempts to solve a crystal structure of a ternary complex between G3P and the holoenzyme where the catalytic cysteine Cys284 was mutated to a serine residue will be undertaken in order to confirm our model.

Table 4. Summary of the crystallographic refinements and model statistics

	Apoenzyme	Holoenzyme
R_{cryst} (%)	22.1 (8.0-1.8 \AA ; $\sigma = 0$)	23.5 (8.0-2.6 \AA ; $\sigma = 0$)
R_{free} (%)	25.6 (8.0-1.8 \AA ; $\sigma = 0$)	28.8 (8.0-2.6 \AA ; $\sigma = 0$)
Bond lengths rmsd (\AA)	0.006	0.008
Bond angles rmsd (deg.)	1.307	2.439
Residues in most favourable region of the Ramachandran plot (%)	92.1	92.2
Residue in the disallowed region of the Ramachandran plot (%)	0.0	0.0
Average B -factor (\AA^2)		
All atoms	13.2	4.9
Protein atoms	12.2	4.6
Water molecules	22.1	12.9
NADP	-	9.9

$R_{\text{cryst}} = \sum |F_{\text{obs}}| - |F_{\text{calc}}| / \sum |F_{\text{obs}}|$. R_{free} (Brünger, 1992) is the same as R_{cryst} , but calculated for $n\%$ of the data omitted from refinements where n is 9.6 for the apoenzyme and 8.9 for holoenzyme.

Material and Methods

Crystallisation, data collection and processing

The protein from *S. mutans* was expressed in *Escherichia coli* and purified as described, but without CaCl_2 (S.M. & G.B., unpublished results). The apoenzyme crystals were grown at 20 °C by the vapour diffusion method in hanging drops from 6 μl of solution containing 2.0 M ammonium sulphate, 2% (w/v) polyethylene glycol 400, 0.1 M Hepes (pH 7.5) and 2 μl of protein at 40 mg/ml in phosphate buffer at 50 mM, pH 6.5. The crystals grew within a few days. A crystal of approximately 1.0 mm \times 0.3 mm \times 0.2 mm was used for data collection at 1.8 Å on the beam line W32 at LURE ($\lambda = 0.96$ Å) at 100 K with 20% ethylene glycol as the cryogenic solution. The data were processed with DENZO and SCALEPACK packages (Otwinowski & Minor, 1997). The holoenzyme crystals were grown in the same conditions in the presence of a large excess of NADP. The diffraction data were collected at 100 K, with the same cryogenic solution as the apoenzyme, using an image plate detector and an Enraf Nonius rotating anode with a wavelength of 1.5418 Å (Table 3).

Molecular replacement, model building and refinement

The apoenzyme structure was solved by molecular replacement using AMoRe (Navaza, 1994) with the model of the ALDH19 tetramer (Johansson *et al.*, 1998) in which more than 300 amino acid residues per monomer were replaced by alanine according to the sequence alignment. Molecular replacement calculations were carried out using data from 9 to 6 Å resolution. The *R* factor and the correlation coefficient for the unique solution are 46.5% and 53.7%, respectively. The model was submitted to rigid body refinement using X-PLOR (Brünger, 1990) followed by several cycles of energy minimisation and a few cycles of simulated annealing. In the final stages of the refinement, the non-crystallographic symmetry constraints were released. The model was rebuilt in $3F_o - 2F_c$ and $F_o - F_c$ electron density maps using TURBO-FRODO (Roussel & Cambillau, 1991) and O (Jones & Kjeldgaard, 1990). The holoenzyme structure was solved by molecular replacement using the apoenzyme structure. The procedure was identical to that described above except that the non-crystallographic symmetry constraints were maintained during the refinement (Table 4).

Structure analysis and final structure parameters

The refined apoenzyme structure contains 1896 residues, 1542 water molecules, and ten sulphate ions clearly identified in the electronic density map. The stereochemical quality of the structure was analysed with programs PROCHECK (Laskowsky *et al.*, 1993) and X-PLOR (Brünger, 1990). The rmsd deviations from ideal values are: bond lengths, 0.006 Å; bond angles, 1.307°; dihedral angles, 23.295°; improper angles, 1.233°.

The refined structure of the holoenzyme contains 469 water molecules and five sulphate ions. The rmsd deviations from ideal values are: bond lengths, 0.008 Å; bond angles, 2.439°; dihedral angles, 25.382°; improper angles, 0.946°.

Sequence alignment and superimposition of structures

The sequence alignment was performed with the GCG package, using 42 sequences of the Protein Identification Resource (Sidman *et al.*, 1998). The sequence alignment was optimised on 15 sequences representative of the solved structures. A total of 14 sequences with the following accession numbers were retrieved from the Protein Identification Resource: *Rattus norvegicus* (A30149); *Rattus norvegicus* (A41028); *Homo sapiens* (A42584); *Pseudomonas oleovorans* (D31266); *Mus musculus* (I48966); *Rattus norvegicus* (S03564); *Mus* species (I57960); *Equus caballus* (S00364); *Bos primigenius taurus* (S09030); *Homo sapiens* (DEIHUE2); *E. coli* (S15181); *Pisum sativum* (S43832); *Zea mays* (S43833); *S. mutans* (A57151). The secondary structures were delineated using DSSP (Kabsh & Sander, 1983).

Protein Data Bank accession numbers

The coordinates of both apo and holo structures have been deposited in the Brookhaven Protein Data Bank (PDB entry 1euh and 2euh, respectively).

Acknowledgements

We are grateful to Professor Vitoux for helpful discussions. We thank Dr Gouet for providing the ESPript package and J.P. Mage from the CURRI Visualisation group at ENSPS in Strasbourg for kind assistance with the drawings. We thank the LURE for providing access to the facilities of the beam line W32. This work was technically supported by the Service Commun de Diffraction X sur Monocristaux at Henri Poincaré University.

References

- Bergdoll, M., Remy, M.-H., Cagnon, C., Masson, J.-M. & Dumas, P. (1997). Proline dependent oligomerization with arm exchange. *Structure*, **5**, 391-401.
- Boyd, D. A., Cvitkovitch, D. G. & Hamilton, I. R. (1995). Sequence, expression, and function of the gene for the non phosphorylating, NADP-dependent glyceraldehyde-3-phosphate dehydrogenase of *Streptococcus mutans*. *J. Bacteriol.* **177**, 2622-2627.
- Brünger, A. T. (1990). *X-PLOR Manual*, Yale University Press, New Haven, CT.
- Brünger, A. T. (1992). The R_{free} value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.
- Brunner, N. A., Brinkmann, H., Siebers, B. & Hensel, R. (1998). NAD⁺-dependent glyceraldehyde-3-phosphate dehydrogenase from *Thermoproteus tenax*. *J. Biol. Chem.* **273**, 6149-6156.
- Crow, V. L. & Wittenberger, C. L. (1979). Separation and properties of NAD⁺- and NADP⁺-dependent glyceraldehyde-3-phosphate dehydrogenases from *Streptococcus mutans*. *J. Biol. Chem.* **254**, 1134-1142.
- Gouet, P., Courcelle, E., Stuart, D. & Metz, F. (1999). SPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305-308.
- Habenicht, A., Hellman, U. & Cerff, R. (1994). Non-phosphorylating GAPDH of higher plants is a member of the aldehyde dehydrogenase superfamily.

- ily with no sequence homology with the phosphorylating GAPDH. *J. Mol. Biol.* **237**, 165-171.
- Hol, W. G. J. (1985). The role of the α -helix dipole in protein function and structure. *Prog. Biophys. Mol. Biol.* **45**, 149-195.
- Johansson, K., El-Ahmad, M., Ramaswamy, S., Hjelmqvist, L., Jörnvall, H. & Eklund, H. (1998). Structure of betaine aldehyde dehydrogenase at 2.1 Å resolution. *Protein Sci.* **7**, 2106-2117.
- Jones, T. A. & Kjeldgaard, M. (1990). "O" Manual, version 6.1, Uppsala University, Sweden and Aarhus University, Denmark.
- Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.
- Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
- Laskowsky, 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.
- Liu, Z. L., Sun, Y.-J., Rose, J., Chung, Y. J., Hsiao, C.-D., Chang, W.-R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J. & Wang, B.-C. (1997). The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nature Struct. Biol.* **4**, 317-326.
- Meritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2. 4-A program for photo-realistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869-873.
- Moore, S. A., Baker, H. M., Blythe, T. J., Kitson, K. E., Kitson, T. M. & Baker, E. N. (1998). Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. *Structure*, **6**, 1541-1551.
- Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C. & Rossmann, M. G. (1975). Studies of asymmetry in the three-dimensional structure of lobster D-glyceraldehyde-3-phosphate dehydrogenase. *J. Biol. Chem.* **250**, 9137-9162.
- Navaza, J. (1994). AMoRe: An automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157-163.
- Otwinowsky, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
- Rossmann, M. G., Moras, D. & Olsen, K. W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *Nature*, **250**, 194-199.
- Roussel, P. A. & Cambillau, C. (1991). *TURBO-FRODO, Silicon Graphics applications directory*, Silicon Graphics, Mountain View, CA.
- Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815.
- Sheikh, S., Ni, L., Hurley, T. D. & Weiner, H. (1997). The potential roles of the conserved amino acids in human liver mitochondrial aldehyde dehydrogenase. *J. Biol. Chem.* **272**, 18817-18822.
- Sidman, K. E., George, D. G., Barker, W. C. & Hunt, L. T. (1998). The protein identification resource (PIR). *Nucl. Acids Res.* **16**, 1869-1871.
- Skarzynsky, T., Moody, P. C. E. & Wonacott, A. J. (1987). Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. *J. Mol. Biol.* **193**, 171-187.
- Steinmetz, C. G., Xie, P., Weiner, H. & Hurley, D. T. (1997). Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. *Structure*, **5**, 701-711.
- Wisconsin Package Version 9.0-UNIX (December 1996). Genetics Computer Group (GCG), Madison, WI.
- Yoshida, A., Rzhetsky, A., Hsu, L. C. & Chang, C. (1998). Human aldehyde dehydrogenase gene family. *Eur. J. Biochem.* **251**, 549-557.

Edited by R. Huber

(Received 11 January 1999; received in revised form 3 May 1999; accepted 3 May 1999)

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

Le mécanisme catalytique des ALDH

A partir des données cristallographiques et biochimiques concernant les ALDH NAD dépendantes et l'ALDH NADP dépendante de *S. mutans*, il est maintenant possible de définir un mécanisme catalytique pouvant être commun uniquement aux ALDH pour lesquelles l'étape limitante de la réaction est la déacylation. Pour L'ALDH3 présentée au début de ce manuscrit, l'étape limitante de la réaction est le transfert d'hydrure et l'acide aminé jouant le rôle de base (Glu250 chez Sm-ALDH) serait le glutamate 333 (Mann & Weiner, 1999) qui correspond à Glu377 chez Sm-ALDH (Glu377 interagit avec les groupements hydroxyle du ribose du nicotinamide).

I- Les acteurs de la réaction

Conservés dans la plupart des séquences d'ALDH, la cystéine, le glutamate et une asparagine jouent les rôles de nucléophile, de base, et pour le dernier, stabiliserait différents intermédiaires de la réaction. Les études d'enzymologie ont mis en évidence deux formes de la cystéine par la tiration de son pKa et un pKa correspondant au glutamate, en fonction de la présence ou non de deux autres acteurs de la réaction: le substrat et le cofacteur (Marchal & Branlant, 1999). L'oxydation de l'aldéhyde en acide nécessite le dernier protagoniste: une molécule d'eau. L'enchaînement réactionnel est alors le suivant: fixation du cofacteur, du

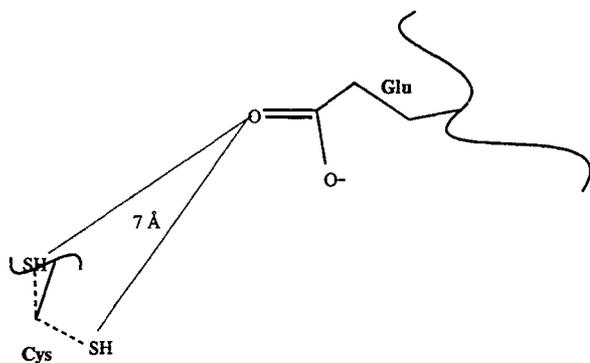
substrat; étape d'acylation et transfert d'hydrure; étape de déacylation: attaque de la molécule d'eau et relargage de l'acide et du cofacteur réduit.

II- Rappel des mécanismes proposés

La participation d'un nucléophile comme la cystéine à une réaction enzymatique nécessite son accessibilité et une baisse de son pKa afin que la réaction se produise à pH physiologique. La formation de l'ion thiolate pourrait alors être assurée par une molécule d'eau activée par le glutamate, ou bien par la présence des dipôles de la chaîne principale et par la charge positive du nicotinamide (Steinmetz *et al.*, 1997). L'attaque nucléophile se produit alors sur la fonction aldéhydique du substrat et la formation de l'intermédiaire tétraédrique précède le transfert d'hydrure de type pro-R. La charge négative de l'intermédiaire tétraédrique est stabilisée par des interactions échangées avec le site oxyanion constitué du groupement amide d'une asparagine conservée et du NH de la cystéine essentielle. L'étape suivante, la déacylation, assurée par une molécule d'eau, nécessiterait des mouvements du cofacteur et du glutamate lui permettant de prendre place en vue de son activation (Moore *et al.*, 1998). Puis, les produits de la réaction, l'acide et le NAD(P)H, sont relargués.

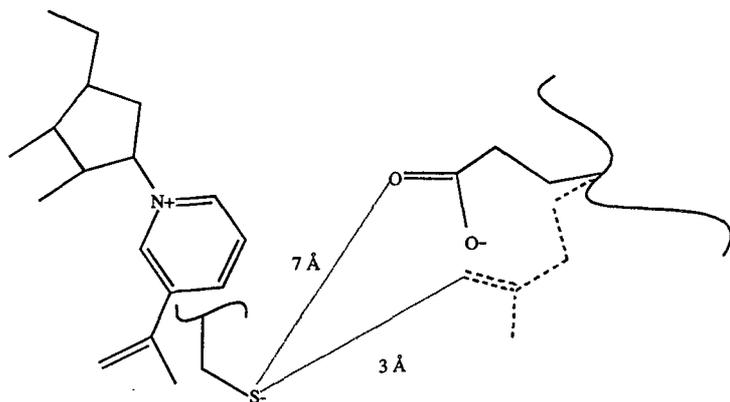
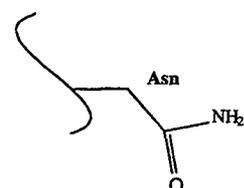
III- Les résultats de la cristallographie et de l'enzymologie: cas de l'ALDH de *S. mutans*

Le schéma du mécanisme catalytique proposé dans les sept dessins suivants est réalisé en se basant sur les structures de l'Apo1, de l'Holo1, de l'Apo2 et du complexe ternaire. Pour chaque étape de l'acte catalytique, les faits les plus importants sont résumés. L'ensemble du mécanisme est décrit dans les pages suivantes.



Apoenzyme: le carboxylate du glutamate se situe à 7 Å de la cystéine. Les orientations différentes de la chaîne latérale de la cystéine observées entre les structures Apo1 et Apo2 d'une part, et les structures apoenzyme des ALDHs 2 et 9, d'autre part, sont représentées en traits pointillés.

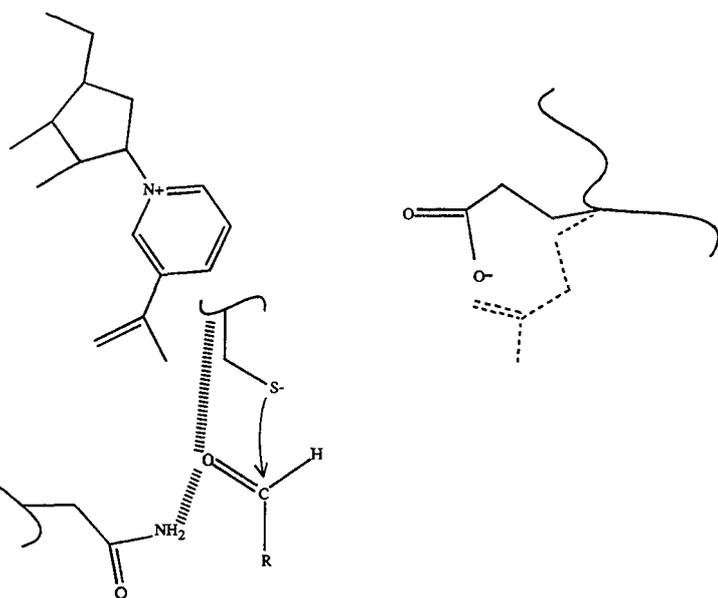
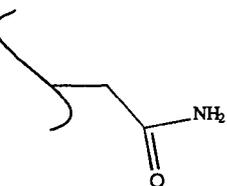
pKa Cys: 8,5 (Marchal & Branlant, 1999)



Holoenzyme: sous l'effet de la fixation du cofacteur, la chaîne latérale de la cystéine change d'orientation. En solution, un second pKa pouvant être celui de Glu250, dont le carboxylate se situerait à moins de 4 Å de Cys284 (Marchal & Branlant, 1999), est mesuré. Cette seconde orientation est représentée en traits pointillés.

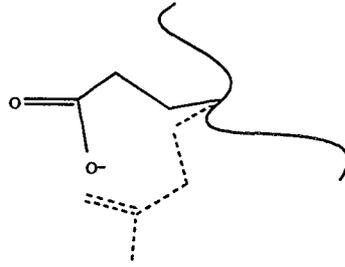
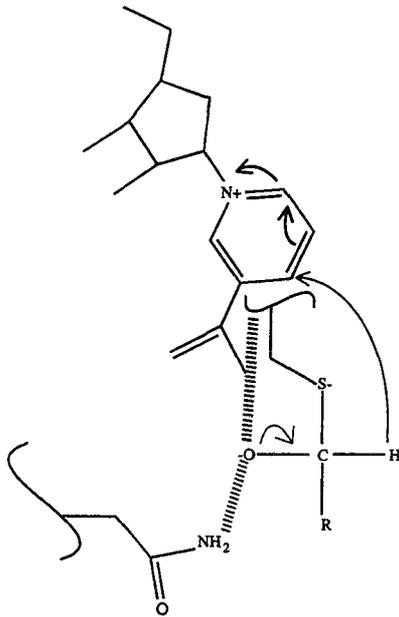
pKa Cys: 6,1 (Marchal & Branlant, 1999)

pKa Glu: 7,6 (Marchal *et al.*, 2000)

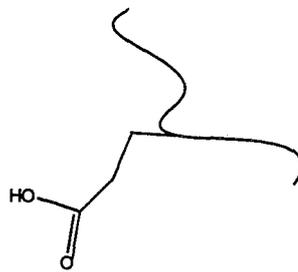
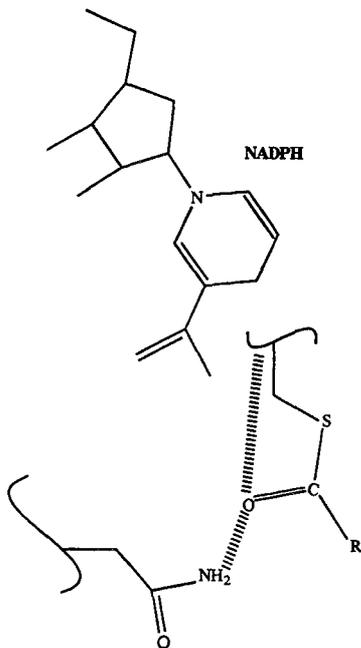


Fixation du substrat: étape d'acylation.

