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

FACULTE DES SCIENCES ET TECHNOLOGIES

Ecole Doctorale Energie Mécanique et Matériaux (EMMA)

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy Université

en sciences physiques

par Dorothee Claudia LIEBSCHNER

Propriétés électrostatiques et structurales des protéines diffractant à haute résolution

Thèse dirigée par Dr. Benoît GUILLOT et Pr. Claude LECOMTE

soutenue le 10 novembre 2010

Président du Jury :

P. Goudeau	Directeur de Recherche CNRS, Université de Poitiers

Rapporteurs :

A. Podjarny	Directeur de Recherche, IGBMC, Strasbourg
D. Vercauteren	Professeur, PCI, Namur (Belgique)

Examinateurs :

N. E. Ghermani	Professeur, PPB, Université Paris XI
E. Chabrière	Professeur, AFMB, Université de la Méditerranée
B. Berche	Professeur, UHP Nancy Université
B. Guillot	Maître de conférences, CRM^2, UHP Nancy Université
C. Lecomte	Professeur, CRM^2, UHP Nancy Université

Laboratoire de Cristallographie, Résonance Magnétique et Modélisations
Groupe Modélisation (CRM^2, UMR 7036)

A ma famille

A Vincent

" Science is facts, just as houses are made of stones,
so is science made of facts ;
but a pile of stones is not a house and
a collection of facts is not necessarily science. "

Henri Poincaré (1824-1912), mathématicien, physicien et philosophe français

Avant-propos

Cette thèse a été dirigée par Monsieur Claude Lecomte qui m'a accueillie au laboratoire CRM^2 dans l'équipe "Modélisation". Je tiens à lui exprimer toute ma reconnaissance car il m'a permis l'opportunié de découvrir la cristallographie et la modélisation des protéines.

Je tiens à remercier particulièrement Monsieur Benoît Guillot qui a encadré de près ce travail depuis mon Master 2. J'ai pu apprécier ses qualités pédagogiques et la richesse de ses connaissances tout au long de ces trois ans. Il a su orienter mes recherches aux bons moments et il a toujours été disponible pour d'intenses et rationnelles discussions. Il m'a été également d'une aide précieuse pour la réalisation de ce manuscrit. Je lui exprime ici toute ma reconnaissance.

Je souhaite vivement remercier Monsieur Eric Chabrière qui m'a introduit à la cristallographie des protéines et qui m'a encadrée pendant mon stage de Master 1 et Master 2. Mêmes après son départ de Nancy, il m'a guidée avec de nombreux conseils et discussions.

Je remercie les rapporteurs de cette thèse Monsieur Alberto Podjarny et Monsieur Daniel Vercauteren d'avoir accepté de juger mon manuscrit et pour l'intérêt qu'ils ont porté à mon travail. Merci également aux autres membres du jury : Monsieur Nour-Eddine Ghermani et Monsieur Philippe Goudeau.

Je souhaite remercier Madame Nouzha Bouhmaida qui m'a aidée à réaliser les sous-routines permettant de calculer le moment quadrupolaire et dipolaire, lors de son séjour à Nancy en avril 2009. Pour les discussions fructueuses et sa patience pour mes questions, je la remercie vivement.

Je remercie Monsieur Yves Bourne, directeur du laboratoire AFMB pour m'avoir accueillie au sein de cette institution en juillet 2008 et juillet 2009. J'ai pu y gagner de l'expérience avec les techniques de purification et du clonage. Je tiens à remercier également Guillaume Gotthard qui m'a aidée à réaliser les expériences de biochimie à l'AFMB.

Je souhaite remercier tous les membres du laboratoire CRM^2 pour leurs conseils, les discussions et leur encouragement. En particulier, je tiens à remercier Madame Anne Clausse et Madame Valérie Richalet pour leur aide et leur sympathie, à Monsieur Enrique Espinosa pour des discussions scientifiques fructueuses et à Monsieur Christian Jelsch pour son aide à l'utilisation du logiciel MoPro et son intérêt pour mon travail.

Mes remerciements vont également à Monsieur Guillermo Mulliert et à Madame Brigitte L'Herbier, qui m'ont confié l'enseignement de leurs étudiants pour des TD en informatique et des TP en physique.

Je souhaite enfin remercier mes collègues doctorants ou post-docs, anciens ou récents, Chasan, Sébastien, Moutsé, Bertrand, Agnieszka, Mariya, Isabella, Ru-Fen, Parthapratim, Slavomir, Dorothea et Maqsood pour des discussions intéressantes et leur encourageante présence pendant ma thèse. En particulier, les discussions et conseils de Mikael Elias m'ont permis d'intégrer les notions de chimie et de biochi-

mie qui me manquaient en tant que physicienne.

Pour leurs encouragements et leur assistance morale qui m'ont permis de faire cette thèse dans de bonnes conditions, je remercie Vincent Ancel ainsi que ma famille.

Table des matières

Introduction **11**

I Rappels théoriques 15

1 Les protéines **17**
1.1 Introduction 17
1.2 La structure des protéines 17
1.3 Motifs de structure secondaire 19
1.3.1 Hélice α 19
1.3.2 Hélice 3_{10} 21
1.3.3 Feuillet β 22

2 La diffraction des rayons X **24**
2.1 Introduction 24
2.2 Principe de l'expérience 24
2.3 Le facteur de structure 26
2.4 Modèles de la densité électronique 28
2.4.1 Le modèle de l'atome indépendant (IAM) 28
2.4.2 Le modèle de Hansen et Coppens 30
2.5 L'affinement cristallographique 32
2.6 La banque de données ELMAM 35
2.7 Topologie de la densité électronique 36
2.8 Calcul du potentiel électrostatique 38

II Moments électrostatiques 39

3 Méthodes de calcul des moments **43**
3.1 Calcul d'un moment électrostatique 43
3.1.1 Définition des moments électrostatiques 43

3.1.2 Dépendance de l'origine 45
3.1.3 Moments électrostatiques atomiques 46
3.1.4 Moments électrostatiques totaux 47
3.2 Calcul de l'axe d'une hélice 48

4 Moments électriques de l'hélice α 50
4.1 Introduction 50
4.2 Historique 52
4.3 Caractérisation du dipôle de l'unité peptidique 54
4.4 Caractérisation du dipôle d'hélices α 57
4.4.1 Hélice de polyalanine idéalisée 57
4.4.2 Hélices réelles 60
4.5 Moment quadrupolaire de l'hélice α idéalisée 70

5 Moments électriques du feuillet β 75
5.1 Le moment dipolaire du feuillet β 75
5.1.1 Feuillet parallèle idéal 78
5.1.2 Feuillet antiparallèle idéal 81
5.1.3 Feuillets β voilés 85
5.1.4 Moments dipolaires de repliements communs 86
5.2 Le quadrupôle du feuillet β 90
5.2.1 Feuillet parallèle idéal 91
5.2.2 Feuillet antiparallèle idéal 93
5.2.3 Feuillets β voilés 95
5.2.4 Moments quadrupolaires de repliements communs . . . 95
5.3 Conclusion 98
5.4 Perspectives 98

III Topologie des interactions dans les motifs de structure secondaire 101

6 Topologie des interactions dans les hélices et les feuillets 103
6.1 Introduction 103
6.2 Description de la liaison hydrogène 104
6.3 Topologie de la densité électronique des liaisons H 106
6.4 Etudes sur les liaisons H des hélices dans la littérature 108
6.5 Matériels et méthodes 109
6.6 Résultats 112
6.6.1 Répartition des points critiques et géométrie des interactions 112

6.6.2 Interactions dans les hélices . . . 114
6.6.3 Liaisons hydrogène dans les feuillets . . . 135
6.7 Conclusion . . . 138
6.8 Perspectives . . . 138

IV Protéines diffractant à haute résolution 141

7 La protéine DFPase 147
7.1 Introduction . . . 147
7.2 La diffraction des neutrons . . . 152
7.3 Manipulations et résultats . . . 153
7.3.1 Affinement : structure à rayons X . . . 153
7.3.2 La densité nucléaire de la DFPase . . . 155
7.3.3 Analyse des liaisons hydrogène . . . 159
7.3.4 Origine des différences . . . 162
7.3.5 Densité nucléaire calculée . . . 164
7.3.6 Etude du site actif de la DFPase . . . 174
7.3.7 Mécanisme catalytique . . . 177
7.4 Conclusion . . . 178
7.5 Perspectives . . . 179

8 La protéine PfluDING 180
8.1 Introduction . . . 180
8.2 Manipulations et résultats . . . 182
8.2.1 Cristallisation - Collecte des données . . . 182
8.2.2 Affinement de la structure . . . 183
8.2.3 Le site de fixation du phosphate . . . 186
8.2.4 L'interaction phosphate O2 - Asp 62 $O\delta 2$ . . . 189
8.2.5 Potentiel électrostatique de PfluDING . . . 193
8.2.6 Mécanisme moléculaire de fixation . . . 194
8.3 Conclusion . . . 196
8.4 Perspectives . . . 197
8.5 Publication . . . 198

Conclusion générale 207

Appendices 212

A Notions utilisées en biologie macromoléculaire **213**
A.1 Abréviations utilisées dans le rapport 213
A.2 La Proteine Data Bank (PDB) 213
A.3 Représentation de la densité électronique 214
A.4 Les 20 acides aminés 216

B Divers **217**
B.1 Détails de calculs 217
B.1.1 Dépendance de l'origine 217
B.1.2 Intégration du moment quadrupolaire 218

Introduction

Le présent travail se situe dans le contexte de la cristallographie des protéines, un domaine scientifique pluridisciplinaire qui regroupe des méthodes venant de la biologie, de la chimie et de la physique et qui vise à déterminer et analyser la structure protéique grâce à une expérience de diffraction. Il s'agit d'un domaine à part entière puisque les méthodes utilisées ont du être adaptées à la nature de l'échantillon, les protéines. Ces macromolécules biologiques comptent des milliers d'atomes et nécessitent un maniement précautionneux, étant sensibles à la température, au pH et à d'autres paramètres.
La première structure protéique a été déterminée par cristallographie des rayons X en 1958 [1]. Le fait que 35 ans ont du s'écouler depuis les premières analyses structurales de modèles obtenus par la méthode de diffraction sur monocristal (W.H. Bragg et W.L. Bragg, 1913) reflète la complexité des protéines. En effet, avant d'effectuer l'analyse de la structure d'une protéine, de nombreuses étapes sont nécessaires : par exemple la production (appelée *expression*), la purification, la cristallisation et la collecte et le traitement des données expérimentales représentent autant d'étapes difficiles et chronophages.
Néanmoins, des progrès technologiques, informatiques et méthodologiques ainsi que l'interconnexion du savoir-faire [1] ont pu stimuler l'avancée du domaine de la cristallographie des protéines. Actuellement, 58500 structures protéiques on été résolues [2] par diffraction des rayons X. Outre le fait que le nombre de structure a rapidement augmenté, la qualité des données de diffraction a pu être améliorée également. De ce fait, il y a de plus en plus de structures déterminées à résolution subatomique. A une telle résolution, le niveau de détail est plus fin et permet de faire une analyse poussée de la structure. Bien que le nombre de protéines diffractant à haute résolution soit encore petit comparé à l'échantillon total, il augmente constamment.
Or, l'existence des structures protéiques obtenues à haute résolution requiert

1. A citer est notamment la "Protein data bank" (PDB) qui rassemble la plupart des structures de protéines résolues par différentes méthodes.
2. En aout 2010.

l'application de nouvelles méthodes d'analyse. Une exemple à citer est certainement l'affinement multipolaire de la densité électronique des protéines Crambine et Aldose Reductase [2], [3]. Cette méthode est normalement utilisée pour les composés de petite taille mais elle a été adaptée aux protéines diffractant à ultra-haute résolution.
Le laboratoire CRM^2 est spécialisé dans le domaine de la densité de charge et de son extension aux macromolécules biologiques. Une banque de données (ELMAM) y a été développée, permettant de transférer les paramètres d'un modèle asphérique de densité électronique sur une structure protéique. De la densité électronique découlent de nombreuses propriétés, comme le potentiel électrostatique ou le moment électrique qui deviennent ainsi accessibles à l'expérimentateur. De plus, une analyse topologique de la densité électronique peut révéler des informations précieuses sur les interactions entre différents atomes, comme par exemple les liaisons hydrogène.
Ce travail porte sur l'application de ces méthodes aux protéines, plus précisément, il traîte des propriétés électrostatiques et structurales des protéines qui diffractent à haute résolution. La compréhension de ces propriétés est essentielle à l'éclaircissement du fonctionnement des protéines puisque leur structure est reliée à leur fonction et leurs propriétés électrostatiques permettent d'en comprendre la réactivité.
Ce manuscrit se divise en 4 grandes parties :

- La première partie est consacrée à des rappels théoriques. Elle commence avec la description de la structure d'une protéine (chapitre 1) et se focalise sur les motifs de structure secondaire. Ensuite, les techniques et méthodes utilisées dans le cadre de la cristallographie macromoléculaire sont regroupées, de la théorie de la diffraction au traitement et à l'analyse des données expériementales (chapitre 2). Enfin, les différents modèles de la densité électronique ainsi que les notions de topologie de la densité électronique ont été abordés.

- La deuxième partie porte sur la caractérisation des moments électrostatiques créés par les motifs de structure secondaire : les hélices et les feuillets (chapitres 3 à 5).

- Dans la troisième partie, l'analyse topologique de la densité électronique est appliquée pour la première fois aux liaisons hydrogènes dans les motifs de structure secondaire.

- Enfin, la quatrième partie regroupe deux applications de la cristallographie macromoléculaire à haute résolution. D'abord, le mode de fixation d'un ion phosphate dans la protéine PfluDING est analysé. Ensuite,

le réseau des liaisons hydrogènes dans la protéine DFPase est comparé entre une structure par rayons X, déterminée à haute résolution, et une structure obtenue par diffraction des neutrons.

Une partie annexe regroupe les notions de biologie macromoléculaire et quelques détails de calculs nécessaires à la compréhension de ce manuscrit.

Première partie

Rappels théoriques

Chapitre 1

Les protéines

1.1 Introduction

Les protéines sont des macromolécules biologiques qui constituent un composant essentiel pour tous les êtres vivants. Leurs fonctions sont très diverses : par exemple, elles transportent les substances primordiales dans les organismes, catalysent des réactions chimiques, régulent la transcription génétique et fabriquent d'autres protéines.
Une protéine est une macromolécule constituée d'une ou plusieurs chaînes d'acides aminés qui se replient et qui ont ainsi une structure tridimensionnelle particulière. La structure détermine les propriétés fonctionnelles de la protéine et explique sa spécificité. Par conséquent, c'est un défi de comprendre comment les protéines se replient et acquièrent leur propriétés bien définies. De plus, il est intéressant d'éclaircir la fonction biologique des protéines pour mieux comprendre leurs mécanismes d'action et pour pouvoir ainsi modifier leurs activités ou leurs spécificités afin de synthétiser des catalyseurs biologiques ou créer des médicaments.

1.2 La structure des protéines

Les protéines sont des molécules comptant de quelques centaines à plusieurs milliers d'atomes, dont principalement les éléments hydrogène, oxygène, carbone, azote et soufre. Ses éléments constitutifs sont les acides aminés (AA) qui sont liés l'un à l'autre par des liaisons peptidiques. Une protéine peut contenir de plusieurs dizaines jusqu'à plusieurs milliers d'acides aminés. Il y a 20 acides aminés différents, une multitude de combinaisons est ainsi disponible pour former la séquence d'une protéine.
Les acides aminés sont des molécules organiques ayant la formule chimique

générale H2N-CH-R-COOH, où R est un groupement variable propre à chacun des 20 acides aminés. La partie H_2N est le groupement amino, la partie COOH est le groupement carboxyl. La structure générale d'un acide aminé est représentée figure 1.1. Les différents acides aminés ne diffèrent que par leur résidu R [1].

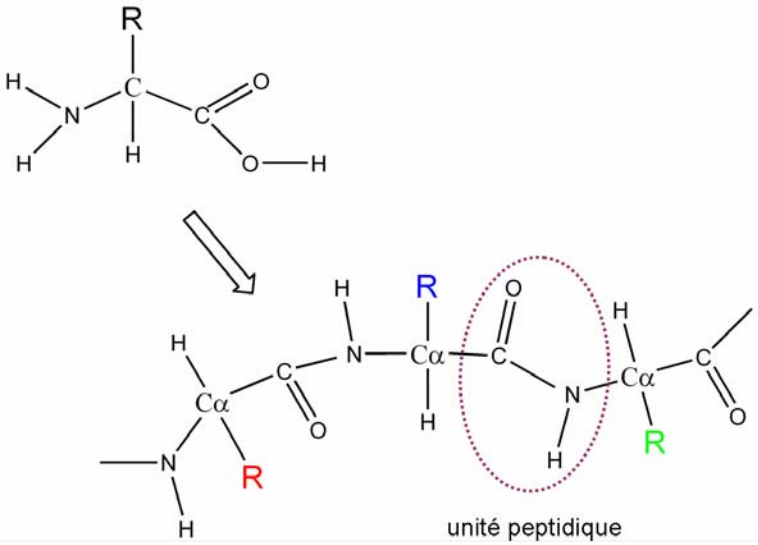


Figure 1.1 – Structures d'un acide aminé et d'une chaîne polypeptidique de trois acides aminés.

Dû à la complexité des protéines, il est pratique de ne pas se focaliser sur chaque atome pour décrire leur structure ou pour faire des comparaisons. On se réfère plutôt au repliement global de la chaîne polypeptidique. Ainsi, pour décrire la structure d'une protéine, il existe 4 niveaux de structuration ou d'échelle (figure 1.2)[4] :

Structure primaire : La structure primaire, ou séquence, correspond à la succession linéaire des acides aminés qui constituent la protéine. Par convention, le premier acide aminé est celui qui possède une extrémité amine libre (= N-terminal), le dernier acide aminé est celui qui possède une extrémité carboxylate libre (C-terminal).

Structure secondaire : La structure secondaire décrit le repliement local de la chaîne principale. Ces repliements sont souvent stabilisés par des liaisons hydrogènes entre les groupements amide (N-H) et carbonyle

1. L'acide aminé proline fait exception à la structure des autres acides aminés, puisque l'anneau pyrrolidine forme une liaison covalente avec l'atome d'azote de l'unité peptidique (l'atome d'hydrogène n'est alors pas présent).

(C=O). Les trois principales structures secondaires sont les hélices α, les feuillets β et les coudes.

Structure tertiaire : Les éléments de structure secondaire sont liés entre eux par des boucles et des tours non structurés. L'arrangement résultant est la structure tertiaire qui définit le repliement global de la protéine.

Structure quaternaire : Il existe des protéines qui sont constituées d'un assemblage de protéines (différentes ou pas). La structure quaternaire décrit l'arrangement des chaînes peptidiques différentes.

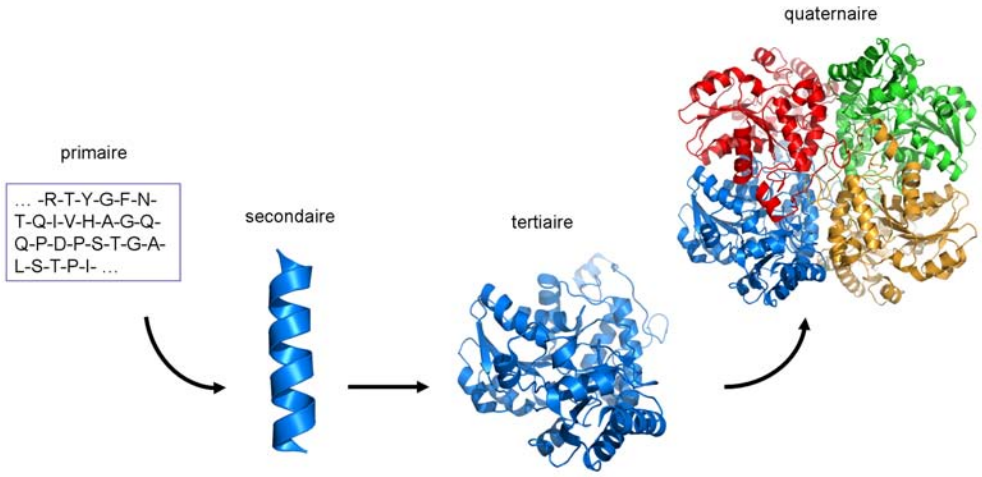


Figure 1.2 – De la structure primaire à la structure quaternaire. (Données de la protéine L-methionine gamma-lyase, code PDB 2RFV)

Une partie de ce travail traitera des propriétés électrostatiques des structures secondaires. Par conséquent, les hélices et les feuillets seront décrits de manière plus détaillée dans les sections suivantes.

1.3 Motifs de structure secondaire

1.3.1 Hélice α

L'hélice α est un motif fréquent dans la structure des protéines. Une étude sur des structures déposées dans la PDB montre que 33% des acides aminés adoptent une géometrie α-hélicoïdale [5]. Dans ce motif, la chaîne principale est enroulée sur elle-même pour former une hélice. Les résidus pointent vers l'extérieur de l'axe de l'hélice, légèrement incliné vers la partie N-terminale (figure 1.3). Des liaisons hydrogènes entre le groupement C=O

d'un acide aminé et le groupement N-H d'un acide aminé situé 4 résidus plus loin stabilisent l'enroulement. A l'extrémité, l'enroulement régulier est souvent perturbé et les liaisons hydrogène se forment plutôt entre les résidus i et i+3 [6].

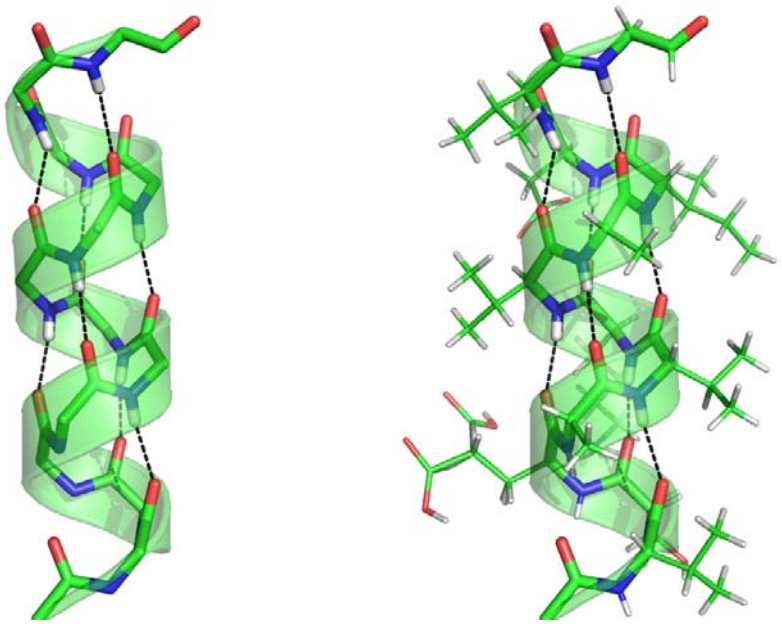


Figure 1.3 – Gauche : Hélice α sans chaînes latérales. Droite : Avec chaînes latérales. Le N-terminal est en bas, le C-terminal est en haut. Les liaisons hydrogène entre les unités peptidiques sont affichées en lignes pointillées (données structurales venant de l'Aldose Reductase [7]).

La translation par acide aminé est 1.5 Å, ce qui correspond à un pas angulaire de 100°. Ainsi, 3.6 résidus sont nécessaires pour faire un tour complet. Les angles dièdres occupent une région précise dans le diagramme de Ramachandran ($\phi \approx -60°$ et $\psi \approx -45°$). Les hélices α peuvent atteindre une longueur de 40 AA, mais en moyenne, elles sont constituées de 12.6 résidus [5]. Un segment hélicoïdal peut être défini soit en se basant sur les angles dièdres soit par l'arrangement des liaisons hydrogène.
De manière intéressante, il y a encore d'autres liaisons hydrogène dans l'hélice α qui stabilisent davantage ce motif. L'oxygène carbonyle peut être impliqué dans des doubles liaisons H, c'est-à-dire qu'il interagit avec deux atomes d'hydrogène [6]. La première liaison est entre l'oxygène O_i et l'hydrogène H_{i+4}, telle qu'elle est dans la définition d'une hélice α. La deuxième interaction est avec un atome hydrogène distant de seulement 3 résidus. Par conséquent, les atomes d'hydrogène forment souvent des liaisons bifurquées et on observe

une succession de liaisons hydrogènes $O_i...H_{i+3}$ et $O_i...H_{i+4}$ (figure 1.4).
Il est à noter que la plupart des hélices α dans les protéines sont courbées, c'est-à-dire que l'axe de l'hélice n'est pas rectiligne. La courbure peut être due à la présence de prolines, qui n'ont pas de groupement N-H et qui interrompent ainsi la succession de liaisons hydrogène dans l'hélice. Une autre raison peut être l'influence du solvant. L'hélice s'incline pour exposer ses groupements C=O au solvant ce qui favorise la formation de liaisons hydrogène.

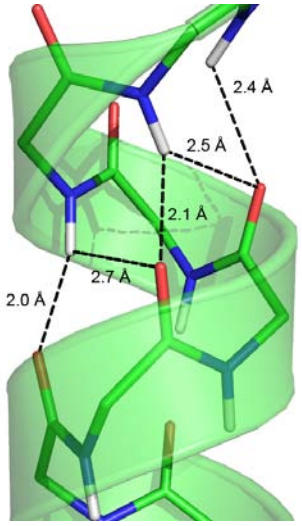


Figure 1.4 – Succession de liaisons hydrogène dans une hélice alpha, montrant les liaisons hydrogène bifurquées. Les liaisons hydrogène sont représentées par des lignes pointillées.

1.3.2 Hélice 3_{10}

Il y a des hélices où la liaison hydrogène se forme entre les résidus de numéro i et i+3. C'est l'hélice[2] 3_{10} dont l'occurrence est beaucoup moins élevée que pour les hélices α, seulement 3.5 % des acides aminés adoptent cette géometrie [5]. Les hélices 3_{10} ont une longueur de 4 à 6 résidus et sont rarement plus longues [5]. Un tour d'hélice correspond à 3 acides aminés, les angles dièdres d'une hélice 3_{10} canonique sont $\phi = -49°$ et $\psi = -26°$ (angles de Perutz). Cependant, une étude sur des structures de protéines déposées dans la PDB montre que les angles dièdres de l'hélice varient en fonction

2. Les hélices sont souvent appelées d'après le schéma n_m, où n est le nombre de résidus par tour et m est le nombre d'atomes dans le tour joint par la liaison hydrogène. L'hélice α est ainsi une hélice 3.6_{13}.

de la longueur [8]. Il paraît alors que la définition de l'hélice 3_{10} en termes d'angles de Ramachandran n'est pas suffisante.
Contrairement à l'hélice α, les liaisons hydrogènes ne sont pas parallèles à l'axe de l'hélice (figure 1.5). Elles sont plutôt inclinées et tendent vers des géométries moins favorables. Donc la distance donneur-accepteur est plus grande et les angles N-H...O sont moins élevés.

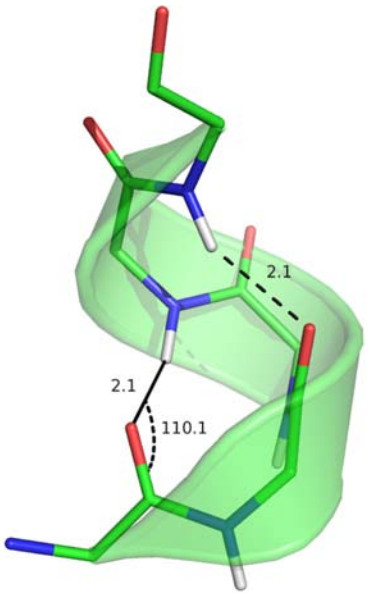


Figure 1.5 – Hélice 3_{10}, les liaisons hydrogène sont représentées en lignes pointillées. La distance H...O et l'angle C-O...H illustrent la géométrie moins favorable de ce type de liaison hydrogène.

1.3.3 Feuillet β

Dans le feuillet β, des portions de la chaîne d'acides aminés sont arrangées parallèlement. Ces portions sont appelées brins. Des liaisons hydrogène entre les unités peptidiques de brins adjacents stabilisent le motif. En général les liaisons hydrogènes ne sont pas entre des acides aminés consécutifs mais entre des AA lointains (contrairement à la situation dans les hélices). Les chaînes latérales pointent alternativement vers le haut et le bas du plan du feuillet. En effet, les chaînes principales ne sont pas exactement dans un plan, mais présentent un plissement sur leur surface. Vue de côté (figure 1.6 b), le repliement ressemble ainsi à des dents de scie. En plus, les brins s'enroulent souvent et, en général, ne sont pas plats. Puisque la chaîne polypeptidique a un sens (allant du N-terminal vers le C-terminal), il y a deux sortes de

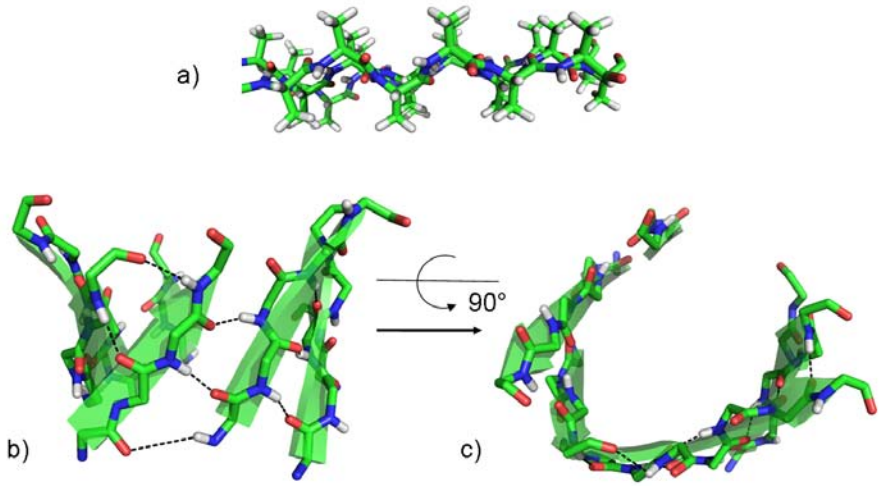


Figure 1.6 – a) Vue de côté d'un feuillet idéal, formé uniquement par des acides aminés alanine. Le repliement de la chaîne principale ressemble à des dents de scie. Les chaînes latérales pointent alternativement vers le haut et vers le bas du feuillet. b) Vue d'un feuillet dans la protéine aldose réductase. Les brins, représentés en flèche, sont enroulés et forment un tonneau. Ils ne sont plus arrangés dans un plan. c) même feuillet que dans b) mais tourné d'un angle droit vers l'horizontale.

feuillets : Dans le feuillet antiparallèle, les brins sont orientés en sens inverse tandis qu'ils sont dans le même sens dans le feuillet parallèle (figure 1.7).

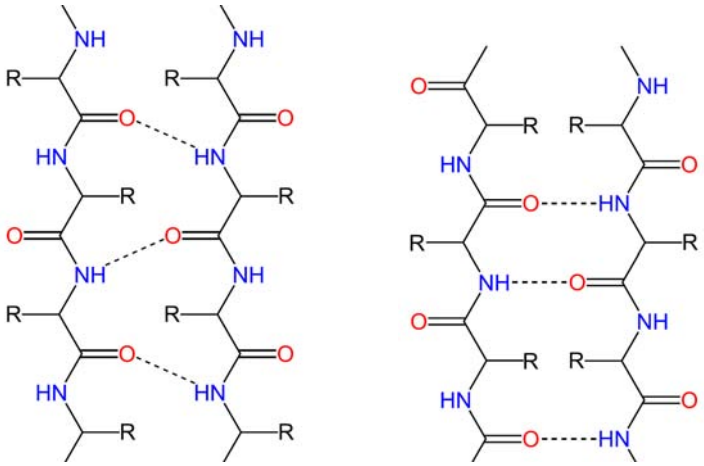


Figure 1.7 – Gauche : Feuillet parallèle. Droite : Feuillet antiparallèle. Les chaînes latérales sortent du plan vers le haut et vers le bas. Les liaisons hydrogène entre les brins sont affichées comme ligne pointillée.

Chapitre 2

La diffraction des rayons X

2.1 Introduction

La diffraction des rayons X est une méthode expérimentale permettant de déterminer la structure d'un échantillon cristallin. Le principe de l'expérience consiste à envoyer un faisceau de rayons X, monochromatique en général, sur un cristal et de détecter l'intensité des rayons diffractés par celui-ci. Les rayons X ont une longueur d'onde de l'ordre de grandeur des distances interatomiques et permettent ainsi d'apporter des informations structurales sur l'échantillon.
Dans ce chapitre, quelques concepts et notions de base, utilisées en cristallographie, seront rappelés.

2.2 Principe de l'expérience

Si un faisceau de rayons X est envoyé sur un cristal, les électrons de l'échantillon commencent à osciller [1] et sont soumis à une accélération. Par conséquent chaque électron émet des ondes électromagnétiques de même énergie, qui interfèrent soit de manière constructive, soit de manière destructive. Les atomes du cristal étant arrangés périodiquement, les rayons transmis sortent du cristal dans des directions privilégiées et leurs intensités sont amplifiées. La condition pour l'observation d'un rayon diffracté dans un

1. Les protons étant également chargés, ils oscillent également. Par contre, leur masse est beaucoup plus élevée que celle des électrons ($\frac{m_e}{m_p} \approx 10^{-4}$) ce qui a pour conséquence de rendre leur signal négligeable par rapport à celui des électrons.

certain angle par rapport au rayon incident est donnée par la loi de Bragg (figure 2.1) :

$$n \cdot \lambda = 2d \cdot sin(\theta) \tag{2.1}$$

n = nombre entier $\in \mathbb{Z}$

λ = longueur d'onde

d = distance interréticulaire (distance entre deux plans réticulaires de la famille de plans (hkl))

θ = angle de Bragg, c'est la moitié de l'angle entre le faisceau incident et le faisceau diffracté

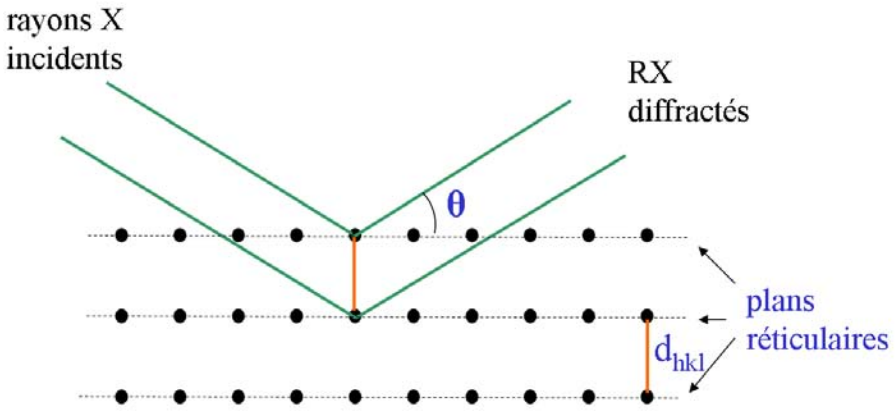


Figure 2.1 – Schéma illustrant la loi de Bragg. Il y a interférence constructive si la différence de marche entre les deux rayons diffractés est un multiple (entier) de λ.

Les rayons diffractés sont détectés sur un cliché de diffraction qui est constitué de taches ou pics de Bragg. Le cliché de diffraction contient plusieurs informations : La position des pics de Bragg est caractéristique des paramètres de maille du cristal tandis que leur intensité est fonction des atomes qui constituent le motif.
On peut déduire de la loi de Bragg que plus la distance interréticulaire d est petite, plus $sin\theta$ (et donc θ) est grand. Ainsi, des plans réticulaires proches l'un de l'autre diffractent à grand angle. Les réflections à grand angle permettent alors de discerner plus de détails dans la densité électronique, ce qui correspond à une résolution élevée des données de diffraction.
En théorie, dans un cristal parfait, des plans réticulaires infiniment proches pourraient donner lieu à la diffraction et satisfaire la condition de Bragg. En réalité, les cristaux présentent souvent des imperfections et il y une limite de

diffraction naturelle (qui varie d'un cristal à l'autre). Les cristaux de protéine diffractent en moyenne [2] autour de 2.2 Å.
L'expérience de diffraction donne alors accès à l'intensité des rayons diffractés par le cristal. Dans la section suivante, il sera expliqué comment l'intensité est liée à la structure de l'échantillon.

2.3 Le facteur de structure

Le facteur de structure $F(\vec{H})$ est un nombre imaginaire qui est proportionnel à l'amplitude du champ électrique dans la direction $\vec{H}$ (figure 2.2), émis par les électrons sous l'influence d'un faisceau de RX incident. Il représente la somme des ondes diffusées par les N_{at} atomes d'une maille. Sa formule est donnée par :

$$F(\vec{H}) = \sum_{j=1}^{N_{at}} f_j(\vec{H}) \cdot exp(2i\pi \vec{H} \cdot \vec{r}_j) \tag{2.2}$$

$\vec{H}$ représente un vecteur du réseau réciproque de coordonnées entières (h, k, l) : $\vec{H} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^*$ et caractérise la direction d'observation (figure 2.2)

$\vec{r}_j$ représente le vecteur position d'un atome j dans la maille, en coordonnées fractionnaires (fractions des paramètres de maille)

$f_j(\vec{H})$ est le *facteur de diffusion atomique* (ou *facteur de forme*), il est donné par la transformée de Fourier de la densité électronique de l'atome :

$$f_j(\vec{H}) = \int_V \rho(\vec{r}) \cdot exp(2i\pi \vec{H} \cdot \vec{r}) dV \tag{2.3}$$

L'intégration s'étend sur le volume de l'atome.

En injectant le facteur de diffusion donné par la formule (2.3) dans l'expression du facteur de structure (2.2), on obtient :

$$F(\vec{H}) = \int_V \rho(\vec{r}) \cdot exp(2i\pi \vec{H} \cdot \vec{r}) dV \tag{2.4}$$

$\rho(\vec{r})$ est dans cette expression la densité électronique de la maille. En appliquant la transformée de Fourier inverse de $F(\vec{H})$, il est donc possible de déterminer la densité électronique à partir de la connaissance des $F(\vec{H})$:

$$\rho(\vec{r}) = \frac{1}{V} \int F(\vec{H}) exp(-2i\pi \vec{H} \cdot \vec{r}) d\vec{H} \tag{2.5}$$

2. La résolution moyenne des structures déposées dans la PDB (structures résolues par rayons X ou par microscopie électronique) est de 2.2 Å au 9/02/2010.

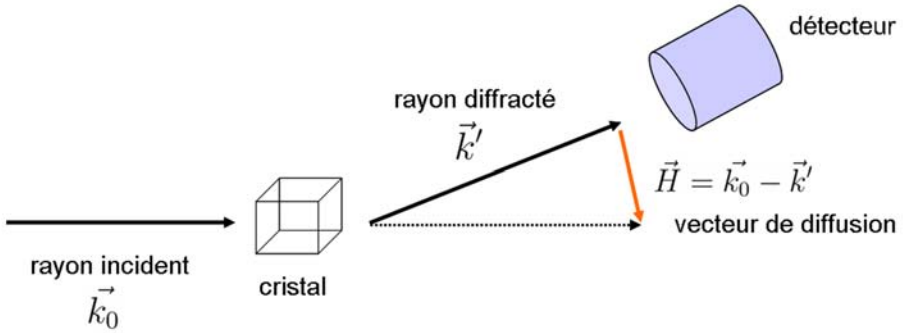


Figure 2.2 – Schéma d'une expérience de diffraction des rayons X. Un faisceau de rayons X de vecteur d'onde $\vec{k}_0$ est envoyé sur le cristal. Le faisceau diffracté par le cristal a un vecteur d'onde $\vec{k'}$. Le vecteur de diffusion est défini par $\vec{H} = \vec{k}_0 - \vec{k'}$.

La densité électronique rapporte des informations sur la structure d'une molécule car un maximum local de densité correspond en général à la position d'un atome [3]. Or, l'intensité des rayons X diffractés par le cristal est proportionnelle au carré du module du facteur de structure :

$$I \propto |F|^2$$

La mesure de $I(\vec{H})$ lors de l'expérience de diffraction donne alors accès au module du facteur de structure $F(\vec{H})$, mais soulignons que la phase de $F(\vec{H})$ est perdue. Afin de pouvoir calculer la densité électronique, ce qui permet ensuite de déterminer la structure de l'échantillon cristallin, il est nécessaire de connaître l'amplitude et la phase de $F(\vec{H})$. Pour résoudre cet enjeu (qui est souvent appelé le "problème de phase") et obtenir une première estimation des phases, il faut se servir soit de méthodes mathématiques, soit de propriétés physiques particulières.
En cristallographie des protéines, les méthodes de phasage les plus utilisées sont le remplacement isomorphe (SIR), le remplacement moléculaire et les méthodes directes [9]. Lors de ma thèse, j'ai utilisé la méthode du remplacement moléculaire, la structure de la protéine (PfluDING) a déjà été résolue à plus basse résolution et pouvait donc servir de modèle homologue.
L'estimation des phases permet ensuite d'établir un premier modèle structural de la molécule cristallisée. Cependant, dans le cas des macromolécules, les phases sont souvent entachées d'erreurs importantes, ce qui empêche l'interprétation fine de la structure obtenue avec une première estimation des phases. Pour cette raison, on adopte un processus appelé "affinement", qui permet de réduire progressivement l'erreur de la phase, en jouant sur les

3. L'atome d'hydrogène est une exception.

paramètres du modèle d'atomes utilisé pour représenter la structure. Ainsi, pour pouvoir faire l'affinement, il faut convenir d'une paramétrisation de la densité électronique. Dans la prochaine section, deux modèles de distribution de la densité électronique seront présentés.

2.4 Modèles de distribution de la densité électronique atomique

Pour calculer les facteurs de structure d'un modèle moléculaire, on utilise un modèle de la densité électronique qui se base sur un certain nombre de paramètres. Deux modèles largement employés en cristallographie sont décrits ci-dessous.

2.4.1 Le modèle de l'atome indépendant (IAM)

Le modèle le plus utilisé dans la cristallographie macromoléculaire est le modèle de l'atome indépendant (**I**ndependent **A**tom **M**odel (IAM)). Dans ce formalisme, les densités atomiques sont supposées de symétrie sphérique, indéformables, neutres et indépendantes. Les interactions entre les atomes, dues aux liaisons chimiques (covalentes, liaisons hydrogènes), sont ainsi négligées. Le modèle est donné par [10] :

$$F_{calc}(\vec{H}) = K \cdot \sum_{j=1}^{N_{at}} q_j f_j(\vec{H}) \cdot exp(-B_j \frac{sin^2\theta}{\lambda^2}) exp(2i\pi \vec{H} \cdot \vec{r}_j) \qquad (2.6)$$

q_j Facteur d'occupation d'un atome

B_j Facteur d'agitation thermique isotrope équivalent, $B_{j,equ} = 8\pi^2 U_{eq}$

K Facteur d'échelle

$\vec{r}_j$ Coordonnées d'un atome

(Pour la signification des autres paramètres, voir les formules 2.1-2.3).

Ce modèle contient les paramètres suivants :

Facteur d'échelle C'est un paramètre global. Il est nécessaire parce que les intensités expérimentales, et celles calculées à partir d'un modèle ne sont pas à la même échelle.

Coordonnées (x,y,z) Ils déterminent la position d'un atome dans la maille.

Occupation Ce facteur reflète le désordre d'un atome ou d'un groupement. En effet, il est possible que la position ou l'orientation d'un groupement

varie d'une maille à une autre et qu'il dispose ainsi d'une conformation multiple. Le facteur d'occupation donne la probabilité de trouver un atome à une certaine position. C'est un chiffre compris entre 0 (l'atome n'est pas présent) et 1 (l'atome a la même position dans toutes les mailles).

Facteur d'agitation thermique Les atomes d'un cristal ne sont pas immobiles. En effet, ayant de l'énergie thermique, ils vibrent autour d'une position d'équilibre. Or, l'effet de l'agitation thermique est visible sur le cliché de diffraction, on observe que l'intensité des rayons diffractés diminue. Le facteur de structure est alors modulé par une fonction qui tient compte de la diminution de l'intensité diffractée à cause du mouvement thermique. Dans le modèle harmonique, il s'agit d'une fonction gaussienne exponentielle décroissante , appelé facteur de Debye-Waller : $exp(-B_{eq}\frac{sin^2\theta}{\lambda^2})$, où $B_{eq} = 8\pi^2 U_{eq}$ est le facteur d'agitation thermique isotrope équivalent.
L'approximation la plus simple du mouvement thermique est l'agitation isotrope, la magnitude du mouvement atomique est alors égale dans toutes les directions. Un seul paramètre par atome suffit pour décrire le phénomène et $U_{eq} = \langle\vec{u}^2\rangle$, $\langle\vec{u}^2\rangle$ étant l'écart quadratique moyen par rapport à la position moyenne.
Cependant, il est plus réaliste qu'un atome ait des directions préférentielles de déplacement et que le mouvement soit anisotrope. La magnitude de l'agitation n'est donc pas égale dans toutes les directions. On décrit alors l'agitation thermique grâce à un tenseur symétrique d'ordre 2 et dans ce cas $U_{eq} = \frac{1}{3}\sum_i\sum_j(U_{ij}\mathbf{a}^j\mathbf{a}^i a_j a_i)$ (les a^i sont les vecteurs de base du réseau direct, les a_i représentent la norme des vecteurs de base du réseau réciproque et les U_{ij} les éléments du tenseur) [11].
Notons que B_j ne représente pas uniquement les vibrations thermiques mais également tous les types de désordre, comme par exemple les conformations multiples, qui sont indépendantes de la température. Le mouvement thermique d'un atome peut être représenté graphiquement par des surfaces de probabilité constante. Dans le cas de l'agitation isotrope, il s'agit alors d'une sphère tandis qu'un ellipsoïde représente le mouvement anisotrope.

Si l'agitation thermique est isotrope, on obtient 1 (occupation) + 3 (coordonnées) + 1 (agitation thermique anisotrope) = 5 paramètres par atome. Dans le cas de l'agitation anisotrope, il y a 10 paramètres.
Le modèle de l'atome indépendant constitue une approximation suffisante pour la densité électronique si les données de diffraction sont de faible ou moyenne résolution. Par contre, à résolution subatomique (à partir de 0.9 Å

environ [12]), le modèle IAM est trop simplifié pour décrire la densité électronique. En effet, à de telles résolutions, des pics de densité apparaissent qui correspondent aux électrons partagés des liaisons covalentes (figure 2.3) ou à la densité des paires libres d'électrons. Pour en tenir compte, une description plus sophistiquée de la densité électronique est nécessaire.

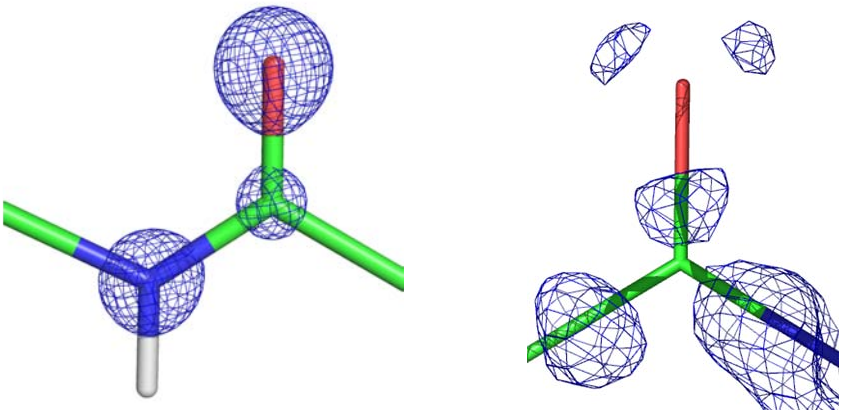


Figure 2.3 – Gauche : Densité électronique sphérique dans l'unité peptidique. Droite : Densité résiduelle après l'affinement utilisant un modèle sphérique (dans la protéine Eel Pout Type III Antifreeze, 0.62 Å de résolution [13]).

2.4.2 Le modèle de Hansen et Coppens

Un modèle qui s'est imposé en cristallographie de petites mailles et dans le domaine de l'analyse de la densité de charge est le modèle multipolaire de Hansen et Coppens [14]. Dans ce modèle, la densité électronique est subdivisée en trois parties : La densité de cœur, la densité de valence et la densité de déformation non sphérique.

$$\rho_{atom}(\mathbf{r}) = \underbrace{\rho_{coeur}(r)}_{coeur} + \underbrace{\underbrace{P_{val}\kappa^3\rho_{val}(\kappa r)}_{\text{valence sphérique}} + \underbrace{\sum_{l=0}^{l_{max}} \kappa'^3 R_{nl}(\kappa' r) \sum_{m=0}^{l} P_{lm\pm} d_{lm\pm}(\theta,\phi)}_{\text{non sphérique}}}_{\text{densité de déformation}} \tag{2.7}$$

ρ_{coeur} densité électronique de cœur

ρ_{val} densité électronique de valence

P_{val}, P_{lm} coéfficient de population

κ, κ' paramètres de contraction-expansion

$d_{lm}(\theta, \phi)$ fonctions harmoniques sphériques réelles (la normalisation est telle que $\int d_{lmp}^2 d\Omega = 2$ pour l>0 et $\int d_{lmp}^2 d\Omega = 1$ pour l=0, $d\Omega$ est un élément de volume dans l'espace $\theta - \phi$) [4]

$R_{nl}(\kappa' r)$ fonction radiale de type Slater

Ce modèle contient plusieurs modifications par rapport au modèle IAM : D'abord, la diffraction des électrons de cœur et de valence est modélisée séparément. On assume que les électrons de cœur ne sont pas perturbés par les interactions interatomiques mais que seule la couche de valence est modifiable. Dans le modèle de Hansen et Coppens, la population et la dépendance radiale de la couche de valence peuvent être ajustées en modélisant les paramètres P_{val} et κ. Par conséquent, un transfert de charge entre des atomes est permis et le paramètre P_{val} donne une estimation de la charge q d'un atome : $q = N_{val} - P_{val}$ (avec N_{val} le nombre d'électrons de valence d'un atome neutre). Le paramètre κ sert à décrire la contraction ou l'expansion de la couche de valence.
Ensuite, le troisième terme de l'équation 2.7 décrit l'aspect non sphérique de la densité électronique. La fonction multipolaire est un produit d'une fonction radiale R_l (dépendant de la coordonnée polaire r) et d'une fonction angulaire d_{lm} (dépendant de θ et ϕ). L'utilisation des fonctions multipolaires permet de modéliser la densité électronique accumulée sur les liaisons covalentes ou celle des paires libres par exemple (figure 2.4). Les paramètres P_{lm} représentent les populations multipolaires, notons qu'ils peuvent être égaux à zéro, l'atome d'hydrogène par exemple n'est en général modélisé qu'avec les dipôles tandis que les autres paramètres multipolaires sont nuls. L'expansion multipolaire est tronquée au niveau hexadécapolaire pour le soufre, au niveau octopolaire pour les atomes O, N et C, et au niveau dipolaire pour l'atome d'hydrogène. La fonction radiale est donnée par une fonction de type Slater :

$$R_{nl} = \frac{(\kappa'\zeta)^{n(l)+3}}{(n(l)+2)!} r^{n(l)} e^{-\kappa'\zeta r} \tag{2.8}$$

où ζ et $n(l)$ sont des paramètres dépendants de l'espèce chimique et de la structure électronique de l'atome.
Notons que le nombre de paramètres de ce modèle est considérablement plus élevé par rapport au modèle IAM. On peut avoir jusqu'à 26 paramètres multipolaires, auxquels s'ajoutent les paramètres classiques (coordonnées,

4. Dans quelques publications, la notation y_{ml} est également utilisée pour d_{ml}. Par contre, dans [15], les fonctions y_{lm} n'ont pas la même normalisation. Pour ne pas confondre ces fonctions, la notation de [15] sera utilisée dans ce rapport.

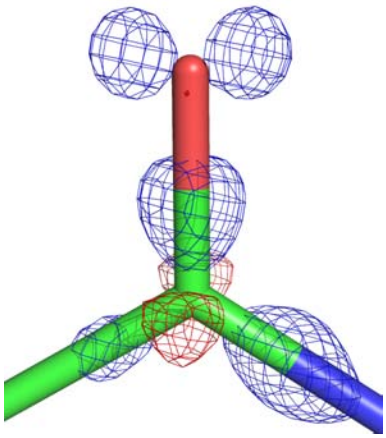


Figure 2.4 – Densité de déformation d'une unité peptidique, calculé à partir des valeurs de la banque de données ELMAM et contourée à 0.3 σ. Bleu : densité positive, rouge : densité négative. Notons la similitude avec la densité résiduelle fig. 2.3 (droite).

agitation thermique, occupation). Au maximum, un atome aura alors 36 paramètres dans ce formalisme.

2.5 L'affinement cristallographique

L'affinement cristallographique consiste à optimiser un modèle (structural, ou de densité électronique) par un procédé mathématique en tenant compte des données de diffraction [16]. Plus précisément, on ajuste les paramètres d'un modèle moléculaire pour qu'il coïncide le plus possible avec les données expérimentales. L'écart entre les modules des facteurs de structure (ou les intensités) expérimentaux et ceux issus du modèle est ainsi minimisé et le modèle ressemble le plus possible à la molécule étudiée (figure 2.5).
Il y a plusieurs méthodes d'affinement, à citer sont notamment la méthode des moindres carrés et le maximum de vraisemblance (aussi appelé "Maximum Likelihood") qui sont utilisées dans les programmes d'affinement macromoléculaire les plus répandus. (**Moindre carrés** : MoPro, Shelxl, Phenix ; **Maximum Likelihood** : Refmac, Phenix). Dans le cas des moindres carrés, la fonction résiduelle à minimiser est donnée par :

$$S = \sum_{h,k,l} w_{hkl}\left(\left|F_{obs}(\vec{H})\right| - \left|F_{calc}(\vec{H},p)\right|\right)^2 \tag{2.9}$$

w_{hkl} Facteur de pondération, affecté à chaque réflexion. w_{hkl} est normalement égal à $1/\sigma(I)^2$ mais il existe d'autres schémas de pondération.

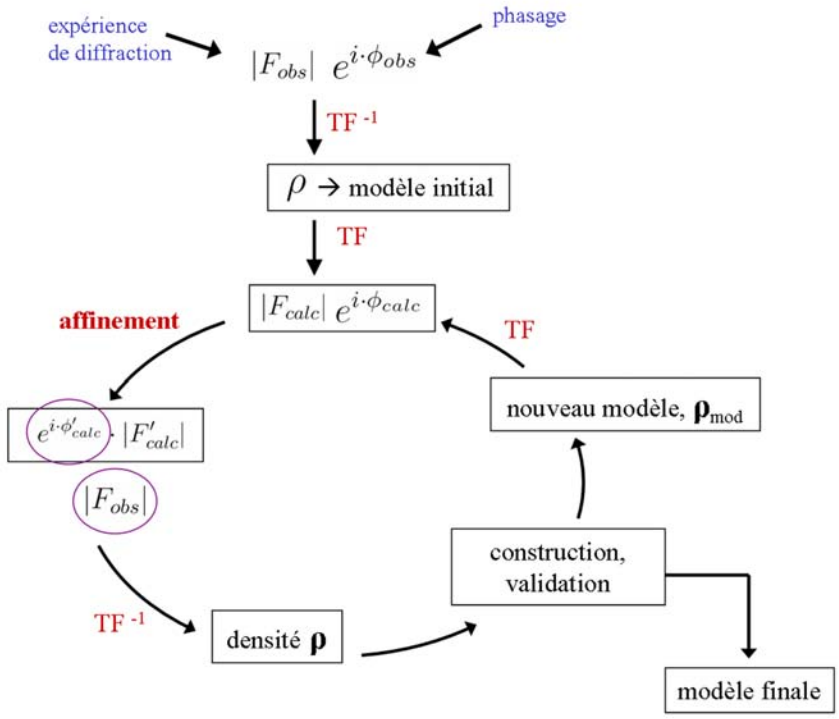


Figure 2.5 – Schéma illustrant le processus d'affinement. Une première estimation de la densité électronique est obtenue avec les modules expérimentaux du facteur de structure et les phases obtenues soit expérimentalement soit par la méthode directe ou le remplacement moléculare (cf section 2.3), ce qui permet la construction d'un premier modèle structural. Ensuite, le modèle est affiné et ses paramètres sont donc changés. Avec les nouvelles phases du modèle ϕ_{calc} et les facteurs de structure observés F_{obs}, on peut calculer une densité électronique. Elle est visualisée pour effectuer des modifications manuelles ou pour valider l'affinement. On passe de nouveau à l'étape d'affinement, et ainsi de suite. Le processus est répété jusqu'à convergence. (TF = transformée de Fourier).

$|F_{obs}|$ Module du facteur de structure observé

$|F_{calc}|$ Module du facteur de structure calculé grâce au modèle

p paramètres du modèle de densité électronique

Des facteurs d'accord statistiques permettent d'estimer la qualité du modèle par rapport aux données de diffraction :

$$R(F) = \frac{\sum_{H_{travail}} (\frac{1}{K} \cdot \left|F_{obs}(\vec{H})\right| - \left|F_{calc}(\vec{H}, p)\right|)}{\sum_{H_{travail}} \frac{1}{K} \cdot \left|F_{obs}(\vec{H})\right|} \tag{2.10}$$

$H_{travail}$ Sous-ensemble de vecteurs du réseau réciproque, normalement 90 à 95% des réflections mésurées

K Facteur d'échelle.

Le facteur $R(F)$ diminue si l'ensemble des facteurs de structure calculés est proche des facteurs de structure observés. Par contre il est possible qu'il diminue artificiellement, par exemple en introduisant un grand nombre de paramètres. Pour contourner cette ambiguïté, le facteur $R_{free}(F)$ a été introduit [17]. Il est calculé de la même manière que le facteur $R(F)$, mais sur un sous-ensemble (normalement 5%) des réflections qui n'ont pas été utilisées pour l'affinement. Ainsi, ce facteur n'est pas atteint par le biais introduit par l'affinement. Un modèle est considéré comme satisfaisant, si les deux facteurs ($R(F)$ et $R_{free}(F)$) ont diminué parallèlement au cours de l'affinement jusqu'à convergence et si leurs valeurs finales ne sont pas trop élevées.
Le nombre de paramètres à affiner dans le cas des protéines est très grand. Pour cette raison, des informations supplémentaires ou des approximations sont appliquées :

Les atomes d'hydrogène ne sont pas affinés L'atome d'hydrogène n'a qu'un seul électron, sa contribution à l'amplitude de l'onde diffractée est donc faible par rapport aux autres espèces chimiques. En plus, le pic de densité correspondant n'est discernable qu'à partir de données de haute résolution, puisqu'il est en général situé à moins d'un Å d'un atome plus lourd. Par conséquent, la présence des atomes d'hydrogène est négligée ou les atomes H à position fixe (comme dans celui dans l'unité peptidique) sont placés. Dans ce cas, leurs paramètres ne seront pas affinés mais ils seront contraints à des valeurs précises.

L'application de restraintes stéréochimiques Une protéine peut être considérée comme un objet composé d'atomes ou bien comme un ensemble de longueurs de liaison interatomiques, d'angles valenciels et d'angles dièdres. Toutes ces valeurs sont connues à partir de structures

à haute résolution de petites polypeptides. Or, on s'attend à ce que ces données ne diffèrent pas considérablement par rapport à celles dans une protéine. On restreint alors les longueurs de liaison et les angles en imposant une pénalité pour des déviations significatives par rapport à une valeur cible.

2.6 La banque de données ELMAM

Les protéines diffractent rarement à résolution subatomique où on peut discerner la densité de déformation et affiner les paramètres multipolaires. Dans les cas rares où un affinement multipolaire a été effectué sur une structure de protéine (par exemple [3], [18]), l'analyse était restreinte sur les régions de faible agitation thermique.
Il est pourtant intéressant d'avoir accès à la densité électronique précise de toute la protéine. Elle permettra l'analyse topologique et le calcul de nombreuses propriétés, comme par exemple le potentiel ou les moments électrostatiques. Un modèle sphérique et neutre de la densité électronique ne pourra pas fournir toutes ces informations, et ces propriétés semblent alors hors de portée de l'expérimentateur.
Or, le principe de transférabilité postule que la densité de déformation dans un groupement chimique est similaire dans différentes molécules et contextes cristallins [19]. Notons que la polarisation des atomes due à l'environnement chimique, comme la liaison hydrogène, est négligée dans cette approximation.
Au laboratoire, les paramètres multipolaires de peptides et d'acides aminés ont été déterminés par analyse de la densité électronique expérimentale et une banque de données (ELMAM - *Experimental Library Multipolar Atom Model*) a été créée. ELMAM contient les paramètres multipolaires de tous les groupements chimiques présents dans les protéines et les polypeptides. Ces paramètres peuvent ainsi être transférés à une protéine quelconque, ce qui permet le calcul de propriétés qui auraient été inaccessibles dans le cas de structures à résolution moyenne.
La librairie ELMAM a été utilisée pour toutes les études décrites dans ce mémoire. Le transfert des paramètres multipolaires permettait le calcul de moments électrostatiques des motifs de structure secondaire (partie II) et l'analyse des liaisons hydrogène dans les hélices de protéine. De plus, le potentiel électrostatique de la protéine PfluDING a été calculé (chapitre 8) ce qui a donné de nouveaux éléments pour expliquer le mode de fixation d'un ion phosphate dans le site actif.

2.7 Topologie de la densité électronique

L'analyse topologique d'une distribution de densité électronique permet de caractériser les interactions interatomiques et intermoléculaires. On s'intéresse par exemple à localiser les maxima, minima et points selle de la densité électronique. La méthode consiste à analyser le champ vectoriel de gradients de la densité électronique totale afin de trouver les points critiques $\vec{r}_c$ où le gradient s'annule [20], [21].

$$\nabla\rho(\vec{r}_c) = \frac{\partial\rho}{\partial x}\vec{i} + \frac{\partial\rho}{\partial y}\vec{j} + \frac{\partial\rho}{\partial z}\vec{k} = 0 \tag{2.11}$$

Pour déterminer la nature du point critique (donc pour savoir s'il s'agit d'un maximum, d'un minimum ou d'un point selle), la courbure de la densité électronique est analysée. La matrice des dérivées secondes $\frac{\partial^2\rho}{\partial x_i \partial x_j}$ ($i \leq j$ = 1,2,3) de la densité électronique est la matrice Hessienne. Une diagonalisation de cette matrice mène à ses vecteurs et valeurs propres qui représentent les axes de courbure principaux et les courbures principales (notés λ_1, λ_2 et λ_3), respectivement. Les valeurs propres λ_1 et λ_2 sont les courbures transversales et λ_3 est la courbure longitudinale (le long du chemin d'interaction).

$$\lambda_1 = \frac{\partial^2\rho}{\partial x'^2}, \qquad \lambda_2 = \frac{\partial^2\rho}{\partial y'^2}, \qquad \lambda_3 = \frac{\partial^2\rho}{\partial z'^2} \tag{2.12}$$

Les valeurs propres caractérisent les points critiques par un couple de valeurs (ω,σ), ω étant le nombre de valeurs propres non nulles (le rang de la matrice hessienne) et σ la somme de leurs signes (la signature). Dans des structures stables, les points critiques mènent à 3 valeurs propres non nulles, ω est donc égal à 3 et σ égale à ±1 ou ±3. On distingue alors quatre types de points critiques :

(3,-3) Point critique atomique (trois courbures négatives). Ce point critique correspond au maximum local de la densité électronique qui se trouve à la position d'un atome.

(3,-1) Point critique de liaison (deux courbures négatives et une positive). C'est un point selle qui est localisé le long d'une liaison (par exemple liaison covalente ou liaison hydrogène).

(3,+1) Point critique de cycle (deux courbures positives et une négative). Il se trouve dans une vallée de densité au sein d'au moins 3 atomes liés.

(3,+3) Point critique de cage (3 courbures positives). C'est un point critique au sein d'une cage d'au moins 4 atomes.

La fonction $\nabla\rho(\vec{r})$ définit un champ de vecteurs qui sont dirigés en chaque point le long du gradient de la densité de charge. Les vecteurs gradient démarrent aux points critiques de courbure positive et terminent aux maxima de la distribution de densité. La région qui contient toutes les lignes de gradient terminant sur un maximum définit un bassin. L'espace est ainsi réparti en bassins atomiques, qui sont définis par la surface de flux nul du gradient (figure 2.6) :

$$\nabla\rho(\vec{r}) \cdot \vec{n}(\vec{r}) = 0 \tag{2.13}$$

Notons que le bassin peut s'étendre jusqu'à l'infini pour certains atomes et

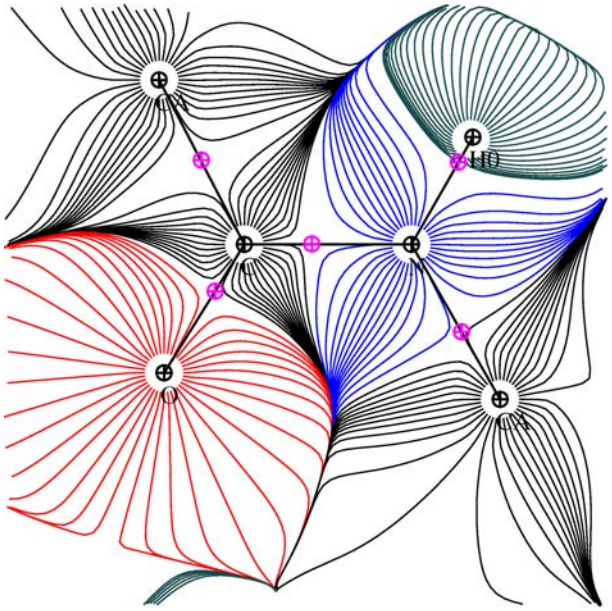


Figure 2.6 – Lignes des gradients résultant de la distribution de charge dans l'unité peptidique ainsi que les points critiques de liaison (croix encerclées en magenta) (calculées avec les paramètres multipolaires de la banque ELMAM).

que cette surface n'est pas franchie par une ligne de gradient, d'où son nom surface de zéro flux (*zero flux surface*).
La trace de la matrice Hessienne correspond au Laplacien qui permet notamment de caractériser la nature d'un point critique (3,-1) :

$$\nabla^2\rho(\vec{r}) = \frac{\partial^2\rho}{\partial x^2} + \frac{\partial^2\rho}{\partial y^2} + \frac{\partial^2\rho}{\partial z^2} \tag{2.14}$$

Le laplacien de la densité électronique est également relié à la densité d'énergie cinétique et à la densité d'énergie potentielle par le théorème du Viriel :

$$\frac{h^2}{(4\pi)^2 m}\nabla^2\rho(\vec{r}) = 2G(\vec{r}) + V(\vec{r}) \tag{2.15}$$

G densité d'énergie cinétique, G≥0

V densité d'énergie potentielle, V≤0

G et V satisfont $G(\vec{r}) + V(\vec{r}) = H(\vec{r})$, où H est la densité d'énergie totale. Un Laplacien négatif implique que la densité soit localement concentrée et un laplacien positif que la densité est localement diluée [5]. Aux points critiques de liaison (3,-1), l'interaction a un caractère covalent de couches partagées si le Laplacien en $\vec{r}_c$ est négatif et si la densité d'énergie potentielle domine l'énergie totale. L'interaction a un caractère ionique à couche fermée si le Laplacien $\nabla^2\rho(\vec{r}_c)$ est positif, c'est alors la densité d'énergie cinétique qui est dominante. Notons que la liaison hydrogène correspond à une liaison à caractère ionique.
Quelques concepts seront repris dans la partie III qui traite de l'analyse topologique des liaisons hydrogènes dans les motifs de structure secondaire.

2.8 Calcul du potentiel électrostatique

Le potentiel électrostatique $V(\vec{r})$ est défini par l'énergie coulombienne entre une densité de charge $\rho(\vec{r})$ et un proton situé au point $\vec{r}$:

$$V(\vec{r}) = \int \frac{\rho(\vec{r'})}{\left|\vec{r} - \vec{r'}\right|} d^3\vec{r'} \tag{2.16}$$

Le potentiel peut être calculé à partir de données RX précises dans l'espace réciproque [22] ($V(\vec{r})$ étant proportionel à la transformée de Fourier de $\frac{F(H)}{H^2}$), ou en utilisant un modèle non sphérique de la densité électronique dans l'espace direct [23]. A partir du modèle multipolaire de Hansen et Coppens, le potentiel s'exprime sous la forme :

$$V(\vec{r}) = V_c(\vec{r}) + V_{val}(\vec{r}) + \Delta V(\vec{r}) \tag{2.17}$$

$V_c(\vec{r})$ correspond au potentiel produit par la charge positive Z du noyau et la densité sphérique de cœur

$V_{val}(\vec{r})$ est le potentiel dû à la densité sphérique de valence ρ_{val}

$\Delta V(\vec{r})$ correspond au potentiel créé par la densité de charge de déformation

(Les calculs de l'intégration sont décrits en détails dans [24])
Notons qu'il est d'usage de neutraliser l'ensemble d'un fragment avant de calculer son potentiel. Sinon, les effets de la densité non-sphérique peuvent être masqués par la charge totale.

