



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 & TECHNIQUES
U.F.R. Sciences & Techniques de la Matière et des Procédés
Ecole Doctorale EMMA
Département de Formation Doctorale STMP

THESE

Présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy I
En Physique et Chimie de la Matière et des Matériaux

Par

Bertrand FOURNIER

TITRE

Modélisation des propriétés électrostatiques des complexes
macromoléculaires à partir des données de diffraction
des rayons X à très haute résolution

Soutenance publique prévue le 6 Juillet 2010

Membres de Jury :

Présidente :	M ^{me} Anne RENAULT	Directeur de Recherche, Université de Rennes I.
Rapporteurs :	M ^r Gérard BRICOGNE	Directeur de Global Phasing Ltd, Cambridge, G-B & Membre correspondant à l'Académie des Sciences.
	M ^r Marc SCHILTZ	Professeur, Ecole Polytechnique Fédérale de Lausanne, Suisse.
Examineurs :	M ^{me} Sine LARSEN	Professeur, University of Copenhagen & Présidente de l'IUCr.
	M ^r Bertrand BERCHE	Professeur, U.H.P., Nancy I.
	M ^r Christian JELSCH	Chargé de Recherche, U.H.P., Nancy, Directeur de Thèse.
	M ^r Claude LECOMTE	Professeur, U.H.P., Nancy I.
Invité :	M ^r Benoît GUILLOT	Maître de conférences, U.H.P., Nancy I.

Remerciements

Ce travail de recherche a été réalisé au laboratoire de Cristallographie, Résonance Magnétique et Modélisations (CRM²) (UMR-CNRS 7036) de la Faculté des Sciences et Techniques de l'université Henri Poincaré, Nancy I.

Je voudrais adresser mes remerciements à Monsieur Claude LECOMTE (Professeur) pour m'avoir accueilli au laboratoire et pour le soutien qu'il m'a apporté.

Je tiens à remercier également Monsieur Christian JELSCH (Chargé de recherche au CNRS), ainsi que Monsieur Benoît GUILLOT (Maître de conférences) pour avoir encadré mon travail de doctorat et m'avoir donné l'opportunité de découvrir la cristallographie.

J'exprime toute ma reconnaissance aux membres du jury :

A Messieurs Gérard BRICOGNE et Marc SHILTZ qui m'ont fait l'honneur de juger ce travail, qu'ils trouvent ici l'expression de mes sincères remerciements.

Je tiens également à remercier Madame Anne RENAULT (Professeur) de l'Institut de Physiques de Rennes pour avoir présidé mon jury de soutenance de thèse, mais aussi Madame Sine LARSEN (Professeur) du département de chimie de l'université de Copenhague et Monsieur Bertrand BERCHE (Professeur) de l'IJL de Nancy pour avoir accepté le rôle d'examineur.

Je tiens à exprimer toute ma reconnaissance à Messieurs El-Eulmi BENDEIF (Maître de conférences) et Enrique ESPINOSA (Professeur) pour leur soutien et leur disponibilité sans faille durant ma thèse.

Que Messieurs Sébastien LEBEGUE (Chargé de recherche au CNRS), Slimane DAHAOUI (Maître de conférences), Claude DIDIERJEAN (Maître de conférences), Janos ANGYAN (Directeur de recherche au CNRS) et Emmanuel AUBERT (Maître de conférences), ainsi que Madame Virginie PICHON-PESME (Maître de conférences) trouvent ici toute ma gratitude pour nos discussions enrichissantes et leurs conseils fructueux.

J'exprime toute ma reconnaissance à Monsieur Alberto PODJARNY (Professeur à l'IGBMC de Strasbourg) pour son aide dans l'étude du fidarestat et des complexes aldo-keto reductase – inhibiteurs.

Je souhaite également remercier Monsieur Angelo GAVEZZOTTI (Professeur au DCSSI de Milan) et Madame Nouzha BOUHMAIDA (Maître de conférences au LSM de Marrakech) pour nos échanges concernant le calcul des énergies d'interaction intermoléculaire.

Je tiens à remercier tout le personnel du laboratoire qui m'a apporté son aide tout au long de ces années de thèse. Que Messieurs Emmanuel WENGER et Abdelatif DOUDOUH soient remerciés pour l'assistance technique qui m'ont apporté dans l'étude du fidarestat.

Je n'oublierai pas mes camarades de bureau qui ont su créer une très bonne ambiance de travail : Maxime, Isabella, Nawel, Mariya, Adellatif, Sophie, William, Elena, Ru-Fen et Slawomir, mais aussi l'ensemble des membres du laboratoire.

Je tiens à saluer également mes camarades de sport, notamment Fabrice et Romain, qui m'ont fait partager leur passion avec simplicité et ouverture d'esprit.

Je remercie chaleureusement ma famille, ma soeur et en particulier mes parents pour leur soutien tout au long de mes études supérieures et leurs encouragements.

Pour finir, je tiens à dédier ce travail à ma grand-mère Gilberte qui a toujours cru en moi, ainsi qu'à mon neveu Baptiste qui nous a quitté trop tôt.