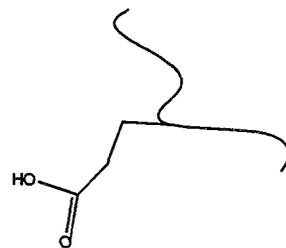
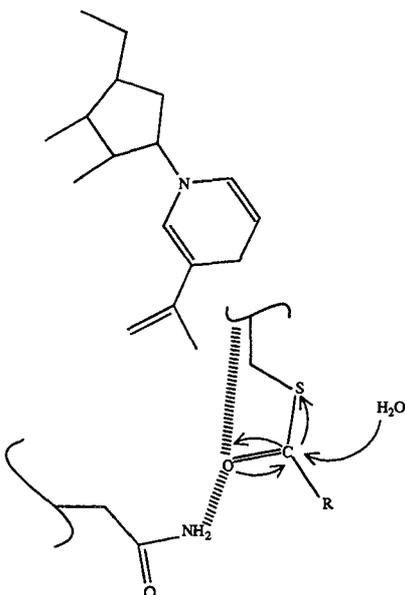
Attaque nucléophile de la cystéine sur la fonction aldéhydique du substrat. L'orientation de la chaîne latérale du glutamate correspondant à celle observée dans les structures du complexe ternaire C284S-NADP-G3P et du mutant C284S est représentée en traits pointillés. Compte tenu de l'effet non négligeable de la mutation sur ce mouvement (Cobessi *et al.*, 2000), l'orientation observée dans l'holoenzyme est également représentée en traits pleins.



Formation de l'intermédiaire tétraédrique et transfert d'hydrure. L'hydrure est transféré en C4 du nicotinamide alors que la charge négative portée par l'intermédiaire tétraédrique est stabilisée par des liaisons hydrogène avec le NH de la cystéine et l'amide de l'asparagine (site oxyanion).



Acylenzyme et cofacteur réduit. Changement de conformation de la partie nicotinamide mononucléotide du cofacteur. Le glutamate peut s'approcher à 3 Å de la cystéine.



Etape de déacylation. Attaque de la molécule d'eau positionnée et activée par le glutamate (Marchal *et al.*, 2000).

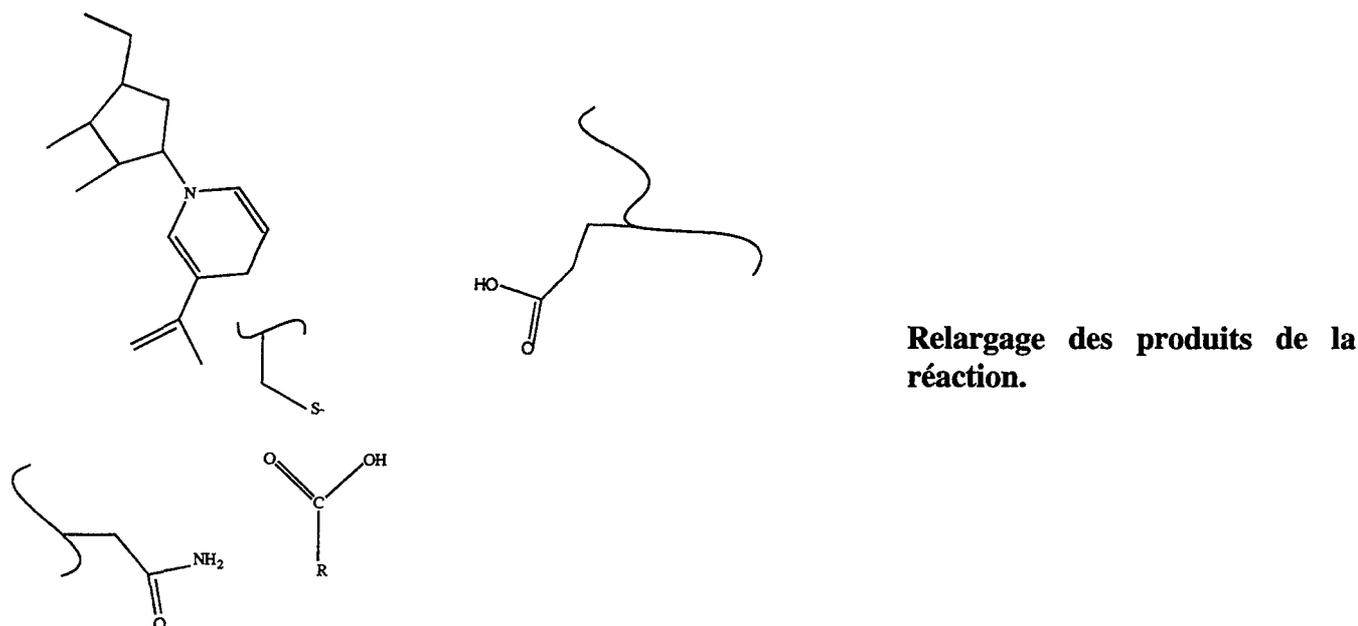


Figure 7.1: schéma de la réaction catalytique catalysée par l'ALDH de *S. mutans*.

A-Apoenzyme, holoenzyme, activation de la cystéine

La comparaison des structures apo et holo des ALDH NAD dépendantes 2 et 9 montre que l'orientation de la cystéine catalytique n'est pas affectée par la fixation du cofacteur. Sa chaîne latérale est déjà dirigée vers le site catalytique dans l'apoenzyme et la cystéine pourrait être sous sa forme thiolate. Dans le cas de l'ALDH de *S. mutans*, deux orientations de la chaîne latérale de la cystéine sont observées. Ainsi, son changement d'orientation lors de la fixation du NADP la positionne dans un environnement protéique différent, pouvant entraîner une baisse de son pKa qui passe de 8,5 à 6,1 (Marchal & Branlant, 1999). Couplé à ce changement de conformation, la chaîne principale de l'ALDH de *S. mutans* dans la zone 208-215 subit un déplacement (4,1 Å de différence entre les positions du C α de Ser211). Ce phénomène, lié à la fixation du cofacteur NADP, pourrait être similaire chez toutes les ALDH NADP dépendantes. Les résultats de l'enzymologie montrent également qu'un second pKa de 7,6 apparaît lors de la fixation du cofacteur. Il est attribué au glutamate Glu250 (Marchal *et al.*, 2000). L'observation de la structure de l'holoenzyme montre qu'aucun changement dans l'orientation de sa chaîne latérale n'intervient lors de la fixation du NADP. Cette divergence

entre la cristallographie et l'enzymologie provient probablement des conditions expérimentales qui diffèrent. En effet, la mesure des pKas dans le complexe binaire est réalisée après incubation de l'enzyme à 4°C, à pH acide et en présence d'un large excès de NADP. Quant à la cristallographie, la protéine est cristallisée à un pH proche de 8,0, et à 20°C. Ainsi, par cette dernière méthode, seule la baisse du pKa de la cystéine dans le complexe binaire peut être interprétée. Il est à noter que la valeur du pKa attribuée à Glu250 est également mesurée à 20°C mais dans le complexe ternaire.

L'étude du mutant E250A a montré que l'étape d'acylation n'était pratiquement pas affectée et a permis d'attribuer un pKa de 7,5 à Glu250 (Marchal *et al.*, 2000). Ainsi, par les résultats de la cristallographie et de l'enzymologie il est possible d'affirmer que le glutamate ne participe pas à la première étape de la réaction catalytique et donc, qu'aucune molécule d'eau n'est nécessaire à l'activation de la cystéine. La baisse de son pKa résulte pour une grande partie (2 unités) de la réorientation de sa chaîne latérale à proximité des groupements NH (Cys284 et Thr285) de la chaîne principale et dans une moindre mesure de la charge positive du cycle du nicotinamide (0,7 unité, Marchal & Branlant, 1999). Donc, à pH physiologique, la forme thiolate peut exister et la réaction se produire.

B- Le complexe ternaire et l'étape d'acylation

L'acylation de l'enzyme se produit à la suite de l'attaque nucléophile de la cystéine sur le carbone du substrat portant la fonction aldéhydique. L'intermédiaire tétraédrique précédant la formation de l'hémithioacétal est alors formé. Le passage de l'intermédiaire tétraédrique à l'hémithioacétal résulte du transfert d'hydrure entre le substrat et le cofacteur. Le transfert est de type Pro-*R* chez les ALDH puisque l'orientation du nicotinamide est *anti*, excepté *a priori* dans le cas de l'ALDH3. Dans ce cas, le nicotinamide mononucléoside est observé dans une conformation *syn* avec un positionnement incompatible pour le transfert d'hydrure en raison de la distance trop importante entre le thiolate et le carbone C4 du nicotinamide.

L'analyse de la structure de l'Apo2 montre des interactions entre l'un de ses ions sulfate (SO₄^b) et le groupement NH de Cys284 et la fonction amide de Asn154, très conservée chez les ALDH. Asn154 participe à la stabilisation de l'intermédiaire tétraédrique.

Dans le complexe ternaire, elle interagit avec l'oxygène du carbone 1 du G3P au même titre que le groupement NH de Cys284. Ces constatations confirment le rôle supposé de l'asparagine. Ces différentes observations sont renforcées par les études de mutagenèse dirigée. La mutation de l'asparagine en thréonine ou alanine a pour effet de transformer l'étape d'acylation en étape limitante alors que la déacylation l'était auparavant (Cobessi *et al.*, 2000). L'étude de la structure de l'holoenzyme C284S montre une distance de 3 Å entre l'un des atomes d'oxygène du carboxylate de Glu250 et l'hydroxyle de Ser284. Cette distance est également retrouvée dans le complexe ternaire. Cette variation de la position de Glu250 est à associer aux trois conformations modélisées du cofacteur: une pour l'holoenzyme, une pour C284S et enfin une dernière pour le complexe ternaire. En plus de ces différentes conformations, la densité électronique de la partie nicotinamide mononucléoside est plus faiblement définie dans la structure de l'holoenzyme en comparaison avec la partie ADP. La densité électronique de la partie nicotinamide mononucléotide ou mononucléoside est très mal définie dans les structures du complexe ternaire et du mutant C284S. Ces différents faits sont en faveur d'une flexibilité de la partie nicotinamide mononucléotide ou mononucléoside nécessaire à l'acte catalytique qui serait couplée à la flexibilité de la chaîne latérale de Glu250. Cette flexibilité du cofacteur est également suggérée par les études de quenching de fluorescence (Marchal & Branlant, 1999) en utilisant comme sonde, le tryptophane W375 dont la position ne varie pas quelles que soient les structures.

C- La déacylation

L'oxydation irréversible de la forme aldéhydique du substrat en son produit de forme acide nécessite l'intervention d'une molécule d'eau chez les GAPDH NADP dépendantes non phosphorylantes contrairement aux GAPDH NAD(P) dépendantes phosphorylantes où un phosphate inorganique est impliqué dans le processus d'oxydo-réduction du G3P .

Chez les ALDH, la molécule d'eau est activée par le glutamate catalytique. Les études d'enzymologie sur l'ALDH de *S. mutans* font apparaître deux rôles pour Glu250. D'une part, il active la molécule d'eau puisque sa mutation en Gln conduit à une diminution de la vitesse de la réaction. D'autre part, il doit contribuer au positionnement de celle-ci, augmentant ainsi l'efficacité de la réaction (Marchal *et al.*, 2000).

Ensuite, le 3-phosphoglycérate puis le NADPH sont relargués de l'enzyme.

IV- Généralisation du mécanisme à l'ensemble des ALDH hors ALDH3

Si les ALDH peuvent être réparties en différentes classes en fonction des pourcentages d'identité de séquences qu'elles partagent, une seconde classification en fonction de l'étape limitante de la réaction peut également être proposée (Wang & Weiner, 1995; Mann & Weiner, 1999). Elle comporte trois classes:

- classe 1: étape limitante: la déacylation (relargage du cofacteur réduit)
- classe 2: étape limitante: la déacylation (hydrolyse du thioacylenzyme)
- classe 3: étape limitante: l'acylation (le transfert d'hydrure)

Les structures tridimensionnelles des enzymes des deux premières classes sont superposables avec de faibles valeurs rms et leur analyse a révélé plusieurs points communs: la distance entre le glutamate et la cystéine se situe en moyenne à 7 Å, le cofacteur dans sa moitié nicotinamide, et le glutamate sont flexibles (Figure 7.2). La flexibilité de ces deux derniers partenaires de l'acte catalytique pourrait être couplée afin de permettre le positionnement de la molécule d'eau lors de l'étape de déacylation (Moore *et al.*, 1998).

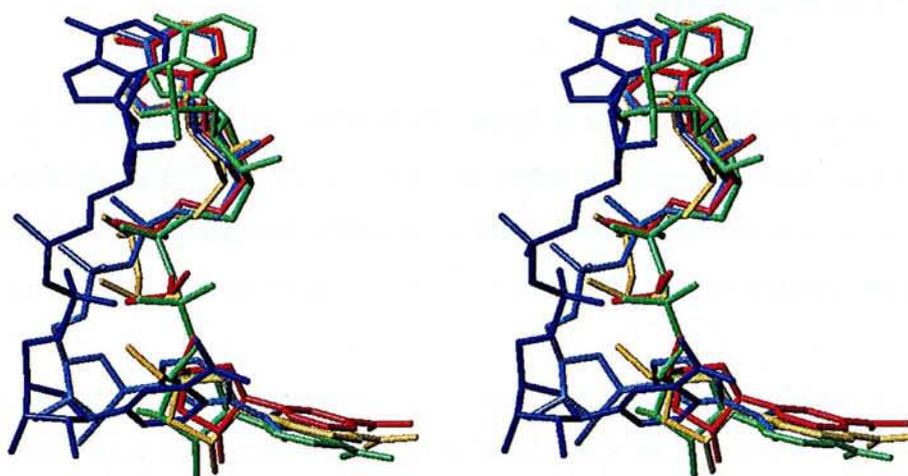


Figure 7.2: vue stéréoscopique des cofacteurs de l'ALDH1 (violet), l'ALDH2 (rouge), l'ALDH3 (bleu), l'ALDH9 (jaune) et de l'ALDH de *S. mutans* (vert) positionnés après superposition des domaines de Rossmann.

Afin d'étendre le mécanisme aux ALDH, autres qu'ALDH3, l'étude des alignements de séquences est rendue nécessaire, notamment celui qui regroupe 145 séquences réalisé par Perozich *et al.* (1999).

Selon les auteurs, la cystéine catalytique est présente chez toutes les ALDH ayant une activité catalytique. Comme mentionné en introduction, elle est mutée en arginine dans l' Ω -cristalline des céphalopodes ce qui aurait pour effet de rendre cette protéine inactive lui conférant un rôle structural (Zinovieva *et al.*, 1993). Ainsi, la cystéine est conservée dans plus de 95 % des séquences alignées ce qui permet d'affirmer que toutes les ALDH présentant une activité catalytique oxydent leur substrat en passant par la formation d'un intermédiaire réactionnel de type thioacylenzyme.

Chez *S. mutans*, l'intermédiaire tétraédrique est stabilisé par Asn154, conservée dans plus de 95 % des séquences étudiées. La mutation de ce résidu est observée dans des séquences d'ALDH répertoriées comme allergènes (Achatz *et al.*, 1995). Cette conservation suggère le même rôle de cet acide aminé chez toutes les ALDH.

Concernant le glutamate catalytique, il est conservé dans plus de 80 % des séquences. Ainsi, son rôle serait moins important que les deux autres résidus. Il semblerait donc que ce résidu intervienne dans le but d'augmenter l'efficacité de la réaction par un meilleur positionnement de la molécule d'eau lors de l'étape de déacylation plutôt que pour activer cette molécule d'eau.

Enfin, concernant la fixation du cofacteur, étant donné que l'ensemble des séquences analysées possède l'empreinte GstXXG, cela suggère que les ALDH fixent les cofacteurs NAD ou NADP d'une façon similaire.

Par conséquent, à l'aide des résultats de la cristallographie, de l'enzymologie obtenus sur l'ALDH de *S. mutans* et des résultats concernant d'autres ALDH, le mécanisme catalytique décrit ici et illustré au début de ce dernier chapitre peut aussi être généralisé à l'ensemble des ALDH des classes 1 et 2 décrites au début de ce paragraphe.

V- Investigations cristallographiques et biochimiques du mécanisme catalytique d'une aldéhyde déshydrogénase NADP dépendante de *Streptococcus mutans*

Structural and biochemical investigations of the catalytic mechanism of an NADP-dependent aldehyde dehydrogenase from Streptococcus mutans (Cobessi et al., 2000).

Cet article accepté à *Journal of Molecular Biology* décrit les études cristallographiques et biochimiques nécessaires à la compréhension du rôle de certains résidus impliqués dans l'acte catalytique. Une partie de l'article est consacrée à l'analyse de la structure de l'Apo2 dans laquelle un ion sulfate mimant la fonction aldéhydique du substrat est fixé au site catalytique de l'enzyme. Il interagit notamment avec l'Asn154 hautement conservée chez les ALDH. La comparaison des sites catalytiques de l'Apo1 et de l'holoenzyme montre la réorientation de la cystéine lors de la fixation du NADP. Ce mouvement de la chaîne latérale de Cys284 positionne alors son groupement sulfhydryle dans un nouvel environnement protéique. L'association des dipôles de la chaîne principale et de la charge positive du NADP serait responsable de la baisse du pKa de Cys284. L'analyse structurale du complexe ternaire enzyme C284S-NADP-substrat montre le rôle important de l'Asn154 comme site oxyanion pour la stabilisation de l'intermédiaire tétraédrique, sa mutation se traduisant par un effet isotopique de substrat montrant un changement dans l'étape limitante de la réaction, qui est alors l'étape d'acylation. Dans ces deux structures, Glu250 est situé à moins de 4 Å de Cys284 ou Ser284 dans un environnement hydrophobe conservé chez les ALDH. Cet environnement pourrait être le facteur structural déterminant de la valeur du pKa de Glu250.