5. C'est valable pour les atomes du premier rang du tableau périodique.

Deuxième partie

Moments électrostatiques des motifs de structure secondaire

Introduction

La distribution de charge de l'unité peptidique est dipolaire et, dans l'hélice α, les dipôles peptidiques sont arrangés parallèlement à l'axe de l'hélice. Par conséquent, cet arrangement régulier des dipôles crée un macrodipôle. De même, on peut attribuer des moments électrostatiques au feuillet β. Or, les moments électrostatiques apparaissent dans les expressions qui quantifient les interactions intermoléculaires. Les moments électrostatiques générés par l'hélice α et le feuillet β sont donc impliqués dans de nombreuses interactions.
Notamment, il a été proposé que le macrodipôle de l'hélice α a les effets suivants [25], [26] :

- le champ du dipôle aide à fixer des substrats ou coenzymes chargés
- le dipôle attire des substrats chargés via une interaction de longue portée et les oriente proprement dans le site actif
- le moment créé par l'hélice contribue à l'accélération de réactions catalysées par des enzymes
- les interactions entre les moments stabilisent le repliement d'une protéine

La connaissance des moments électriques est alors particulièrement intéressante pour pouvoir estimer l'ordre de grandeur de ces interactions. Il y a plusieurs méthodes expérimentales qui permettent de déterminer les moments [27], comme par exemple l'effet Stark ou la permittivité électrique pour la mesure du dipôle ou l'effet Zeeman et la birefringence pour déterminer le moment quadrupolaire. Or, ces méthodes ne donnent que le moment de la molécule entière tandis qu'on ne s'intéresse ici dans un premier temps qu'aux moments générés par les motifs de structures secondaires. Un autre désavantage est que certaines méthodes (comme par exemple la permittivité électrique) fournissent seulement le module du dipôle, sa direction reste alors inconnue. Or, il est possible de calculer des moments électriques à partir de la connaissance précise de la distribution de charge, donc à partir de données de diffraction. Cette méthode donne accès au module et aux composantes vectorielles ou éléments tensoriels des moments électrostatiques et elle per-

met de choisir un fragment sur lequel le moment peut être calculé.
Ce chapitre tend à caractériser et à quantifier les moments électrostatiques des motifs de structure secondaire (hélice α et feuillet β) en utilisant la bibliothèque de paramètres multipolaires ELMAM. Dans un premier temps, nous détaillerons comment calculer les moments électriques à partir de la densité électronique, puis nous décrirons nos résultats.

Chapitre 3

Méthodes de calcul des moments

3.1 Calcul d'un moment électrostatique à partir d'une distribution de charge

Dans cette section, il sera éclairci comment on peut calculer un moment électrostatique à partir d'une distribution de charge qui est exprimée dans le formalisme de Hansen et Coppens [14]. Les expressions obtenues ont été utilisées pour implémenter de nouvelles sous-routines dans VMoPro qui permettent le calcul du moment dipolaire et quadrupolaire d'une molécule ou d'un fragment.
La nomenclature suit celle utilisée par P. Coppens dans le livre "X-ray charge densities and chemical bonding" [15].

3.1.1 Définition des moments électrostatiques

Les moments électrostatiques cartésiens d'une distribution de charge $\rho(\vec{r})$ sont donnés par [15]

$$\mu_{\alpha_1,\alpha_2,\alpha_3...\alpha_l} = \int_{V_T} \rho(\vec{r}) r_{\alpha_1} r_{\alpha_2} r_{\alpha_3} ... r_{\alpha_l} d\vec{r} \tag{3.1}$$

Les termes r_α sont les 3 composantes du vecteur position $\vec{r}$ ($\alpha_i = 1, 2, 3$), l'intégrale s'étend sur le volume V_T de la distribution de charge $\rho(\vec{r})$. Pour l

de 0 à 2, on obtient :

$$l = 0 \quad \text{monopôle} \qquad q = \int_{V_T} \rho(\vec{r}) d\vec{r} \tag{3.2}$$

$$l = 1 \quad \text{dipôle} \qquad \mu_{\alpha_i} = \int_{V_T} \rho(\vec{r}) r_{\alpha_i} d\vec{r} \tag{3.3}$$

$$\text{par exemple } \mu_x = \int_{V_T} \rho(\vec{r}) x d\vec{r}$$

$$l = 2 \quad \text{quadrupôle} \qquad \theta_{\alpha_i \alpha_j} = \int_{V_T} \rho(\vec{r}) r_{\alpha_i} r_{\alpha_j} d\vec{r} \tag{3.4}$$

Le monopole représente la charge totale de la distribution $\rho(\vec{r})$ dans le volume d'intégration V_T tandis que les moments d'ordre supérieur mesurent la séparation de la charge. Un dipôle simple par exemple est formé par deux charges q de signes opposés et séparées d'un vecteur $\vec{d}$. Le moment dipolaire est alors donné par $q \cdot \vec{d}$, il augmente si la distance entre les deux charges devient plus longue. Pour des distributions de charge plus compliquées, il s'en suit que des molécules longues vont avoir un grand moment dipolaire, surtout si elles portent des groupements de charges opposées aux extrémités. De même, des moments quadrupolaires importants vont caractériser une distribution de charge ayant des groupements de charge opposée et bien séparés.

Notons que le choix du volume d'intégration V_T est délicat. Dans une molécule, les densités de charge des différents pseudo-atomes, liés de manière covalente, se chevauchent. On peut établir un critère qui permet d'assigner à chaque atome un volume précis, sur lequel l'intégration va être effectuée. Il y a différentes méthodes de partition, comme par exemple celle de Bader [20], qui est basée sur le volume délimité par la surface de flux nul du gradient de la densité électronique (section 2.7).
Pour certaines applications, il est avantageux d'utiliser l'expression des moments électrostatiques issue du formalisme des tenseurs sphériques. Ils sont définis par

$$\Theta_{lmp} = \int_{V_T} \rho(\vec{r}) c_{lmp}(\vec{r}) r^l d\vec{r} \tag{3.5}$$

Les fonctions $c_{lmp}(\vec{r})$ sont des fonctions harmoniques sphériques où la dépendance angulaire est exprimée par les cosinus directeurs [15] et p est égal à ± 1.
Les moments définis par l'équation (3.1) sont les moments électrostatiques dits non abrégés. Il existe une définition alternative pour des moments d'ordre supérieur à 1 où les moments sont définis de sorte que la trace du tenseur

soit égale à 0 (pour le moment quadrupolaire : $\Theta_{xx} + \Theta_{yy} + \Theta_{zz} = 0$). Le moment quadrupolaire sans trace s'écrit [27] [15] [1] :

$$\Theta_{\alpha\beta} = \frac{1}{2}\int_{V_T} \rho(\vec{r})\Big[3r_\alpha r_\beta - r^2\delta_{\alpha\beta}\Big]d\vec{r} \tag{3.6}$$

$\delta_{\alpha\beta}$ est le symbole de Kronecker : $\delta_{\alpha\beta} = \begin{cases} 1 & \text{si } \alpha = \beta \\ 0 & \text{si } \alpha \neq \beta \end{cases}$.
On peut exprimer les éléments tensoriels des moments cartésiens sans trace en fonction de ceux des moments sphériques sans trace (équation 3.5) [15].

$$\Theta_{xx} = \tfrac{1}{2}(3\Theta_{22+} - \tfrac{1}{2}\Theta_{20}) \tag{3.7}$$

$$\Theta_{yy} = \tfrac{1}{2}(-3\Theta_{22+} - \tfrac{1}{2}\Theta_{20}) \tag{3.8}$$

$$\Theta_{zz} = \tfrac{1}{2}\Theta_{20} \tag{3.9}$$

$$\Theta_{xy} = \tfrac{3}{2}\Theta_{22-} \tag{3.10}$$

$$\Theta_{xz} = \tfrac{3}{2}\Theta_{21+} \tag{3.11}$$

$$\Theta_{yz} = \tfrac{3}{2}\Theta_{21-} \tag{3.12}$$

3.1.2 Dépendance de l'origine

Les moments multipolaires, à l'exception du moment d'ordre zéro, dépendent du choix de l'origine. Une translation d'origine O vers O' par le vecteur $\vec{R} = (X, Y, Z)$ (figure 3.1) implique que les vecteurs $\vec{r_{\alpha_i}}$ dans l'expression 3.1 du moment dipolaire cartésien doivent être substitués par $r'_{\alpha_i} = r_{\alpha_i} - R_{\alpha_i}$.

Ainsi, par exemple, pour le moment dipolaire non abrégé (équation 3.3), il en suit

$$\mu'_x = \int \rho(\vec{r})(x - X)d\vec{r} = \int \rho(\vec{r})xd\vec{r} - \int \rho(\vec{r})Xd\vec{r} = \mu_x - Xq \tag{3.13}$$

$$\mu'_y = \mu_y - Yq \tag{3.14}$$

$$\mu'_z = \mu_z - Zq \tag{3.15}$$

Ces expressions montrent que le moment dipolaire est indépendant de l'origine pour une distribution de charge neutre, où q=0. Notons que les fragments utilisés dans cette étude ont été neutralisés afin de garantir que leurs

1. Remarque : Il existe une expression générale pour les éléments tensoriels, voir [15].

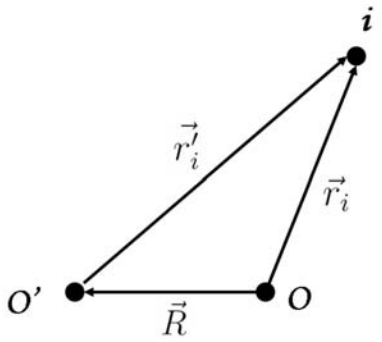


Figure 3.1 – Translation d'origine de O vers O' par un vecteur $\vec{R}$.

moments dipolaires soient indépendants de l'origine. Cette démarche a été également appliquée dans la littérature, par exemple dans [25] et [28].
Pour le moment quadrupolaire sans trace (equation 3.6), on obtient après un changement d'origine (le calcul est détaillé en annexe B.1.1) :

$$\begin{aligned}\Theta'_{\alpha\beta} &= \frac{1}{2}\int \rho(\vec{r})\Big[3r'_\alpha r'_\beta - r'^2\delta_{\alpha\beta}\Big]d\vec{r} \\ &= \frac{1}{2}\int \rho(\vec{r})\Big[3(r_\alpha - R_\alpha)(r_\beta - R_\beta) - (\vec{r}-\vec{R})^2\delta_{\alpha\beta}\Big]d\vec{r} \\ &= \Theta_{\alpha\beta} + \frac{1}{2}\Big(3R_\alpha R_\beta - R^2\delta_{\alpha\beta}\Big)q - \frac{3}{2}\Big(R_\beta\mu_\alpha + R_\alpha\mu_\beta\Big) + \sum_{i=1}^{3} R_i\mu_i\delta_{\alpha\beta}\end{aligned} \tag{3.16}$$

Comme pour le dipôle, les éléments tensoriels du moment quadrupolaire sont indépendant de l'origine, si tous les moments d'ordre inférieur (la charge q et les dipôles μ_i) sont égaux à zéro. Dans ce cas, les trois derniers termes de l'équation 3.16 s'annulent. Cette équation permet de calculer les moments quadrupolaires par rapport à une origine de choix. En particulier, il est possible grâce à ces expressions de calculer la somme des moments quadrupolaires de différents atomes par rapport à une origine commune. Une possibilité serait de choisir l'origine au centre de masse du fragment considéré, comme proposé par Spackman [27].

3.1.3 Moments électrostatiques atomiques

Le moment électrostatique d'un atome est obtenu par intégration sur sa distribution de charge totale. La distribution de charge d'un atome correspond à la somme de la distribution nucléaire et électronique : $\rho_{total} = \rho_{\text{nucléaire}} + \rho_{\text{électronique}}$. Or, si l'on l'on choisit l'origine à la position atomique,

le moment nucléaire s'annule [2], et seule la distribution électronique contribue à l'intégrale. On peut alors substituer $\rho(\vec{r})$ dans la formule (3.1) par l'expression de la densité électronique du modèle de Hansen et Coppens :

$$\begin{aligned}\mu_{\alpha_1,\alpha_2,\alpha_3...\alpha_j} &= -\int_{V_T}\Big[\rho_{core} + P_{val}\kappa^3\rho_{valence}(\kappa r) + \\ &\quad + \sum_{l=0}^{l_{max}} \kappa'^3 R_{nl}(\kappa' r)\sum_{m=0}^{l} P_{lmp} d_{lmp}(\theta\phi)\Big] r_{\alpha_1} r_{\alpha_2}...r_{\alpha_j} d\vec{r}\end{aligned} \tag{3.17}$$

Il est utile de passer au formalisme des harmoniques sphériques et d'utiliser les moments sans trace pour faire l'intégration. En effet, pour les moments sans trace, seulement les multipôles atomiques d'ordre l vont contribuer au moment d'ordre l [15]. Par exemple, ce ne sont que les termes quadrupolaires de l'expansion multipolaire de la distribution de charge qui vont contribuer au moment quadrupolaire. Ainsi, on peut simplifier l'expression 3.17 :

$$\begin{aligned}\Theta_{lmp} &= -\int_{V_T} r^l c_{lmp}\Big[\sum_{l'=0}^{l'_{max}} \kappa'^3 R_{l'} \sum_{m'=0}^{l'} P_{l'm'p} d_{l'm'p}\Big] d\vec{r} \\ &= -P_{lmp}\frac{L_{lm}}{M_{lm}^2}\frac{(n(l)+l+2)!}{(n(l)+2)!}\frac{1}{(\kappa'\zeta)^l}\end{aligned} \tag{3.18}$$

(Le calcul est détaillé section B.1.2.)
Cette expression contient les valeurs de P_{lmp} qui correspondent aux populations multipolaires du formalisme de Hansen et Coppens, obtenues par affinement ou par transfert. De même, κ' est le paramètre de contraction-expansion (équation 2.7). M_{lm} et L_{lm} sont des facteurs de conversion entre différentes fonctions harmoniques sphériques. ζ et $n(l)$ sont des paramètres de la fonction radiale $R_l(r)$. Ainsi, le moment électrique sans trace de l'atome peut être calculé en coordonnées sphériques à partir de la connaissance des P_{lmp}. Notons qu'il s'agit du moment d'un atome seul, qui est propre à son système d'axes tel que défini dans le formalisme de Hansen et Coppens.

3.1.4 Moments électrostatiques totaux

Afin d'obtenir le moment qui correspond à une molécule (constituée de plusieurs atomes), il faut additionner les moments des atomes individuels **et** les contributions dues au changement d'origine (équation 3.16).
En résumé, le calcul du moment quadrupolaire, tel qu'il a été implémenté dans VMoPro, contient les étapes suivantes [3] :

2. Les r_{α_i} des protons s'annulent dans ce cas.
3. La succession des étapes est en fait valable pour tous les moments $l \geq 2$.

choix de l'origine le centre de masse des atomes qui contribuent au moment électrique est choisi comme origine.

rotation des multipôles souvent, les multipôles sont exprimés dans un repère local (ce qui permet l'application de contraintes chimiques lors du procédé d'affinement), il faut donc les exprimer dans un répère commun pour qu'ils aient la même orientation.

intégration des moments atomiques Calcul du moment atomique sphérique sans trace à partir des paramètres multipolaires (équation 3.18).

calcul des moments cartésiens les moments atomiques sphériques sans trace sont transformés en moments cartésiens sans trace (équations 3.7 - 3.12).

changement d'origine les moments atomiques sont transformés afin d'avoir une origine commune, les expressions de changement d'origine (équations 3.15 et 3.16) sont donc appliquées.

sommation des moments atomiques Les contributions de chaque atome, exprimé selon l'origine choisie, sont additionnées pour calculer le moment électrique total.

Le calcul du moment dipolaire est similaire, mais il n'y a pas les étapes de transformation entre les tenseurs avec et sans trace.

3.2 Calcul de l'axe d'une hélice

Afin de caractériser le dipôle, il s'est avéré intéressant de mesurer sa déviation par rapport à l'axe de l'hélice ou de déterminer sa sensibilité à la courbure globale de l'hélice. Il existe différents algorithmes pour calculer l'axe d'une hélice [29] [30] [31], et on peut trouver des logiciels où ces méthodes sont implémentées (par exemple HELFIT [8]). Or, l'utilisation d'un logiciel externe est chronophage si un grand nombre de calculs est effectué : pour chaque hélice, le dipôle est calculé avec VMoPro, son axe est déterminé par le logiciel externe et les résultats respectifs sont exploités.
Pour cette raison, j'ai décidé d'implémenter une sous-routine dans MoPro qui permet le calcul de l'axe d'une hélice. Ainsi, la comparaison dipôle/axe peut être effectuée instantanément. J'ai choisi un algorithme basé sur de simples opérations vectorielles (algorithme 2, proposé par P. Kahn dans [29]) qui utilise les coordonnées des atomes Cα d'un fragment d'hélice pour le calcul de l'axe. Deux segments d'axe sont calculés à partir des coordonnées des 4 premiers et des 4 derniers atomes Cα de l'hélice. Puisqu'une hélice est par définition un objet géométrique régulier, on peut obtenir les deux segments d'axe en utilisant des opérations mathématiques simples comme le

produit scalaire et le produit vectoriel. Cet algorithme produit des résultats comparables à ceux d'un ajustement par les moindres carrés [29]. La routine a été implémentée dans VMoPro.
Notons que l'angle entre les deux segments qui sont utilisés pour déterminer l'axe reflète l'irrégularité de l'hélice (figure 3.2). Ainsi, un angle important indique que l'hélice a une courbure ou qu'un des résidus à l'extrémité n'a pas une conformation habituelle.

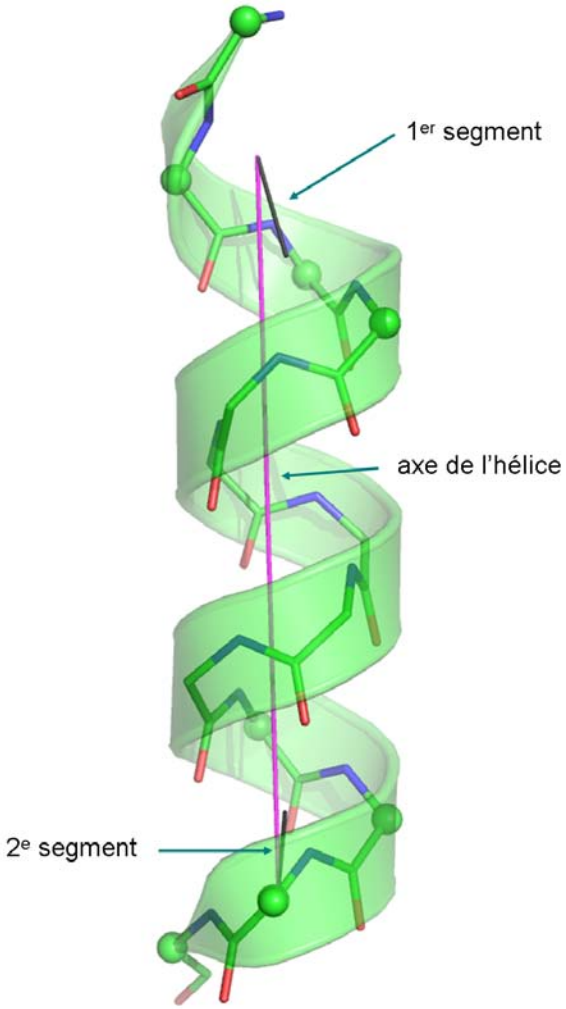


Figure 3.2 – Calcul de l'axe d'une hélice. Un premier segment d'hélice est calculé à partir des coordonnées des 4 premiers atomes Cα (représentés comme sphères vertes). De même, on obtient un deuxième segment d'axe avec les 4 derniers atomes Cα (représentés comme sphères vertes). La droite qui lie le point de départ du premier segment et l'extrémité du deuxième segment correspond à l'axe de l'hélice (représenté en rose).

Chapitre 4

Moments électriques de l'hélice α

4.1 Introduction

La structure de l'hélice α est telle que les liaisons covalentes C=O et N-H de l'unité peptidique s'alignent parallèlement à l'axe de l'hélice. Or, les atomes C, O, N et H de l'unité peptidique sont d'électronégativité différente, des charges positives sont donc attribuées aux atomes de carbone et d'hydrogène alors que des charges négatives sont attribuées aux atomes d'oxygène et d'azote (figure 4.1). La distribution de charge des unités peptidiques est alors dipolaire et ces dipôles s'accumulent dans l'hélice pour former un macrodipôle le long de l'axe, de l'extrémité C-terminale au N-terminale [32] (figure 4.2).

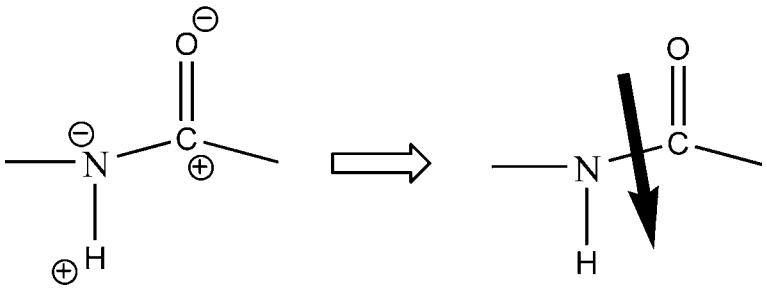


Figure 4.1 – Distribution de charge dans l'unité peptidique qui mène à un dipôle électrique.

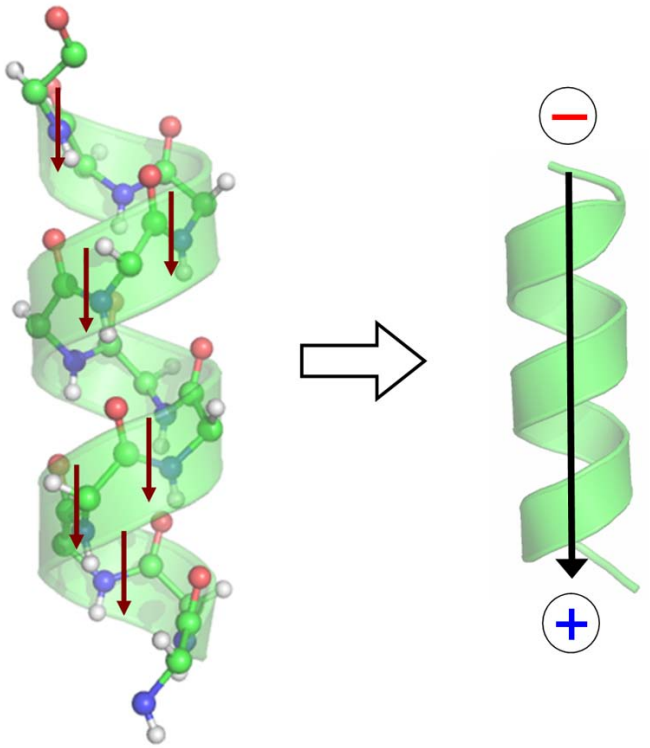


Figure 4.2 – *Gauche* : Arrangement des dipôles (flèches brunes) des unités peptidiques. Remarque : Pour une meilleure visualisation, les dipôles peptidiques ne sont pas tous représentés. *Droite* : Macrodipôle le long l'axe de l'hélice.

4.2 Historique

Dans la littérature, on trouve de nombreuses études sur le dipôle de l'hélice alpha, quelques résultats seront brièvement présentés ci-dessous.
L'existence du macrodipôle de l'hélice α a été confirmée expérimentalement en effectuant des mesures diélectriques sur des polypeptides hélicoïdaux synthétiques [32]. Le dipôle de l'unité peptidique seule a été quantifié par des mesures de spectroscopie micro-onde sur formamide [1] [33] menant à un module [2] de 3.71 D. En utilisant ce moment dipolaire pour les unités peptidiques de l'hélice α, on obtient un dipôle de 3.22 D par unité peptidique le long de l'axe [32]. Or, dans l'hélice, le groupement amide N-H d'une unité peptidique forme une liaison hydrogène avec le groupement carboxyl C=O appartenant à une autre unité peptidique. Sous l'influence de cette interaction, les atomes sont alors davantage polarisés, ce qui augmente le moment dipolaire du groupement. Se référant à des calculs théoriques [34], le dipôle de l'unité peptidique au sein d'une hélice est alors estimé atteindre jusqu'à 5.0 D [32].
Hol *et al.* proposaient un module de 3.5 D pour le dipôle de l'unité peptidique [25], se basant sur des charges de Mulliken [35] (figure 4.3). Ils effectuaient

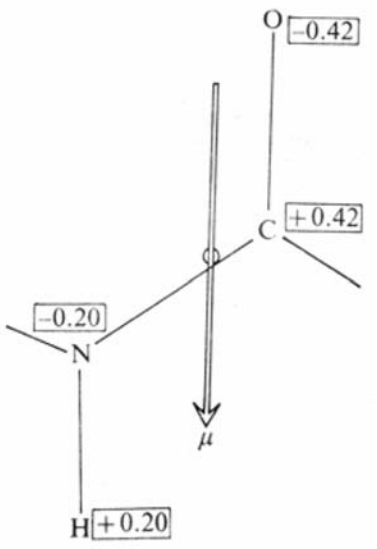


Figure 4.3 – Dipôle de l'unité peptidique, en utilisant des charges de Mulliken. Notons que la somme des charges est égale à zéro, donc l'unité est globalement neutre. La figure est extraite de [25].

1. Formamide est une molécule simple qui possède le groupement H-N-C=O de l'unité peptidique (Formule chimique : CH_3NO).
2. Le dipôle est souvent indiqué en Debye (D), 1 D = $3.33564 \cdot 10^{-30}$ Cm = $\frac{1}{4.803}$eÅ.

plusieurs calculs pour caractériser le macrodipôle de l'hélice α. Ainsi, ils pouvaient montrer que le champ électrique d'une hélice correspond approximativement à un dipôle, avec une charge positive à l'extrémité amino et une charge négative à l'extrémité carboxyl (figure 4.4). Hol *et al.* suggéraient également que le dipôle de l'hélice α soit impliqué dans de nombreux processus, comme la fixation d'un substrat chargé [25] ou le repliement protéique [26]. En effet, ils ont pu mettre en évidence que l'énergie d'interaction électrostatique entre les motifs de structure secondaire dans 20 protéines différentes est stabilisante [26].
Sheridan a montré que le potentiel électrostatique d'une hélice de polyalanine ressemble à celui d'un dipôle (figure 4.4). Cependant, après l'inclusion d'un résidu chargé (acide aspartique), le caractère dipolaire est presque entièrement masqué par le potentiel de la charge supplémentaire [36].

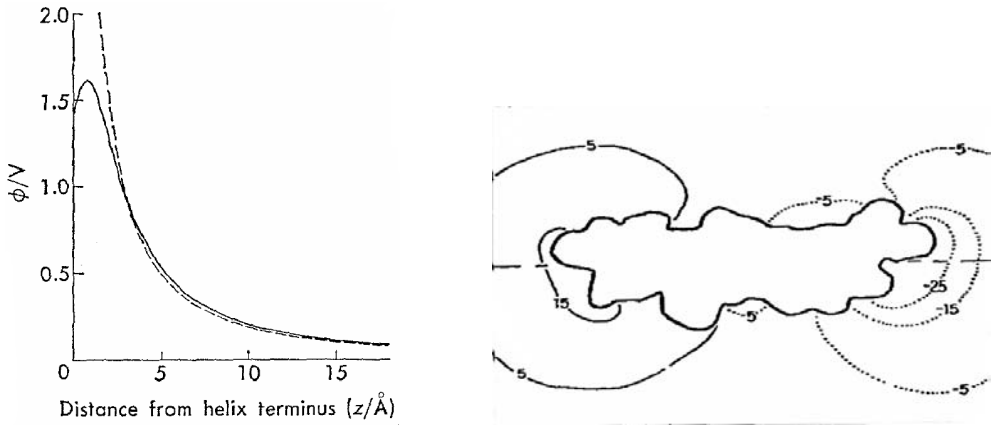


Figure 4.4 – *A gauche* : Potentiel en Volt, à une distance de z sur l'axe de l'hélice à partir de l'extrémité N-terminale. Ligne noire : potentiel à partir de 7 unités peptidiques, ligne pointillée : potentiel d'un dipôle rectiligne, qui constitue une bonne approximation du dipôle calculé à partir des unités peptidiques (la figure est extraite de [25]). *A droite* : potentiel électrostatique d'une hélice α d'une longueur de 15 résidus. Les potentiels positifs et négatifs sont représentés par des lignes noires et des lignes pointillées respectivement (la figure est extraite de [36]).

Lockhart et Kim ont effectué des expériences d'effet Stark sur des polypeptides hélicoïdaux synthétiques. Ils ont pu ainsi montrer expérimentalement que le champ électrique à l'extrémité de l'hélice, engendré par le macrodipôle de l'hélice α, n'augmente pas avec la longueur de l'hélice [37] (au contraire du dipôle, qui augmente avec la longueur). De manière intéressante, la présence d'un résidu chargé dans le polypeptide n'avait pas d'influence remarquable sur le champ électrique, les chaînes latérales n'interviennent alors pratiquement pas dans l'interaction entre une hélice et un substrat. Finalement, She-

ridan et collaborateurs estimaient que ce ne sont que les deux premiers tours d'hélice qui contribuent à ≈90% de l'énergie d'interaction.
Le transfert des paramètres multipolaires de la banque ELMAM, permettant une description précise de la densité de charge, offre une nouvelle possibilité de quantifier la valeur du macrodipôle de l'hélice α. Afin d'aller plus loin de ce qui a été fait dans la littérature, les feuillets β ont été analysés également et en particulier, les moments quadrupolaires ont étés calculés. En conjonction avec les autres méthodes implémentées dans la suite MoPro, cela permettra notamment de calculer les interactions électrostatiques, dans lesquelles le macrodipôle est impliqué.

4.3 Caractérisation du dipôle de l'unité peptidique

Dans un premier temps, le dipôle de l'unité peptidique seule, calculé à partir des paramètres multipolaires d'ELMAM, est analysé.
Les paramètres multipolaires ont été transférés de la banque ELMAM vers un modèle d'une unité peptidique. Notons que l'unité peptidique n'est pas neutre d'après les paramètres de la banque, elle porte une charge globale de -0.28 |e| (tableau 4.1). Pour cette raison, l'unité a été neutralisée, en ajoutant 0.07 |e| à la charge de chaque atome [3]. Ensuite, le dipôle a été calculé avec MoPro.

atome	**ELMAM** (|e|)	**ELMAM neutralisé** (|e|)	**Mulliken** (|e|)
C	0.02	0.09	0.42
O	-0.31	-0.24	-0.42
N	-0.31	-0.24	-0.2
H	0.32	0.39	0.2

Table 4.1 – Charges atomiques (ELMAM) de l'unité peptidique avant et après neutralisation ainsi que les charges de Mulliken utilisées dans [25].

Le dipôle généré par les charges atomiques ELMAM est de 3.9 D, il s'étend de l'atome d'oxygène vers l'atome d'hydrogène de l'unité peptidique. Notons qu'il n'est pas tout à fait parallèle aux liaisons N-H et C=O, il est plutôt

3. Notons que c'est une méthode de neutralisation simple mais il existe des possibilités plus sophistiquées. Par exemple, on pourrait pondérer la variation par le poids ou la charge initiale de chaque atome.

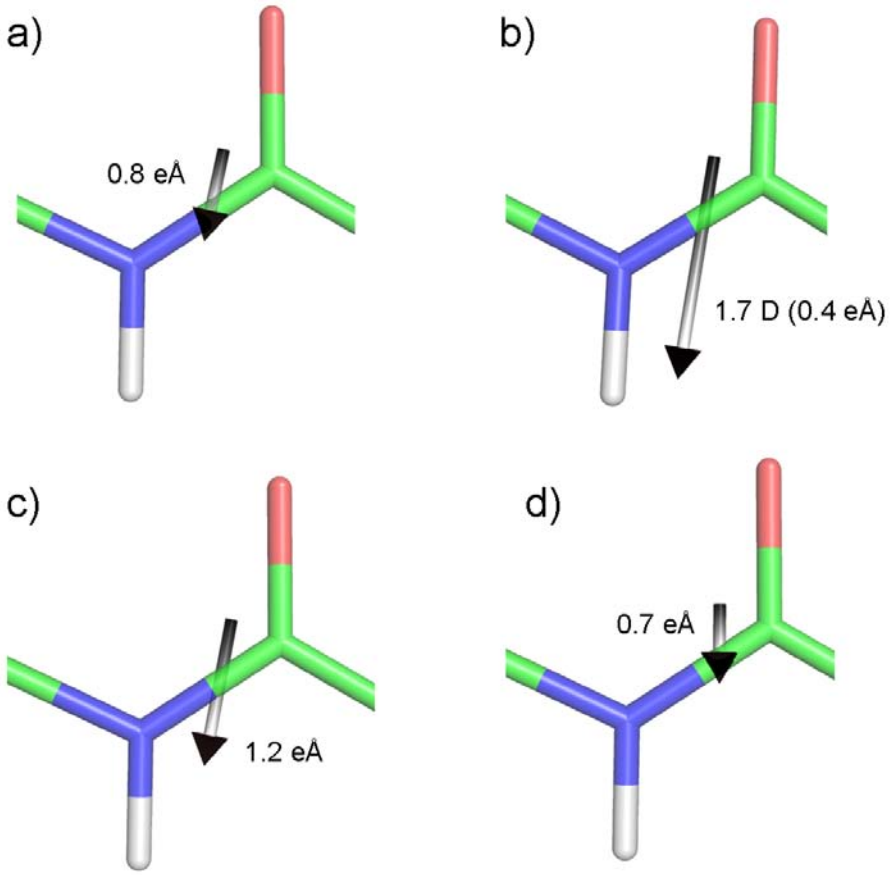


Figure 4.5 – Dipôles de l'unité peptidique, formé par les atomes N (bleu), H (blanc), C (vert) et O (rouge). (a) Charges atomiques, ELMAM (b) dipôle généré par les dipôles atomiques, ELMAM (c) dipôle total, ELMAM (d) Charges de Mulliken. Les dipôles sont représentés en eÅ, sauf pour le dipôle atomique (b), qui est représenté en Debye. Ils émergent du centre de masse du système O=C-N-H.

incliné dans la direction H-O (figure 4.5 a). Le dipôle calculé à partir des dipôles atomiques uniquement est plus faible, avec un module de 1.7 D, mais son orientation est similaire (figures 4.5 b et 4.6). L'angle entre le moment dû aux dipôle atomiques et celui dû aux charges est de 4.7°.

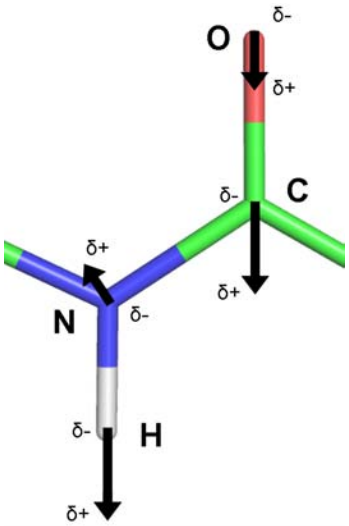


Figure 4.6 – Dipôles atomiques des atomes dans l'unité peptidique, représentés comme flèches noires. Ils sont alignés parallèlement, à l'exception du dipôle de l'atome d'azote. Par conséquent, le dipôle total est légèrement incliné.

Afin de comparer les dipôles avec les valeurs de la littérature, les charges de Mulliken (tableau 4.1) ont été assignées aux atomes de l'unité peptidique et leur dipôle a été calculé. Tandis que son module, de 3.5 D, est comparable au dipôle calculé à partir des charges atomiques d'ELMAM, son orientation est différente, il est presque parallèle aux liaisons C=O et N-H (figure 4.5 d). En effet, la charge de l'atome de carbone dans les deux cas n'est pas la même : l'atome est presque neutre (0.093 |e|) pour ELMAM et il porte une charge de Mulliken très importante (0.42 |e|). En effet, Hol *et al.* changeaient la charge de l'atome de carbone de 0.32 |e| [35] à 0.42 |e| afin de neutraliser le groupement. Cette différence a nécessairement un effet sur l'orientation du dipôle.
Le module du dipôle total ELMAM, calculé à partir des charges *et* des dipôles atomiques, est de 5.6 D, ce qui est plus élevé que le moment dipolaire basé sur les charges de Mulliken (3.5 D). Les dipôles atomiques représentent alors une partie non négligeable du dipôle total. L'angle entre le moment dipolaire dû aux charges de Mulliken et le dipôle totale d'ELMAM est de 10.4°. Notons que les effets de polarisation dûs aux liaisons hydrogènes entre les unités

peptidiques ne sont pas pris en compte dans ces calculs [4]. Il est alors certain que le dipôle en serait renforcé et qu'il soit en vérité plus fort, comme il a été montré par [32] et [33].
Afin de vérifier l'impact de la neutralisation du fragment sur le calcul du dipôle, le moment dipolaire de l'unité peptidique a été calculé également sans avoir été neutralisé préalablement. Prenant la même origine, c'est-à-dire le centre de masse, le dipôle total a un module de 4.9 D et l'angle entre les dipôle calculé à partir du fragment neutralisé et non-neutralisé est de 1.2°. Le dipôle du fragment neutralisé est alors légèrement sur-estimé tandis que sa direction reste similaire.

4.4 Caractérisation du dipôle d'hélices α

4.4.1 Hélice de polyalanine idéalisée

Dans un premier temps, le dipôle d'une hélice idéalisée est caractérisé. Une hélice de polyalanine a été générée avec le logiciel Moleman [39], qui utilise une géométrie de $\phi = -60°$ et $\psi = -50°$ pour les angles dièdres de la chaîne principale. Les atomes d'hydrogène n'étant pas présents dans ce modèle d'hélice, ils ont été ajoutés avec Molprobity [40]. Ensuite, les paramètres multipolaires ont été transférés de ELMAM au modèle et les hélices ont été neutralisées. Le moment dipolaire total, engendré par les atomes de l'unité peptidique, ainsi que l'angle entre le dipôle et l'axe de l'hélice ont été calculés avec MoPro.
Dans l'hélice idéale, le dipôle croît linéairement avec la longueur de l'hélice, la croissance est de 5.1 D par unité peptidique (figure 4.7) ce qui est moins élevé que le dipôle de l'unité peptidique seule, qui est de 5.6 D. La diminution peut être due à l'inclinaison des moments dipolaires qui ne sont pas parfaitement parallèles les uns par rapport aux autres, et, par conséquent, la somme vectorielle ne correspond pas à la somme de leurs modules.
Ensuite, l'angle entre l'axe de l'hélice α et le dipôle a été analysé. Dans la littérature, il a été postulé que l'arrangement régulier des dipôles peptidiques résulte en un macrodipôle qui est parallèle à l'axe. Dans l'hélice idéalisée, l'angle entre le dipôle et l'axe est en effet très faible, ils est de moins de 6° en général et tend vers $\approx 1.5°$ pour les hélices très longues (> 15 unités peptidiques) (figure 4.8). Les calculs confirment alors que le dipôle est parallèle

4. Les valeurs moyennes dans la banque ELMAM incluent de la polarisation. Cependant, les composés qui ont servi à construire la banque n'adoptent généralement pas de géométrie hélicoïdale. L'effet coopératif dans les hélices, qui est proposé par la littérature [38], est donc négligé. Notons toutefois que la polarisation est à la limite de l'observable.

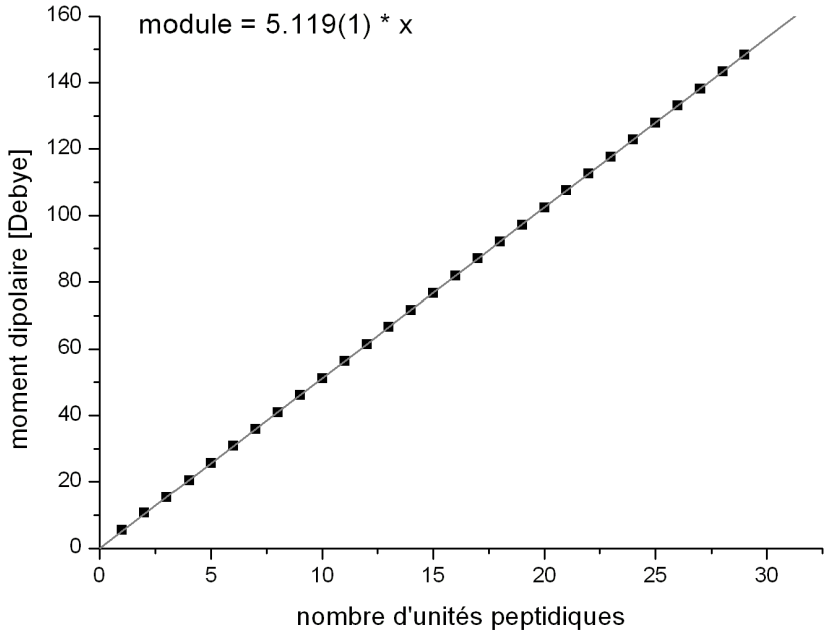


Figure 4.7 – Module du dipôle total en fonction de la longueur de l'hélice. La relation est linéaire, le dipôle augmente de 5.1 D par unité peptidique.

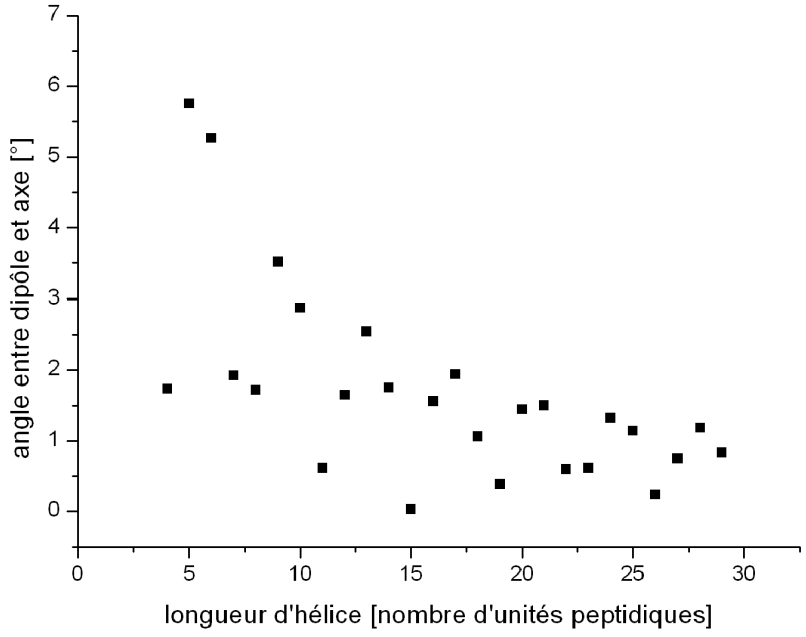


Figure 4.8 – L'angle entre le dipôle et l'axe de l'hélice en fonction de sa longueur. Les angles sont plus importants pour des petites hélices ($<$ 10 Å) mais ils sont généralement faibles.

à l'axe de l'hélice.
Cependant, les hélices dans les structures protéiques sont moins ordonnées que l'hélice de polyalanine idéalisée. On peut constater les déviations suivantes par rapport à une géometrie idéale (figure 4.9) :

- Les hélices sont courbées, elles n'ont donc pas un axe rectiligne.
- L'enroulement n'est pas aussi régulier que celui de l'hélice parfaite. On observe par exemple des liaisons hydrogènes i –> i+3 (section 1.3.1) à l'extrémité qui interrompent l'arrangement des unités peptidiques.
- On peut observer qu'une hélice contient un résidu proline. Ce résidu est considéré comme "casseur" du repliement des structures secondaires car il ne possède pas de groupement amino (N-H) pour former une liaison hydrogène, ce qui modifie ses angles de Ramachandran, ϕ et ψ,

Pour cette raison, des calculs ont été effectués avec des hélices provenant de structures réelles.

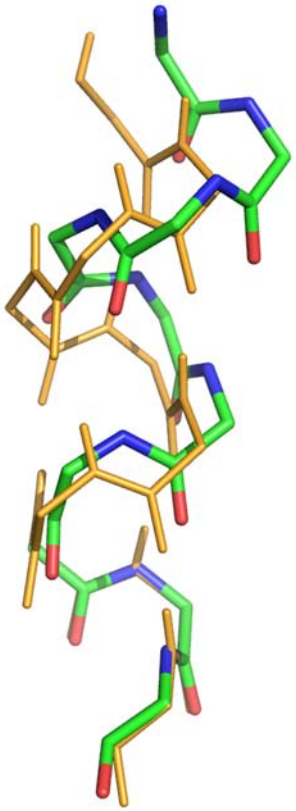


Figure 4.9 – Superposition d'une hélice idéale de polyalanine (en jaune) et d'une hélice dans la protéine aldose réductase (résidus 162-171) (coloré). Un seul tour d'hélice suffit pour observer une déviation par rapport à la géométrie idéale.

4.4.2 Hélices réelles

L'étude a été effectuée sur les hélices issues des trois structures protéiques suivantes [5] :

- Aldose réductase humaine (modèle fourni par B. Guillot, 0.66 Å de résolution) [7]
- Xylanase thermostable de *thermoascus aurantiacus* (code PDB : 1I1W, 0.89 Å de résolution)
- Beta glucosidase de *thermus thermophilus* (code PDB : 1UG6, 0.99 Å de résolution)

Les trois protéines adoptent une structure en tonneau $(\alpha/\beta)_8$ (figure 4.10) et elles diffractent à haute résolution. L'avantage de la haute résolution est que la position des atomes "lourds" est déterminée avec plus de précision.

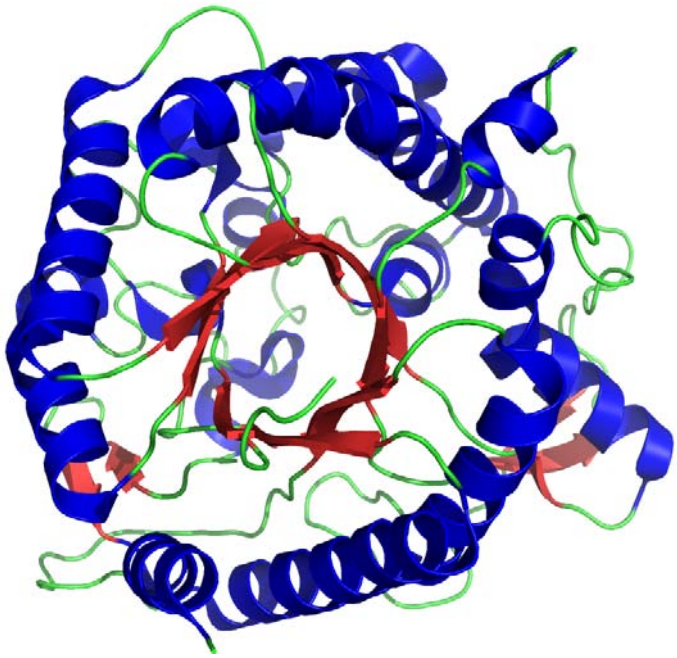


Figure 4.10 – Structure de la protéine beta-glucosidase. Seule la chaîne principale est représentée, les hélices sont colorées en bleue, les feuillets en rouge.

Afin de savoir quels résidus adoptent une géométrie en hélice α, le logiciel

5. Les critères de choix étaient les suivants : la résolution des données de diffraction est supérieure à 1 Å, le facteur d'accord statistique R est inférieur à 15%, et la différence entre R et R_{free} n'est pas trop élevée. De plus il ne doit pas s'agir d'une structure trop ancienne (après 1995).

STRIDE [6] a été utilisé [41]. En plus, les structures ont été inspectées (avec Coot ou Pymol) pour ajuster l'assignement. Afin d'exclure le désordre, les acides aminés ayant une conformation multiple n'ont pas été pris en compte. Si les atomes d'hydrogène n'étaient pas présents dans le modèle protéique, ils ont été ajoutés avec Molprobity [40]. Ensuite, les paramètres multipolaires ont été transférés de la banque ELMAM au modèle. Pour tous les calculs suivants, le dipôle total a été calculé à partir des atomes de l'unité peptidique, ce qui correspond à un fragment dont la charge globale a été neutralisée préalablement (utilisant la même méthode comme précédemment pour l'unité peptidique).
D'abord, l'agrandissement du dipôle total en fonction de la longueur a été analysé. Dans un premier temps, le calcul a été effectué avec une seule hélice qui se trouve dans la protéine beta-glucosidase (résidus 135-153). Le dipôle a été calculé en augmentant successivement le nombre d'unités peptidiques du fragment pris en compte. De plus, l'angle entre l'axe de l'hélice (déterminé à partir des résidus 135-153) et le dipôle des différents fragments a été déterminé. Comme pour l'hélice de polyalanine, on remarque une relation

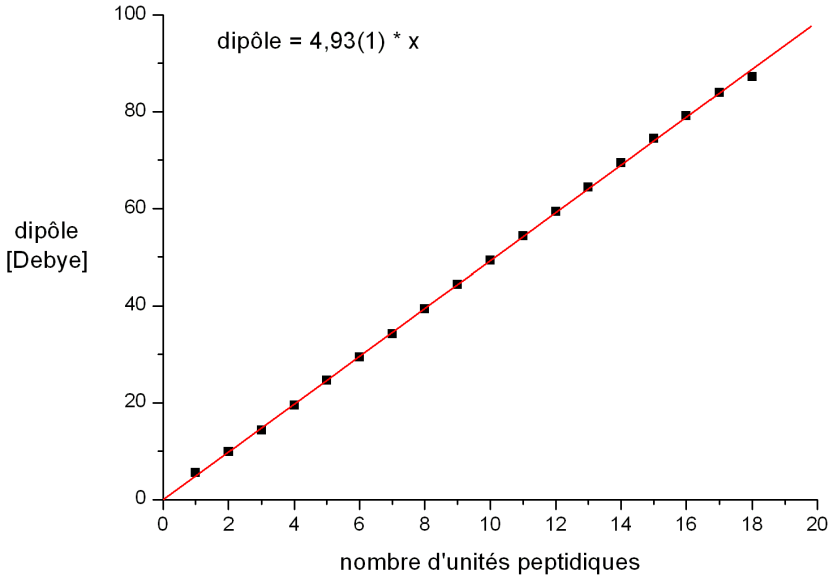


Figure 4.11 – Moment dipôlaire en fonction de la longueur du fragment ainsi qu'un ajustement linéaire (ligne rouge) passant par l'origine.

linéaire entre le module du dipôle et le nombre d'unités peptidiques (figure

6. STRIDE est un logiciel permettant d'assigner les acides aminés qui font partie d'un motif de structure secondaire. Son algorithme est basé sur l'énergie des liaisons hydrogène ainsi que les angles dièdres de la chaîne principale.

4.11). D'après l'ajustement linéaire, le module augmente de 4.9 D par unité peptidique, ce qui est un peu plus faible que pour le dipôle de l'hélice de polyalanine (5.1 D). L'accroissement plus faible peut être expliqué par une géométrie moins régulière par rapport à l'hélice idéale.
Ensuite, l'angle entre l'axe de l'hélice et le moment dipolaire a été analysé en fonction du nombre d'unités peptidiques (figure 4.12). On peut noter qu'à partir de 6 unités peptidiques, l'angle reste inférieur à 5°. Donc malgré la courbure, le moment dipolaire s'aligne à l'axe de l'hélice. L'angle est toutefois plus élevé que pour l'hélice idéalisée.

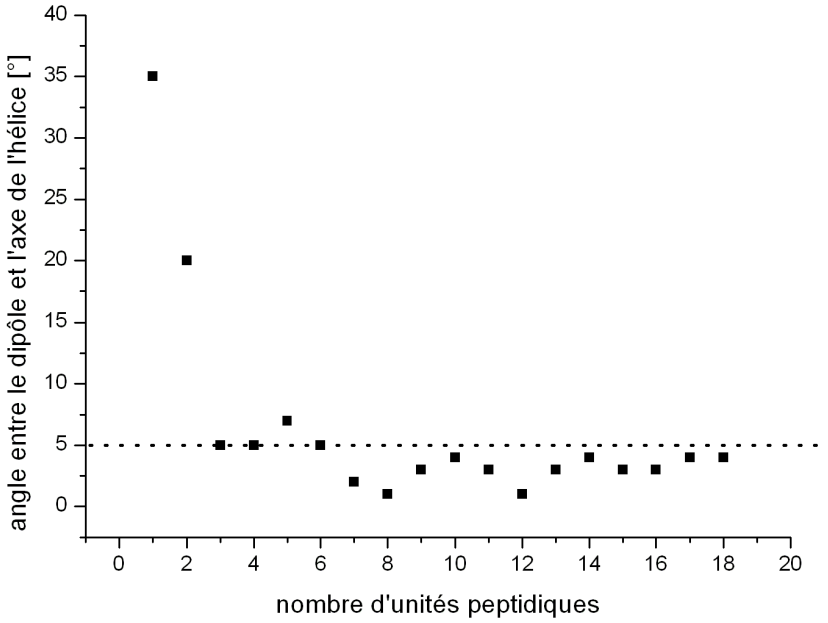


Figure 4.12 – L'angle entre l'axe de l'hélice et le moment dipolaire en fonction de la longueur d'hélice. La ligne pointillée est à 5°.

Relation entre courbure et dipôle

Le calcul sur une hélice réelle ci-dessus a montré que son dipôle est moins élevé que celui d'une hélice idéale. Pour cette raison, j'ai analysé s'il y a une relation entre la courbure d'une hélice et son moment dipolaire.
Pour des fragments d'hélices de même longueur, le module du moment dipolaire a été déterminé en fonction de la courbure de l'hélice. La courbure est représentée ici comme l'angle entre les deux segments d'axe qui ont été utilisés pour le calcul de l'axe de l'hélice (figure 4.13).

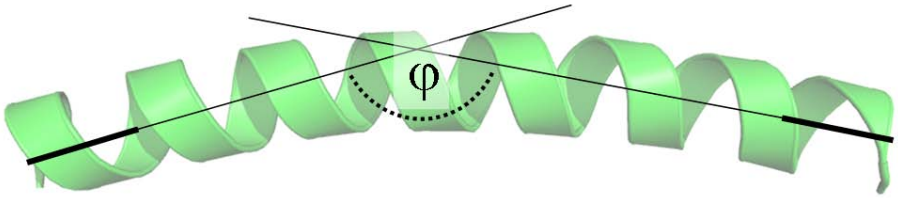


Figure 4.13 – Hélice courbée dans la protéine calmodulin (code PDB 1EXR, résidu 64-93) L'algorithme permettant le calcul de l'axe fournit deux segments d'hélice, et l'angle entre ces segments en reflète la courbure.

Les calculs ont été effectués sur des fragments hélicoïdaux d'une longueur de 5, 10 et 15 acides aminés. Une hélice de 17 AA par exemple est utilisée 3 fois : les 5 premiers résidus sont dans le premier groupe, les 10 premiers AA sont dans le deuxième et les résidus 1 à 15 sont dans le troisième. Si l'hélice contenait l'acide aminé proline, ce résidu ainsi que ses voisins directs n'ont pas été pris en compte dans les calculs. On peut ainsi déconvoluer la distortion structurale induite par le résidu proline de la courbure "naturelle" de l'hélice.
Pour des hélices longues de 5 acides aminés, on ne peut pas voir de relation apparente (figure 4.14) à cause de la dispersion des points. A une valeur constante de 20 D du moment dipolaire par exemple correspond une courbure entre 5 et 25°. La dispersion des points pourrait résulter de la géométrie irrégulière des petites hélices où l'arrangement des unités peptidiques est moins parallèle que celui des longues hélices. Par conséquent, le dipôle total des hélices de longueur 5 AA varie entre 18 et 21 D.
Pour les hélices de longueur 10 AA, on observe également une dispersion des couples dipôle-courbure. La courbure varie entre 0 et 20°, tandis que le moment dipolaire se situe entre 43 et 46 D. Rappelons que le dipôle d'une hélice idéale de longueur 10 AA est de 46.5 D.
Finalement, pour les hélices de 15 AA, la relation entre courbure et moment dipolaire semble avoir un caractère linéaire. Il serait donc possible d'exprimer la courbure en fonction du module du dipôle. La courbure maximale est de

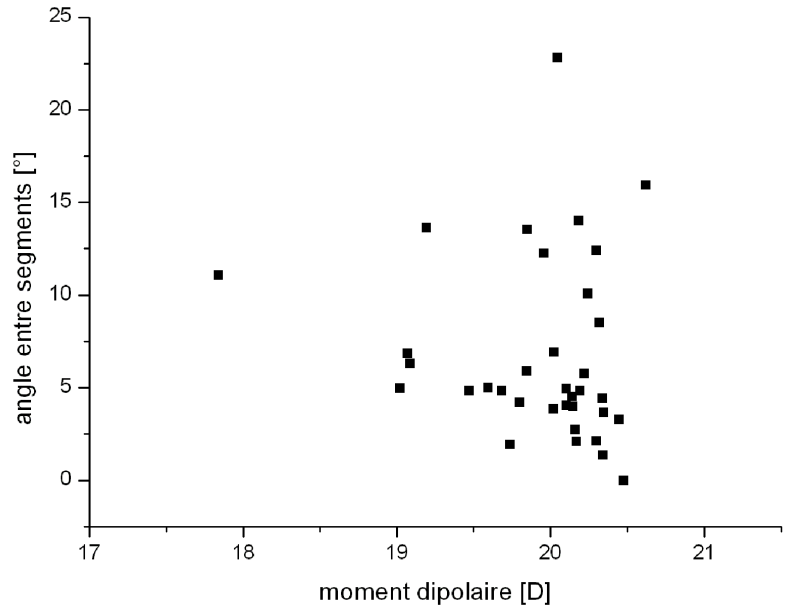


Figure 4.14 – Courbure de l'hélice en fonction du module du dipôle, longueur d'hélice 5 acides aminés.

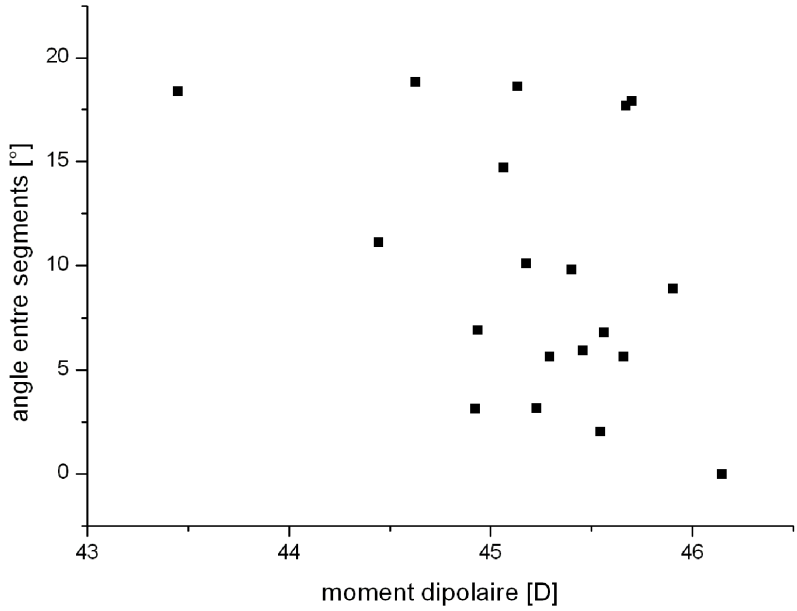


Figure 4.15 – Courbure de l'hélice en fonction du module du dipôle, longueur d'hélice 10 acides aminés.

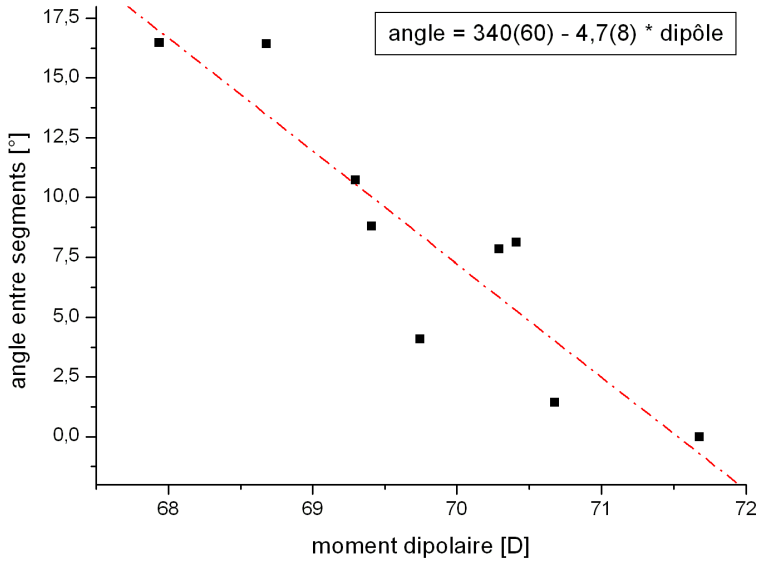


Figure 4.16 – Courbure de l'hélice en fonction du module du dipôle, longueur d'hélice 15 acides aminés.

16° et le dipôle varie entre 68 et 72 D. La courbure maximale est plus faible que pour les fragments de longueur 5 ou 10 AA. Il apparaît alors qu'une géométrie régulière soit requise pour une longue hélice.
Pour conclure, le module du dipôle est influencé par la courbure de l'hélice, il est donc moins élevé que celui d'une hélice idéalisée. Pour les hélices de 15 AA, la courbure peut être exprimée en fonction du dipôle par une relation linéaire. Cependant, il n'était pas possible d'élaborer une relation claire entre la courbure et le module du dipôle pour les hélices longues de 5 ou 10 AA.
Afin d'élaborer une équation simple permettant d'estimer le module du dipôle en fonction de sa longueur, les moment dipolaires de plusieurs hélices réelles ont été calculés. Comme pour l'analyse ci-dessus, les conformations multiples et les résidus proline ont été exclus des fragments. En tout, 31 hélices, ou fragments d'hélice, ont été utilisées pour l'analyse [7].
La figure 4.17 montre le module du dipôle en fonction de la longueur de l'hélice. La relation est linéaire et correspond à une droite ayant une ordonnée à l'origine nulle. Utilisant l'équation de cette droite pour une hélice quelconque qui ne contient pas le résidu proline, il est possible d'estimer son moment dipolaire.

7. Remarque : Pour cette étude, la structure des protéines aldose réductase, beta glucosidase, xylanase (décrit ci-dessus) et de CTX-M-9 beta-lactamase (code PDB : 1YLJ, résolution : 0.98) ont été utilisées.

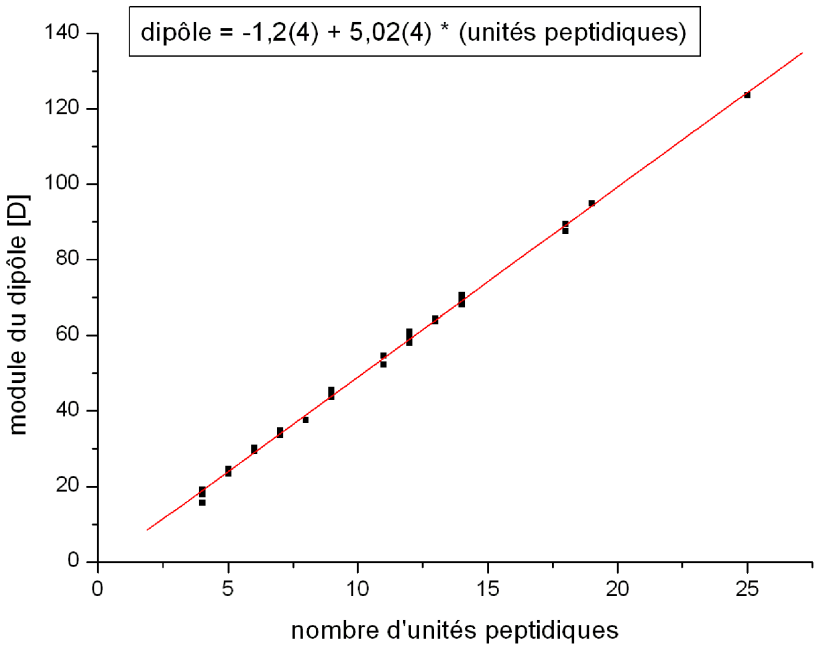


Figure 4.17 – Module du dipôle en fonction du nombre d'unités peptidiques dans des hélices courbées ainsi qu'un ajustement linéaire.

Une autre information importante est l'orientation du dipôle. La figure 4.18 montre l'angle entre l'axe de l'hélice et le moment dipolaire en fonction du nombre d'unités peptidiques présentes dans l'hélice. Pour des petites hélices (d'une longueur inférieure à 8 acides aminés), l'angle peut être important. Cependant, à partir d'une longueur de 11 AA, l'angle est toujours inférieur à 7°, le dipôle est alors pratiquement parallèle à l'axe.

Influence du résidu proline

Le résidu proline est considéré comme un "casseur" d'hélice puisque sa chaîne latérale forme une liaison covalente avec l'atome d'azote de la chaîne principale. Ainsi, cet acide aminé ne possède pas le groupement amide N-H qui peut former les liaisons hydrogènes typiques de l'hélice. Pourtant, il est possible de trouver un nombre substantiel de prolines dans les hélices [42]. De manière intéressante, le résidu proline est situé souvent à l'extrémité N-terminale de l'hélice [43].
Il est alors intéressant d'analyser l'influence du résidu proline sur le module du dipôle d'une hélice. Puisque l'atome d'hydrogène du groupement amide n'est pas présent, seulement les atomes N, C et O font partie de l'unité peptidique dans la librairie ELMAM. La charge positive de l'hydrogène est

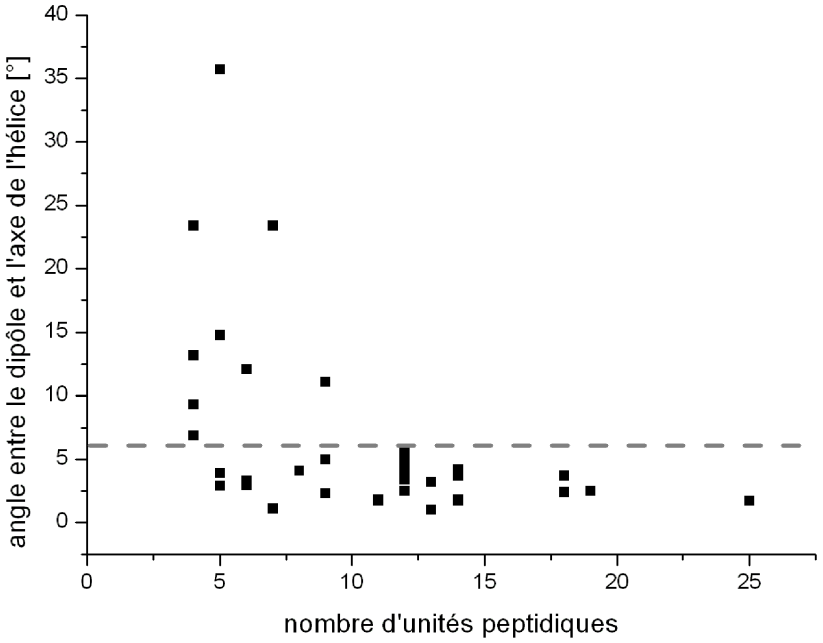


Figure 4.18 – Angle entre l'axe de l'hélice et le moment dipolaire pour plusieurs hélices de longueurs différentes.

alors manquante [8], le centre des charges positives se trouve alors à une autre position par rapport à une hélice de même longueur, formé par des unités peptidiques H-N-C=O. Le dipôle devrait donc être affecté par la présence du résidu proline.
Parmi les hélices dans les protéines aldose réductase, beta lactamase, beta glucosidas et xylanase, 14 contiennent un résidu proline. Dans la majorité des cas, il se trouve à l'extrémité N-terminale mais dans deux hélices, il est situé à l'extrémité C-terminale. Dans l'aldose réductase, une proline forme la 9e unité peptidique dans une hélice longue de 13 AA. Dans la protéine xylanase, on trouve une proline à la 10e position parmi 14 acides aminés.
Pour des fragments d'hélice longs de 10 unités peptidiques (UP), le module du dipôle a été calculé, en neutralisant préalablement la charge. Pour les deux hélices contenant le résidu proline à l'extrémité C-terminale, le début et la fin du fragment ont été changés, en gardant la longueur de 10 UP. Par exemple dans la protéine xylanase, le fragment constitué des acides aminés 214-224 contient le résidu proline à la 9e position, dans le fragment 215-225, la proline est à la 8e position.

8. La charge de l'atome d'azote est de -0.252 |e|, ce qui est similaire à celle de l'atome d'azote du motif N-H-C=O de l'unité peptidique complète.

La figure 4.19 montre le module du moment dipolaire en fonction de la position du résidu proline.