SOMMAIRE

Introduction	1
I. La diffraction des rayons X, Techniques & Méthodes	9
I.1. La diffraction des rayons X	11
I.1.1. Pourquoi la diffraction des rayons X ?	11
I.1.2. La clé de la diffraction des rayons X sur un monocristal	11
I.1.3. Facteur d'agitation thermique ou de Debye-Waller	13
I.2. L'affinement cristallographique et les différents modèles couramment utilisés	16
I.2.1. Qu'est-ce qu'un affinement cristallographique	16
I.2.2. Les modèles de distribution de charges	17
I.2.3. Stratégies d'affinement cristallographique	24
I.2.4. Critères de validation des affinements cristallographiques	33
II. Caractérisation de la distribution de charges d'un composé et propriétés électrostatiques dérivées	43
II.1. Propriétés électrostatiques	46
II.1.1. Le potentiel électrostatique	46
II.1.2. Le Laplacien de densité électronique	48
II.1.3. L'analyse topologique de la distribution de charges	49
II.2. Energies d'interaction intermoléculaire	55
II.2.1. Expression de l'énergie d'interaction totale entre deux composés	55
II.2.2. Calcul des énergies d'interaction électrostatique	60
II.2.3. La banque de paramètres multipolaires, ELMAM	85

III. Etude de la densité de charges du fidarestat, un inhibiteur de l'aldose réductase humaine	91
III.1. Introduction	93
III.2. Matériels et méthodes	95
III.2.1. Cristallisation	95
III.2.2. Collecte des données	95
III.2.3. Réduction des données	96
III.2.4. Affinement du modèle IAM	97
III.2.5. Affinement multipolaire	97
III.2.6. Stratégie d'affinement	97
III.3. Densité électronique du fidarestat	100
III.3.1. Affinement cristallographique	100
III.3.2. Structure et empilement cristallins	100
III.3.3. Densité électronique de déformation	105
III.3.4. Topologie des liaisons covalentes	107
III.3.5. Charges atomiques nettes	109
III.4. Propriétés électrostatiques du fidarestat	111
III.4.1. Potentiel électrostatique du fidarestat	111
III.4.2. Topologie des liaisons hydrogène H...O	112
III.4.3. Les liaisons hydrogène π ...H—X	113
III.4.4. L'empilement cristallin et les énergies d'interaction électrostatique	117
III.5. Conclusion	119

IV. Estimation de l'incertitude statistique des énergies d'interaction électrostatique au sein des systèmes macromoléculaires	125
IV.1. Motivations et objectifs	127
IV.2. Matériels et méthodes	132
IV.2.1. Outils d'affinement cristallographique	133
IV.2.2. Outils d'analyse statistique	136
IV.2.3. Fonction de perturbation	140
IV.3. Etude préliminaire	143
IV.3.1. Présentation du complexe aldose réductase – sorbinil	143
IV.3.2. Préparation « Shake »	145
IV.3.3. Etudes statistiques des indicateurs de qualité	147
IV.3.4. Etudes statistiques des coordonnées atomiques	152
IV.3.5. Détermination des distributions d'énergies d'interaction électrostatique	162
IV.3.6. Discussion des conditions du test	169
IV.3.7. Discussion des résultats obtenus	191
IV.4. Applications à d'autres complexes macromoléculaires	197
IV.4.1. Choix des complexes	197
IV.4.2. Cas de la structure de l'aldose réductase complexée avec l'inhibiteur fidarestat	198
IV.4.3. Cas de la structure de l'aldéhyde réductase complexée avec l'inhibiteur fidarestat	206
IV.4.4. Cas de la structure de l'aldose réductase complexée avec le stéréoisomère (2 <i>R</i> ,4 <i>S</i>) de l'inhibiteur fidarestat	220

IV.5. Discussion & Perspectives	232
Conclusion et perspectives	239
Annexes	249

Introduction

La cristallographie consiste en l'étude des cristaux à leur échelle atomique. Elle est fondée sur différentes méthodes expérimentales (diffraction des rayons X sur monocristal et sur poudre, diffraction des neutrons) permettant d'obtenir des informations structurales sur les composés sous forme cristalline, et plus particulièrement, l'agencement des atomes dans l'espace. Ces informations sont d'une importance fondamentale car elles conditionnent les propriétés des matériaux qu'ils soient organiques ou inorganiques.

Les progrès techniques et méthodologiques en diffraction des rayons X sur monocristal ont permis d'accéder à l'échelle atomique (0,9-1,5 Å) qui permet de positionner les atomes des molécules à l'état cristallin. Puis, l'échelle subatomique (0,6-0,9 Å) a pu également être atteinte ce qui permet de percevoir la déformation du nuage électronique des atomes due aux liaisons chimiques et aux interactions intermoléculaires. Cependant, ce n'est qu'à une résolution ultra-haute (0,3–0,6 Å) que la cristallographie des rayons X permet de quantifier cette déformation.

En réponse à cette avancée technique, de nouveaux modèles de distribution de charges permettant de prendre en compte la déformation du nuage électronique ont été proposés. Parmi ces derniers, le modèle atomique de Hansen & Coppens [5] donne une représentation analytique en termes de multipôles de la densité de charges expérimentale. Disposer d'un tel modèle permet non seulement d'étudier la structure atomique du composé comme cela est le cas usuellement, mais aussi d'analyser la distribution de charges du composé et d'estimer ses propriétés dérivées : transfert de charges, potentiel et énergie d'interaction électrostatique.

Depuis 1990, le nombre de structures cristallographiques de macromolécules biologiques telles que les protéines à haute résolution déposées dans la Protein Data Bank (PDB) [1] augmente de façon significative. Une résolution ultra-haute a même, dans quelques cas, pu être atteinte. On peut donner comme exemples de structures : la crambin de résolution 0,54 Å [9], une protéine antigél à 0,62 Å [10], la protéine lysozyme à 0,65 Å [14], aldose réductase humaine (hAR) à 0,66 Å [13;6].

La faisabilité d'un affinement multipolaire pour des structures de protéines a été prouvée [8;9], dans le cas de données de diffraction X sur monocristal à résolution subatomique et d'une agitation thermique moléculaire faible. Malgré les améliorations significatives des dispositifs expérimentaux, les affinements multipolaires de systèmes macromoléculaires restent rarement possibles. Seulement deux affinements de modèle de charges non-sphérique de protéines ont été rapportés dans la littérature [9;4].

Afin de pallier ces limites techniques, la transférabilité des paramètres de charges est un moyen fiable pour obtenir un modèle estimé de distribution de charges pour un système macromoléculaire, et dériver de ce modèle ses propriétés électrostatiques.