L'analyse structurale de l'holoenzyme, du mutant C284S et du complexe ternaire montre différentes conformations du NADP. Cette analyse associée aux connaissances acquises sur les ALDH suggère une flexibilité du cofacteur couplée à celle du glutamate catalytique lors de la réaction. La comparaison des domaines de Rossmann de Sm-ALDH, des autres ALDH et des déshydrogénases non ALDH révèle la conservation de la longueur de la première boucle du repliement de Rossmann plus longue chez les ALDH. Cette caractéristique structurale pourrait être responsable du nouveau mode de fixation du NAD(P) au domaine de Rossmann.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Running Head : ALDH catalytic mechanism

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Structural and biochemical investigations of the catalytic mechanism of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*.

David Cobessi¹, Frédérique Tête-Favier¹, Stéphane Marchal², Guy Branlant² and André Aubry^{1,*}

Addresses:

1 : Laboratoire de Cristallographie et Modélisation des Matériaux Minéraux et Biologiques, Groupe Biocristallographie, ESA 7036, Faculté des Sciences, BP 239, 54506 Vandoeuvre-lès-Nancy, France

2 : Laboratoire de Maturation des ARNs et Enzymologie Moléculaire, UMR 7567 CNRS-UHP, Faculté des Sciences, BP 239, 54506 Vandoeuvre-lès-Nancy, France

* : to whom correspondence should be addressed

Fax number: (33) 03-83-40-64-92

Email: aubry@lcm3b.u-nancy.fr

Keywords: ALDH X-ray structure, oxyanion hole, NADP flexibility, ternary complex, isotope effect.

Abstract

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

The NADP-dependent non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Streptococcus mutans* (abbreviated Sm-ALDH) belongs to the aldehyde dehydrogenase (ALDH) family. Its catalytic mechanism proceeds *via* two steps: acylation and deacylation. Its high catalytic efficiency at neutral pH implies prerequisites relative to the chemical mechanism. First, the catalytic Cys284 should be accessible and in a thiolate form at physiological pH to attack efficiently the aldehydic group of the glyceraldehyde-3-phosphate (G3P). Second, the hydride transfer from the hemithioacetal intermediate toward the nicotinamide ring of NADP should be efficient. Third, the nucleophilic character of the water molecule involved in the deacylation should be strongly increased. Moreover, the different complexes formed during the catalytic process should be stabilised.

The crystal structures presented here (an apoenzyme named Apo2 with two sulphate ions bound to the catalytic site, the C284S mutant holoenzyme and the ternary complex composed of the C284S holoenzyme and G3P) together with biochemical results and previously published apo and holo crystal structures (named Apo1 and Holo1 respectively) contribute to the understanding of the ALDH catalytic mechanism.

Comparison of Apo1 and Holo1 crystal structures shows a Cys284 side-chain rotation of 110°, upon cofactor binding, which is probably responsible for its pKa decrease. In the Apo2 structure an oxygen atom of a sulphate anion interacts by hydrogen bonds with the NH₂ group of a conserved asparagine (Asn154 in Sm-ALDH) and the Cys284 NH group. In the ternary complex, the oxygen atom of the aldehydic carbonyl group of the substrate interacts with the Ser284 NH group and the Asn154 NH₂ group. A substrate isotope effect on acylation is observed for both the wild type and the N154A and N154T mutants. The rate of the acylation step strongly decreases for the mutants and becomes limiting. All these results suggest the involvement of Asn154 in an oxyanion hole in order to stabilise the tetrahedral intermediate and likely the other intermediates of the reaction. In the ternary complex, the cofactor conformation is shifted in comparison with its conformation in the C284S holoenzyme structure likely resulting from its peculiar binding mode to the Rossmann fold (*i.e.* non perpendicular to the plane of the β sheet). This change is likely favoured by a characteristic loop of the Rossmann fold, longer in ALDHs than in other dehydrogenases, whose orientation could be constrained by a conserved proline. In the ternary and C284S holoenzyme structures, as well as in the Apo2 structure, the Glu250 side chain is situated less

than 4 Å from Cys284 or Ser284 instead of 7 Å in the crystal structure of the wild-type holoenzyme. It is now positioned in a hydrophobic environment. This supports the pKa assignment of 7.6 to Glu250 as recently proposed from enzymatic studies.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

Introduction

S.C.D. - U.N.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

Aldehyde dehydrogenases (ALDH) belong to a family of enzymes which oxidise aldehydes into their carboxylic acids in the presence of NAD(P). For example, they participate in ethanol metabolism by conversion of acetaldehyde into acetic acid (Steinmetz *et al.*, 1997), or in the cellular differentiation as they produce retinoic acid by retinal oxidation (Moore *et al.*, 1998). In *Streptococcus mutans*, an ALDH (abbreviation: Sm-ALDH) catalyses the NADPH production during the irreversible oxidation of glyceraldehyde-3-phosphate (G3P) into 3-phosphoglycerate (3-PGA) (Boyd *et al.*, 1995).

Also found in photosynthetic organisms, this enzyme is defined as a NADP-dependent non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPDH) on the basis of its reaction. In spite of its name, this 200 kDa enzyme has to be classified in the ALDH family as it does not possess sequence identity with the other phosphorylating GAPDHs for which the molecular weight is 145 kDa.

In ALDHs, the catalytic reaction occurs in two steps. Firstly, the acylation step occurs *via* the nucleophilic attack of the catalytic cysteine on the substrate (Cys284 in Sm-ALDH) leading to formation of the hemithioacetal intermediate next to the hydride transfer and the thioacylenzyme and NAD(P)H. Secondly, the deacylation step involves a water molecule. To be efficient, these acylation and deacylation steps require cysteine and water molecule activations at physiological pH as well as stabilisation of the different intermediates of the reaction.

Concerning the cysteine activation, it was proposed that a conserved glutamate (Glu250 in Sm-ALDH) might act as the general base. Its putative role might be to activate a water molecule observed in the ALDH2 holoenzyme structure (Steinmetz *et al.*, 1997). This water molecule might then deprotonate the cysteine to form the thiolate ion (Steinmetz *et al.*, 1997). However, the thiolate formation might also be due to the proximity of the main chain NH groups on both sides of the catalytic cysteine (*i.e.* the peptide nitrogens of Cys284 and Thr285 in Sm-ALDH) (Steinmetz *et al.*, 1997) and/or to the positive charge of the NAD(P) nicotinamide ring (Steinmetz *et al.*, 1997; Johansson *et al.*, 1998) thus decreasing the cysteine pKa. Recently, it was proposed that Cys284 activation is accompanied by a slight conformational change during cofactor binding (Marchal & Branlant 1999). A careful investigation of the comparison of the previously solved Sm-ALDH apo and holoenzyme

structure catalytic sites (Cobessi *et al.*, 1999) (named Apo1 and Holo1) shows that a Cys284 side chain rotation accompanies cofactor binding.

In the deacylation, the thioacylenzyme undergoes a nucleophilic attack by a water molecule likely positioned and/or activated by its protein environment. Recently, it was proposed that Glu250 in Sm-ALDH could be the residue involved (Marchal *et al.*, 2000). Movements of the reduced cofactor are also postulated to be necessary to position the water molecule between the cysteine and the glutamate (Moore *et al.*, 1998). In the Sm-ALDH Apo1 and Holo1 structures (Cobessi *et al.*, 1999), the Glu250 carboxylate group is situated approximately 7 Å from the Cys284 sulphur atom, as observed in the other ALDH structures. The present analyses of the catalytic sites of the Apo1, Holo1, Apo2 and ternary complex structures of Sm-ALDH in relation with the biochemical results (Marchal & Branlant, 1999; Marchal *et al.*, 2000) suggest a catalytic mechanism in which the glutamate Glu250 has no role in the acylation. The data presented here highlight the role of an oxyanion hole in the stabilisation of the tetrahedral intermediate and likely of the other intermediates of the reaction. We also compare the Rossmann fold of the ALDHs and other dehydrogenases in order to understand the structural determinants of the new NAD(P) binding mode in ALDHs which seem important for an efficient catalytic reaction.

Results and discussion

The Apo2 enzyme structure

Apo2 crystals were obtained in the same crystallisation conditions as Apo1 and Holo1. The structure was solved at 2.5 Å resolution (see Material and Methods) and revealed a fold similar to the two other reported Sm-ALDH structures (Cobessi *et al.*, 1998) (rms between Apo1 and Apo2 tetramers: 0.51 Å). The major difference concerns the position of a few side chains in the catalytic site, most notably Glu250.

In the catalytic site of Apo2, the side chain orientation of Cys284 remains the same as in Apo1 while the Glu250 side chain orientation is modified. Its carboxylate group is now positioned 3.7 Å from the Cys284 sulphur atom (Figure 1) instead of 7 Å. This shift results from the Glu250 side chain rotation around its χ_1 angle (the omit map does not present any alternative conformation, data not shown). Therefore, the interaction between the Glu250

carboxylate group and the Arg437 guanidinium group mentioned in the Apo1 structure (Cobessi *et al.*, 1999) is lost. No water molecule is observed between Glu250 and Cys284. Another difference is the presence of two sulphate ions in the Apo2 structure instead of one in the catalytic site of Apo1 (Figure 1). Their mean temperature factor is 25.5 \AA^2 (B overall: 10.0 \AA^2) and these anions are clearly defined in the electron density. The first anion (SO_{4a}) observed in Apo2 corresponds to that observed in the Apo1 and Holo1 structures. In Apo2, this anion is present in all monomers with an occupancy of 0.5. It interacts through hydrogen bonds with the Arg103, Arg283 guanidinium groups and the Arg437 and Gly438 NH groups. The crystallographic occupancy of the second sulphate ion (SO_{4b}) is different (1 or 0.5) depending on the monomer. Its stabilisation is ensured by the orientation of the Arg437 side chain which does not interact with Glu250 anymore, and by the Arg283 side chain. SO_{4b} is also in interaction with the Asn154 amide group, the Thr285, Cys284 NH groups, and the Thr285 hydroxyl group.

The Glu250 side chain is situated in a hydrophobic environment, close to SO_{4b}. Its carboxylate group is situated 3.7 \AA from the Cys284 thiolate group. The Glu250 carboxylate group is surrounded by several hydrophobic groups such as the Thr285, Thr229, Leu159 and the Leu405 methyl groups, and the Phe444 aromatic ring. In such an environment, the Glu250 pKa is probably shifted toward a basic value such as the pKa value of 7.6 titrated by Marchal & Branlant (1999) and Marchal *et al.* (2000). It is likely protonated and interacts with SO_{4b} by hydrogen bonds (no other hydrogen bond is observed with surrounding residues). As the distance between Glu250 and Cys284 cannot be related to the crystallisation conditions which are identical for Apo1 and Apo2, the rotation of the Glu250 side chain could be related to the binding of the second sulphate ion SO_{4b}. Both observed sulphate anions could mimic the substrate (G3P) binding in the catalytic site. SO_{4a} could correspond to the C3 phosphate group of G3P. Taking into account the sulphate spatial geometry, one oxygen atom of SO_{4b} looks like the oxyanion of the tetrahedral intermediate. Among the above mentioned interactions concerning SO_{4b}, two must be pointed out: the conserved Asn154 amide group and the Cys284 NH group which interact with the same SO_{4b} oxygen atom. Their important role was already postulated on the basis of substrate binding modelling (Steinmetz *et al.*, 1997; Johansson *et al.*, 1998). The oxyanion hole is also similar to that in papain-like cysteine proteases, where it involves Gln side chain amide and the peptide NH of the catalytic Cys (Schroder *et al.*, 1993).

At first sight, the short distance between Cys284 and Glu250 observed in Apo2 suggests that Glu250 could directly deprotonate the cysteine residue without the water

molecule involvement. The achievement of the carboxylic form is envisaged from the hydrophobic environment of Glu250. But, due to the fact that the pKa of 7.6 likely corresponds to Glu250 and that the acylation step is not affected when Glu250 is mutated into Ala or Gln (Marchal *et al.*, 2000), the cysteine activation thus requires another pathway. In the Holo1 structure, the Cys284 side chain rotation toward the catalytic site, during the cofactor binding, positions it near the Cys284 and Thr285 NH groups as displayed in Figure 2. The orientation change of Cys284 side chain is very clear and unambiguous as confirmed by the omit map. It is likely responsible for its increased accessibility in comparison with Apo1, where the Cys284 pKa is 8.5, and for its pKa decrease. Although no hydrogen bond is observed between Cys284 and either NH284 or NH285 main chain, the crystallographic and biochemical studies imply that it is these groups, together with NADP, that cause the pKa decrease for Cys284. We note that the positively charged nicotinamide ring also influences the Cys284 pKa as it is 6.8 in the presence of NADPH instead of 6.1 when NADP is added (Marchal & Branlant, 1999).

The C284S holoenzyme structure

The overall structure of the C284S mutant holoenzyme displays no difference with Holo1 (rms value between tetramers: 0.43 Å). The Ser284 side chain is directed toward the catalytic site like the side chain of Cys284 in Holo1. Although the resolution is low (3.15 Å), the electron density clearly shows that the interaction between Glu250 and Arg437 is lost and the Glu250 carboxylate group is situated 3 Å from the Ser284 hydroxyl group. One sulphate ion is bound to each monomer at the position where the G3P phosphate group is supposed to interact with the enzyme (Cobessi *et al.*, 1999). It is stabilised by the Arg103 and Arg283 guanidinium groups, Tyr155 and Thr285 hydroxyl groups, and the Arg437 NH group. Its occupancy is 0.5. The electron density around the nicotinamide mononucleoside is very poorly defined in comparison with the ADP part. NADP occupancy is 0.5 and its mean B value is 3.7 Å² (B overall: 13.1 Å²). The observed disorder concurs with the cofactor flexibility already described in ALDHs (Moore *et al.*, 1998; Johansson *et al.*, 1998; Lamb & Newcomer, 1999) and with the low occupancy in the C284S structure. In the C284S mutant, the observed conformational change of the nicotinamide mononucleotide could result from the loss of the electrostatic interaction between the negative charge of the sulphhydryl group and the positive charge of the nicotinamide ring as Cys284 is mutated to a serine residue. This cofactor conformation allows the Glu250 side chain rotation and its positioning 3 Å away

from the Ser284 hydroxyl group. Despite the low resolution (3.15 Å), our interpretation of these movements is confirmed in a second crystal form of the C284S mutant holoenzyme, obtained under the same crystallisation conditions as for Holo1 (data not shown).

C284S ternary complex analysis

As already observed in the structure of C284S holoenzyme, the electron density is mostly continuous for the cofactor ADP part, but is very poor for the nicotinamide and ribose portions, indicating conformational flexibility in the latter. The NADP occupancy is 0.5 and its mean temperature factor is 6.6 \AA^2 (B overall: 14.6 \AA^2). The interactions between the substrate and protein are summarised in Table 2 and Figure 3a shows G3P in an omit electron density map. The occupancy of G3P is 0.5 (mean B value: 6.7 \AA^2). The G3P phosphate group is notably stabilised by Arg103 and Arg283 guanidinium groups as we proposed on the basis of the Holo1 structure (Cobessi *et al.*, 1999). The aldehyde carbonyl oxygen of the substrate interacts with the Asn154 amide and Ser284 NH groups (Figure 3b) and thereby likely with the Cys284 NH group of the wild-type enzyme as the rms value between their positions is small (0.27 \AA). This result strengthens the essential role of the conserved Asn154 in the ALDH family which is also assumed to stabilise the oxyanion of the tetrahedral intermediate as proposed on the basis of the Apo2 structure and modelling (Steinmetz *et al.*, 1997). The Glu250 side chain is situated 3 \AA from the Ser284 hydroxyl group in a hydrophobic environment as described in Apo2. Due to substrate binding, the observed NADP conformation differs from those observed in the holo C284S and Holo1 structures (Figure 4a and b). The ternary complex structure was solved at 3 \AA from hexagonal crystals obtained in the same crystallisation condition as C284S crystals. Therefore, only the substrate binding can be responsible for the new cofactor conformation in these crystallisation conditions. Moreover, differences are observed between the monomers in the ternary complex. Indeed, the electron densities around G3P and nicotinamide mononucleoside differ according to the monomers suggesting slightly different conformations. The different observed NADP geometries show that cofactor movements probably occur during the catalytic reaction. They could be favoured by the low amount of interactions between the cofactor and Rossmann fold as well as by the new binding mode of the cofactor to the nucleotide binding domain. The observed NADP conformational change during the substrate binding in the C284S mutant might also occur in the wild-type enzyme.

The Glu250 and NAD(P) flexibilities

There is no structural evidence about the wild-type holoenzyme providing an explanation to the Glu250 pKa shift toward a basic value (7.6) as postulated to occur in solution when the cofactor binds (Marchal & Branlant, 1999). Indeed, Glu250 has the same position in Apo1 and Holo1 crystal structures and is at least 6.9 Å away from Cys284. Therefore, the Glu250 pKa measurement suggests a crystal holoenzyme conformation differing from that existing in solution by the Glu250 orientation. Another unambiguously defined Glu250 side chain orientation is now observed in the C284S binary and ternary complexes, at 3.15 and 3 Å resolution respectively, in comparison with that observed in Apo1 and Holo1. Its carboxylate group is situated 3 Å from Ser284 oxygen atom. This could be related with the mutation as already mentioned. However, the shift is also observed in the Apo2 structure where Glu250 is situated 3.7 Å from Cys284. ALDH structure comparisons (ALDH1, RalDH2, ALDH2, ALDH3, ALDH9, Apo1, Holo1, Apo2) reveal the same characteristics, *i.e.* NAD(P) and glutamate flexibilities, required for the catalytic mechanism. In ALDH3, the NAD conformation is *syn* in contrast with the *anti* conformation observed in the other structures; the cofactor is buried less in comparison with the other ALDH structures. Two NAD conformations are observed in ALDH1 where the Glu268 electron density (corresponding to Glu250 in Sm-ALDH) is poorly defined and its side chain accompanied by high B factors (Moore *et al.*, 1998). The electron density is also reported to be stronger in the cofactor ADP part in ALDH9 (Johansson *et al.*, 1998), while the nicotinamide moiety is disordered in an ALDH2 structure (Steinmetz *et al.*, 1997). All these observations are consistent with Glu250 mobility and then with its ability to move closer to Cys284 as observed in the C284S binary and C284S ternary complex mutant where Ser284 is situated 3 Å from Glu250 and also in Apo2 where Glu250 is located 3.7 Å from Cys284. Moreover, the Glu250 side chain rotation can occur with only a few changes in the catalytic site like the SO₄^b binding observed in Apo2. In the ternary complex, C284S and Apo2 structures, the Glu250 carboxylate group no longer interacts with the Arg437 guanidinium group. Moreover, the distances with Leu159, Thr229, Thr285, Leu405 methyl groups and the Phe444 aromatic ring are less than 5 Å (Table3). The association of the salt bridge loss and hydrophobic environment may lead to the Glu250 pKa increase to 7.6. Thr229, Leu251, Phe444 and Leu405 are highly conserved in the ALDH family. Leu159 is mainly replaced by methionine or isoleucine residues (Perozich *et al.*, 1999). This suggests a similar role of all these hydrophobic residues in ALDH (Figure 1).