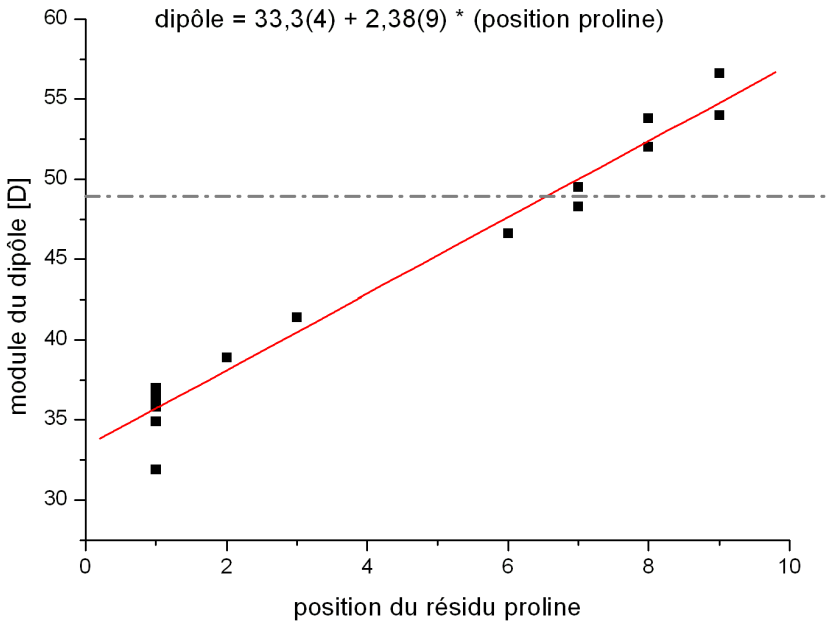


Figure 4.19 – Module du moment dipolaire en fonction de la position du résidu proline, pour des fragments d'hélice longs de 10 unités peptidiques. La ligne pointillée grise est à 49 D, ce qui correspond au dipôle d'une hélice de longueur 10 UP, ne contenant pas de résidu proline.

Notons tout d'abord que le dipôle varie entre 30 et 60 D, pour un nombre fixe d'atomes. La variation est due à la position du résidu proline dans le fragment. Le moment dipolaire d'une hélice de même longueur, sans le résidu proline, est de 49 D (ligne grise dans la figure 4.19). La position de la proline à l'extrémité N-terminale permet alors de diminuer le moment dipolaire par rapport à une hélice "normale", sans résidu proline, tandis que le dipôle est renforcé quand la proline se trouve à l'extrémité C-terminale. Il semble alors que l'inclusion d'un résidu proline permette de moduler la magnitude du dipôle de l'hélice alpha.
L'équation établie à partir du graphe (figure 4.19) n'est valable que pour les hélices longues de 10 unités peptidiques. Afin de généraliser, le moment dipolaire d'hélices de différentes longueurs et dans lesquelles le résidu proline forme la 1ère unité peptidique a été calculé.
D'après le graphique (figure 4.20), la relation entre la longueur d'hélice et le module du moment dipolaire est linéaire, comme il a été observé pour les hélices sans proline. Cependant, les coefficients directeurs des deux droites

sont différents, ils sont de 3.7(2) et 5.02(4) pour l'hélice avec et sans proline, respectivement. Cette différence est probablement due au manque de la charge positive de l'atome d'hydrogène dans l'unité peptidique, ce qui influence la charge globale et ainsi la neutralisation. L'équation de la droite permet de calculer le module du dipôle pour une hélice quelconque dont le 2e acide aminé est un résidu proline [9].

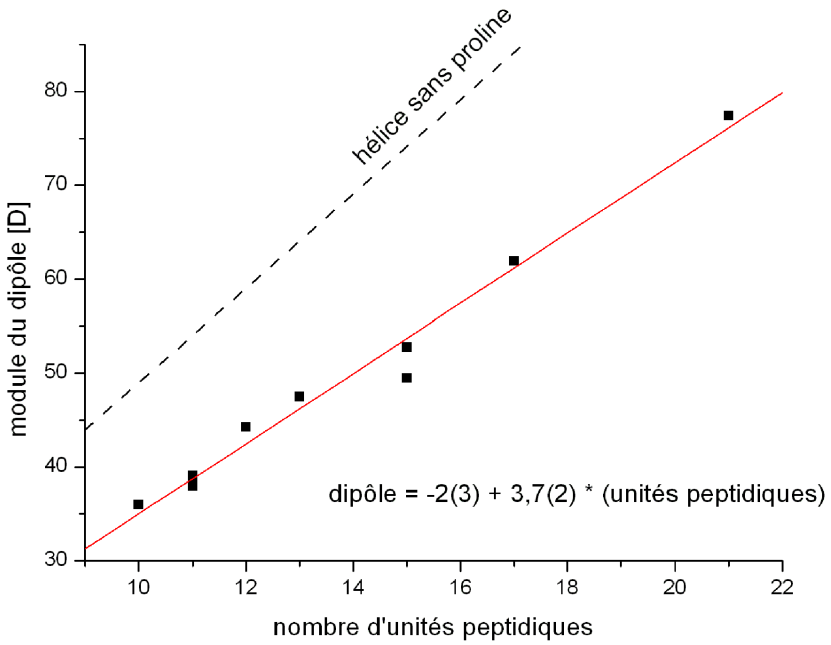


Figure 4.20 – Module du dipôle en fonction de la longueur d'hélice. Notons que le résidu proline forme la première unité peptidique et correspond donc au 2ème acide aminé du fragment. Pour comparaison, l'équation caractérisant le module du dipôle pour une hélice sans proline a été tracée (ligne pointillée).

La figure 4.21 montre l'angle entre l'axe de l'hélice et le dipôle en fonction du nombre d'unités peptidiques. Les points sont assez dispersés et l'échantillon n'est pas très grand mais on observe que les angles sont un peu plus élevés que pour une hélice sans résidu proline. La direction du dipôle est alors affectée bien que l'influence n'est pas autant de changer complètement d'orientation. Toutefois, le calcul de l'axe de l'hélice est probablement biaisé par la géométrie tordue introduite par la proline (son atome C_α est utilisé pour calculer le premier segment d'axe).

9. Si le premier acide aminé est une proline, le dipôle n'est pas affecté, c'est son groupement carbonyl C=O qui fait alors partie de la première unité peptidique.

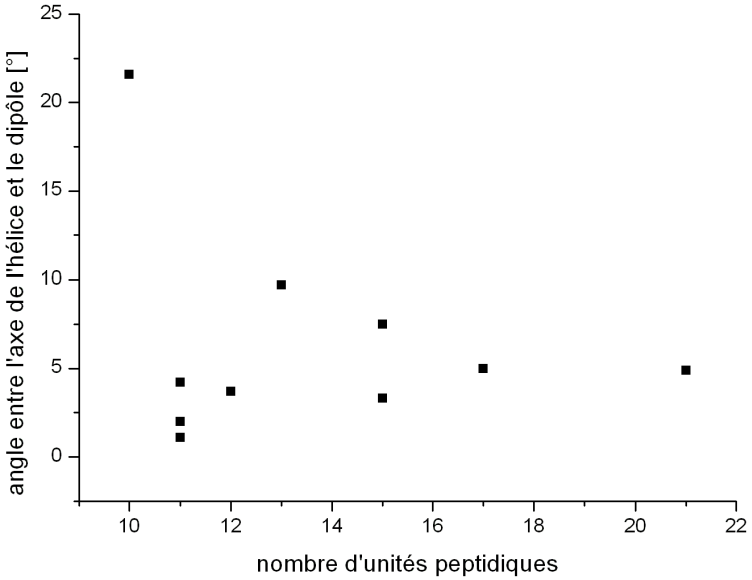


Figure 4.21 – Angle entre l'axe de l'hélice et le dipôle en fonction de la longueur d'hélice en nombre d'unités peptidiques.

4.5 Moment quadrupolaire de l'hélice α idéalisée

De manière intéressante, seul un moment dipolaire a été associé aux motifs de structure secondaire et les moments électriques d'ordre supérieur n'ont pas été considérés dans la littérature. Par conséquent, j'ai entrepris l'analyse du moment quadrupolaire des hélices et des feuillets. La figure 4.22 montre l'arrangement de charges ponctuelles qui génèrent un quadrupôle parfait (avec un dipôle nul).
Le moment quadrupolaire est représenté par un tenseur symétrique d'ordre 2. Celui-ci peut être diagonalisé afin d'obtenir trois vecteurs propres (orthogonaux entre eux) et trois valeurs propres, dont seulement deux sont indépendantes pour le moment quadrupolaire sans trace. Il existe une convention permettant de poser :

$$|\theta_{xx}| \leq |\theta_{yy}| \leq |\theta_{zz}| \tag{4.1}$$

et d'appeler θ_{zz} le moment quadrupolaire. Le paramètre d'asymétrie η,

$$\eta = -\frac{\theta_{yy} - \theta_{xx}}{\theta_{zz}}, \tag{4.2}$$

mesure la déviation du moment quadrupolaire par rapport à une symétrie axiale [44]. Par exemple, $\eta = 0$ pour un quadrupôle longitudinal et $\eta = 1$

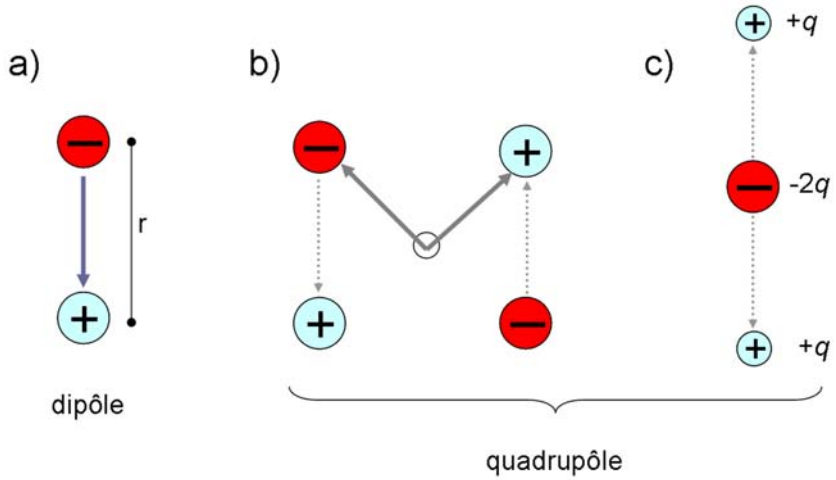


Figure 4.22 – (a) Représentation schématique d'un moment dipolaire parfait, qui est constitué de deux charges ponctuelles opposées ($+q$ et $-q$) se trouvant à une distance r l'une par rapport à l'autre. La flèche grise représente le dipôle qui est donné par $q \cdot \vec{r}$. (b) Quadrupôle transversal et (c) quadrupôle longitudinal. Les deux distributions de charge sont constituées de deux dipôles opposés (flèches pointillées). Les vecteurs propres du tenseur quadrupôle (lignes grises) émergent du centre de masse du système.

pour un quadrupôle transversal (distributions c et b respectivement, figure 4.22). La dimension des éléments tensoriels est une charge multipliée par une longueur au carré, par exemple eÅ^2. Par ailleurs, le moment dipolaire a la dimension d'une longueur multipliée par une charge, par exemple eÅ. Il se pose alors la question de savoir comment comparer les deux moments pour attribuer à un fragment un caractère majoritairement dipolaire, ou quadrupolaire.
Une possibilité consiste à analyser le potentiel électrostatique pour révéler le caractère d'une distribution de charge. En effet, les distributions créent des potentiels caractéristiques (figure 4.23) qui peuvent aider à faire la distinction entre elles.
Le potentiel électrostatique calculé à partir des atomes de l'unité peptidique d'une hélice de polyalanine idéalisée est représenté figure 4.24. On reconnait clairement les lobes positifs et négatifs typiques du potentiel d'un dipôle. Pour des hélices polyalanine de taille croissante, les valeurs propres du quadrupôle, et le paramètre η sont résumés tableau 4.2.
Les valeurs de θ_{yy} et θ_{xx} sont du même ordre de grandeur, θ_{zz} est plus élevé. Le paramètre d'asymétrie η fluctue pour des hélices de petite taille, mais quand le nombre d'unités peptidiques augmente, il est inférieure à 0.2. Les vecteurs

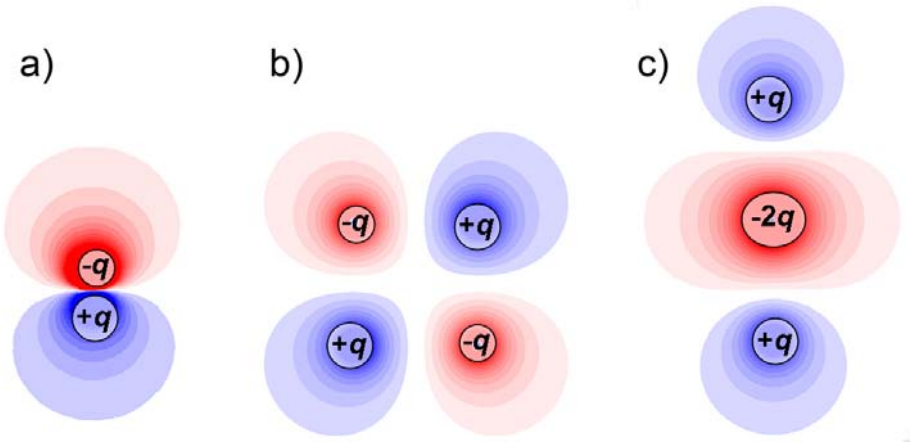


Figure 4.23 – Potentiel électrostatique du dipôle (a) et des deux configurations quadrupolaires (b et c) qui sont représentées figure 4.22.

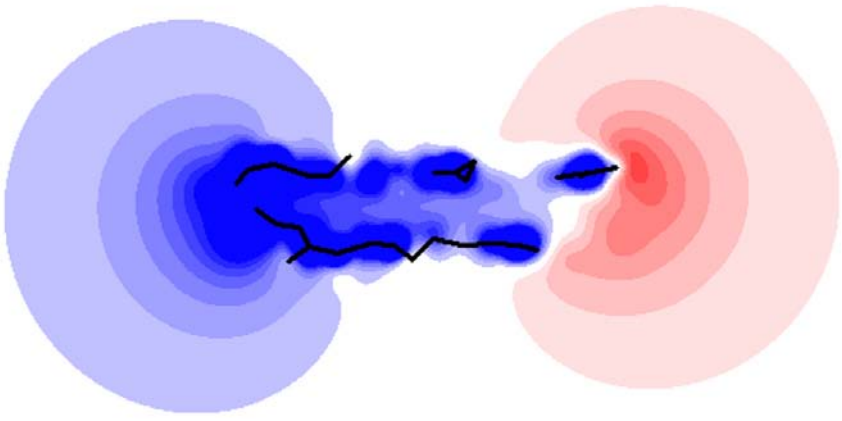


Figure 4.24 – Potentiel total électrostatique d'une hélice idéalisée. Les unités peptidiques ont étés neutralisées préalablement et le calcul a été effectué dans le plan formé par l'axe de l'hélice ainsi que de l'atome d'oxygène de la première unité peptidique. Ce plan correspond alors à une coupe à travers l'hélice.

nombre UP	quadrupôle [eÅ^2]			
	θ_{zz}	θ_{yy}	θ_{xx}	η
1	1,4	-1,0	-0,5	0,36
2	3,0	-2,8	-0,2	0,87
3	4,8	-3,8	-1,0	0,58
4	6,3	-3,5	-2,8	0,11
5	7,3	-4,3	-3,0	0,18
6	8,9	-5,5	-3,4	0,24
7	11,2	-6,8	-4,4	0,21
8	12,5	-6,9	-5,6	0,10
9	13,1	-7,0	-6,1	0,07
10	15,1	-8,8	-6,3	0,17
14	21,5	-12,3	-9,2	0,14

Table 4.2 – Moments quadrupolaires des hélices polyalanine de différentes longueurs.

propres du quadrupôle sont représentés figure 4.25. Le vecteur propre de θ_{zz} est orienté dans le sens de l'axe de l'hélice et il presque parallèle au dipôle. Les deux autres vecteurs se trouvent dans le plan perpendiculaire à l'axe et pointent vers l'extérieur. L'orientation des vecteurs propres reflète clairement la symétrie de l'hélice.
Le dipôle d'une hélice idéale de 14 unités peptidiques est de 15.1 eÅ, tandis que θ_{zz} est de 21.5 eÅ^2. Le potentiel électrostatique de l'hélice a clairement un caractère dipolaire, le moment quadrupolaire est alors négligeable.

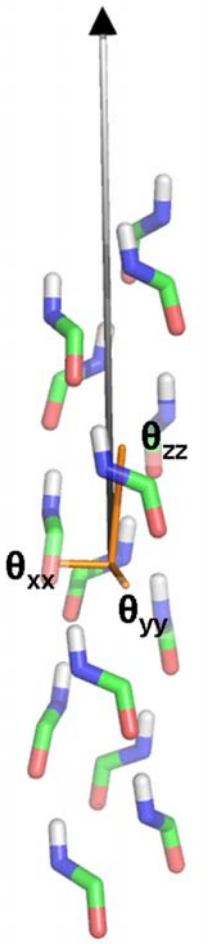


Figure 4.25 – Vecteurs propres du moment quadrupolaire (lignes oranges) et dipôle (flèche noire) d'une hélice composée de 14 unités peptidiques.

Chapitre 5

Moments électriques du feuillet β

5.1 Le moment dipolaire du feuillet β

Tandis que le concept du dipôle de l'hélice α est largement accepté, on peut noter que moins d'attention a été prêtée aux propriétés électrostatiques des feuillets β. Par conséquent, on trouve peu de références sur ce sujet.
Dans un premier temps, on peut citer les travaux de Hol *et al.* [26]. Se basant sur le modèle de l'unité peptidique décrit dans le chapitre précédent (charges ponctuelles de Mulliken, section 4.2), il a été proposé que les feuillets β parallèles portent un moment dipolaire le long du feuillet, pointant de l'extrémité C-terminale au N-terminal. Par contre, le dipôle du feuillet antiparallèle serait négligeable puisque les dipôles des unités peptidiques s'annulent mutuellement (figures 5.2 et 5.1) [26].
Or, les résultats de calculs *ab initio*, effectués par VanDuijnen et collaborateurs, ont remis en question les estimations de Hol *et al.*. Ils ont pu montrer que les moments dipolaires des feuillets parallèles et antiparallèles de tetraglycine sont similaires : il n'y aurait donc pas de différence significative entre la distribution de charge des deux types de feuillet [45].
A part les calculs de vanDuijnen [45] et de Hol [26], les moments électrostatiques (dipolaires ou autres) du feuillet β n'ont pas été analysés à notre connaissance. Les deux études menaient à des résultats différents, il est alors intéressant d'utiliser la librairie ELMAM afin de caractériser les moments électrostatiques des feuillets β. L'idée initiale était de rechercher une relation entre le moment dipolaire et le nombre d'unités peptidiques, comme elle a été trouvée pour le macrodipôle de l'hélice α.
Soulignons que les feuillets β sont des éléments de structure qui sont plus difficile à décrire que les hélices. Dans un premier temps, les brins peuvent être arrangés de manière parallèle ou antiparallèle. De plus, un feuillet peut

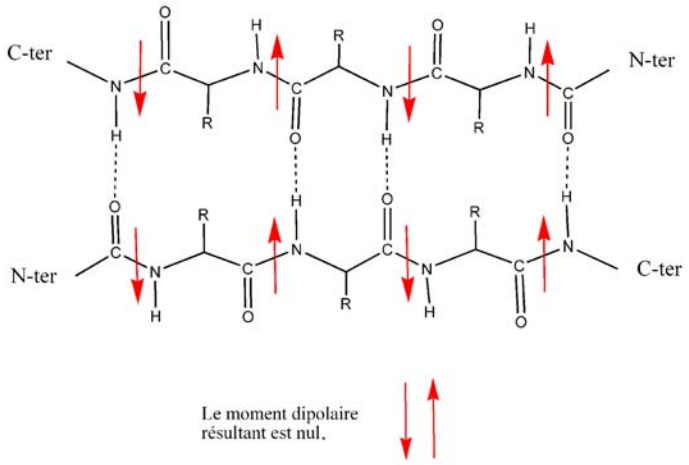


Figure 5.1 – Annulation du moment dipolaire du feuillet β antiparallèle. Les unités peptidiques sont arrangées de façon à ce que leur dipôles soient opposés, et par conséquent, s'annulent.

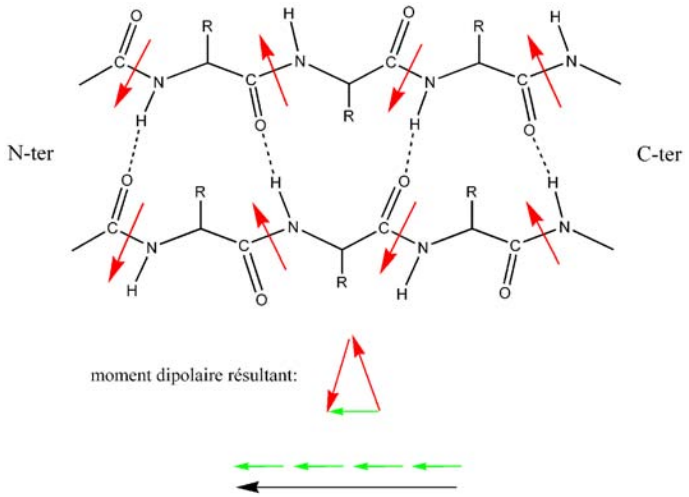


Figure 5.2 – Origine du moment dipolaire du feuillet β parallèle selon [26]. Les unités peptidiques sont légèrement inclinées par rapport à la direction des brins afin de former des liaisons hydrogènes. La somme vectorielle des dipôles générés par des unités peptidiques voisines a alors une composante non nulle le long du feuillet. Hol *et al.* ont proposé que la composante le long du brin correspond à 1/3 du dipôle d'une unité peptidique (=0.23 eÅ). Ces dipôles sont espacé de 3.5 Å, le feuillet entier peut alors être considéré comme dipôle ayant une charge de +1/15 électrons (e) à l'extrémité N-terminale et -1/15 e au C-terminal. Par conséquent, le dipôle du feuillet parallèle est 5 fois plus faible que celui d'une hélice de même longueur.

être constitué en même temps de brins arrangés parallèlement et antiparallèlement, une multitude de combinaisons est alors possible. Ensuite, le moment dipolaire du feuillet dépend non seulement du nombre de brins le composant mais aussi de la longueur de ces brins, donc du nombre d'unités peptidiques qui le forme. Par exemple, pour un feuillet antiparallèle constitué de deux brins, longs de 3 unités peptidiques chacun, les dipôles ne s'annulent pas tous (figure 5.3). Notons que les brins d'un feuillet peuvent ne pas tous avoir exactement la même longueur, ce qui complique davantage la description. Finalement, les feuillets que l'on trouve dans les protéines présentent généra-

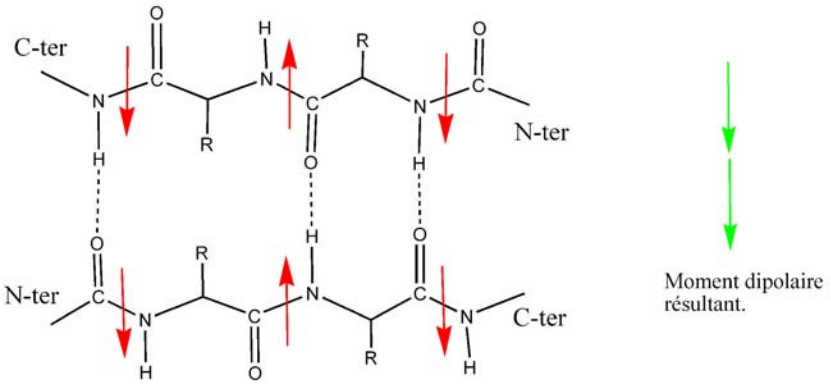


Figure 5.3 – Le dipôle effectif dépend du nombre d'unités peptidiques. Les 4 unités peptidiques des deux premières colonnes s'annulent tandis que les moments dipolaires des deux unités restantes s'accumulent pour former un dipôle perpendiculaire au feuillet.

lement une torsion et ne sont donc pas plans. Par exemple, ils peuvent être arrangés en tonneaux (comme dans le repliement du tonneau $(\alpha/\beta)_8$, figure 4.10) ou en hélice de bateau (comme la protéine DFPAse, section 7.1).
Pour cette raison, j'ai procédé pas à pas pour la caractérisation du moment dipolaire des feuillets β. Dans un premier temps, des feuillets idéalisés plans sont analysés dans la conformation parallèle et antiparallèle. Comme mentionné ci-dessus, le dipôle dépend apparemment de la parité du feuillet (donc du nombre de brins ainsi que du nombre d'unités peptidiques par brin). Les moments dipolaires seront alors calculés pour des feuillets de différentes tailles et parités. Ensuite, l'effet de la torsion sur le dipôle sera analysé. Finalement, la caractérisation trouve une application dans le calcul de moments quadrupolaires de quelques repliements protéiques.

La création d'un feuillet β s'est avérée plus compliquée que la construction de l'hélice α idéalisée. Il y a un grand nombre de logiciels permettant de créer un brin singulier (comme par exemple Pymol [46], Moleman [39]),

mais pas la création d'un feuillet. Cela est probablement dû à la discontinuité de ce motif : tandis que l'hélice est formée par une succession continue d'acides aminés, le feuillet est constitué de différents brins. J'ai adopté alors le protocole suivant pour la création des feuillets :

- Création d'un brin polyalanine avec des angles de Ramachandran d'un feuillet étendu (ϕ=180°, ψ=180°), parallèle voilé ($\phi = -120°$, $\psi = 113°$) ou antiparallèle voilé ($\phi = -140°$, $\psi = 135°$).
- Le brin a été copié et il a été positionné à côté du précédent, parallèle ou antiparallèle. Ainsi, un feuillet "approximatif" de 5 brins, chacun long de 20 AA, a été créé.
- Notons que ce sont les liaisons hydrogènes H...O entre les brins qui sont typiques du repliement feuillet β. Avec le logiciel MOE [47], l'énergie potentielle du système moléculaire (le feuillet de 5 brins) a été minimisée en imposant une distance H...O de 2.0 Å entre les unités peptidiques [1]. Cette distance correspond à la longueur moyenne des liaisons hydrogènes dans les feuillets [49].
- Les angles et les longueurs de liaisons des acides aminés ont été optimisés, se basant sur les valeurs standards [50].

Notons que les angles de Ramachandran du modèle final peuvent être différents de la conformation initiale. Ce protocole constitue alors un compromis entre la définition d'un feuillet en terme d'angles de Ramachandran et en terme de liaisons hydrogène. Pour les analyses suivantes, le protocole ci-dessus a été appliqué et la notation suivante a été utilisée : un fragment (n x m) est constitué de n brins où chaque brin contient m unités peptidiques. Comme il a été mentionné ci-dessus, le moment dipolaire est apparemment sensible au nombre d'unités peptidiques par brin. Pour cette raison, le dipôle est calculé pour différentes tailles (n x m) du feuillet.

5.1.1 Feuillet parallèle idéal

Le feuillet idéal utilisé pour l'analyse est représenté figure 5.4. Le module du moment dipolaire a été calculé en fonction du nombre d'unités peptidiques, le fragment a été neutralisé préalablement.
La figure 5.5 montre les moments dipolaires calculés à partir d'un feuillet de 5 brins de longueur m. Si la longueur m des brins, c'est-à-dire le nombre d'unités peptidiques par brin, est paire, le dipôle est parallèle au feuillet, et il s'étend de l'extrémité N-terminale vers la C-terminale (ligne noire). Par ailleurs, si m est impair, le dipôle a également une composante perpendicu-

1. La minimisation d'énergie est basée sur les paramètres du champ de force MMFF94 [48].

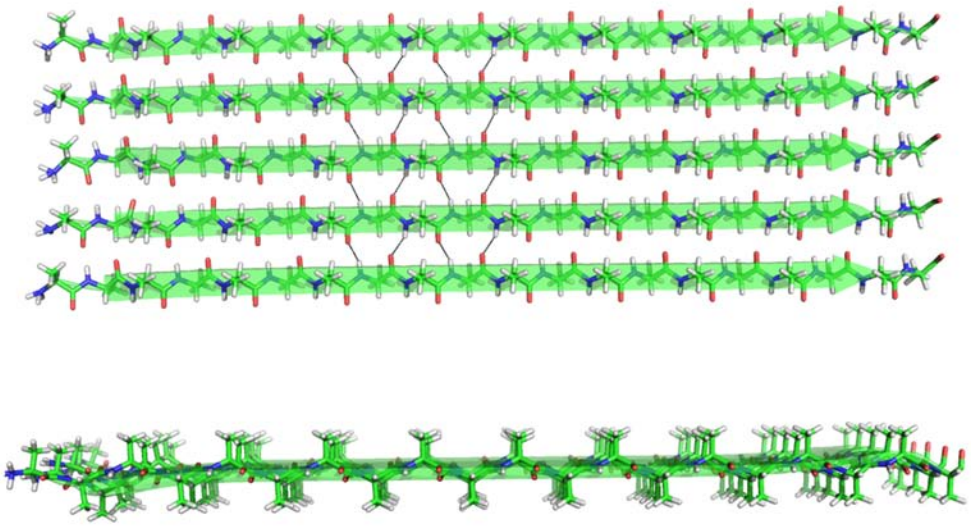


Figure 5.4 – Représentation du feuillet β parallèle idéal utilisé pour les calculs des moments électrostatiques. En haut : vue de haut. En bas : vue latérale. Quelques liaisons hydrogène entre les brins sont représentées en lignes noires.

laire aux brins (lignes colorées figure 5.5).
Le module du dipôle du feuillet (5 x 11), contenant 55 unités peptidiques, est de 52.4 D. Un tel moment dipolaire peut être trouvé dans une hélice constituée de 11 UP seulement (calculé à partir de l'équation établie pour les dipôles des hélices). Par conséquent, l'arrangement des UP dans un feuillet est donc moins efficace que dans l'hélice pour créer un moment dipolaire.
Les résultats sont similaires pour les feuillets parallèles de taille (4 x m), (3 x m) et (2 x m).

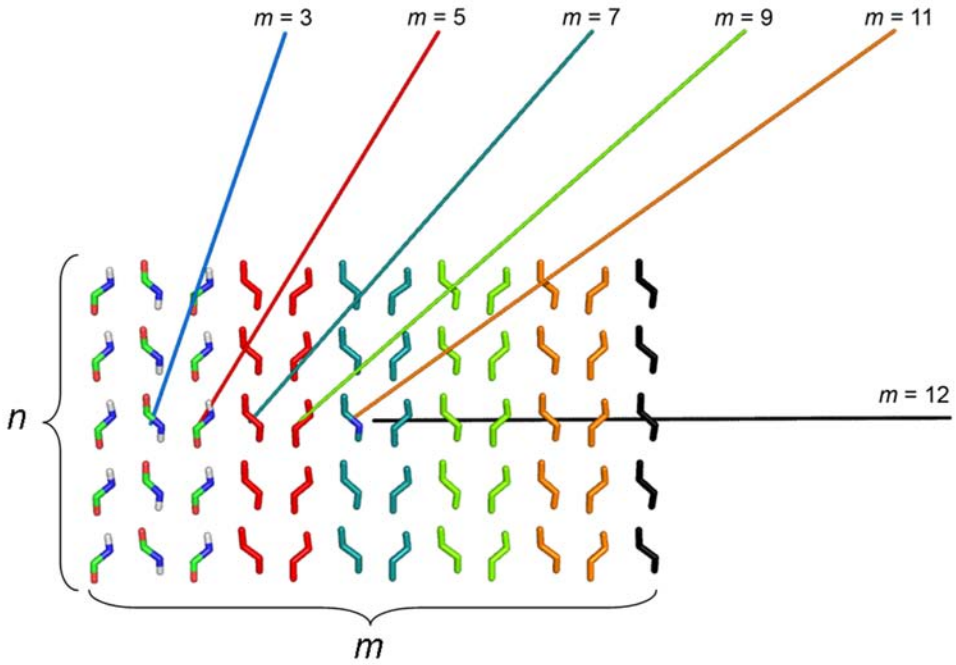


Figure 5.5 – Dipôles d'un feuillet parallèle (5 x m) constitué de 5 brins de longueur m. Seules les unités peptidiques sont représentées. Chaque dipôle est représenté par une ligne colorée et prend son origine au centre de masse du fragment utilisé pour le calcul. Bleu : dipôle calculé à partir du fragment (5x3) constitué des unités peptidiques colorées. Rouge : dipôle calculé à partir du fragment (5x5) constitué des unités peptidiques colorées et rouges, etc. Noir : dipôle du fragment (5x12). C'est le seul dipôle de parité paire qui est représenté puisque les autres vecteurs se superposent. Leur module est plus faible et ils ont la même orientation que le dipôle du fragment (5x12).

5.1.2 Feuillet antiparallèle idéal

Le feuillet antiparallèle (figure 5.6) a été créé selon le protocole décrit section 5.1.

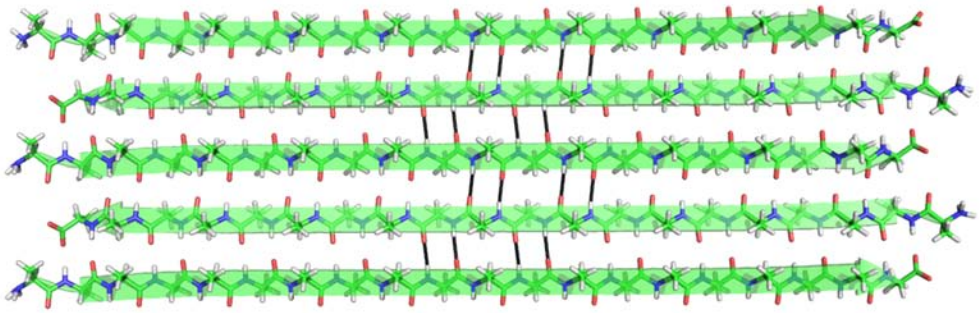


Figure 5.6 – Représentation du feuillet antiparallèle utilisé pour l'analyse des moments dipolaires. Quelques liaisons hydrogène entre les brins sont représentées en lignes noires.

La figure 5.7 montre les dipôles du feuillet de taille (4 x m). De manière intéressante, le dipôle est proche de zéro si le nombre d'unités peptidiques par brin est pair (n et m pair). Par ailleurs, si la parité des UP est impaire (n pair, m impair), le dipôle est orienté perpendiculairement au feuillet. Notons qu'il est plus faible que le moment dipolaire d'un feuillet parallèle de même taille : des feuillets parallèle et antiparallèle de taille (4 x 9) génèrent un dipôle de 36,3 D et 21.4 D respectivement.
Le feuillet (2 x m) a le même comportement, par contre, les dipôles des feuillets (3 x m) et (5 x m) ressemblent à ceux d'un feuillet parallèle. Ainsi, les moments dipolaires sont parallèles aux brins pour m impair et ils sont diagonaux pour m pair.
Dans un premier temps, on peut constater que les dipôles des feuillets β parallèles et antiparallèles diffèrent soit en magnitude (pour n impair et m quelconque), soit en magnitude et orientation (n pair et m quelconque). La parité du fragment utilisé a alors une grande influence sur l'orientation et le module du moment dipolaire.

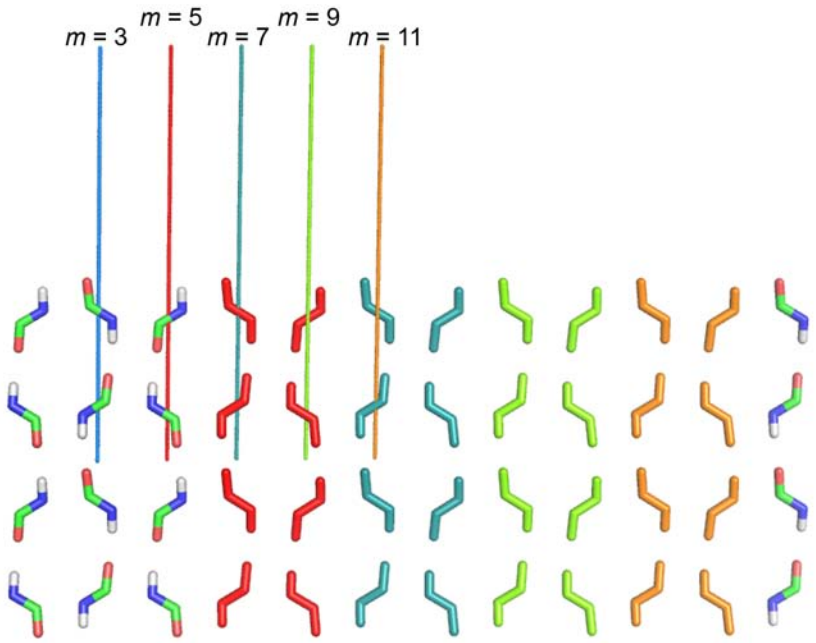


Figure 5.7 – Moments dipolaires d'un feuillet antiparallèle, constitué de 4 brins de longueur m. La signification des couleurs est comme dans la figure 5.5. Les dipôles calculés à partir d'un nombre pair d'unités peptidiques par brin (m pair) sont proches de zéro et ne sont pas représentés.

Il se pose la question de savoir pourquoi les feuillets ont un comportement différent pour un nombre pair de brins (n pair). Rappelons que le dipôle généré par l'unité peptidique est incliné, il s'étend approximativement de l'atome d'oxygène vers l'atome d'hydrogène. Or, les unités peptidiques sont arrangées de manières différentes dans les feuillets parallèles et antiparallèles (figures 5.8 et 5.9). Dans le feuillet β parallèle, les UP ont la même orientation dans chaque brin tandis que dans le feuillet antiparallèle, l'orientation des UP alterne d'un brin à l'autre. Cette alternance d'orientation des dipôles de l'UP explique pourquoi le dipôle total d'un feuillet antiparallèle de parité paire (n et m pairs) s'annule (figure 5.9).

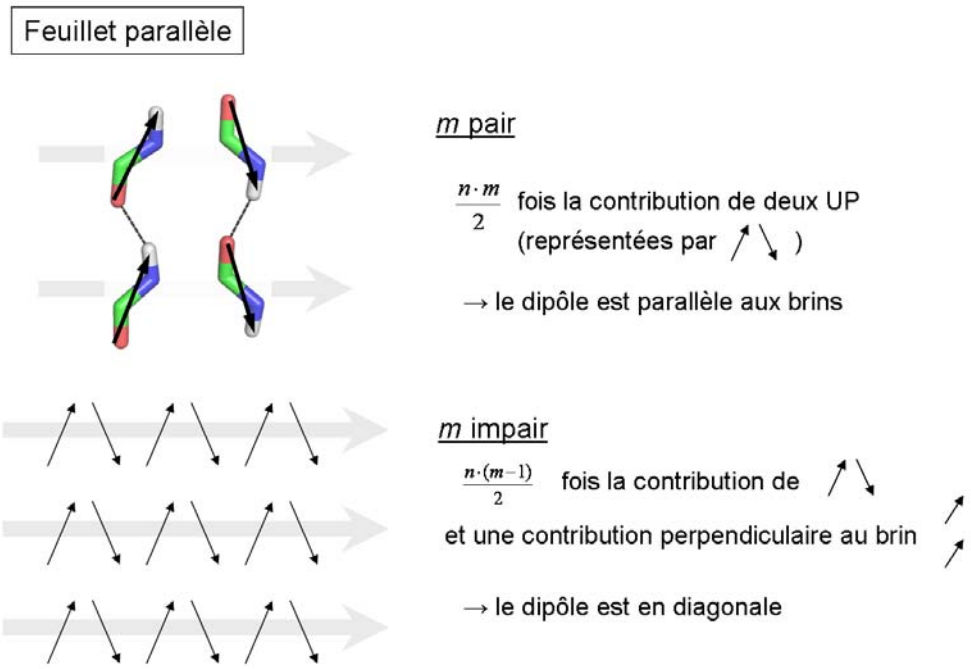


Figure 5.8 – Schema d'un feuillet parallèle et des dipôles des unités peptidiques. Les flèches grises horizontales représentent le sens des brins.

Les estimations du moment dipolaire des feuillets β de Hol *et al.* étaient basées sur un dipôle de l'unité peptidique dont l'orientation est perpendiculaire à la liaison N-C. Or, des calculs *ab initio* [45] et des mesures expérimentales [32] sur la molécule formamide suggèrent une inclinaison du moment dipolaire, équivalente à celle observée lorsque le dipôle est calculé avec les charges de la librairie ELMAM. En effet, l'angle d'inclinaison par rapport à la liaison C-N est de 39,6°, 35° et 46° pour les mesures sur formamide, le calcul théorique et le dipôle ELMAM, respectivement. Les moments dipolaires calculés

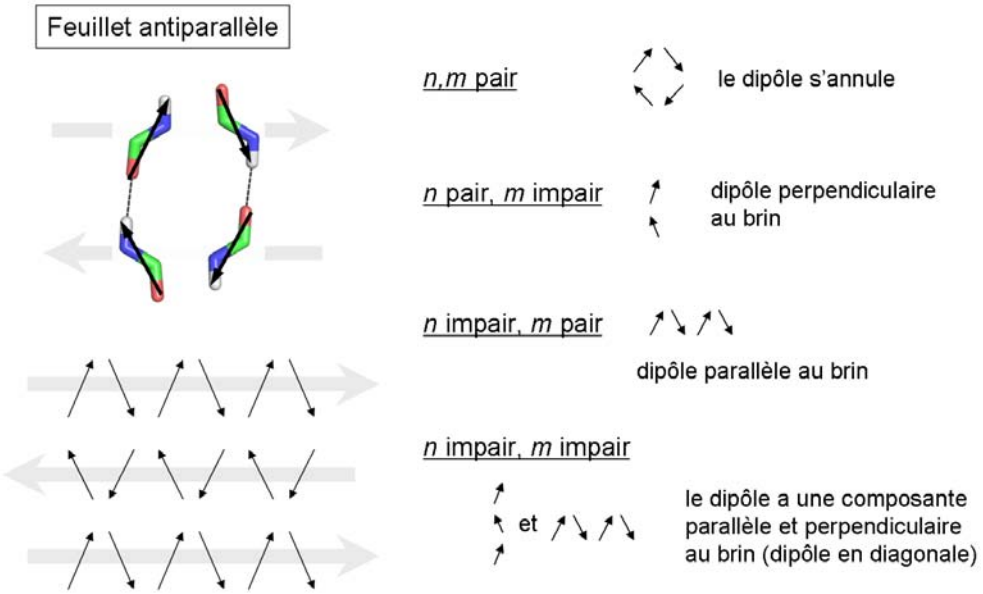


Figure 5.9 – Schema d'un feuillet antiparallèle et des dipôles des unités peptidiques.

à partir d'ELMAM sont alors cohérents avec ces données expérimentales et avec ces calculs théoriques.

5.1.3 Feuillets β voilés

Feuillet parallèle

Le feuillet parallèle voilé (figure 5.10) a été créé avec des brins dont les angles de Ramachandran ϕ et ψ sont de -120° et 113° respectivement. Comme pour les feuillets idéaux, les dipôles ont été calculés pour différentes tailles (n x m) du feuillet. Le fragment a été neutralisé préalablement.

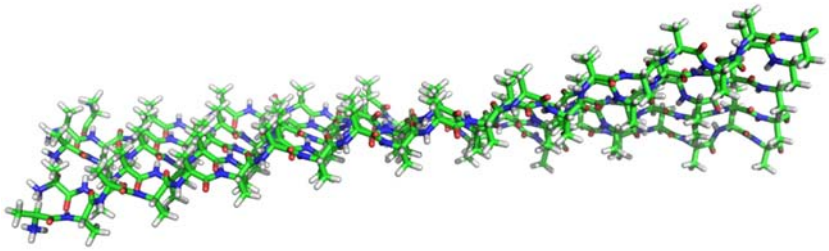


Figure 5.10 – Vue de côté d'un feuillet parallèle voilé qui présente une torsion entraînant une déviation par rapport au feuillet plan.

De manière intéressante, on constate des différences par rapport aux moments dipolaires du feuillet idéal. D'abord, pour les feuillets dont le nombre d'UP par brin est pair (n quelconque, m pair), les dipôles changent d'orientation et de magnitude. Tandis qu'ils étaient parallèles aux brins dans le feuillet idéal, ils sont maintenant orientés au dessus du feuillet et par ailleurs, leur magnitude est plus faible. Par exemple, le dipôle généré par un feuillet idéal de taille (5 x 10) est de 40.5 D, tandis qu'il est de 9.1 D dans le feuillet voilé. Par conséquent, la composante le long des brins pour le dipôle des feuillets de parité impaire (n quelconque, m impair) est également plus faible et approximativement, les moments dipolaires sont orientés perpendiculairement aux brins. Les déviations par rapport au feuillet idéal sont clairement dues à la torsion du feuillet.

Feuillet antiparallèle

Le feuillet antiparallèle enroulé a été créé avec des brins dont les angles de Ramachandran ϕ et ψ sont de -140° et 135° respectivement. Pour différentes tailles (n x m) du feuillet, le moment dipolaire a été calculé.
Cette fois-ci, il n'y a pas de différences nettes entre les dipôles des feuillets idéaux ou enroulés. La magnitude et l'orientation est similaire dans les deux cas, pour toutes les parités. En effet, les angles dièdres du feuillet antiparallèle sont plus proches de ceux d'un feuillet étendu ($\phi = \psi$ =180°), la torsion du feuillet est alors moins élevée.

5.1.4 Moments dipolaires de repliements communs

Les feuillets β peuvent adopter une multitude d'arrangements dans les protéines. Afin d'appliquer les résultats des sections précédentes, les moments dipolaires de quelques repliements communs seront analysés.

Tonneau $(\alpha/\beta)_8$

Le repliement du tonneau $(\alpha/\beta)_8$ est tel que les feuillets forment un tonneau au centre de la protéine (figure 4.10).
Pour les protéines aldose réductase, xylanase et beta glucosidase, le moment dipolaire des unités peptidiques formant le tonneau a été calculé. L'assignement des motifs a été effectué avec le logiciel STRIDE [41] et il a été vérifié manuellement avec Pymol [46]. Le fragment a été neutralisé pour le calcul du dipôle. Si un résidu disposait de conformations multiples, la conformation avec l'occupation la plus élevée a été utilisée (en posant l'occupation à 1).

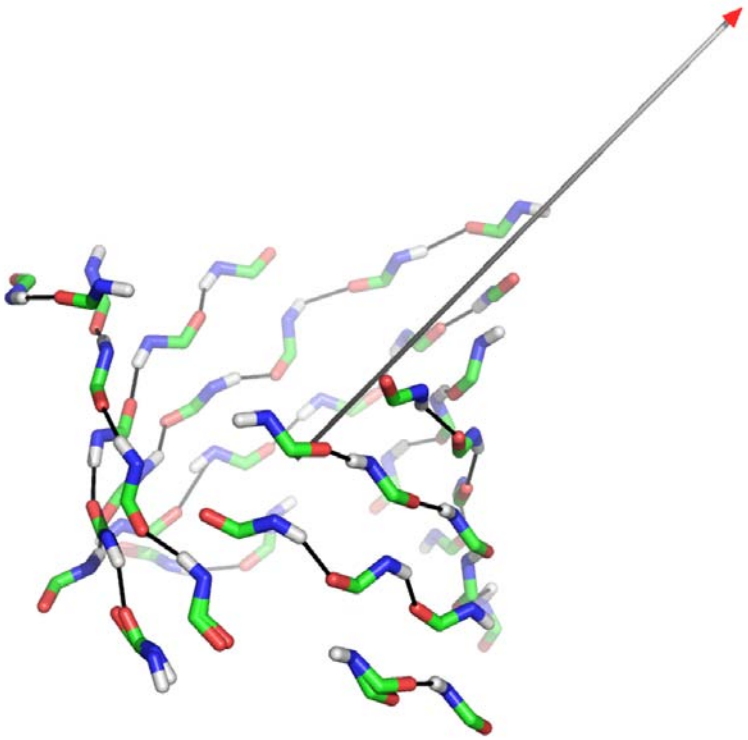


Figure 5.11 – Moment dipolaire généré par les unités peptidiques du tonneau au centre de la protéine aldose réductase. Le dipôle, représenté comme flèche noire, prend son origine au centre de masse du fragment. Les liaisons hydrogène entre les unités peptidiques sont représentées en lignes noires.

La figure 5.11 montre le dipôle du tonneau de feuillets β dans la protéine aldose réductase. Il est de 22.6 D pour 164 atomes, ce qui correspond à 41 unités peptidiques [2]. Le dipôle prend son origine au centre de masse du fragment qui se trouve au milieu du tonneau, et il est orienté en diagonale vers l'extérieur. Notons que le tonneau n'a pas une forme circulaire (figure 5.12) : vu de haut, il a une forme ovale et le dipôle est orienté en direction du petit axe.

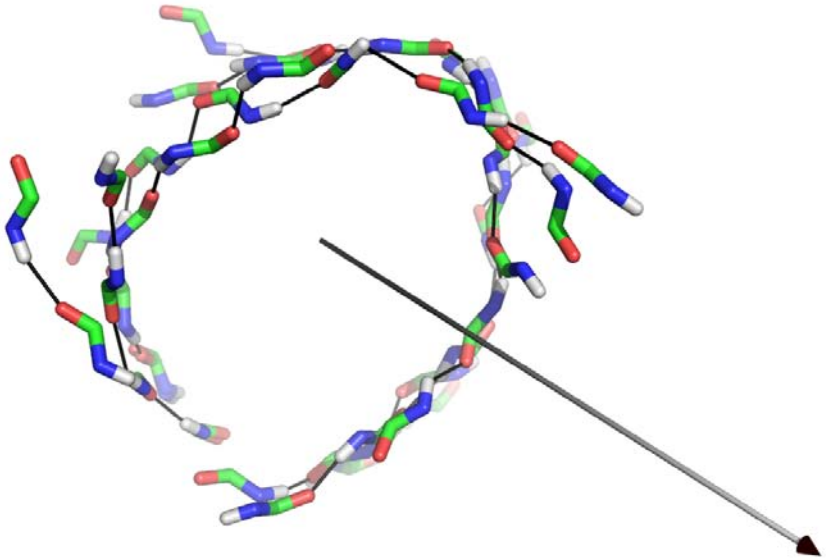


Figure 5.12 – Vue de haut du fragment en forme de tonneau dans la protéine aldose réductase. Le dipôle est représenté comme flèche noire. Notons que le tonneau a une forme ellipsoïdale.

Dans les protéines xylanase et beta-glucosidase, l'orientation du dipôle est comparable à celle de l'aldose reductase. Ils prennent leurs origines au centre du tonneau et s'orientent vers l'extérieur, en diagonale par rapport à l'axe du tonneau. Le tonneau dans la protéine xylanase est formé par 227 atomes et le moment dipolaire est de 46.3 D. Dans la protéine beta glucosidase, 183 atomes forment le tonneau central qui possède un moment dipolaire de 43.9 D. Notons que quelques feuillets contiennent un ou plusieurs résidus proline. S'il n'est pas pris en compte pour le calcul, le dipôle est un peu moins élevé mais l'orientation reste similaire.
Les trois protéines possèdent alors un dipôle d'orientation comparable et d'un même ordre de grandeur. Soulignons que la taille du tonneau et l'arrangement des brins sont différents. Cependant, le nombre d'échantillons est trop

2. Afin de comparer : une hélice longue de 5 UP crée un dipôle de 24 D.

petit pour faire des conclusions générales, donc pour conclure que le moment dipolaire du tonneau dans les repliements $(\alpha/\beta)_8$ soit sensiblement conservé.

β-propeller

Le *β-propeller* est un repliement constitué de feuillets β, antiparallèles en général, qui s'arrangent pour former un motif qui ressemble à une hélice de bateau, avec un tunnel central au milieu. Le site actif se trouve souvent dans la cavité centrale. Un représentant de ce repliement est la protéine DFPase (voir chapitre 7).
Le même protocole de préparation que ci-dessus (section 5.1.4) a été appliqué. Notons que 4 unités peptidiques non complètes (N-C=O) sont incluses dans le fragment du propeller correspondant à des résidus proline.

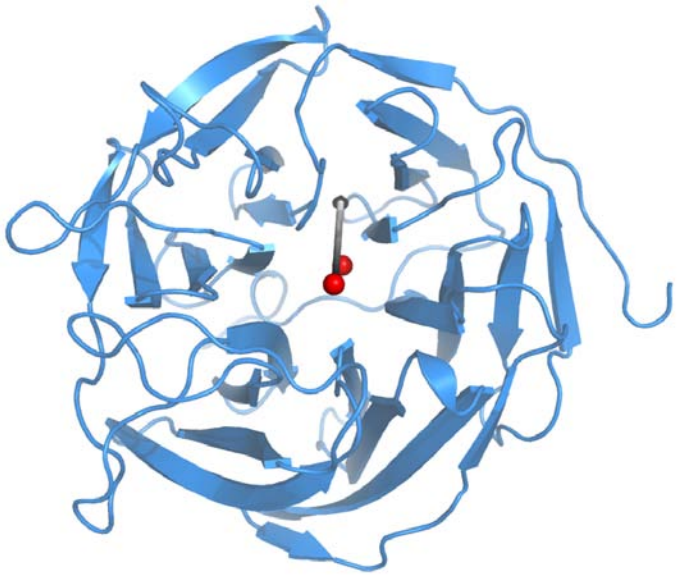


Figure 5.13 – Moment dipolaire (flèche noire) des feuillets de la protéine DFPase. La chaîne principale et les deux ions calcium sont représentés en bleu et en rouge, respectivement. Le dipôle est représenté en eÅ.

Le moment dipolaire total des feuillets formant l'hélice de bateau est représenté figures 5.13 et 5.14. Le dipôle prend son origine au centre de masse du fragment (situé entre deux ions calcium [3]) et il est orienté vers le côté fermé du tunnel central. Notons qu'il est incliné par rapport à l'axe de l'hélice de bateau, il ne reflète donc pas la symétrie structurale du fragment. Le motif

3. Voir chapitre 7 pour plus d'information sur les ions calcium de la protéine DFPase.

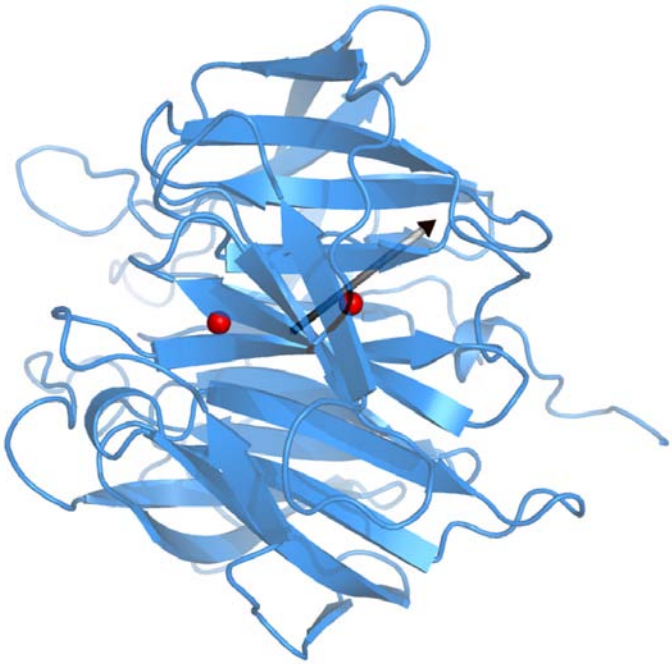


Figure 5.14 – Vue latérale de la protéine DFPase. Le moment dipolaire est représenté comme flèche noire (en eÅ).

est alors symétrique en terme de masse, le centre de masse se trouvant au centre du tunnel, mais pas en terme d'agencement de charges.

5.2 Le quadrupôle du feuillet β

Comme il a été discuté dans la section précédente, le dipôle calculé à partir d'un feuillet β est nettement plus faible que le moment dipolaire d'une hélice de même taille (pour le même nombre d'UP). Rappelons que les unités peptidiques dans le feuillet sont arrangées dans un plan, les unes à côté des autres. Il se pose alors la question si le feuillet n'est pas plus adapté pour créer un moment quadrupolaire plutôt qu'un dipôle.
La figure 5.15 représente le potentiel électrostatique d'un feuillet β parallèle (2x2). On reconnait les lobes de signes opposés aux quatre coins du fragment qui sont typiques du potentiel d'un moment quadrupolaire transversal (figure 4.23). Par conséquent, ce feuillet a un caractère quadrupolaire. Rappelons que le moment dipolaire des feuillets dépend de sa parité, il en est certainement de même pour le quadrupôle. J'ai donc effectué une étude sur le moment quadrupolaire des feuillets β idéaux et voilés pour analyser son évolution en fonction de la parité des nombres n et m. J'ai procédé de la même manière pour tous les feuillets : le quadrupôle est calculé pour un fragment feuillet β de taille (n x m), qui a été préalablement neutralisé.

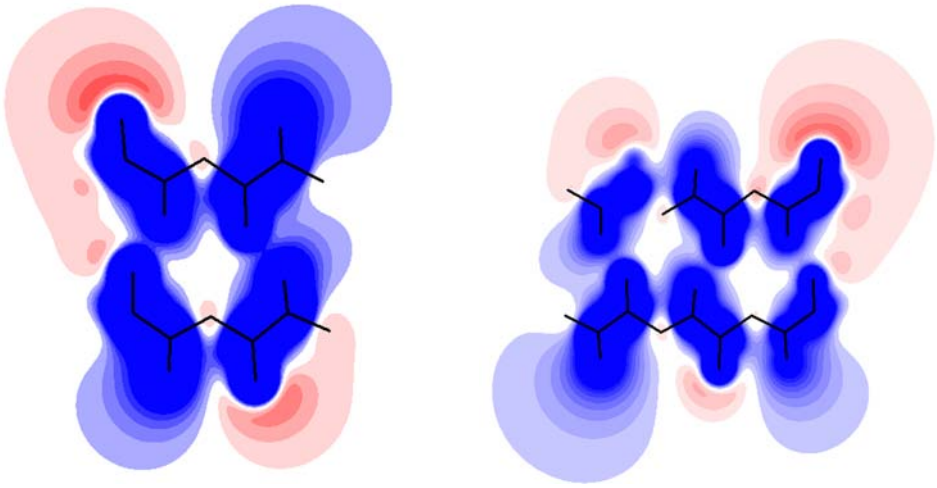


Figure 5.15 – *A gauche* : Potentiel électrostatique total d'un feuillet β parallèle de taille (2x2). Le fragment a été neutralisé préalablement. Les lignes noires correspondent aux liaisons entre les atomes de l'unité peptidique. *A droite* : Potentiel électrostatique total d'un feuillet de taille (2x3). Le potentiel électrostatique a été calculé avec VMoPro en utilisant la densité de charge transférée d'ELMAM

5.2.1 Feuillet parallèle idéal

Le tableau 5.1 récapitule les moments du feuillet (4xm) en fonction de sa parité, les axes principaux des moments quadrupolaires sont représentés figure 5.16. Notons tout d'abord que la valeur de θ_{xx} est très faible comparée

(n x m)	quadrupôle [eÅ^2]				dipôle [eÅ]
	θ_{zz}	θ_{yy}	θ_{xx}	η	
4x2	-24,6	22,8	1,8	0,85	1,3
4x3	-5,3	3,9	1,4	0,46	5,0
4x4	-49,2	45,6	3,6	0,86	2,6
4x5	-7,6	5,1	2,5	0,34	5,6
4x6	-73,8	68,5	5,4	0,85	3,8
4x7	-10,0	6,8	3,2	0,36	6,4
4x8	-98,4	91,3	7,1	0,85	5,1
4x9	-12,5	8,8	3,7	0,41	7,4
4x10	-123,1	114,1	9,0	0,85	6,4

Table 5.1 – Moments dipolaires et quadrupolaires du feuillet (4xm).

aux autres éléments tensoriels, c'est clairement du à la géométrie plane du feuillet. De manière intéressante, θ_{zz} et θ_{yy} ont le même ordre de grandeur qui alterne avec la parité. Les deux valeurs sont grandes pour une longueur paire de brins et elles sont faibles pour m impair. Le paramètre d'asymétrie η alterne également : il est de 0.85 pour m pair, ce qui est proche de la valeur d'un quadrupôle transversal ($\eta \approx 1$), tandis qu'il est compris entre 0.34 et 0.46 pour m impair. Par conséquent, la distribution de charge a le caractère d'un quadrupôle transversal pour m pair. En effet, le potentiel électrique d'un feuillet (2x2) (figure 5.15) est similaire au potentiel d'un quadrupôle transversal parfait (figure 4.23).
La relation entre θ_{zz} et m est linéaire (figure 5.17), le quadrupôle augmente alors avec la taille du feuillet.
Pour les feuillets de m pair, le vecteur propre associé à θ_{zz} pointe vers le coin du feuillet qui correspond au lobe négatif du potentiel électrostatique, tandis que le vecteur propre de θ_{yy} pointe vers le coin du lobe positif (figure 5.16).
Le potentiel électrostatique d'un feuillet parallèle de taille (2x3) est représenté figure 5.15 (à droite). Il ressemble à un dipôle le long de la diagonale du feuillet, ce qui est cohérent avec le fait que le moment quadrupolaire soit faible.
Le comportement des feuillets (2 x m), (3 x m) et (5 x m) est similaire. Ainsi, le feuillet parallèle possède un moment quadrupolaire et un moment

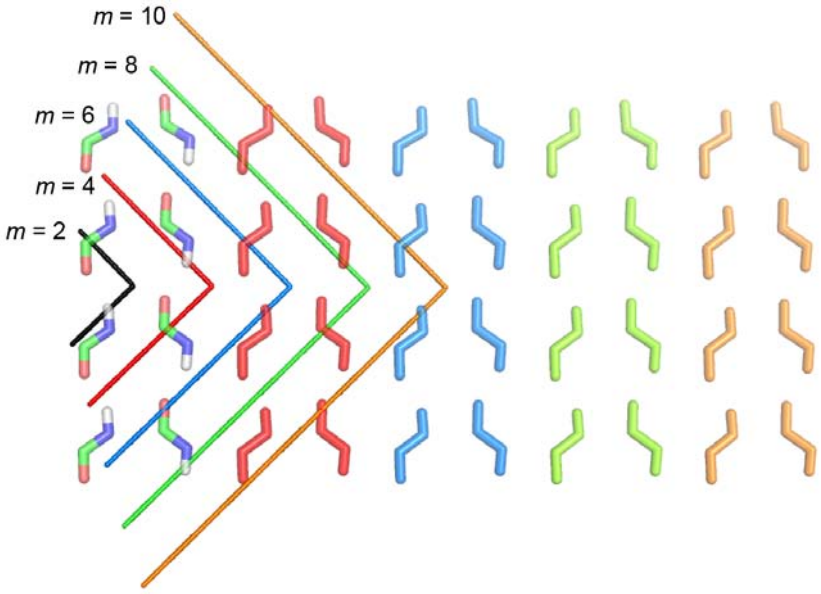


FIGURE 5.16 – Unités peptidiques du feuillet parallèle (4xm) et axes principaux θ_{yy} et θ_{zz} du moment quadrupolaire. Chaque quadrupôle est représenté par une ligne colorée et prend son origine au centre de masse du fragment utilisé pour le calcul. Noire : quadrupôle calculé à partir du fragment (4x2) constitué des unités peptidiques colorées. Rouge : quadrupôle calculé à partir du fragment (4x4) constitué des unités peptidiques colorées en rouges, etc. Les moments quadrupolaires des brins de parité impaire ainsi que l'axe principal correspondant à θ_{xx} ne sont pas représentés puisqu'ils sont négligeables.

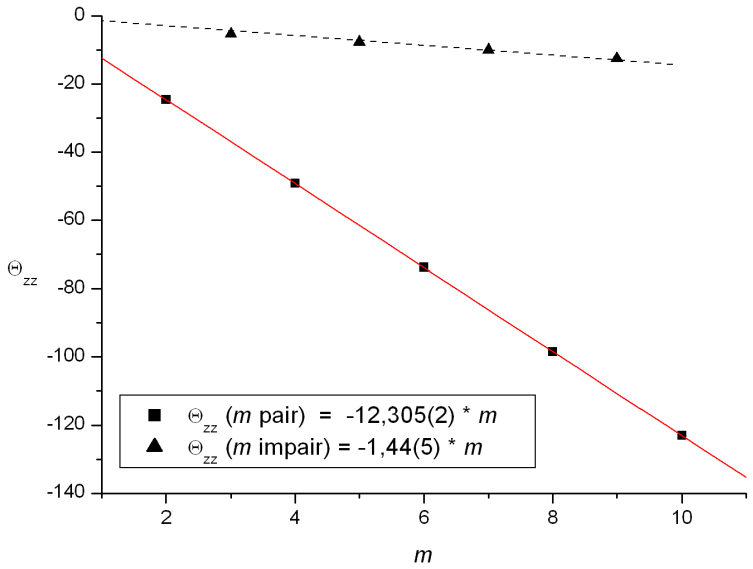


FIGURE 5.17 – θ_{zz} contre la parité m du feuillet, pour m pair et impair.

dipolaire si la longueur des brins est paire. Pour des brins de longueur impaire, le moment quadrupolaire est plutôt faible et son moment dipolaire devient plus important.

5.2.2 Feuillet antiparallèle idéal

Le feuillet antiparallèle idéal a le même comportement que le feuillet parallèle pour un nombre impair de brins (n impair) : il existe un moment quadrupolaire élevé pour m pair et il est négligeable pour m impair.
Cependant, la situation change pour n pair. Le moment quadrupolaire ainsi que la valeur de η d'un feuillet de taille (4 x m) sont récapitulés tableau 5.2. Les axes principaux du quadrupôle sont représentés figure 5.18.

(n x m)	quadrupôle [eÅ^2]				dipôle [eÅ]
	θ_{zz}	θ_{yy}	θ_{xx}	η	
4x2	-16,9	15,8	2,0	0,82	0,0
4x3	-12,7	10,4	1,6	0,69	4,5
4x4	-35,7	31,6	4,1	0,77	0,0
4x5	-20,5	16,4	4,1	0,60	4,5
4x6	-53,6	47,4	6,2	0,77	0,1
4x7	-28,9	22,6	6,2	0,57	4,5
4x8	-71,5	63,2	8,3	0,77	0,1
4x9	-37,2	28,9	8,3	0,55	4,5
4x10	-89,3	78,9	10,4	0,77	0,1

Table 5.2 – Moments dipolaires et quadrupolaires du feuillet (4xm).

Comme pour le feuillet parallèle, la valeur de θ_{xx} est plus faible que les autres éléments du tenseur quadrupôle. Par ailleurs, θ_{zz} et θ_{yy} sont du même ordre de grandeur. On observe une faible alternance de θ_{zz} et θ_{yy} en fonction de la parité, elles sont plus élevés pour m pair que pour m impair. Cependant, la différence n'est pas assez marquante pour dire que le quadrupôle soit négligeable pour m impair. Le paramètre η alterne aussi légèrement, mais moins que dans le cas du feuillet parallèle, η se situant entre 0.6 et 0.8. La direction des vecteurs propres est similaire pour toutes les parités de m.
La figure 5.19 représente le potentiel électrostatique total de deux feuillets antiparallèles. Le potentiel du feuillet de taille (2x2) ressemble à celui d'un quadrupôle transversal tandis que le potentiel du feuillet (2x3) a un fort caractère dipolaire à l'extrémité représentée à droite sur la figure 5.19.
Notons que les éléments tensoriels des feuillets antiparallèles sont moins éle-

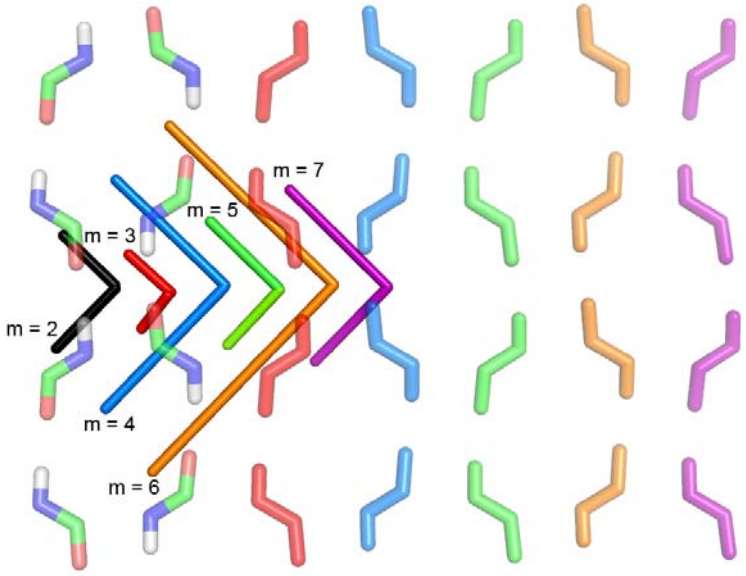


Figure 5.18 – Axes principaux des moments quadrupolaires d'un feuillet antiparallèle de taille (4 x m). Noir : quadrupôle du fragment (4x2). Rouge : quadrupôle du fragment (4x3), etc.