Les paramètres de distribution de charges non-sphérique peuvent être obtenus à partir de différentes banques de paramètres. Récemment, le principe de transférabilité a été employé dans plusieurs études de complexes enzyme-inhibiteur. Nous avons construit des modèles de distribution de charges pour des complexes aldose réductase humaine (hAR) à partir de la banque expérimentale ELMAM (Experimental Library of Multipolar Atom Model) [12;15] de notre laboratoire [3]. Le même type d'étude a été mené sur des complexes neuraminidase-inhibiteur en utilisant la banque théorique UBDB [11] (University of Buffalo DataBase) [2]. Des modèles de distribution de charges non-sphérique apportent une meilleure description des propriétés électrostatiques qui sont la clé de la reconnaissance ligand-récepteur compte tenu de la directionnalité des interactions électrostatiques et le rôle majeur des liaisons hydrogène dans les systèmes biologiques.

Les résultats présentés dans ce manuscrit correspondent à une partie des travaux que j'ai réalisés ou auxquels j'ai participé au cours de mon doctorat au laboratoire de Cristallographie, Résonance Magnétique et Modélisations (CRM2). Ce travail s'intègre dans une dynamique visant à étendre les méthodes et outils initialement développés pour l'étude des distributions de charges de petites molécules aux systèmes macromoléculaires. Il s'articule autour de la faisabilité du transfert de paramètres multipolaires issus de la banque ELMAM pour l'étude des énergies d'interaction électrostatique au sein du site actif de complexes enzyme-inhibiteur et tente également d'en discuter la légitimité.

Le chapitre I de ce manuscrit est une présentation générale de la cristallographie des rayons X, des méthodologies et outils employés pour permettre l'affinement cristallographique de petites molécules et des macrosystèmes.

Le chapitre II est plus spécifique au modèle de distribution de charges non-sphérique, et présente les différents calculs de propriétés électrostatiques (potentiel électrostatique, Laplacien,...) et analyses pouvant être menés sur un modèle de charges afin d'étudier les propriétés du système. Les améliorations logicielles que j'ai réalisées dans le suite de logiciels pour la cristallographie MoPro [7] de calcul, car essentielles à ce travail, sont à cette occasion exposées.

Le chapitre III expose l'étude du fidarestat, un inhibiteur de l'aldose réductase humaine, à partir de données obtenues à très haute résolution par diffraction des rayons X sur monocristal. Le chapitre IV présente les premiers résultats de calcul d'énergies d'interaction électrostatique au sein du site actif de complexes enzyme-inhibiteur, à partir de modèles estimés de distribution de charges non-sphérique définis par transfert de paramètres multipolaires depuis la banque ELMAM [12;15], et propose une méthode permettant d'estimer l'incertitude sur ces énergies d'interaction électrostatique par le biais d'une approche statistique.

Je terminerai ce manuscrit par un bilan du travail exposé, et par une discussion sur les perspectives envisagées.

- [1] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne. *Nucleic Acids Research*, **28**, 235–242, 2000.
- [2] P. M. Dominiak, A. Volkov, A. P. Dominiak, K. N. Jarzemska, P. Coppens. *Acta Crystallographica*, **D65**, 485–499, 2009.
- [3] B. Fournier, E-E Bendeif, B., Guillot, A. Podjarny, C. Lecomte, C. Jelsch. *Journal of American Chemistry Society*, **131**(31), 10929–10941, 2009.
- [4] B. Guillot, C. Jelsch, A. Podjarny, C. Lecomte. *Acta Crystallographica*, **D64**, 567–588, 2008.
- [5] N. K. Hansen & P. Coppens. *Acta Crystallography*, **A34**, 909–921, 1978.
- [6] E. R. Howard, R. Sanishvili, R. E. Cachau, A. Mitschler, B. Chevrier, P. Barth, V. Lamour, M. Van Zandt, E. Sibley, C. Bon, D. Moras, T. R. Schneider, A. Joachimiak, A. Podjarny. *Proteins*, **55**, 792–804, 2004.
- [7] C. Jelsch, B. Guillot, A Lagoutte, C. Lecomte. *Journal of Applied Crystallography*, **38**, 38–54, 2005.
- [8] C. Jelsch, V. Pichon-Pesme, C. Lecomte, A. Aubry. *Acta Crystallographica*, **D54**, 1306–1318, 1998.
- [9] C. Jelsch, M. M. Teeter, V. Lamzin, V. Pichon-Pesme, R. H. Blessing, C. Lecomte. *Proceedings of the National Academy of Sciences (USA)*, **97**, 3171–3176, 2000.
- [10] T.-P. Ko, H. Robinson, Y. G. Gao, C. H. Cheng, A. L. DeVries, A. H. Wang. *Biophysical Journal*, **84**, 1228–1237, 2003.
- [11] T. S. Koritsanszky, A. Volkov, P. Coppens. *Acta Crystallographica*, **A58**, 464–472, 2002.
- [12] V. Pichon-Pesme, C. Lecomte, H. Lachekar. *Journal of Physical Chemistry*, **99**, 6242–6250, 1995.
- [13] A. Podjarny, E. Howard, A. Mitschler, B. Chevrier, C. Lecomte, B. Guillot, V. Pichon-Pesme, C. Jelsch. *Europhysics News*, **33**, 113–117, 2003.
- [14] J. Wang, M. Dauter, R. Alkire, A. Joachimiak, Z. Dauter. *Acta Crystallographica*, **D63**, 1254–1268, 2007.
- [15] B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte, C. Jelsch. *Acta Crystallographica*, **A63**, 108–125, 2007.

Chapitre I

La diffraction des rayons X

Techniques & Méthodes

I.1. La diffraction des rayons X

I.1.1. Pourquoi la diffraction des rayons X ?