As the above mentioned pKa measurements were performed at 4°C after an enzyme incubation of 18 hours in the presence of a large NADP excess (Marchal & Branlant, 1999) at pH 5-6, it is likely that the structures in solution at 4°C differed from those determined by X-ray crystallography at 20°C. However, the Glu250 pKa is also measured at 20°C when G3P and NADP were added. Thereby the Glu250 pKa increase could be due to its positioning in a hydrophobic environment as we observed in Apo2 (wild-type apoenzyme) and in the C284S binary and ternary structures. Nevertheless, a Cys284-Ser284 mutation effect cannot be totally excluded in this case.

Biochemical and enzymatic properties of N154A, N154T and C284S mutants

Over a pH range from 5 to 9.5, substitution of Asn154 by Thr or Ala results in a decrease of the k_{cat} value by two or three orders of magnitude, respectively (Table 4). No changes in the apparent K_M values for substrate and cofactor are observed (data not shown). These mutations also suppress the presteady state burst of NADPH production observed for the wild-type enzyme (Marchal *et al.*, 2000) indicating that the rate-limiting step is shifted from the deacylation step in the wild type to the acylation step in the mutant enzyme. The acylation is at least 10^3 to 10^4 -fold less efficient compared to the wild type. The fact that an isotope effect is observed for the N154A or N154T mutants when D-[1- ^2H]G3P is used as the substrate proves that hydride transfer has become rate-limiting within acylation. An isotope effect is also observed for the wild type in the acylation step (Table 4).

The combined observations of a substrate isotope effect on the acylation for both the wild type and the mutant enzymes, and of the very low efficiency of the acylation step for the mutants suggest a direct interaction between the Asn154 amide group and the oxygen atom of the thiohemiacetal intermediate and/or the oxygen atom of the G3P aldehyde group. This hypothesis is strongly supported by inspection of the Apo2 and the ternary complex C284S/NADP/G3P structures which show the NH_2 group of the Asn154 side chain pointing towards one of the SO_4b oxygen atoms and the oxygen atom of the G3P aldehyde group, respectively. Altogether, these results give evidence of an oxyanion hole composed of at least residue Asn154. Within the acylation, Asn154 would first stabilise the substrate binding and tetrahedral intermediate and secondly, decrease the pK_{app} of the thiohemiacetal hydroxyl group such that it is deprotonated at physiological pH, thus permitting an efficient hydride transfer towards the C4 position of the oxidised cofactor without base catalyst assistance.

The fact that for the C284S mutant the acylation is rate-limiting, but that no isotope effect is observed with D-[1-²H]G3P indicates that the rate-limiting step occurs prior to the hydride transfer. Thus, the very low catalytic efficiency of the C284S mutant is likely due to poor nucleophilicity of the Ser284 hydroxyl group within the Sm-ALDH protein context which in turn should prevent efficient formation of the hemiacetal intermediate. This probably explains why under the experimental X-ray conditions used, the non covalent ternary complex C284S-NADP-G3P was chemically stable within the crystal.

NAD(P) binding mode in ALDHs

The analysis of the ALDH holoenzyme structures reveals a new binding mode of the cofactor to the Rossmann fold in comparison with the usual binding mode where the cofactor is perpendicular to the plane of the β -sheet. This latter NAD(P) binding mode and the Rossmann fold characteristics were extensively described (Wierenga *et al.*, 1986; Carugo *et al.*, 1997; Bellamacina 1996), especially the interactions of the first loop containing the famous GXGXXG fingerprint with the NAD(P) pyrophosphate group. In ALDHs, the cofactor straddles the Rossmann fold. The fingerprint GX₄G (Liu *et al.*, 1997) in ALDHs is found in the loop connecting the fourth β -strand to the fourth α -helix. In contrast with the other dehydrogenases, no interaction between an α -helix dipole and the NAD(P) pyrophosphate group is observed except in mitochondrial ALDH (Steinmetz *et al.*, 1997). The low number of interactions between the cofactor and Rossmann fold in ALDHs could be related to required movements of NAD(P) during the catalytic reaction (Moore *et al.*, 1998).

In order to study the new binding mode of the cofactor to ALDHs, nine NADP-dependent and eleven NAD-dependent proteins possessing a Rossmann fold were selected in a representative way in the Protein Data Bank (Bernstein *et al.*, 1977). Their cofactor binding domains were superimposed onto the Sm-ALDH NADP-binding domain taking the three central β -strands of the Rossmann fold as the common and invariant structural features in these enzymes (Figure 5a and b). The Rossmann folds differ in the number of β -strands and α -helices and by the length of the secondary structure elements. However, the standard NAD(P) binding mode to the Rossmann fold *i.e.* perpendicular to the plane of the β -sheet is observed in all the selected proteins except Sm-ALDH. The most striking point concerns the length of the first loop connecting the first β -strand to the first α -helix of the domain. The loop in Sm-ALDH is longer than in the other selected NAD(P)-dependent dehydrogenases.

The superposition of the known ALDH Rossmann fold structures and the sequence alignments show that all ALDHs possess this long loop which contains a strictly conserved proline (Pro156 in Sm-ALDH). Figure 6 displays the virtual interaction of the Sm-ALDH Rossmann fold with NAD as it binds in a classical manner to the Rossmann fold of *Bacillus stearothermophilus* GAPDH. Bad contacts occurred between the adenosine moiety and the conserved long loop. The orientation of the loop in the ALDH family could arise from the presence of the conserved proline. Therefore, the conjunction of the conserved proline and length of the loop could be structural determining factors responsible for the cofactor binding mode to the Rossmann fold observed in ALDHs. These structural characteristics probably have a role in the NADP conformational change observed in the Sm-ALDH ternary complex and both NAD positions observed in ALDH1 (Moore *et al.*, 1998), thereby enabling the efficiency of the catalytic reaction.

Conclusion

Surprisingly, two Sm-ALDH apoenzyme structures show catalytic sites differing by the side chain positions of a few residues, notably the catalytic glutamate Glu250 and the arginine Arg437 observed in interaction with Glu250 in Apo1. The side chain rotation, due to the binding of a second sulphate anion in Apo2, reveals Glu250 flexibility. This flexibility is confirmed by comparing its observed position in the Apo1, Holo1, binary and ternary complex of C284S mutant structures. The present structural studies combined with previous biochemical studies (Marchal & Branlant, 1999; Marchal *et al.*, 2000) show a local structural reorganisation associated with cofactor and substrate binding which could explain the pKa variations of Cys284 and Glu250 and the change in Cys284 accessibility necessary for the catalytic reaction. The chemical activation of Cys284 could be due to the main chain NH groups on both sides of Cys284 (*i.e.* the peptide nitrogens of Cys284 and Thr285) while the 7.6 value for Glu250 pKa could result from a conserved hydrophobic environment in ALDHs.

The results of the Apo2 and ternary complex structure comparison are in agreement with the presence of an oxyanion hole composed of the invariant Asn154 and the catalytic Cys284 NH group which both interact by hydrogen bonding with one of the four sulphate oxygen atoms of Apo2 and the aldehyde carbonyl group of the substrate. Our biochemical data on the Asn154 mutants show a strong k_{cat} decrease and a substrate isotope effect which support this conclusion. When the substrate binds, a different nicotinamide mononucleoside

conformation is observed while several NAD conformations are found in the NAD-dependent ALDH structures. This NAD(P) flexibility could be due to its peculiar binding mode to the Rossmann fold. The comparison with several dehydrogenases shows that a conserved loop length between the first β strand and the first α helix in association with a conserved proline could be the structural determining factors of its new binding mode and therefore of the NAD(P) movements required for the catalytic mechanism of ALDHs.

Material and Methods

Material

C284S, N154A and N154T mutants were produced and purified following a procedure similar to that described for wild type and other Sm-ALDH mutants (Marchal & Branlant, 1999).

The Apo2 structure

Crystallisation and data collection

Apo2 was crystallised in two conditions using the hanging drop method at 20°C. 2 μ l of protein solution at a concentration of 5 mg/ml was mixed with 2 μ l of reservoir containing 2.0 M ammonium sulphate, 0.1 M cacodylate buffer pH 6.5 and over a range of 1 to 5 % acetone. These Apo2 crystals poorly diffract at 3 Å resolution. The data were collected at 100 K using glucose as cryogenic agent. The second Apo2 crystal form was obtained with a protein concentration of 5 mg/ml. 2 μ l of protein solution was mixed with 2 μ l of reservoir containing 2.0 M ammonium sulphate, 0.1 M HEPES pH 7.5 and 2 % PEG 400. The data were collected at 100 K using an image plate detector and an Enraf Nonius rotating anode with a wavelength of 1.54 Å. The cryogenic solution was composed of 20 % glycerol. The crystals diffracted at 2.5 Å resolution. The space group of both crystal forms is P2₁2₁2. These latter crystals were used to solve the structure.

Structure resolution

The Apo2 structure was solved by molecular replacement using AMoRe (Navaza, 1994) with the Apo1 model (Cobessi *et al.*, 1999) (PDB entry: 1euh). The structure was refined using X-

PLOR (Brünger, 1990) and rebuilt using O (Jones & Kjeldgaard, 1990) and TURBO-FRODO (Roussel & Cambillau, 1991) in 3Fo-2Fc and Fo-Fc electron density maps. After a few cycles of energy minimisation and molecular dynamics, the R-factor and free R-factor (Brünger, 1992) including all data at 2.5 Å resolution were 23.2 % and 28 % respectively.

Binary and ternary complex of C284S mutant

Crystallisation

The purified C284S mutant was dialysed in a phosphate buffer pH 6.4, 50 mM containing 1mM dithiothreitol, at 4°C. The enzyme was next concentrated at 40 mg/ml. The holoenzyme was obtained by addition of NADP at a final concentration of 16 mM. The droplet consisted in 2 µl protein and 6 µl of ammonium sulphate 2.0 M. The crystals reached their maximal size in few days using the hanging drop method at 20°C. Next, D-G3P was added to obtain a droplet at 2.5 mM D-G3P.

Data collection and structure resolution

Data for the crystals of the binary and ternary complex of C284S mutant were collected at 100 K using a DIP2030 image plate detector and an Enraf Nonius rotating anode. Data were processed and merged using DENZO and SCALEPACK packages (Otwinowsky & Minor, 1997). Crystals of binary and ternary complex of C284S mutant belong to space group P3₂21 (Table 1).

The structures were solved by molecular replacement using AMoRe (Navaza, 1994) and the holoenzyme structure (Cobessi *et al.*, 1999) (PDB entry: 2euh) as template. As in the orthorhombic form, one tetramer is present in the asymmetric unit. Models were rebuilt using TURBO-FRODO (Roussel & Cambillau, 1991), and O (Jones & Kjeldgaard, 1990). Successive energy minimisation using a conjugate gradient minimisation, molecular dynamics at 3000 K by simulated annealing, occupancy refinement and finally grouped B-factor refinement were performed using X-PLOR (Brünger, 1990), working with 75.5 % (56846 reflections) and 81.1 % (51464 reflections) of data for the ternary and binary C284S complexes respectively, *i. e.* a cutoff of 2 σ on F. At the end of the procedure, the free R-factor (Brünger, 1992) was 28.3 % and the R-factor was 24.2 %, at 2 σ (F) and 3 Å resolution ,

for the ternary complex. The free R-factor (Brünger, 1992) and R-factor for the C284S mutant holostructure were 27.3 % and 22.8 % respectively at $2 \sigma(F)$ and 3.15 Å resolution. For these structures, a non-crystallographic symmetry restraint was applied during the whole refinement. Indeed, its absence would lead to a divergent evolution of R_{free} and R_{crist} (Kleywegt & Jones, 1995) due to the high number of refined parameters compared with the number of reflections present at such resolutions.

Note: in the binary and ternary C284S structures, the nicotinamide mononucleoside part of NADP was positioned as best as possible in the poor quality electron density. All the mentioned observations, distances, and comparisons were performed from the monomer A as the D-G3P density is correctly defined inside it. In the other monomers the G3P electron density is poorly defined.

Atomic coordinates of Apo2 and the ternary complex were deposited in the Protein Data Bank, accession numbers 1qi6 and 1qi1 respectively.

Steady state kinetics

Initial rate measurements were carried out on a Kontron Uvikon 933 spectrophotometer by following the NADPH appearance at 340 nm and 25°C. The standard experimental conditions were 1 mM NADP, 0.2 mM D-G3P and 15 μN C284S, 4 μN N154A or 0.66 μN N154T mutant enzyme in 50 mM TES buffer 5 mM, β-mercaptoethanol, pH 8.2.

Deuterium isotope effects

D-[1-²H]G3P was prepared as already described (Corbier *et al.*, 1990). Its concentration was enzymatically determined. Isotope effects on wild type, N154A/T mutants were measured in 50 mM TES buffer, 5 mM β-mercaptoethanol, pH 8.2, by direct comparison of the initial velocities measured with D-G3P or D-[1-²H]G3P (0.1 mM and 0.2 mM) in the presence of 1 mM NADP at 25°C. Isotope measurement of the wild type presteady state kinetics was carried out on a SFM 3 Biologic Instrument stopped flow apparatus (Biologic, Grenoble France) using a Tungsten lamp and a 5 nm slit as light source. As for steady state kinetics, time-courses for the reaction of wild-type enzyme with D-G3P or D-[1-²H]G3P and NADP were followed by the appearance of NADPH monitored at 340 nm, at a constant temperature of 10°C, pH 8.2. The experiment was set up with one syringe filled with wild-type enzyme

(16 μ N) and the other one containing 1 mM NADP and 0.1 or 0.2 mM D-G3P or D-[1-²H]G3P. The continuously measured absorbance was digitised and directly analysed as the sum of exponential and linear components with the Biokine program employing non-linear regression analysis.

Acknowledgements

We gratefully acknowledge Pr Vitoux and Dr Rahuel-Clermont for helpful discussions. This work was technically supported by the Service Commun de Diffraction X sur Monocristaux at the Henri Poincaré University. We thank E. Habermacher, S. Boutserin and S. Azza for their efficient technical help. This work was financially supported by the Centre National de la Recherche Scientifique, the University Henri Poincaré Nancy I and the Institut Fédératif de Recherche Protéines.

S.C.D. - U.N.P. NANCY I
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

S.C.D. - U.N.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

References

- Bellamacina, C. R. (1996). The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. *FASEB J.*, **10**, 1257-1269.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archive-file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.
- Boyd, D. A., Cvitkovitch D. G. & Hamilton I. R. (1995). Sequence, expression, and function of the gene for the non phosphorylating, NADP-dependent glyceraldehyde-3-phosphate dehydrogenase of *Streptococcus mutans*. *J. Bacteriol.*, **177**, 2622-2627.
- Brünger, A. T. (1990a). *X-PLOR Manual*, Yale University Press, New Haven, CT.
- Brünger, A. T. (1992b). The Rfree value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.
- Carugo, O. & Argos, P. (1997). NADP-dependent enzymes. I: conserved stereochemistry of cofactor binding. *Proteins. Struct. Fonct. Genet.*, **28**, 10-28.
- Cobessi, D., Tete Favier, F., Marchal, S., Azza S., Branlant, G. & Aubry, A. (1999). Apo and holo crystal structures of a NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*. *J. Mol. Biol.*, **290**, 161-173.
- Corbier, C., Mouglin, A., Mely, Y., Adolph, H.W., Zeppezauer, M., Gerard, D., Wonacott, A. & Branlant, G. (1990). The nicotinamide subsite of glyceraldehyde-3-phosphate dehydrogenase studied by site-directed mutagenesis. *Biochimie*, **72**, 545-554.
- Johansson, K., El-Ahmad, M., Ramaswamy, S., Hjelmqvist, L., Jörnvall, H. & Eklund, H. (1998). Structure of betaine aldehyde dehydrogenase at 2.1 Å resolution. *Protein Sci.*, **7**, 2106-2117.
- Jones, T. A. & Kjeldgaard, M. (1990). "O" Manual, version 6.1, Uppsala University Sweden, and Aarhus University, Denmark.
- Kleywegt, G. J. & Jones, T., A. (1995). Where freedom is given, liberties are taken. *Structure*, **3**, 535-540.
- Kraulis, P. J. (1991). MOLSCRIPIT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, **24**, 946-950.
- Lamb, A.L. & Newcomer, M.E. (1999). The structure of retinal dehydrogenase type II at 2.7 Å resolution: implications for retinal specificity. *Biochemistry*, **38**, 6003-6011.

- Liu, Z. L., Sun, Y.-J., Rose, J., Chung, Y. J., Hsiao, C.-D., Chang, W.-R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J. & Wang, B.-C. (1997). The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nat. Struc. Biol.*, **4**, 317-326.
- Marchal, S. & Branlant, (1999). Evidence for the chemical activation of essential Cys-302 upon cofactor binding to nonphosphorylating glyceraldehyde 3-phosphate dehydrogenase from *Streptococcus mutans*. *Biochemistry*, **38**, 12950-12958.
- Marchal, S., Rahuel-Clermont, S. & Branlant, G. (2000). The role of Glu268 in the catalytic mechanism of non-phosphorylating glyceraldehyde dehydrogenase from *Streptococcus mutans*. *Biochemistry*, **39**, 3327-3335.
- Meritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.4 – A program for photo-realistic molecular graphics. *Acta Crystallogr.*, **D50**, 869-873.
- Moore, S. A., Baker, H. M., Blythe, T. J., Kitson, K. E., Kitson, T. M. & Baker, E. N. (1998). Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. *Structure*, **6**, 1541-1551.
- Navaza, J. (1994). AMoRe: An automated package for molecular replacement. *Acta Cryst.*, **A50**, 157-163.
- Otwinowsky, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307-326.
- Perozich, J., Nicholas, H., Wang, B.-C., Lindahl, R. & Hempel, J. (1999). Relationships within the aldehyde dehydrogenase extended family. *Protein Sci.*, **8**, 137-146.
- Roussel, P. A. & Cambillau, C. (1991). TURBO-FRODO Silicon Graphics applications directory, Silicon Graphics, Mountains View, CA.
- Schroder, E., Phillips, C., Garman, E., Harlos, K. & Crawford, C. (1993). X-ray crystallographic structure of a papain-leupeptin complex. *FEBS Lett.*, **315**, 38-42.
- Steinmetz, C. G., Xie, P., Weiner, H., & Hurley, D. T. (1997). Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. *Structure*, **5**, 701-711.
- Wierenga, R. K., Terpstra, P. & Hol, W. G. J. (1986). Prediction of the occurrence of the ADP binding $\beta\alpha\beta$ -fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.*, **187**, 101-107.