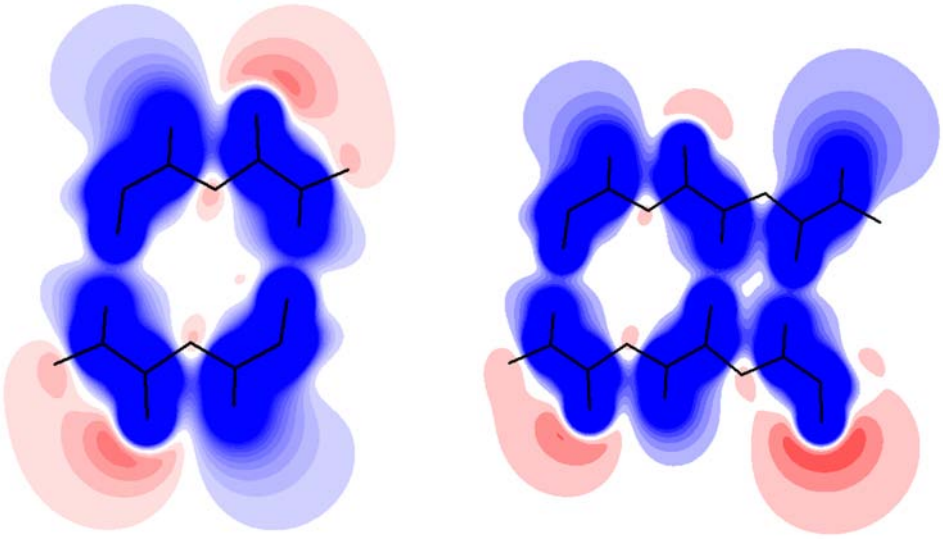


Figure 5.19 – Potentiel électrostatique total calculé à partir des feuillets antiparallèles de taille (2x2) (à gauche) et (2x3) (à droite).

vés que ceux d'un feuillet parallèle de même taille. Par exemple, θ_{zz} vaut respectivement -35.7 eÅ^2 et -49.2 eÅ^2 pour un feuillet de taille (4 x 4).

5.2.3 Feuillets β voilés

Le feuillet parallèle voilé a le même comportement que le feuillet parallèle idéal. Les axes principaux pointent dans les mêmes directions et les valeurs des éléments tensoriels sont comparables. Rappelons que le moment dipolaire des feuillets de parité paire devient négligeable dans le feuillet voilé (section 5.1.3). Par conséquent, le feuillet parallèle possède soit un moment dipolaire (m impair), soit un moment quadrupolaire (m pair).
Entre le feuillet antiparallèle voilé et idéal, on n'observe pas de différences marquantes. Les éléments tensoriels n'ont pas tout à fait les mêmes valeurs, mais les proportions restent conservées. La direction des vecteurs propres est également similaire.
Pour conclure, le feuillet β possède un moment quadrupolaire qui peut être plus important que son dipôle, se basant sur l'analyse du potentiel électrostatique. A notre connaissance, personne n'a associé jusqu'à maintenant des quadrupôles aux motifs de structure secondaire. Afin d'appliquer ces résultats, les moments quadrupolaires de quelques repliements, les mêmes que ceux utilisés pour le dipôle, sont calculés.

5.2.4 Moments quadrupolaires de repliements communs

La préparation des structures est identique à celle décrite section 5.1.4.

Tonneau $(\alpha/\beta)_8$

Les éléments tensoriels du moment quadrupolaire des fragments en forme de tonneau sont résumés tableau 5.3.

protéine	quadrupôle [eÅ^2]				dipôle [eÅ]	nombre d'atomes
	θ_{zz}	θ_{yy}	θ_{xx}	η		
Aldose réductase	-37,1	34,6	2,5	0,87	4,7	164
Beta glucosidase	-51,5	42,8	8,7	0,66	9,1	183
Xylanase	-105,5	75,1	30,4	0,42	9,6	227

Table 5.3 – Moments dipolaires et quadrupolaires de repliements en tonneau.

Dans les trois protéines, la valeur de θ_{xx} est plus faible que les autres composantes. Les axes principaux du moment quadrupolaire des protéines aldose

réductase et xylanase sont représentés figure 5.20. Bien que la direction des axes soit comparable, leurs valeurs propres associées (reflétées par la longueur du vecteur) ne le sont pas. Notons que l'orientation et la valeur du quadrupôle est sensible à la selection du fragment. Une unité peptidique de plus ou de moins influence la direction des axes principaux et les élements tensoriels.

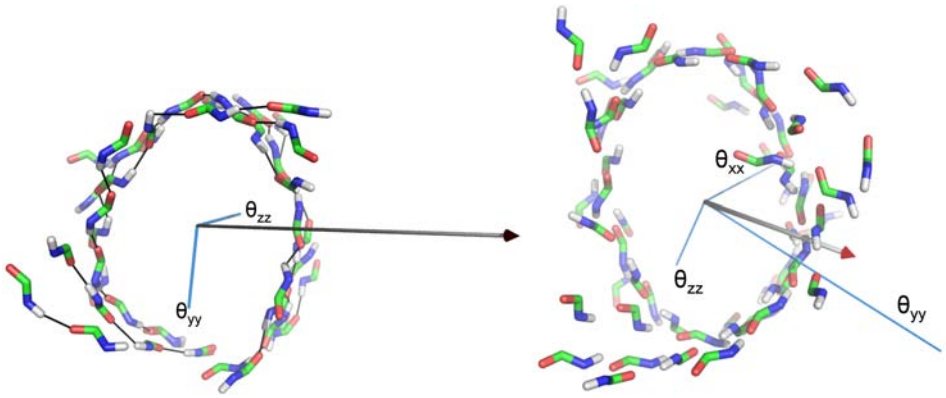


Figure 5.20 – *Gauche* : axes principaux (lignes bleues) et moment dipolaire (flèche noire) du tonneau dans l'aldose reductase. Le vecteur propre de θ_{xx} n'est pas représenté car il est très court. *Droite* : axes principaux et dipôle du tonneau de la protéine xylanase, avec même code de couleurs.

β-propeller

Les axes principaux du quadrupôle, calculés à partir des feuillets de la protéine DFPase, sont représentés figure 5.21. Les valeurs propres sont $\theta_{zz} = 242$ eÅ^2, $\theta_{yy} = -177$ eÅ^2, $\theta_{xx} = -65$ eÅ^2 et $\eta = 0.46$.
Le vecteur propre de θ_{xx} est proche de l'axe du β-propeller. Les deux autres vecteurs propres forment un plan légèrement incliné par rapport à une coupe à travers les feuillets. Notons également que θ_{xx}, la valeur propre la moins élevée, est associé à un vecteur propre qui suit une dimension spatiale du propeller qui est la moins grande. Sa profondeur (≈20Å) est en effet moins élevée que son diamètre (≈40 Å). D'un autre côté, les vecteurs propres associés à θ_{zz} et θ_{yy} suivent le diamètre du propeller et leurs valeurs propres sont plus élevées. Le moment quadrupolaire reflète alors, mieux que le dipôle, la symétrie du fragment.

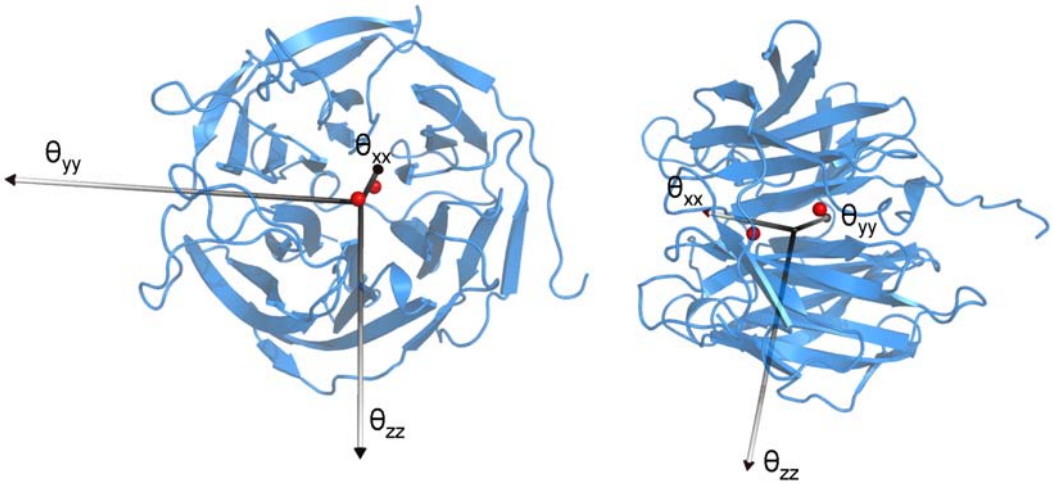


Figure 5.21 – *Gauche* : axes principaux (flèches grises) du quadrupôle des feuillets qui forment l'hélice de bateau. Ils prennent leur origine au centre de masse du fragment. *Droite* : Vue de côté du repliement.

5.3 Conclusion

Les moments électrostatiques créés par les motifs de structure secondaire ont été quantifiés en utilisant les paramètres multipolaires de la bibliothèque ELMAM. Dans un premier temps, le dipôle créé par les hélices dans plusieurs structures protéiques a été analysé et une équation linéaire a été établie, qui met en relation le nombre d'unités peptidiques et le module du moment dipolaire. Ainsi, le dipôle d'une hélice quelconque peut être estimé uniquement sur la base de sa longueur. Par ailleurs, l'influence du résidu proline sur le dipôle a été étudiée. En fonction de sa position dans l'hélice, le module du dipôle peut être significativement atténué ou renforcé.
Les moments dipolaires des feuillets β dépendent du sens (parallèle ou antiparallèle), du nombre de brins, et de la longueur de brins. Bien qu'ils soient plus faibles que le dipôle d'une hélice de même taille, ce dipôle n'est souvent pas négligeable. En particulier, le feuillet antiparallèle possède un dipôle [4] ce qui est contraire à ce qui a été proposé par la littérature [26].
Le moment quadrupolaire des feuillets a été calculé : Il peut être plus important que le moment dipolaire, comme par exemple pour le feuillet parallèle voilé qui possède soit un moment dipolaire (m impair) soit un moment quadrupolaire (m pair).
Enfin, les moments dipolaires et quadrupolaires de deux repliement communs, le tonneau $(\alpha/\beta)_8$ et le β-propeller, ont été analysés. Pour trois protéines différentes, les moments créés par les unités peptidiques du tonneau n'ont pas de rapport apparent avec la symétrie du repliement. Pour la protéine adoptant le repliement du β-propeller, la symétrie du fragment est reflétée approximativement par le moment quadrupolaire des unités peptidiques.

5.4 Perspectives

Les analyses ci-dessus tenaient à caractériser et à quantifier les moments électriques associés aux motifs de structure secondaire. Elles peuvent être considérées comme un travail préliminaire à de futurs études portant sur l'analyse des aspects énergétiques dans les protéines.
Dans un premier temps, on envisage de suivre la caractérisation des moments électriques :

- Il est intéressant d'étudier plus l'influence du résidu proline sur le module du moment dipolaire d'une hélice. Une équation a été établie entre le nombre d'unités peptidiques et le dipôle pour des hélices contenant la

4. Sauf pour les feuillets de parité (n x m), où n et m sont pairs.

proline à la première position. Une étude plus large permettrait d'élaborer une équation plus générale, paramétrisant le module du dipôle en fonction de la position de la proline et de la longueur. De plus, un autre manière de neutralisation peut être testée. Dans la présente étude, l'hélice entière contenant l'acide aminé proline a été neutralisée. Une autre possibilité consiste à neutraliser les acides aminés individuellement.

- Afin d'obtenir des résultats plus généraux, les calculs de moments électrostatiques peuvent être effectués sur un plus grand nombre de structures afin de vérifier si le dipôle est conservé d'une protéine à l'autre. Il serait ensuite intéressant d'étudier s'il y a un lien entre l'orientation d'un ligand ou la localisation du site actif et les moments électrostatiques des motifs à proximité.

- On peut tester si les hélices ou les feuillets possèdent des moments supérieurs non négligeables, comme par exemple un octupôle. En effet, les feuillets idéaux, dans lesquelles les unités peptidiques sont arrangés dans un plan, peuvent avoir un fort caractère quadrupolaire. Cependant, les motifs de feuillets dans les protéines sont souvent plus compliqués, comme des repliements en tonneau, en hélice ou en propeller, qui pourraient engendrer des moments électriques d'ordre supérieur.

- Un autre point intéressant à suivre est l'effet de coopération des liaisons hydrogène dans les hélices. Car les unités peptidiques s'arrangent de manière parallèle et créent un macrodipôle, elles sont polarisées davantage. La charges des atomes est alors plus élevée. Cet effet est difficile à observer dans les protéines mais il existe des polypeptides stables en solution qui ont été cristallisés avec succès [51]. Ils peuvent être utilisés pour un affinement multipolaire afin de révéler la polarisation des atomes de l'unité peptidique sous l'influence du macrodipôle de l'hélice.

Les moments électrostatiques peuvent être utilisés afin d'analyser les aspects énergétiques dans les protéines. En effet, les moments électriques apparaissent dans les termes décrivant l'énergie d'interaction de Buckingham (interaction multipôle-multipôle) entre deux distributions de charge A et B, qui est donnée pour les moments sans trace par [52]

$$\begin{aligned} U &= q^A T q^B + \mu_\alpha^A T_\alpha q^B - q^A T_\alpha \mu_\alpha^B - \\ &\quad - \mu_\alpha^A T_{\alpha\beta} \mu_\alpha^B + \tfrac{1}{3}\theta_{\alpha\beta}^A T_{\alpha\beta} q^B + \tfrac{1}{3} q^A T_{\alpha\beta} \theta_{\alpha\beta}^B \end{aligned} \tag{5.1}$$

Les tenseurs T sont les tenseurs cartésiens d'interaction, $T = \frac{1}{R}$, $T_\alpha = \nabla_\alpha(\frac{1}{R})$, $T_{\alpha\beta} = \nabla_\alpha \nabla_\beta(\frac{1}{R})$, les moments appartenant à la distribution de charge A ont

un exposant A (idem pour la distribution B).
Cette expression permet par exemple de calculer l'énergie d'interaction entre une charge d'un ligand et le moment dipolaire d'un motif de structure secondaire :

$$U_{dipolaire} = -q\frac{\vec{\mu} \cdot \vec{R}}{R^3} \tag{5.2}$$

Si les moments électriques des éléments de structure secondaire d'une protéine sont connus, il est alors possible de calculer l'énergie d'interaction entre ceux-ci ou entre un ligand et une hélice d'une protéine. Il serait notamment intéressant de savoir si l'énergie électrostatique entre les motifs de structure secondaire de repliements communs, comme le tonneau $(\alpha/\beta)_8$ ou le *helix-bundle*, est favorable. Cela pourra fournir un élément important pour le design de nouvelles enzymes. Une autre application peut être le calcul de l'énergie d'interaction entre les motifs de structure secondaire dans un complexe protéine-protéine, où les protéines en interactions sont représentées par la somme des moments de leurs éléments de repliement. De plus, le calcul d'énergie permettra également de savoir quel moment électrique a une contribution prépondérante (à un certain point). L'analyse du potentiel électrostatique permet de qualifier approximativement la nature de la distribution de charge, mais l'énergie d'interaction est capable de quantifier les différentes contributions.

Troisième partie

Topologie des interactions dans les motifs de structure secondaire

Chapitre 6

Topologie des interactions dans les hélices et les feuillets

6.1 Introduction

En cristallographie des petites molécules, il est courant d'analyser la topologie [1] des liaisons hydrogène afin de caractériser en détail cette interaction (section 2.7). Dans les protéines, les liaisons hydrogène jouent un rôle important afin de maintenir la cohésion globale de la macromolécule et elles stabilisent notamment les motifs de structure secondaire. Or, les données de diffraction de cristaux protéiques, souvent de moyenne résolution, ne permettent pas d'accéder à la densité électronique non-sphérique et de faire une analyse topologique précise. L'analyse des liaisons hydrogène dans les structures expérimentales se restreint alors souvent à l'étude de leur géometrie.
Au laboratoire, des outils informatiques pour l'analyse topologique (logiciel NEWPROP [21], software InteGriTy [53], logiciel VMoPro [18]) de la densité électronique sont disponibles. En plus, la bibliothèque ELMAM offre la possibilité de transférer les paramètres multipolaires vers les modèles de structures protéiques. Une analyse topologique de la densité électronique d'une protéine est alors envisageable. L'idée initiale de ce projet consistait à faire une analyse topologique d'un type particulier de liaisons H dans les protéines : les liaisons hydrogène dans les motifs de structure secondaire (hélice et feuillet).

1. Il s'agit d'une analyse topologique dans les deux sens du terme : la géométrie ainsi que la topologie de la densité électronique des liaisons hydrogène sont analysées.

6.2 Description de la liaison hydrogène

La liaison hydrogène a été découverte il y a 100 ans et elle fait toujours l'objet de nombreuses recherches. Néanmoins, le phénomène de cette interaction n'est pas encore compris complètement [54] et par conséquent, on trouve différentes définitions dans la littérature. Dans la suite, la définition de Steiner et de Desiraju, proposée dans [55] et [56] est utilisée.
Une liaison hydrogène est une interaction attractive, non covalente, entre un donneur de proton, noté D-H et un accepteur A (figure 6.1 à gauche) :

$$D - H \cdots A \tag{6.1}$$

Le donneur $D - H$ est un groupement polaire ($D^{\delta-} - H^{\delta+}$), l'accepteur A ($A^{\delta-}$) est un atome possédant une concentration de charge négative qui est stériquement accessible. Ainsi, ce sont surtout les atomes comme l'azote ou l'oxygène, qui jouent le rôle du donneur, mais des études ont montré que même un groupement $C - H$ peut être impliqué dans une liaison hydrogène [57]. L'interaction peut avoir lieu entre différentes molécules ou au sein de la même molécule.
L'énergie de la liaison hydrogène est plus faible que celle de la liaison covalente ou ionique mais plus forte que la liaison de van der Waals. Elle peut varier de -1 kcal/mol (liaison faible) jusqu'à -40 kcal/mol (liaison forte) [55]. De manière intéressante, l'énergie de Gibbs du repliement protéique est estimée de -5 à -15 kcal/mol [4], ce qui est de l'ordre de grandeur d'une seule liaison hydrogène.
La liaison H a une préférence pour la linéarité, la directionalité de cette interaction est alors importante. Par conséquent, son énergie dépend de la distance et de l'angle de liaison [58]. Il est alors commode de caractériser une liaison hydrogène par ses paramètres géométriques (figure 6.1). Dans la littérature, les paramètres suivants sont souvent utilisés :

- la distance H...A
- l'angle D-H...A (φ)
- la distance D...A
- l'angle D...A-D' (φ')

Notons que pour les deux premiers paramètres, la position de l'atome hydrogène doit être connue.

Les paramètres géométriques peuvent également être utilisés comme critère permettant de valider la présence d'une liaison hydrogène. Certaines définitions utilisent le "Van der Waals cutoff", donc ils considèrent que la liaison existe si la distance entre le donneur D-H et l'accepteur A est plus courte

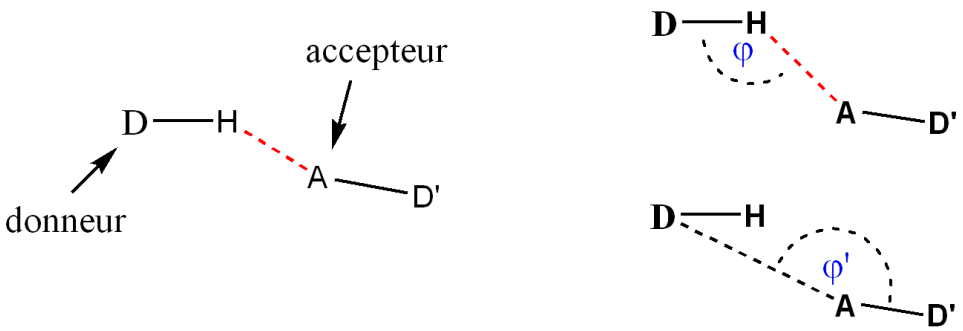


Figure 6.1 – A gauche : schéma d'une liaison hydrogène. L'interaction entre l'atome d'hydrogène et l'accepteur est représentée en ligne pointillée. A droite : Paramètres géométriques d'une liaison H. φ représente l'angle D-H...A, φ' représente l'angle D...A-D'.

que la somme de leur rayons de van der Waals respectifs. Cependant, ce critère semble trop restrictif [55], Steiner par exemple propose alors les limites suivantes :

- la distance H...A est plus courte que 3.0 voir même de 3.2 Å
- l'angle D-H...A est plus grand que 110° voir même de 90°
- les angles D-H...A linéaires sont favorisés statistiquement

Notons que des configurations plus compliquées de la liaison H peuvent exister (figure 6.2). Si un accepteur interagit avec deux atomes d'hydrogène en même temps, on parle d'une liaison double. Dans le cas ou un atome d'hydrogène interagit avec deux accepteurs, la liaison est dite bifurquée.

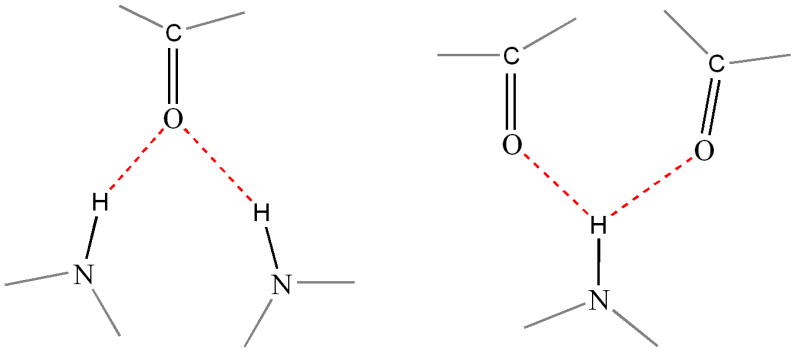


Figure 6.2 – *A gauche* : liaison H double (lignes pointillées). *A droite* : liaison H bifurquée.

La liaison hydrogène est une interaction complexe composée de plusieurs constituants de nature différente. L'énergie totale est divisé en contributions électrostatiques, de polarisation, de transfert de charge, de dispersion et de

répulsion d'échange. L'interaction n'est pas dominée par un terme particulier dans tous les cas, mais notons que le potentiel à longue distance est dominé par la nature électrostatique.
En conséquence, j'ai entrepris d'étudier la topologie des interactions dans les hélices et les feuillets dans plusieurs structures de protéine.

6.3 Topologie de la densité électronique des liaisons hydrogène

L'analyse topologique de 83 points critiques dans des composés de petite taille [2] a montré que les liaisons hydrogènes peuvent être décrites et classifiées de manière adéquate par les propriétés topologiques de la densité électronique, aux points critiques de type (3, -1) (section 2.7) [59]. Remarquons que pour les interactions N-H...O, également présentes dans les hélices, les distances accepteur-atome d'hydrogène varient entre 1.65 et 2.65 Å. De manière intéressante, on trouve des relations entre les paramètres géométriques de la liaison et les propriétés de la densité électronique au point critique. Par exemple, la densité d'énergie cinétique [3] $G(\vec{r}_{CP})$, ainsi que la densité d'énergie potentielle [4] $V(\vec{r}_{CP})$, dépendent exponentiellement de la distance H...O. Les deux quantités, G et V, reflètent comment les électrons autour du point critique $\vec{r}_{CP}$ sont affectés par l'interaction.
$G(\vec{r}_{CP})$ représente la tendance du système à diluer les électrons, ce qui correspond à la pression exercée par les électrons sur le système de liaison hydrogène. $V(\vec{r}_{CP})$ reflète la capacité du système de concentrer des électrons au point critique, ce qui correspond à la pression du système de liaison H sur les électrons autour du point critique.
De plus, la densité d'énergie cinétique $G(\vec{r}_{CP})$ est proportionnelle à la valeur propre positive de la densité au point critique, λ_3, tandis que $V(\vec{r}_{CP})$ est proportionnelle à la somme des valeurs propres négatives, $\lambda_1 + \lambda_2$ [60].
Finalement, il a été montré que λ_3 est lié de manière exponentielle à la distance H...O (figure 6.3) [61]. Rappelons que λ_3 correspond à la courbure de la densité électronique le long de l'interaction, elle représente donc la concentration de la densité autour des points critiques lors de l'interaction à couche fermée.

2. Les structures ont été déterminées par diffraction des rayons X ou neutrons, les liaisons H sont de type X-H...O avec X = C, N, O.

3. $G(\vec{r}_{CP})$ est calculé à partir de la connaissance de la densité électronique (et de son Laplacien) au point critique

4. $V(\vec{r}_{CP})$ est liée par la forme locale du théorème du Viriel à $G(\vec{r}_{CP})$ et au Laplacien de la densité électronique.

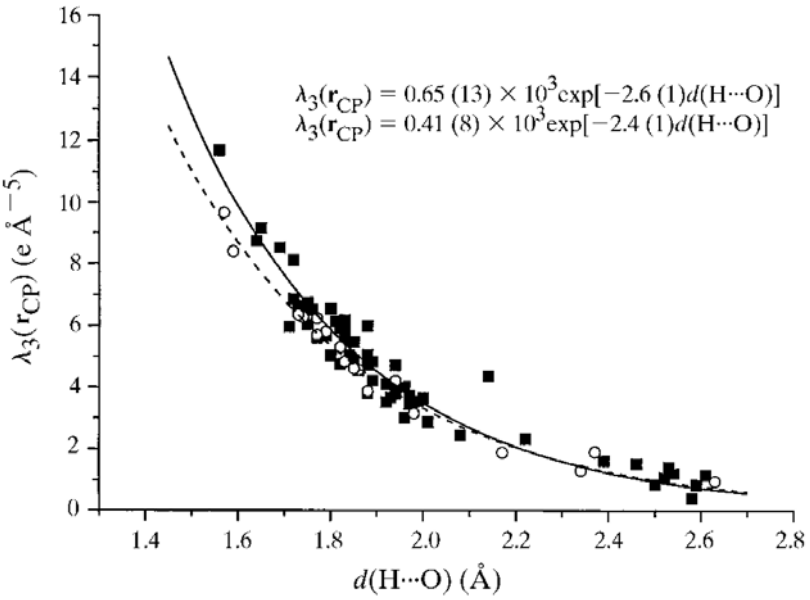


Figure 6.3 – Valeur propre positive de la matrice Hessienne au point critique en fonction de la distance H...O dans les liaisons H de type X-H...O. Les carrés noirs et les cercles représentent respectivement les points dérivés de données de diffraction à RX seuls et de données neutrons et RX joints. La courbe noire (1ere équation) correspond au fit des points dérivés de données RX, la courbe pointillé (deuxième équation) correspond au fit des points dérivés de données RX et neutrons. La figure est extraite de [61].

6.4 Etudes sur les liaisons H des hélices dans la littérature

Les liaisons hydrogènes dans les motifs de structure secondaire ont déjà attiré l'attention ; dans la littérature on trouve plusieurs études sur les géométries des liaisons H dans les hélices et les feuillets.
Les paramètres géométriques des liaisons H dans les motifs de structures secondaires, tirés de calculs théoriques ([62], [63]) ou d'analyses sur un grand nombre de structures protéiques ([5], [49], [64]), sont résumés tableau 6.1.

motif	**distance** [Å]		**angle** [°]	**source**
	N...O	H...O	N...O=C	
hélice α	3.0	2.10	147	[49]
	2.99(2)		153(2)	[64]
	2.98(16)			[5]
		2.1		[62]
		2.3		[63]
hélice 3_{10}		2.15	153	[49]
	3.13(6)		106-122	[64]
feuillet parallèle	2.93	2.05	157	[49]
	2.94(5)		157-162	[64]
		1.9		[62]
feuillet antiparallèle	2.93	2.06	154	[49]
	2.94(3)		153	[64]

Table 6.1 – Résultats de différentes études sur la géométrie des liaisons H impliquant les unités peptidiques dans les motifs de structure secondaire. Les valeurs surlignées en bleu sont issues de calculs théoriques.

L'étude de Thomas *et al.* sur 131 protéines, dont les structures ont été déterminées par diffraction des rayons X, révélait que, dans l'hélice α, les atomes d'hydrogènes des groupements amide (N-H) ne sont pas situés dans le plan des paires libres d'électrons de l'atome d'oxygène des groupements carbonyl (C=O). L'atome H attaque alors l'atome d'oxygène en moyenne avec un angle de 58° au dessus du plan formé par les paires libres. Par contre, dans les feuillets, les groupements NH forment des liaisons H canoniques, permettant l'attaque des paires libres dans le plan peptidique [49].
Fain *et al.* ont analysé la géométrie des liaisons hydrogène dans 1723 structures protéiques non-homologues [5]. Ils montraient qu'un grand nombre de liaisons H dans les motifs de structure secondaire a le caractère d'une liaison double ou bifurquée. En effet, on trouve un nombre important de liaisons H

entre les résidus i et $i+3$ (noté $O_i...N_{i+3}$) dans les hélices α bien que ce type de liaison soit typique pour les hélices 3_{10}. Dans les hélices α, les distances moyenne des liaisons $O_i...N_{i+4}$ et $O_i...N_{i+3}$ sont de 2.98(16) et de 3.22(18) Å, respectivement. Notons que cette dernière valeur est légèrement plus élevée que la distance N...O dans les hélices 3_{10}, qui est de 3.13(6) Å(tableau 6.1). Parthasarathi *et al.* ont entrepris une analyse topologique sur une structure de polyalanine obtenue par calculs *ab initio* [65]. D'après leur étude, il y a une relation linéaire entre l'énergie relative de l'hélice et la valeur totale[5] de la densité électronique $\rho(\vec{r}_{CP})$ ainsi que du Laplacien total de la densité ($\nabla^2\rho(\vec{r}_{CP})$) aux points critiques. Ceci souligne l'importance des interactions sur la stabilité d'une hélice.
Une étude théorique effectuée par LaPointe *et al.* a mis en évidence une autre interaction dans l'hélice, qui implique l'atome d'azote du résidu i et l'atome d'oxygène du résidu i+3 [63]. Ce type d'interaction a surtout lieu au centre de l'hélice.

6.5 Matériels et méthodes

L'étude a été effectuée sur les hélices et les feuillets issus de plusieurs structures protéiques tirées de la PDB [66]. Les critères de choix étaient la résolution (supérieure à 1 Å), les facteurs d'accord statistiques (R est inférieur à 15% et $R - R_{free} \leq 5\%$) et la date de déposition (pas de structure datant d'avant 1995). Certaines structures ont déjà été utilisées pour la partie II :

- Aldose réductase humaine (modèle fourni par B. Guillot, 0.66 Å de résolution) [7]
- Beta glucosidase de *thermus thermophilus* (code PDB : 1UG6, 0.99 Å de résolution)
- CTX-M-9 beta lactamase (code PDB : 1YLJ, 0.98 Å de résolution)
- Xylanase thermostable de *thermoascus aurantiacus* (code PDB : 1I1W, 0.89 Å de résolution)
- Endopolygalacturonase I (code PDB : 1K5C, 0.96 Å de résolution, repliement en feuillet parallèle)
- Concavanalin A, 0(code PDB : 1NLS, 0.94 Å de résolution, repliement en feuillet antiparallèle)

Les conformations multiples des protéines aldose réductase, beta lactamase, endopolygalacturonase et xylanase ont été enlevées avec le module "pdbcure" du logiciel CCP4 [67].
L'utilisation des structures obtenues par diffraction des RX à haute réso-

5. La valeur totale correspond à la somme de $\rho(\vec{r}_{CP})$ de toutes les liaisons H dans l'hélice.

lution offre l'avantage que les coordonnées atomiques sont déterminées avec plus de précision et que certains atomes d'hydrogènes peuvent être placés sur la base de la densité électronique. Cependant, les atomes H étaient absents dans la plupart des modèles [6] déposés sur la PDB. Par conséquent, ils ont été ajoutés avec Molprobity [40]. Puisque les atomes d'hydrogènes importants pour cette étude sont localisés sur les atomes N de la chaîne principale, leur position n'est pas ambiguë et ne dépend que de la stéréochimie de l'unité peptidique et de la position de l'atome d'azote. Les structures n'ont pas été affinées après l'ajout des atomes d'hydrogène.
Ensuite, la structure a été importée au format Mopro et les paramètres multipolaires ont été transférés de la banque ELMAM au modèle. Les liaisons covalentes X-H ont été allongées à la distance observée par diffraction des neutrons. Les molécules d'eau et les ligands ont été supprimés dans les modèles protéiques, donc la recherche des points critiques a été effectué avec la densité électronique de la protéine uniquement.
L'analyse topologique des liaisons hydrogène de type N-H...O=C a été réalisée avec VMoPro [18]. Le calcul est basé sur la minimisation de la norme du gradient de la densité électronique. Par ailleurs, un critère géométrique a été appliqué : un point critique a été cherché seulement si la distance entre le donneur et l'accepteur (ce qui correspond à la distance H...O) est inférieure à 2.9 Å. En plus, les deux critères suivants, basés sur une étude théorique des liaisons H [57], ont été appliqués afin de considérer un point critique comme valide. (1) Le point critique (CP) doit satisfaire une topologie correcte, donc le CP est situé entre les atomes d'oxygène et d'hydrogène et un chemin de liaison (*bond path*) existe. (2) Les valeurs de la densité électronique $\rho(\vec{r}_{CP})$ et du Laplacien $\Delta\rho(\vec{r}_{CP})$ sont plus élevées que 0.013 e/Å^3 et 0.57 e/Å^5, respectivement.
Ce calcul fournit toutes les interactions N-H...O=C dans la protéine. Les liaison H ont été classées d'après leur appartenance à un motif de structure secondaire (hélice, feuillet parallèle, feuillet antiparallèle). L'assignement des hélices et des feuillets a été effectué avec STRIDE [41], une tolérance de ± 1 résidu a été acceptée [7] En effet, les différents logiciels d'assignement fournissent souvent des résultats qui se distinguent légèrement [68], surtout aux extrémités des motifs.
Pour chaque point critique, les paramètres topologiques et géométriques "classiques" ainsi que certains paramètres moins répandus (figure 6.4) ont été déterminés :

6. Les modèles beta glucosidase, beta lactamase, endopolygalacturonase et de concavanalin A ne contenaient pas d'atomes d'hydrogènes.

7. Exemple : L'hélice s'étend du résidu 5 à 10. Si un point critique $O_7...H_{11}$ existe, il a été considéré comme faisant partie de l'hélice.

- les distances H...O et N...O.
- les angles N-H...O, H...O=C, N...O=C.
- les valeurs de la densité électronique $\rho(\vec{r}_{CP})$ et du Laplacien $\Delta\rho(\vec{r}_{CP})$ au point critique.
- les valeurs propres de la matrice Hessienne au point critique et les valeurs de $V(\vec{r}_{CP})$ ainsi que de $G(\vec{r}_{CP})$.
- la distance entre l'atome d'hydrogène et la paire libre d'électrons la plus proche de l'atome d'oxygène, noté H...LP (LP pour "lone pair"), et l'angle N-H...LP (afin d'estimer la position des paires libres, la position des deux maxima du Laplacien de $\rho(\vec{r})$ autour de l'atome d'oxygène carbonyle a été utilisée).
- la distance entre l'atome d'hydrogène et le plan des paires libres d'électron de l'atome d'oxygène
- les atomes situés aux extrémités des chemins de liaison afin de déterminer quels atomes sont impliqués dans l'interaction.
- pour les hélices, les interactions sont triés d'après leur type : Les liaisons hydrogène formées par les groupements N-H et C=O séparés de 3 et 4 acides aminés sont notées $i \rightarrow i+3$ et $i \rightarrow i+4$, respectivement.

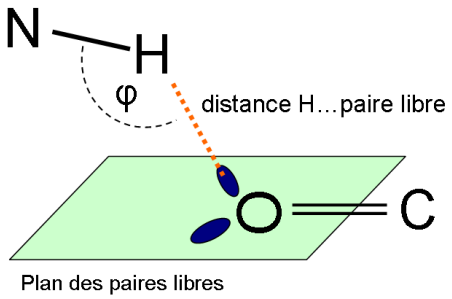


Figure 6.4 – Paramètres géométriques non standards utilisés dans cette étude. Il est intéressant de considérer la distance H...LP puisque l'atome d'hydrogène interagit notamment avec les paires libres d'électrons.

Il est intéressant de déterminer les atomes situés aux extrémités du chemin de liaison afin de savoir s'il s'agit vraiment d'une liaison hydrogène H...O ou s'il s'agit d'une interaction N...O [8]. Pour cette raison, un test automatique du chemin de liaison a été implémenté.
De même, il est intéressant de faire la distinction entre les liaisons H de type O_i...H_{i+3} et O_i...H_{i+4} dans les hélices. L'étude de Fain *et al.* a montré que les deux types sont présents dans les hélices α, souvent on trouve une succession de liaisons H bifurquées ou doubles (figure 6.5). Dans les feuillets,

8. Cette interaction a été reportée par LaPointe *et al.* [63]

ce phénomène peut être également observé. Cependant, la dinstinction étant plus difficile à réaliser automatiquement, elle n'est pas appliquée aux feuillets.

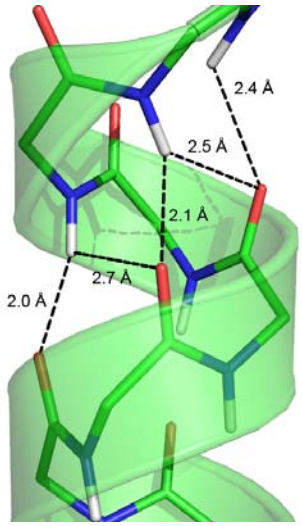


Figure 6.5 – Succession de liaisons H de type $O_i...H_{i+3}$ et $O_i...H_{i+4}$ dans une hélice α.

6.6 Résultats

6.6.1 Répartition des points critiques et géométrie des interactions

Dans les 6 protéines, 1328 points critiques (CP) ont été trouvés, le tableau 6.2 montre la distribution des CPs parmi les motifs de structure secondaire. 50,7 % des interactions ont lieu dans les hélices, 10,8 % et 22,3 % dans les feuillets antiparallèles et parallèles, respectivement. 16,2 % des liaisons sont entre des unités peptidiques qui n'adoptent pas un motif spécifique. Notons que cette repartition n'est pas représentative pour les protéines en général puisque les modèles ont été choisis en fonction de leur structure secondaire.
Les CP ont ensuite été classés de manière suivante : $i \rightarrow i+4$ signifie que la liaison hydrogène se situe entre l'atome d'oxygène O du résidu i et l'atome d'hydrogène du résidu $i+4$. Pareillement, $i \rightarrow i+3$ représente une liaison $O_i...H_{i+3}$. Si l'interaction a lieu entre l'atome d'oxygène et l'atome d'azote, elle est de type N...O.
Parmi les 1328 CP, 182 correspondent à des interactions N...O. De manière

Type	Helix	Feuillet AP	Feuillet P	Reste
Tout	673	144	296	215
$i \rightarrow i+4$	364	1	0	20
$i \rightarrow i+3$	141	0	4	52
N...O	163	1	3	15

Table 6.2 – Distribution des points critiques des liaisons hydrogène parmi les motifs de structures secondaire. P = parallèle, AP = antiparallèle, Reste = n'appartient pas à un motif de structure secondaire.

intéressante, 163 de ces interactions, donc 90,0 %, ont lieu dans les hélices. Elles sont pratiquement absentes dans les feuillets et il y en a très peu dans le reste de la protéine. Ce type d'interaction semble alors favorisé par la géométrie particulière que les acides aminés adoptent dans une hélice. Rappelons que cette interaction a été trouvé également par une étude topologique d'une structure hélicoïdale théorique [63]. Cette interaction n'est donc apparemment pas un artefact dû à l'analyse basé sur le transfert des paramètres multipolaires (et non sur des paramètres affinés).
Les valeurs moyennes des paramètres géométriques sont résumées tableau 6.3. Soulignons qu'elles correspondent à l'échantillon complet des points critiques, sans distinction entre les liaisons bifurquées ou doubles.

	distance [Å]		**angle** [°]		
motif	N...O	H...O	N...O=C	N-H...O	H...O=C
$i \rightarrow i+4$	2.98(13)	2.03(16)	155(7)	158(10)	149(8)
$i \rightarrow i+3$	3.09(13)	2.22(19)	118(9)	147(15)	110(11)
feuillet (P)	2.95(14)	2.03(19)	154(14)	154(12)	149(15)
feuillet (AP)	2.93(12)	1.99(15)	154(12)	156(9)	149(12)
N...O	3.16(9)	2.66(11)	113(16)	110(6)	97(17)

Table 6.3 – Paramètres géométriques des liaisons H et des interactions N...O trouvés avec VMoPro. P = parallèle, AP = antiparallèle. Pour l'interaction N...O, les paramètres géométriques impliquant l'atome d'hydrogène sont indiqués en surlignés afin de comparer la géométrie par rapport aux liaisons H.

Les valeurs moyennes des distances H...O et N...O ainsi que de l'angle N...O=C sont en accord avec les valeurs que l'on trouve dans la littérature (tableau 6.1). Toutefois, les déviations standards sont plus élevées, ce qui est dû à l'échantillon de structures protéiques, plus petit pour cette étude.
Les paramètres géométriques sont comparables pour les liaisons H dans les feuillets ou pour ceux de type $i \rightarrow i+4$ dans les hélices. Pour les liaisons

$i \rightarrow i+3$, les distances de liaison sont plus longues et les angles N...O=C et H...O=C significativement moins favorables.
Pour l'interaction N...O, la distance moyenne entre les atomes d'azote et d'oxygène est significativement plus élevée (de 0.07 à 0.23 Å) que celle des liaisons H dans les hélices ou les feuillets. La distance H...O de l'interaction N...O est encore plus élevée (différente de 0.44 à 0.67 Å) que celle dans les liaisons hydrogène et les angles de liaison sont clairement plus petits que dans les liaisons H des hélices ou des feuillets. L'interaction N...O semble alors se former si l'atome d'hydrogène est éloigné de l'atome d'oxygène et si son orientation n'est pas favorable pour interagir avec une paire libre d'électrons.

6.6.2 Interactions dans les hélices

Schémas des interactions

La géométrie particulière que les acides aminés adoptent dans une hélice engendre la présence de 3 interactions, qui se distinguent soit par leur nature ou par leur géométrie :

1. Liaisons hydrogène $i \rightarrow i+4$
2. Liaisons hydrogène $i \rightarrow i+3$
3. Interaction N...O

Les trois interactions sont illustrées au sein d'une hélice dans la protéine Beta-Lactamase (figure 6.6).
La liaison H $i \rightarrow i+4$ a lieu entre les groupements peptidiques éloignés de 4 acides aminés. Elle est caractérisée par des distances H...O relativement courtes, et par des angles N-H...O et H...O=C élevés. En plus, la distance entre l'atome d'hydrogène et le plan des paires libres d'électrons de l'atome d'oxygène est faible en moyenne. La figure 6.7 représente une liaison de ce type dans la protéine Beta-Lactamase.
Ensuite, il est possible qu'une liaison hydrogène se forme entre des résidus éloignés seulement de 3 acides aminés ($i \rightarrow i+3$). Les paramètres géométriques de cette interaction sont significativement différents de ceux de la liaison $i \rightarrow i+4$. L'angle H...O=C est moins élevé et la distance entre l'atome H et le plan des paires libres est plus importante. Ce changement d'orientation a pour conséquence que la surface délimitant les bassins des atomes H et O n'est pas la même que pour les liaisons H $i \rightarrow i+4$ (figure 6.8).
Finalement, l'analyse topologique a pu mettre en évidence l'interaction N...O, qui a lieu entre l'atome d'oxygène carbonyl et l'atome d'azote amide (figure 6.9). Dans la plupart des cas, l'interaction implique des atomes éloignés de 3 acides aminés.

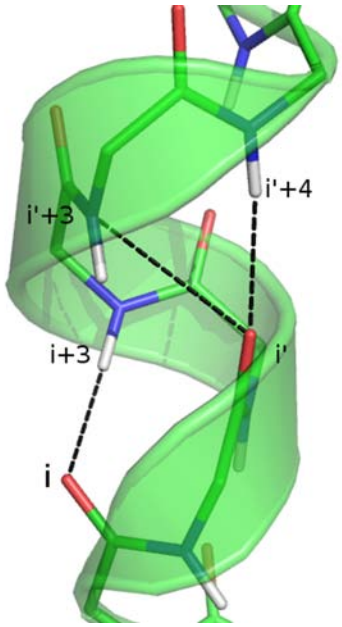


Figure 6.6 – Liaisons hydrogène $i \rightarrow i+4$ et $i \rightarrow i+3$ ainsi que l'interaction N...O dans une hélice de la protéine Beta-Lactamase.

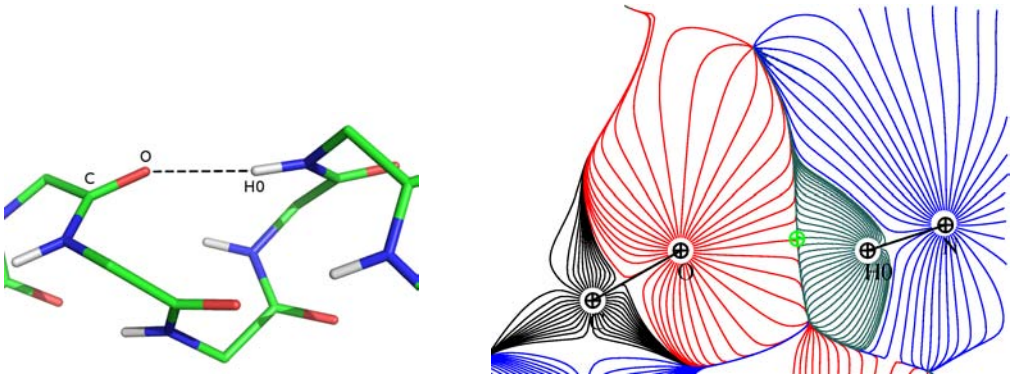


Figure 6.7 – *A gauche* : Représentation de la liaison hydrogène $i \rightarrow i+4$ au sein de l'hélice. *A droite* : lignes de gradient de la même interaction (même orientation), calculées dans le plan H...O=C. La position du point critique est marquée par une croix verte.

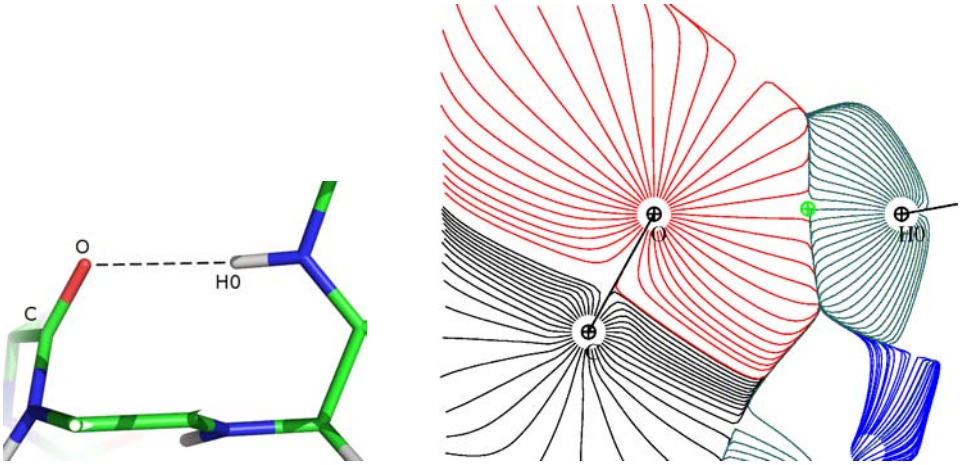


Figure 6.8 – *A gauche* : Représentation de la liaison hydrogène $i \rightarrow i+3$ au sein de l'hélice. Notons que le plan peptidique est presque perpendiculaire par rapport au plan de la figure. *A droite* : lignes de gradient de la même interaction (même orientation), calculées dans le plan H...O=C. La position du point critique est marquée par une croix verte.

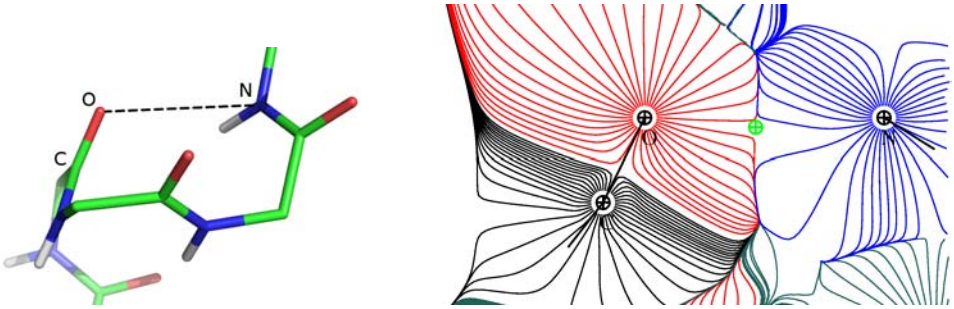


Figure 6.9 – *A gauche* : Représentation de l'interaction N...O au sein de l'hélice. *A droite* : lignes de gradient de la même interaction (même orientation), calculées dans le plan N...O=C. La position du point critique est marquée par une croix verte.

Afin d'illustrer la différence de l'interaction N...O par rapport à une liaison hydrogène, les lignes de gradient sont également représentées dans le plan N-H...O (figure 6.10). L'orientation du groupement N-H par rapport à l'atome d'oxygène est telle que le point critique n'est pas situé sur la surface interatomique qui est partagé par les atomes d'hydrogène et d'oxygène. Par conséquent, l'interaction n'est pas une liaison hydrogène.

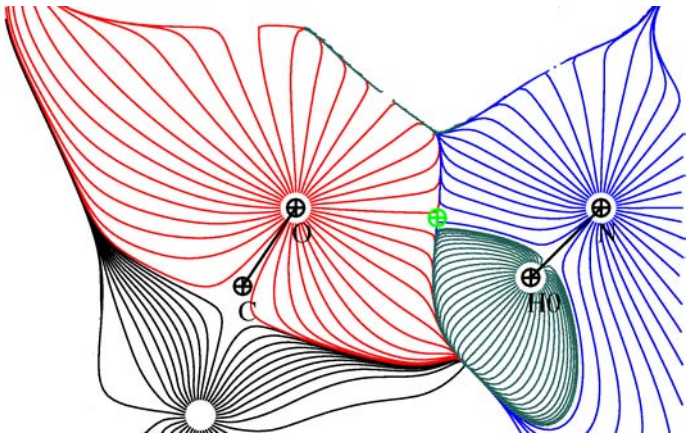


Figure 6.10 – Lignes de gradient de l'interaction N...O, calculées dans le plan N-H...O.

Liaisons hydrogène

La figure 6.11 représente la courbure longitudinale λ_3 de la densité électronique au point critique en fonction de la distance H...O.
La distance des liaisons H de type $i \rightarrow i+4$ se situe entre 1.7 et 2.6 Å, tandis que les liaisons H $i \rightarrow i+3$ occupent une région s'étendant de 1.9 à 2.7 Å. Les deux types de liaisons n'occupent alors pas la même gamme de distances. Dans les deux cas, les distances H...O n'atteignent pas 2.9 Å, ce qui correspond à la valeur limite appliquée lors de la recherche des points critiques. Ce critère n'est donc apparemment pas trop fort et n'entraîne pas l'omission de points critiques.
Les deux courbes se superposent bien et sont régulières[9] jusqu'à la distance H...O de 2.2 Å. A partir de cette distance, les points sont plus dispersés et il semble y avoir deux courbes parallèles.
L'ajustement exponentiel obtenu à partir des 83 liaisons H dans des petites

9. A l'exception d'un outlier, qui correspond à une liaison H $i \rightarrow i+4$ dans une hélice de l'aldose réductase. La valeur de $\rho(r_{\vec{CP}})$ est très élevée à ce point et la géométrie de cette liaison est très favorable.

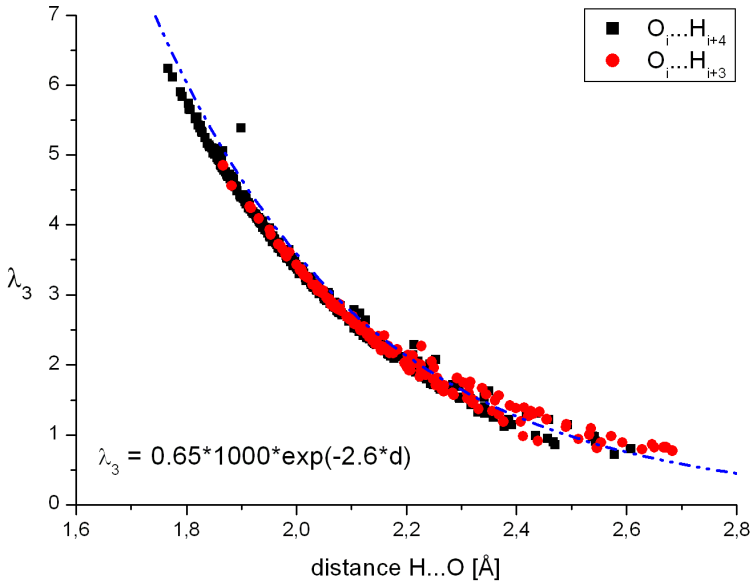


Figure 6.11 – Courbure positive de la densité électronique au point critique en fonction de la distance H...O. Cercles rouges : liaisons H de type $i \rightarrow i+3$. Carrés noirs : liaisons de type $i \rightarrow i+4$. Ligne bleue pointillée : Fit obtenu à partir des 83 liaisons hydrogènes dans les composés de petite taille (section 6.3) [61], d représente la distance H...O.

mailles (section 6.3) [61]) est tracé également figure 6.11. La fonction est de la forme :

$$\lambda_3 = a \cdot 1000 \cdot exp\,[b \cdot d] \tag{6.2}$$

d correspond à la distance de liaison H...O, les paramètres à ajuster sont a et b ($a > 0$ et $b < 0$).
L'ajustement représente bien les points entre 2.2 et 2.7 Å mais il dévie pour les courtes distance H...O. Rappelons que cette étude incluait également les donneurs C-H et N-H, ce qui pourrait expliquer la faible déviation par rapport à l'analyse actuelle. Dans les petites molécules, les atomes d'hydrogène ont étés placés d'après des pics de densité électronique tandis que pour la présente étude, ils ont étés générés, dans la plupart des cas. Il est alors possible que leur position soit légèrement différente (par exemple ils sortent légèrement du plan peptidique).
Un ajustement de courbe a été effectué séparément pour les deux types de liaison hydrogène ($i \rightarrow i+3$ et $i \rightarrow i+4$), en utilisant la même fonction exponentielle décroissante. Les paramètres a et b sont résumés tableau 6.4. L'ajustement des points correspondant aux liaisons de type $i \rightarrow i$+4 concorde bien avec celui basé sur les liaisons hydrogène dans les composés de petite

taille [61]. Toutefois, les paramètres ajustés de la courbe $i \rightarrow i+3$ sont sensiblement différents.

Courbe	a	b [Å^{-1}]
[61]	0.65(13)	-2.6(1)
$i \rightarrow i+4$	0.65(2)	-2.62(1)
$i \rightarrow i+3$	0.45(2)	-2.43(2)

Table 6.4 – Ajustement des paramètres pour la courbe représentée figure 6.11, dans l'équation 6.2.

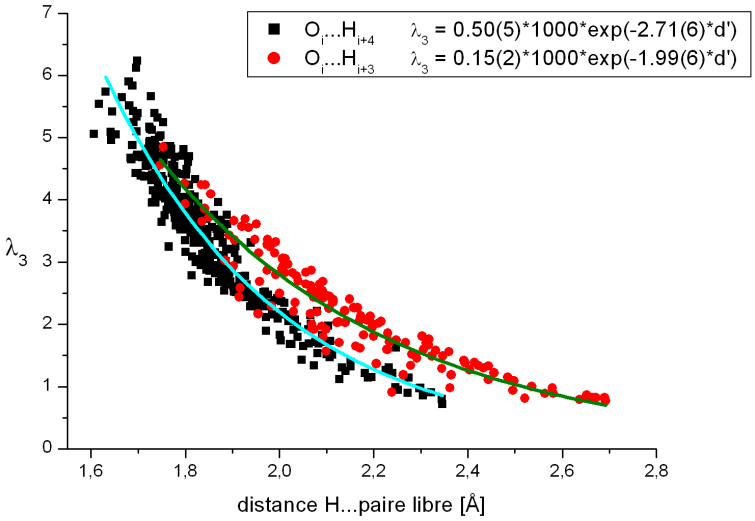


Figure 6.12 – Courbure positive de la densité électronique au point critique en fonction de la distance entre l'atome d'hydrogène et la paire libre d'électrons la plus proche. Cercles rouges : liaisons H de type $i \rightarrow i+3$. Cercles noirs : liaisons de type $i \rightarrow i+4$.

La figure 6.12 représente la courbure de la densité électronique en fonction de la distance entre l'atome d'hydrogène et la paire libre d'électrons la plus proche. En utilisant ce nouveau paramètre géométrique, la courbe garde son caractère exponentiel mais elle devient moins régulière et les points sont plus dispersés. De manière surprenante, les points appartenant aux interactions $i \rightarrow i+3$ et $i \rightarrow i+4$ occupent maintenant différentes régions (à l'exception de quelques points de type $i \rightarrow i+3$ qui entrent dans la zone des liaisons $i \rightarrow i+4$). Un ajustement de courbe séparé, utilisant la même fonction exponentielle que ci-haut (mais en remplacant la distance H...O par H...LP), donne des paramètres assez différents : a et b sont égaux à 0.50(5) et 2.71(6)

pour les liaisons $i \rightarrow i+4$ tandis qu'ils sont de 0.15(2) et 1.99(6) pour les liaisons $i \rightarrow i+3$.
A notre connaissance, c'est la première fois que la distance entre l'atome d'hydrogène et la paire libre d'électrons a été utilisé comme paramètre lors d'une analyse topologique.

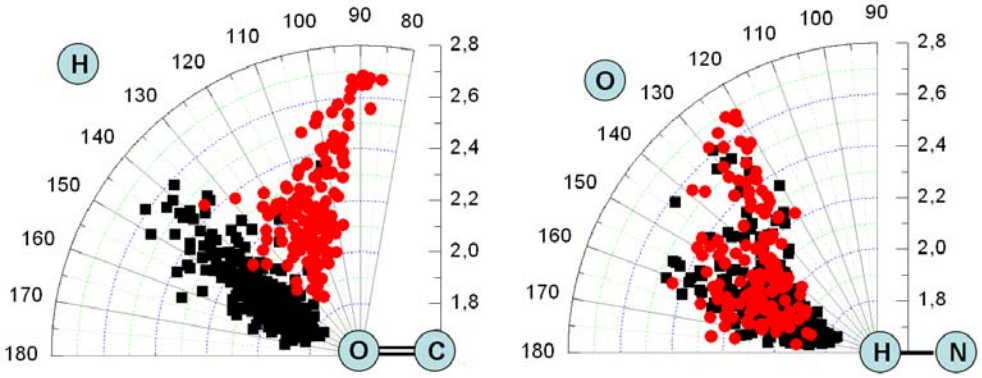


Figure 6.13 – Représentation polaire des paramètres géométriques. *A gauche* : L'angle H...O=C et la distance H...O pour les interactions de type $i \rightarrow i+3$ (cercles rouges) et $i \rightarrow i+4$ (carrés noirs). Le rayon correspond à la distance H...O, l'angle polaire correspond à l'angle H...O=C. *A droite* : L'angle N-H...O et la distance H...O, même code couleurs.

Quelques paramètres géométriques des liaisons H dans les hélices sont représentés figure 6.13. La distribution des angles N-H...O, classiquement utilisée pour caractériser géométriquement les liaisons H, est pratiquement identique pour les liaisons de type $i \rightarrow i+4$ et $i \rightarrow i+3$. Ils varient entre 180 et 115°, notons que pour des angles inférieurs à 140°, les distances de liaison correspondantes sont plus longues. Quant à la distribution des angles H...O=C, on observe de nouveau des accumulations différentes des points. Pour les liaisons H $i \rightarrow i+4$, l'angle H...O=C varie entre 180 et 130° tandis qu'il est situé entre 80 et 130° pour les liaisons $i \rightarrow i+3$. Par conséquent, les deux types de liaison ne se distinguent pas par l'orientation du groupement amide, mais par celui du groupement C=O qui forme alors un angle important par rapport à la direction N-H.
Ce résultat est illustré par la figure 6.14 qui représente l'angle H...O=C en fonction de l'angle N-H...O pour les deux types de liaison hydrogène. De nouveau, on observe une accumulation assez distincte des points. Pour un même angle N-H...O, l'angle H...O=C correspondant est systématiquement plus élevé pour les liaisons $i \rightarrow i+4$, pour lesquelles les groupements N-H et C=O sont alors plus alignés. Les distances moyennes entre l'atome d'hydrogène et le plan des paires libres de l'atome d'oxygène sont de 0.79(30)

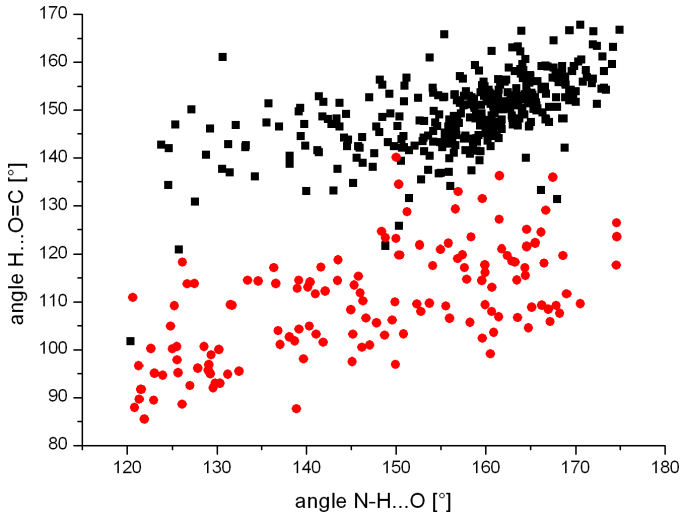


Figure 6.14 – L'angle H...O=C en fonction de l'angle N-H...O des liaisons H dans les hélices. Cercles rouges : liaisons H de type $i \rightarrow i+3$. Cercles noirs : liaisons de type $i \rightarrow i+4$.

et de 2.01(33) Å pour les liaisons $i \rightarrow i+4$ et $i \rightarrow i+3$, respectivement. Par conséquent, les liaisons $i \rightarrow i+3$ adoptent une géométrie où l'atome d'hydrogène est significativement éloigné du plan des paires libres.

Interaction N...O

Parmi les 163 interactions N...O dans les hélices, 139 ont lieu entre les atomes d'oxygène et d'azote éloignés de 3 acides aminés ($N_{i+3}...O_i$). De manière intéressante, l'interaction peut également avoir lieu entre des résidus séparés de 4 acides aminés ($N_{i+4}...O_i$) : on trouve 16 liaisons de ce type. Les interactions restantes (8 interactions) impliquent encore d'autres combinaisons.
Dans un premier temps, il est intéressant d'étudier ce qui distingue la géométrie des interactions N...O de celle des liaisons hydrogène. Les valeurs moyennes ont été comparées ci-dessus, rappelons que c'étaient les angles impliquant l'atome d'hydrogène qui étaient significativement plus faibles que ceux des liaisons H. La figure 6.15 représente les angles N...O=C (à gauche) et N-H...O (à droite) en fonction des distance N...O et H...O, respectivement.

La distribution des angles N...O=C des interactions N...O chevauche avec celle des liaisons H $i \rightarrow i+3$ (figure 6.15 à gauche). Quelques points sont dans la région des liaisons hydrogènes $i \rightarrow i+4$, ils correspondent aux inter-

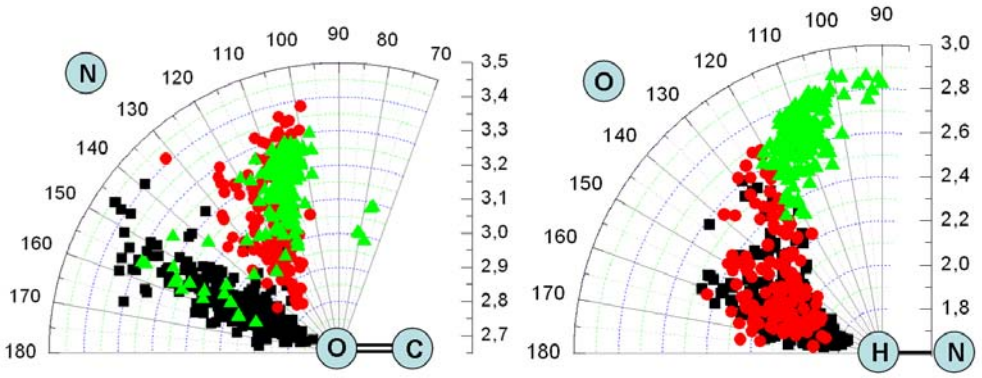


Figure 6.15 – Représentation polaire des paramètres géométriques. *A gauche* : L'angle N...O=C et la distance N...O pour les interactions de type $i \rightarrow i+3$ (cercles rouges), $i \rightarrow i+4$ (carrés noirs) et N...O (triangles verts). Le rayon correspond à la distance N...O, l'angle polaire correspond à l'angle N...O=C. *A droite* : L'angle N-H...O et la distance H...O, même code couleurs.

actions N...O qui ont lieu entre les atomes séparés de 4 acides aminés. L'angle N...O=C ne semble alors pas être le paramètre crucial qui distingue l'interaction N...O d'une liaison hydrogène. Toutefois, la distribution des angles N-H...O en fonction des distances H...O est différente pour les interactions N...O et les liaisons H (figure 6.15 à droite). On observe une accumulation des points qui correspond à des distances H...O élevés et à des angles N-H...O faibles. Notons que cette fois-ci, il n'y a pas d'outliers, les points sont concentrés dans une région précise. La valeur minimale de l'angle N-H...O est de 120° pour les liaisons H ($i \rightarrow i+3$ et $i \rightarrow i+4$), tandis que la valeur maximale du même angle est de 125° pour les interactions N...O. L'angle de 120° constitue alors une valeur limite qui pourrait discriminer une liaison H d'une interaction N...O dans une hélice, sans effectuer une analyse topologique.
Ceci pourrait servir à des études futures qui se basent uniquement sur des critères géométriques pour déterminer la présence d'une liaison hydrogène dans les hélices. Par exemple dans l'étude de *Koch et al.* [64], la distance N...O doit être inférieure à 3.5 et l'angle N...O=C doit être supérieur à 90° pour qu'une interaction soit validée comme liaison H. La figure 6.16 représente l'angle N...O=C en fonction de la distance N...O sur la base de l'échantillon utilisé pour notre étude. Puisque l'interaction N...O satisfait également les critères, il n'est pas possible de la discriminer des liaisons hydrogènes.
L'interaction N...O se distingue alors d'une liaison H $i \rightarrow i+3$ par la grandeur de l'angle N-H...O. De plus, la distance moyenne entre l'atome d'hydrogène et le plan des paires libres d'électrons est plus élevée (2.4(5) Å) que celle des liaisons $O_i...H_{i+3}$ (2.0(3) Å). Pour l'interaction N...O, cette différence

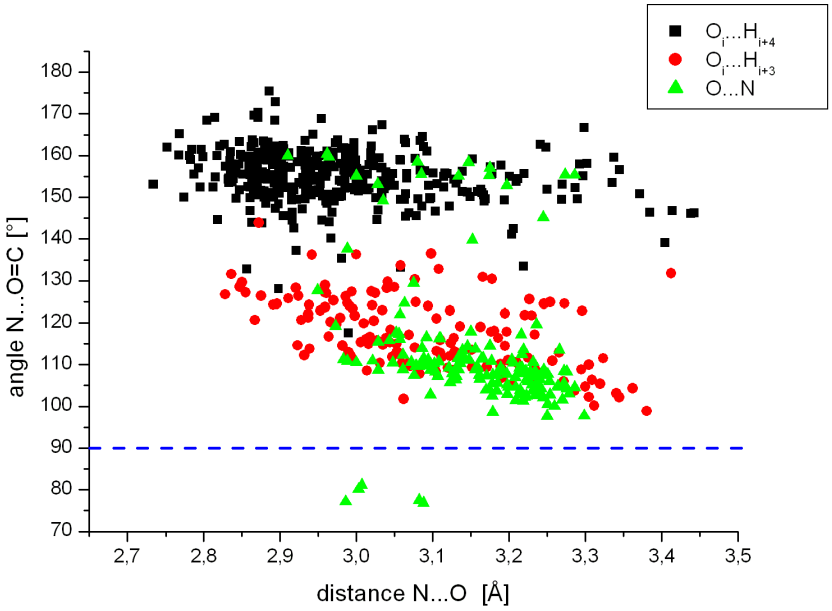


FIGURE 6.16 – Représentation de l'angle N...O=C en fonction de la distance N...O. Carrés noirs : liaison H $i \rightarrow i+4$, cercles rouges : liaison H $i \rightarrow i+3$, triangles verts : interaction N...O. La valeur limite de l'angle N...O=C utilisé comme critère dans l'étude [64] est représenté comme ligne pointillée bleue.

géométrique est suffisante pour qu'un point critique ne soit pas partagé par les bassins interatomiques des atomes H et O et qu'un chemin de liaison ne puisse pas être créé.
Afin d'analyser la différence en terme de topologie, la relation entre la densité d'énergie cinétique $G(\vec{r}_{CP})$ est tracée en fonction de la courbure positive de la densité électronique au point critique (λ_3) (figure 6.17).
Pour les différentes liaisons, les distributions des points de $G(\vec{r}_{CP})$ se superposent et forment une courbe linéaire et très régulière. Les courbes ont été ajustées séparément avec la fonction linéaire $G(\vec{r}_{CP}) = b \cdot \lambda_3$, passant par l'origine. Les paramètres (15.200(5), 15.24(1) et 15.42(9)) ne diffèrent pas significativement et ressemblent au fit obtenu à partir des 83 liaisons H dans les composés de petite taille qui obtenait la valeur 15.3(1) pour le paramètre b [60].
Avec 11 kJ/mol (par volume d'unité atomique), la valeur minimale de $G(\vec{r}_{CP})$ est la même pour les trois interactions. Néanmoins, les valeurs maximales sont différentes : elle sont de 96, 74 et 29 kJ/mol (par volume d'unité atomique) pour les liaisons $i \rightarrow i+4$, $i \rightarrow i+3$ et N...O, respectivement. Des petites valeurs de $G(\vec{r}_{CP})$ et de λ_3 représentent des liaisons faibles. Par conséquent, le chevauchement entre les distributions électroniques des atomes d'azote et

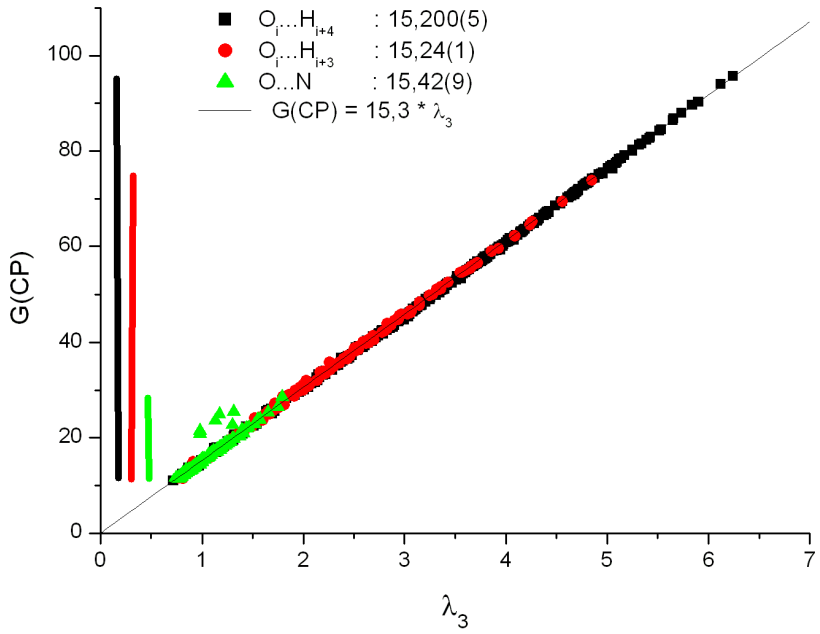


Figure 6.17 – Densité d'énergie cinétique (en kJ/mol par volume d'unité atomique) en fonction de la courbure positive au point critique pour les liaisons H $i \to i+3$ (cercles rouges), $i \to i+4$ (carrés noirs) et l'intereaction N...O (triangles verts). Les lignes verticales désignent la gamme de $G(\vec{r}_{CP})$ (qui n'est pas bien visible car les courbes se superposent). Le fit linéaire des 83 liaisons H [60] est tracé afin de comparer avec les valeurs de la littérature. Les paramètres du fit linéaire sont donnés sur la légende.

d'oxygène est faible et l'interaction N...O a plutôt un caractère électrostatique.
Remarquons qu'il y a quelques outliers pour l'interaction N...O qui sont légèrement décalés de la droite. Ils correspondent à des interactions de type $i \to i+2$ qui ont des angles H...O=C inférieur à 67°.
Ainsi, bien que l'interaction N...O n'implique pas un atome d'hydrogène, elle suit la même caractéristique topologique que les liaisons hydrogène.