La diffraction des rayons X permet d'obtenir des informations structurales sur des matériaux moléculaires via la détermination de leur distribution électronique. Elle peut être utilisée pour différents types d'études : les minéraux et sels inorganiques, les molécules organiques, organométalliques, protéines dont la taille peut aller de quelques atomes à plusieurs milliers. Comme toute méthode expérimentale, la diffraction des rayons X repose sur des propriétés physiques de la matière ce qui induit des conditions requises et par là même des limites d'applications.

I.1.2. La clé de la diffraction des rayons X sur un monocristal

La diffraction des rayons X par un monocristal est un phénomène d'interaction particule - matière. L'échantillon est en effet exposé à un faisceau de rayons X qui va interagir avec la distribution électronique du matériau.

Les rayons X peuvent, comme toute onde électromagnétique, interagir avec la matière selon deux phénomènes : la diffusion et l'absorption. Le phénomène de diffusion est la clé des expériences de diffraction des rayons X. Dans les expériences de diffraction des rayons X, la diffusion est essentiellement élastique mais peut cependant être inélastique selon les espèces chimiques rencontrées dans le matériau et la longueur d'onde des rayons X utilisés. La diffraction des rayons X par un échantillon conduit à la diffusion dans toutes directions de rayons X qui, bien que détectables, ne peuvent être exploités car constituant sur la surface d'un détecteur un continuum de faisceaux diffusés qui ne peuvent être distingués les uns des autres.

La diffraction des rayons X repose également sur le fait que les échantillons utilisés sont des cristaux en 3 dimensions et présentent, par définition, une périodicité spatiale. Cette périodicité de l'échantillon affecte le schéma de diffraction et impose des directions préférentielles de diffusion.

Afin d'expliquer le principe de la diffraction des rayons X par un monocristal, l'on se limitera à la diffusion élastique ou Thomson. La démonstration de ce principe est faite sous certaines approximations connues sous le nom d'approximation cinématique.

Dans un cas à une dimension, l'amplitude du faisceau diffusé $E(\mathbf{H})$ dans la direction du vecteur de diffusion $\mathbf{H} = \frac{\mathbf{S} - \mathbf{S}_0}{\lambda}$, par l'ensemble des atomes présent dans la maille de l'échantillon périodique peut s'exprimer comme le produit du facteur de structure global $F(\mathbf{H})$ et le terme $\Gamma(\mathbf{H})$ appelé « fonction d'interférence » du cristal.

$$E(\mathbf{H}) = E_e(\mathbf{H}) \sum_{j=1}^{n_{at}} f_j \exp(2\pi i \mathbf{H} \cdot (x_j) \mathbf{a}) \times \Gamma(\mathbf{H}) = E_e(\mathbf{H}) \times F(\mathbf{H}) \times \Gamma(\mathbf{H}) \quad (1.1)$$

avec f_j , le facteur de diffusion atomique du $j^{\text{ème}}$ atome et $E_e(\mathbf{H})$, l'amplitude du faisceau diffracté par un électron. On note également $I_e(\mathbf{H})$, l'intensité correspondante.

$F(\mathbf{H})$ dépend seulement de la nature et des positions des atomes dans la maille. C'est l'amplitude de ce terme qui est mesurée expérimentalement. La connaissance de ce terme permet de connaître le motif.

$\Gamma(\mathbf{H})$ ne dépend pas de la composition de la maille mais uniquement de ses dimensions. Il permet de caractériser le réseau.

L'intensité d'un faisceau diffracté devient :

$$I(\mathbf{H}) = I_e(\mathbf{H}) \times (F(\mathbf{H}) \times \overline{F(\mathbf{H})}) \times \left(\frac{\sin(\pi V \mathbf{H} \cdot \mathbf{a})}{\sin(\pi \mathbf{H} \cdot \mathbf{a})} \right)^2 \quad (1.2)$$

Cette formule (1.2) nous montre que la périodicité de l'échantillon implique que les intensités ne seront significatives que lorsque $\pi \mathbf{H} \cdot \mathbf{a} = k\pi$ avec k , un entier. Les rayons diffusés interfèrent de manière constructive que dans ces directions spécifiques de l'espace.

Dans le cas d'un échantillon cristallin réel et donc à 3 dimensions ($\mathbf{a}, \mathbf{b}, \mathbf{c}$), il y aura 3 conditions au lieu d'une seule. Ces conditions sont dites « de Laue » :

Si h, k, l des entiers, $\mathbf{H} \cdot \mathbf{a} = h$, $\mathbf{H} \cdot \mathbf{b} = k$ et $\mathbf{H} \cdot \mathbf{c} = l$

De plus amples détails concernant l'histoire et la théorie de la diffraction des rayons X par un monocristal sont communiqués en annexe. (**Annexe A**)

I.1.3. Facteur d'agitation thermique ou de Debye-Waller

Les expressions précédentes ont été définies sous l'hypothèse que les atomes sont gelés, c'est-à-dire qu'ils n'ont pas de déplacements thermiques et que la structure est parfaitement ordonnée. Dans la réalité, les atomes sont toujours animés d'un mouvement. L'amplitude de ces mouvements va dépendre de la température de l'expérience mais aussi de la nature des interactions dans l'environnement cristallin. Les déplacements thermiques réduisent les amplitudes des facteurs de structures, donc les intensités diffractées, et doivent être pris en compte.

L'on peut voir ces déplacements comme des oscillations des atomes autour de leurs positions d'équilibre. Les déplacements thermiques d'un atome peuvent être vus comme une incertitude sur la position de cet dernier.