Legends

Figure 1: stereo view of the superposition of the Apo1 and Apo2 catalytic sites. The Asn154, Glu250, Cys284, SO_{4a} and SO_{4b} are displayed in ball-and-stick using MOLSCRIPT (Kraulis, 1991) and Raster3D (Meritt & Murphy, 1994). Traces of the described residues forming the hydrophobic environment of Glu250 in Apo2 are displayed. The Glu250 side chain in Apo1 (Apo2) is coloured green (red). Accessibility of Glu250 in Apo2 is 2 Å².

Figure 2: stereo view illustrating the orientation shift of the Cys284 side chain under the cofactor binding drawn by MOLSCRIPT (Kraulis, 1991) and Raster3D (Meritt & Murphy, 1994). For clarity, only the nicotinamide mononucleotide is displayed. The Holo1 position of Cys284 is labelled Holo.

Figure 3: a) stereo view of D-G3P in a 3Fo-2Fc omit electron density map contoured at 1.2σ generated using TURBO-FRODO (Roussel & Cambillau, 1991). Surrounding residues (Arg283, Ser284, Thr285, Arg437) and the nicotinamide mononucleoside are also displayed.

b) stereo view of the interactions between D-G3P and the amide group of the conserved asparagine in ALDHs (Asn154 in Sm-ALDH) and the NH group of the serine residue (Ser284 in Sm-ALDH). Arg283 backbone is also displayed. For clarity, only the nicotinamide mononucleotide is drawn. Note that no hydrogen bond is drawn between the cofactor and substrate as the electron density is poorly defined around the nicotinamide mononucleoside. Figure was prepared using MOLSCRIPT (Kraulis, 1991) and Raster3D (Meritt & Murphy, 1994).

Figure 4: a) stereo view of NADP in a 3Fo-2Fc electron density map contoured at 1.2σ generated using TURBO-FRODO (Roussel & Cambillau, 1991) for the C284S structure.

b) stereo view of NADP in a 3Fo-2Fc electron density map contoured at 1.2σ generated using TURBO-FRODO (Roussel & Cambillau, 1991) for the ternary complex.

Figure 5 a and b: stereo view of the Rossmann fold superposition of several NAD (5a) and NADP (5b) dependent dehydrogenases from the PDB using TURBO-FRODO (Roussel & Cambillau, 1991). For clarity only the region of the first loop is shown. The Sm-ALDH loop is coloured red and its residues are labelled. The NAD-dependent dehydrogenases selected are

available in the PDB under the accession codes: 1cdo, 1dru, 1emd, 1enp, 1eny, 1fmc, 1gd1, 1ldg, 1ra1, 1xel, 2npx. The accession codes for the NADP-dependent dehydrogenases selected are: 1cyd, 1fdt, 1get, 1pgo, 1qor, 1typ, 1tdf, 1ykf, 2dpg.

Figure 6: stereo view of the superposition of NAD on the Rossmann fold of Sm-ALDH as it interacts with the Rossmann fold of *B. stearrowophilus* GAPDH. The conserved proline side chain in ALDHs is drawn in ball-and-stick. NAD trace is displayed. The figure was prepared using MOLSCRIPT (Kraulis, 1991) and Raster3D (Meritt & Murphy, 1994).

Figure 7: schematic representation of G3P.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

S.C.D. - U.H.P. NANCY 1
 BIBLIOTHÈQUE DES SCIENCES
 Rue du Jardin Botanique
 54000 VILLERS-LES-NANCY

Table 1

Data collection and refinement statistics

Data set	Apo2	C284S	C284S ternary complex
1- Data collection			
Resolution (Å)	2.5	3.15	3.0
Cell parameters	140.8 157.5 110.3	157.7 157.7 280.9	158.5 158.5 282.8
Space group	P2 ₁ 2 ₁ 2	P3 ₂ 21	P3 ₂ 21
Completeness (%)	96	91.5	80.3
R _{sym}	0.06	0.127	0.097
2- Refinement statistics			
R-factor value	0.232	0.228	0.242
R-free value	0.280	0.273	0.283
Rms deviation			
Bond lengths (Å)	0.007	0.009	0.009
Bond angles (Å)	1.295	2.454	2.403
Improper angles (°)	1.169	1.326	0.938

Table2

Substrate-protein interactions by hydrogen bonds observed in monomer A of Sm-ALDH.

Protein	Substrate	Distance (Å)
Arg103 NH1	OC3	3.2
Asn154 ND2	O1	3.2
Arg283 NH2	OC2	2.7
Ser284 N	O1	3.2
Thr285 OG1	OPC	3.0
Arg437 N	OC2	3.1
Tyr155 OH	OPC	3.1

Table 3

Distances between Glu250 carboxylate group and CH2/CH3 groups of its hydrophobic environment in the C284S ternary complex.

Glu250 carboxylate	CH2/CH3 group	Distance (Å)
OE1	Leu159 CD1	4.7
OE1	Leu159 CD2	4.5
OE1	Thr229 CG2	3.3
OE2	Thr229 CG2	4.4
OE2	Thr285 CG2	4.8
OE2	Leu405 OE2	4.6
OE1	Phe444 CZ	4.8
OE1	Phe444 CE2	4.9
OE1	Phe444 CD2	4.6
OE2	Phe444 CZ	3.7
OE2	Phe444 CE2	3.4
OE2	Phe444 CD2	4.7

Table 4

Parameters of the mutant catalysed reaction and substrate isotope effect

	Rate limiting step	k_{ss}^{\max} (s^{-1}) ^a	K.I.E.
<i>Wild-type</i> ^c	<i>deacylation</i>	68	1.1 (2.7 ^b)
N154A	<i>acylation</i>	0.07	3.7
N154T	<i>acylation</i>	0.7	2.1
C284S	<i>acylation</i>	0.005	1.2

^a k_{ss}^{\max} value corresponds to the steady state rate constant

^b The K.I.E. (kinetic isotope effect) value was determined under presteady state conditions at 10°C (see also Material and Methods). It is defined by the ratio:

$$\frac{[D-G3P]_{k_{cat}}}{[H2-D-G3P]_{k_{cat}}}$$

^cThe acylation rate constant is 210 s^{-1} at pH 8.2 and 10°C .

Abbreviations

ALDH aldehyde dehydrogenase

Sm-ALDH *Streptococcus mutans* aldehyde dehydrogenase

D-G3P D-glyceraldehyde-3-phosphate

3-PGA 3-phosphoglycerate

K.I.E. kinetic isotope effect

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Figure 1

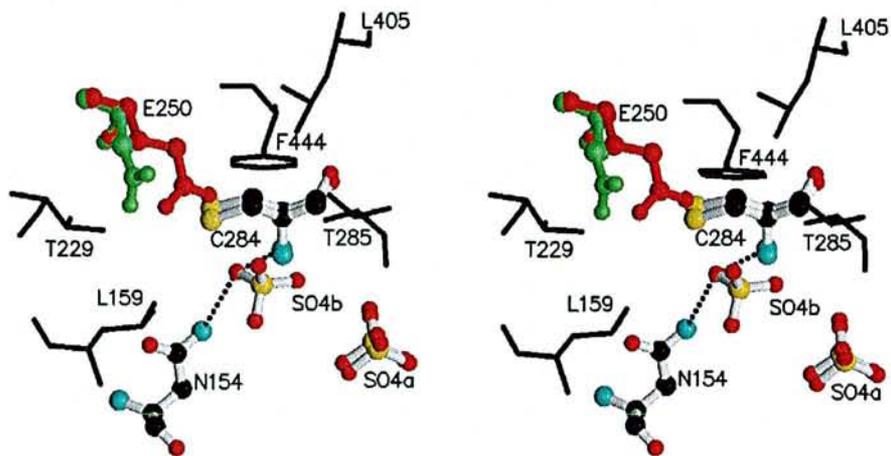
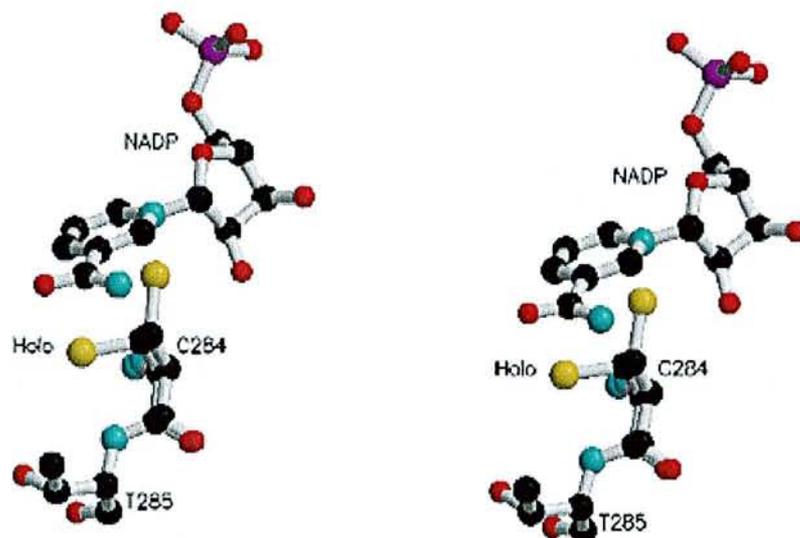


Figure 2



S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Figure 3a

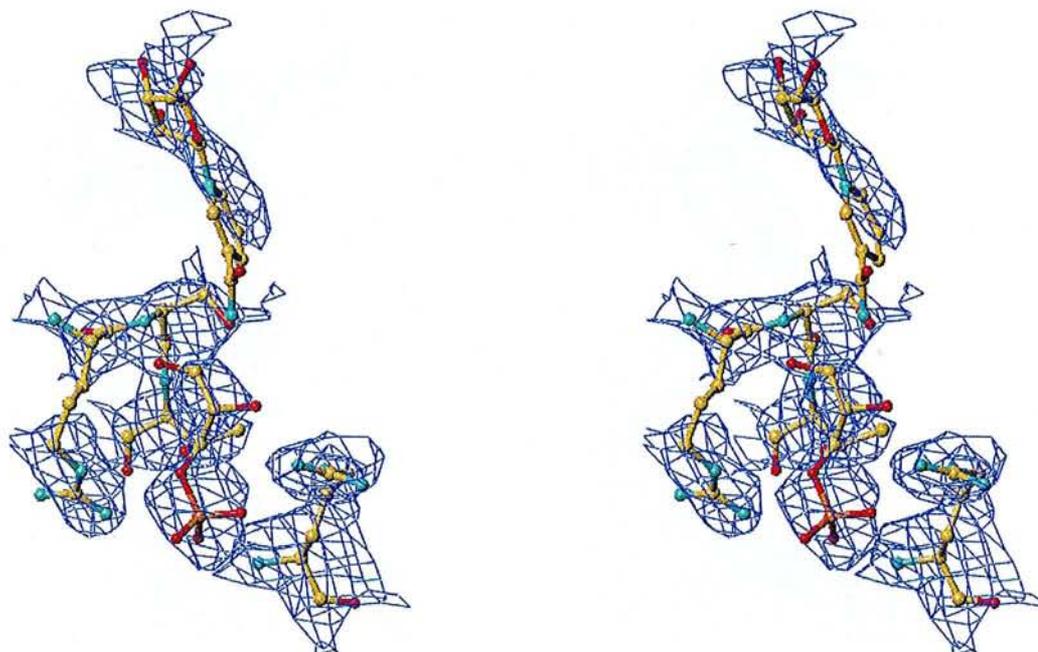


Figure 3b

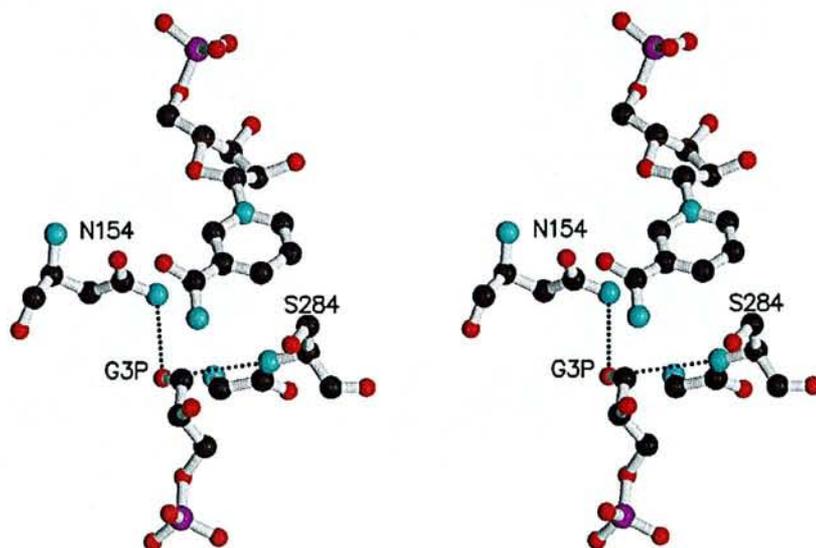


Figure 4a

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

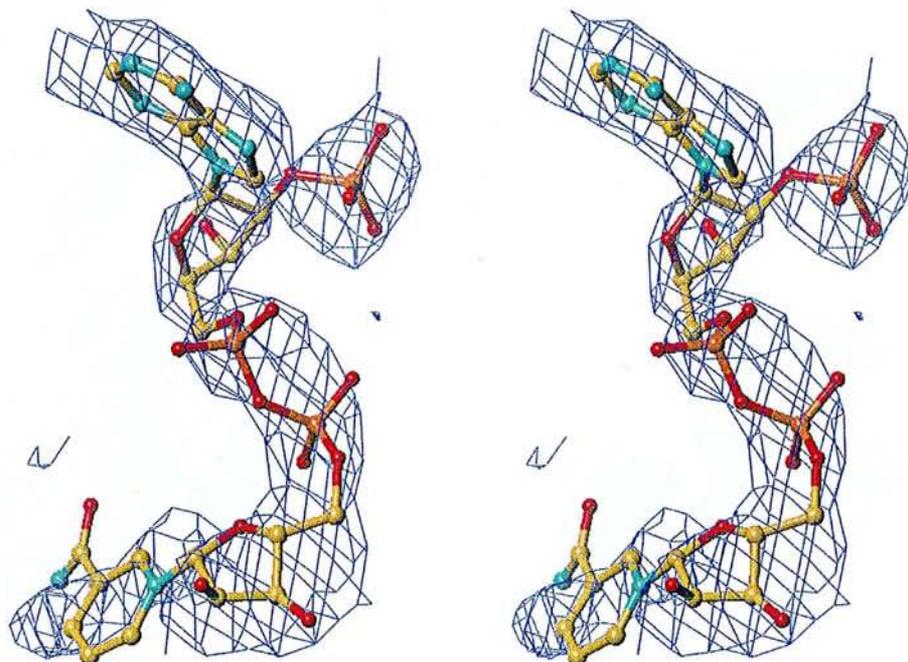
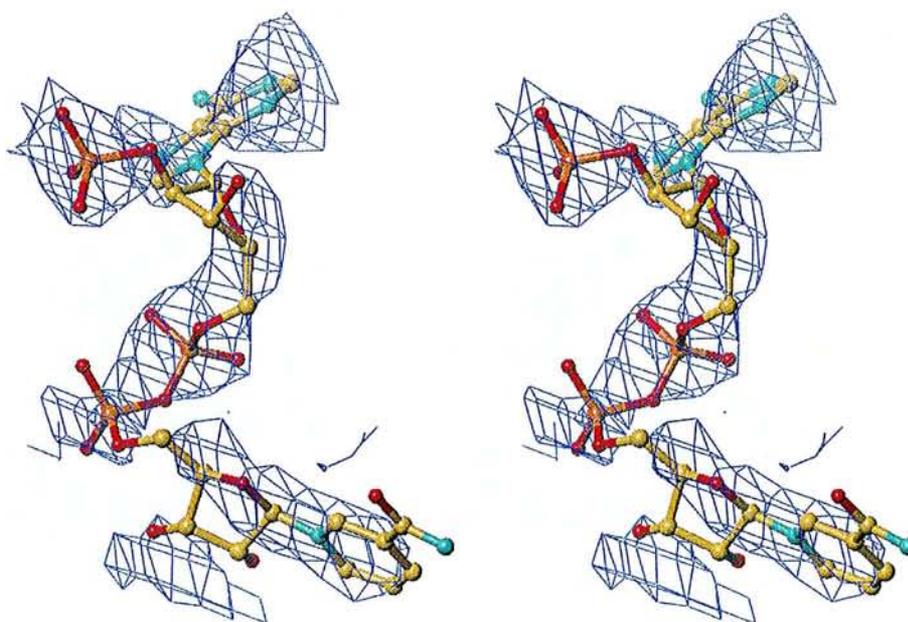


Figure 4b



S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Figure 5a

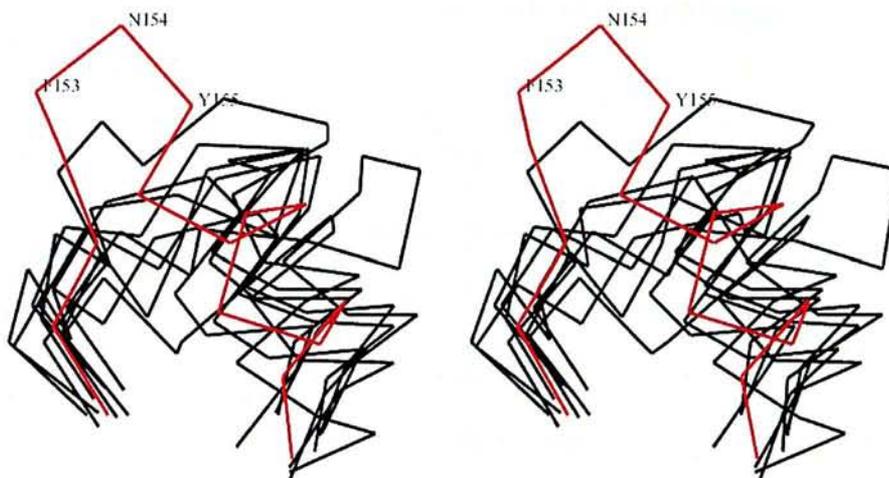


Figure 5b

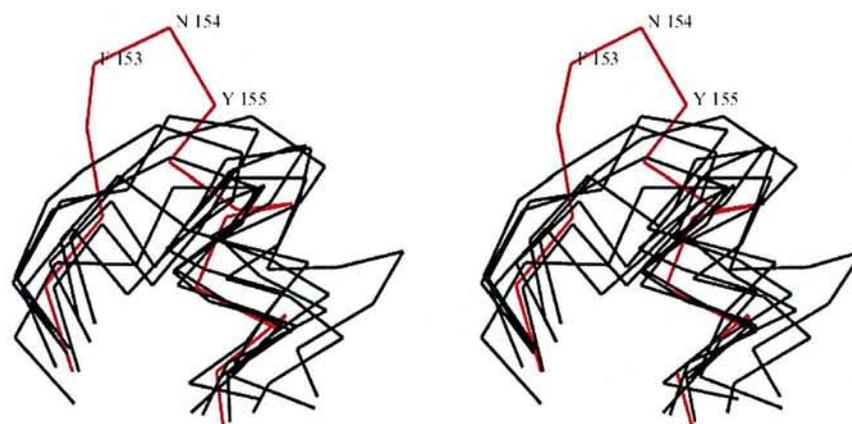


Figure 6

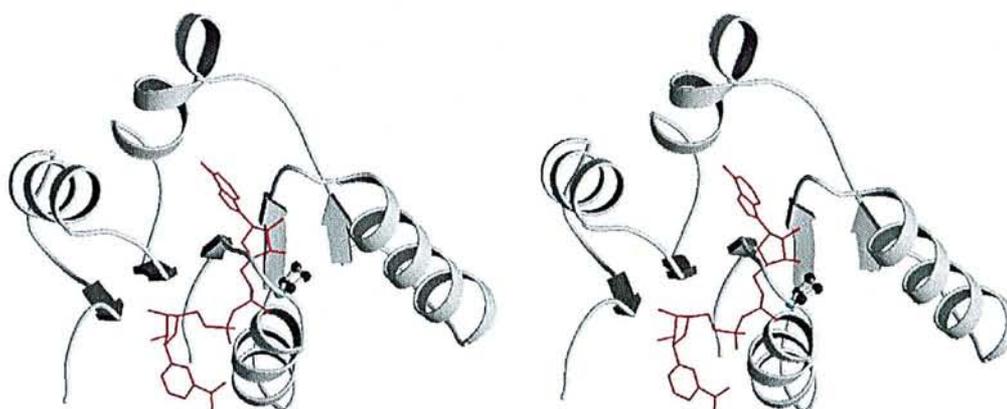
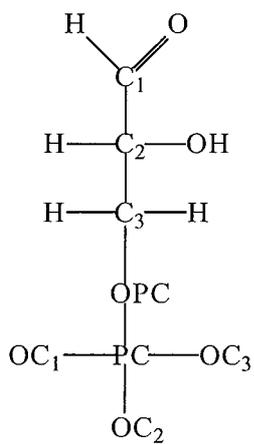