Population des liaisons $i \to i+3$ et $i \to i+4$

A la différence de la définition "traditionnelle" d'une hélice, l'arrangement des liaisons hydrogènes (entre les unités peptidiques) n'est pas régulier et parfait. Comme il a été mentionné ci-haut (section 6.5 et figure 6.5) on trouve des liaisons de type $i \to i+3$ dans une hélice α, qui, par définition, est constituée seulement de liaisons $i \to i+4$. Par conséquent, il existe des atomes d'oxygène ou hydrogène qui sont impliqués dans plusieurs interactions

en même temps. Dans ce cas, il est intéressant d'y attribuer un pourcentage, par exemple, de déterminer si une interaction a un caractère $i \rightarrow i+4$ de 100% ou moins. On peut ainsi calculer la population des liaisons H $i \rightarrow i+3$ et $i \rightarrow i+4$ pour toutes les hélices.
Le pourcentage d'un type de liaison a été attribué sur la base de sa géométrie. La figure 6.18 représente la distribution des angles H...O=C ainsi que de la distance entre l'atome d'hydrogène et le plan des paires libres d'électrons de l'atome d'oxygène, noté dH...plan, pour les deux types de liaisons H ($i \rightarrow i+3$ et $i \rightarrow i+4$).

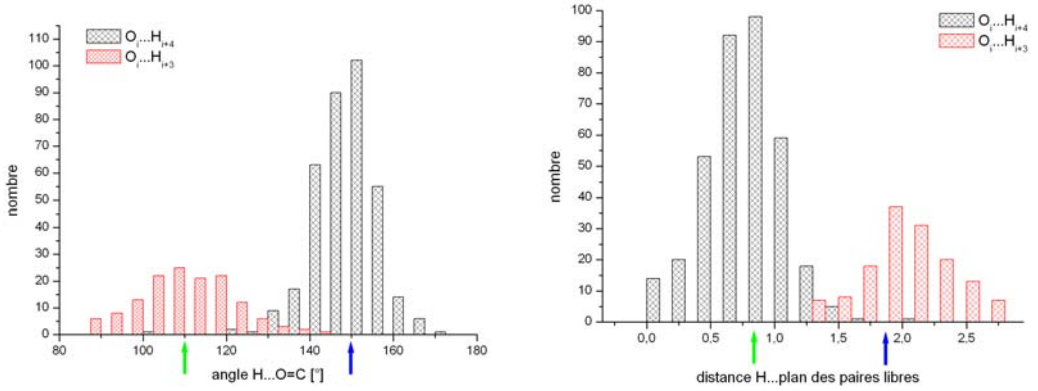


FIGURE 6.18 – *A gauche* : Distribution de l'angle H...O=C. En noir : Liaison hydrogène O_i...H_{i+4}. En rouge : liaison hydrogène O_i...H_{i+3}. Les valeurs moyennes sont marqués par des flèches, en vert et en bleu pour les interactions $i \rightarrow i+4$ et $i \rightarrow i+3$, respectivement. *A droite* : Distribution de la distance H...plan des paires libres, même code couleurs.

Les deux distributions sont bimodales et montrent clairement la distribution entre les liaisons $i \rightarrow i+3$ et $i \rightarrow i+4$. Elles peuvent ainsi servir comme critère géométrique. Puisque la valeur moyenne calculée à partir de l'échantillon total serait biaisé par la géométrie des liaisons doubles ou bifurqués, elle a été déterminé pour les liaisons hydrogènes uniques seulement. Elles sont de 115(10) et 150(8)° (angle H...O=C) et 1,9(3) et 0,8(3) Å (distance H...plan des paires libres) pour les liaisons $i \rightarrow i+3$ et $i \rightarrow i+4$, respectivement. Les valeurs maximales ou minimales d'une distribution ne dépassent pas la valeur moyenne de l'autre [10]. Le pourcentage a été calculé de manière suivante :
Le paramètre p_1 mesure la déviation par rapport à la géométrie d'une liaison

10. A l'exception d'un outlier qui correspond à une liaison H $i \rightarrow i+4$ dans la protéine xylanase

$i \rightarrow i+4$ unique en fonction de l'angle H...O=C. Il est donné par

$$p_1 = \begin{cases} 0 & \text{si } x < m_{3/10} \\ 1 & \text{si } x > m_\alpha \\ p_1 = \dfrac{x - m_{3/10}}{m_\alpha - m_{3/10}} & \text{sinon} \end{cases} \tag{6.3}$$

x correspond à l'angle H...O=C, $m_{3/10}$ et m_α sont les valeurs moyennes des liaisons $i \rightarrow i+3$ et $i \rightarrow i+4$ (flèches verts et bleues, figure 6.18), respectivement. Le paramètre p_1 est alors égal à 1 si la géométrie de l'interaction correspond à celle d'une liaison $i \rightarrow i+4$, et il est égal à 0, si elle ressemble la liaison $i \rightarrow i+3$.
Afin d'affiner le calcul du pourcentage, un deuxième paramètre, p_2, est introduit. Une combinaison de deux paramètres permet de prendre en compte plus de dispersion [11]. p_2 est calculé à partir de la distance entre l'atome H et le plan des paires libres, noté y. Il mesure la déviation de la géométrie d'une liaison H par rapport à la valeur moyenne de y des interactions $i \rightarrow i+4$, donc il prend la valeur de 1 si la géométrie est de type $i \rightarrow i+4$ et 0, si elle est de type $i \rightarrow i+3$ pur. p_2 est défini par :

$$p_2 = 1 - \begin{cases} 1 & \text{si } y > m'_{3/10} \\ 0 & \text{si } y < m'_\alpha \\ \dfrac{y - m'_{3/10}}{m'_\alpha - m'_{3/10}} & \text{sinon} \end{cases} \tag{6.4}$$

$m'_{3/10}$ et m'_α sont les valeurs moyennes des liaisons $i \rightarrow i+3$ et $i \rightarrow i+4$, respectivement.
Finalement, le pourcentage final P est égal à

$$P = \frac{p_1 + p_2}{2} \tag{6.5}$$

La distribution des pourcentages est représentée figure 6.19.
De manière intéressante, la population des pourcentages "mixtes" est très faible, la grande majorité des liaisons ont alors un caractère $i \rightarrow i+3$ ou $i \rightarrow i+4$ pur. La géométrie des interactions $i \rightarrow i+4$ semble néanmoins plus souple car le nombre réduit graduellement, tandis qu'il y a une coupure nette pour les liaisons $i \rightarrow i+3$. En effet, un nombre important de liaisons avec le

11. La population de p_1 est basée sur une distribution bimodale, donc ce n'est pas surprenant si l'on obtient des maxima pour des pourcentages de 0 et de 100. En ajoutant un paramètre, il faut que les deux critères géométriques soient satisfaits simultanément pour obtenir un pourcentage de 0 ou de 100.

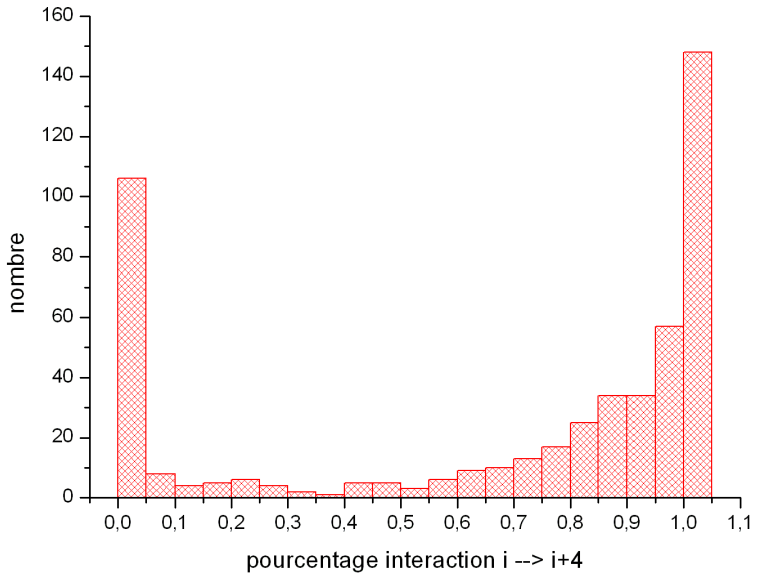


FIGURE 6.19 – Distribution du pourcentage géométrique $P = \frac{p_1 + p_2}{2}$ des liaisons H dans les hélices. 0 correspond à une liaison qui adopte une géométrie de type $i \rightarrow i + 3$

pourcentage 0 implique que la majorité des interactions ont non seulement l'angle H...O=C mais encore la distance dH...plan, qui sont inférieurs aux valeurs moyennes.
La structure hélicoïdale est alors assez rigide, puisque elle ne permet pas la formation de liaisons hydrogène avec une géométrie quelconque. Ceci souligne l'importance de la régularité des liaisons H pour la stabilité des hélices.

Impact sur le bilan énergétique des hélices

L'étude précédente a montré que de nombreuses interactions N...O dans les hélices peuvent être interprétées comme liaison hydrogène si l'on se base uniquement sur des critères géométriques classiques. Le but de cette section est d'évaluer quel serait l'impact sur le bilan énergétique si les interactions N...O étaient considérées comme liaison H. En effet, la méthode de mécanique moléculaire (MM) utilise des fonctions d'énergie potentielle afin de calculer des propriétés comme l'affinité ou afin de prédire des sites de fixation, le repliement tridimensionnel protéique ou des interfaces d'interaction protéine-protéine. Or, les liaisons hydrogène apportent une contribution significative à la stabilité d'une protéine et elles peuvent représenter la clé pour des phénomènes de fixation ou des catalyses.
La détermination d'une fonction d'énergie potentielle précise décrivant une

liaison hydrogène est difficile, ce qui est dû à la diversité de ses composantes énergétiques (électrostatique, transfert de charge, polarisation, dispersion et répulsion d'échange) [69]. Par conséquent, de nombreux champs de force ne possèdent pas de terme explicite pour l'énergie d'une liaison H (Amber [70], OPLSA [71]) ou n'incluent qu'un terme radial qui dépend uniquement de la distance entre le donneur et l'accepteur (MM2 [72] et MM3 [73]). Or, l'énergie d'une liaison H dépend également de l'orientation respective des groupements donneur et accepteur, ce qui est cohérent avec les critères géométriques (angle N-H...O par exemple). Pour cette raison, j'ai choisi d'utiliser une expression qui tient compte de la distance **et** de l'orientation, proposée par Dahiyat *et al.* [74] (On trouve d'autres exemples de fonctions plus élaborées dans [75] et [76]) :

$$V_{HB}(R,\theta,\phi,\varphi) = f(R)g(\theta,\phi,\varphi) \tag{6.6}$$

La fonction d'énergie potentielle peut être exprimée comme le produit d'une partie radiale et d'une partie angulaire. La fonction radiale correspond à un potentiel de Lennard-Jones avec les exposants -10 et -12 :

$$f(R) = D_0 \left\{ 5(\frac{R_0}{R})^{12} - 6(\frac{R_0}{R})^{10} \right\} \tag{6.7}$$

D_0 (8 kcal/mol), R_0 (2.8 Å) et R correspondent à la profondeur du puit de potentiel, à la distance d'équilibre et à la distance N...O, respectivement. Les paramètres du modèle sont représentés figure 6.20.

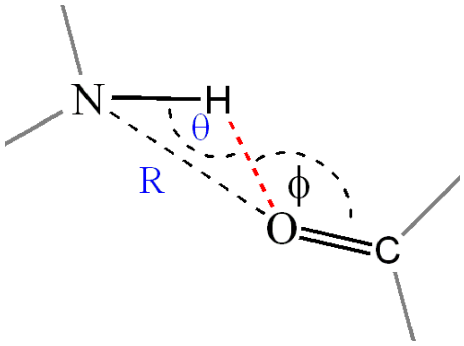


Figure 6.20 – Paramètres du modèle d'énergie potentielle de Dahiyat *et al.* [74].

Le terme angulaire dépend de l'état d'hybridation du donneur et de l'accepteur. Pour les groupements N-H et C=O, il est de la forme :

$$g(\theta,\phi,\varphi) = cos^2(\theta) \cdot cos^2(max[\phi,\varphi]) \tag{6.8}$$

θ et ϕ correspondent aux angles N-H...O et H...O=C respectivement et φ est l'angle entre les normales des plans définis par les 6 atomes liés aux centres

sp2 (ce qui correspond à l'angle entre les plans peptidiques dans cette étude). Notons que si $\varphi < 90°$, l'angle complémentaire de 180° est utilisé.
Dans un premier temps, analysons plus en détail cette fonction d'énergie potentielle. La figure 6.21 représente la partie radiale de l'énergie potentielle d'une liaison hydrogène en fonction de la distance N...O. L'énergie diminue quand la distance N...O devient plus petite, le minimum étant à 2.8 Å. Ensuite, l'énergie augmente rapidement et devient défavorable, ce qui limite la distance minimale entre le donneur et l'accepteur.

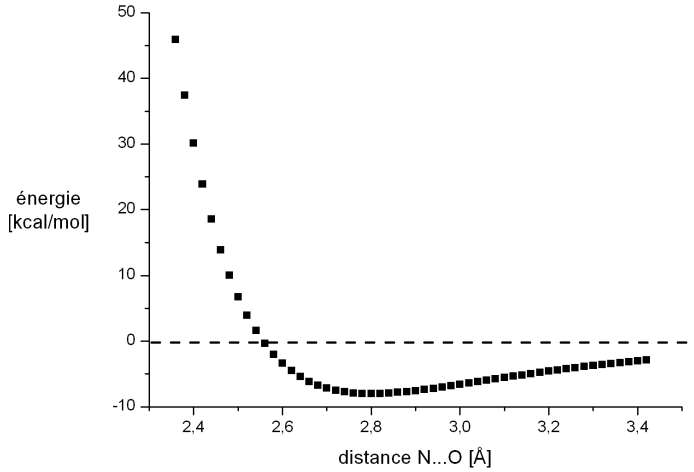


Figure 6.21 – Terme radial de l'énergie potentielle d'une liaison H en fonction de la distance N...O.

Cette fonction seule n'est pas suffisante pour décrire l'énergie d'une liaison hydrogène. En effet, en se basant uniquement sur la distance, une configuration non alignée, avec un angle de 90° entre le groupement donneur et accepteur (figure 6.22, à gauche), serait la plus favorable. Pour cette raison, un terme angulaire est nécessaire qui favorise l'arrangement linéaire des groupements N-H et C=O. Le potentiel angulaire est représenté figure 6.23 en fonction de l'angle θ, la distance R et les angles ϕ et φ étant fixés à 2.8 Å et à 180° chacun, respectivement. La variation géométrique est représentée schématiquement figure 6.22.
La partie angulaire du potentiel a un maximum, si l'angle θ est égal à 180°, par conséquent, l'énergie totale est maximale si les atomes N, H et O sont alignés. La deuxième partie du terme angulaire donne un nombre entre zéro et un (φ étant plus élevé que 90°), donc elle ne change pas le signe du produit. Par contre, des angles ϕ ou φ petits (ce qui correspond à des géométries moins favorables d'une liaison H) diminuent la contribution de la partie angulaire du potentiel.

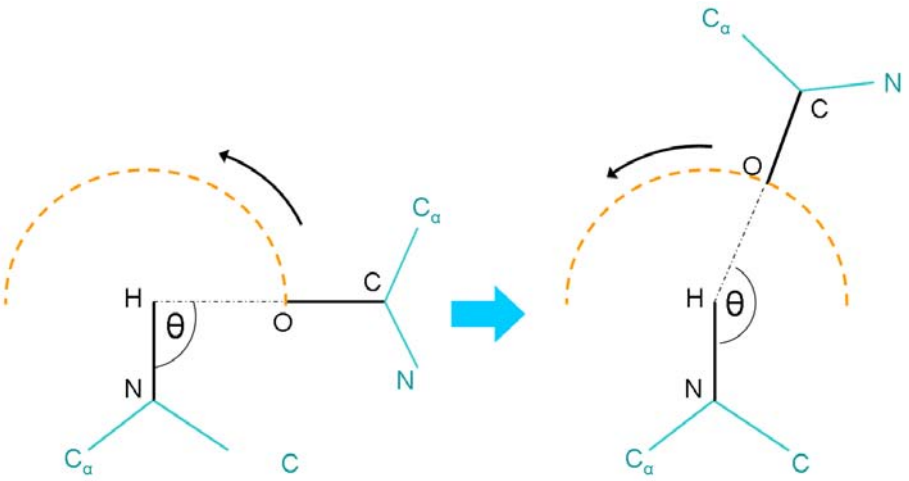


Figure 6.22 – Schéma de la liaison H entre les groupements N-H et C=O utilisée pour calculer l'énergie de la partie angulaire. *A gauche* : Configuration initiale, θ=90°. *A droite* : Configuration ultèrieure.

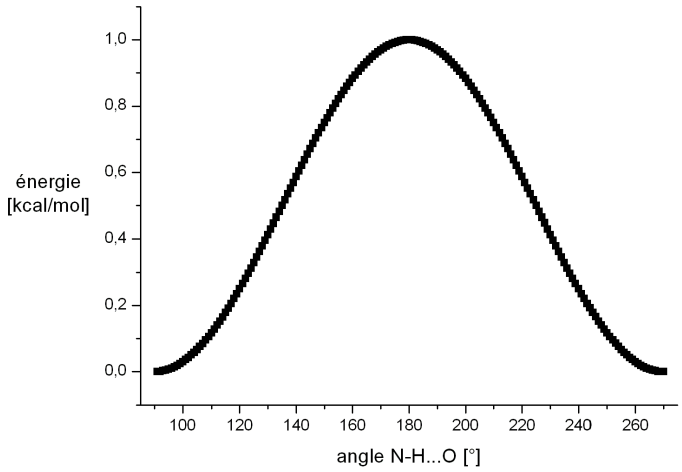


Figure 6.23 – Terme angulaire de l'énergie potentielle d'une liaison H en fonction de l'angle N-H...O (angle θ).

L'énergie totale en fonction de l'angle N-H...O est représentée figure 6.24 (avec $R = 2.8$ Å, $\phi = \varphi = 180°$). Le minimum d'énergie est à 180° et l'énergie augmente si l'angle est moins grand.

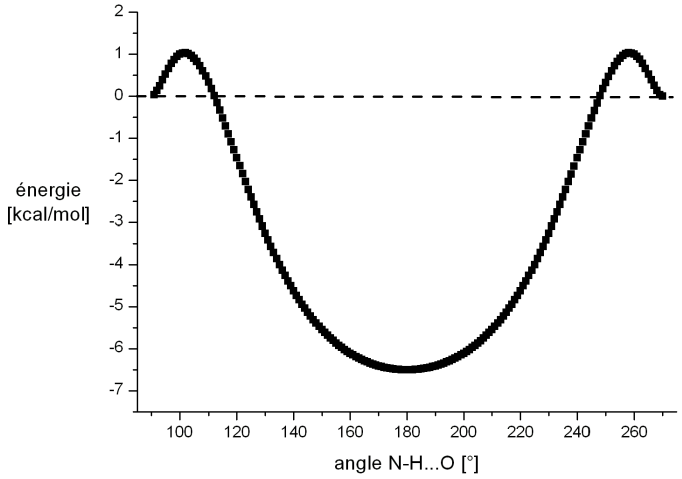


Figure 6.24 – Energie potentielle totale d'une liaison H en fonction de l'angle N-H...O (angle θ).

Grâce à cette fonction, l'énergie des liaisons H dans les hélices de la protéine Beta-Glucosidase a pu être estimée. De plus, en appliquant l'expression du potentiel aux interactions N...O, il est possible d'estimer l'impact sur le bilan énergétique si elles sont mal-interprétées comme liaison hydrogène. Le tableau 6.5 résume les résultats pour les différents types d'interaction.
Afin d'avoir une deuxième estimation de l'énergie des liaisons H, les calculs ont été effectués avec un modèle récent [75]. Le potentiel est aussi modélisé comme produit d'une partie radiale et d'une partie angulaire, mais la forme des deux fonctions est différente par rapport au modèle de Dahiyat. La partie radiale est exprimé avec une fonction de Morse :

$$f(R) = D_e \left(1 - e^{-\alpha(R-R_0)}\right) \tag{6.9}$$

D_e est la profondeur du puits de potentiel (7,785 kcal/mol), α reflète la courbure de la fonction $f(R)$ (1.234 1/Å), R_0 est la distance d'équilibre (2.0 Å) et R correspond à la distance H...O. Les paramètres du modèle sont représentés figure 6.25. La fonction de Morse a l'avantage que l'énergie devient moins rapidement défavorable en s'éloignant de la distance d'équilibre. Ainsi, des distances courtes ou longues sont moins pénalisées qu'avec un potentiel de Lennard-Jones. Le potentiel en fonction de la distance H...O est représenté figure 6.26.

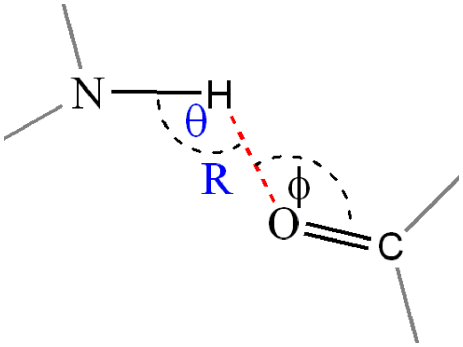


Figure 6.25 – Paramètres d'énergie potentielle d'une liaisons H dans le modèle de Choi *et al.* [75].

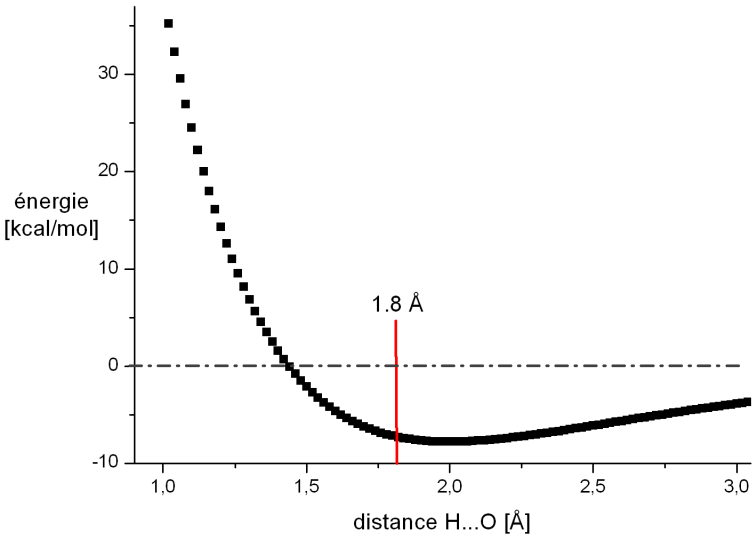


Figure 6.26 – Energie potentielle radiale d'une liaison H en fonction de la distance H...O.

La partie angulaire correspond à un produit dont les facteurs sont des séries de cosinus et sinus :

$$\begin{aligned} g(\theta,\phi) &= \sum_{i=1}^{3}\left(a_i sin^i(\theta-\theta_0)+b_i cos^i(\theta-\theta_0)\right)\cdot \\ &\quad \cdot\sum_{j=1}^{3}\left(c_j sin^j(\phi-\phi_0)+d_j cos^j(\phi-\phi_0)\right) \end{aligned} \tag{6.10}$$

θ : l'angle N-H...O

ϕ : l'angle H...O=C

θ_0 : 170°, l'angle d'équilibre N-H...O

ϕ_0 : 200°, l'angle d'équilibre H...O=C

a_i, b_i, c_j, d_j : coefficients

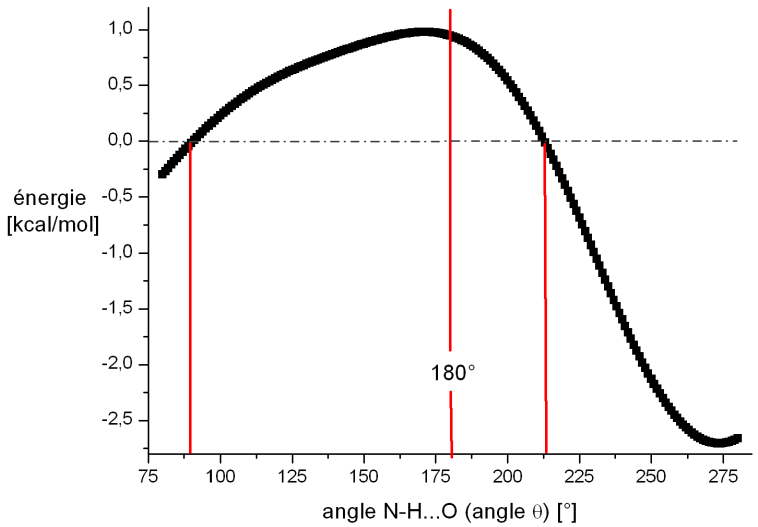


Figure 6.27 – Energie potentielle angulaire d'une liaison H en fonction de l'angle N-H...O (noté θ).

Pour $\phi = 180°$, la partie angulaire du potentiel est représentée en fonction de θ figure 6.27. Le potentiel est favorable à partir de 90°, il augmente jusqu'à 170° et tombe ensuite rapidement vers des énergies défavorables [12]. La partie du potentiel de 90 à 180° semble alors mieux modéliser le potentiel angulaire que la partie de 180 à 360°. Puisque les angles θ et ϕ sont inférieurs à 180° dans le cadre de mon étude, la fonction proposée par Choi *et al.* peut néanmoins être utilisée.

12. Ce potentiel ne ressemble pas tout à fait à celui présenté par Choi *et al.*. En effet, il y avait des erreurs dans le tableau 2 de [75], résumant les valeurs des coefficients a_i etc. Les coefficients qui ont été utilisés pour cette étude ont été communiqués par les auteurs de [75].

Le potentiel total est représenté figure 6.28, la distance H...O et l'angle H...O=C étant fixés à 1.8 Å et à 180°, respectivement (marquée par une ligne sur les figures 6.26 et 6.27).

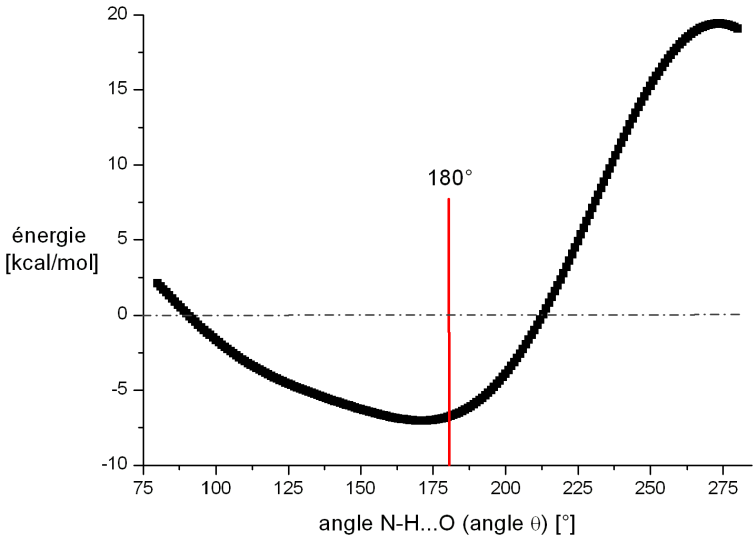


Figure 6.28 – Energie potentielle totale d'une liaison H en fonction de l'angle N-H...O (noté θ). La distance R est fixée à 1.8 Å, l'angle ϕ est fixé à 180°.

L'énergie moyenne ainsi que la somme de toutes les interactions respectives est résumée tableau 6.5.

Pour le modèle de Dahiyat *et al.*, les liaisons H de type $i \rightarrow i+4$ ont en

interaction	**moyenne** [kcal/mol]	**somme** [kcal/mol]	**modèle**
$i \rightarrow i+4$	-4.1	-504	
$i \rightarrow i+3$	-1.5	-55	Dahiyat
N...O	-0.5	-30	[74]
$i \rightarrow i+4$	-5,3	-646	
$i \rightarrow i+3$	-4,2	-151	Choi
N...O	-1.8	-96	[75]

Table 6.5 – Energie moyenne des différentes interactions dans les hélices de la protéine Beta-Glucosidase ainsi que leurs sommes respectives.

moyenne une énergie de -4.1 kcal/mol, ce qui correspond à -504 kcal/mol pour toutes les liaisons hydrogènes dans les hélices. Notons que cette valeur correspond à la moitié de l'énergie d'une liaison H parfaitement alignée (qui est de -8 kcal/mol). D'après le modèle de Choi *et al.*, l'énergie moyenne est

de -5.3 kcal/mol, ce qui est sensiblement plus élevé, et la somme est de -646 kcal/mol.
Les liaisons H $i \rightarrow i+3$, qui sont beaucoup moins nombreuses, sont moins énergétiques, la moyenne est de -1.5 kcal/mol pour le modèle de Dahiyat et de -4,2 kcal/mol pour le modèle de Choi. La différence est assez élevée, ce qui peut être expliqué par deux raisons : d'abord, la partie angulaire du potentiel décroit moins rapidement en fonction de l'angle θ pour le modèle de Choi. Ensuite, ce modèle n'inclut pas de terme qui pénalise une géométrie où les atomes ne sont pas alignés dans un plan, contrairement au modèle de Dahiyat. Le fait que les liaisons H $i \rightarrow i+3$ soient moins énergétiques est en accord avec la géométrie de ces liaisons, qui est caractérisée par des distances donneur-accepteur plus longues et des angles moins favorables.
Enfin, les interactions N...O dans les hélices, interprétés comme liaison hydrogène, donneraient -0.5 kcal/mol (modèle de Dahiyat) et -1.8 kcal/mol (modèle de Choi) chacune. Bien que l'énergie totale, qui est de -30 kcal/mol ou de -96 kcal/mol, soit beaucoup moins élevée que la contribution des liaisons H $i \rightarrow i+4$, elle n'est pas négligeable, l'énergie de Gibbs du repliement protéique étant de l'ordre de grandeur -5 à -15 kcal/mol [4]. Il est alors important de distinguer les interactions N...O d'une véritable liaison hydrogène, puisqu'elles peuvent influencer considérablement le bilan énergétique total d'une protéine.
Notons un autre point qui souligne l'importance de distinguer les liaisons H des interactions N...O. Les paramètres géométriques des liaisons hydrogènes dans les structures protéiques sont souvent utilisés afin d'obtenir les distances et les angles d'équilibres d'une fonction d'énergie potentielle (par exemple [75], [76], [77]). Si les interactions N...O sont inclues dans l'échantillon, elles vont influencer la valeur d'équilibre.

6.6.3 Liaisons hydrogène dans les feuillets

La figure 6.29 représente la courbure positive en fonction de la distance H...O pour les liaisons hydrogènes dans les feuillets parallèles et antiparallèles. Les distances de liaison dans les feuillets parallèles varient de 1.7 à 2.8 Å. Pour les feuillets antiparallèles, les distances H...O occupent une région entre 1.7 et 2.6 Å, mais on oberve que la densité de points est faible à partir de 2.2 Å. Ceci peut être dû au fait que le nombre de CP correspondant aux feuillets parallèles est le double de celui des feuillets antiparallèles, qui sont alors moins bien représentés dans l'échantillon.
La courbure positive en fonction de la distance H...LP pour les feuillets β est représentée figure 6.30. Les points occupent les mêmes régions, un ajustement de courbe séparé pour les feuillets parallèles et antiparallèles donne

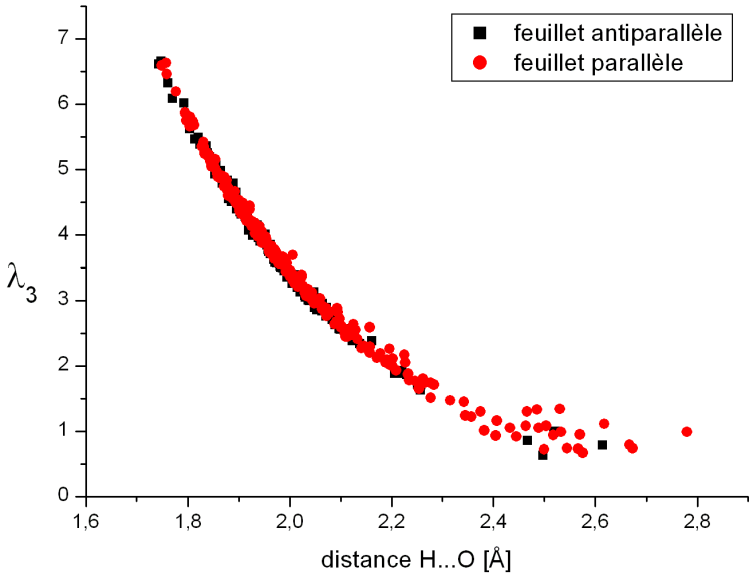


Figure 6.29 – Courbure positive de la densité électronique au point critique en fonction de la distance H...O des liaisons hydrogène dans les feuillets. Carrés noirs : feuillet antiparallèle. Cercles rouges : feuillet parallèle.

$a = 0.28(3)$, $b = 2.41(7)$ et $a = 0.21(3)$, $b = 2.28(10)$, respectivement. Les deux courbes ne se distinguent donc pas significativement et on ne retrouve pas le même comportement topologique que pour les hélices.
Notons que les distances moyennes de l'atome d'hydrogène par rapport au plan des paires libres de l'atome d'oxygène sont de 0.57(44) et de 0.64(49) Å pour les feuillets antiparallèle et parallèle, respectivement. Dans les feuillets les atomes d'hydrogènes sont alors plus proches du plan des paires libres que dans les hélices. Ceci est en accord avec les résultats de Thomas *et al.* [49], qui postulent que les liaisons hydrogènes sont canoniques dans les feuillets tandis qu'elles ne le sont pas dans les hélices.
Rappelons que les paramètres géométriques, donc les valeurs moyennes des angles et des distances, sont similaires pour les feuillets parallèles et antiparallèles. La figure 6.31 montre les angles H...O=C et N-H...O en fonction de la distance H...O. Les distributions ont le même comportement, ce qui souligne la nature similaire des liaisons hydrogène dans les feuillets antiparallèles et parallèles.

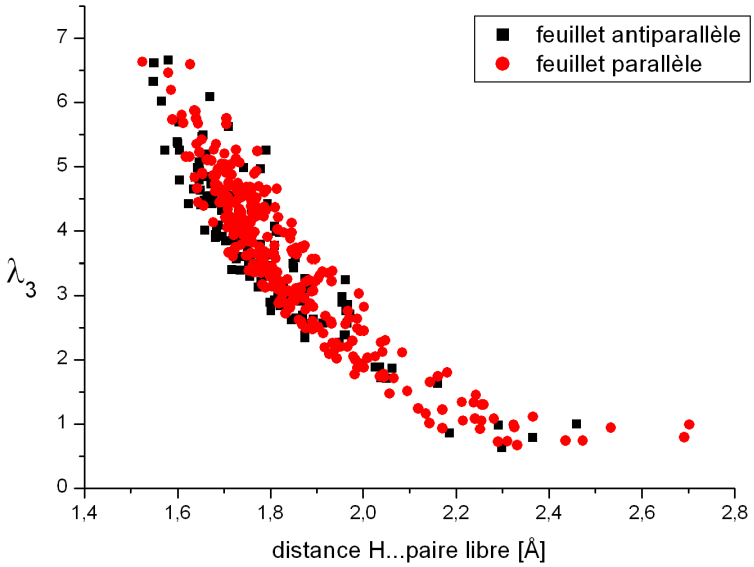


Figure 6.30 – Courbure positive de la densité électronique au point critique en fonction de la distance H...LP. Carrés noirs : feuillet antiparallèle. Cercles rouges : feuillet parallèle.

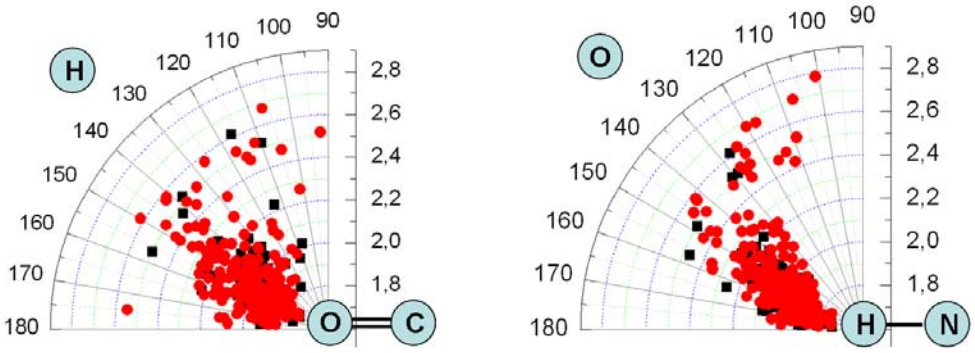


Figure 6.31 – Représentation polaire des paramètres géométriques. *A gauche* : L'angle H...O=C et la distance H...O pour les feuillets parallèles (cercles rouges) et antiparallèles (carrés noirs). Le rayon correspond à la distance H...O, l'angle polaire correspond à l'angle H...O=C. *A droite* : L'angle N-H...O et la distance H...O, même code couleurs.

6.7 Conclusion

L'analyse topologique de la densité électronique dans 6 protéines a montré que les groupements N-H et C=O de la chaîne principale dans les hélices peuvent interagir de 3 manières différentes. Ils forment des liaisons hydrogène entre des groupements éloignés de trois ($i \rightarrow i+3$) ou quatre acides aminés ($i \rightarrow i+4$) ou l'interaction a lieu entre les atomes d'azote et d'oxygène. Cette interaction N...O a été souvent interprétée comme liaison hydrogène dans la litérature.
Les trois interactions ont été analysées en terme de géométrie et de topologie de la densité électronique. De ce fait, les paramètres géométriques ainsi que topologiques sont caractéristiques de chaque interaction. Cette analyse a également permis d'identifier l'angle N-H...O comme critère qui permet de discriminer entre une liaison H et une interaction N...O.
L'analyse a pu montrer que les liaisons hydrogène dans les hélices adoptent soit la géométrie d'une liaison $i \rightarrow i+4$ soit celle d'une liaison $i \rightarrow i+3$ pur, donc on ne trouve pas de configuration qui est entre les deux géométries (en particulier pour les angles H...O=C et la distance H...plan, figure 6.19).
Si les interactions N...O sont interprétés comme liaison hydrogène, leur énergie totale représente une valeur non-négligeable (par exemple -30 kcal/mol pour la protéine Beta-Glucosidase) par rapport à l'énergie de repliement d'une protéine, supposée d'être de l'ordre de grandeur d'une liaison hydrogène.
Les liaisons hydrogènes dans les feuillets β ne se distinguent pas fondamentalement, leurs paramètres géométriques et leur topologie ont le même comportement. On ne trouve pas d'interaction N...O dans les feuillets, et ce en raison de leur géométrie.

6.8 Perspectives

Des études théoriques sur des fragments de structure secondaire ont menées à des énergies entre -4 et -10 kcal/mol pour les liaisons hydrogène [62], ce qui est plus élevé que les estimations sur la base de la fonction d'énergie potentielle utilisée dans cette étude. Il est alors possible que la fonction d'énergie potentielle sous-estime l'énergie des liaisons hydrogène. En utilisant les résultats ci dessus, il est envisageable de proposer une nouvelle fonction qui tienne compte de l'existence des interactions N...O, par exemple en imposant un cutoff de 120° pour l'angle N-H...O.
Il serait intéressant d'élargir l'étude sur les autres liaisons hydrogènes dans une protéine. En effet, les atomes d'hydrogènes des chaînes latérales sont éga-

lement impliqués dans des liaisons H. En plus, même les groupements C-H peuvent jouer le rôle de donneur afin de former des faibles liaisons hydrogène [78] qui peuvent être trouvées surtout dans les feuillets et dans les hélices [56]. Il serait alors particulièrement intéressant d'effectuer une étude topologique sur ces interactions faibles.
Pour les liaisons hydrogène dans les feuillets β, aucune différence a été faite dans la présente étude entre les liaisons bifurquées ou doubles. Il serait intéressant de les distinguer afin de savoir si leur géométrie ou leur topologie est différente.

Quatrième partie

Protéines diffractant à haute résolution

Introduction

La première structure tridimensionnelle d'une protéine, résolue par cristallographie, a été publiée en 1958 par John Kendrew. C'était la structure de la protéine myoglobine, à 6 Å de résolution [1]. Depuis, des progrès méthodologiques et technologiques ont fait évoluer la cristallographie macromoléculaire. On peut citer notamment les nouvelles méthodes de phasage, comme les méthodes directes [79] ou la dispersion anomale à plusieurs longueurs d'onde (*Multi-wavelength anomalous dispersion*). Du côté technologique, on peut citer le synchrotron qui permet de faire l'expérience avec des longueurs d'onde variables et avec un rayonnement de brilliance élevée, ou les progrès en informatique, qui ont facilité le traitement des données de diffraction et la visualisation des objets macromoléculaires. Ainsi, le nombre de structures de protéines résolues augmente constamment.
Alors que la grande majorité des protéines n'atteint actuellement [13] qu'une résolution de l'ordre de 2 à 3 Å, résolution à laquelle on peut déterminer l'allure générale de la chaîne d'acides aminés mais pas des détails de la structure, on peut observer que de plus en plus de protéines diffractent à résolution atomique ou subatomique.
Les données de diffraction à haute résolution offrent l'avantage d'analyser des détails assez fins de la structure d'une protéine. On peut citer notamment les points suivants :

- La densité électronique des atomes d'hydrogène peut apparaître dans les cartes de densité électronique (figure 6.32). Ainsi, des états de protonation ou l'orientation d'un groupement hydroxyl O-H par exemple (présents dans les acides aminés sérine, thréonine et tyrosine) peuvent être déduits. Ces informations sont primordiales pour la compréhension d'un mécanisme enzymatique, car les atomes d'hydrogène y sont souvent impliqués. En outre, en plaçant les atomes d'hydrogène, il est possible de déterminer l'orientation de molécules d'eau. Ceci permet

13. La résolution moyenne des structures déposées dans la PDB (structures résolues par rayons X ou par microscopie électronique) est de 2.2 Å au 9/02/2010.

de construire leur réseau de liaisons hydrogène, ce qui peut également jouer un rôle important pour le mécanisme d'une protéine.

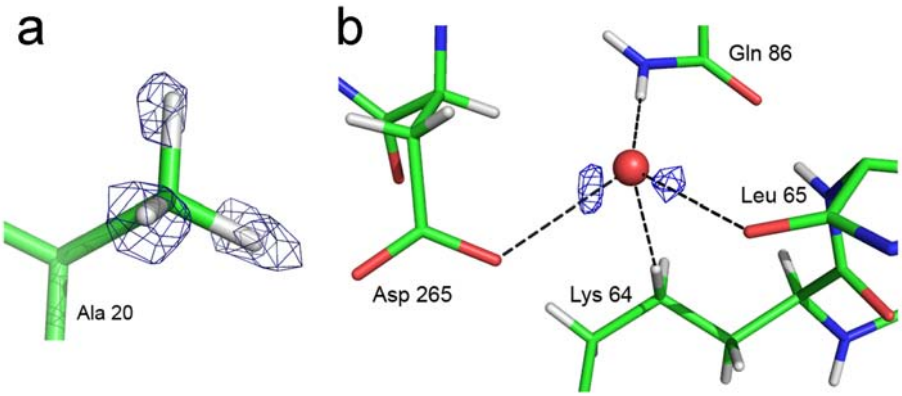


Figure 6.32 – (a) Groupement méthyl dans la protéine DFPase (0.85 Å de résolution) et carte de densité électronique résiduelle (en bleu) dans laquelle les atomes d'hydrogène ont été omis (la carte est contourée à 2.5σ). (b) Molécule d'eau dans la structure de PfluDING (0.98 Å de résolution) et carte de densité électronique omit H contourée à 2.9 σ en bleu. Le réseau de liaisons hydrogène est clairement visible.

- Les atomes dits "lourds" (espèces chimiques C,O,N,S) peuvent être placés précisément ce qui permet de déterminer les déviations par rapport à la stéréochimie standard. En effet, pour pouvoir effectuer l'affinement d'une structure à résolution moyenne, il est nécessaire d'apporter des informations supplémentaires et d'appliquer des contraintes ou des restrictions stéréochimiques qui imposent des distances interatomiques et des angles valenciels au modèle. Ces valeurs sont souvent obtenues par l'étude cristallographique par rayons X ou neutrons de petits composés d'acides aminés [50]. A haute résolution, il est possible de "relâcher" ces contraintes et les atomes sont placés sans biais. En conséquence, dans plusieurs structures à haute résolution, des déviations par rapport aux valeurs standards ont été observées [80].

- Les conformations multiples peuvent être observées et affinées. Dans des régions d'agitation thermique élevée, comme par exemple à la surface d'une protéine, il peut arriver que des acides aminés n'aient pas la même orientation dans toutes les mailles. Tandis que ce phénomène n'est pas visible à moyenne résolution, le signal de chaque conformation étant trop faible, on le trouve souvent dans les structures à haute résolution. Ainsi, à haute résolution, le désordre est plus visible. On peut même

observer des conformations multiples de la chaîne principale, comme par exemple dans les structures de la protéine aldose réductase (résolue à 0.66 Å) [7] ou de la protéine lysozyme (résolue à 0.65 Å) [81].

- Le potentiel électrostatique ainsi que les propriétés dérivées de la densité électronique précise peuvent être calculés à partir d'une densité électronique transférée. Dans ce cas, la haute résolution apporte la position des atomes d'hydrogène et la position précise des autres atomes de la protéine. Ceci est important puisque le potentiel électrostatique est très dépendant de la position des atomes du modèle. Le potentiel électrostatique peut apporter de précieuses informations sur l'interaction ligand-protéine. Par exemple, il a été montré pour la protéine aldose réductase que le potentiel électrostatique du ligand et du cofacteur $NADP^+$ de la protéine aldose réductase sont complémentaires [82].

De plus, à partir d'une résolution de $\sim$0.7 Å, dite "sub-atomique" ou "ultra-haute", le niveau de détails devient encore plus grand :

- Il est possible d'observer la densité électronique de déformation, comme par exemple la densité des paires libres d'électrons ou la densité sur les liaisons covalentes. Dans ce cas, il devient envisageable d'effectuer un affinement multipolaire de la densité électronique, comme c'est le cas pour la protéine aldose réductase [3]. Par contre, un tel affinement est effectué seulement dans une partie de la protéine, par exemple dans le site actif, puisque le nombre de paramètres à affiner augmente considérablement. De plus, il faut que l'agitation thermique des atomes soit très basse ($\approx 5\text{Å}^2$), sinon la densité de déformation est diluée et donc peu visible.

- Un affinement à grand angle (high order refinement), c'est-à-dire contre les données de haute résolution uniquement, devient envisageable. Ce type d'affinement sert à déterminer la position précise des atomes lourds et à mieux décrire leur agitation thermique. En effet, les rayons X interagissent avec les électrons de cœur ainsi qu'avec le cortège électronique de valence. A basse résolution, la contribution des électrons de valence au facteur de diffusion est plus élevée que celle des électrons de cœur [14]. A haute résolution, la situation est inversée, c'est-à-dire ce sont les électrons de cœur qui ont une contribution plus élevée (fig 6.33). Ainsi, la position du noyau d'un atome, trouvée par affinement contre les données de toute la gamme de résolution, peut être biaisée [83]. En plus,

14. Ceci est valable si l'agitation thermique de l'atome n'est pas trop élevée.

les ellipsoïdes d'agitation thermique peuvent être allongés artificiellement pour tenir compte de la densité de déformation. En affinant les coordonnées atomiques (non atomes hydrogènes) contre les données de diffraction de très haute résolution (∼0.8 Å), on inclut uniquement le signal des électrons de cœur, ce qui correspond mieux à la position du noyau atomique. Ainsi, l'effet de la densité de déformation peut être déconvolué de celui de l'agitation thermique. L'affinement à grand angle est souvent utilisé pour les composés de petite taille, mais il a déjà été appliquées aux protéines [81] [3].

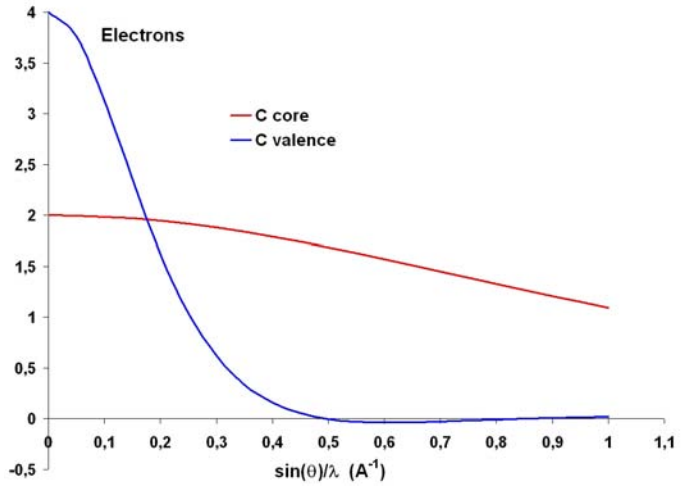


Figure 6.33 – Facteur de diffusion statique en fonction de $\frac{sin\theta}{\lambda}$ pour les électrons de cœur (en rouge) et les électrons de valence (en bleu) de l'atome de carbone. Le facteur de diffusion des électrons de valence tend plus rapidement vers zéro que celui des électrons de cœur quand $\frac{sin\theta}{\lambda}$ augmente. Notons que la densité de déformation n'est visible qu'à résolution subatomique et, cependant, sa contribution au facteur de diffusion est élevée à basse résolution.

Ce chapitre portera sur les deux protéines PfluDING et DFPase qui illustrent comment des données de haute résolution peuvent apporter des informations précieuses pour la compréhension de mécanismes moléculaires. Les deux études ont fait l'objet de publications [84] [85].

Remarque : Dans ce chapitre, le texte se réfèrera plusieurs fois aux cartes de densité (électroniques ou nucléaires) de différents types, comme par exemple $F_{obs} - F_{calc}$ ou "carte de densité omit H". La différence entre ces cartes de densité est rappelée dans section A.3.

Chapitre 7

La protéine DFPase

Dans le cadre d'une collaboration entre J. Koepke de l'institut Max-Planck de biophysique à Francfort, E. Chabrière de l'institut AFMB et l'équipe Modélisation du laboratoire CRM^2, nous disposons des facteurs de structure et du modèle de la protéine DFPase, issus d'une expérience de diffraction des rayons X (0.85 Å de résolution). L'idée initiale était d'effectuer un affinement multipolaire d'une partie de la structure, profitant de la résolution élevée des données de diffraction. De plus, au cours de ma thèse, la structure de la DFPase issue d'une expérience de neutrons a été publiée (2.2 Å de résolution) [86]. Ainsi, il a été possible de confronter les résultats obtenus sur la base de l'analyse de la structure à RX avec ceux issus des données de diffraction des neutrons.

7.1 Introduction

La protéine Diisopropylfluorophospatase (DFPase) a été initialement isolée à partir du ganglion nerveux du calmar *Loligo vulgaris*. Elle se compose de 314 acides aminés et compte 35 kDa. Le rôle naturel de la DFPase est jusqu'à maintenant inconnu mais elle est capable d'hydrolyser des agents neurotoxiques (organophosphorés) comme le soman, le sarin, le tabun et le DFP (figure 7.1).
Les organophosphorés (OP) sont des substances très toxiques qui résultent de la réaction entre l'acide phosphorique et des alcools. Pour leur pouvoir toxique, ils sont utilisés comme pesticides ou armes chimiques : ils inhibent l'acétylcholinestérase, une protéine dégradant le neurotransmetteur acétylcholine. Si l'acétylcholine n'est pas dégradé, des impulsions nerveuses (par exemple entre un nerf et un muscle) sont transmises constamment, ce qui

entraîne des convulsions et peut aboutir à la mort.

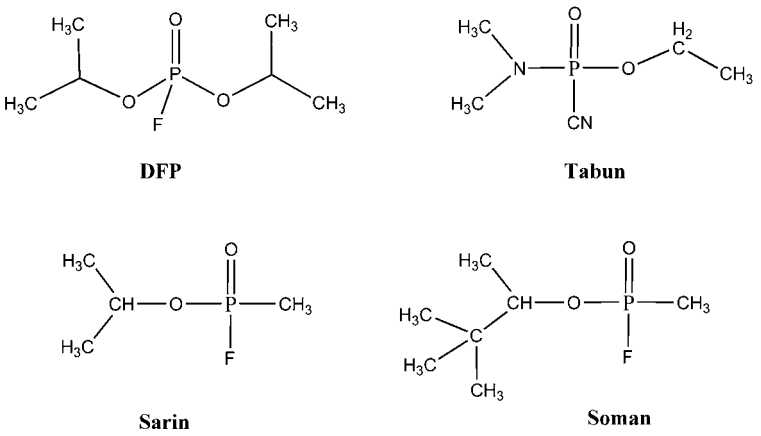


Figure 7.1 – Quelques représentants des organophosphorés, utilisés comme agents neurotoxiques de guerre.

La figure 7.1 montre quelques représentants des organophosphorés. Les dérivés comme le sarin, tabun, soman ou VX sont des agents chimiques de guerre. Lors de la guerre froide, les Etats-Unis et la Russie en ont produit de grandes quantités, et jusqu'à présent il en existe des stocks dans le monde entier.
Pour cette raison, de nombreuses recherches sont effectuées pour trouver une possibilité de détruire le pouvoir toxique des organophosphorés. Bien qu'il y ait des méthodes décontaminantes, comme par exemple la chaleur ou l'utilisation de soude, elles ne peuvent pas être appliquées en grande quantité sur l'homme et nuisent souvent à l'environnement [87]. Il existe en fait une alternative plus douce : des enzymes sont en effet capables de dégrader les agents neurotoxiques. De plus, il est intéressant d'utiliser un procédé enzymatique puisqu'on a seulement besoin d'une petite quantité de la protéine qui est régénérée après chaque catalyse.
La structure de la DFPase a été résolue par diffraction des rayons X [88]. Sa topologie ressemble à une hélice de bateau[1] avec 6 feuillets β arrangés autour d'un tunnel qui est rempli de molécules d'eau (figure 7.2) et où se trouvent deux cations calcium. Le premier ion calcium a une affinité élevée et est enfoui à l'intérieur du tunnel. Il est supposé avoir un rôle structural [89]. Le deuxième ion calcium est impliqué dans le mécanisme d'hydrolyse et a une affinité plus faible [89].
Il y a plusieurs propositions sur le fonctionnement de la DFPase dans son rôle de dégradation des OP : Le premier mécanisme (figure 7.3) propose l'attaque nucléophile du substrat par le groupement carboxyle d'un résidu voisin,

1. En anglais, le repliement est appelé *β-propeller*.

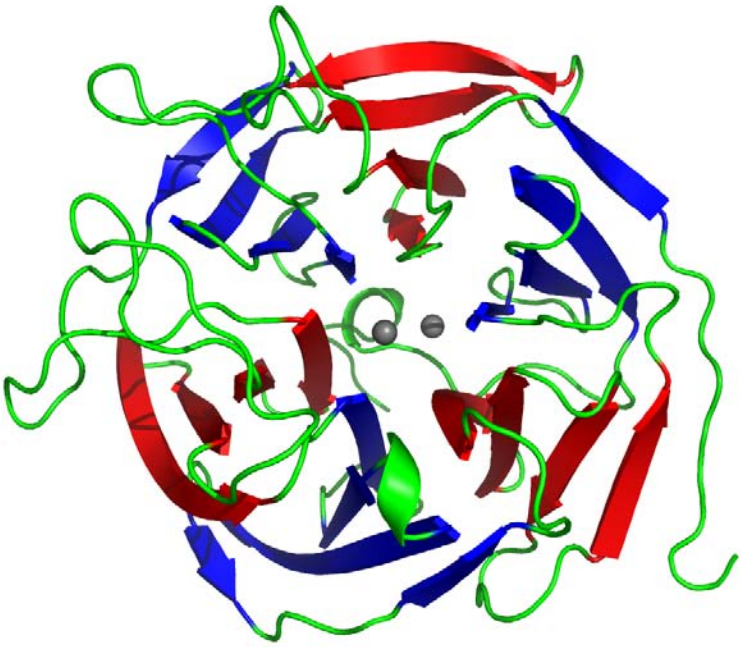


Figure 7.2 – Représentation de la protéine DFPase. La chaîne principale est dessinée en vert. Pour une meilleure visualisation, les 6 feuillets beta sont représentés en bleu et en rouge. Les deux ions calcium, situés dans le tunnel central, sont figurés comme sphères grises.

Asp 229. Il se crée un intermédiaire covalent phospho-enzyme et le fluor du substrat est libéré sous forme F^-. La régénération de l'enzyme est ensuite assurée par une molécule d'eau du solvant. Le second mécanisme (figure 7.4) implique l'activation d'une molécule d'eau par le calcium catalytique. Cette fois-ci, c'est l'ion hydroxyde résultant de cette activation qui prend le rôle du nucléophile et qui attaque le substrat.

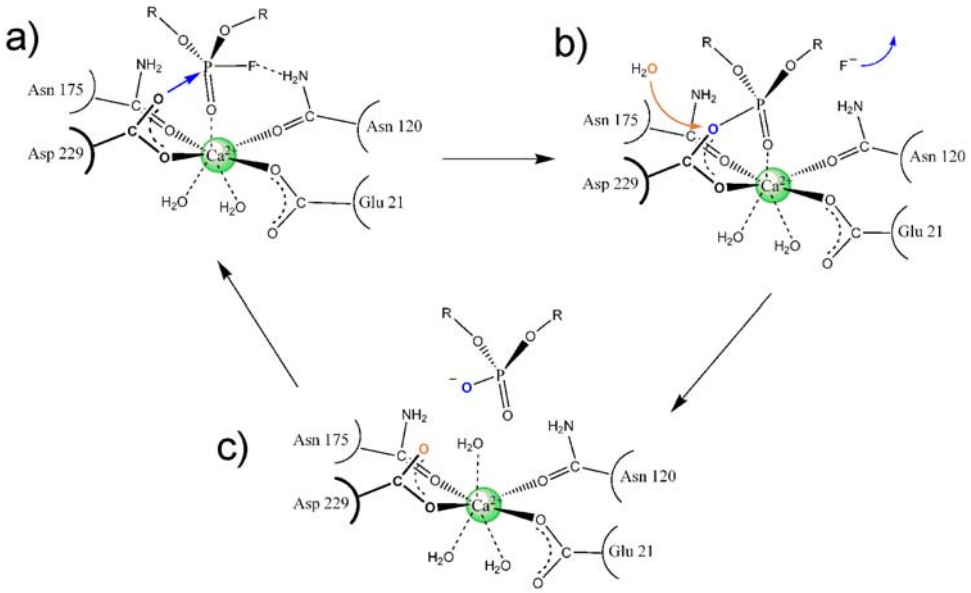


Figure 7.3 – Premier mécanisme d'hydrolyse proposé. (a) Le substrat arrive dans le site actif et se coordonne au calcium catalytique. (b) L'oxygène du résidu Asp 229 attaque le phosphore du substrat et il se crée un intermédiaire phospho-enzyme. En même temps, le fluor se libère. Une molécule d'eau attaque ensuite l'acide aspartique et (c) la protéine se régénère en libérant le produit Diisopropyl-phosphate.

Pour discriminer les deux mécanismes, la connaissance précise des états de protonation dans le site actif est nécessaire. L'étude des données de diffraction à haute résolution pourrait alors donner de précieuses informations sur le mécanisme catalytique.
Avant de passer à la partie expérimentale de ce projet ainsi qu'à la comparaison des modèles à RX et neutrons, quelques principes de base d'une expérience de diffraction des neutrons seront rappelés.

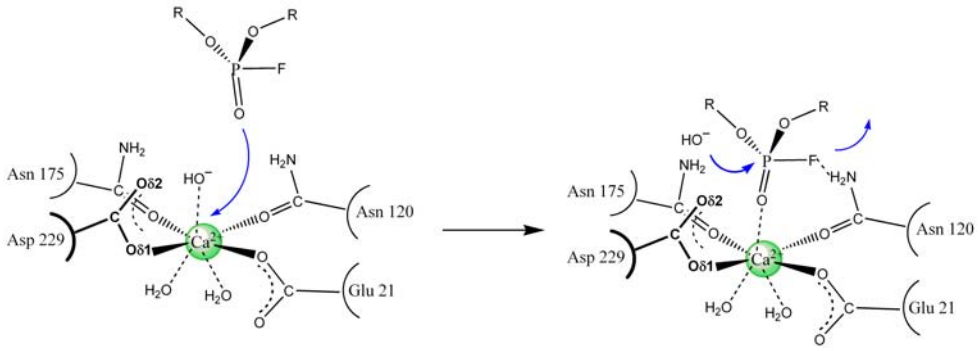


Figure 7.4 – Deuxième mécanisme d'hydrolyse proposé. Au début, il se trouve déjà un ion hydroxyde dans le site actif, coordonné à l'atome de calcium. Le substrat arrive dans le site actif, se coordonne au calcium catalytique et est attaqué par l'ion hydroxyde. Le fluor est libéré et la protéine se régénère.

7.2 La diffraction des neutrons

Le principe de base d'une expérience de diffraction des neutrons est similaire à celui des rayons X : Un faisceau de neutrons est envoyé sur un cristal et l'intensité des rayons diffractés est mesurée. Par contre, les neutrons n'interagissent pas avec le nuage électronique mais avec le noyau de l'atome, donc la section efficace d'interaction atome-neutron est indépendante du nombre d'électrons de l'atome. Chaque isotope d'une espèce chimique interagit alors différemment avec les neutrons (table 7.1), c'est pourquoi la diffraction des neutrons est intéressante pour la détermination d'une structure de protéine. Ainsi, l'atome d'hydrogène ou de deutérium ont une section efficace de l'ordre de grandeur de celle des espèces carbone, azote et oxygène, ce qui n'est pas le cas pour la diffraction des rayons X où le facteur de diffusion atomique augmente avec le nombre d'électrons de l'atome. Des espèces chimiques comme l'hydrogène ont un facteur de diffusion faible, ce qui se traduit par un faible pic de densité dans des cartes de densité électronique et donc par la difficulté d'observer les atomes H à basse résolution. La diffraction des neutrons est alors considérée comme méthode expérimentale de choix permettant de déterminer la position des atomes d'hydrogènes dans une structure cristallisée.

Espèce chimique	**longueur de diffusion [fm]**
O	5.803(4)
H	-3.7406(11)
D	6.671(4)
N	9.360
C	6.6511(16)
S	2.804(2)
Ca	4.99(3)

Table 7.1 – Longueur de diffusion des neutrons de quelques atomes [90]. Elle varie de façon non-linéaire et peut avoir une valeur négative.

Néanmoins, il y a un inconvénient : la longueur de diffusion des atomes d'hydrogène est négative et sa diffusion incohérente est élevée. C'est pourquoi on cherche souvent à échanger les atomes d'hydrogène par leur isotope deutérium. Ceci est possible soit par la production de la protéine dans un milieu deutéré (toutes les étapes d'expression, de purification et de cristallisation se font avec des solutions deutérées), soit par le trempage d'un cristal de protéine dans une solution deutérée. En l'occurrence, dans le cas du trempage, ce ne sont que certains atomes d'hydrogènes qui sont susceptibles de

s'échanger, comme ceux des groupements N-H ou O-H. Par contre, les atomes d'hydrogène liés à un carbone ne s'échangent pas facilement. Il s'en suit que la densité positive du carbone et la densité négative de l'atome d'hydrogène se chevauchent et que le signal du groupement peut devenir inférieur au bruit. En particulier, les groupements CH_2 sont souvent invisibles dans les cartes de densité nucléaire [2].