En première approximation, les atomes sont supposés être des oscillateurs harmoniques non couplés. Les différents déplacements \mathbf{u}_j du $j^{\text{ème}}$ atome autour de sa position d'équilibre résultant de l'agitation thermique suivent une loi gaussienne dont la densité de probabilité est donnée par :

$$P(\mathbf{u}_j) = \frac{e^{-\frac{|\mathbf{u}_j|^2}{2\langle \mathbf{u}_j^2 \rangle}}}{\left(\sqrt{2\pi \langle \mathbf{u}_j^2 \rangle}\right)^3} \quad (1.3)$$

avec $\langle \mathbf{u}_j^2 \rangle$, la moyenne quadratique du vecteur déplacement.

La fonction de distribution $P(\mathbf{u}_j)$ (1.3) des déplacements est supposée isotrope : la probabilité d'un déplacement \mathbf{u}_j n'est fonction que de son amplitude et non de sa direction et de son sens.

Une expérience de diffraction des rayons X sur monocristal permet d'obtenir des informations sur la densité électronique moyenne de la maille cristallographique.

La densité électronique du $j^{\text{ème}}$ atome de la maille $\rho_{dyn,j}$ est définie comme le produit de convolution de sa densité statique $\rho_{stat,j}$ et de la fonction de distribution $P(\mathbf{u}_j)$:

$$\rho_{dyn,j} = \rho_{stat,j} \otimes P(\mathbf{u}_j) \quad (1.4)$$

Le facteur de température ou de Debye-Waller correspondant est défini comme la transformée de Fourier de la fonction de distribution $P(\mathbf{u}_j)$:

$$T_j(\mathbf{H}) = \int P(\mathbf{u}_j) e^{2\pi i(\mathbf{H} \cdot \mathbf{u}_j)} d\mathbf{u}_j = e^{-2\pi^2 |\mathbf{H}|^2 \langle \mathbf{u}_j^2 \rangle} = e^{-B_j \frac{\sin^2 \theta}{\lambda^2}} \quad (1.5)$$

et intégré à chaque contribution atomique dans l'expression du facteur de structure :

$$F_{cell}(\mathbf{H}) = \sum_{j=1}^{n_{at}} f_j e^{-B_j \frac{\sin^2 \theta}{\lambda^2}} e^{2\pi i \mathbf{H} \cdot \mathbf{r}_j} \quad (1.6)$$

avec $B_j = 8\pi^2 \langle \mathbf{u}_j^2 \rangle$, le facteur B isotrope du $j^{\text{ème}}$ atome

L'on peut aussi introduire une nouvelle expression pour le facteur de diffusion incluant le facteur de Debye-Waller :

$$f_j(\mathbf{H}) = f_j^0(\mathbf{H}) e^{-B_j \frac{\sin^2 \theta}{\lambda^2}} \quad (1.7)$$

Les déplacements thermiques d'un atome sont en réalité restreints selon certaines directions de l'espace en raison de l'environnement chimique, en particulier des atomes qui lui sont liés de manière covalente. Pour cette raison, l'approximation isotrope est trop basique pour modéliser les déplacements thermiques réels. Il est possible d'utiliser un modèle d'agitation thermique anisotrope.

Ce modèle est également un oscillateur harmonique non-couplé mais utilisant trois gaussiennes pour reproduire les déplacements thermiques.

La fonction des déplacements thermiques devient alors une fonction ellipsoïde centrée sur la position à l'équilibre.

Six paramètres sont nécessaires à ce modèle : 3 paramètres déterminant l'orientation des axes et 3 autres pour la largeur de l'ellipsoïde le long de chaque axe.

Dans ce cas, le facteur de déplacement thermique du $j^{\text{ème}}$ atome est :

$$T_j(\mathbf{H}) = e^{-\sum_{k,l} 2\pi^2 h_k h_l (\mathbf{a}_k \cdot \mathbf{a}_l)^* (U_{kl})_j} = e^{-2\pi^2 \mathbf{H} \mathbf{U}_j \mathbf{H}} \quad (1.8)$$

avec $(U_{kl})_j$, l'élément kl du tenseur de déplacement atomique du $j^{\text{ème}}$ atome.

L'amplitude diffusée par le cristal peut être écrite :

$$E(\mathbf{H}) = E_e(\mathbf{H}) \times F(\mathbf{H}) \times \Gamma(\mathbf{H}) \quad (1.9)$$

$$\text{avec } F(\mathbf{H}) = \sum_{j=1}^{n_a} f_j(\mathbf{H}) e^{2\pi i \mathbf{H} \cdot \mathbf{r}_j} T_j(\mathbf{H}) \quad (1.10)$$

Les intensités diffusées par diffraction des rayons X permettent d'obtenir les modules des facteurs de structure :

$$I(\mathbf{H}) = K |F(\mathbf{H})|^2 \quad (1.10)$$

avec K , le facteur de mise à l'échelle.

L'expression sous forme complexe d'un facteur de structure est :

$$F(\mathbf{H}) = |F(\mathbf{H})| e^{i\varphi(\mathbf{H})} \quad (1.11)$$

Les phases $\varphi(\mathbf{H})$ des facteurs de structure sont une information critique mais qui ne peut pas être obtenue directement par l'expérience.

Or, les facteurs de structure complets sont nécessaires pour obtenir la distribution électronique dans la maille, celle-ci étant la transformée de Fourier des facteurs de structure.

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{H}} |F(\mathbf{H})| e^{i\varphi(\mathbf{H})} e^{-2\pi i \mathbf{H} \cdot \mathbf{r}} \quad (1.12)$$

Leur absence rend l'obtention de la distribution électronique d'un composé impossible à partir d'un jeu de données expérimentales. C'est pour cette raison qu'une détermination approximative des phases des facteurs de structures est nécessaire avant tout affinement.

Les méthodes directes [5;26] sont généralement utilisées pour obtenir une bonne approximation des phases de facteurs de structure. Dans le cas d'un affinement de macromolécule, le remplacement moléculaire ou isomorphe sont efficaces [24;2].