Figure 7

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY



S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Conclusion et perspectives

Les objectifs de cette étude étaient multiples. Tout d'abord, la résolution des structures de l'apo (Apo1) et de l'holoenzyme de la glycéraldhyde-3-phosphate déshydrogénase NADP dépendante non phosphorylante de *Streptococcus mutans*, premières structures d'une ALDH NADP dépendante, a permis de montrer quels étaient les déterminants structuraux de la spécificité de cofacteur. Des déterminants de l'état oligomérique des ALDH ont également pu être mis en évidence. Par la suite, l'étude structurale d'une seconde forme d'apoenzyme (Apo2), d'un mutant (C284S) et d'un complexe ternaire (enzyme C284S-NADP-G3P), en association avec des résultats de l'enzymologie, ont apporté des informations sur le mécanisme catalytique des ALDH. Enfin, l'analyse de différents alignements de séquences a révélé la conservation de plusieurs résidus impliqués dans la fixation du NADP et du G3P chez les GAPDH NADP dépendantes non phosphorylantes.

L'obtention régulière de cristaux de qualité satisfaisante pour des études cristallographiques et l'utilisation du remplacement moléculaire, ont permis la résolution de la première structure d'une GAPDH NADP dépendante non phosphorylante, ainsi que celle de différentes structures de mutants et de complexes. La structure de Sm-ALDH, enzyme tétramérique d'environ 200 kDa, peut être décrite comme un dimère de dimère présentant une symétrie non cristallographique 222. Le monomère de l'enzyme se divise en trois domaines: le domaine de fixation du cofacteur (qui contient le domaine de Rossmann), le domaine

catalytique et un petit domaine d'oligomérisation. Les deux premiers domaines sont composés de brins β et d'hélices α alors que le troisième domaine est un feuillet β antiparallèle.

La comparaison de structures d'ALDH dimérique (ALDH3) et tétramériques (ALDH2, et ALDH9) montre un positionnement différent de leurs domaines. Ce phénomène est probablement dû à une orientation différente du domaine d'oligomérisation entre les ALDH tétramériques et l'ALDH3. L'analyse des alignements de séquences et les comparaisons de structures ont permis de proposer un rôle de déterminant de l'état oligomérique à certains acides aminés. La mutation de trois d'entre eux simultanément (L116P, G120E, M118F) a permis l'obtention d'une solution protéique polydisperse.

L'apoenzyme de Sm-ALDH est en interaction avec plusieurs ions sulfate dont certains miment les groupements phosphate du NADP et du substrat, le G3P. L'étude structurale de l'holoenzyme a permis de mettre en évidence le nouveau mode de fixation du NADP au domaine de Rossmann (*i.e.* enjambant le feuillet β parallèle) et les facteurs moléculaires majeurs responsables de la spécificité de cofacteur pour le NADP. Elle est due en partie à la présence d'une thréonine (en position 180) ou d'une sérine pour ce type d'ALDH. Ces résidus sont mutés en proline, en glutamate, ou en isoleucine chez les ALDH NAD dépendantes. Ainsi la spécificité de cofacteur pour le NADP serait due à l'absence de charge et à l'absence de contraintes stériques à proximité du 2'-phosphate du NADP. La comparaison structurale de l'apoenzyme avec l'holoenzyme montre deux changements locaux de conformation dus à la fixation du NADP, non observés dans les structures d'ALDH NAD dépendantes. Un premier se produit au niveau de la zone Gly208-Asp215 du domaine de Rossmann, constituée d'une boucle et de la partie N-terminale d'une hélice α . Il se traduit par une amplification de la courbure de l'hélice α nécessaire à la stabilisation du NADP par son dipôle. Le second est la rotation de la chaîne latérale de la cystéine 284. Cette rotation positionne son groupement sulfhydryle à proximité des dipôles de la chaîne principale et l'oriente vers le site catalytique. Ce changement pourrait être à l'origine de la baisse de son pKa et de son changement d'accessibilité nécessaire à l'étape d'acylation.

La comparaison des différentes structures d'ALDH a permis de mettre en évidence la flexibilité du NAD(P). Les différentes positions et conformations observées pour le cofacteur résultent probablement de son mode particulier de fixation au domaine de Rossmann. Ce mode de fixation serait dû à la longueur conservée chez les ALDH de la première boucle du domaine de Rossmann et dont l'orientation serait contrainte par la présence d'une proline conservée. La flexibilité du cofacteur suggère son mouvement au cours de l'acte catalytique.

Au cours de cet acte, différents intermédiaires réactionnels sont formés et stabilisés par l'environnement protéique. L'étude d'une seconde structure d'apoenzyme a révélé la présence d'un ion sulfate supplémentaire dans le site catalytique en interaction avec la fonction amide de l'asparagine conservée Asn154 et avec le groupement NH de Cys284 (site oxyanion). Une interaction de même type est également observée avec la fonction aldéhydique du substrat dans la structure du complexe C284S-NADP-G3P. Dans cette structure, le carboxylate du glutamate Glu250 se situe à 3 Å de Ser284 au lieu de 7 Å (observé dans les structures de l'holoenzyme et de l'Apo1) ou de 3,7 Å (structure de l'Apo2). La rotation de la chaîne latérale de Glu250 entraîne la perte du pont salin avec Arg437 et positionne son groupement carboxylate dans un environnement hydrophobe. Ces deux phénomènes sont probablement à l'origine de la valeur de pKa de 7,6 attribuée à Glu250. La comparaison des différentes structures d'ALDH et de Sm-ALDH a permis de mettre également en évidence la flexibilité de Glu250 suggérant son mouvement couplé à celui du NAD(P) au cours de l'acte catalytique.

La cristallographie, en association avec certains résultats de l'enzymologie, a apporté des éléments de réponse pour la première partie de la réaction. La seconde étape de l'acte catalytique, la déacylation, nécessite l'activation d'une molécule d'eau par le glutamate. La molécule d'eau n'est pas visible dans les structures de Sm-ALDH. *In vitro*, d'autres molécules, telles que l'hydroxylamine et l'hydrazine peuvent se fixer aux mutants E250A et E250Q de l'ALDH de *S. mutans* et attaquer le thioacylenzyme. L'étude structurale du mutant E250A qui présente une activité catalytique en présence d'hydroxylamine a été entreprise. Elle devrait permettre d'observer les résidus avec lesquels l'hydroxylamine interagit. Il sera peut-être possible d'assimiler sa position à celle de la molécule d'eau.

L'étude cristallographique d'une forme dimérique de Sm-ALDH sera entreprise, si d'une part les mutations permettant son obtention n'abolissent pas l'activité enzymatique, et si d'autre part le dimère reste stable en solution. Dans cette optique, de nouvelles mutations ont été proposées. Il s'agit des mutations des résidus Arg117, Glu119, Glu124, Lys133 et Lys134.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

Bibliographie

- Achatz, G., Oberkofler, H., Lechenauer, E., Simon, B., Unger, A., Kandler, D., Ebner, C., Prillinger, H., Kraft, D. & Breitenbach, M. (1995). Molecular cloning of major and minor allergens of *Alternaria alternata* and *Cladosporium herbarum*. *Mol. Immunol.*, **32**, 213-227.
- Adams, M. J., Ellis, G. H., Gover, S., Naylor, C. E. & Phillips, C. (1994). Crystallographic study of coenzyme, coenzyme analogue and substrate binding in 6-phosphogluconate dehydrogenase: implications for NADP specificity and the enzyme mechanism. *Structure*, 1994, **2**, 651-668.
- Argawall, R. C. (1978). A new least-square refinement technique based on the fast Fourier transform algorithm. *Acta Cryst.*, **A34**, 791-809.
- Auerbach, G., Herrmann, A., Gutlich, M., Fischer, M., Jacob, U., Bacher, A. & Huber R. (1997). The 1.25 Å crystal structure of sepiapterin reductase reveals its binding mode to pterins and brain neurotransmitters. *EMBO J*, **16**, 7219-7230.
- Baker, H. M., Brown, R. L., Dobbs, A. J., Blackwell, L. F., Buckley, P. D., Hardman, M. J., Hill, J. P., Kitson, K. E., Kitson, T. M. & Baker, E. N. (1994). Crystallisation and preliminary x-ray diffraction studies on cytosolic (Class1) aldehyde dehydrogenase from sheep liver. *J. Mol. Biol.*, **241**, 263-264.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E. Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archive-file for macromolecular structures. *J. Mol. Biol.*, **112**, 535-542.
- Blundell, T. L. & Johnson, L. N. (1976). Protein Crystallography. Academic Press.

- Boyd, D. A., Cvitkovitch D. G. & Hamilton I. R. (1995). Sequence, expression, and function of the gene for the non phosphorylating, NADP-dependent glyceraldehyde-3-phosphate dehydrogenase of *Streptococcus mutans*. *J. Bacteriol.*, **177**, 2622-2627.
- Breton, R., Housset D., Mazza, C. & Fontecilla-Camps, J.C. (1996) The structure of a complex of human 17-beta-hydroxysteroid dehydrogenase with estradiol and NADP⁺ identifies two principal targets for the design of inhibitors. *Structure*, **4**, 905-915.
- Brünger, A. T. (1990a). *X-PLOR Manual*, Yale University Press, New Haven, CT.
- Brünger, A. T. (1992b). The Rfree value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.
- Brunner, N. A., Brinkmann, H., Siebers, B. & Hensel R. (1998). NAD⁺-dependent glyceraldehyde-3-phosphate dehydrogenase from *Thermoproteus tenax*. *J. Biol. Chem.*, **273**, 6149-6156.
- Brunner, N. A., Lang, D. A., Willmanns, M. & Hensel, R. (2000). Crystallization and preliminary X-ray analysis of the NAD-dependent non-phosphorylating GAPDH of the hyperthermophilic archaeon *Thermoproteus tenax*. *Acta Cryst.*, **D56**, 89-91.
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M. & Venter, J.C. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058-1073.
- Burmeister, W. P. (2000). Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Cryst.*, **D56**, 328-341.
- Carugo, O. & Argos, P. (1997). NADP-dependent enzymes. I: conserved stereochemistry of cofactor binding. *Proteins*, **28**, 10-28.
- CCP4. Collaborative Computational Project, Number 4. (1994). "The CCP4 suite: programmes for protein crystallography". *Acta Cryst.*, **D50**, 760-763.
- Charron, C., Talfournier, F., Isupov, M. N., Branlant, G., Littlechild, J. A., Vitoux, B., & Aubry, A. (1999). Crystallization and preliminary X-ray diffraction studies of D-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeon *Methanothermus fervidus*. *Acta Cryst.*, **D55**, 1353-1355.
- Charron, C., Talfournier, F., Isupov, M. N., Branlant, G., Littlechild, J. A., Vitoux, B., & Aubry, A. (2000). The crystal structure of D-glyceraldehyde-3-phosphate

- dehydrogenase from the hyperthermophilic archeon *Methanothermobacter thermophilus* at 2.1 Å resolution. *J. Mol. Biol.*, **297**, 481-500.
- Cobessi, D., Tête Favier, F., Marchal, S., Azza S., Branlant, G. & Aubry, A. (1999). Apo and holo crystal structures of a NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*. *J. Mol. Biol.*, **290**, 161-173.
- Cobessi, D., Tête Favier, F., Marchal, S., Branlant, G. & Aubry, A. (2000). Structural and biochemical investigations of the catalytic mechanism of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*. *J. Mol. Biol.*, *submitted*.
- Cody, V., Galitsky, N., Luft, J. R., Pangborn, W., Rosowsky A. & Blakley R. L. (1997). Comparison of two independent crystal structures of human dihydrofolate reductase ternary complexes reduced with nicotinamide adenine dinucleotide phosphate and the very tight-binding inhibitor PT523. *Biochemistry*, **36**, 13897-13903.
- Cosgrove, M. S., Naylor, C., Paludan, S., Adams, M. J. & Levy, H. R. (1998). On the mechanism of the reaction catalyzed by glucose 6-phosphate dehydrogenase. *Biochemistry*, **37**, 2759-67.
- Cowtan, K. D., & Main, P. (1996). Phase combination and cross validation in iterated density-modification calculations. *Acta Cryst.*, **D52**, 43-48.
- Croteau, N., Vevadi, M., Delarge, M., Meighen, E., Abu-Abed, M., Lynne Howell, P. & Vrielink A. (1996). Crystallisation and preliminary X-ray analysis of aldehyde dehydrogenase from *Vibrio harveyi*. *Protein Sci.*, **5**, 2130-2132.
- Crow, V. L. & Wittenberger, C. L. (1979). Separation and properties of NAD⁺ - and NADP⁺-dependent glyceraldehyde-3-phosphate dehydrogenases from *Streptococcus mutans*. *J. Biol. Chem.*, **254**, 1134-1142.
- Diamond, R. (1969). Profile analysis in single crystal diffractometry. *Acta Cryst.*, **A25**, 43-55.
- Eyck, L. F. T. (1973). Crystallographic fast Fourier transform. *Acta Cryst.*, **A29**, 183-191.
- Farrès, J., Guan, K. L. & Weiner, H. (1989). Primary structures of rat and bovine liver mitochondrial aldehyde dehydrogenases deduced from cDNA sequences. *Eur. J. Biochem.*, **180**, 67-74.
- Farrès, J., Wang, T. Y., Cunningham, S. J. & Weiner, H. (1995). Investigation of the active site cysteine residue of rat liver mitochondrial aldehyde dehydrogenase by site directed mutagenesis. *Biochemistry*, **34**, 2592-2598.

- Gouet, P., Courcelle, E., Stuart, D. and Metoz, F. (1999). ESPript: analyse of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305-308.
- Graham, C., Hodin, J. & Wistow, G. (1996). A retinaldehyde dehydrogenase as a structural protein in a mammalian eye lens. Gene recruitment of η -crystallin. *J. Biol. Chem.*, **271**, 15623-15628.
- Habenicht, A. (1997). The non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase: biochemistry, structure, occurrence and evolution. *Biol. Chem.*, **378**, 1413-1419.
- Habenicht, A., Hellman, U. & Cerff, R. (1994). Non-phosphorylating GAPDH of higher plants is a member of the aldehyde dehydrogenase superfamily with no sequence homology with the phosphorylating GAPDH. *J. Mol. Biol.*, **237**, 165-171.
- Habenicht, A., Quesada, A. & Cerff, R. (1997). Sequence of the non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Nicotiana plumbaginifolia* and phylogenetic origin of the gene family. *Gene*, **198**, 237-243.
- Hammen, P. K., Heard T. S., Waltner, M. & Weiner, H. (1999). The loss in hydrophobic area resulting from a Leu to Val mutation at the N-terminus of the aldehyde dehydrogenase presequence prevents import of the protein into mitochondria. *Protein Sci.*, **8**, 890-896.
- Hempel, J., Kaiser, R. & Jörnvall, H. (1985). Mitochondrial aldehyde dehydrogenase from human liver. Primary structure, difference in relation to the cytosolic enzyme, and functional correlations. *Eur. J. Biochem.*, **153**, 13-28.
- Hempel, J., Harper, K. & Lindahl, R. (1989). Inducible (Class 3) aldehyde dehydrogenase from rat hepatocellular carcinoma and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin-treated liver: distant relationship to the class 1 and class 2 enzyme from mammalian liver cytosol/mitochondria. *Biochemistry*, **28**, 1160-1167.
- Hooft, R. W. W., Vriend G., Sander, C. & Abola, E. E. (1996). Errors in protein structures. *Nature*, **381**, 272.
- Hsu, L. C., Chang, W. C., Shibuya, A. & Yoshida, A. (1992). Human stomach aldehyde dehydrogenase cDNA and genomic cloning, primary structure, and expression in *Escherichia coli*. *J. Biol. Chem.*, **267**, 3030-3037.
- Hsu, L. C. & Chang W. C. (1991). Cloning and characterization of a new functional human aldehyde dehydrogenase gene. *J. Biol. Chem.*, **266**, 12257-12265.
- Hsu, L. C., Tani, K., Fujiyoshi, T., Kurachi, K. & Yoshida, A. (1985). Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2. *Proc. Natl. Acad. Sci.*, **82**, 3771-3775.