7.3 Manipulations et résultats

Pour alléger l'écriture, les abréviations suivantes seront utilisées :

carte omit X carte de densité omettant l'atome "X"

structure$_{RX}$ structure de la protéine DFPase, obtenue par diffraction des rayons X

structure$_{N}$ structure de la protéine DFPase, obtenue par diffraction des neutrons

7.3.1 Affinement de la structure obtenue par diffraction des rayons X

Les données de diffraction ont été collectées sur la ligne de lumière BW7B, au synchrotron Hasylab à Hamburg (Allemagne). Pour améliorer la mosaïcité et pour pousser la limite de résolution, les cristaux ont été soumis au processus dit de "annealing" [3]. Notons que 4 cristaux différents ont été utilisés pour avoir un jeu complet des données de diffraction. Les conditions expérimentales et quelques données cristallographiques sont résumées table 7.2.
Le modèle fourni dans le cadre de la collaboration avec J. Koepke était issu d'un affinement restreint conventionnel effectué avec SHELXL [91], les atomes d'hydrogène des acides aminés étaient placés dans ce modèle. Parmi ces atomes d'hydrogènes, un grand nombre est clairement visible dans des cartes de densité électronique omit H (figure 7.5 a), ce qui démontre la qualité des données de diffraction. Des tests d'affinement multipolaire ont été effectués dans la région du site actif avec le logiciel MoPro mais n'ont pas abouti à un résultat satisfaisant car ils menaient à des paramètres multipolaires irréalistes. Une raison possible est que le mouvement thermique des

2. Le rapport de la longueur de diffusion du carbone et de l'atome d'hydrogène est $-\frac{2}{1}$.
3. Le cristal a été soumis plusieurs fois à un changement de température entre 120 K et température ambiante.

longueur d'onde	0.84 Å
température	120 K
résolution	0,85 Å
complétude	93.8 %
groupe d'espace	$P2_12_12_1$
a,b,c [Å]	43.114, 81.849, 86.467
molécules d'eau placées	496
R_{cryst} (SHELXL)	11,1 %
R_{free} (SHELXL)	12,8 %

Table 7.2 – Récapitulatif des conditions expérimentales ainsi que quelques statistiques sur les données de diffraction (RX) de la protéine DFPase.

atomes dans la protéine soit trop élevé, ce qui a pour conséquence de masquer le caractère non-sphérique de la densité électronique.
En revanche, les paramètres de solvant (modèle de solvant exponentiel [92]) ont pu être affinés avec MoPro et le modèle a été amélioré : Le facteur R a baissé de 11.1% à 10.3%, parallèlement le facteur R_{free} a diminué de 12.8% à 12.1%. Dans les cartes de densité électronique résiduelle $F_{obs} - F_{calc}$ qui ont été calculées avec le nouveau modèle structural, des pics de densité correspondant aux atomes d'hydrogène des molécules d'eau sont devenus clairement visibles (figure 7.5 b).
Par conséquent, les positions de nombreux atomes d'hydrogène des molécules d'eau dans la structure de la DFPase ont pu être déterminées expérimentalement, d'après des pics de densité électronique significatifs. Le noyau de l'atome d'hydrogène a été positionné au maximum du pic de densité résiduelle, en utilisant le logiciel Coot [93]. Notons qu'il n'était pas nécessaire d'utiliser des considérations stérérochimiques pour placer les atomes H, c'est-à-dire qu'il ne fallait pas analyser les acides aminés voisins pour chercher des donneurs ou des accepteurs de liaisons H permettant de conclure indirectement sur la position d'un atome d'hydrogène. Les atomes d'hydrogène ont alors été placés d'après le signal expérimental, mais notons que cela ne correspond pas à la véritable position du noyau (voir aussi section 7.3.3). En somme, 31 atomes d'hydrogènes ont été ajoutés de cette manière, dans le tunnel central et autour du calcium catalytique.
Connaissant la position d'un atome d'hydrogène, on peut en déduire les interactions dans lesquelles il est impliqué. Ainsi, il a été possible de déterminer le réseau de liaisons hydrogènes formé par les nombreuses molécules d'eau dans le tunnel central de la DFPase (figure 7.5 b).
De manière surprenante, l'orientation des molécules d'eau dans la structure

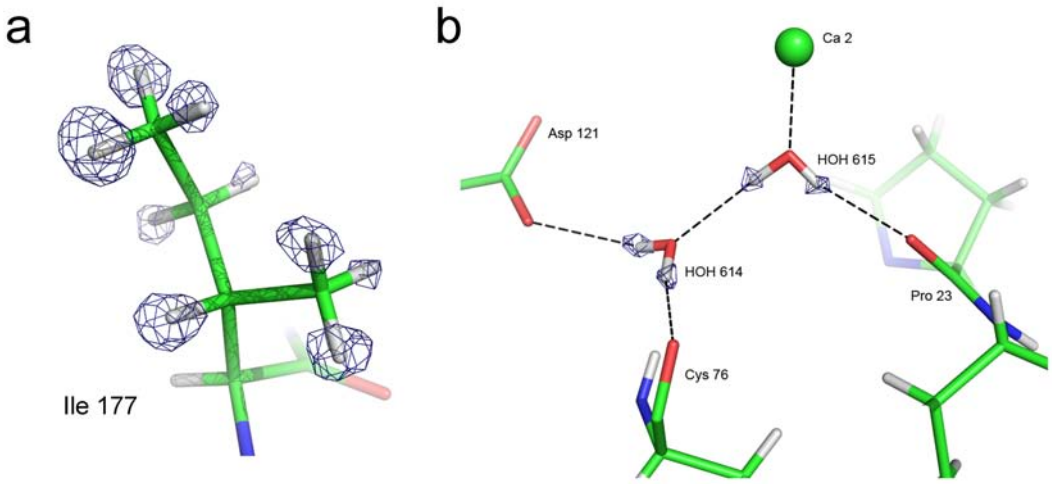


Figure 7.5 – Densité électronique résiduelle observée par diffraction des rayons X révélant la position de quelques atomes d'hydrogène. a) Résidu Ile 177 et carte résiduelle. b) Molécules d'eau 614 et 615 ainsi que les liaisons hydrogènes, représentées en ligne pointillée. Le niveau de contour des cartes de densité est 2.7σ dans a) et 2.5σ dans b), les atomes d'hydrogène ont été omis pour le calcul des cartes.

obtenue par diffraction des rayons X n'est pas toujours la même que celle visible dans la structure déterminée par diffraction des neutrons (figure 7.6) [4]. Pour cette raison, la structure neutrons a été analysée plus en détails afin d'expliquer ces déviations. La première étape consiste à calculer des cartes de densité nucléaire à partir des facteurs de structure et du modèle déposés sur la PDB (code 3BYC).

7.3.2 La densité nucléaire de la DFPase

L'expérience de diffraction de neutrons a été effectuée au Los Alamos Neutron Science Center (LANSCE) [94]. Le cristal utilisé pour l'expérience a été deutéré par trempage pendant une semaine. La structure a été affinée avec le logiciel CNS [95] dans un processus d'affinement joint, c'est-à-dire en utilisant simultanément les données de diffraction des neutrons et des rayons X [5]. De nombreux atomes d'hydrogènes des molécules d'eau sont présents dans la structure déposée sur la PDB. Il a été supposé par Blum *et al.* que

4. En effet, les expériences ont été effectuées à différentes températures. Pour cette raison, l'agitation thermique des atomes d'oxygènes de certaines molécules d'eau a été comparée dans section 7.3.4

5. de résolution de 1,84 Å [86]

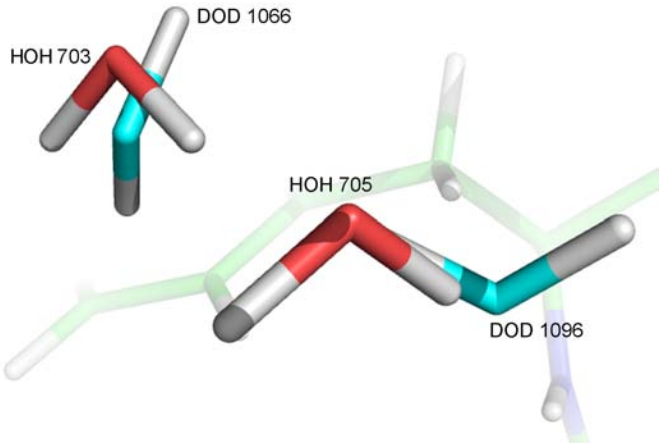


Figure 7.6 – Superposition des structures issues d'une expérience des rayons X et des neutrons, respectivement. Les molécules d'eau vues par diffraction des rayons X (rouge-blanc, notation HOH) et par diffraction des neutrons (cyan-blanc, notation DOD) ont des orientations différentes.

tous les atomes H des molécules d'eau se sont échangés contre des atomes de deutérium. Les conditions expérimentales et quelques données cristallographiques sont résumées table 7.3.

longueur d'onde	0,8 - 4,5 Å
taille du cristal	0.43 mm^3
température	293 K
résolution	2,2 Å
complétude	82 %
groupe d'espace	$P2_12_12_1$
a, b, c [Å]	43.441, 83.291, 87.509
molécules d'eau placées	738
molécules d'eau (D_2O)	246
R_{cryst}	26,4 %
R_{free}	31,6 %

Table 7.3 – Récapitulatif des conditions expérimentales (diffraction de Laue) ainsi que quelques statistiques sur les données de diffraction des neutrons de la protéine DFPase.

Les phases ont été calculées avec le programme *Phenix* v.1-4.138 [96], après un cycle d'affinement du facteur d'échelle et des paramètres de solvant. La

résolution et le seuil des données de diffraction ($\frac{F}{\sigma(F)}$), telles qu'elles ont été communiquées dans la publication correspondante [86], ont été appliquées.
Il peut être constaté que la densité nucléaire est moins riche en détail que la densité électronique. En général, les pics de densité nucléaire ne peuvent pas être attribués incontestablement à un certain atome. Entre autres, l'étalement de la densité peut être attribué à la résolution moyenne des données de diffraction et au mouvement thermique des atomes[6].
On peut observer également les effets du trempage du cristal : Les atomes d'hydrogène liés aux carbones ne s'étant pas échangés, la densité nucléaire des groupements CH_2 n'est souvent pas visible (figure 7.7 a).

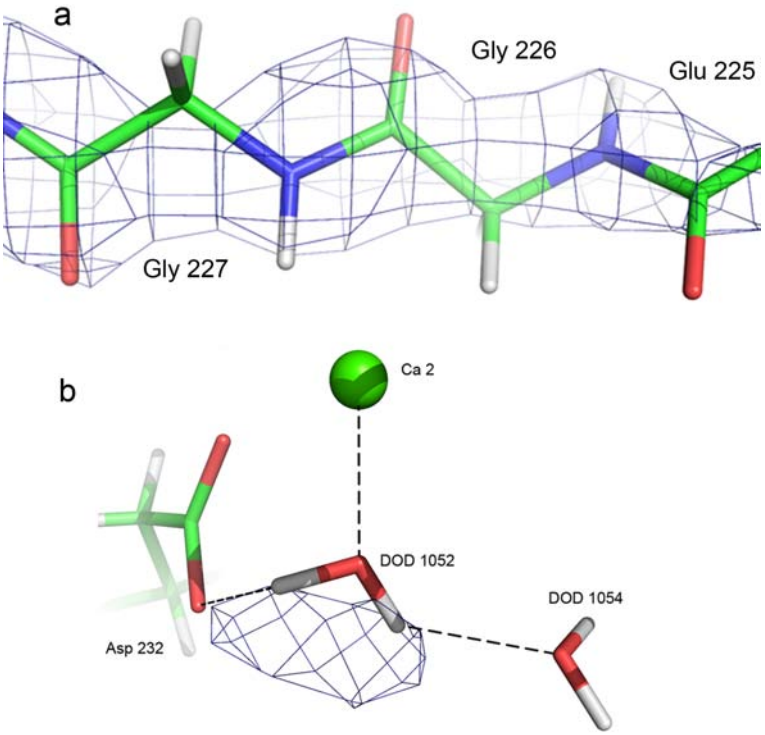


Figure 7.7 – a) Chaîne principale des résidus 225 à 227 et carte de densité nucléaire $2F_o - F_c$ contourée à 1.5 σ. b) Molécule d'eau DOD 1052 et réseau de liaison hydrogène possible (ligne pointillée) et $F_{obs} - F_{calc}$, contourée à 3.9σ, omettant les deux atomes de deutérium.

La densité nucléaire des molécules d'eau se présente souvent sous une forme ellipsoïdale. Il est donc difficile de déterminer l'orientation exacte de la molécule à partir de ce pic puisque sa forme peut être également influencée par l'agitation thermique. Pour cette raison, des cartes de densité nucléaire "omit

6. Rappel : L'expérience a été effectuée à température ambiante.

D" ont été calculées (figure 7.7 b). Dans ces cartes, la densité générée par les atomes D de la molécule d'eau devrait être visible. Il s'avère que la densité nucléaire des deux atomes de deutérium mène à un unique pic de densité résiduelle. Par conséquent, il est difficile de déterminer la position exacte d'un atome D à partir d'un pic qui résulte de la superposition de deux atomes.
La difficulté de placer correctement les atomes de deutérium d'une molécule d'eau (et de déterminer ainsi son orientation) uniquement sur la base de la densité nucléaire, est appuyée par l'observation de nombreuses distances courtes H...H, qui ne sont pas conformes à la stéréochimie standard des liaisons hydrogène. Par exemple, la molécule DOD 1084 est orientée de telle façon que ses deux atomes de deutérium sont proches des atomes d'hydrogène de la chaîne principale, eux-mêmes clairement localisés (figure 7.8). Les distances H...H sont de 1.6 et 2.1 Å, respectivement. De plus, le maximum du pic de densité nucléaire est localisé à côté de la molécule, hors du plan D-O-D, où il est attendu théoriquement.

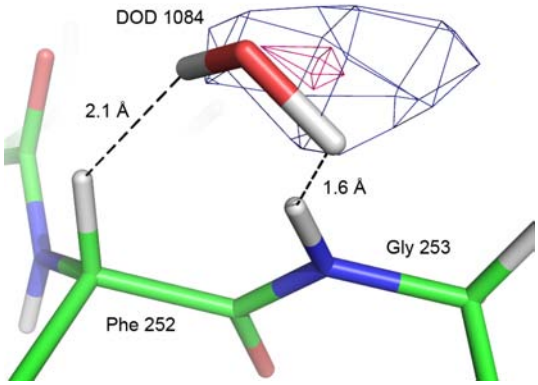


Figure 7.8 – Molécule d'eau DOD 1084 et carte de densité nucléaire $F_o - F_c$, omettant la molécule entière (atomes H et O). Les niveaux de contour sont 2.2 σ (bleu) et 2.65 σ (magenta), respectivement.

En résumé, la densité nucléaire de la protéine DFPase ne justifie souvent pas l'orientation des molécules d'eau dans le modèle. En particulier, on peut observer plusieurs distances courtes H...H. Ensuite, dans les cartes de densité nucléaire, il y a un seul pic correspondant aux atomes de deutérium des molécules d'eau au lieu de deux pics distincts. Ainsi, la localisation exacte d'un atome de deutérium n'est pas facile à réaliser.
Ces observations n'ont pour le moment qu'un caractère qualitatif. Afin de quantifier ces résultats, une analyse des géométries des liaisons hydrogènes a été effectuée pour les structures obtenues par diffraction des RX et neutrons.

7.3.3 Analyse des liaisons hydrogène

Avant de faire l'analyse, la géométrie des molécules d'eau obtenues après construction manuelle dans la structure à rayons X à été optimisée pour que les angles et distances de liaisons soient conformes aux valeurs standards. Plus précisément, l'angle O-H-O a été contraint à 104.5°. La distance O-H d'une molécule d'eau, observée par diffraction des rayons X, est normalement égale 0.81 Å [97]. Afin de pouvoir comparer la distance donneur-accepteur des liaisons hydrogènes dans la structure à RX et à neutrons, l'atome d'hydrogène a été déplacé le long de la liaison O-H afin de l'allonger à 0.96 Å [97], ce qui correspond à la distance O-H observée par diffraction des neutrons. Cette différence est observée car le seul électron de l'atome H est impliqué dans la liaison covalente avec l'atome d'oxygène de la molécule d'eau. La densité électronique est donc décalée en direction de la liaison et le pic de densité de l'électron n'est pas situé au même endroit que le noyau de l'atome d'hydrogène. Par conséquent, les distances O-H, comme toutes les distances C-H, apparaissent systématiquement plus courtes par diffraction des rayons X que par diffraction des neutrons. Après élongation, la position de l'atome d'hydrogène n'est plus biaisée par l'électron partagé dans la liaison, ce qui rend possible la comparaison des longueurs des liaisons hydrogène.
L'analyse a été réalisée avec le logiciel MoPro. Seules les liaisons hydrogènes H...O ont été analysées, où l'atome d'hydrogène H appartient à une molécule d'eau et l'oxygène O fait partie soit de la protéine soit d'une autre molécule d'eau. Si la distance entre les deux atomes est plus courte que 2.5 Å, et si l'angle O-H...O est plus grand que 90°, l'interaction a été interprétée comme liaison hydrogène [7].
L'analyse a été effectuée sur les molécules d'eau de la structure résolue par diffraction des RX dont au moins un atome d'hydrogène a été localisé par la méthode décrite dans la section 7.3.2. Dans la structure obtenue par diffraction des neutrons, seules ont été incluses les molécules D_2O, dont les atomes d'oxygène se trouvent à la même position que celle utilisée dans la structure$_{\mathrm{RX}}$. De cette manière, 30 et 38 liaisons hydrogène ont été trouvées pour les molécules d'eau dans la structure$_{\mathrm{RX}}$ et la structure$_{\mathrm{N}}$, respectivement. D'après les résultats de l'analyse, la moyenne des distance H...O dans la structure obtenue par diffraction des RX est de 1.93(18) Å, ce qui est plus court que celle observée dans la structure$_{\mathrm{N}}$, où la moyenne est de 2.12(19) Å. La moyenne des distances RX est certes dans l'intervalle de dispersion des distances neutrons, mais la distribution est différente, favorisant des distances

7. La subroutine writehbond.f de MoPro a été changée pour écrire les valeurs géométriques de ces liaisons hydrogènes sous forme de tableau, permettant le traitement avec un tableur (Origin).

donneur-accepteur plus courtes dans la structure$_{RX}$ (figure 7.9) [8]. Cette déviation n'est pas due à la différence intrinsèque entre les deux techniques puisque les distances O-H des molécules d'eau ont été allongées dans la structure$_{RX}$.

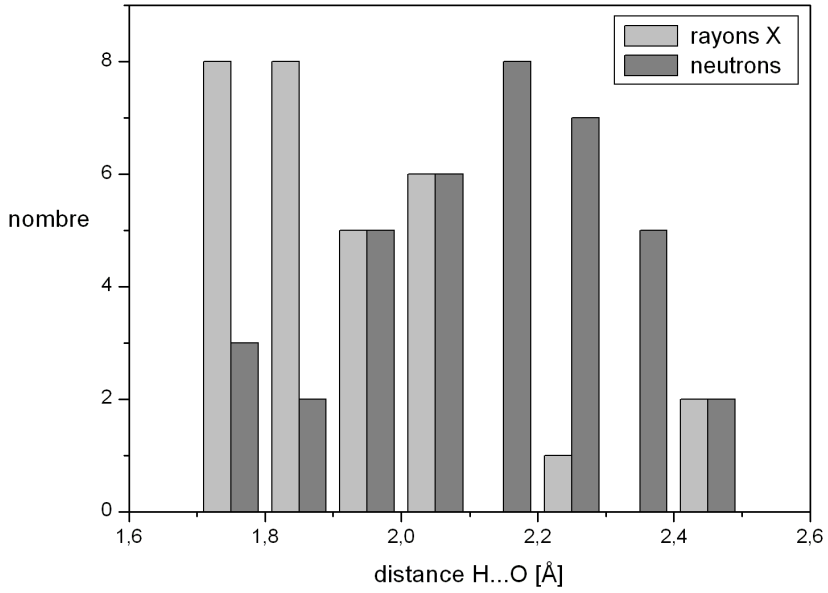


Figure 7.9 – Histogramme des distances donneur-accepteur (H_w...O) des liaisons hydrogènes impliquant des molécules d'eau dans la structure à rayons X (gris clair) et neutrons (gris foncé).

Une analyse similaire a été effectuée pour dénombrer les distances courtes hydrogène-hydrogène dans la structure obtenue par diffraction des neutrons [9]. Parmi ces interactions, 59 contacts courts entre deux atomes d'hydrogène ont été trouvés, avec une distance H...H inférieure à 2 Å.
La molécule d'eau HOH 511 illustre les différences trouvées entre les structures obtenues par diffraction des RX et des neutrons (figure 7.10). HOH 511 convient bien à son environnement chimique. Ses deux atomes d'hydrogène pointent en direction d'atomes d'oxygène (un glutamate et une autre molécule d'eau) et forment clairement des liaisons hydrogène. En revanche, les deux atomes D de la molécule d'eau DOD 1010, dont l'atome d'oxygène est à la position correspondante dans la structure obtenue par diffraction des neutrons, ont été placés différemment. Il n'y a pas d'accepteurs de liaison hydrogène dans le voisinage, et les atomes de deutérium sont positionnés à

8. Les σ correspondent à la dispersion des distances H...O autour de la valeur moyenne.
9. La subroutine writehbon.f de MoPro a été modifiée pour donner les contacts H...H, avec les mêmes contraintes géométriques que mentionné ci-dessus.

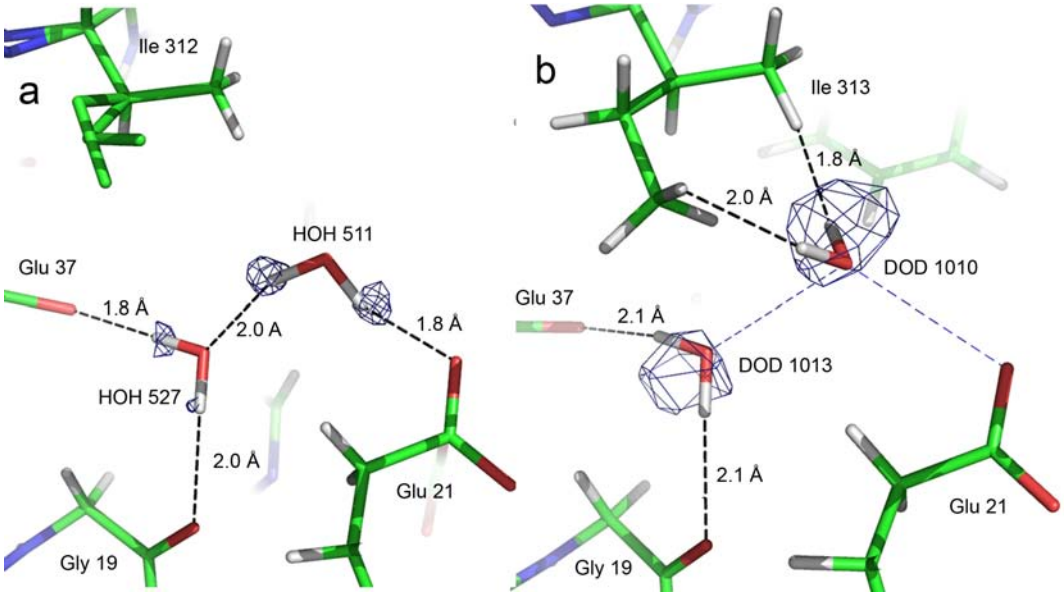


Figure 7.10 – (a) Vue de la molécule d'eau HOH 511 dans la structure obtenue par diffraction des rayons X et carte de densité résiduelle $F_{obs} - F_{calc}$ omettant les atomes d'hydrogènes, contourée à $2,4\sigma$. *Remarque* : Le résidu Ile 312 étant désordonné dans la structure à RX, les atomes d'hydrogène n'ont pas été placés. (b) Vue de la molécule d'eau DOD 1010 dans la structure obtenue par diffraction des neutrons et carte de densité nucléaire $2F_{obs} - F_{calc}$ contourée à $1,4\sigma$. L'atome d'oxygène de DOD 1010 est à la même position que celui de la molécule HOH 511. Pour comparer, les liaisons hydrogènes de HOH 511 sont représentées comme fine ligne pointillée bleue.

proximité d'autres d'atomes H de la protéine. Les distances D...H sont de 1.8 et 2.0 Å, respectivement.

7.3.4 Origine des différences entre le réseau de liaisons hydrogène des molécules d'eau

Les déviations de la géométrie des liaisons hydrogène impliquant les molécules d'eau peuvent trouver leur origine dans une différence intrinsèque entre les informations venant des neutrons ou des rayons X, dans la qualité des données de diffraction ou dans le dispositif expérimental (p.ex. température).
Concernant le dispositif expérimental, les expériences ont été effectuées à différentes températures, 100 K pour les RX, et température ambiante (298 K) pour les neutrons. Le désordre statique et dynamique est alors plus élevé dans la structure$_{\mathrm{N}}$ et il est possible que quelques molécules d'eau soient plus agitées et moins confinées dans une liaison hydrogène. Afin de vérifier si la différence d'orientation est corrélée à l'agitation thermique des molécules d'eau, les facteurs d'agitation thermique isotrope équivalents (B_{eq}) de leurs atomes d'oxygène ont été calculés (table 7.4). Les molécules d'eau dont l'atome d'oxygène se trouve à position équivalente dans les deux structures ont été classées selon leur orientation. Le premier sous-ensemble contient les molécules d'eau orientées différemment, le deuxième sous-ensemble contient celles qui ont une orientation comparable. Les facteurs B_{eq} ont été calculés pour les atomes d'oxygène de chaque sous-ensemble. De cette façon, il peut être mis en évidence si les atomes O sont plus agités dans un des deux sous-ensembles.
Les valeurs moyennes de B_{eq} dans la structure obtenue par diffraction des neutrons sont de $26 \pm 8\text{Å}^2$ et $26 \pm 12\text{Å}^2$ pour le sous-ensemble 1 et 2, respectivement, et de $8 \pm 2\text{Å}^2$ et $6 \pm 1\text{Å}^2$ dans la structure$_{\mathrm{RX}}$. Les facteurs B_{eq} ne sont pas significativement différents dans les deux sous-ensembles, ce qui implique que les déviations du réseau de liaisons H dans les deux structures ne sont pas corrélées à l'agitation thermique des atomes.
Finalement, les données de diffraction ont été analysées. Notons tout d'abord que la résolution des données de diffraction des neutrons est 2.2 Å, tandis qu'elle est 0.85 Å pour les données des rayons X. Une résolution moyenne ne permet pas de discerner les détails de la densité. La complétude est également différente, elle est de 93% pour les données à RX, et 82% pour les données à neutrons.
Les facteurs de structure des données neutrons ont été analysés avec le logiciel Truncate [98] de CCP4 [67] et avec Xtriage de *Phenix* [96]. Les résultats

rayons X		**neutrons**		**sous-ensemble**
HOH	B_{eq} [Å^2]	DOD	B_{eq} [Å^2]	
509	6,3	1052	15,2	
509	6,3	1052	15,2	
511	7,8	1010	45,0	
604	5,4	1034	18,3	
524	6,9	1033	26,0	1
703	7,9	1066	24,2	orientation
613	8,8	1089	23,7	différente
607	7,2	1016	23,9	
505	6,1	1095	22,0	
543	7,2	1057	23,0	
705	10,0	1096	33,7	
538	8,9	1064	29,6	
moyenne	7.5(1.4)		25.9(8)	
683	7,1	1054	24,1	
527	7,6	1013	51,9	
502	5,6	1032	15,7	
501	5,1	1050	14,4	2
623	5,5	1055	26,9	orientation
615	5,8	1015	14,9	similaire
614	5.7	1036	22,7	
580	7,9	1076	32,8	
512	6,9	1037	25,7	
moyenne	6.4(1.1)		25.5(12)	

Table 7.4 – Facteurs d'agitation thermique isotrope équivalent (B_{eq}) d'atomes d'oxygènes, localisés à position correspondante dans les structures à RX et neutrons. Notons qu'une moyenne de $B_{eq} = 8$Å^2 et $B_{eq} = 26$Å^2 correspondent approximativement à un écart de 0.3 Å et de 0,6 Å par rapport à la position d'équilibre. Rappelons que $B_{eq} = 8\pi^2 \left\langle \vec{u}^2 \right\rangle$.

suggèrent que le cristal utilisé pour l'expérience de diffraction des neutrons était maclé (tableau 7.5). En effet, le L-test [99] et l'analyse des moments de Wilson donnent des valeurs qui sont attendues pour des cristaux maclés.

méthode	logiciel	résultat	non maclé	macle parfaite
$\langle I^2 \rangle / \langle I \rangle^2$	Xtriage	1.786	2.0	1.5
$\langle F^2 \rangle / \langle F \rangle^2$	Xtriage	0.872	0.785	0.885
L-test	Xtriage	0.333	0.5	0.375
L-test	Truncate	0.324	0.5	0.375

Table 7.5 – Récapitulatif de l'analyse des données de diffraction des neutrons. Pour plusieurs méthodes de détection de macles, la valeur calculée est donnée (colonne "résultat") ainsi que les valeurs attendues pour une macle parfaite ou un cristal non maclé.

Il semble alors que c'est la qualité des données de diffraction des neutrons qui est à l'origine des différences observées entre rayons X et neutrons. Des cartes de densité ont été calculées pour tester l'effet de la résolution et de la complétude sur l'apparence et donc l'interprétabilité des pics de densité nucléaire et électronique des atomes d'hydrogène (ou de deutérium) d'une molécule d'eau. En plus, l'effet de l'occupation des atomes H/D dans des groupements chimiques comprenant des atomes d'hydrogène a été étudié.

7.3.5 Densité nucléaire calculée de groupements chimiques comprenant des atomes H

Quelques groupements dans les protéines sont souvent impliqués dans les mécanismes catalytiques ou dans la fixation d'un substrat, comme par exemple le groupement hydroxyl (présents dans les acides aminés, sérine, thréonine et tyrosine) ou l'acide aminé histidine. Leurs états de protonation sont alors particulièrement intéressants. Or, si la protéine a été deutérée par trempage, il est possible que les atomes d'hydrogène ne se soient pas échangés complètement contre des atomes D. Cela mène donc à une superposition d'atomes H et D qui sont à la même position. Leurs longueurs de diffusion étant de signes opposés, il est possible que leur signaux respectifs s'annulent. En effet, pour trouver la position d'un atome hydrogène, il est usage d'analyser des cartes de densité nucléaire (ou électronique) résiduelle. Si le rapport des occupations H et D est tel que leurs densités s'annulent, il n'y aura pas de pic de densité dans une carte résiduelle, ce qui rend difficile, voire impossible leur construction.

Un autre constituant important des protéines sont les molécules d'eau. Les protéines solubles sont entourées d'une couche d'hydratation et contiennent souvent des molécules d'eau enfouies à l'intérieur, parfois situées dans le site actif de la macromolécule. Les atomes d'hydrogène des molécules d'eau sont supposés s'échanger facilement contre les atomes de deutérium, l'effet d'annulation peut alors négligé. Par contre, il est intéressant de comparer l'apparence des pics de densité électronique et nucléaire, à différentes résolutions. Dans un premier temps de tels calculs sont utiles pour élucider les déviations de géométrie des liaisons H dans le cas de la DFPase, mais ils sont aussi applicables à toutes les protéines contenant des molécules H_2O.

Molécules d'eau

Les facteurs de structure ont été calculés [10] avec MoPro [18], en utilisant une molécule D_2O (neutrons) ou une molécule H_2O (rayons X). La même résolution et complétude des données de diffraction que celles reportées dans les publications ont été appliquées, c'est-à-dire une résolution de 0.85 Å et une complétude de 93% pour les données RX calculées d'une molécule H_2O, et une résolution de 2.2 Å et 82% de complétude pour les données neutrons calculées d'une molécule D_2O. L'occupation de l'atome d'oxygène a été fixée à 0 tandis que celle des atomes H et D a été fixée à 1. Ainsi, les cartes calculées miment les caractéristiques d'une carte de densité résiduelle $F_{obs} - F_{calc}$ d'un modèle où l'atome d'oxygène a été placé et qui est utilisée pour déterminer la position des atomes H (ou D). Les facteurs d'agitation thermique isotrope équivalent des atomes H et D ont été fixés à 7.5Å^2 et 25Å^2, respectivement, ce qui correspond aux moyennes observées dans les deux modèles. Les cartes Fourier de densité $F_{calc}e^{i\phi_{calc}}$ ont été calculées avec le module VMoPro [18]. Dans la carte de densité électronique à 0.85 Å (figure 7.11), on peut voir, comme attendu, deux pics distincts correspondant aux atomes d'hydrogène de la molécule d'eau. Par contre, dans la carte de densité nucléaire de la molécule D_2O, un seul pic de densité nucléaire de forme ellipsoïdale est visible. Ce pic est localisé entre les deux atomes de deutérium et rappelle clairement les pics de densité observés avec les données expérimentales (figure 7.7). Il est probable que le mouvement thermique (pour les calculs, l'agitation à été supposée identique pour les deux atomes de deutérium) et l'imperfection des données de diffraction expérimentales puissent changer davantage la forme du pic de densité ou la position du maximum. Par conséquent, il semble difficile de déterminer la position exacte des deux atomes de deutérium uniquement

10. Il s'agit de facteurs de structure F_{calc}, calculés à partir d'un modèle en utilisant la transformée de Fourier inverse ($\vec{F} = TF^{-1}(\rho)$).

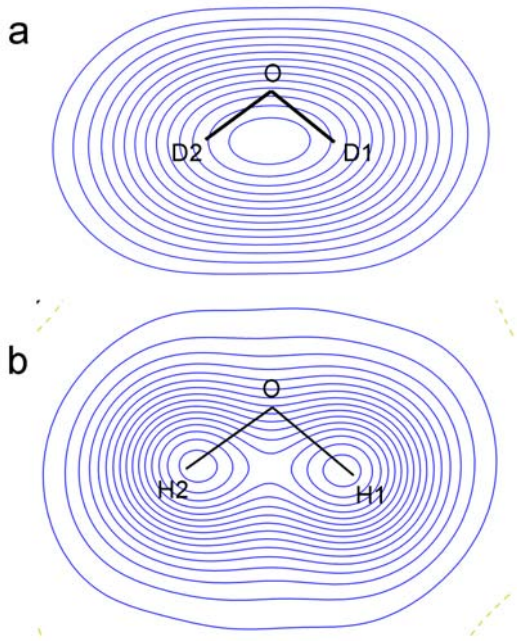


Figure 7.11 – (a) Densité nucléaire calculée d'une molécule D_2O à 2.2 Å de résolution. Les lignes de contour sont de 0.1 $fm/Å^3$. (b) Densité électronique calculée d'une molécule H_2O à 0.85 Å de résolution. Les lignes de contour sont de 0.025 $e/Å^3$. Les deux figures ont été calculées dans le plan H-O-H (ou D-O-D).

sur la base d'un seul pic de densité nucléaire.
Afin de déterminer la résolution, ou la complétude limite, à partir desquelles on peut observer deux pics de densité distincts pour les atomes de deutérium, des cartes de densité nucléaire ont été calculées en faisant varier ces deux paramètres.
Le profil de la complétude des données de diffraction nucléaire a été tracé avec le module Truncate du logiciel CCP4 [67]. Ce profil a ensuite été paramétré pour pouvoir être appliqué à des gammes de résolution plus élevées. Les facteurs de structure et les phases ont été calculés avec MoPro. Les calculs ont été effectués avec une molécule d'eau D_2O avec des facteurs d'agitation thermique isotrope équivalent de $15Å^2$ et $25Å^2$. La limite inférieure de la résolution a été fixée à 20.0 Å tandis que la limite de la haute résolution a été changée. En utilisant les paramètres obtenus par le profil de complétude des données neutrons, des réflections ont été supprimées de façon aléatoire afin de mimer l'incomplétude de données de diffraction expérimentales. La densité nucléaire a été tracée dans le plan D-O-D de la molécule et la forme du pic a été examinée si deux pics sont présents.

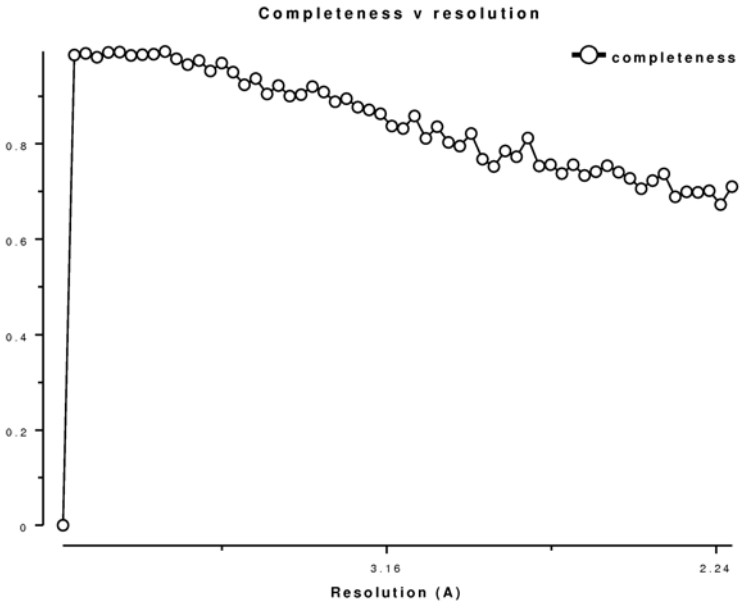


Figure 7.12 – Profil de complétude pour les données de diffraction des neutrons. La complétude est tracée en fonction de la résolution. La complétude diminue de façon quasi linéaire à partir de 5.5 Å.

Selon ces cartes, une résolution de 1.6 Å est nécessaire pour, dans l'idéal, commencer à discerner les pics de densité nucléaire des deux atomes de deutérium ayant un B_{eq} égal à 25Å^2 (figure 7.13). Pour la molécule d'eau moins agitée (B_{eq} = 15Å^2), la résolution limite est de 1.7 Å. L'effet de la complétude des données se manifeste par la diminution de la hauteur du pic mais sa forme reste similaire. A résolution moyenne, les deux pics se chevauchent et résultent en un seul pic de densité nucléaire[11], localisé entre les deux atomes D. Par conséquent il est nécessaire d'avoir une bonne résolution et complétude des données de diffraction des neutrons pour observer le réseau de liaisons hydrogène. Dans le cas de la DFPase, la résolution et la complétude ne sont pas très élevées, rendant difficile l'interprétation des cartes de densité nucléaire. Ceci explique probablement les nombreuses déviations par rapport à la stéréochimie standard et par rapport à la structure à rayons X.

11. Au sens strict, il n'est pas nécessaire d'observer deux pics distincts pour pouvoir placer les atomes de deutérium. Une carte de densité nucléaire résiduelle comme sur la figure 7.13 (a) permettra sans doute de déduire l'orientation de la molécule d'eau. Par contre, il est discuté de décider à partir de quelle forme exacte du pic la localisation devient difficile. En effet, quand la résolution diminue, le pic de densité devient de plus en plus étroit, entravant son interprétation. Le critère que j'ai choisi est alors basé sur l'observation de deux pics distincts.

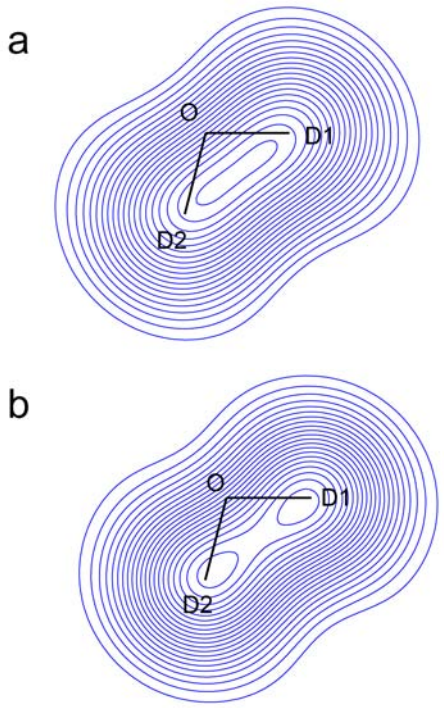


Figure 7.13 – Densité nucléaire calculée d'une molécule D_2O ($B_{eq} = 25\text{Å}^2$). La résolution est limitée à 1.7 Å dans (a) et à 1.5 Å dans (b). Les lignes de contour sont de 0.1 $fm/\text{Å}^3$.

Par conséquent, la distance élevée des longueurs de liaison hydrogène dans la structure à neutrons est vraisemblablement due à la modélisation incorrecte de certaines molécules d'eau.

Groupement hydroxyl (O-H)

Le groupement hydroxyl a un degré de rotation autour de l'axe C-O, ce qui permet à l'atome d'hydrogène de s'orienter différemment. Il y a encore une autre particularité : l'atome d'oxygène est capable d'accepter une liaison hydrogène tandis que l'atome d'hydrogène peut être un donneur. Le groupement O-H peut alors jouer les deux rôles simultanément, de donneur et d'accepteur de liaisons H. Par conséquent, la détermination de la position de l'atome d'hydrogène est utile pour connaître précisément les interactions du groupement car des considérations stéréochimiques peuvent être ambigües.
Un modèle contenant uniquement l'acide aminé sérine a été créé avec le logiciel Pymol [46]. A la même position de l'atome d'hydrogène du groupement hydroxyl, un atome de deutérium a été ajouté. Les occupations respectives

des atomes H et D du groupement hydroxyl ont été changées, en gardant leur somme égale à 1. L'occupation de tous les autres atomes a été fixée à 0 de façon à mimer une carte résiduelle $F_{obs} - F_{calc}$. Pour trouver la combinaison des deux occupations qui annule le facteur de structure, il suffit de calculer :

$$occ_H \cdot b_H + occ_D \cdot b_D = 0$$

Avec occ_H et occ_D les occupations des atomes d'hydrogène et de deutérium, respectivement, et $b_H = -3.74$ fm et $b_D = 6.671$ fm leurs longueurs de diffusion. Supposant que la somme des deux occupations donne 1 ($occ_H = 1 - occ_D$) et que les atomes H et D sont à la même position, on peut en déduire que quand l'occupation de l'atome de deutérium est égale à 36% et celle de l'atome d'hydrogène est de 64%, le facteur de structure s'annule. Par conséquent, il n'y aura pas de pic visible dans une carte de densité nucléaire résiduelle.
L'effet de l'occupation sur la forme du pic de densité qui résulte de la superposition d'un atome H et D a été analysé. Les facteurs de structure et les phases ont été calculés à 1.5 Å de résolution, en changeant le taux d'occupation H/D.
Le pic de densité des atomes H et D subit l'évolution suivante (figure 7.14) : Quand l'occupation de l'atome de deutérium est nulle, le pic est fort et le contour est négatif. En diminuant l'occupation de l'atome H, la hauteur du pic diminue d'autant et il disparaît complètement quand l'occupation est égale à 64%. Ensuite, le pic devient positif et sa hauteur augmente au fur et à mesure que l'occupation de l'atome D domine. Il est ainsi difficile de voir l'état de protonation d'un groupement hydroxyl quand l'occupation de l'atome d'hydrogène est compris entre 50% et 75 % environ, car la densité nucléaire résultante est très faible et il est probable qu'elle ne sera pas distinguable du bruit expérimental. Ceci souligne qu'il est préférable d'effectuer la deuteration depuis l'étape d'expression de la protéine plutôt que par trempage si l'on veut déterminer la position exacte des atomes d'hydrogène (ou de deutérium) dans une structure protéique.

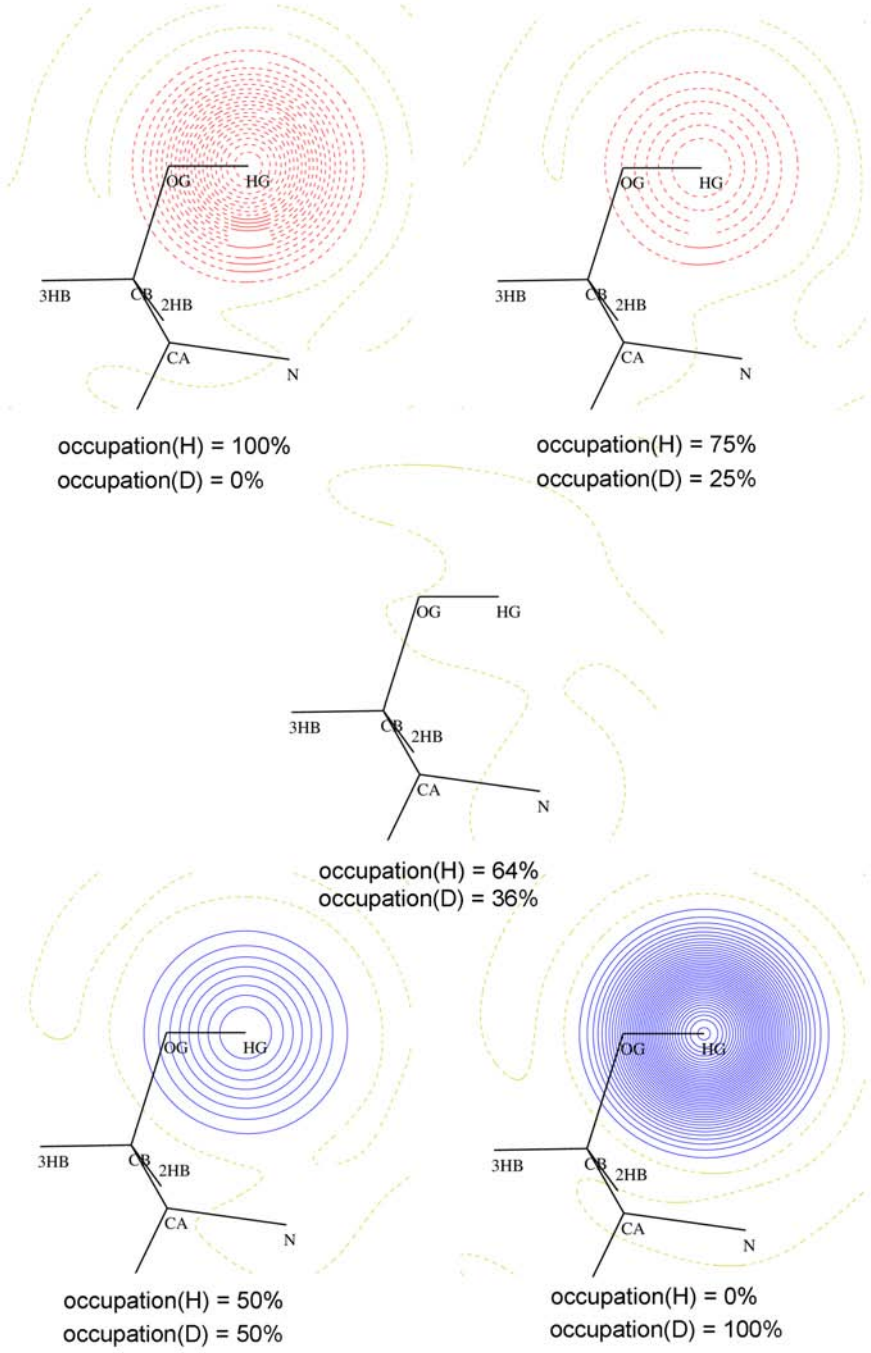


Figure 7.14 – Densité nucléaire calculée qui résulte de la superposition d'un atome d'hydrogène et d'un atome de deutérium dans un groupement hydroxyl. Le contour est de 0.05 $fm\mathrm{A}^{-3}$. Les cartes de densité ont été calculées dans le plan $\mathrm{C}_\beta - O\gamma - H\gamma$.

Histidine

Le résidu histidine a la particularité que l'azote basique de la chaîne latérale imidazole peut capter un proton à pH physiologique. Des petites variations de pH peuvent donc changer la charge du résidu. En plus, dans l'histidine non chargée, l'imidazole est un tautomère, l'atome d'hydrogène peut donc être localisé sur l'atome d'azote Nδ1 ou sur Nϵ2.
La chaîne latérale de l'acide aminé histidine est souvent impliquée dans la coordination de cations métalliques, on trouve ce résidu alors dans les sites catalytiques de métalloprotéines. Dans une triade catalytique, l'histidine basique déprotone un groupement hydroxyle pour l'activer et créer un nucléophile. La connaissance de l'état de protonation d'une histidine peut alors avoir un intérêt crucial pour la compréhension d'un mécanisme catalytique.
Pour l'acide aminé histidine, il est intéressant d'analyser des cartes de densité tridimensionnelles car l'imidazole contient des atomes d'hydrogène capable de s'échanger et des groupements C-H dont l'atome H s'échange moins facilement.
Un modèle contenant uniquement l'acide aminé histidine, simplement protonée à l'atome d'azote Nϵ2, a été créé avec le logiciel Pymol [46]. A la même position de l'atome d'hydrogène, un atome de deutérium a été ajouté. Pour différents rapports d'occupation H/D, deux cartes de densité nucléaire ont été calculées à 2 Å de résolution [12].
La première carte reproduit les caractéristiques d'une carte de densité nucléaire expérimentale $F_{obs} - F_{calc}$ d'un modèle, où il n'y a pas d'atome d'hydrogène (ou de deutérium) lié à Nϵ2. L'occupation de tous les atomes de l'acide aminé a été fixée à 0, sauf l'occupation des atomes H et D de l'azote Nϵ2. Des facteurs de structure et des phases ont été calculés avec des occupations 100%, 64% ou 0% de l'atome d'hydrogène et en fixant l'occupation de l'atome de deutérium complémentaire à 100%.
La deuxième carte a pour but de mimer une carte $2F_{obs} - F_{calc}$. Dans le cas des facteurs de structure calculés, F_{obs} est obtenue à partir du modèle de l'histidine, où Nϵ2 est protoné tandis que F_{calc} vient de la structure où cet atome d'hydrogène n'est pas placé. Ainsi, on peut reproduire les cartes de densité nucléaires auxquelles on est confronté avant avoir placé les atomes d'hydrogène.
La figure 7.15 montre les deux cartes de densités nucléaires superposées à la structure de l'acide aminé histidine. Quand l'occupation de l'atome d'hydrogène est de 100%, l'atome H crée, comme attendu, un pic de densité nucléaire négatif dans une carte de densité résiduelle. Quant à la densité $2F_{obs} - F_{calc}$ (calculée), on observe qu'elle est aplatie autour de Nϵ2. A l'occupation de

12. Souvent, la résolution des données expérimentales neutrons est autour de 2 Å.

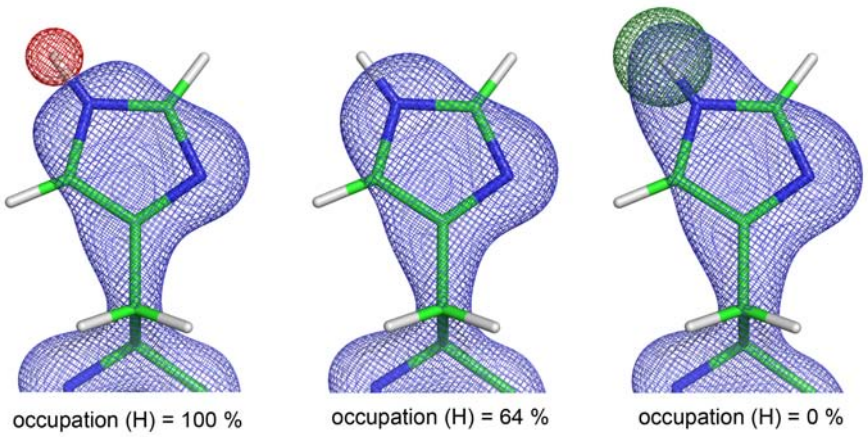


Figure 7.15 – En bleu : Densité nucléaire calculée mimant une carte $2F_{obs} - F_{calc}$ du résidu histidine. Le niveau de contour est de 2.0 σ. La densité nucléaire négative et positive mimant une carte $F_{obs} - F_{calc}$ est en rouge et en vert, respectivement. Les deux cartes sont contourées à 1.5 σ.

64%, il n'y a pas de densité résiduelle dans la carte nucléaire car la densité des atomes H et D s'annule. En plus, la densité $2F_{obs} - F_{calc}$ (calculée) autour de Nϵ2 est similaire à celle observée autour de Nϵ1. Ainsi, en se basant uniquement sur la densité nucléaire et lorsque le taux d'occupation de l'atome d'hydrogène est de 64%, il n'est pas possible de savoir si l'atome Nϵ2 ou Nδ1, ou les deux, sont protonés. Par contre, quand l'occupation de l'atome de deutérium est de 100%, la situation est très claire : Il y a un fort pic de densité positif et la densité mimant $2F_{obs} - F_{calc}$ (calculée) s'étend autour de l'atome d'hydrogène. Dans toutes les cartes mimant la densité nucléaire $2F_{obs} - F_{calc}$, on remarque également l'annulation de la densité nucléaire du groupement $C\beta H_2$.
Les deux exemples sur l'état de protonation de l'imidazole et l'orientation du groupement hydroxyl montrent que la localisation d'un atome d'hydrogène (ou de deutérium) peut être également difficile par diffraction des neutrons. Si le cristal de protéine a été deutéré par trempage, le taux d'échange pour certains atomes d'hydrogène peut être tel que la densité des atomes H et D s'annule ou ne sera pas distinguable du bruit expérimental.
Or, les logiciels d'affinement offrent souvent la possibilité d'affiner l'occupation des atomes d'hydrogène dans une structure à neutrons. Si la densité superposée des atomes H et D s'annule ou si elle est faible, un affinement pourra donner une position quelconque sans pénalisation.

Influence du facteur de Debye-Waller

L'estimation du rapport d'occupation ci-dessus n'est valable que pour des températures très basses (T≈ 0) ou si le facteur de Debye-Waller (DW) est le même pour l'atome d'hydrogène ou de deutérium. En général, ce n'est pas le cas puisque l'écart quadratique moyen du déplacement atomique (qui apparaît dans le facteur de DW : $exp(-B_{eq}\frac{sin^2\theta}{\lambda^2}) = exp(-8\pi^2 \langle\vec{u}^2\rangle \frac{sin^2\theta}{\lambda^2})$) dépend de la masse de l'atome diffractant.
Dans l'approximation harmonique, l'écart quadratique moyen $\langle\vec{u}^2\rangle$ est inversement proportionnel à la masse [100] :

$$\langle\vec{u}^2\rangle \propto \frac{1}{m} \tag{7.1}$$

Par conséquent, le facteur d'agitation thermique isotrope équivalent d'un atome d'hydrogène est deux fois plus élevé que celui d'un atome de deutérium.

$$B_{eq,H} = 2 \cdot B_{eq,D} \tag{7.2}$$

Il s'en suit pour l'annulation de la densité nucléaire :

$$\begin{aligned} 0 &= occ_H \cdot b_H \cdot exp(-2B_{eq,D}\frac{sin^2\theta}{\lambda^2}) + occ_D \cdot b_D \cdot exp(-B_{eq,D}\frac{sin^2\theta}{\lambda^2}) \\ &= occ_H \cdot b_H \cdot exp(-B_{eq,D}\frac{sin^2\theta}{\lambda^2}) + occ_D \cdot b_D \end{aligned} \tag{7.3}$$

Le rapport d'occupation dépend alors de l'agitation thermique. Par exemple, si la résolution est de 2 Å, et $B_{eq,D} = 20\text{Å}^2$, la fonction exponentielle est égale à 0,3. L'occupation de l'atome H, pour laquelle la densité nucléaire résultante s'annule, est alors de 86 % ($occ_D = 14\%$). Si l'agitation est encore plus élevée, $B_{eq,D} = 40\text{Å}^2$, son occupation atteint 95 %. Ceci démontre la forte corrélation entre l'agitation thermique et l'occupation dont il faut tenir compte lors d'un affinement d'une structure neutrons.

7.3.6 Etude du site actif de la DFPase

Le site actif de la DFPase se trouve dans le tunnel central de la protéine. Les atomes d'oxygène de 4 résidus différents (Glu 21, Asn 120, Asn 175 et Asp 229) ainsi que 3 molécules d'eau sont coordonnés au calcium catalytique (figure 7.16).
D'après la structure à neutrons, une de ces molécules d'eau (DOD 1033) a une orientation particulière, avec un de ses atomes d'hydrogène pointant directement vers l'ion calcium. La distance H...Ca2+ est 2.0 Å et l'angle formé par les atomes O-H...Ca2+ est 112°. Puisque les deux atomes sont chargés positivement, la configuration n'est pas favorable énergétiquement et sa validité est douteuse.

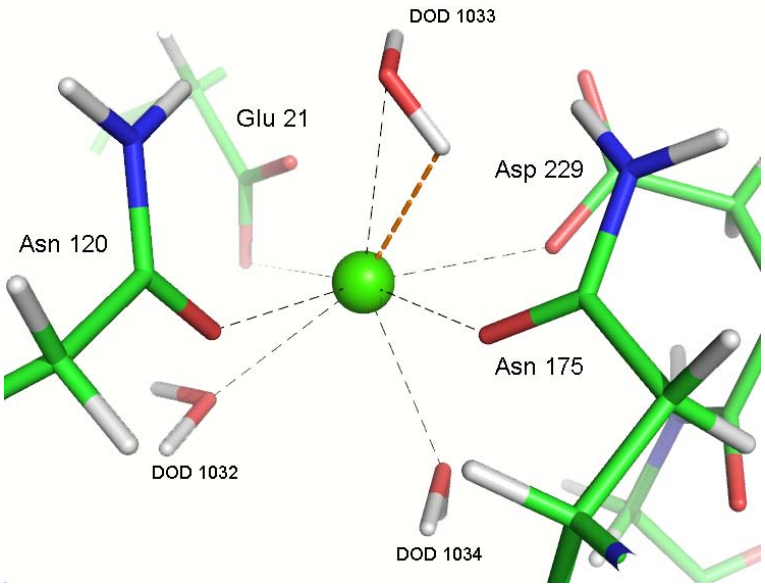


Figure 7.16 – Site actif dans la structure à neutrons. La coordination du calcium catalytique est représentée par des lignes pointillées noires, en somme, l'ion calcium est coordonné à sept atomes d'oxygène voisins. L'interaction courte H...Ca2+, telle que modélisée dans [86] est représentée par la ligne pointillée orange.

Dans la structure à rayons X, la coordination du calcium catalytique est similaire et on trouve l'atome d'oxygène d'une molécule d'eau (HOH 524) à la même position que celui de DOD 1033. Dans la carte de densité électronique résiduelle, il y a un pic fort (niveau de contour : 4σ) à côté de cette molécule d'eau, à 1.1 Å de distance, qui peut être attribué à un atome d'hydrogène. De manière intéressante, la position du pic ne correspond pas à la position d'un des atomes de deutérium de DOD 1033 dans la structure à neutrons. Notons en plus que le pic de densité a une forme nettement ellipsoïdale, quoique tous

les autres pics d'atomes H dans le site actif soient sphériques.
L'analyse de l'environnement chimique de la molécule d'eau HOH 524 montre qu'elle est impliquée dans les interactions suivantes (figure 7.17) :

1. Son atome d'hydrogène, localisé d'après le fort pic de densité électronique décrit ci-dessus, pointe vers HOH 511 et forme une liaison hydrogène canonique (distance O...O = 2.82 Å, angle O-H...O = 136°).
2. Son atome d'oxygène accepte une liaison H du groupe NH_2 appartenant au résidu voisin Asn 175 (distance H...O = 2.23 Å, angle N-H...O = 139°).
3. L'atome O de HOH 524 est coordonné au calcium catalytique (distance O...Ca2+ = 2.5 Å).
4. L'atome oxygène est en interaction avec la chaîne latérale du résidu Asp 229.

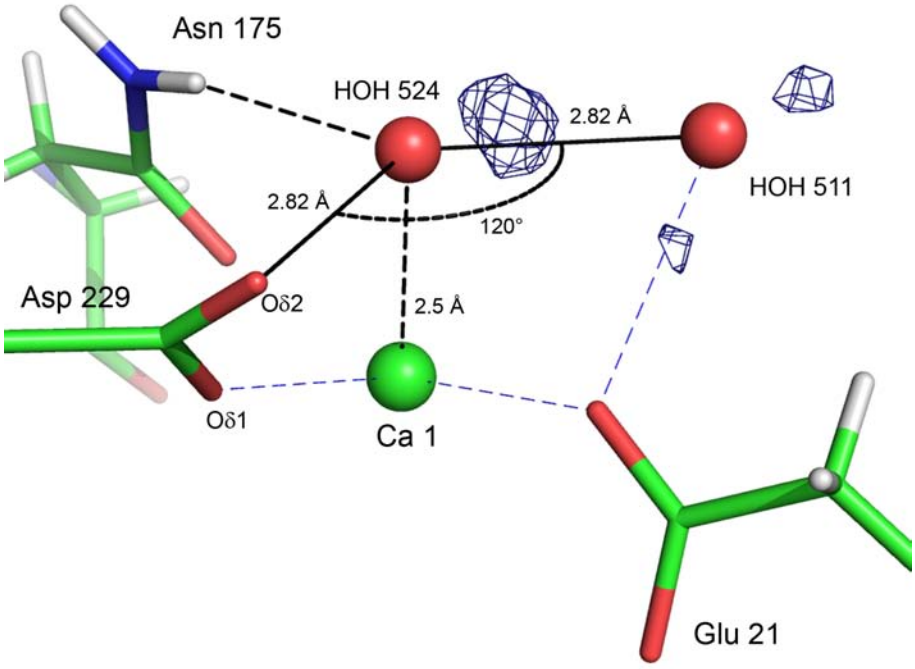


Figure 7.17 – Environnement de la molécule d'eau HOH 524 dans la structure à rayons X. La carte de densité électronique résiduelle $F_{obs} - F_{calc}$ est contourée à 2.8 σ. Les interactions impliquant HOH 524 et correspondant aux positions proposées de ses atomes d'hydrogène sont représentées par des lignes noires.

Cette dernière interaction, entre HOH 524 et le groupement carboxylate de Asp 229, présente des distances O...Oδ2 et O...Oδ1 de 2.82 Å, ce qui correspond aux distances donneur-accepteur standards pour les liaisons H...O dans

les protéines [101]. Toutefois, la localisation du second atome d'hydrogène de HOH 524 entre l'oxygène et l'atome Oδ1 de Asp 229 est moins favorable, du à la proximité du cation calcium. Par conséquent, l'atome Oδ2 de Asp 229 est le seul possible accepteur d'une liaison hydrogène pour HOH 524. Dans ce cas, l'angle O(511)...O(524)...Oδ2(229) est 120°, compatible avec les liaisons hydrogène possible des 2 atomes d'hydrogène de la molécule d'eau HOH 524. Malgré cette configuration géométrique idéale, il n'y a pas de preuve expérimentale pour un second atome d'hydrogène dans les cartes de densité résiduelle $F_{obs} - F_{calc}$, à part un très faible pic de densité électronique visible à un niveau de contour égal à 1 σ, localisé entre l'atome d'oxygène de HOH 524 et Oδ2 de Asp 229 (figure 7.18).

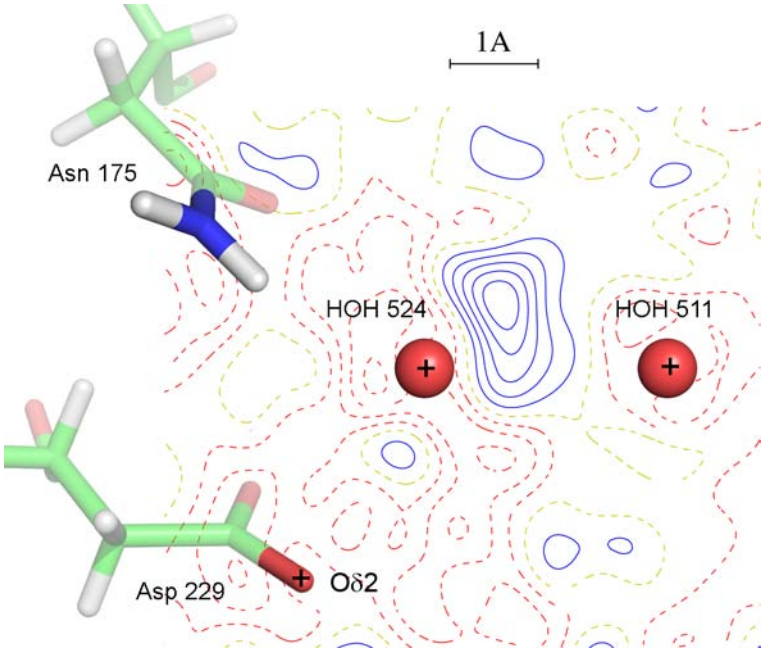


Figure 7.18 – Carte de densité électronique résiduelle $F_{obs} - F_{calc}$, calculée dans le plan des trois atomes d'oxygène O(524), O(511) et Oδ2(229) (marqués avec une croix) montrant le pic de densité ellipsoïdal entre O(524) et O(511) ainsi que le faible pic de densité visible entre O(524) et Oδ2. L'atome d'oxygène de HOH 524 accepte une liaison hydrogène d'un groupement NH_2 qui est localisé légèrement au dessus du plan. Le contour des lignes de niveau est égal à 0.05 $e\text{Å}^{-3}$.

Or, on ne peut pas conclure de l'absence d'un pic significatif dans une carte de densité résiduelle H omit qu'un atome hydrogène ne soit pas présent. Le pic de densité électronique d'un atome d'hydrogène est déjà faible, mais il peut diminuer encore d'avantage en raison d'une résolution insuffisante, d'une agitation thermique élevée, de conformations multiples ou de positions

diffuses comme une "low-barrier-hydrogen-bond" [84]. En tenant compte des autres interactions dans lesquelles HOH 524 est impliquée, comme le contact O...Oδ2 (distance : 2.82 Å), on pourrait s'attendre à ce que ses deux atomes d'hydrogène soient bien définis et que leurs densités électroniques soient visibles, comme celles des autres atomes H dans le site actif.
Ces observations peuvent cependant être expliquées par un mécanisme de délocalisation impliquant l'atome d'hydrogène de HOH 524. Plus précisément, Oδ2 agit comme base et accepte le proton de HOH 524, qui est ensuite activé en ion hydroxyde. Ce mécanisme est compatible avec la présence de l'ion Ca2+, qui agit comme acide de Lewis en favorisant la libération du proton de HOH 524 et qui interagit avec la paire d'électrons libres de la molécule d'eau activée. Puisque le proton délocalisé n'est pas nettement visible sur Asp 229 dans la structure obtenue par diffraction des RX de la DFPase, on observerait alors une superposition cristallographique des deux états de HOH 524, ce qui expliquerait le faible pic de densité électronique entre l'oxygène de la molécule d'eau et Oδ2 de Asp 229. Ceci est également compatible avec la forme ellipsoïdale du pic de densité qui correspond au deuxième atome H de HOH 524. L'ellipticité résulte de la superposition des deux états chimiques de la molécule d'eau HOH 524. En effet, le mécanisme de transfert de proton impose une légère rotation de HOH 524 pour favoriser l'interaction entre la paire libre d'électrons qui vient d'être libérée et l'ion Ca2+. Ce mécanisme va maintenir la neutralité formelle du site actif de la DFPase, la charge négative étant transférée de Asp 229 vers la molécule d'eau activée. Il a été montré par des études de mutagenèse dirigée que la neutralité du site actif est nécessaire pour maintenir l'activité enzymatique [102].

7.3.7 Mécanisme catalytique

Deux mécanismes catalytiques ont été suggérés pour la DFPase. Le premier mécanisme suppose l'attaque du centre phosphorique par un ion hydroxyde. Ce mécanisme est similaire à celui d'autres hydrolases capables de neutraliser les organophosphorés, comme par exemple la phosphotriesterase (PTE) [103] ou la paraoxonase humaine [104], qui est un homologue proche de la DFPase.
Suite à une expérience de labelling d'isotope $H_2^{18}O$, Blum *et al.* proposaient que le résidu Asp 229 agisse comme nucléophile et attaque le centre phosphorique du substrat via la formation d'un intermédiaire phospho-enzyme [105]. Ce mécanisme semble maintenant bien établi mais il manque d'une hypothèse satisfaisante expliquant comment l'enzyme est régénérée. En effet, il a été proposé que l'enzyme soit régénérée par une molécule d'eau du solvant (non activée), ce qui est improbable pour les raisons suivantes : d'une part, la

DFPase est une enzyme très rapide et d'autre part, le site actif d'une structure DFPase avec un mime de substrat est enfoui et il n'est pas accessible à des molécules d'eau du solvant [105], suggérant que le site actif de DFPase soit inaccessible quand l'intermédiaire phospho-enzyme est formé.
La régénération de l'enzyme pourrait être expliquée sur la base d'un réarrangement du site actif après la formation de l'intermédiaire, et de l'existence d'une molécule d'eau partiellement activée. En effet, la coordination du calcium catalytique dans une métalloprotéine peut varier, comme il a été observé dans la protéine gluconolactonase [106]. Un repositionnement de la molécule d'eau partiellement activée pendant que le substrat est fixé et une attaque de l'intermédiaire par cette molécule d'eau constituent une hypothèse qui pourrait expliquer la régénération de l'enzyme.
Ainsi, la présence possible d'une molécule d'eau partiellement activée dans le site actif constitue un nouvel élément et devrait relancer le débat sur le mécanisme de la DFPase.