I.2. L'affinement cristallographique et les différents modèles couramment utilisés

Pour déduire des informations structurales à partir de données de diffraction des rayons X, il faut ajuster un modèle de distribution de charges contre ces données. Cette démarche est appelée affinement cristallographique.

I.2.1. Qu'est-ce qu'un affinement cristallographique ?

Comme expliqué dans la partie précédente, les données expérimentales collectées par diffraction des rayons X sont les intensités des faisceaux diffractés proportionnelles aux carrés des amplitudes des composantes de la transformée de Fourier de la distribution électronique de l'échantillon.

Dans une composante de Fourier d'un signal, l'amplitude correspond à l'importance de sa contribution au signal global et la phase à sa « localisation ». Cette localisation inconnue ne peut pas être prise en charge aisément dans la grande majorité des cas. Une stratégie d'affinement cristallographique est nécessaire.

Les données expérimentales sont des informations exprimées dans l'espace réciproque (ou également appelé espace dual ou de Fourier). L'objectif d'un affinement est de réduire la différence entre les amplitudes des composantes Fourier observées et calculées.

En pratique, la méthode de minimisation utilisée est le plus souvent la méthode des moindres carrés et la quantité à minimiser est :

$$\chi^2 = \sum_{\mathbf{H}} w_{\mathbf{H}} \left[\frac{1}{K} |F_{obs}(\mathbf{H})| - |F_{calc}(\mathbf{H})| \right]^2 \quad (1.13)$$

Le facteur de mise à l'échelle K permet quant à lui de mettre les amplitudes des facteurs de structure calculés et observés à la même échelle, ces derniers étant mesurés à une échelle arbitraire.

Le coefficient $w_{\mathbf{H}}$ est un terme de pondération défini pour chaque réflexion $|F_{obs}(\mathbf{H})|$. Son rôle est de prendre en compte la fiabilité (ou la qualité) de l'amplitude du facteur de structure expérimental. Pour expliquer les choses simplement, le schéma de pondération est choisi de

manière à rendre χ^2 plus sensible aux amplitudes des facteurs de structure bien mesurés en leur attribuant des coefficients de pondération plus larges.

Pour les affinements de petites molécules, le schéma de pondération standard est :

$$w_{\mathbf{H}} = \frac{1}{\sigma^2(F_{obs}(\mathbf{H}))} \quad (1.14)$$

Pour les affinements de macro-systèmes comme les protéines, les schémas de pondération peuvent être plus sophistiqués pour faciliter l'affinement du modèle. Le choix du schéma de pondération $w_{\mathbf{H}}$ n'est pas aisé.

Par exemple, dans le suite logiciel d'affinement cristallographique SHELX [27], le schéma de pondération, connu sous le nom de « robust-resistant weighting », inclut les valeurs de $|F_{obs}(\mathbf{H})|$ et $|F_{calc}(\mathbf{H})|$:

$$w_{\mathbf{H}} = \frac{1}{\sigma^2(F_{obs}(\mathbf{H})) + (aP)^2 + bP} \quad (1.15)$$

$$\text{avec } P = \frac{2F_{calc}(\mathbf{H}) + \text{Max}(F_{obs}(\mathbf{H})^2, 0)}{3} \quad (1.16)$$

De plus, la définition de la quantité χ^2 (1.13), peut présenter des termes additionnels qui permettent d'inclure des connaissances *a priori* sur le système étudié.

Ces termes supplémentaires peuvent être des restrictions de stéréochimie portant sur les distances interatomiques ou angles valencielles, la planéité des groupements chimiques, les volumes chiraux, mais aussi sur les paramètres de déplacements atomiques d'agitation thermique.

I.2.2. Les modèles de distribution de charges

Le modèle de distribution électronique utilisé durant un affinement dépend du type de système, de la résolution et de la qualité des données. Le modèle tente de s'ajuster contre les données expérimentales avec plus ou moins de degrés de liberté et donc plus ou moins de précision (ou détails).

L'on peut considérer deux modèles couramment utilisés :

- Le modèle d'atomes neutres sphériques et indépendants appelé communément I.A.M. (« Independent Atom Model ») qui est utilisé dans les affinements macromoléculaires ainsi que dans les affinements de petites molécules pour obtenir leurs structures.
- Le modèle multipolaire défini par Hansen & Coppens [15] et qui est la concrétisation de nombreux travaux menés notamment par Stewart [30] et Hirshfeld [17]. Ce modèle permet de modéliser l'asphéricité de la distribution électronique d'un atome due à son environnement chimique.

I.2.2.A. Le modèle «I.A.M.»

Le modèle «I.A.M.» suppose que l'environnement chimique n'a aucun effet sur la distribution électronique d'un atome. Cela revient à considérer d'un point de vue électronique que les atomes se comportent comme s'ils étaient indépendants ou isolés.

Par conséquent, la distribution de charges moléculaire sera définie comme la somme des distributions neutres sphériques atomiques :

$$\rho(\mathbf{r}) = \sum_{\text{atomes}} \rho_{i,\text{sph}}(\mathbf{r} - \mathbf{R}_i) \quad (1.17)$$

avec \mathbf{R}_i , le vecteur position du noyau du $i^{\text{ème}}$ atome.

Pour chaque atome, les paramètres relevant de la distribution de charges sphérique dépendent de l'espèce chimique.

Le modèle I.A.M. est largement utilisé dans les affinements structuraux en cristallographie et en biocristallographie car la résolution maximale atteinte et la qualité des données permettent rarement d'accéder aux détails subatomiques.

Néanmoins, les équipements utilisés en diffraction des rayons X et les techniques de cristallisation actuelles permettent dans le cas de petites molécules d'atteindre des résolutions et des qualités de données révélant des détails de la distribution électronique qui ne peuvent être gérés par le modèle I.A.M.