- Hunter, W. N., Bailey, S., Habash, J., Harrop, S. J., Helliwell, J. R., Aboagye-Kwarteng, T., Smith, K. & Fairlamb A. H. (1992). Active site of trypanothione reductase. A target for rational drug design. *J. Mol. Biol.*, **227**, 322-333.
- Hurley, T. & Weiner, H. (1992). Crystallization and preliminary X-ray investigation of bovine liver mitochondrial aldehyde dehydrogenase. *J. Mol. Biol.*, **227**, 1255-1257.
- Iglesias, A. A. & Losada, E. (1988). Purification and kinetic and structural properties of spinach leaf NADP-dependent nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase. *Arch. Biochem. Biophys.*, **260**, 830-840.
- Iglesias, A. A., Serrano, A., Guerrero, M. G. & Losada, M. (1987). Purification and properties of NADP-dependent non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase from the green alga *Chlamydomonas reinhardtii*. *Biochim. Biophys. Acta*, **925**, 1-10.
- Isupov, M. N., Fleming T. M., Dalby, A. R., Crowhurst, G. S., Bourne, P. C. & Littlechild, J. A. (1999). Crystal structure of the glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archeon *Sulfolobus solfataricus*. *J. Mol. Biol.*, **291**, 651-660.
- Jancarik, J. & Kim, S. H. (1991). Sparse matrix sampling: a screening method for crystallisation of proteins. *J. Appl. Crystallog.*, **24**, 409-411.
- Johansson, J., Bahr-Lindström, H., Jeck, R., Woenckhaus, C., & Jörnvall, H. (1988). Mitochondrial aldehyde dehydrogenase from horse liver. Correlation of the same species variants for both the cytosolic and the mitochondrial forms of an enzyme. *Eur. J. Biochem.*, **172**, 527-533.
- Johansson, K., El-Ahmad, M., Ramaswamy, S., Hjelmqvist, L., Jörnvall, H. & Eklund, H. (1998). Structure of betaine aldehyde dehydrogenase at 2.1 Å resolution. *Protein Sci.*, **7**, 2106-2117.
- Johnson, L. N. (1992). Time-resolved protein crystallography. *Protein Sci.*, **10**, 1237-1243
- Jones, E. J., Brennan, M. D., Hempel, J. & Lindahl, R. (1988). Cloning and complete nucleotide sequence of a full-length cDNA encoding a catalytically functional tumor-associated aldehyde dehydrogenase. *Proc. Natl. Acad. Sci.*, **85**, 1782-1786.
- Jones, K. H., Lindahl, R., Baker, D. C. & Timkovich, R. (1987). Hydride transfert stereospecificity of rat liver aldehyde dehydrogenase. *J. Biol. Chem.*, **262**, 10911-10913.
- Jones, T. A. & Kjeldgaard, M. (1990). "O" Manual, version 6.1, Uppsala University, Sweden and Aarhus University, Denmark.
- Kabsch, W. (1988). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constant. *J. Appl. Cryst.*, **26**, 795-800.

- Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.
- Kedishvili, N. Y., Popov, K. M., Rougraff, P. M., Zhao, Y., Crabb, D. W. & Harris, R. A. (1992). Co-A-dependent methylmalonate-semialdehyde dehydrogenase, a unique member of the aldehyde dehydrogenase superfamily. *J. Biol. Chem.*, **267**, 19724-19729.
- Kleywegt, G. J. (1997). Validation of protein models from CA coordinates alone. *J. Mol. Biol.*, **273**, 371-376.
- Kleywegt, G. J. & Jones, T. A. (1993). Masks Made Easy. *ESF/CCP4 Newsletter*, **28**, 1993, 56-59.
- Kleywegt, G. J. & Jones, T. A. (1994). A superposition. *ESF/CCP4 Newsletter*, **31**, 9-14.
- Kleywegt, G. J. & Read, R. J. (1997). Not your average density. *Structure*, **5**, 1557-1569.
- Korkhin, Y., Kalb(Gilboa), A. J., Peretz, M., Bogin, O., Burstein, Y. & Frolow F. (1998). NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of *Clostridium beijerinckii* and *Thermoanaerobacter brockii*. *J. Mol. Biol.*, **278**, 967-981.
- Kraulis, P. J. (1991). MOLSCRIPIT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, **24**, 946-950.
- Krupenko, S. A., Wagner, C. & Cook, R. J. (1995). Cysteine 707 is involved in the dehydrogenase active site of rat 10-formyltetrahydrofolate dehydrogenase. *J. Biol. Chem.*, **270**, 519-522.
- Kurys, G., Shah, P. C., Kikonyogo, A., Reed, D., Ambroziak, W. & Pietruszko, R. (1993). Human aldehyde dehydrogenase. cDNA cloning and primary structure of the enzyme that catalyzes dehydrogenation of 4-aminobutyraldehyde. *Eur. J. Biochem.*, **218**, 311-320.
- Lamb, A. L. & Newcomer, M. E. (1999). The structure of retinal dehydrogenase type II at 2.7 Å resolution: implications for retinal specificity. *Biochemistry*, **38**, 6003-6011.
- Lamzin, V. S. & Wilson, K. S. (1993). Automated refinement of protein models. *Acta Cryst.*, **D49**, 129-147.
- Laskowsky, 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. Cryst.*, **26**, 283-291.
- Legaria, J., Rajsbaum, R., Munoz-Clares, R. A., Villegas-Sepulveda, N., Simpson, J. & Iturriaga, G. (1998). Molecular characterization of two genes encoding betaine aldehyde

- dehydrogenase from amaranth. Expression in leaves under short-term exposure to osmotic stress or abscisic acid. *Gene*, **218**, 69-76.
- Liu, Z. L., Sun, Y.-J., Rose, J., Chung, Y. J., Hsiao, C.-D., Chang, W.-R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J. & Wang, B.-C. (1997). The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nat. Struc. Biol.*, **4**, 317-326.
- Mann, C. J. & Weiner, H. (1999). Differences in the roles of conserved glutamic acid residues in the active sites of human class 3 and class 2 aldehyde dehydrogenase. *Protein Sci.*, **8**, 1922-1922.
- Marchal, S. & Branlant, (1999). Evidence for the chemical activation of essential Cys-302 upon cofactor binding to nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Streptococcus mutans*. *Biochemistry*, **38**, 12950-12958.
- Marchal, S., Rahuel-Clermont, S. & Branlant, G. (2000). The role of Glu268 in the catalytic mechanism of nonphosphorylating glyceraldehyde dehydrogenase from *Streptococcus mutans*. *Biochemistry*, **39**, 3327-3335.
- Matthews, B. W. (1968). Solvent content in protein crystals. *J. Mol. Biol.*, **33**, 491-497.
- McPherson, A. (1990). Current approaches to macromolecular crystallization. *Eur. J. Biochem.*, **189**, 1-23.
- Meritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.4 – A program for photo-realistic molecular graphics. *Acta Cryst.*, **D50**, 869-873.
- Mittl, P. R. & Schulz, G. E. (1994). Structure of glutathione reductase from *Escherichia coli* at 1.86 Å resolution: comparison with the enzyme from human erythrocytes. *Protein Sci.*, **3**, 799-809
- Miyauchi, K., Masaki, R., Taketani, S., Yamamoto, A., Akayama, M. & Tashiro, Y. (1991). Molecular cloning, sequencing, and expression of cDNA for rat liver microsomal aldehyde dehydrogenase. *J. Biol. Chem.*, **266**, 19536-19542.
- Moore, S. A., Baker, H. M., Blythe, T. J., Kitson, K. E., Kitson, T. M. & Baker, E. N. (1998). Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. *Structure*, **6**, 1541-1551.
- Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C. & Rossmann, M. G. (1975). Studies of asymmetry in the three-dimensional structure of lobster D-glyceraldehyde-3-phosphate dehydrogenase. *J. Biol. Chem.*, **250**, 9137-9162.

- Nakajima, K., Yamashita, A., Akama, H., Nakatsu, T., Kato, H., Hashimoto, T., Oda, J. & Yamada, Y. (1998). Crystal structures of two tropinone reductases: different reaction stereospecificities in the same protein fold. *Proc. Natl. Acad. Sci.*, **95**, 4876-4881.
- Navaza, J. (1994). AMoRe: An automated package for molecular replacement. *Acta Cryst.*, **A50**, 157-163.
- Ni, L., Sheikh, S. & Weiner, H. (1997). Involvement of glutamate 399 and lysine 192 in the mechanism of human liver mitochondrial aldehyde dehydrogenase. *J. Biol. Chem.*, **272**, 18823-18826.
- Ni, L., Jianzhong, Z., Hurley, T. & Weiner, H. (1999). Human liver mitochondrial aldehyde dehydrogenase: three-dimensional structure and the restoration of solubility and activity of chimeric form. *Protein Sci.*, **8**, 2784-2790.
- Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281-296
- Otwinowsky, Z. & Minor, W. (1997). Processing of X-Ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307-326.
- Peat, T. S. , Newman, J., Waldo, G. S. & Terwilliger, T. C. (1999). The crystal structure of a putative formate dehydrogenase from *Pyrobaculum aerophilum*. *Non publié*.
- Perozich, J., Nicholas, H., Wang, B.-C., Lindahl, R. & Hempel, J. (1999). Relationships within the aldehyde dehydrogenase extended family. *Protein Sci.*, **8**, 137-146.
- Racker, E. (1949). Bovine liver aldehyde dehydrogenase. *J. Biol. Chem.*, **177**, 883-892.
- Ravelli, B. G. & McSweeney, S. M. (2000). The "fingerprint" that X-rays can leave on structure. *Structure*, **8**, 315-328.
- Read, R. J. (1986). Improved Fourier coefficients for map using phases from partial structures with errors. *Acta Cryst.*, **A42**, 140-149.
- Rodgers, D. W. (1994). Cryocrystallography. *Structure*, **2**, 1135-1140.
- Rose, J. P., Hempel, J., Kuo, I., Lindahl, R. & Wang, B. C. (1990). Preliminary crystallographic analysis of class 3 rat liver aldehyde dehydrogenase. *Proteins*, **8**, 305-308.
- Rossmann, M. G. & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.*, **15**, 24-31.
- Rossmann, M. G., Moras, D. & Olsen, K. W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *Nature*, **250**, 194-199.
- Roussel, P. A. & Cambillau, C. (1991). *TURBO-FRODO Silicon Graphics applications directory*, Silicon Graphics, Mountains View, CA.

- Sabatino, P., Fermani, S., Ripamonti, A., Casseta, A., Scagliarini, S. & Trost, P. (1999). Crystallization and preliminary X-ray study of chloroplast glyceraldehyde-3-phosphate dehydrogenase. *Acta Cryst*, **D55**, 566-567.
- Scapin, G., Reddy, S. G. & Blanchard, J. S. (1996). Three-dimensional structure of meso-diaminopimelic acid dehydrogenase from *Corynebacterium glutamicum*. *Biochemistry*, **35**, 13540-13551.
- Serrano, A., Mateos, M. I. & Losada, M. (1991). Differential regulation by trophic conditions of phosphorylating and non-phosphorylating NADP⁺-dependent glyceraldehyde-3-phosphate dehydrogenases in *Chlorella fusca*. *Biochem. Biophys. Res. Commun.*, **181**, 1077-1083.
- Sheikh, S., Ni, L., Hurley, T. D. & Weiner, H. (1997). The potential roles of the conserved amino acids in human liver mitochondrial aldehyde dehydrogenase. *J. Biol. Chem.*, **272**, 18817-18822.
- Sidman, K. E., George, D.G., Barker, W. C., & Hunt, L. T. (1998). The protein identification resource (PIR). *Nucleic Acids Res.*, **16**, 1869-71.
- Skarzynsky, T., Moody, P.C. E. & Wonacott, A. J. (1987). Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. *J. Mol. Biol.*, **193**, 171-187.
- Smith, D.R., Doucette-Stamm, L. A., Deloughery, C., Lee, H.-M., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicare, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrovski, S., Church, G. M., Daniels, C. J., Mao, J.-I., Rice, P., Nolling, J. & Reeve, J. N. Complete genome sequence of *Methanobacterium thermoautotrophicum delta H*: functional analysis and comparative genomics. *J. Bacteriol.* **179**, 7135-7155 (1997).
- Smith, J. L. & Thompson, A. (1998). Reactivity of selenomethionine-dents in the magic bullet? *Structure*, **6**, 815-819.
- Steinmetz, C. G., Xie, P., Weiner, H., & Hurley, D. T. (1997). Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. *Structure*, **5**, 701-711.
- Tanaka, N., Nonaka, T., Nakanishi, M., Deyashiki, Y., Hara, A. & Mitsui, Y. (1996). Crystal structure of the ternary complex of mouse lung carbonyl reductase at 1.8 Å resolution:

- the structural origin of coenzyme specificity in the short-chain dehydrogenase/reductase family. *Structure*, **4**, 33-45.
- 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. *Nucleic Acids Res.*, **22**, 4673-4680.
- Thorn, J. M., Barton, J. D., Dixon, N. E., Ollis, D. L. & Edwards K. J. (1995). Crystal structure of *Escherichia coli* QOR quinone oxidoreductase complexed with NADPH. *J. Mol. Biol.*, **249**, 785-799.
- Tomarev, S. I., Zinovieva, R. D. & Piatigorsky, J. (1991). Crystallins of the octopus lens. Requirement from detoxification enzymes *J. Biol. Chem.*, **266**, 24226-24231.
- Tong & Rossmann, (1990). The locked rotation function. *Acta Cryst.*, **A46**, 783-792.
- Vedadi, M. & Meighen, E. (1997). Critical glutamic acid residues affecting the mechanism and nucleotide specificity of *Vibrio harveyi* aldehyde dehydrogenase. *Eur. J. Biochem.*, **246**, 698-704.
- Waksman, G., Krishna, T. S., Williams, C. H. Jr. & Kuriyan, J. (1994) Crystal structure of *Escherichia coli* thioredoxin reductase refined at 2 Å resolution. Implications for a large conformational change during catalysis. *J. Mol. Biol.*, **236**, 800-816
- Wang, X. & Weiner, H. (1995). Involvement of glutamate 268 in the active site of human liver mitochondrial (Class 2) aldehyde dehydrogenase as probed by site-directed mutagenesis. *Biochemistry*, **34**, 237-243.
- Weil, (1983). *Biochimie générale*, 4^e édition, Masson. Paris.
- Weretilnyk, E. A. & Hanson, A. D. (1990). Molecular cloning of a plant betaine-aldehyde dehydrogenase, an enzyme implicated in adaptation to salinity and drought. *Proc. Natl. Acad. Sci.*, **87**, 2745-2749.
- Wisconsin Package Version 9.0-UNIX, December 1996. Genetics Computer Group (GCG), Madison, Wisc.
- Wierenga, R. K., Terpstra, P. & Hol, W. G. J. (1985). Prediction of the occurrence of the ADP binding βαβ-fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.*, **187**, 101-107.
- Yang, W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. (1990). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science*, **249**, 1398-1405.

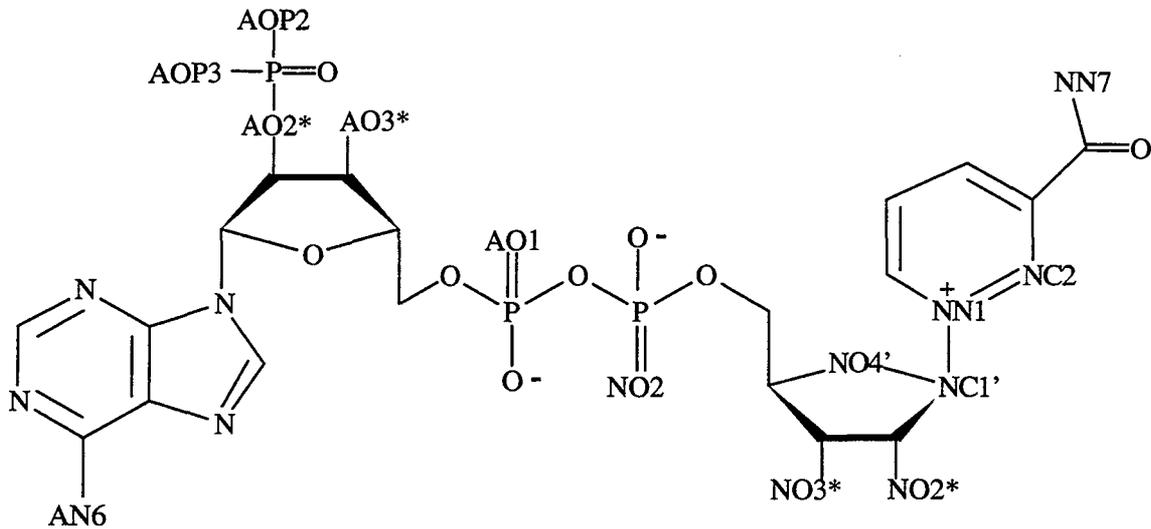
- Yoshida, A., Rzhetsky, A., Hsu, L. C. & Chang C. (1998). Human aldehyde dehydrogenase gene family. *Eur. J. Biochem.*, **251**, 549-557.
- Zhang, L., Ahvazi, B., Szittner, R., Vrieling, A. & Meighen E. (1999). Change of nucleotide specificity and enhancement in single point mutants of *Vibrio harveyi* aldehyde dehydrogenase. *Biochemistry*, **38**, 11440-11447.
- Zinovieva, R. D., Tomarev, S. I. & Piatigorsky, J. (1993). Aldehyde dehydrogenase-derived Ω -crystallin of squid and octopus. Specialization for lens expression. *J. Biol. Chem.*, **268**, 11449-11455.