7.4 Conclusion

L'analyse de la structure subatomique de la DFPase a permis la localisation de nombreux atomes d'hydrogènes, notamment ceux de 21 molécules d'eau. Ainsi, le réseau de liaison H dans le tunnel central de la protéine a été établi en se basant uniquement sur l'information expérimentale directe constituée par les pics de densité électronique. Il était donc possible de comparer le réseau de liaison hydrogène dans la structure déterminée par diffraction des rayons X et une structure neutrons publiée récemment [106]. De manière surprenante, il s'avérait que les orientations de plusieurs molécules d'eau sont différentes et ces différences ne pouvaient pas être attribuées à l'agitation thermique ou au montage expérimental. Des cartes Fourier de densité nucléaire calculée, utilisant la complétude et la résolution expérimentales, ont été analysées. Elles montrent dans ce cas que les deux atomes de deutérium d'une molécule d'eau résultent en un unique pic de densité, ce qui entrave la localisation précise de ces atomes. De plus, il a été montré que, pour des données de diffraction des neutrons, une résolution d'au moins 1.7 Å est nécessaire pour pouvoir discerner les deux pics des atomes D.
L'analyse du site actif de la DFPase suggère que la molécule d'eau HOH 524, qui est coordonnée à l'ion calcium catalytique, soit partiellement activée. L'activation est réalisée par une délocalisation d'un proton vers l'atome d'oxygène Oδ2 de l'acide aminé voisin, Asp 229. L'hypothèse est supportée par la présence d'un faible pic de densité électronique qui est compatible avec une position diffuse du proton entre la molécule d'eau et la chaîne latérale de

Asp 229. D'un autre coté, la configuration défavorable de la molécule d'eau DOD 1033 (à position équivalente par rapport à HOH 524) qui se trouve dans la structure à neutrons, ne pouvait pas être confirmée. La présence d'une molécule d'eau partiellement activée dans le site actif de la DFPase constitue un nouveau élément qui devrait être prise en compte dans le débat sur le mécanisme de dégradation des organophosphorés.
Finalement, ces résultats montrent que des données de diffraction des rayons X à résolution subatomique peuvent apporter des informations précises sur les états de protonation dans les protéines, et concurrencent clairement les données de diffraction des neutrons, quand celles-ci sont de moyenne résolution.

7.5 Perspectives

Il est envisageable de faire une nouvelle expérience de diffraction des rayons X de la DFPAse sur un synchrotron de 3e génération pour essayer d'améliorer encore la résolution. Ainsi, il serait possible d'observer encore mieux les atomes d'hydrogènes et l'analyse de la molécule d'eau HOH 524 serait très intéressant. De plus, un affinement multipolaire de la densité électronique pourrait être tenté.
La comparaison des deux structures de la DFPase a montré qu'il y avait plusieurs incohérences dans la structure à neutrons : L'orientation de plusieurs molécules d'eau était douteuse et l'analyse des facteurs de structure suggérait que le cristal utilisé pour l'expérience était maclé. Pour cette raison, il ne devrait pas être exclu de refaire une expérience de diffraction des neutrons et de confronter à nouveau ces données aux résultats des rayons X.
Il serait intéressant de généraliser les calculs sur la limite de résolution qui permet de discerner deux pics distincts pour les atomes d'hydrogènes d'une molécule d'eau dans le cas de la diffraction des neutrons. En effet, les facteurs de structure calculés qui ont étés utilisées dans cette étude ne sont pas soumis à une détérioration comme c'est le cas pour les facteurs de structure expérimentales (bruit, mosaïcité du cristal, anneau de glace, etc.). En mimant un bruit expérimental, la limite de résolution peut être déterminée dans des conditions plus réalistes.

Chapitre 8

La protéine PfluDING

La protéine PfluDING a été cristallisée au laboratoire CRM² et sa structure a été initialement résolue à 1.43 Å [107]. Au début de ma thèse, des données de diffraction à résolution subatomique (0.98 et 0.88 Å), collectées par S. Moniot, étaient disponibles. Dans le site actif de PfluDING se trouve un ion phosphate, supposé être fixé par des liaisons hydrogène. L'idée initiale était d'élucider le mode de fixation du phosphate en utilisant l'information apportée par les données de résolution subatomique.

8.1 Introduction

La protéine PfluDING, issue de la bactérie *Pseudomonas fluorescens*, fait partie de la famille de protéine DING. Le nom de cette famille intéressante vient des quatre acides aminés conservés sur la partie N-terminale de la séquence (acide aspartique (D), isoleucine (I), asparagine (N), glycine (G)). Bien que les protéines DING soient présentes chez les animaux, les plantes et les champignons [108], elles sont systématiquement absentes des banques de données des génomes connus chez les eukaryotes [1] [109]. Pour cette raison, leur séquencage est difficile et seulement une séquence d'acides aminés complète d'une protéine de cette famille et venant d'un organisme eukarotique est connue. Il s'agit de la protéine "Human phosphate binding protein" (HPBP) [110].
On observe une similarité topologique entre les protéines DING isolées des eukaryotes et les protéines solubles bactériennes fixant le phosphate ("phosphate solute binding proteins" - SBP) [109]. Les protéines SBP adoptent un repliment appelé "piège à mouches" ("Venus flytrap"), où deux domaines globulaires sont reliés par une charnière et forment au centre le site de fixa-

1. Eukaryote = organisme dont les cellules comportent un noyau.

tion d'un ion phosphate [111] (figure 8.1). Les 8 acides aminés impliqués dans le mécanisme de fixation sont conservés. La structure de HPBP a été résolue et elle confirme que les protéines DING adoptent une topologie similaire à celle des protéines SBP [112].
La protéine bactérienne PfluDING (poids moléculaire : 39 kDa) a une forte similarité de séquence (70 %) avec HPBP et elle est active contre les fibroblastes humains, comme les protéines DING eukaryotes [113]. PfluDING a la capacité de former de gros cristaux à fort pouvoir diffractant. Ainsi, elle offre la possibilité de poursuivre la recherche sur la structure et le fonctionnement des protéines DING en général. En effet, puisque les séquences des protéines DING eukaryotes sont inconnues, leur analyse structurale est entravée. Concernant la protéine HPBP, dont la séquence et la structure sont déterminées, elle est associée à la protéine paraoxonase humaine, ce qui rend difficile sa production.

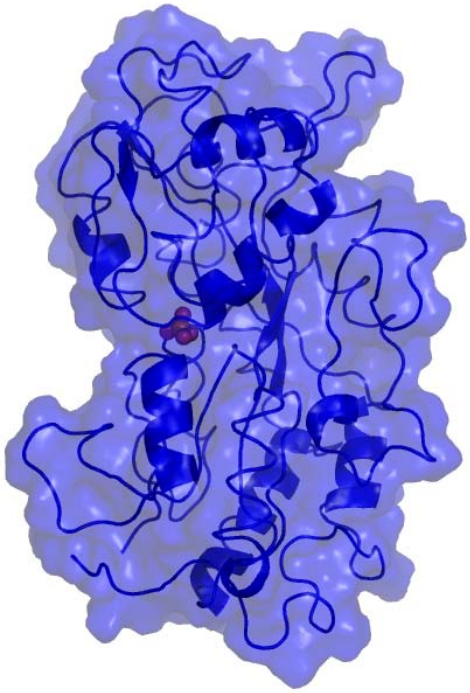


Figure 8.1 – Structure de la protéine PfluDING, seulement la chaîne principale ainsi que la surface moléculaire est représentée. L'ion phosphate est représenté par des sphères rouges et oranges.

Le mécanisme de fixation du phosphate dans les protéines SBP a été analysé, elles fixent le phosphate à pH 4,5 et 8. Notons que l'état de protonation de celui-ci change en fonction du pH, à pH 8, c'est la forme simplement pro-

tonée (HPO_4^{2-}) qui est en excès tandis que c'est le phosphate monobasique à pH 4.5 ($H_2PO_4^-$). Il a été proposé que la forme dibasique du phosphate (HPO_4^{2-}) soit fixée dans le site actif par 12 liaisons hydrogènes [111]. Les interactions phosphate-protéine sont réalisées par 11 groupements dipolaires, de type N-H ou O-H, et par la liaison H entre le seul proton du phosphate et le carboxylate d'un acide aspartique. Ce résidu est le seul accepteur de liaison hydrogène dans le site de fixation et il a été proposé qu'il joue un rôle crucial pour la discrimination entre phosphate dibasique et l'ion sulfate. En effet, le sulfate ne possédant aucun proton à pH physiologique, il ne pourra pas former la liaison hydrogène avec l'acide aspartique et sera repoussé par sa charge négative. D'après les recherches sur les SBP, la distance de liaison H entre le résidu Asp et le phosphate est très courte (distance O...O = 2.45 Å dans une structure atomique de résolution 0,98 Å [111]) et, de manière surprenante, l'énergie de liaison de cette interaction n'est cependant pas plus élevée que celle des liaisons moyennes dans une protéine [114]. Quant au mode de fixation pour le phosphate monobasique ($H_2PO_4^-$), il a été proposé que le groupement hydroxyl (O-H) du résidu Ser 32 s'oriente différemment pour changer de rôle : de donneur en accepteur de liaison H. Ainsi, le deuxième proton du phosphate peut s'intégrer dans le réseau de liaisons hydrogène du site actif. Le potentiel électrostatique de la protéine a été calculé et, de manière surprenante, il est négatif dans le site actif [115], bien que la charge du ligand soit également négative. Il a été proposé que c'est le potentiel négatif qui permet de discriminer entre les substrats phosphate et sulfate : seulement les ligands qui peuvent s'intégrer parfaitement au réseau de liaisons hydrogène sont fixés à la protéine. Le sulfate ne possédant aucun proton se fixe alors moins facilement.

8.2 Manipulations et résultats

8.2.1 Cristallisation - Collecte des données

La protéine a été fournie dans le cadre d'une collaboration par K. Scott de l'université d'Auckland, Nouvelle-Zélande. Les expériences de cristallisation et de collecte des données ont été effectuées par S. Moniot, alors doctorant au laboratoire CRM2. Des monocristaux de PfluDING ont été obtenus par la méthode de la goutte suspendue, la solution de cristallisation se composait de 24-28 % (w/v) PEG [2] 8000, de tampon acétate 100 mM à pH 4.5 et 200 mM Li_2SO_4. Quelques cristaux ont été trempés ensuite pendant 24 h dans une solution à pH 8.5, contenant le tampon Tris-Hcl 0.5M à pH 8.5, 26%

2. Polyéthylène glycol : polymère linéaire d'éthylène glycol.

PEG et 200 mM Li_2SO_4. Notons que l'état de protonation du phosphate, qui est dans le centre de fixation de PfluDING, change en fonction du pH.
La collecte des données de diffraction a été effectuée à l'ESRF (European Synchrotron Radiation Facility, Grenoble), aux lignes de lumière ID14-4 et ID-29. Pour chaque cristal, deux jeux de données ont été collectés. Le premier contient les réflections à basse et moyenne résolution (< 1.5 Å), le deuxième celles de haute résolution. Cette stratégie assure que les réflections à basse résolution soient mesurées correctement. Les données de diffraction ont été indexées et mises à l'échelle avec le logiciel XDS [116]. Des phases initiales ont été obtenues par remplacement moléculaire avec la structure de PfluDING résolue à 1.43 Å[107]. Soulignons que deux jeux de données sont ainsi disponibles, celui de la protéine cristallisée à pH 4.5 (noté aussi structure$_{pH4.5}$) et celui du cristal trempé à pH 8.5 (noté aussi structure$_{pH8.5}$). Ainsi, le mode de fixation peut être analysé à des valeurs de pH différents.

	cristal$_{pH4.5}$	cristal$_{pH8.5}$
longueur d'onde [Å]	0.953	0.750
température [K]	100	100
résolution [Å]	0.98	0.88
complétude [%]	98.1	97.4
redondance	5.4	6.1
groupe d'espace	$P2_1$	$P2_1$
R_{merge}[%]	4.3	5.3
$\left\langle \frac{I}{\sigma(I)} \right\rangle$	24.4	16.6

Table 8.1 – Statistiques sur les données de diffraction de la protéine PfluDING.

Des premiers affinements ont été effectués par S. Moniot avec le logiciel REFMAC [117] pour la structure à pH 4.5. La structure à pH 8.5 a été affiné avec REFMAC ainsi que SHELXL [118]. Plus précisément, les atomes des chaînes principales et latérales ont été placés en densité électronique et leur agitation thermique a été affinée avec le modèle isotrope. De plus, quelques conformations multiples étaient présentes dans le modèle.

8.2.2 Affinement de la structure

Le processus d'affinement est une répétition des étapes de modification du modèle, affinement et vérification du modèle après affinement. Il n'est pas rare de revenir à une étape antérieure lorsqu'une modification s'est avérée

comme incorrecte. Pour cette raison, seulement la démarche générale d'affinement est décrite. L'affinement et les résultats pour les deux structures de PfluDING étant similaires, sauf mention contraire, le texte réfèrera à la structure cristallisée à pH 4.5.

Dans les cartes de densité résiduelle du modèle initial, certains pics de densité électronique indiquaient la présence de nombreuses conformations multiples, en majorité des chaînes latérales mais également de la chaîne principale (figure 8.2). Ces conformations multiples ont donc été identifiées et affinées. Au total, 95 conformations multiples ont été modélisées. Par rapport aux facteurs d'accord statistiques initiaux, on observe une légère amélioration (tableau 8.2).

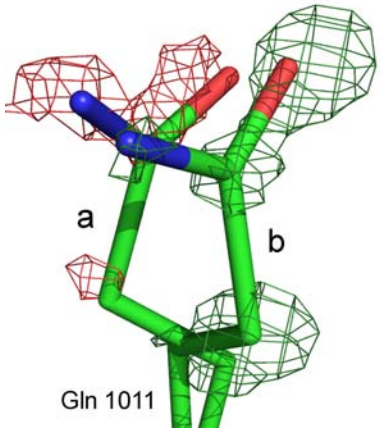


Figure 8.2 – La double conformation du résidu Gln 1011 dans la structure à pH 4.5 de PfluDING. Conformation **a** est modélisée avec une occupation de 100%, celle de conformation **b** est de 0%. Rouge : densité électronique résiduelle négative. Verte : densité électronique positive. Pour une meilleure visualisation, les atomes d'hydrogène ne sont pas représentés.

Ensuite, l'agitation thermique des atomes non-H de la protéine ainsi que de certaines molécules d'eau a été affinée avec un modèle anisotrope (figure 8.3). En effet, le mouvement thermique d'un atome n'est pas forcément isotrope mais peut varier en fonction des directions. L'introduction des paramètres de déplacement thermique anisotrope mène alors à une modèle plus réaliste permettant une meilleure description du mouvement thermique des atomes. On observe effectivement que les facteurs d'accord statistiques baissent chacun de 1 % environ, ce qui est significatif.
Les prochaines étapes de l'affinement sont effectuées avec le logiciel SHELXL

Affinement	R_{cryst} [%]	R_{free} [%]
REFMAC		
modèle initial	14.9	15.7
ajout doubles conformations	14.4	15.3
agitation thermique anisotrope	13.2	14.4
SHELXL		
agitation thermique isotrope	13.9	15.9
agitation thermique anisotrope	9.8	12.4
ajout atomes hydrogènes	9.1	11.3
MoPro		
facteur d'échelle et paramètres de solvant	8.5	10.7

Table 8.2 – Evolution des facteurs d'accord statistiques (R_{cryst}) pour l'affinement de la protéine PfluDING. Les facteurs d'accord statistique de SHELXL ont été calculés avec une limite de $F_o > 4\sigma(F_o)$ (172.600 réflections) et ceux de MoPro avec une limite de $\frac{F_{obs}-k\cdot F_{calc}}{\sigma(F_{obs})} > 100$ (176.518 réflections).

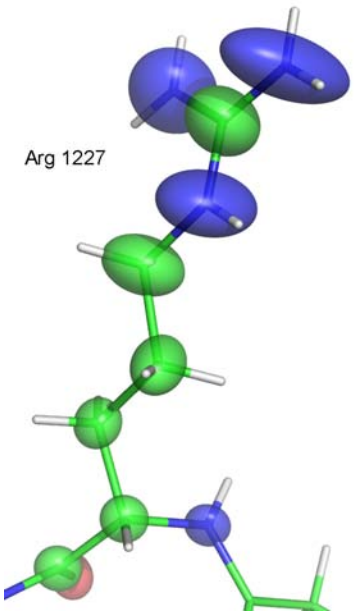


Figure 8.3 – Ellipsoïdes d'agitation thermique anisotrope des atomes non H du résidu Arg 1227 dans la structure cristallisée à pH 4.5. Notons que l'aspect anisotrope de l'agitation thermique augmente lorsque les atomes sont loin de la chaîne principale.

[118]. Initialement prévu pour l'affinement de petites molécules, SHELXL est également utilisé pour affiner les structures de protéines diffractant à haute résolution. Le logiciel utilise la méthode des moindres carrés, tandis que REFMAC utilise la méthode de maximum de vraisemblance [119].
Avec SHELXL, l'affinement a été effectué d'abord avec le modèle d'agitation thermique isotrope, et ensuite avec le modèle anisotrope. Notons que le passage est de nouveau accompagné par une baisse des facteurs d'accord statistique.
L'étape finale consistait à ajouter les atomes d'hydrogènes dits "fixes", donc qui n'ont pas de degrés de liberté de rotation. On observe encore une fois que les facteurs d'accord diminuent nettement, surtout le facteur R_{free}.
Enfin, la structure a été affinée avec MoPro. Des tentatives ont été effectuées pour affiner sans restrictions les paramètres d'agitation thermiques après un transfert des paramètres multipolaires, mais elles n'ont pas abouti à des résultats réalistes. Pour cette raison, un affinement conventionnel a été effectué. Dans un premier temps, le facteur d'échelle a été affiné, suivi par les paramètres de solvant (utilisant le modèle d'enveloppe de solvant [120]). En effet, dans un cristal de protéine, il y a une grande quantité de solvant qui diffracte surtout à basse résolution, le module du facteur de structure de la protéine y est donc surestimé. Or, l'utilisation d'un modèle de solvant permet de prendre en compte la contribution du solvant au facteur de structure observé (F_{obs}) et le modèle peut être amélioré. Concernant les facteurs d'accord statistiques, on n'observe pas de changement, mais notons que moins de réflexions sont rejetées, d'après le critère $\frac{F_{obs}-k\cdot F_{calc}}{\sigma(F_{obs})} > 100$ et $I/\sigma(I) = 0$.
Ensuite, les coordonnées et l'agitation thermique de l'ion phosphate ainsi que de la chaîne latérale du résidu Asp 62 ont été affinées sans contraintes ou restrictions, pour obtenir des paramètres de stéréochimie non biaisés des groupements du site actif impliqués dans le mécanisme.
Après l'affinement des deux structures de PfluDING, la densité électronique de nombreux atomes d'hydrogène, surtout de ceux localisés dans les régions de faible agitation thermique, devient visible dans les cartes de densité résiduelles (figure 8.4). Ceci a permis de positionner les atomes H fixes ou mobiles du site actif sur la base de pics significatifs de densité électronique.

8.2.3 Réseau de liaisons hydrogène dans le site de fixation du phosphate

Dans la structure à pH 4.5, on observe des forts pics de densité électronique, autour de l'atome d'oxygène O3 du phosphate, qui correspondent à des atomes d'hydrogène (figure 8.5 a). Notons que même la densité électro-

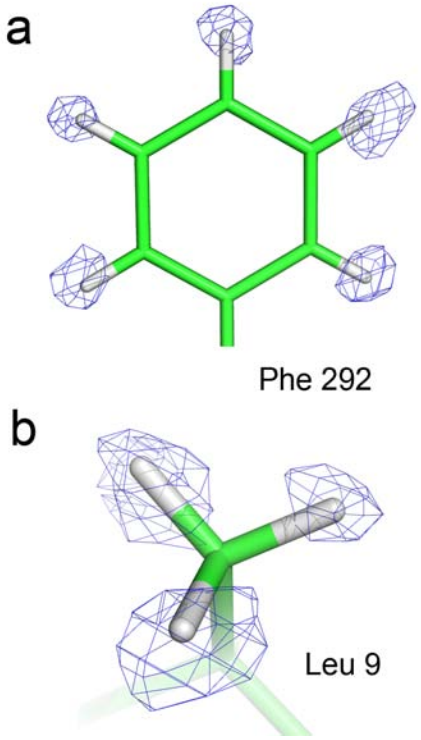


Figure 8.4 – Visibilité des atomes d'hydrogène dans les structures de PfluDING. (a) chaîne latérale de Phe 292 dans la structure à pH 4.5 et carte de densité électronique omit H, contourée à 3.6 σ. (b) Leu 9 dans la structure à pH 8.5, la carte de densité omit H est contourée à 3.1 σ. (La figure est partiellement extraite de [84]).

nique de l'atome d'hydrogène mobile ($H\gamma$) de Ser 32 est clairement visible, pointant vers l'atome O3. De plus, il n'y a pas de pics de densité significatifs aux alentours de l'atome d'oxygène. Ainsi, O3 accepte trois liaisons hydrogène dans une coordination tetraédrique, les groupements N-H de la chaîne principale de Gly 146 et de Ser 32 ainsi que la chaîne latérale de Ser 32 ont le rôle de donneurs. On peut donc conclure que l'atome O3 n'est pas protoné, étant accepteur de trois liaisons hydrogène.
Une analyse similaire peut être effectuée pour les atomes d'oxygène O1 et O4 de l'ion phosphate. L'atome O1 accepte 3 liaisons hydrogène dans une coordination tetraédrique. Les donneurs sont un groupement amide de la chaîne principale de Thr 8 ainsi que l'atome Hγ mobile du même résidu, et la partie terminale N-H de Arg 141. De même, l'atome d'oxygène O4 est impliqué dans 4 interactions, il accepte des liaisons hydrogène par l'amide N-H de la chaîne principale de Thr147, les groupements hydroxyl de Thr 147

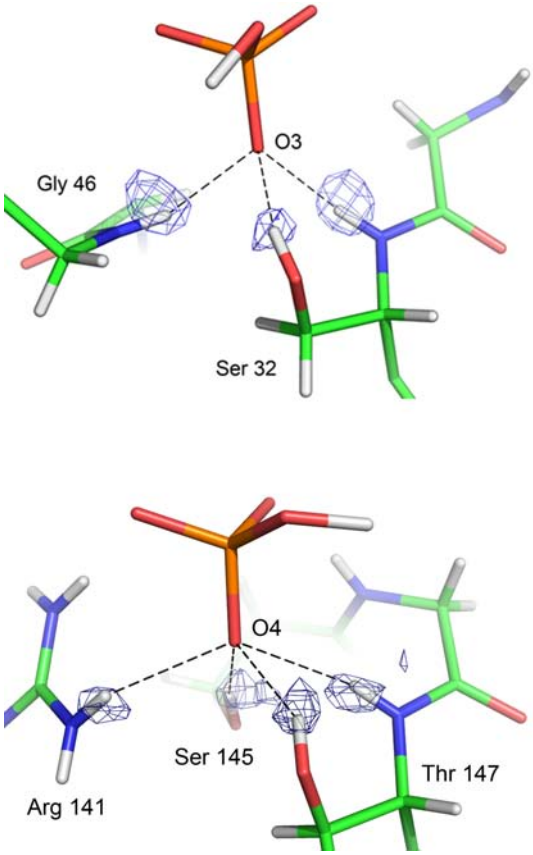


Figure 8.5 – (a) coordination de l'atome d'oxygène O3 du phosphate dans la structure à pH 4.5. La carte de densité omit H est contourée à 2.5 σ. Les interactions phosphate-protéine sont représentées par des lignes pointillées. (b) coordination de l'atome O4 dans la structure à pH 8.5. Le niveau de contour de la carte omit H est de 2.5 σ. (La figure (a) est extraite de [84]).

et de Ser 145, et la partie N-H de la chaîne latérale de Arg 141. Notons que la dernière interaction, impliquant Arg 141, est moins forte, puisque l'angle de liaison N-H...O de 133° est nettement plus petit que 180°. En résumé, les atomes d'oxygène O1, O3 et O4 sont impliqués dans au moins trois liaisons hydrogènes et ne peuvent pas être protonés [3]. Le cas de l'atome O2 est moins évident et va être discuté dans la prochaine section.

3. Les distances et les angles des interactions de l'ion phosphate sont résumés dans la publication à la fin de ce chapitre.

8.2.4 L'interaction phosphate O2 - Asp 62 $O\delta 2$

L'atome d'oxygène O2 du phosphate accepte une liaison H du groupement amide de Leu 9. L'atome Oδ2 de la chaîne latérale de Asp 62 est situé à proximité de O2, la distance O2...Oδ2 est remarquablement courte : Elle est de 2.502(6) et 2.495(7) Å dans les structure$_{pH4.5}$ et structure$_{pH8.5}$, respectivement. Notons que toutes les autres distances O...X dans le site actif, impliquant un atome d'oxygène du phosphate et son donneur de liaison hydrogène, sont plus longues. Dans la carte de densité électronique résiduelle de la structure à pH 4.5, on observe un faible pic de densité entre les atomes d'oxygène O2 et Oδ2. Il est distant de 0.88 Å par rapport à O2 et le maximum du pic a un contour de 2.2 σ. Dans la structure à pH 8.5 se trouve également un faible pic de densité entre O2 et Oδ2. Le pic est plus distant de O2 (à 1 Å de distance) et il apparaît à un niveau de contour de 2.3 σ. La position des pics est compatible avec les géométries habituelles des liaisons hydrogènes [55], mais puisqu'ils sont nettement plus faibles que ceux des autres atomes H dans le site actif, on ne peut pas les attribuer incontestablement à des atomes d'hydrogène. En effet, on pourrait s'attendre à ce qu'un atome H lié à O2 soit bien défini, la distance O...O étant courte et la géométrie favorable, et que sa densité électronique soit visible, comme celle des autres atomes d'hydrogène interagissant avec le phosphate.

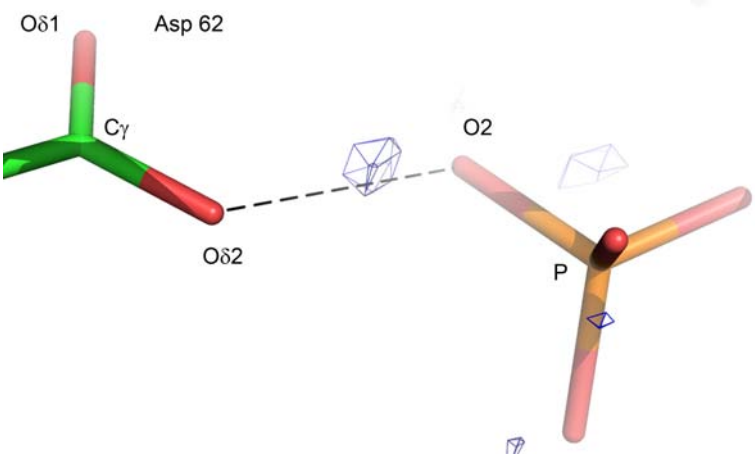


Figure 8.6 – Interaction entre l'atome O2 de l'ion phosphate et l'atome d'oxygène Oδ2 de la chaîne latérale de Asp 62 dans la structure à pH 4.5. La carte de densité électronique résiduelle en bleue est contourée à 2 σ. (La figure est extraite de [84]).

Cependant, plusieurs phénomènes peuvent être à l'origine d'un affaiblissement de la densité électronique d'un atome d'hydrogène : Par exemple, si

l'agitation thermique de l'atome H est élevée, sa densité électronique est étalée et elle peut être difficilement distinguable du bruit expérimental. L'agitation des atomes dans le site actif a été analysée (table 8.3), et il s'avère qu'elle est plutôt faible pour une protéine. Pour cette raison, il est peu probable que l'agitation thermique soit responsable de l'affaiblissement de la densité électronique.

	atom	**structure**$_{pH4.5}$	**structure**$_{pH8.5}$
		Agitation thermique [Å^2]	
phosphate	P	3.16(8)	3.3(1)
	O1	3.4(3)	3.9(3)
	O2	2.8(2)	3.5(3)
	O3	3.1(2)	2.8(3)
	O4	3.2(3)	3.2(3)
Asp 62	Oδ2	4.2(3)	3.9(3)
	Oδ1	4.3(3)	3.8(3)
	Cγ	3.3(5)	3.1(5)
		Longueurs de liaison [Å]	
phosphate	O1	1.522(6)	1.529(6)
distances P-O	O2	1.561(7)	1.559(7)
	O3	1.533(7)	1.528(7)
	O4	1.555(8)	1.541(8)
Asp 62	Oδ2	1.26(1)	1.26(1)
distances Cγ-O	Oδ1	1.26(1)	1.22(1)
distance Oδ2-O2		2.502(6)	2.495(7)

Table 8.3 – Facteurs d'agitation thermique isotrope équivalent (B_{eq}) et longueurs de liaison de l'ion phosphate ainsi que de la chaîne latérale d'Asp 62, issu de l'affinement non-restreint du site actif. *Remarque* : Les erreurs sont sous-estimées car l'affinement a été effectué seulement sur les atomes du site acif.

D'un autre côté, l'atome d'hydrogène pourrait être désordonné et avoir une double conformation [121] le long de l'interaction O2...Oδ2. La distance courte de 2.5 Å entre les deux atomes d'oxygène est en effet typique pour les "Low Barrier Hydrogen Bond" (LBHB) [122] [123].
Dans ce type particulier de liaison hydrogène, l'atome H est distribué entre les deux atomes d'oxygène et sa position est diffuse (figure 8.7). Par conséquent, il est difficile d'observer la densité de cet atome H par diffraction des

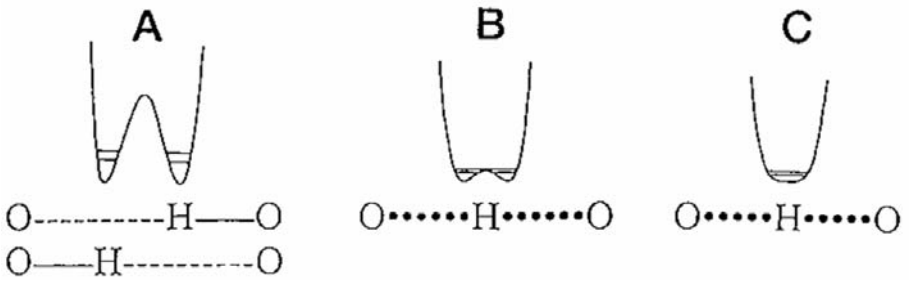


Figure 8.7 – Diagrammes d'énergie de liaisons hydrogènes pour différentes distances O...O entre les atomes d'oxygène. Les lignes horizontales représentent l'énergie du point zéro. (A) 2.8 Å. Il y a deux puits de potentiel, l'atome d'hydrogène est donc lié à un des deux atomes d'oxygène car la barrière centrale entre les positions d'équilibre est élevée. (B) 2.55 Å. La barrière centrale est pratiquement au même niveau que l'énergie zéro. L'atome H est libre de se déplacer dans les puits entre les deux atomes d'oxygène. (C) 2.3 Å. Il n'y a qu'un seul minimum et l'atome est partagé. (La figure est extraite de [123]).

RX ou même par diffraction des neutrons [123]. Dans le cas de PfluDING, la densité électronique faible de l'atome d'hydrogène suggère que sa position soit diffuse. Ainsi, l'atome H est probablement distribué entre deux puits de potentiel asymétriques, la position principale étant plus près de l'atome O2 de l'ion phosphate.
Les longueurs de liaisons P-O dans le phosphate soutiennent ce résultat. D'après une analyse de structures de phosphates dibasiques (HPO_4^{2-}), déposées dans la banque de données structurales de Cambridge [124], la longueur de liaison P-O de l'atome d'oxygène protoné est plus élevée. L'analyse a été effectuée sur 31 structures, résolues par diffraction des RX, qui contiennent un phosphate simplement protoné. Un critère supplémentaire était que le facteur d'accord statistique devrait être meilleur que 5%. La distance moyenne de la longueur de liaison P-O de l'atome d'oxygène protoné est de 1.59(1) Å, celle des atomes O non protonés est de 1.52(1) Å.
Dans la structure de PfluDING (pH 4.5), la longueur de liaison P-O2 est de 1.561(7) Å (tableau 8.3), elle est alors moins élevée que celle d'un atome d'oxygène complètement protoné. Ce résultat supporte donc l'hypothèse d'une protonation partielle de l'atome O2. Quant aux distances P-O1 et P-O3, elles sont de 1.522(6) et 1.533(7) Å, respectivement, ce qui correspond à la distance P-O impliquant des atomes d'oxygène non protonés. On observe que la longueur P-O4 est moins élevée que celle de P-O2 mais cependant plus élevée que les distances P-O1 ou P-O3. Notons que O4 est l'atome oxygène qui accepte 4 liaisons hydrogènes, il est ainsi probable que O4 soit plus polarisé et, par conséquent, que la distance P-O4 soit allongée.

Ainsi, l'étude structurale de la protéine PfluDING montre que l'ion phosphate possède partiellement un proton, même à pH 4.5, où on s'attendrait à avoir la forme doublement protoné du phosphate. PfluDING fixe alors le phosphate dibasique au deux pH (4.5 et 8.5) et le réseau de liaisons H dans le site de fixation est identique dans les deux cas. Le mode de fixation du phosphate dibasique est d'ailleurs identique à celui reporté pour les SBP. Cependant, l'hypothèse sur le mode de fixation du phosphate monobasique ($H_2PO_4^-$), proposant qu'un groupement hydroxyl (qui correspond à Ser 32 dans PfluDING) s'oriente différemment pour accepter le deuxième proton du phosphate, ne pouvait pas être confirmée. A pH 4.5, l'atome d'hydrogène Hγ de la chaîne latérale de Ser 32 pointe vers l'atome O3 du phosphate et forme clairement une liaison hydrogène (figure 8.5 a). Le mode de fixation du phosphate par PfluDING est résumé figure 8.8.

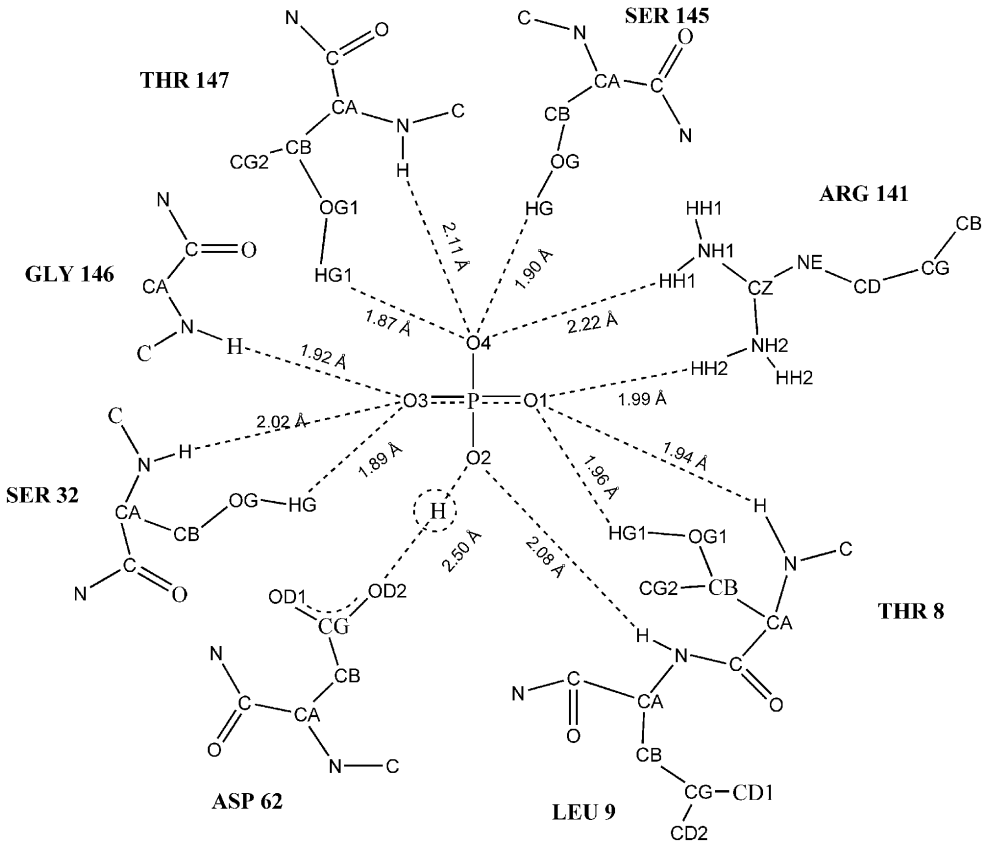


Figure 8.8 – Diagramme schématique du réseau de liaison hydrogène, impliquant l'ion phosphate, qui a été déterminé expérimentalement. Les distances entre les accepteurs et les atomes H sont indiquées. (La figure est extraite de [84]).

8.2.5 Potentiel électrostatique de PfluDING

Pour calculer le potentiel électrostatique de la protéine, le modèle obtenu en fin d'affinement avec MoPro a été utilisé. Les charges atomiques et les paramètres multipolaires ont été transférés de la banque de données ELMAM [125] aux atomes de la protéine. Les longueurs de liaison covalentes X-H ont été allongées à des valeurs standards obtenues par diffraction des neutrons [97] (voir aussi section 7.3.3). Ensuite, la protéine a été neutralisée pour que la somme de toutes les charges soit égale à zéro. La modification de la charge appliquée à chaque atome était $2{\cdot}10^{-3}$ électrons par atome, ce qui est inférieur à la déviation standard observée normalement pour les populations de valence affinées des structures de petites molécules ou polypeptides. Finalement, le potentiel électrostatique de déformation a été calculé avec VMoPro [18]. Les molécules d'eau et de solvant ainsi que l'ion phosphate n'ont pas été inclus dans le calcul, de façon à obtenir le potentiel électrostatique généré par le site actif uniquement.

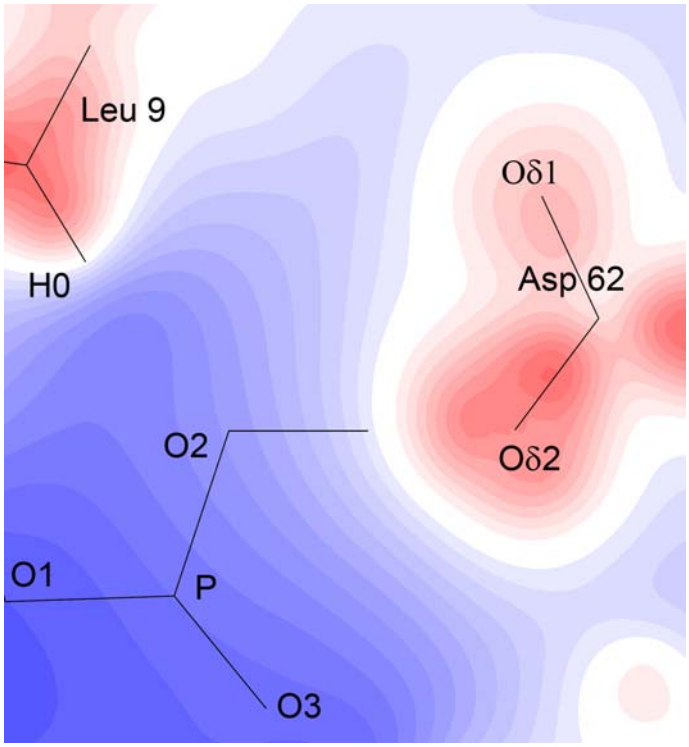


Figure 8.9 – Potentiel électrostatique de déformation dans le plan H0...O2...Oδ2. L'échelle s'étend de -1 $|e|$ /Å (rouge) à +1 $|e|$ /Å (bleu), les lignes de contour sont de 0.05 $|e|$ /Å. Le potentiel zéro est représenté en blanc. (La figure est extraite de [84]).

D'après les calculs, le potentiel électrostatique dans le site de fixation de PfluDING est largement positif (figure 8.9). Notons que même le potentiel autour du résidu négatif Asp 62 est presque neutre. Il y a plusieurs dipôles locaux (groupements N-H et O-H) et deux résidus chargés dans le site actif : un acide aspartique (Asp 62) et une arginine (Arg 141). Le potentiel positif est partiellement dû aux charges positives des atomes H interagissant avec le phosphate. La présence du résidu arginine et de parties N-terminales de 4 hélices α amplifient cet effet. Le potentiel électrostatique positif de la cavité de fixation est en accord avec la capacité de PfluDING de fixer un ion négatif. De manière intéressante, il a été proposé par Ledvina *et al.* que le potentiel du site actif dans les protéines SBP soit négatif [126]. Le calcul a été effectué avec le logiciel GRASP [127], qui utilise l'équation de Poisson-Boltzmann pour obtenir le potentiel. Dans le cas de la PBP, les valeurs de défaut pour les charges (full charges) et de concentration de sel ont été appliquées. Par conséquent, des charges formelles (+1 ou -1) ont été utilisées pour les acides aminés lysine, arginine, acide glutamique, acide aspartique et les acides aminés N et C-terminaux uniquement. Les charges partielles positives des atomes d'hydrogène n'ont donc pas été considérées explicitement, ce qui est contraire au calcul de VMoPro qui inclut toutes les charges atomiques partielles ainsi que les mulipôles. Comme mentionné ci-dessus, de nombreux dipôles locaux de groupements N-H et O-H se trouvent dans le site actif et le potentiel positif dans la cavité de fixation est essentiellement dû à ces groupements polarisés ainsi qu'au résidu arginine positif. Notons que le site actif ne contient qu'un seul acide aminé négatif (Asp 62). Puisque dans [126] les dipôles locaux sont négligés et uniquement les résidus à charges formelles sont pris en compte, il est probable que le potentiel calculé avec GRASP soit trop simplifié, ce qui résulte en un potentiel globalement négatif. Le potentiel de la protéine PBP a été recalculé avec MoPro (en procédant de la même manière que pour le calcul du potentiel de PfluDING) en incluant cette fois-ci des charges partielles pour tous les atomes de la protéine, et d'après ces nouveaux calculs, le potentiel est positif (figure 8.10). Par conséquent, il est trop approximatif de considérer seulement les chaînes latérales chargés pour calculer le potentiel électrostatique dans ce type de cavité [4].

8.2.6 Mécanisme moléculaire de fixation

PfluDING fixe le phosphate dibasique (HPO_4^{2-}) à pH 4.5 et 8.5, bien que le phosphate monobasique soit prépondérant en solution à pH acide. Ce

4. *Remarque :* Le calcul de potentiel sur la protéine PBP a été effectué par B. Fournier, alors doctorant du laboratoire.

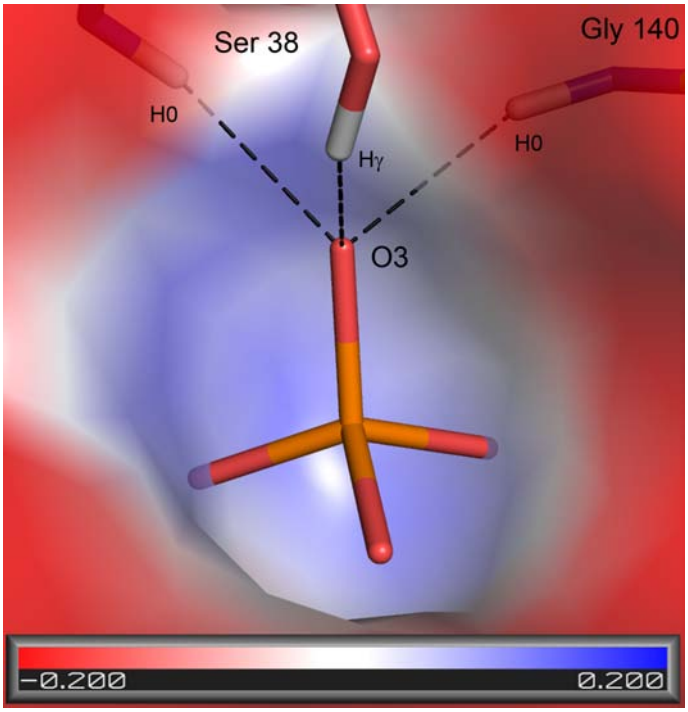


Figure 8.10 – Potentiel électrostatique de déformation dans le site actif de la protéine PBP (code PDB 1IXH), résolue à 0.98 Å. Le potentiel est représenté sur la surface de quelques résidus uniquement (pour une meilleure visualisation de la cavité), le rayon est 1.4 Å. L'échelle s'étend de -0.2 $|e|$ /Å (rouge) à 0.2 $|e|$ /Å (bleu). *Remarque :* Le potentiel autour de la cavité apparait négatif artificiellement puisque cette région correspond à la partie du modèle qui a été tronquée pour visualiser la cavité.

résultat ainsi que le potentiel positif du site actif suggèrent que le site de fixation de PfluDING serait capable d'attirer et de stabiliser le phosphate dibasique. Les nombreux donneurs de liaisons hydrogène dans le site actif favorisent l'absence d'un proton sur les atomes d'oxygène O1, O3 et O4 du phosphate. En plus, la formation d'une LBHB entre Asp 62 et l'atome O2 du phosphate suggère que le pKa des deux partenaires soit similaire [128]. En solution, le pKa de $H_2PO_4^-/HPO_4^{2-}$ est plus élevé que celui de la chaîne latérale d'un acide aspartique, ce qui souligne la forte influence du potentiel électrostatique positif qui baisse le pKa de l'ion phosphate. D'un autre côté, on s'attend à ce que le pKa du carboxylate de Asp 62 augmente en raison de l'effet de désolvatation. Les résultats sur PfluDING sont en accord avec des calculs d'énergie dans les protéines SBP. Pour déterminer la contribution de la LBHB entre O2 et Oδ2, un mutant D56N a été créé [114] qui remplace l'acide aminé chargé Asp 56, impliqué dans cette courte

liaison H, par une asparagine. Dans le mutant [5], la distance O...Oδ2(Asn 56) a augmenté et est de 2.6 Å, tandis qu'elle était 2.5 Å dans la protéine sauvage [6]. La LBHB s'est alors transformé en liaison hydrogène "classique". L'énergie de liaison a changé seulement légèrement, $\Delta\Delta G$ entre la protéine sauvage et le mutant est de +0.3 kcal/mol, bien que des énergies de 12-24 kcal/mol soient attendus pour des LBHB [129]. Or, l'énergie de liaison entre l'ion phosphate et la protéine est constituée de liaisons hydrogènes et d'attraction électrostatique. Dans le mutant, la contribution des liaisons H a été diminuée puisque la LBHB n'existe plus. D'un autre côté, le résidu négatif (Asp 56) a été remplacé par un acide aminé neutre (Asn) et il est probable que le potentiel électrostatique positif dans la cavité augmente. L'interaction avec l'ion phosphate, qui est chargé négativement, est ainsi renforcée. De cette manière, la modification des charges dans le site actif du mutant compense la perte de la LBHB et explique pourquoi l'énergie libre du mutant et de la forme sauvage était similaire.

8.3 Conclusion

Dans les structures atomiques de PfluDING, obtenus à différents pH (4.5 et 8.5), de nombreux atomes d'hydrogènes pouvaient être localisés. Il était ainsi possible de déterminer expérimentalement le mécanisme moléculaire responsable de la fixation du phosphate. En particulier, PfluDING fixe l'ion phosphate dibasique au deux pH et en utilisant le même réseau de liaison hydrogènes. D'après des calculs de potentiel électrostatique, le potentiel du site actif est positif, contrairement à des résultats publiés sur les protéines SBP. Il est possible que le potentiel positif et de nombreux donneurs de liaisons H favorisent le phosphate sous la forme simplement protonée ce qui explique pourquoi PfluDING fixe le phosphate dibasique même à un pH acide. L'ion phosphate accepte 11 liaisons hydrogène et partage son seul proton via une interaction LBHB avec Asp 62. Cette LBHB dans PfluDING est probablement responsable de sa spécificité à fixer le phosphate dibasique et permet de discriminer des molécules similaires comme le sulfate.

5. La structure a été résolue a 1.9 Å et le code PDB est 1IXI
6. La résolution de cette structure est 0.98 Ået le code PDB est 1IXH

8.4 Perspectives

La protéine PfluDING a la capacité de former de gros monocristaux qui ont le potentiel de diffracter à résolution subatomique. Des tests préliminaires au synchrotron ont montré des taches de diffraction des rayons X jusqu'à 0.7 Å. Il est possible d'optimiser les conditions de cristallisation, par exemple en utilisant des techniques d'ensemencement, pour obtenir encore de meilleurs cristaux. Avec des données de diffraction d'une résolution subatomique, le mode de fixation peut être analysée encore plus en détail. De plus, un affinement des paramètres multipolaires peut être envisagé.
Une autre perspective est de faire une expérience de diffraction des neutrons. Cela permettra en particulier d'analyser plus en détail la LBHB. La longueur de diffusion des atomes d'hydrogène est comparable à celle des autres atomes protéiques, et, probablement, la conformation multiple de l'atome H serait ainsi mieux visible que par la diffraction des rayons X.

Published on Web 05/15/2009

Elucidation of the Phosphate Binding Mode of DING Proteins Revealed by Subangstrom X-ray Crystallography

Dorothee Liebschner,[†] Mikael Elias,[‡] Sèbastien Moniot,[II,†] Bertrand Fournier,[†] Ken Scott,[§] Christian Jelsch,[†] Benoit Guillot,[†] Claude Lecomte,[†] and Eric Chabrière*[,‡]

Cristallographie Resonnance Magnetique et Modelisations, CNRS UMR 7036 Nancy-Universite, 54506 Vandoeuvre-les-Nancy, France, Architecture et Fonction des Macromolecules Biologiques, CNRS-Universite de la Mediterranee, 13288 Marseille, France, and School of Biological Sciences, The University of Auckland, Auckland, New Zealand

Received March 11, 2009; E-mail: eric.chabriere@afmb.univ-mrs.fr

Downloaded by UNIV HENRI POINCARE on August 10, 2009
Published on May 15, 2009 on http://pubs.acs.org | doi: 10.1021/ja901900y

Abstract: PfluDING is a bacterial protein isolated from *Pseudomonas fluorescens* that belongs to the DING protein family, which is ubiquitous in eukaryotes and extends to prokaryotes. DING proteins and PfluDING have very similar topologies to phosphate Solute Binding Proteins (SBPs). The three-dimensional structure of PfluDING was obtained at subangstrom resolution (0.88 and 0.98 Å) at two different pH's (4.5 and 8.5), allowing us to discuss the hydrogen bond network that sequesters the phosphate ion in the binding site. From this high resolution data, we experimentally elucidated the molecular basis of phosphate binding in phosphate SBPs. The phosphate ion is tightly bound to the protein via 12 hydrogen bonds between phosphate oxygen atoms and OH and NH groups of the protein. The proton on one oxygen atom of the phosphate dianion forms a 2.5 Å low barrier hydrogen bond with an aspartate, with the energy released by forming this strong bond ensuring the specificity for the dianion even at pH 4.5. In particular, contrary to previous theories on phosphate SBPs, accurate electrostatic potential calculations show that the binding cleft is positively charged. PfluDING structures reveal that only dibasic phosphate binds to the protein at both acidic and basic phosphate, suggesting that the protein binding site environment stabilizes the HPO_4^{2-} form of phosphate.

Introduction

DING proteins constitute an intriguing protein family that was identified in a wide range of organisms, from prokaryotes to eukaryotes.[1−3] The name of these 38 kDa proteins comes from the sequence of their four conserved N-terminal residues. Despite the seemingly ubiquitous occurrence of DING proteins in animals, plants, and fungi, they are systematically absent from eukaryotic genome databases.[4] Consequently, available eukaryotic DING protein sequences are essentially constituted of N-terminal sequencing and internal peptide sequences. The sequence of the Human Phosphate Binding Protein (HPBP) is the only complete eukaryotic DING protein sequence and was obtained by a tandem use of X-ray crystallography and mass spectrometry.[5] HPBP is a hydrophobic plasmatic apolipoprotein associated with human paraoxonase, for which heterologous production remains problematic.[6] The tertiary structure of HPBP confirmed that DING proteins possess very similar topologies as bacterial phosphate solute binding proteins (SBPs). They fit a model known as the "Venus flytrap", in which two globular domains hinge together to form the phosphate-binding site, with eight conserved residues H-bonded to phosphate.[7,8] Recently, a bacterial DING protein from *Pseudomonas fluorescens* was isolated and successfully expressed. It possesses a high sequence similarity with eukaryotic DING proteins,[9] suggesting that the protein family extends to the prokaryotes.[4] This protein has mitogenic activity toward human fibroblasts, which is consistent with similar activity ascribed to human DING proteins.[10−12] There is only one complete DING gene sequence from eukaryotes. As PfluDING yields crystals diffracting to very high resolution, it offers the most convenient model for pursuing

[II] Present address: Institut für Chemie and Biochemie/Kristallographie, Freie Universität Berlin, 14195 Berlin, Germany.
[†] CNRS UMR 7036 Nancy-Universitè.
[‡] CNRS-Universitè de la Mèditerranèe.
[§] The University of Auckland.

(1) Berna, A.; Bernier, F.; Scott, K.; Stuhlmüller, B. *FEBS Lett.* **2002**, *524*, 6–10.
(2) Maro, A. D.; Maio, A. D.; Castellano, S.; Parente, A.; Farina, B.; Faraone-Mennella, M. R. *J. Biol. Chem.* **2009**, *390*, 27–30.
(3) Pantazaki, A. A.; Tsolkas, G. P.; Kyriakidis, D. A. *Amino Acids* **2008**, *34*, 437–448.
(4) Berna, A.; Bernier, F.; Chabrière, E.; Perera, T.; Scott, K. *Int. J. Biochem. Cell Biol.* **2008**, *40*, 170–175.
(5) Diemer, H.; Elias, M.; Renault, F.; Rochu, D.; Contreras-Martel, C.; Schaeffer, C.; Dorsselaer, A. V.; Chabriere, E. *Proteins* **2008**, *71*, 1708–1720.
(6) Rochu, D.; Chabriere, E.; Renault, F.; Elias, M.; Clery-Barraud, C.; Masson, P. *Biochem. Soc. Trans.* **2007**, *35*, 1616–1620.
(7) Morales, R.; et al. *Ann. Pharm. Fr.* **2007**, *65*, 98–107.
(8) Morales, R.; et al. *Structure* **2006**, *14*, 601–609.
(9) Scott, K.; Wu, L. *Biochim. Biophys. Acta* **2005**, *1744*, 234–244.
(10) Hain, N.; Stuhlmuller, B.; Hahn, G.; Kalden, J.; Deutzmann, R.; Burmester, G. *J. Immunol.* **1996**, *157*, 1773–1780.
(11) Adams, L.; Davey, S.; Scott, K. *Biochim. Biophys. Acta* **2002**, *1586*, 254–264.
(12) Belenky, M.; Prasain, J.; Kim, H.; Barnes, S. *J. Nutr.* **2003**, *133*, 2497S–2501S.

10.1021/ja901900y CCC: $40.75 © 2009 American Chemical Society

Downloaded by UNIV HENRI POINCARE on August 10, 2009
Published on May 15, 2009 on http://pubs.acs.org | doi: 10.1021/ja901900y

structure−function relationships in DING proteins. The PfluDING structure was determined and found to be highly similar to that of HPBP.[13,14] Mutant proteins deficient in phosphate-binding ability and truncated forms of PfluDING suggest that the mitogenic activity of PfluDING is related to the Venus flytrap motion of the structure induced by the phosphate binding.[13]

The binding mechanism of phosphate SBPs was previously investigated, showing that the phosphate is sequestered in a conserved cleft, completely buried in the closed form of the protein.[8] The dibasic phosphate anion was described as being bound by a rich network comprising 12 hydrogen bonds. Eleven of them involve dipolar donor groups with five backbone N−H groups, mostly located at the N-terminal end of four helices; two N−H side chains from an Arg residue; and four from side chain O−H hydroxyl groups (e.g., Ser 32 in HPBP and PfluDING). An Asp carboxylate, the only acceptor group in the pocket, makes a hydrogen bond with the only proton available on dibasic phosphate.[8,13] This key residue is believed to be responsible for the discrimination by 5 orders of magnitude between the closely related chemical species, dibasic phosphate and sulfate.[15] Indeed, sulfate at physiological pH does not possess any proton and will be repelled by the aspartate residue. This hydrogen bond between the anion and the Asp carboxylate is unusual, as it was described to be particularly short in phosphate binding proteins[13,16] including an atomic resolution structure at 0.98 Å.[17] Despite the short O...O distance, this hydrogen bond, surprisingly, is claimed not to be stronger than other hydrogen bonds.[17] Since phosphate SBPs have been shown to bind phosphate with high affinity at pH 8 and 4.5, the binding mode of monobasic phosphate was hypothesized.[15] Because the γ-OH of Ser 32 is the only other group favorably positioned to accept an oxygen electron lone pair of a second phosphate proton (on O3), it was proposed that this hydroxyl group flipped to accept the second hydrogen of monobasic phosphate.[15] The electrostatic potential of the protein was calculated, and surprisingly, the binding cleft was described as negative although it binds a negatively charged ligand.[18] It has been proposed that this peculiar feature of an anion bound within a negative environment would increase the discrimination between negative substrates. In this way, it was hypothesized that only anions perfectly matching the hydrogen bond network are capable of binding and compensating the unfavorable electrostatic interaction of the anion.[18]

Here, we report the subangstrom structure of PfluDING at two different pH values (4.5 and 8.5). The quality of the obtained diffraction data gives new insights into the molecular basis of phosphate binding via the direct observation of hydrogen atoms. Moreover, this highly accurate model has also permitted us to calculate precise electrostatic potential maps of the phosphate binding cavity, showing that the binding cleft has a globally positive potential.

Table 1. Data Collection and Refinement Statistics

	$crystal_{pH4.5}$	$crystal_{pH8.5}$
wavelength (Å)	0.953	0.750
space group	$P2_1$	$P2_1$
unit cell parameters (Å, deg)	$a = 36.97$	$a = 36.96$
	$b = 124.39$	$b = 124.90$
	$c = 41.02$	$c = 40.92$
	$\beta = 116.55$	$\beta = 116.31$
resolution (Å)	50−0.98	50−0.88
no. of unique reflections (last bin)	185 558 (44 846)	253 101 (42 675)
redundancy (last bin)	5.4 (3.1)	6.1 (3.1)
completeness (last bin)	98.1 (94.0)	97.4 (91.8)
R_{merge}[a] (last bin)	4.3 (7.5)	5.3 (43.8)
$\langle I/\sigma(I)\rangle$ (last bin)	24.4 (10.4)	16.6 (2.6)
R_{cryst}[b] (%)	8.44	12.76
R_{free}[c] (%)	10.71	14.81
water molecules	1234	1368
double conformations	95	97
average B factor (Å^2) (only protein atoms)	6.4	6.7
rmsd[d] bond length (Å)	0.014	0.014
rmsd[d] bond angle (deg)	2	2
Ramachandran analysis:[e]		
favored regions (%)	98.7	98.5
allowed regions (%)	1.3	1.5
outliers (%)	0	0

[a] . [b] after *MoPro* refinement. [c] Same definition as for R_{cryst}, but for 5% of the reflections which were not used for refinement. [d] Root mean square deviation from the Engh & Huber standard dictionary. [e] Calculated with Molprobity.

Materials and Methods

Crystallization and Data Collection. The recombinant DING protein from *P. fluorescens* was overexpressed in *Escherichia coli* and purified as described previously.[9] Crystals of PfluDING at pH 4.5 were obtained as described by Moniot et al.[14] Some of them were soaked for 24 h into a Tris-HCl 0.5 M buffer solution at pH 8.5 containing 26% PEG 8000 and Li_2SO_4 200 mM. Diffraction data collections were performed at the ESRF (Grenoble, France) at beamlines ID14-4 and ID-29. The mother liquor was used as a cryoprotectant, and all crystals were flash frozen in liquid nitrogen at 100 K prior to data collections. Two data collections were performed for one crystal of each pH value. One data set concerns the high resolution reflections ($d < 1.5$ Å), and the second data set collects the low resolution reflections. This strategy prevents overloaded measurements. X-ray diffraction data were integrated, scaled, and merged with the program *XDS*[19] (Table 1).

Refinement Procedure. Initial phases were obtained using the model of PfluDING deposited in the PDB (code: 2q9t).[14] The refinement was initially performed using *REFMAC*[20] and subsequently with *SHELXL 97*[21] and finally with *MoPro*.[22] The structure was manually built using the graphics program *Coot*.[23] Temperature factors for heavy atoms were refined anisotropically. To improve the model, hydrogen atoms were automatically generated using *SHELXL*. Hydrogen atoms able to rotate and located in the phosphate-binding site were omitted to avoid model bias. Thereafter, these hydrogen atoms were positioned manually according to significant electron density peaks in the $F_{obs} - F_{calc}$ residual map.

(13) Ahn, S.; Moniot, S.; Elias, M.; Chabrière, E.; Dim, D.; Scott, K. *FEBS Lett.* **2007**, *581*, 3455–3460.
(14) Moniot, S.; Elias, M.; Kim, D.; Scott, K.; Chabrière, E. *Acta Crystallogr., Sect. F* **2007**, *63*, 590–592.
(15) Luecke, H.; Quiocho, F. A. *Nature* **1990**, *347*, 402–406.
(16) Vyas, N. K.; Vyas, M. N.; Quiocho, F. A. *Structure* **2003**, *11*, 765–774.
(17) Wang, Z.; Luecke, H.; Yao, N.; Quiocho, F. A. *Nat. Struct. Biol.* **1997**, *4*, 519–522.
(18) Ledvina, P. S.; Tsai, A. L.; Wang, Z.; Koehl, E.; Quiocho, F. A. *Protein Sci.* **1998**, *7*, 2550–2559.
(19) Kabsch, W. *J. Appl. Crystallogr.* **1993**, *26*, 795–800.
(20) Murshudov, G. N.; Vagin, A. A.; Lebedev, A.; Wilson, K. S.; Dodson, E. J. *Acta Crystallogr., Sect. D* **1999**, *55*, 247–255.
(21) Sheldrick, G. M.; Schneider, T. R. *Methods Enzymol.* **1997**, *277*, 319–343.
(22) Jelsch, C.; Guillot, B.; Lagoutte, A.; Lecomte, C. *J. Appl. Crystallogr.* **2005**, *38*, 38–54.
(23) Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect. D* **2004**, *60*, 2126–2132.

During refinement, no constraints or restraints were applied on the geometry and bond lengths of the phosphate molecule. The last refinement steps were performed with *MoPro* using the ELMAM aspherical scattering factor library. First, the scaling factor and solvent parameters of the bulk solvent model[24] were refined. Then, the position and the thermal displacement parameters of the phosphate ion and side chain atoms of residue Asp 62 ($O_{\delta 1}$, $O_{\delta 2}$, C_{γ}, C_{β}, and C_{α}) were refined with *MoPro* in several least-squares refinement cycles. Neither stereochemical nor thermal motion restraints were applied to the refined atoms. Final refinement statistics are shown in Table 1.

Calculation of the Electrostatic Potential. The electrostatic potential can be calculated from accurate X-ray diffraction data, being the Fourier transform of $F(H)/H^2$ in reciprocal space,[25] or in direct space if the data are refined using a nonspherical atom model of the electron density.[26] The ELMAM experimental electron density database[27] was created by modeling the electron density of peptide crystals using the multipolar atom model (Hansen and Coppens model[28]). In this formalism, the multipolar atom electron density is described as a sum of three parts: the core, spherical valence, and multipolar electron density:

$$\rho_{\mathrm{atom}}(\vec{r}) = \rho_{\mathrm{core}}(r) + P_{\mathrm{val}}\kappa^3\rho_{\mathrm{val}}(\kappa r) + \sum_{l=0}^{l_{\max}} \kappa'^3 R_{nl}(\kappa' r) \sum_{m=-l}^{l} P_{lm} y_{lm}(\theta, \varphi) \quad (1)$$

ρ_{core} and ρ_{val} are the spherical core and valence electron density, respectively. The last term describes the nonspherical deformation density. P_{val} and P_{lm} are population coefficients, κ and κ' represent expansion/contraction parameters. R_{nl} are radial functions of Slater type, and y_{lm} are real spherical harmonic functions, permitting us to model the deviation from the spherical symmetry of the valence electron density.

Consequently, the electrostatic potential can also be subdivided:[29]

$$V(\mathbf{r}) = V_{\mathrm{neu}} + \Delta V_{\mathrm{val}} + \Delta V_{\mathrm{mul}} \quad (2)$$

V_{neu} is the potential computed from the spherical neutral atoms. ΔV_{val} is derived from the spherical valence atomic charges ($N_{\mathrm{val}} - P_{\mathrm{val}}$). ΔV_{mul} corresponds to the multipolar part of the electrostatic potential.

This formalism is implemented in the program *VMoPro*,[22] and electrostatic potential maps can be generated for any protein refined at atomic resolution. The structural model after the *MoPro* refinement has been used for the potential calculations. Atomic charges and multipole parameters (up to octupoles for C, O, and N atoms and to dipoles for H atoms) were transferred from the ELMAM database. The covalent X−H bonds have been elongated to standard neutron distances.[30] Afterward, the protein was neutralized electrically, so that the sum of all charges was zero. The charge shift was 2×10^{-3}electrons per atom, which is below the standard deviation usually observed for refined valence populations in small molecules. Finally, the electrostatic deformation potential $\Delta V(\mathbf{r})$ of the protein was calculated using only ΔV_{val} and ΔV_{mul} from eq 2. The phosphate moiety, water, and solvent molecules were not included to estimate only the protein potential.

The figures superposing the electron density or the electrostatic potential on the structural model were prepared with *Pymol*.[31]

Results

Structures of PfluDING. Two structures of PfluDING were determined at subangstrom resolution. One was obtained from crystals grown at pH 4.5 (0.98 Å, PDB code 3G62), and the other from crystals grown at pH 4.5 and subsequently soaked at pH 8.5 (0.88 Å, PDB code 3G63). Both structures are highly similar, so that in the following tables and figures, we refer to the structure solved at pH 4.5, if not otherwise indicated. The major advantage of subatomic resolution for protein structures is that most hydrogen atoms can be located in difference Fourier electron density maps. This is the case for the present structures of PfluDING. Electron density peaks of a large part of the omitted hydrogen atoms, especially in regions of low thermal motion, appear clearly in $F_{\mathrm{obs}} - F_{\mathrm{calc}}$ maps. In the active site, hydrogen atom positions were determined using a minimum of the 3σ level in the Fourier difference map and considering geometrical parameters lying within the usual ranges.[32] Thus, most of the fixed and mobile hydrogen atoms can be placed according to the electron density (Figure 1).

Phosphate Binding Site. The hydrogen atoms in the phosphate binding site are clearly visible in electron density maps. Hydrogen bonds that sequester the phosphate molecule are summarized in Table 2.

Around the oxygen atom O3 of the phosphate ion, there are clear electron density peaks for interacting hydrogen atoms (Figure 2a). Notably, the electron density corresponding to the mobile hydrogen atom H_{γ} of Ser 32 is clearly visible, pointing toward O3. Furthermore, no significant electron density peak appears close to O3. Thus, O3 accepts three hydrogen bonds in tetrahedral coordination; the donors are the main chain NH group of Gly 146 and Ser 32 as well as the Ser 32 side chain. Hence, it can be concluded that the phosphate O3 oxygen atom is unambiguously involved as the acceptor in three hydrogen bonds, meaning that O3 cannot be protonated. A similar analysis can be made for the oxygen atoms O1 and O4 of the phosphate moiety. The atom O1 accepts three hydrogen bonds in a tetragonal coordination. The backbone amide of Thr 8, the hydroxyl group of the same residue, and the N−H terminal group of Arg141 are donors. Concerning the oxygen atom O4, it interacts with four donors: the main chain N−H group of Thr 147 and the hydroxyl groups of Thr 147 and Ser 145 (Figure 2b). A fourth donor, a N−H of an Arg 141 side chain, forms a noncanonical hydrogen bond with O4 because the D−H...A angle (132.5°) is far from 180° (see Table 2).

The phosphate atom O2 accepts one hydrogen bond from the backbone N−H of Leu 9. Furthermore, O2 is in close proximity to $O_{\delta 2}$ from the carboxylate side chain of Asp 62. The oxygen−oxygen distance is 2.502(6) and 2.495(7) Å in the PfluDING structures at pH 4.5 and 8.5, respectively. A small peak in the residual electronic density map between phosphate O2 and Asp 62 $O_{\delta 2}$ atoms is visible at a 2.2 σ level and at a distance of 0.88 Å from atom O2 (see Table 2) but cannot be attributed with certainty to a hydrogen atom. A similar feature can be observed in the structure at pH 8.5, the electron density peak being more distant from atom O2 (1.00 Å) (Figure 3). Since the electron density can be smeared out by thermal motion,

(24) Fokine, A.; Urzhumtsev, A. *Acta Crystallogr., Sect. D* **2002**, *58*, 1387–1392.
(25) Stewart, R. F. *God. Jugosl. Cent. Kristalogr.* **1982**, *17*, 1–24.
(26) Ghermani, N.; Lecomte, C.; Bouhmaida, N. *Z. Naturforsch. A* **1993**, *48*, 91–98.
(27) Zarychta, B.; Pichon-Pesme, V.; Guillot, B.; Lecomte, C.; Jelsch, C. *Acta Crystallogr., Sect. A* **2007**, *63*, 108–125.
(28) Hansen, N. K.; Coppens, P. *Acta Crystallogr., Sect. A* **1978**, *34*, 909–921.
(29) Ghermani, N.; Bouhmaida, N.; Lecomte, C. *Acta Crystallogr., Sect. A* **1993**, *49*, 781–789.
(30) Allen, F. H. *Acta Crystallogr., Sect. B* **1986**, *42*, 515–522.
(31) DeLano, W. L. *Pymol*; DeLano Scientific LLC: San Carlos, CA, 2008 (http://www.pymol.org).
(32) Steiner, T. *Angew. Chem., Int. Ed.* **2002**, *41*, 48–76.