Le modèle I.A.M. ne dispose pas de suffisamment de degrés de liberté pour reproduire les détails de la distribution accessible à haute résolution ($> 0.7 \text{ \AA}^{-1}$). Concrètement, les liaisons chimiques entre les atomes altèrent leurs distributions électroniques qui deviennent

asphériques. Par définition, le modèle I.A.M. néglige les effets de l'environnement sur les distributions de charges atomiques.

Pour reproduire ces déformations, différents modèles ont été proposés, mais c'est sur le modèle de Hansen & Coppens [15] que nous nous attarderons ici. Ce modèle est celui utilisé dans l'ensemble des études décrites dans ce rapport.

I.2.2.B. Le modèle multipolaire de Hansen & Coppens

L'expansion de la distribution de densité de charges d'un atome en termes de somme de fonctions harmoniques sphériques réelles centrées sur le noyau atomique a été l'objet de nombreux travaux [10;9;19;22;17;16]. Le modèle multipolaire de Hansen & Coppens [30;31;32] se fonde sur le formalisme généralisé d'atome asphérique de Stewart.

Le modèle de Hansen & Coppens [15] est défini en partant du modèle I.A.M. et peut donc être perçu comme une extension de ce dernier. Comme dans le modèle I.A.M., la distribution électronique d'un atome est centrée sur son noyau. Elle peut être décrite comme la somme de trois termes : le terme à symétrie sphérique des électrons de cœur $\rho_{core,sph}(\mathbf{r})$, le terme à symétrie sphérique des électrons de valence $\rho_{val,sph}(\mathbf{r})$ et le terme à symétrie non-sphérique des électrons de valence $\rho_{val,mul}(\mathbf{r})$.

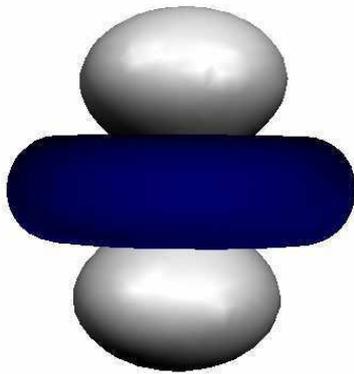
$$\rho_{electron}(\mathbf{r}) = \rho_{core,sph}(\mathbf{r}) + \rho_{val,sph}(\mathbf{r}) + \rho_{val,mul}(\mathbf{r}) \quad (1.18)$$

Le terme à symétrie sphérique des électrons de cœur $\rho_{core,sph}(\mathbf{r})$ est le même que dans le modèle I.A.M. Il est indéformable et a une charge électronique fixe.

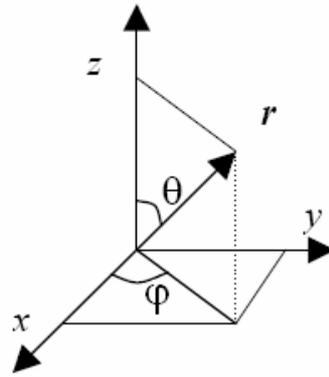
Le terme à symétrie sphérique des électrons de valence $\rho_{val,sph}(\mathbf{r})$ a été amélioré selon le formalisme Kappa proposé par Becker et Coppens [8]. Cette amélioration permet d'affiner une population de valence P_{val} mais aussi un coefficient d'expansion-contraction radiale k du nuage électronique de valence (expansion pour $k < 1$ et contraction pour $k > 1$). Suivant ce formalisme, la distribution de charges reste sphérique mais n'est plus nécessairement neutre.

$$\rho_{val,sph}(\mathbf{r}) = k^3 P_{val} \rho_{val,sph}(kr) \quad (1.19)$$

Dipôles

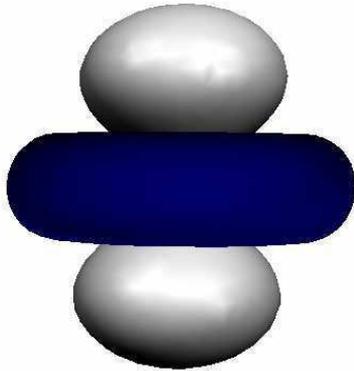


$$y_{10} = z$$

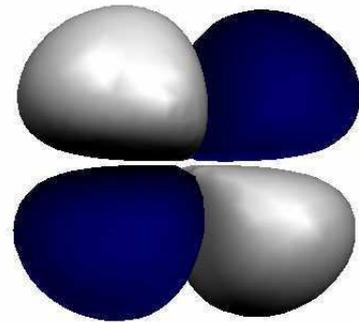


Définition du repère sphérique, centré sur l'atome considéré, sur lequel les harmoniques $y_{lm}(\theta, \phi)$ sont calculées.