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS LES NANCY

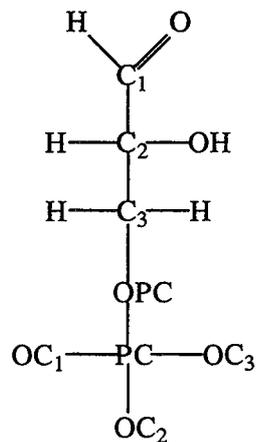
Annexes

S.C.D. - U.H.P. NANCY 1
 BIBLIOTHÈQUE DES SCIENCES
 Rue du Jardin Botanique
 54600 VILLERS-LES-NANCY

Formule développée du NADP



Formule développée du D-G3P



Listes des organismes dont les différentes séquences de l'alignement ont été isolées dans la banque PIR (protein identification resource; Sidman *et al.*, 1998)

1, *Homo sapiens* (EC 1.2.1.3); 2, *Homo sapiens* (EC 1.2.1.3); 3, *Equus caballus* (EC 1.2.1.3); 4, *Ovis orientalis aries*, *Ovis ammon aries* (EC 1.2.1.3); 5, *Rattus norvegicus* (EC 1.2.1.5); 6, *Rattus norvegicus* (EC 1.2.1.3); 7, *Mus musculus* (EC 1.2.1.3); 8, *Rattus norvegicus* (EC 1.2.1.3); 9, *Gallus gallus* (EC 1.2.1.3); 10, *Mus musculus* (EC 1.2.1.3); 11, *Homo sapiens* (EC 1.2.1.3); 12, *Rattus norvegicus* (EC 1.2.1.3); 13, *Mus musculus* (EC 1.2.1.3); 14, *Equus caballus* (EC 1.2.1.3); 15, *Homo sapiens* (EC 1.2.1.3); 16, *Bos primigenius taurus* (EC 1.2.1.3); 17, *Homo sapiens* (EC 1.2.1.3); 18, *Enchytraeus buchholzi* (EC 1.2.1.3); 19, *Octopus dofleini* (omega-crystallin); 20, *Alternaria alternata* (EC 1.2.1.3); 21, *Cladosporium herbarum* (EC 1.2.1.3); 22, *Emericella nidulans*, *Aspergillus nidulans* (EC 1.2.1.3); 23, *Nicotiana tabacum* (EC 1.2.1.3); 24, *Arabidopsis thaliana* (protéine T17F15.130); 25, *Leishmania tarentolae* (EC 1.2.1.3); 26, *Bacillus subtilis* (aldéhyde déshydrogénase dhaS); 27, *Saccharomyces cerevisiae* (EC 1.2.1.3); 28, *Saccharomyces cerevisiae* (EC 1.2.1.5); 29, *Alcaligenes eutrophus* (EC 1.2.1.3); 30, *Escherichia coli* (EC 1.2.1.22); 31, *Pseudomonas putida* (EC 1.2.1.-); 32, *Pseudomonas putida* (EC 1.2.1.-); 33, *Pseudomonas* sp. (EC 1.2.1.45); 34, *Escherichia coli* (EC 1.2.1.8); 35, *Homo sapiens* (EC 1.2.1.3); 36, *Beta vulgaris* (EC 1.2.1.8); 37, *Spinacia oleracea* (EC 1.2.1.8); 38, *Atriplex hortensis* (EC 1.2.1.8); 39, *Hordeum vulgare* (EC 1.2.1.8); 40, *Bacillus subtilis* (gbsA); 41, *Zea mays* (EC 1.2.1.9); 42, *Pisum sativum* (EC 1.2.1.9); 43, *Streptococcus mutans* (EC 1.2.1.9); 44, *Bacillus subtilis* (aldY); 45, *Bacillus subtilis* (yfmt); 46, *Bacillus stearothermophilus* (EC 1.2.-.-); 47, *Bacillus subtilis* (ycbD); 48, *Bacillus subtilis* (protéine ipa-76d); 49, *Bacillus subtilis* (ycgN); 50, *Bos primigenius taurus* (EC 1.2.1.27); 51, *Rattus norvegicus* (EC 1.2.1.27); 52, *Pseudomonas aeruginosa* (EC 1.2.1.27); 53, *Brassica napus* (btg-26); 54, *Pisum sativum*; 55, *Homo sapiens* (antiquine); 56, *Escherichia coli* (EC 1.2.1.16); 57, *Bacillus subtilis* (ycnH); 58, *Rattus norvegicus* (EC 1.2.1.4); 59, *Homo sapiens* (EC 1.2.1.5); 60, *Rattus norvegicus* (EC 1.2.1.3); 61, *Homo sapiens* (ALDH7); 62, *Bacillus subtilis* (protéine ipa-58r); 63, *Bacillus subtilis* (aldX); 64, *Pseudomonas oleovorans* (EC 1.2.1.3); 65, *Saccharomyces cerevisiae* (EC 1.2.1.3); 66, *Saccharomyces cerevisiae* (EC 1.2.1.-); 67, *Escherichia coli* (EC 1.2.1.3); 68, *Escherichia coli* (EC 1.2.1.-); 69, *Escherichia coli* (EC 1.2.1.22); 70, *Methanobacterium thermoautotrophicum* (GAPDH); 71, *Aquifex aeolicus*; 72, *Methanococcus jannaschii* (EC 1.2.1.-); 73, *Pseudomonas* sp. (EC 1.2.1.-); 74, *Escherichia coli* (b1525).

Le consensus entre les séquences est matérialisé par une séquence de caractères qui, selon la situation rencontrée en une position donnée, seront les suivants :

"." si le caractère majoritaire a une représentation $\geq 20\%$

":" si le caractère majoritaire a une représentation $\geq 40\%$

"+" si le caractère majoritaire a une représentation $\geq 60\%$

"*" si le caractère majoritaire a une représentation $\geq 80\%$

le caractère lui-même, si celui-ci est identique pour l'ensemble des séquences.

	100	110	120	130	140	150	160	170	180					
1	-----	KKFPVFN	PATEEELC	--QVEEGDK	EDVDKAVKAARQAFQIGS	--PWRTMD	ASERGRLLYKLADL	IERDRLL	LATMES	108				
2	-----	KKFPVFN	PATEEELC	--QVEEGDK	EDVDKAVKAARQAFQIGS	--PWRTMD	ASERGRLLYKLADL	IERDRLL	LATMES	108				
3	-----	KKFPVFN	PATEEELC	--EVEEGDK	EDVKAARQAFQIGS	--PWRTMD	ASERGRLLYKLADL	IERDRLL	LATMES	107				
4	-----	KKFPVFN	PATEEELC	--EVEEGDK	EDVDKAVKAARQAFQIGS	--PWRTMD	ASERGRLLNKLADL	IERDRLL	LATMEA	108				
5	-----	KKFPVFN	PATEEELC	--HVEEGDK	ADVDKAVKAARQAFQIGS	--PWRTMD	ASERGRLLNKLADL	MERDRLL	LATIEA	108				
6	-----	KKFPVFN	PATEEELC	--HVEEGDK	ADVDKAVKAARQAFQIGS	--PWRTMD	ASERGRLLNKLADL	MERDRLL	LATIEA	108				
7	-----	KKFPVFN	PATEEELC	--HVEEGDK	ADVDKAVKAARQAFQIGS	--PWRTMD	ASERGRLLNKLADL	MERDRLL	LATIEA	108				
8	-----	KKFPVFN	PATEEELC	--HVEEGDK	ADVDKAVKAARQAFQIGS	--PWRTMD	ASERGRLLNKLADL	MERDRLL	LATIEA	108				
9	-----	KKFPVFN	PATEEELC	--EVAEGDK	ADVDKAVKAARQAFQIGS	--PWRTMD	ASERGRLLNKLADL	MERDRLL	LATIEA	116				
10	-----	RVFPVFN	PATEEELC	--EVQEGDK	ADVDKAVKAARQAFQIGS	--VWRRMD	ASERGRLLNKLADL	VERDRAT	LATMES	106				
11	-----	KKFPVFN	PATEEELC	--EVEEGDK	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRAT	LAALET	119				
12	-----	KTFFPTV	NPSTGEVIC	--QVAEGNK	EDVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLYRLADL	IERDRTY	LAALET	126				
13	-----	KTFFPTV	NPSTGEVIC	--QVAEGNK	EDVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLYRLADL	IERDRTY	LAALET	126				
14	-----	KTFFPTV	NPSTGEVIC	--QVAEGNK	EDVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLYRLADL	IERDRTY	LAALET	107				
15	-----	KTFFPTV	NPSTGEVIC	--QVAEGNK	EDVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLYRLADL	IERDRTY	LAALET	124				
16	-----	KTFFPTV	NPSTGEVIC	--HVAEGDK	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	IERDRTY	LAALET	127				
17	-----	KTFFPTV	NPSTGEVIC	--HVAEGDK	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	124				
18	-----	KTFFPTV	NPSTGEVIC	--EVQEGDK	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	105				
19	-----	KAYSVIN	PSTGKIC	--DVQEGDK	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	103				
20	-----	KTFFPTV	NPSTGEVIC	--SVQEAET	EKDVDAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	103				
21	-----	KTFFPTV	NPSTGEVIC	--SVQEAET	EKDVDAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	103				
22	-----	KTFFPTV	NPSTGEVIC	--SVQEAET	EKDVDAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	104				
23	-----	KTFFPTV	NPSTGEVIC	--HVAEGDK	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	148				
24	-----	KTFFPTV	NPSTGEVIC	--HVAEGDK	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	144				
25	-----	KTFFPTV	NPSTGEVIC	--HVAEGDK	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	104				
26	-----	ATFDTPN	PATGETLM	--TLYEAQA	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	106				
27	-----	KTFFPTV	NPSTGEVIC	--HVAEGDK	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	106				
28	-----	KTFFPTV	NPSTGEVIC	--HVAEGDK	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	106				
29	-----	KTFFPTV	NPSTGEVIC	--HVAEGDK	ADVDKAVKAARQAFQIGS	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	103				
30	-----	EYQNLTP	PVTGQLLC	--EVAASGR	KRIDLALDAHAKVVDK	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	139				
31	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	89				
32	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	89				
33	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	88				
34	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	91				
35	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	95				
36	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	97				
37	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	94				
38	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	96				
39	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	98				
40	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	89				
41	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	102				
42	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	100				
43	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	87				
44	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	91				
45	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	92				
46	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	95				
47	-----	KLFDVNP	PANGQVIG	--RVHEAGR	AEVDAAVRAARAALKG	--PWRRMD	ASERGRLLNKLADL	VERDRVY	LASLET	95				
48	LVINGEKIE	--TDRKI	IISINP	ANKEEIG	--YASTADQ	GELEAKAMQAAALQAFDS	--WKRQ	PEHRANIL	FKAAAL	LRRRKH	EFSYLV	120		
49	LVINGERVE	--TEAKI	VSINP	ADKEEIVG	--RVSKASQ	EHAEQAIQAAAKAFEE	--WRYT	SPEERAAV	LFRAAK	VRRRKH	HEFSALLV	120		
50	LFIDGKFIESKSDKWDIHN	PATNEVIG	--RVPEST	KAEMDA	AVSSCKRTFPA	--WADT	SILSRQ	QVLLRY	QQLIKEN	LKEIAR	LIM	126		
51	LFIDGKFVIESKSDKWDIHN	PATNEVIG	--RVPEST	KAEMDA	AVSSCKRTFPA	--WADT	SILSRQ	QVLLRY	QQLIKEN	LKEIAR	LIM	124		
52	HLIAGAFVEGLGAPRIPVSN	PLDNSTLA	--ELACAS	AEBQEA	AVASARFAS	--WKET	PVSEAR	VMLRYQ	ALLIKEN	LKEIAR	LIM	89		
53	NYVGGKWL	--NGPLV	STLNP	PANNQVLP	PIAQVVEAS	LEDEYI	GLKACE	EAAKT	--WMQV	PAPKRG	DIVRQ	IGDALRS	KLDYLGRLLS	106
54	SFINGQWKA	--NGPTV	HVSNP	STNQVI	--ASVTEA	TLDYEBGL	RASSEAAKT	--WRTV	PAPKRG	GEIVR	QIGDAL	RAKLDPLGR	LVA	103
55	GVYNGSWG	--RGEVIT	TYCPAN	NEPI	--ARVRQAS	VADYET	VKKAREAWKI	--WADI	PAPKRG	GEIVR	QIGDAL	REKIQV	LGSLVS	105
56	-----	EAIDVTN	PANGDKLG	--SVPKMG	AEDRAA	IDAANRALPA	--WRAL	TAKERATI	LNRW	NFMM	HEHDD	LARLMT	95	
57	-----	DQLTVYN	PATGEEIK	--TIPQGS	ATEVEE	AIERSHQAFK	--WSKT	SANERTS	LLK	KWYEL	IVEH	KEELADLIT	72	
58	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	49
59	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	49
60	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	46
61	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	49
62	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	50
63	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	54
64	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	65
65	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	129
66	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	121
67	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	106
68	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	86
69	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	91
70	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	83
71	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	86
72	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	80
73	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	87
74	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	84
C	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

.....+*:::.. +:.....+::*::+::+::.. +.....:~*..*::+:::.....+:::..

	640	650	660	670	
1	-----	-----	-----	-----	502
2	-----	-----	-----	-----	501
3	-----	-----	-----	-----	500
4	-----	-----	-----	-----	501
5	-----	-----	-----	-----	501
6	-----	-----	-----	-----	501
7	-----	-----	-----	-----	501
8	-----	-----	-----	-----	501
9	-----	-----	-----	-----	509
10	-----	-----	-----	-----	499
11	-----	-----	-----	-----	512
12	-----	-----	-----	-----	519
13	-----	-----	-----	-----	519
14	-----	-----	-----	-----	500
15	-----	-----	-----	-----	517
16	-----	-----	-----	-----	520
17	-----	-----	-----	-----	517
18	-----	-----	-----	-----	497
19	-----	-----	-----	-----	496
20	-----	-----	-----	-----	495
21	-----	-----	-----	-----	496
22	-----	-----	-----	-----	497
23	-----	-----	-----	-----	542
24	-----	-----	-----	-----	538
25	-----	-----	-----	-----	498
26	-----	-----	-----	-----	495
27	-----	-----	-----	-----	506
28	-----	-----	-----	-----	506
29	-----	-----	-----	-----	506
30	-----	-----	-----	-----	542
31	-----	-----	-----	-----	486
32	-----	-----	-----	-----	486
33	-----	-----	-----	-----	484
34	-----	-----	-----	-----	490
35	-----	-----	-----	-----	493
36	-----	-----	-----	-----	500
37	-----	-----	-----	-----	497
38	-----	-----	-----	-----	502
39	-----	-----	-----	-----	505
40	-----	-----	-----	-----	490
41	-----	-----	-----	-----	498
42	-----	-----	-----	-----	496
43	-----	-----	-----	-----	475
44	-----	-----	-----	-----	485
45	-----	-----	-----	-----	485
46	-----	-----	-----	-----	488
47	-----	-----	-----	-----	488
48	-----	-----	-----	-----	515
49	-----	-----	-----	-----	515
50	MGR-----	-----	-----	-----	537
51	MGR-----	-----	-----	-----	535
52	-----	-----	-----	-----	497
53	-----	-----	-----	-----	494
54	-----	-----	-----	-----	508
55	-----	-----	-----	-----	511
56	-----	-----	-----	-----	482
57	-----	-----	-----	-----	462
58	KMPRH-----	-----	-----	-----	453
59	KMTQH-----	-----	-----	-----	453
60	SKVSWSKFFLLKQFNKGRLOLLLLVCLVAVAAVIVKDQL	-----	-----	-----	484
61	RRLRMLLVAMEAQGCSTLL-----	-----	-----	-----	468
62	RMIRKILK-----	-----	-----	-----	457
63	-----	-----	-----	-----	445
64	QLIGITVKYLS-----	-----	-----	-----	483
65	-----	-----	-----	-----	511
66	RWPIH-----	-----	-----	-----	533
67	-----	-----	-----	-----	495
68	-----	-----	-----	-----	468
69	-----	-----	-----	-----	479
70	-----	-----	-----	-----	455
71	-----	-----	-----	-----	476
72	-----	-----	-----	-----	463
73	-----	-----	-----	-----	483
74	-----	-----	-----	-----	470
C					

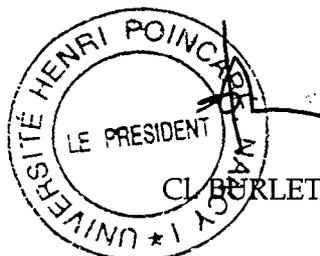
Monsieur COBESSI David

DOCTORAT de l'UNIVERSITE HENRI POINCARÉ, NANCY-I
en BIOLOGIE STRUCTURALE

VU, APPROUVÉ ET PERMIS D'IMPRIMER

Nancy, le 26/06/00

Le Président de l'Université



☞☞☞☞☞☞☞☞☞☞

La structure de l'ALDH de *Streptococcus mutans*, enzyme tétramérique de 200 kDa, peut être décrite comme un dimère de dimère présentant une symétrie non cristallographique 222. Le monomère de l'enzyme se divise en trois domaines: le domaine de fixation du cofacteur (qui contient le domaine de Rossmann), le domaine catalytique et un petit domaine d'oligomérisation. Les deux premiers domaines sont composés de brins β et d'hélices α alors que le troisième domaine est un feuillet β mixte. L'apoenzyme est en interaction avec plusieurs ions sulfate dont certains miment les groupements phosphate du NADP et du substrat, le D-G3P. La comparaison de différentes structures d'ALDH (ALDH2, ALDH3 et ALDH9) montre un positionnement des domaines différents entre les ALDH. Ce phénomène est probablement dû à une orientation différente du domaine d'oligomérisation entre les ALDH tétramériques et les ALDH dimériques. L'analyse des alignements de séquences et des comparaisons de structures ont permis de proposer un rôle de déterminant de l'état oligomérique à certains acides aminés.

L'étude structurale de l'holoenzyme a permis de mettre en évidence le nouveau mode de fixation du NADP au domaine de Rossmann (*i.e.* non perpendiculaire au plan du feuillet β parallèle) et les facteurs moléculaires majeurs responsables de la spécificité de cofacteur. La spécificité de cofacteur pour le NADP est due en partie à la présence d'une thréonine ou d'une sérine pour ce type d'ALDH. Ces résidus sont mutés en proline ou en glutamate chez les ALDH NAD-dépendantes. Ainsi la spécificité de cofacteur serait due à l'absence de charge et de contrainte stérique à proximité du 2'-phosphate du NADP. La comparaison structurale de l'apoenzyme et de l'holoenzyme montre deux changements locaux de conformation dus à la fixation du NADP et non observés dans les structures d'ALDHs NAD-dépendantes. Un premier se produit au niveau de la boucle qui connecte le deuxième brin β du domaine de Rossmann à la seconde hélice α de ce domaine. Le second est la rotation de la chaîne latérale de la cystéine 284. Cette rotation positionne son groupement sulphydrique à proximité des dipôles de la chaîne principale et l'oriente vers le site catalytique. Ce changement serait à l'origine de la baisse de son pKa et de son changement d'accessibilité nécessaires à l'étape d'acylation. Au cours de l'acte catalytique, différents intermédiaires réactionnels sont formés et stabilisés par l'environnement protéique. L'étude d'une seconde structure d'apoenzyme a révélé la présence d'un second ion sulfate dans le site catalytique en interaction avec la fonction amide de l'asparagine conservée Asn154 et avec le groupement NH de Cys284. Cette interaction est également observée avec la fonction aldéhydique du substrat dans la structure du complexe C284S-NADP-G3P suggérant la présence d'un site oxyanion stabilisant différents intermédiaires réactionnels. Dans cette structure, le glutamate Glu250 se situe à 3 Å de Ser284 au lieu de 7 Å (observé dans les structures de l'holoenzyme et de l'apoenzyme de type sauvage) ou de 3,4 Å (structure de la seconde apoenzyme). La rotation de la chaîne latérale de Glu250 entraîne la perte du pont salin avec Arg437 et positionne son groupement carboxylate dans un environnement hydrophobe. Ces deux phénomènes sont probablement à l'origine de la valeur de pKa de 7,6 attribuée à Glu250.

La comparaison des différentes structures d'ALDH a permis de mettre en évidence la flexibilité du NAD(P) en plus de celle de Glu250. Les différentes positions et conformations observées pour le cofacteur chez les ALDH résultent probablement de son mode particulier de fixation au domaine de Rossmann et du faible nombre d'interactions qu'ils échangent. Le mode de fixation serait dû à la longueur de la première boucle du domaine de Rossmann conservée chez les ALDH et dont l'orientation serait contrainte par la présence d'une proline conservée. La flexibilité du cofacteur et du glutamate suggèrent leurs mouvements au cours de l'acte catalytique.

Mots clé : ALDH, structure cristalline, remplacement moléculaire, spécificité de cofacteur, mécanisme catalytique, mutant, apoenzyme, holoenzyme.