Downloaded by UNIV HENRI POINCARE on August 10, 2009
Published on May 15, 2009 on http://pubs.acs.org | doi: 10.1021/ja901900y

Downloaded by UNIV HENRI POINCARE on August 10, 2009
Published on May 15, 2009 on http://pubs.acs.org | doi: 10.1021/ja901900y

Figure 1. Seeing a clear view of the hydrogen atoms in PfluDING structures. (a) Water 5022 and its hydrogen bond network in the structure obtained at pH 4.5. (b) Close-up of Phe 292 side chain in the pH 4.5 structure. (c) Close-up of Leu 9 in the structure obtained at pH 8.5 The maps show the $F_{obs} - F_{calc}$ difference electron density, the contour level is 2.9 σ for (a), 3,6 σ for (b), and 3.1 σ for (c). Hydrogen atoms have been omitted in the structure factor calculation. The electron density of fixed (Phe 292) and mobile hydrogen atoms (Leu 9) sticks out clearly and evidences the quality of the diffraction data.

the phosphate ion and the aspartic acid have been further inspected. It appears that the thermal displacement parameters of the atoms are very low for a protein (Table 3). The B_{eq} factor of phosphate O2 and Asp 62 $O_{\delta 2}$ are 2.8(2) and 4.2(3) Å^2, respectively. Thus, it is surprising that this hydrogen atom does not appear more clearly in the electron density map. The P−O bond lengths of the phosphate molecule have been investigated. In phosphate moieties, longer P−O distances are found for protonated oxygen atoms. As observed in Table 3, the P−O bond length is the longest in both structures. Finally, the obtained structural data of PfluDING show clearly 11 hydrogen atoms involved in hydrogen bonds with the phosphate molecule (Figure 4). The O1, O3, and O4 phosphate oxygen atoms are each involved as acceptor atoms in at least three hydrogen bonds. Hence, these oxygen atoms cannot be protonated. The case of O2 will be discussed below.

Electrostatic Potential. The experimental attribution of hydrogen atom location allows us to calculate the precise electrostatic potential of PfluDING (Figure 5). The potential in the cavity turns out to be largely positive and complementary to the negatively charged phosphate species. Notably, the potential around the negatively charged residue Asp 62 is close to neutral.

Discussion

Hydrogen Bond Network in the Phosphate Binding Site. The subangstrom X-ray structures of PfluDING give precious information concerning the phosphate binding mode. The phosphate ion is interacting via numerous hydrogen bonds with protein side chains or main chain amide groups (Table 2). The phosphate oxygen atom O1, as well as O3, accepts three hydrogen bonds. O4 is involved in four hydrogen bonds. Therefore, O1, O3, and O4 are not protonated. The interaction between the phosphate oxygen atom O2 and $O_{\delta 2}$ from the carboxylate side chain of Asp 62 is noteworthy. As previously said, this H-bond is very short (2.50 Å). An electron density peak appears at a low sigma level in $F_{obs} - F_{calc}$ Fourier maps. This peak could correspond to a hydrogen atom bound to atom O2, which is compatible with the bond lengths and geometry observed locally. However, such a hydrogen atom cannot be placed with certainty due to the weakness of the density peak. Regarding the short O2...$O_{\delta 2}$ distance and ideal H-bond geometry, one could expect that a hydrogen atom bound to O2 would be well-defined, similarly to the other hydrogen atoms H-bonded to the phosphate ion (Table 2). Different phenomena can weaken the electron density of a hydrogen atom in omit-H difference maps. However, the thermal motion in the active site of PfluDING is very low (see Table 3). It is thus unlikely that thermal motion is the cause of the difficulty to see this hydrogen atom in the electronic density map. The hydrogen atom could, on the other hand, be disordered and display alternative sites along the O2...$O_{\delta 2}$ interaction.[33] The short distance O2...$O_{\delta 2}$ (2.50 Å) is typical for low barrier hydrogen bonds (LBHBs).[17,34,35] In this particular type of H-bond, the hydrogen atom should be dynamically distributed between the two implicated oxygen atoms. This makes it more difficult to observe the hydrogen atom via X-ray or even neutron diffraction.[36] In the case of PfluDING, the low electron density of this H-atom strongly suggests that its position is diffuse. In addition, the weak electron density peak suggesting that the phosphate O2 is protonated is compatible with the LBHB concept, as the 2.5 Å LBHBs are known to be asymmetric. The hydrogen atom is thus likely to be distributed between two low energy positions, the main being close to the phosphate O2 atom and the second closer to the Asp 62 $O_{\delta 2}$ atom.

The P−O bond lengths of the phosphate ion are consistent with this hypothesis, as seen in Table 3. According to the structural data deposited in the Cambridge Data Bank on the HPO_4^{3-} anion, longer P−O distances are observed for protonated oxygen atoms. In 31 structures of HPO_4^{2-} with a crystallographic agreement factor better than 5%, the average P−O distance for protonated oxygen atoms is 1.59(1) Å, whereas it is 1.52(1) Å for the nonprotonated oxygen atoms. (The average value is over X-ray structures deposited in the Cambridge Data Bank, containing a dibasic phosphate moiety

(33) Boukhris, A. *Elaboration et ètude cristallographique haute rèsolution de phosphate diacide d'ammonium et de potassium (ADP et KDP) et de leur solution mixte K1-x(NH4)xH2PO4*, Thesis, University of Marrakesh, Morocco, and Nancy Universite, France, 1995.
(34) Gilli, G.; Gilli, P. *J. Mol. Struct.* **2000**, *552*, 1–15.
(35) Cleland, W. W.; Frey, P. A.; Gerlt, J. A. *J. Biol. Chem.* **1998**, *273*, 25529–25532.
(36) Cleland, W. W. *Arch. Biochem. Biophys.* **2000**, *382*, 1–5.

Table 2. Summary of Hydrogen Bonds That Sequester the Phosphate Ion

oxygen	residue	H[a]	D[a]	sigma[b]	distance H...A[a] (Å)	distance D...A (Å)	angle D−H...A (deg)	B_{eq}[c] of D (Å^2)
O1	ARG 141	HH2	NH2	4.3	1.99	2.84	166.4	4.3
	THR 8	$H_{\gamma 1}$	$O_{\gamma 1}$	3.4	1.96	2.69	147.1	3.6
	THR 8	H0	N	5.1	1.94	2.77	161.4	3.4
O2	LEU 9	H0	N	5.5	2.08	2.93	168.0	3.1
	ASP 62	H	$O_{\delta 2}$	2.2	1.59[d]	2.50	168.5[d]	4.2
O3	SER 32	H0	N	4.5	2.02	2.77	145.4	3.2
	SER 32	Hγ	Oγ	3.2	1.89	2.66	156.4	3.7
	GLY 146	H0	N	4.5	1.92	2.75	161.5	3.2
O4	THR 147	$H_{\gamma 1}$	$O_{\gamma 1}$	3.2	1.87	2.67	168.2	3.5
	THR 147	H0	N	5.0	2.11	2.94	163.3	3.1
	SER 145	H_{γ}	O_{γ}	3.4	1.90	2.69	163.2	3.6
	ARG 141	HH1	NH1	4.5	2.22	2.87	132.5	4.3

[a] H, D, and A are the hydrogen, donor, and acceptor atoms in the hydrogen bond, respectively. [b] Sigma is the minimum contour level of electron density (in σ units) permitting the location of the hydrogen atom in the omit-H $F_{obs} - F_{calc}$ electron density map. [c] Equivalent isotropic B_{eq} factor of the donor atom. [d] The H...A distance and the angle are according to the electron density peak visible in the difference density map.

Downloaded by UNIV HENRI POINCARE on August 10, 2009
Published on May 15, 2009 on http://pubs.acs.org | doi: 10.1021/ja901900y

Figure 2. View of the phosphate binding site in the structure obtained at pH 4.5 and $F_{obs} - F_{calc}$ difference electron density map (blue). (a) Coordination of phosphate oxygen atom O3. (b) Coordination of phosphate oxygen atom O4. The contour levels are 2.5 σ and 2.6 σ in (a) and (b), respectively. In the model used for the difference map calculation, the hydrogen atoms which are H-bonded to the phosphate oxygen atoms are omitted.

Figure 3. Interaction between oxygen atom O2 of the phosphate ion and Oδ2 of Asp 62 in (a) the pH 4.5 structure of PfluDING and (b) the pH 8.5 structure of PfluDING. The positive difference Fourier map is contoured at 2 σ in (a) and 1.8 σ in (b). The electron density peak is located at a distance of 0.88 and 1.00 Å from O2 in the pH 4.5 and pH 8.5 structures, respectively.

noncovalently bonded to other molecules and having a crystallographic *R* factor better than 5%.) In the PfluDING case, the P−O2 distance is 1.56 Å, which is below 1.59 Å for a totally protonated oxygen atom. This also supports that O2 is partially protonated. Furthermore, the P−O1 and P−O3 distances are 1.52 and 1.53 Å, respectively, suggesting that the phosphate double bond is delocalized between these atoms.

This structural study shows that the bound phosphate ion partly possesses one hydrogen atom; thus, PfluDING binds dibasic phosphate even at pH 4.5. Moreover, since the hydrogen bond network is identical in both PfluDING structures at pH 4.5 and 8.5, this demonstrates that PfluDING binds dibasic phosphate (HPO_4^{2-}) at both pH values. Different binding modes for monobasic and dibasic phosphate were hypothesized for phosphate-SBPs. For dibasic phosphate, the hydrogen bond network in phosphate-SBPs is identical to that found in PfluDING.[15] For monobasic phosphate ($H_2PO_4^-$) binding, it has

Table 3. Thermal Displacement Parameters B_{eq} (Å^2) and Bond Lengths (Å) of the Phosphate Ion and the Asp 62 Side Chain[a]

		Thermal Motion	
	atom	$crystal_{pH4.5}$	$crystal_{pH8.5}$
Phosphate	P	3.16(8)	3.3(1)
	O1	3.4(3)	3.9(3)
	O2	2.8(2)	3.5(3)
	O3	3.1(2)	2.8(3)
	O4	3.2(3)	3.2(3)
Asp 62	$O_{\delta 2}$	4.2(3)	3.9(3)
	$O_{\delta 1}$	4.3(3)	3.8(3)
	$C\gamma$	3.3(5)	3.1(5)
		Bond Lengths	
Phosphate P−O distances	O1	1.522(6)	1.529(6)
	O2	1.561(7)	1.559(7)
	O3	1.533(7)	1.528(7)
	O4	1.555(8)	1.541(8)
Asp 62 C_γ−O distances	$O_{\delta 2}$	1.26(1)	1.26(1)
	$O_{\delta 1}$	1.26(1)	1.22(1)
$O_{\delta 2}$−O2 distance		2.502(6)	2.495(7)

[a] The values and their errors (between parentheses) are taken from the *MoPro* last least-squares refinement. No restraints or constraints have been applied on the listed atoms during the refinement. The errors are underestimated here as the refinement was performed only on local atoms.

been hypothesized that a donor hydroxyl group (which corresponds to Ser 32 in PfluDING) flips and becomes the acceptor for the second phosphate hydrogen atom.[15] However, in the PfluDING structure at pH 4.5, the hydrogen atom on γ-OH of Ser 32 is still H-bonded to the phosphate ion (Figure 2). Indeed, PfluDING at pH 4.5 binds dibasic phosphate, and no flip of Ser32 γ-OH is required.

Electrostatic Potential of the Phosphate Binding Cleft. In the phosphate binding site, the electrostatic potential turns out to be largely positive (Figure 4). The electrostatic potential in the vicinity of the carboxylate side chain of Asp 62 is almost neutral. There are local dipoles (O−H and N−H groups) and two charged residues in the binding cleft: an aspartic acid (Asp 62) and an arginine (Arg 141). The positive potential is partly due to the positive charges of the hydrogen atoms lining the phosphate binding cavity. The arginine side chain and four nearby N-termini of α-helices increase this effect. The positive potential in the phosphate binding site region is in accordance with the ability of PfluDING to bind negatively charged ions. Former calculations of the electrostatic potential of phosphate binding protein (PBP) led to a negative potential in the active site.[37] The program used for the calculation was GRASP,[38] which calculates the electrostatic potential using the Poisson−Boltzmann equation. In the case of PBPs, default values for charges (full charges) and salt concentration (0.0) were applied.[37] This means that formal charges were used for lysines, arginines, glutamic acids, aspartic acids, and the termini. Contrary to the *VMoPro* calculation which takes into account all atom charges and higher multipoles for example, the positive partial charges of hydrogen atoms have not been taken into account explicitly. As already shown, there are numerous local dipoles of N−H amide and O−H hydroxyl groups in the active site. In particular, three α-helices directly interact with the phosphate through amide groups, and one α-helix is in close proximity. The positive potential of the cleft is mainly due to these polarized groups and to the positively charged Arg

Downloaded by UNIV HENRI POINCARE on August 10, 2009
Published on May 15, 2009 on http://pubs.acs.org | doi: 10.1021/ja901900y

Figure 4. Schematic diagram of the experimentally determined hydrogen bond network involving the phosphate molecule sequestered in the PfluDING binding site. Distances between the acceptor atoms (phosphate oxygen atoms) and the hydrogen atoms are indicated.

Downloaded by UNIV HENRI POINCARE on August 10, 2009
Published on May 15, 2009 on http://pubs.acs.org | doi: 10.1021/ja901900y

Figure 5. (a) Deformation electrostatic potential without the phosphate moiety shown as color-coded map displayed on the surface in the phosphate binding cleft (0.7 Å probe radius). Negative, zero, and positive potential are displayed in red, white, and blue, respectively. The scale goes from −0.2 |e|/Å to + 0.2 |e|/Å. (b) Electrostatic potential in the plane through H0 (Leu 9) − O2 − $O_{\delta 2}$. The scale goes from −1.0 |e|/Å (red) to +1.0 |e|/Å (blue); the increment is 0.05 |e|/Å. Zero potential is displayed in white.

141. Only one negatively charged residue is in the binding cleft (Asp 62). It is likely that a general negative shift occurred in the final electrostatic potential, when only this negative charge was taken into account, but not the local dipoles. The electrostatic potential of PBP was recalculated with *VMoPro*, which uses experimental charge densities for all protein atoms,[27] and led, similarly to the case of PfluDING, to a positive potential in the binding cleft (see supplementary data). Thus, the approximation of formal charges for only the charged side chains seems not sophisticated enough for electrostatic potential calculations in this type of binding cavity.

The high resolution data allow us to calculate the electrostatic deformation potential for PfluDING. For the first time, a protein structure of atomic resolution has been interpreted using transferred charges and multipoles from the ELMAM data bank. As the resolution usually obtained in protein diffraction experiments is not high enough to refine experimentally the charge density parameters, the use of an appropriate electron density description offers the possibility to calculate easily electrostatic properties which would be otherwise inaccessible.

Molecular Mechanism of the Phosphate Binding. According to the subangstrom crystal structures, PfluDING sequesters dibasic phosphate even at acidic pH 4.5, although monobasic phosphate is predominant in solution. In relation to the positive electrostatic potential of the binding cleft, this strongly suggests that the PfluDING binding site is able to attract and stabilize selectively dibasic phosphate. The nonprotonation of the phosphate oxygen atoms O1, O3, and O4 is favored in the binding cleft due to the presence of numerous donating hydrogen bonds. In addition, the formation of an LBHB between Asp62 and the phosphate O2 atom suggests that the pK_a values of the two hydrogen bond partners are in the same range.[35] Since the pK_a of $H_2PO_4^-$/HPO_4^{2-} in solution is higher by more than 3 units than the pK_a of the aspartic acid side chain, this confirms the strong influence of the positive electrostatic potential of the binding cleft decreases the pK_a of the bound phosphate ion. On the other hand, as the Asp62 carboxylate group is located in the protein binding cleft, its pK_a value is expected to be increased due to the desolvation effect.

Our results are in agreement with energy determinations for PBPs.[17] To determine the energy contribution of the short hydrogen bond $O_{\delta 2}$...O2, a mutant D56 N was designed.[17] In this mutant, the distance between O2 and $O_{\delta 2}$ of N56 has increased to 2.6 Å (structure resolution 1.9 Å, PDB code 1ixi) compared to the wild type distance of 2.5 Å (structure resolution 0.98 Å, 1ixh) and has shifted from an LBHB to a normal H-bond. The phosphate binding free energy was slightly changed; $\Delta\Delta G$ between wild type and the mutant is +0.3 kcal/mol, albeit energies of 12−24 kcal/mol are expected for an LBHB.[39] The binding energy of the phosphate ion to the protein is constituted by hydrogen bonds and electrostatic attraction. In the mutant, the energy contribution due to hydrogen bonds decreases, because the LBHB no longer exists. On the other hand, the electrostatic contribution has been largely modified. Indeed, the negative charge (Asp) in the binding site has been replaced by a neutral residue (Asn). Thus, the positive global electrostatic potential of the binding cleft is likely to increase, and the interaction with the negatively charged phosphate ion is reinforced. The shift in the electrostatic charge of the binding cleft is likely to compensate for the loss of the LBHB and explains why the free energy was similar for the wild type and the mutant proteins.

Conclusion

The subangstrom resolution structure of PfluDING at two different pH values (4.5 and 8.5) allowed us to identify a large part of hydrogen atom positions in the PfluDING structure. More precisely, we experimentally determined the molecular mechanism of phosphate binding in a phosphate-SBP. Particularly, we found that, at either pH value (4.5 or 8.5), PfluDING binds dibasic phosphate using the same mechanism. The binding cleft of this highly accurate model, contrary to previously published results,[37] was found to have a positive electrostatic potential, as determined for the first time from the ELMAM database. The positive electrostatic potential and numerous H-bond donors in the binding cleft

(37) Ledvina, P. S.; Yao, N.; Choudhary, A.; Quiocho, F. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6786–6791.
(38) Nicholls, A.; Sharp, K. A.; Honig, B. *Proteins* **1991**, *11*, 281–296.

could decrease the proton affinity of the phosphate ion and explain why PfluDING binds dibasic phosphate even at an acidic pH. The phosphate molecule accepts 11 hydrogen atoms and donates its only hydrogen atom to Asp 62, forming an LBHB. This hydrogen bond network is probably the key that explains the very high affinity of phosphate-SBPs for their ligand. The LBHB is likely to be responsible for the dibasic phosphate binding specificity, allowing it to discriminate against closely related chemical species like sulfate.

(39) Garcia-Viloca, M.; Gonzalez-Lafont, A.; Lluch, J. M. *J. Am. Chem. Soc.* **1997**, *119*, 1081–1086.

Acknowledgment. The authors thank Prof. Enrique Espinosa for fruitful discussions. This work was partly funded by the Agence Nationale de la Recherche, programme blanc (Grant LibrarEnergy). D.L. and M.E. are doctoral fellows supported by the French Ministry of Research.

Supporting Information Available: Complete refs 7 and 8. This material is available free of charge via the Internet at http://pubs.acs.org.

JA901900Y

Downloaded by UNIV HENRI POINCARE on August 10, 2009
Published on May 15, 2009 on http://pubs.acs.org | doi: 10.1021/ja901900y

Conclusion générale

Ce travail portait sur les propriétés électrostatiques et structurales des protéines résolues par diffraction X à haute résolution. Deux aspects principaux ont été abordés : d'une part, l'analyse d'un point de vue fondamental des propriétés dérivées de la distribution de charge des structures des protéines, d'autre part, l'application pratique des méthodes de la cristallographie haute résolution des macromolécules à deux enzymes présentant un intérêt majeur.

L'analyse structurale et électrostatique de la protéine PfluDING, un transporteur de phosphate, a permis de mettre en évidence le mode de fixation du phosphate à deux pH différents. En particulier, la présence d'une liaison hydrogène diffuse entre l'ion phosphate et la protéine a été démontrée. Cette interaction assure que l'état de protonation du phosphate est identique à pH 4.5 et 8.5. Le calcul du potentiel électrostatique dans le site de fixation, se basant sur les paramètres multipolaires transférés de la bibliothèque ELMAM, a montré l'influence importante des atomes d'hydrogènes sur celui-ci. Ce résultat souligne le gain d'information obtenu grâce à la diffraction X haute résolution et qui consiste à localiser les atomes H d'après la densité électronique.

La seconde enzyme étudiée était la protéine DFPase, capable de dégrader les organophosphorés (agents neurotoxiques de guerre). Nous avons comparé la structure de la DFPase obtenue par diffraction X haute résolution et la structure affinée contre des données neutroniques de moyenne résolution. Le travail pointe les différences d'interprétation du réseau de liaisons hydrogène des molécules d'eau dans les deux modèles, et permet de proposer un nouveau mécanisme d'action de cette protéine. Ce travail a démontré que les données de diffraction X à haute résolution peuvent défier des données de diffraction des neutrons, obtenues à moyenne résolution.

Les deux aspects plus fondamentaux portaient sur les éléments de structure secondaire des protéines : les propriétés électrostatiques et le réseau d'interactions, notamment des liaisons H, assurant la cohésion des motifs de structure secondaire. En particulier, l'analyse des moments électriques a mis en évi-

dence l'existence de moments dipolaires et quadrupolaires des feuillets β. En tant qu'application, la caractérisation des moments électriques devrait permettre de décrire une protéine comme un ensemble de macro-moments afin de calculer leur énergie d'interaction.
L'analyse topologique des liaisons hydrogène dans les motifs de structure secondaire a permis de démontrer que certaines interactions entre groupements peptidiques au sein d'hélices, qui sont classiquement représentées comme des liaisons hydrogène, devaient en fait être considérés comme des contacts purement électrostatiques. Ce résultat pourra aider à mieux estimer l'énergie potentielle des hélices et ainsi de toute la protéine.
En résumé, ce travail souligne l'importance de la diffraction X à haute résolution pour le développement de nouvelles méthodes et pour les applications dans le domaine de la cristallographie des protéines. En particulier, il a tracé la voie pour des perspectives intéressantes :

- La réduction des protéines à un ensemble de moments électriques représente un moyen de caractériser l'énergie d'interaction protéine-substrat ou protéine-protéine. Cela permettra de mieux comprendre les mécanismes de fonctionnement des enzymes et pourra trouver son application dans le design de ligands.

- Grâce aux résultats sur la topologie des interactions dans les hélices, une fonction potentielle plus précise, décrivant l'énergie des liaisons H, pourra être développée. Les fonctions d'énergie potentielle sont utilisées afin de calculer des propriétés comme l'affinité ou afin de prédire des sites de fixation, le repliement tridimensionnel protéique ou des interfaces d'interaction protéine-protéine.

- L'étude du mode de fixation de l'ion phosphate à la protéine PfluDING a montré la spécificité d'interaction entre une protéine et un ligand via des liaisons hydrogène. Récemment, il a été montré par simulations Monte-Carlo que l'interaction entre une protéine et l'ADN est également régie par une forte spécificité, le contact étant réalisé seulement dans le cas où la complémentarité des charges et des donneurs/accepteurs de liaisons hydrogène est garantie [130]. L'application des méthodes utilisées dans ce travail (calcul du potentiel électrostatique et des moments électriques, topologie des liaisons hydrogènes) pourra fournir des informations complémentaires sur cette interaction.

- Il est intéressant d'étendre l'étude de la protéine PfluDING. En effet, puisqu'elle est capable de former des cristaux de grande taille, une ex-

périence de diffraction des neutrons pourra être envisagée. L'analyse de la structure neutrons de PfluDING devrait aider à caractériser encore mieux le mécanisme de fixation et une comparaison avec la structure X haute résolution pourra apporter des éléments intéressants à la méthodologie en cristallographie nucléaire macromoléculaire. En particulier, l'analyse de l'agitation thermique dans le cas RX/neutrons nous semble intéressant. En conjonction avec des données structurales de petites mailles (tirées par exemple de la Cambridge data base), nous pourrons nous focaliser sur l'agitation thermique des atomes d'hydrogène et des atomes de deutérium, qui, théoriquement, devrait être différente pour les deux espèces atomiques.

- L'étude de la structure de la protéine DFPase, obtenue par diffraction des neutrons, a montré qu'il y a des différences marquées entre le réseau de liaisons hydrogène observé dans cette structure et celle résolue par diffraction X haute résolution. Les données neutrons étant vraisemblablement issues d'un cristal maclé, il nous semble nécessaire qu'une nouvelle expérience de diffraction des neutrons soit effectuée.

- Afin de vérifier expérimentalement la présence de l'interaction N...O dans les hélices, il serait intéressant d'effectuer des études sur des polypeptides hélicoïdaux, capables de cristalliser et de diffracter à résolution suffisamment haute pour pouvoir entreprendre une analyse topologique de la densité électronique sur la base de paramètres affinés contre les données X, plutôt que, comme ça a été le cas dans la présente étude, en utilisant un modèle de densité de charge transférée.

Appendices

Annexe A

Notions utilisées en biologie macromoléculaire

A.1 Abréviations utilisées dans le rapport

AA Acide aminé
ADP *anisotropic displacement parameter*
CP *critical point* (point critique)
DW Debye-Waller
liaison H Liaison hydrogène
MM Mécanique moléculaire
OP organophosphoré
PDB Protein Data Bank
PEG Polyéthylene glycol
RMN Résonance Magnétique Nucléaire
rms *root mean square*
rmsd *root mean square deviation*
RX rayons X
TF Transformée de Fourier
UP Unité peptidique

A.2 La Proteine Data Bank (PDB)

La "proteine data bank" [66] est une banque de données de structures tridimensionnelles de macromolécules biologiques (protéines et acides nucléiques). Les données structurales viennent en majorité des expérience de

diffraction des rayons X et de RMN. Les modèles peuvent êtres déposés après avoir satisfaits des tests de validité (p.ex. stéréochimie, indices statistiques sur l'affinement). Les structures sont disponibles gratuitement et peuvent être téléchargées sur internet (http ://www.pdb.org/pdb/home/home.do). Chaque structure a un identifiant, se composant de 4 lettres ou chiffres.

A.3 Représentation de la densité électronique

Au cours de l'analyse cristallographique, il y a de nombreuses occasions où une visualisation (ou représentation) de la densité électronique ou nucléaire peut s'avérer utile. Par exemple, entre les étapes d'affinement, on examine la densité électronique pour voir si l'affinement a amélioré le modèle. Dans les publications scientifiques, il y a des figures d'un modèle protéique auquel la densité électronique est superposée. Ainsi, la pertinence du modèle peut être montrée en affirmant que les données expérimentales soutiennent le modèle présenté. Ce sont souvent des ligands, des atomes d'hydrogène ou des résidus catalytiques qui sont représentés de cette manière.
La représentation de la densité est souvent appelée "carte", suivi d'une précision, à partir de quel facteur de structure elle est calculée. En général, on obtient la densité à partir de la transformée de Fourier d'un facteur de structure $\vec{F}$ ayant le module F, qui peut aussi être une somme de plusieurs termes, et une phase ϕ. Le module du facteur de structure peut être calculé (noté F_{calc} dans ce cas) ou observé lors de l'expérience de diffraction (F_{obs}), les phases sont en général issues du modèle, donc ce sont des phases calculées ϕ_{calc}. Ainsi, on distingue les représentations suivantes :

carte calculée (F_{calc}) Elle montre la densité électronique totale calculée uniquement à partir du modèle.

carte Fourier différence ($F_{obs} - F_{calc}$) ou résiduelle Elle est calculée à partir de la différence des facteurs de structure observés et calculés (avec la phase du modèle). Cette carte représente alors la différence entre le modèle et la molécule cristallisé qui donnait lieu à la diffraction. Il peut y avoir de la densité résiduelle positive et négative. La densité positive représente ce qui n'a pas été pris en compte dans le modèle. C'est-à-dire, un atome qui manque dans le modèle va se présenter comme pic de densité positif. Par contre, une densité négative représente ce qui est de trop dans le modèle. Plus précisément, un atome mal placé [1] va engendrer un pic de densité négatif dans une carte Fourier différence.

1. Un atome mal placé est un atome qui est normalement à un autre endroit ou qui n'est pas présent dans la structure.

Ainsi, une telle carte est souvent utilisé pour vérifier les progrès d'un affinement puisqu'on peut visualiser l'impact d'une modification du modèle.

carte Fourier différence ($2F_{obs} - F_{calc}$) Cette carte de densité est utilisée pour construire le modèle, c'est-à-dire les pics de densité non modélisés sont interprétés s'ils correspondent à des atomes, des ligands ou des groupements. La validité des modifications sera vérifiée après l'affinement, avec les cartes Fourier différence $F_{obs} - F_{calc}$.

carte omit ($F_{obs} - F_{calc}$) Une carte omit est un cas particulier de la carte Fourier différence, calculé à partir d'un modèle dans lequel des atomes ou des molécules ont été supprimés (par exemple en posant leur occupation égale à zéro). La différence par rapport à la carte Fourier différence est que ces atomes étaient auparavant présent dans le modèle. La démarche est alors la suivante : On construit le modèle moléculaire jusqu'à l'obtention d'un modèle satisfaisant. Si l'on veut montrer la densité électronique qui correspond à une partie du modèle, p.ex. un ligand, on l'enlève du modèle et on calcule une carte omit. Si le ligand est présent dans la structure, sa densité sera visible comme pic positif dans la carte omit. Ainsi, une carte de densité "omit H" est une carte de densité résiduelle, dans laquelle les atomes d'hydrogène ont été omis.

Les cartes de densité électronique ou nucléaire peuvent être représentées en deux ou trois dimensions [2]. Les plans sont utilisés pour visualiser la densité à un endroit précis (par exemple sur un plan, comprenant le chemin d'interaction entre deux atomes) et ils permettent de montrer des détails (par exemple densité de déformation) et de les quantifier. La densité est en effet visualisée en lignes de niveaux de valeur constante et adaptée [3] comme $0.05e^{-}/\text{Å}^{-3}$.
Les cartes tridimensionnelles sont utilisées pour la construction d'un modèle moléculaire. Des programmes de visualisation (par exemple Pymol [46]) permettent d'exporter ces vues dans une image 2d (en utilisant des ombres qui créent une vue 3d apparente). La densité est visualisée comme isosurface. Le contour est choisi comme un multiple du bruit expérimental (souvent noté σ). En fonction de la hauteur d'un pic de densité en termes de σ, il est considéré comme significatif. Par exemple, dans une carte Fourier différence, des pics à partir de 2.5 à 3 σ sont considérés comme significatifs, puisque égaux à environ 3 fois la déviation standard de la densité.

2. Il est également possible de représenter la densité en une dimension, le long d'une liaison covalente par exemple.

3. C'est le même principe que celui utilisé pour représenter les lignes de nivellation sur les cartes d'état-Major.

A.4 Les 20 acides aminés

Il y a deux systèmes d'abréviation pour les acides aminés : le code à trois lettres tandis le code à une lettre(tableau A.1).

Acide aminé	code 3 lettres	code 1 lettre
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Acide aspartique (aspartate)	Asp	D
Cystéine	Cys	C
Acide glutamique (glutamate)	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Méthionine	Met	M
Phénylalanine	Phé	F
Proline	Pro	P
Sérine	Ser	S
Thréonine	Thr	T
Tryptophane	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table A.1 – Abréviations des acides aminés.

Annexe B

Divers

B.1 Détails de calculs

B.1.1 Dépendance de l'origine du moment quadrupolaire sans trace

$$\begin{aligned}
\Theta'_{\alpha\beta} &= \frac{1}{2}\int \rho(\vec{r})[3r'_\alpha r'_\beta - r'^2\delta_{\alpha\beta}]d\vec{r} \\
&= \frac{1}{2}\int \rho(\vec{r})[3(r_\alpha - R_\alpha)(r_\beta - R_\beta) - (\vec{r}-\vec{R})^2\delta_{\alpha\beta}]d\vec{r} \\
&= \frac{1}{2}\int \rho(\vec{r})[3r_\alpha r_\beta - 3(r_\alpha R_\beta + r_\beta R_\alpha) + 3R_\alpha R_\beta - \delta_{\alpha\beta}(r^2 + R^2 - 2\vec{r}\vec{R})]d\vec{r}
\end{aligned}$$

On peut réarranger la somme pour simplifier

$$\begin{aligned}
\Theta'_{\alpha\beta} &= \underbrace{\frac{1}{2}\int \rho(\vec{r})[3r_\alpha r_\beta - r^2\delta_{\alpha\beta}]d\vec{r}}_{\Theta_{\alpha\beta}} - \tfrac{3}{2}R_\beta \underbrace{\int \rho(\vec{r})r_\alpha d\vec{r}}_{\mu_\alpha} - \tfrac{3}{2}R_\alpha \underbrace{\int \rho(\vec{r})r_\beta d\vec{r}}_{\mu_\beta} + \\
&\quad + \tfrac{3}{2}R_\alpha R_\beta \underbrace{\int \rho(\vec{r})d\vec{r}}_{q} - \tfrac{1}{2}\delta_{\alpha\beta}R^2 \underbrace{\int \rho(\vec{r})d\vec{r}}_{q} + \delta_{\alpha\beta}\underbrace{\int \rho(\vec{r})\sum_{i=1}^{3} r_i R_i}_{\sum_{i=1}^{3} R_i\mu_i} \\
&= \Theta_{\alpha\beta} + \tfrac{1}{2}(3R_\alpha R_\beta - R^2\delta_{\alpha\beta})q - \tfrac{3}{2}(R_\beta\mu_\alpha + R_\alpha\mu_\beta) + \textstyle\sum_{i=1}^{3} R_i\mu_i\delta_{\alpha\beta}
\end{aligned}$$

B.1.2 Intégration du moment quadrupolaire

$$\Theta_{lmp} = -\int_{V_T} r^l c_{lmp}[\sum_{l'=0}^{l'_{max}} \kappa'^3 R_{l'} \sum_{m'=0}^{l'} P_{l'm'p} d_{l'm'p}]d\vec{r} \tag{B.1}$$

Les harmoniques sphériques étant orthogonales, tous les produits $c_{lmp} \cdot d_{l'm'p}$ s'annulent, sauf dans le cas $l = l'$ et $m = m'$. Il en suit pour le moment sphérique, en utilisant que $R_l = \frac{\kappa'^{n(l)}\zeta^{n(l)+3}}{(n(l)+2)!} r^{n(l)} e^{-\kappa'\zeta r}$, $c_{lmp} = \frac{1}{M_{lm}} y_{lm}$ et $d_{lmp} = \frac{L_{lm}}{M_{lm}} y_{lmp}$. :

$$\begin{aligned}
\Theta_{lmp} &= -\int_{V_T} r^l c_{lmp} \kappa'^3 R_l P_{lmp} d_{lmp} d\vec{r} \\
&= -P_{lmp}\kappa'^3 \int_{V_T} r^l y_{lm}^2 \frac{L_{lm}}{M_{lm}^2} \frac{\kappa'^{n(l)}\zeta^{n(l)+3}}{(n(l)+2)!} r^{n(l)} e^{-\kappa'\zeta r} d\vec{r} \\
&= -P_{lmp} \frac{L_{lm}}{M_{lm}^2} \frac{(\kappa'\zeta)^{n(l)+3}}{(n(l)+2)!} \int_{V_T} r^{n(l)+l} e^{-\kappa'\zeta r} y_{lmp}^2 r^2 dr sin\theta d\theta d\phi \\
&= -P_{lmp} \frac{L_{lm}}{M_{lm}^2} \frac{(\kappa'\zeta)^{n(l)+3}}{(n(l)+2)!} \int_0^\infty r^{n(l)+l+2} e^{-\kappa'\zeta r} dr \int y_{lmp}^2 sin\theta d\theta d\phi
\end{aligned} \tag{B.2}$$

Les fonctions y_{lmp} étant normalisées, le dernier intégrale est égale à 1. L'intégrale sur r peut être calculé en utilisant $\int_0^\infty x^n e^{-ax} dx = n! a^{-n-1}$ et en utilisant que la densité électronique tend vers zéro quand on s'éloigne vers l'infini par rapport à l'atome [1], on obtient alors :

$$\Theta_{lmp} = -P_{lmp} \frac{L_{lm}}{M_{lm}^2} \frac{(n(l)+l+2)!}{(n(l)+2)!} \frac{1}{(\kappa'\zeta)^l} \tag{B.3}$$

1. C'est une approximation qui permet de contourner le choix du volume d'intégration V_T.

Bibliographie

[1] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis," *Nature*, vol. 181, pp. 662–666, Mar. 1958.

[2] C. Jelsch, M. M. Teeter, V. Lamzin, V. Pichon-Pesme, R. H. Blessing, and C. Lecomte, "Accurate protein crystallography at ultra-high resolution : Valence electron distribution in crambin," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 7, pp. 3171–3176, 2000.

[3] B. Guillot, C. Jelsch, A. Podjarny, and C. Lecomte, "Charge-density analysis of a protein structure at subatomic resolution : the human aldose reductase case," *Acta Crystallographica Section D*, vol. 64, pp. 567–588, 2008.

[4] C. Branden and J. Tooze, *Introduction à la structure des protéines*. De Boeck, 1996.

[5] A. V. Fain, D. L. Ukrainskii, S. A. Dobkin, A. V. Galkin, and N. G. Esipova, "Abundance and arrangement of single, double and bifurcated hydrogen bonds in protein alpha-helices : Statistical analysis of pdbselect," *Biophysics*, vol. 53, pp. 125–133, 2008.

[6] W. Kabsch and C. Sander, "Dictionary of protein secondary structure : Pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, 1983.

[7] E. Howard, R. Sanishvili, R. Cachau, A. Mitschler, B. Chevrier, P. Barth, V. Lamour, M. V. Zandt, E. Sibley, C. Bon, D. Moras, T. Schneider, A. Joachimiak, and A. Podjarny, "Ultrahigh resolution drug design i : Details of interactions in human aldose reductase-inhibitor complex at 0.66 angstrom," *Proteins : Structure, Function, and Bioinformatics*, vol. 55, no. 4, pp. 792–804, 2004.

[8] P. Enkhbayar, K. Hikichi, M. Osaki, R. H. Kretsinger, and N. Matsushima, "3/10 helices in proteins are parahelices," *Proteins : Structure, Function, and Bioinformatics*, vol. 64, no. 3, pp. 691–699, 2006.

[9] G. Taylor, "The phase problem," *Acta Crystallographica Section D*, vol. 59, pp. 1881–1890, Nov 2003.

[10] J. Cavarelli, "Cristallographie des macromolécules biologiques," *Techniques de l'ingénieur*, vol. P2, no. P1090, 2000.

[11] K. N. Trueblood, H.-B. Bürgi, H. Burzlaff, J. D. Dunitz, C. M. Gramaccioli, H. H. Schulz, U. Shmueli, and S. C. Abrahams, "Atomic Dispacement Parameter Nomenclature. Report of a Subcommittee on Atomic Displacement Parameter Nomenclature," *Acta Crystallographica Section A*, vol. 52, pp. 770–781, Sep 1996.

[12] P. V. Afonine, V. Y. Lunin, N. Muzet, and A. Urzhumtsev, "On the possibility of the observation of valence electron density for individual bonds in proteins in conventional difference maps," *Acta Crystallographica Section D*, vol. 60, pp. 260–274, Feb 2004.

[13] T.-P. Ko, H. Robinson, Y.-G. Gao, C.-H. C. Cheng, A. L. DeVries, and A. H.-J. Wang, "The Refined Crystal Structure of an Eel Pout Type III Antifreeze Protein RD1 at 0.62-Å Resolution Reveals Structural Microheterogeneity of Protein and Solvation," *Biophysical Journal*, vol. 84, no. 2, pp. 1228 – 1237, 2003.

[14] N. K. Hansen and P. Coppens, "Testing aspherical atom refinements on small-molecule data sets," *Acta Crystallographica Section A*, vol. 34, pp. 909–921, 1978.

[15] P. Coppens, *X-Ray Charge Densities and Chemical Bonding (International Union of Crystallography Texts on Crystallography , No 4)*. Oxford University Press, 1997.

[16] D. E. Tronrud, "Introduction to macromolecular refinement," *Acta Crystallographica Section D*, vol. 60, pp. 2156–2168, 2004.

[17] A. T. Brunger, "Free r value : a novel statistical quantity for assessing the accuracy of crystal structures," *Nature*, vol. 355, pp. 472–475, Jan. 1992.

[18] C. Jelsch, B. Guillot, A. Lagoutte, and C. Lecomte, "Advances in protein and small-molecule charge density refinement methods using mopro," *Journal of Applied Crystallography*, vol. 38, pp. 38–54, 2005.

[19] C. P. Brock, J. D. Dunitz, and F. L. Hirshfeld, "Transferability of deformation densities among related molecules : atomic multipole parameters from perylene for improved estimation of molecular vibrations in naphthalene and anthracene," *Acta Crystallographica Section B*, vol. 47, pp. 789–797, Oct 1991.

[20] R. W. F. Bader, *Atoms in Molecules. A Quantum Theory.* Oxford University Press, 1990.

[21] M. Souhassou and R. H. Blessing, "Topological analysis of experimental electron densities," *Journal of Applied Crystallography*, vol. 32, pp. 210–217, Apr 1999.

[22] R. F. Stewart, "Mapping electrostatic potentials from diffraction data," *God. Jugosl. Cent. Kristalogr*, vol. 17, pp. 1–24, 1982.

[23] N. Ghermani, N. Bouhmaida, and C. Lecomte, "Modelling electrostatic potential from experimentally determined charge densities. i. spherical-atom approximation," *Acta Crystallographica Section A*, vol. 49, pp. 781–789, 1993.

[24] N. Bouhmaida, *Modélisation du potentiel électrostatique calculé à partir de la densité électronique de déformation X-X dans les cristaux moléculaires. Application aux composés peptidiques.* PhD thesis, G.F.D. Sciences et Génie des Matériaux, 1993.

[25] W. Hol, P. van Duijnen, and H. Berendsen, "The alpha-helix dipole and the properties of proteins," *Nature*, vol. 273, pp. 443–446, 1978.

[26] W. Hol, L. Halie, and C. Sander, "Dipoles of the alpha-helix and beta-sheet : their role in protein folding," *Nature*, vol. 294, pp. 532–536, 1981.

[27] M. A. Spackman, "Molecular electric moments from x-ray diffraction data," *Chemical Reviews*, vol. 92, pp. 1769–1797, Dec. 1992.

[28] D. Sengupta, R. N. Behera, J. C. Smith, and G. M. Ullmann, "The alpha helix dipole : Screened out ?," *Structure*, vol. 13, pp. 849–855, June 2005.

[29] P. C. Kahn, "Defining the axis of a helix," *Computers Chem.*, vol. 13, no. 3, pp. 185–189, 1989.

[30] P. C. Kahn, "Simple methods for computing the least squares line in three dimensions," *Computers Chem.*, vol. 13, no. 3, pp. 191–195, 1989.

[31] J. A. Christopher, R. Swanson, and T. O. Baldwin, "Algorithms for finding the axis of a helix : Fast rotational and parametric least-squares methods," *Computers & Chemistry*, vol. 20, no. 3, pp. 339 – 345, 1996.

[32] A. Wada, "The alpha-helix as an electric macro-dipole," *Adv. Biophys.*, vol. 9, pp. 1–63, 1976.

[33] R. J. Kurland and J. E. Bright Wilson, "Microwave spectrum, structure, dipole moment, and quadrupole coupling constants of formamide," *The Journal of Chemical Physics*, vol. 27, no. 2, pp. 585–590, 1957.

[34] F. A. Momany, R. F. McGuire, J. F. Yan, and H. A. Scheraga, "Energy parameters in polypeptides. iii. semiempirical molecular orbital calculations for hydrogen-bonded model peptides," *The Journal of Physical Chemistry*, vol. 74, pp. 2424–2438, June 1970.

[35] D. Poland and H. A. Scheraga, "Energy parameters in polypeptides. i. charge distributions and the hydrogen bond," *Biochemistry*, vol. 6, pp. 3791–3800, Dec. 1967.

[36] R. P. Sheridan and L. C. Allen, "The electrostatic potential of the alpha helix," *Biophysical Chemistry*, vol. 11, no. 2, pp. 133 – 136, 1980.

[37] D. Lockhart and P. Kim, "Internal stark effect measurement of the electric field at the amino terminus of an alpha helix," *Science*, vol. 257, no. 5072, pp. 947–951, 1992.

[38] C. Park and W. A. Goddard, "Stabilization of alpha-helices by dipole-dipole interactions within alpha-helices," *The Journal of Physical Chemistry B*, vol. 104, pp. 7784–7789, Aug. 2000.

[39] G. J. Kleywegt, "Validation of protein models from c[alpha] coordinates alone," *Journal of Molecular Biology*, vol. 273, no. 2, pp. 371 – 376, 1997. Moleman(2).

[40] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, I. Arendall, W. Bryan, J. Snoeyink, J. S. Richardson, and D. C. Richardson, "Molprobity : all-atom contacts and structure validation for proteins and nucleic acids," *Nucl. Acids Res.*, vol. 35, no. suppl 2, pp. W375–383, 2007.

[41] M. Heinig and D. Frishman, "Stride : a web server for secondary structure assignment from known atomic coordinates of proteins," *Nucl. Acids Res.*, vol. 32, pp. W500–502, 2004.

[42] P. Chakrabarti and S. Chakrabarti, "C–H...O hydrogen bond involving proline residues in alpha-helices," *Journal of Molecular Biology*, vol. 284, no. 4, pp. 867 – 873, 1998.

[43] J. Richardson and D. Richardson, "Amino acid preferences for specific locations at the ends of alpha helices," *Science*, vol. 240, no. 4859, pp. 1648–1652, 1988.

[44] J. C. P. Poole, *The Physics Handbook : Fundamentals and Key Equations*. Wiley-VCH, 2007.

[45] P. T. V. Duijnen, J. C. D. Jager, and B. T. Thole, "Do parallel beta-strands have dipole moments? an ab initio molecular-orbital-direct reaction field study," *Biopolymers*, vol. 24, no. 5, pp. 735–745, 1985.

[46] W. L. DeLano, "The PyMOL Molecular Graphics System.," *DeLano Scientific LLC, San Carlos, CA, USA. http ://www.pymol.org*, 2008.

[47] "http ://www.chemcomp.com/, c.c. group, molecular operating environment."

[48] T. A. Halgren, "Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94," *J. Comput. Chem.*, vol. 17, no. 5-6, pp. 490–519, 1996.

[49] A. Thomas, N. Benhabiles, R. Meurisse, R. Ngwabije, and R. Brasseur, "Pex, analytical tools for pdb files. ii. h-pex : Noncanonical h-bonds in alpha-helices," *Proteins : Structure, Function, and Genetics*, vol. 43, no. 1, pp. 37–44, 2001.

[50] R. A. Engh and R. Huber, "Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement," *Acta Crystallographica Section A*, vol. 47, pp. 392–400, 1991.

[51] M. Crisma, M. Saviano, A. Moretto, Q. B. Broxterman, B. Kaptein, and C. Toniolo, "Peptide alpha/3-10-helix dimorphism in the crystal state," *Journal of the American Chemical Society*, vol. 129, no. 50, pp. 15471–15473, 2007. PMID : 18027948.

[52] A. D. Q. Buckingham, "Molecular quadrupole moments," *Rev. Chem. SOC.*, vol. 13, pp. 183–214, 1959.

[53] C. Katan, P. Rabiller, C. Lecomte, M. Guezo, V. Oison, and M. Souhassou, "Numerical computation of critical properties and atomic basins from three-dimensional grid electron densities," *Journal of Applied Crystallography*, vol. 36, pp. 65–73, Feb 2003.

[54] G. Gilli and P. Gilli, "Towards an unified hydrogen-bond theory," *Journal of Molecular Structure*, vol. 552, no. 1-3, pp. 1 – 15, 2000.

[55] T. Steiner, "The hydrogen bond in the solid state," *Angew. Chem. Int. Ed.*, vol. 41, pp. 48–76, 2002.

[56] G. R. Desiraju and T. Steiner, *The weak hydrogen bond in structural chemistry and biology*. Oxford, 2001.

[57] U. Koch and P. L. A. Popelier, "Characterization of C-H...O Hydrogen Bonds on the Basis of the Charge Density," *The Journal of Physical Chemistry*, vol. 99, pp. 9747–9754, June 1995.

[58] S. J. Grabowski, ed., *Hydrogen Bonding - New Insights*. Springer, 2006.

[59] E. Espinosa, E. Molins, and C. Lecomte, "Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities," *Chemical Physics Letters*, vol. 285, no. 3-4, pp. 170 – 173, 1998.

[60] E. Espinosa, C. Lecomte, and E. Molins, "Experimental electron density overlapping in hydrogen bonds : topology vs. energetics," *Chemical Physics Letters*, vol. 300, no. 5-6, pp. 745 – 748, 1999.

[61] E. Espinosa, M. Souhassou, H. Lachekar, and C. Lecomte, "Topological analysis of the electron density in hydrogen bonds," *Acta Crystallographica Section B*, vol. 55, pp. 563–572, Aug 1999.

[62] Z.-X. Wang, C. Wu, H. Lei, and Y. Duan, "Accurate ab initio study on the hydrogen-bond pairs in protein secondary structures," *Journal of Chemical Theory and Computation*, vol. 3, pp. 1527–1537, July 2007.

[63] S. M. LaPointe, S. Farrag, H. J. Bohorquez, and R. J. Boyd, "Qtaim study of an alpha-helix hydrogen bond network," *The Journal of Physical Chemistry B*, vol. 113, pp. 10957–10964, Aug. 2009.

[64] O. Koch, M. Bocola, and G. Klebe, "Cooperative effects in hydrogen-bonding of protein secondary structure elements : A systematic analysis of crystal data using secbase," *Proteins : Structure, Function, and Bioinformatics*, vol. 61, no. 2, pp. 310–317, 2005.

[65] R. Parthasarathi, S. S. Raman, V. Subramanian, and T. Ramasami, "Bader's electron density analysis of hydrogen bonding in secondary structural elements of protein," *The Journal of Physical Chemistry A*, vol. 111, no. 30, pp. 7141–7148, 2007.

[66] H. Berman, K. Henrick, and H. Nakamura, "Announcing the worldwide protein data bank," *Nature Structural Biology*, vol. 10, p. 980, 2003.

[67] Collaborative, "The ccp4 suite : programs for protein crystallography," *Acta Crystallographica Section D*, vol. 50, no. 5, pp. 760–763, 1994.

[68] J. Martin, G. Letellier, A. Marin, J.-F. Taly, A. de Brevern, and J.-F. Gibrat, "Protein secondary structure assignment revisited : a detailed analysis of different assignment methods," *BMC Structural Biology*, vol. 5, no. 1, p. 17, 2005.

[69] R. S. Paton and J. M. Goodman, "Hydrogen bonding and pi-stacking : How reliable are force fields ? a critical evaluation of force field descriptions of nonbonded interactions," *Journal of Chemical Information and Modeling*, vol. 49, pp. 944–955, Apr. 2009.

[70] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, "Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules," *Computer Physics Communications*, vol. 91, no. 1-3, pp. 1 – 41, 1995.

[71] W. L. Jorgensen and J. Tirado-Rives, "The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin," *Journal of the American Chemical Society*, vol. 110, no. 6, pp. 1657–1666, 1988.

[72] N. L. Allinger, "Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms," *Journal of the American Chemical Society*, vol. 99, no. 25, pp. 8127–8134, 1977.

[73] N. L. Allinger, Y. H. Yuh, and J. H. Lii, "Molecular mechanics. the mm3 force field for hydrocarbons. 1," *Journal of the American Chemical Society*, vol. 111, no. 23, pp. 8551–8566, 1989.

[74] B. I. Dahiyat, D. Benjamin Gordon, and S. L. Mayo, "Automated design of the surface positions of protein helices," *Protein Science*, vol. 6, no. 6, pp. 1333–1337, 1997.

[75] H. Choi, H. Kang, and H. Park, "New angle-dependent potential energy function for backbone-backbone hydrogen bond in protein-protein interactions," *J. Comput. Chem.*, vol. 31, no. 5, pp. 897–903, 2010.

[76] H. Choi, H. Kang, and H. Park, "Extended morse function model for angle-dependent hydrogen bond in protein-protein interactions," *The Journal of Physical Chemistry B*, vol. 114, pp. 2980–2987, Mar. 2010.

[77] F. Fabiola, R. Bertram, A. Korostelev, and M. S. Chapman, "An improved hydrogen bond potential : Impact on medium resolution protein structures," *Protein Science*, vol. 11, pp. 1415–1423, 2002.

[78] Z. S. Derewenda, L. Lee, and U. Derewenda, "The Occurence of C-H...O Hydrogen Bonds in Proteins," *Journal of Molecular Biology*, vol. 252, no. 2, pp. 248 – 262, 1995.

[79] H. Hauptman and J. Karle, "Solution of the phase problem. i. the centrosymmetric crystal.," *A.C.A. Monograph No. 3*, 1953.

[80] V. König, L. Vértesy, and T. R. Schneider, "Structure of the alpha-amylase inhibitor tendamistat at 0.93Å," *Acta Crystallographica Section D*, vol. 59, pp. 1737–1743, Oct 2003.

[81] J. Wang, M. Dauter, R. Alkire, A. Joachimiak, and Z. Dauter, "Triclinic lysozyme at 0.65 Å resolution," *Acta Crystallographica Section D*, vol. 63, pp. 1254–1268, 2007.

[82] N. Muzet, B. Guillot, C. Jelsch, E. Howard, and C. Lecomte, "Electrostatic complementarity in an aldose reductase complex from ultra-high-resolution crystallography and first-principles calculations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 15, pp. 8742–8747, 2003.

[83] P. Coppens, T. M. Sabine, G. Delaplane, and J. A. Ibers, "An experimental determination of the asphericity of the atomic charge distribution in oxalic acid dihydrate," *Acta Crystallographica Section B*, vol. 25, pp. 2451–2458, Dec 1969.

[84] D. Liebschner, M. Elias, S. Moniot, B. Fournier, K. Scott, C. Jelsch, B. Guillot, C. Lecomte, and E. Chabriére, "Elucidation of the phosphate binding mode of ding proteins revealed by subangstrom x-ray crystallography," *Journal of the American Chemical Society*, vol. 131, pp. 7879–7886, June 2009.

[85] M. Elias, D. Liebschner, G. Gotthard, and E. Chabriere, "Structural insights and ab initio sequencing within the DING proteins family," *Journal of Synchrotron Radiation*, vol. 18, Jan 2011.

[86] M.-M. Blum, M. Mustyakimov, H. Rüterjans, K. Kehe, B. P. Schoenborn, P. Langan, and J. C.-H. Chen, "Rapid determination of hydrogen positions and protonation states of diisopropyl fluorophosphatase by joint neutron and X-ray diffraction refinement," *Proceedings of the National Academy of Sciences*, vol. 106, no. 3, pp. 713–718, 2009.

[87] A. Richardt, M.-M. Blum, and S. Mitchell, "Was wissen calamari über sarin ? enzymatische dekontamination von nervenkampfstoffen," *Chemie in unserer Zeit*, vol. 40, no. 4, pp. 252–259, 2006.

[88] E. I. Scharff, C. Lücke, G. Fritzsch, J. Koepke, J. Hartleib, S. Dierl, and H. Rüterjans, "Crystallization and preliminary X-ray crystallographic analysis of DFPase from *Loligo vulgaris*," *Acta Crystallographica Section D*, vol. 57, pp. 148–149, Jan 2001.

[89] J. Hartleib, S. Geschwindner, E. I. Scharff, and H. Rüterjans, "Role of calcium ions in the structure and function of the diisopropylfluorophosphatase from loligo vulgaris.," *Biochem. J.*, vol. 353, no. 3, pp. 579–589, 2001.

[90] A. J. C. Wilson, ed., *International Tables for crystallogaphy*. Kluwer Academic Publishers, 1992.

[91] G. M. Sheldrick, "A short history of shelx," *Acta Crystallographica Section A*, vol. 64, pp. 112–122, 2007.

[92] D. E. Tronrud, "Tnt refinement package," in *Macromolecular Crystallography Part B* (C. W. C. Jr. and R. M. Sweet, eds.), vol. 277 of *Methods in Enzymology*, pp. 306 – 319, Academic Press, 1997.

[93] P. Emsley and K. Cowtan, "*Coot* : model-building tools for molecular graphics," *Acta Crystallographica Section D*, vol. 60, pp. 2126–2132, Dec 2004.

[94] M.-M. Blum, A. Koglin, H. Rüterjans, B. Schoenborn, P. Langan, and J. C.-H. Chen, "Preliminary time-of-flight neutron diffraction study on diisopropyl fluorophosphatase (DFPase) from *Loligo vulgaris*," *Acta Crystallographica Section F*, vol. 63, pp. 42–45, Jan 2007.

[95] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren, "Crystallography & NMR System : A New Software Suite for Macromolecular Structure Determination," *Acta Crystallographica Section D*, vol. 54, pp. 905–921, Sep 1998.

[96] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart, "PHENIX : a comprehensive Python-based system for macromolecular structure solution," *Acta Crystallographica Section D*, vol. 66, pp. 213–221, Feb 2010.

[97] F. H. Allen, "A systematic pairwise comparison of geometric parameters obtained by X-ray and neutron diffraction," *Acta Crystallographica Section B*, vol. 42, pp. 515–522, 1986.

[98] S. French and K. Wilson, "On the treatment of negative intensity observations," *Acta Crystallographica Section A*, vol. 34, pp. 517–525, 1978.

[99] J. E. Padilla and T. O. Yeates, "A statistic for local intensity differences : robustness to anisotropy and pseudo-centering and utility for detecting twinning," *Acta Crystallographica Section D*, vol. 59, pp. 1124–1130, Jul 2003.

[100] S. C. Capelli, H. B. Bürgi, and H. Birkedal, "Anharmonicity in anisotropic displacement parameters," *Acta Cryst. A*, vol. 56, no. 5, pp. 425–435, 2000.

[101] D. F. Stickle, L. G. Presta, K. A. Dill, and G. D. Rose, "Hydrogen-bonding in globular-proteins," *Journal of Molecular Biology*, vol. 226, pp. 1143–1159, Aug. 1992.

[102] M.-M. Blum and J. C.-H. Chen, "Structural characterization of the catalytic calcium-binding site in diisopropyl fluorophosphatase (dfpase) - comparison with related beta-propeller enzymes," *Chemico-Biological Interactions*, vol. In Press, Corrected Proof, pp. –, 2010.

[103] E. Ghanem and F. M. Raushel, "Detoxification of organophosphate nerve agents by bacterial phosphotriesterase," *Toxicol Appl Pharmacol*, vol. 207, pp. 459–70, Sep 1 2005.

[104] M. Harel, A. Aharoni, L. Gaidukov, B. Brumshtein, O. Khersonsky, R. Meged, H. Dvir, R. B. Ravelli, A. McCarthy, L. Toker, I. Silman, J. L. Sussman, and D. S. Tawfik, "Structure and evolution of the serum

paraoxonase family of detoxifying and anti-atherosclerotic enzymes," *Nat Struct Mol Biol*, vol. 11, pp. 412–9, may 2004.

[105] M.-M. Blum, F. Lohr, A. Richardt, H. Ruterjans, and J. C.-H. Chen, "Binding of a designed substrate analogue to diisopropyl fluorophosphatase : Implications for the phosphotriesterase mechanism," *Journal of the American Chemical Society*, vol. 128, pp. 12750–12757, Oct. 2006.

[106] C.-N. Chen, K.-H. Chin, A. H.-J. Wang, and S.-H. Chou, "The first crystal structure of gluconolactonase important in the glucose secondary metabolic pathways," *Journal of Molecular Biology*, vol. 384, no. 3, pp. 604 – 614, 2008.

[107] S. Moniot, M. Elias, D. Kim, K. Scott, and E. Chabriére, "Crystallization, diffraction data collection and preliminary crystallographic analysis of ding protein from pseudomonas fluorescens," *Acta Crystallographica Section F*, vol. 63, pp. 590–592, 2007.

[108] A. Berna, F. Bernier, K. Scott, and B. Stuhlmüller, "Ring up the curtain on ding proteins," *FEBS Letters*, vol. 524, pp. 6–10, 2002.

[109] A. Berna, F. Bernier, E. Chabriére, T. Perera, and K. Scott, "Ding proteins ; novel members of a prokaryotic phosphate-binding protein superfamily which extends into the eukaryotic kingdom," *The International Journal of Biochemistry & Cell Biology*, vol. 40, no. 2, pp. 170 – 175, 2008.

[110] H. Diemer, M. Elias, F. Renault, D. Rochu, C. Contreras-Martel, C. Schaeffer, A. V. Dorsselaer, and E. Chabriére, "Tandem use of x-ray crystallography and mass spectrometry to obtain ab initio the complete and exact amino acids sequence of hpbp, a human 38-kda apolipoprotein," *Proteins*, vol. 71, pp. 1708–1720, 2008.

[111] H. Luecke and F. A. Quiocho, "High specificity of a phosphate transport protein determined by hydrogen bonds," *Nature*, vol. 347, pp. 402–406, 1990.

[112] R. Morales, A. Berna, P. Carpentier, C. Contreras-Martel, F. Renault, M. Nicodeme, M. Chesne-Seck, F. Bernier, J. Dupuy, C. Schaeffer, H. Diemer, A. Van-Dorsselaer, J. C. Fontecilla-Camps, P. Masson, D. Rochu, and E.Chabriére, "Serendipitous discovery and x-ray structure of a human phosphate binding apolipoprotein," *Structure*, vol. 14, pp. 601–609, 2006.

[113] M. Belenky, J. Prasain, H. Kim, and S. Barnes, "Ding, a genistein target in human breast cancer : A protein without a gene," *J Nutr.*, vol. 133, pp. 2497S–2501S, 2003.

[114] Z. Wang, H. Luecke, N. Yao, and F. A. Quiocho, "A low energy short hydrogen bond in very high resolution structures of protein receptor-phosphate complexes," *nature structural biology*, vol. 4, pp. 519–522, 1997.

[115] P. S. Ledvina, A. L. Tsai, Z. Wang, E. Koehl, and F. A. Quiocho, "Dominant role of local dipolar interactions in phosphate binding to a receptor cleft with an electronegative charge surface : Equilibrium, kinetic, and crystallographic studies," *Protein Sci.*, vol. 7, pp. 2550–2559, 1998.

[116] W. Kabsch, "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants," *J. Appl. Cryst.*, vol. 26, pp. 795–800, 1993.

[117] G. N. Murshudov, A. A. Vagin, A. Lebedev, K. S. Wilson, and E. J. Dodson, "Effcient anisotropic refinement of macromolecular structures using fft," *Acta Crystallographica Section D*, vol. 55, pp. 247–255, 1999.

[118] G. M. Sheldrick and T. R. Schneider, "Shelxl : High-resolution refinement," *Methods in Enzymology*, vol. 277, pp. 319–343, 1997.

[119] G. N. Murshudov, A. A. Vagin, and E. J. Dodson, "Refinement of Macromolecular Structures by the Maximum-Likelihood Method," *Acta Crystallographica Section D*, vol. 53, pp. 240–255, May 1997.

[120] A. Fokine and A. Urzhumtsev, "Flat bulk-solvent model : obtaining optimal parameters," *Acta Crystallographica Section D*, vol. 58, pp. 1387–1392, 2002.

[121] A. Boukhris, M. Souhassou, C. Lecomte, B. Wyncke, and A. Thalal, "Evolution of the structural and mean square displacement parameters in solid solutions versus concentration and temperature," *Journal of Physics : Condensed Matter*, vol. 10, no. 7, pp. 1621–1641, 1998.

[122] P. Gilli, V. Bertolasi, L. Pretto, V. Ferretti, and G. Gilli, "Covalent versus electrostatic nature of the strong hydrogen bond : discrimination among single, double, and asymmetric single-well hydrogen bonds by variable-temperature x-ray crystallographic methods in beta-diketone enol rahb systems," *Journal of the American Chemical Society*, vol. 126, no. 12, pp. 3845–3855, 2004.

[123] W. W. Cleland, "Low-barrier hydrogen bonds and enzymatic catalysis," *Archives of Biochemistry and Biophysics*, vol. 382, no. 1, pp. 1 – 5, 2000.

[124] F. H. Allen, "The Cambridge Structural Database : a quarter of a million crystal structures and rising," *Acta Crystallographica Section B*, vol. 58, pp. 380–388, Jun 2002.

[125] B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte, and C. Jelsch, "On the application of an experimental multipolar pseudo-atom library for accurate refinement of small-molecule and protein crystal structures," *Acta Crystallographica Section A*, vol. 63, pp. 108–125, Mar 2007.

[126] P. S. Ledvina, N. Yao, A. Choudhary, and F. A. Quiocho, "Negative electrostatic surface potential of protein sites specific for anionic ligands," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 13, pp. 6786–6791, 1996.

[127] A. Nicholls, K. A. Sharp, and B. Honig, "Protein folding and association : insights from the interfacial and thermodynamic properties of hydrocarbons," *Proteins*, vol. 11, pp. 281–296, 1991.

[128] W. W. Cleland, P. A. Frey, and J. A. Gerlt, "The Low Barrier Hydrogen Bond in Enzymatic Catalysis," *J. Biol. Chem.*, vol. 273, no. 40, pp. 25529–25532, 1998.

[129] M. Garcia-Viloca, A. Gonzalez-Lafont, and J. M. Lluch, "Theoretical study of the low-barrier hydrogen bond in the hydrogen maleate anion in the gas phase. comparison with normal hydrogen bonds," *Journal of the American Chemical Society*, vol. 119, no. 5, pp. 1081–1086, 1997.

[130] V. Dahirel, F. Paillusson, M. Jardat, M. Barbi, and J.-M. Victor, "Nonspecific dna-protein interaction : Why proteins can diffuse along dna," *Phys. Rev. Lett.*, vol. 102, p. 228101, Jun 2009.

Cette étude est consacrée à l'analyse des propriétés électrostatiques et structurales des protéines diffractant à haute résolution. Les travaux se déclinent en deux aspects principaux qui sont, d'une part, l'analyse d'un point de vue fondamental des propriétés dérivées de leur distribution de charge des structures des protéines, et, d'autre part, l'application pratique des méthodes de la cristallographie haute résolution des macromolécules à deux enzymes.
Grâce à une analyse structurale et électrostatique de la protéine PfluDING, le mode de fixation du phosphate a été mis en évidence à deux pH différents. En particulier, cette étude a permis de démontrer la présence d'une liaison hydrogène diffuse assurant un mode de fixation identique quelque soit le pH. La seconde enzyme étudiée est la protéine DFPase, capable de dégrader les organophosphorés. La structure de la DFPase obtenue par diffraction X haute résolution et la structure affinée contre des données neutroniques ont été comparées. L'analyse pointe les différences d'interprétation des deux modèles, et permet de proposer un nouveau mécanisme d'action.
Les aspects plus fondamentaux de cette étude portent sur les éléments de structure secondaire des protéines : leurs propriétés électrostatiques en termes de moments électriques (moments dipolaires et quadripolaires des hélices α et des feuillets β), et le réseau des interactions (dont les liaisons H) assurant la cohésion des hélices ont été analysés. Il a été démontré comment certaines interactions entre liaisons peptidiques au sein d'hélices, classiquement représentées comme des liaisons hydrogène, devaient en fait être considérés comme des contacts purement électrostatiques.

MOTS CLES :
liaison hydrogène, analyse topologique, cristallographie macromoléculaire, haute résolution, moment électrostatique, hélice, feuillet

This study is about the electrostatic and structural properties of proteins diffracting at high X-ray resolution. Two different aspects are tackled which are 1) the analysis of properties derived from their charge distribution, parting from a fundamental point of view and 2) the application of methods used in high resolution macromolecular crystallography to two enzymes of major interest.
After the analysis of the structure and electrostatic properties of PfluDING protein, the binding mode of a phosphate ion, located in the active site, was elucidated at two different pH values. Particularly, this study demonstrates that a diffuse hydrogen bond assures the protonation state of the phosphate ion, which is thus identical at each pH value. The second enzyme studied is the proteine DFPase which is capable of hydrolysing nerve agents. A high resolution X-ray structure and a medium resolution neutron structure where compared. The analysis points out the differences between the two models and a new catalytic mechanism could be proposed.
The more fundamental aspects of this study are about the secondary structural elements of proteins : their electrostatic properties in terms of electrostatic moments (dipole and quadrupole moments of helices and sheets) as well as the hydrogen bond network assuring the cohesion of helices have been analyzed. It has been shown that certain interactions between peptide units within helices, represented usually as hydrogen bonds, should actually be considered as pure electrostatic contacts.

KEY WORDS :
hydrogen bond, topological analysis, macromolecular crystallography, high resolution, electrostatic moment, helix, sheet