Quadripôles

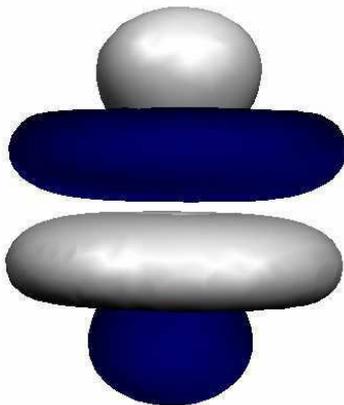


$$y_{20} = 2z^2 - (x^2 + y^2)$$

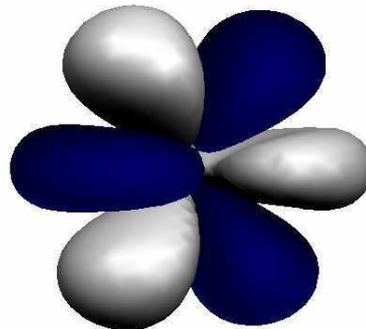


$$y_{21} = xz$$

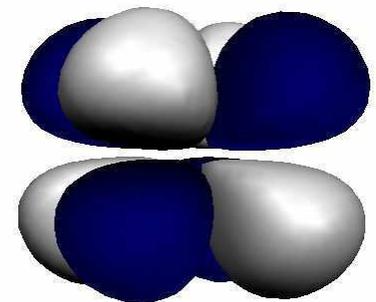
Octupôles



$$y_{30} = 2z^3 - 3z(x^2 + y^2)$$

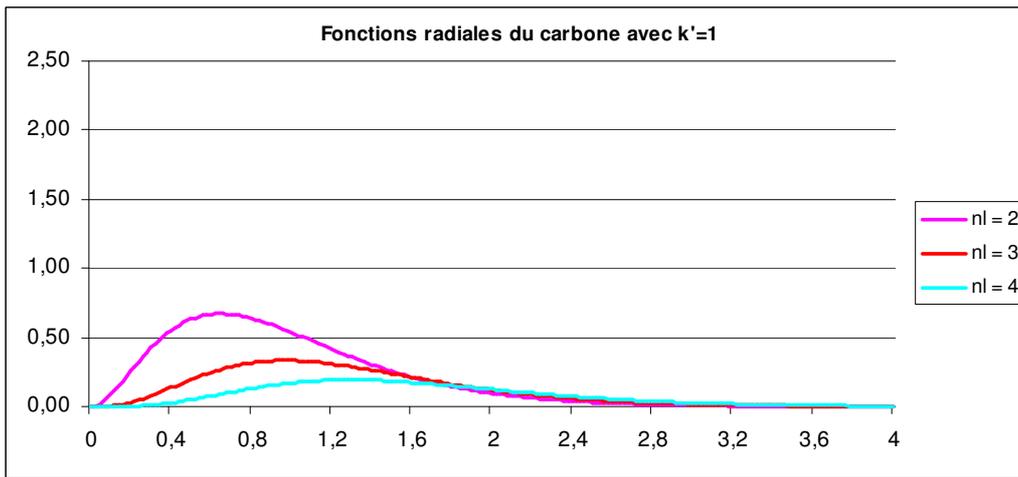


$$y_{31} = x[4z^2 - (x^2 + y^2)]$$

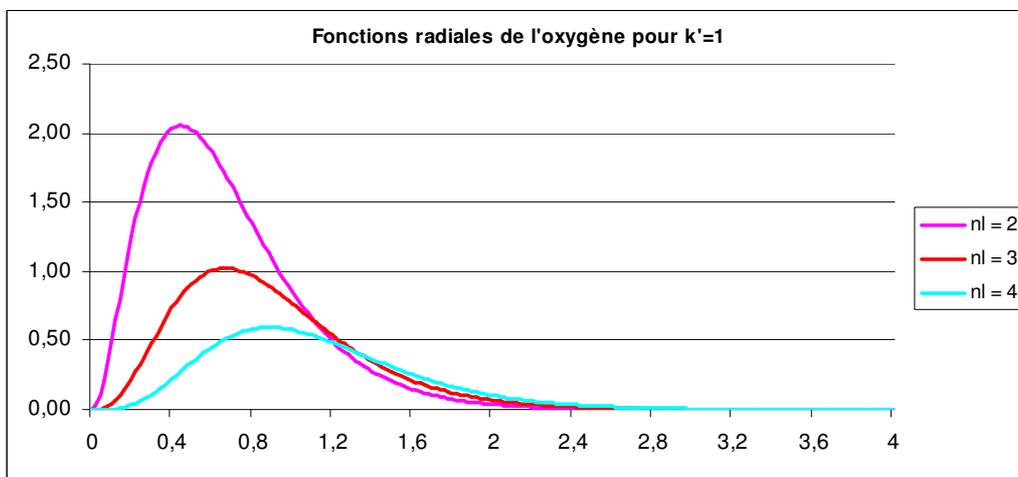


$$y_{32} = 2xyz$$

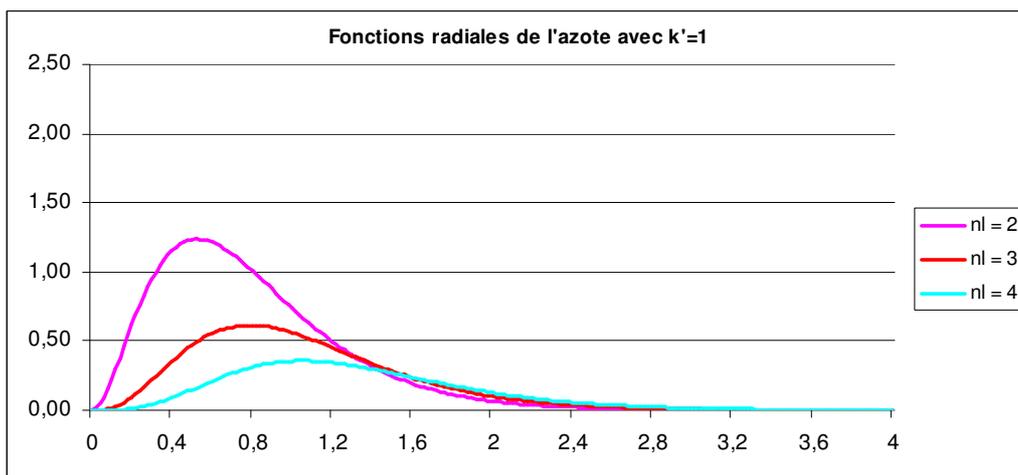
Figure 1 : Représentation de la dépendance angulaire des fonctions multipolaires sous forme de surfaces de densité électronique constante (lobes de densité positive en gris clair, négative en bleu). L'expression de cette dépendance est indiquée dans un repère cartésien.



(a)



(b)



(c)

Figure 2 : Courbes des fonctions radiales type Slater des atomes de carbone (a) ($\zeta=3,1$), d'oxygène (b) ($\zeta=4,5$) et d'azote (c) ($\zeta=3,8$) pour différentes valeurs de n_l .

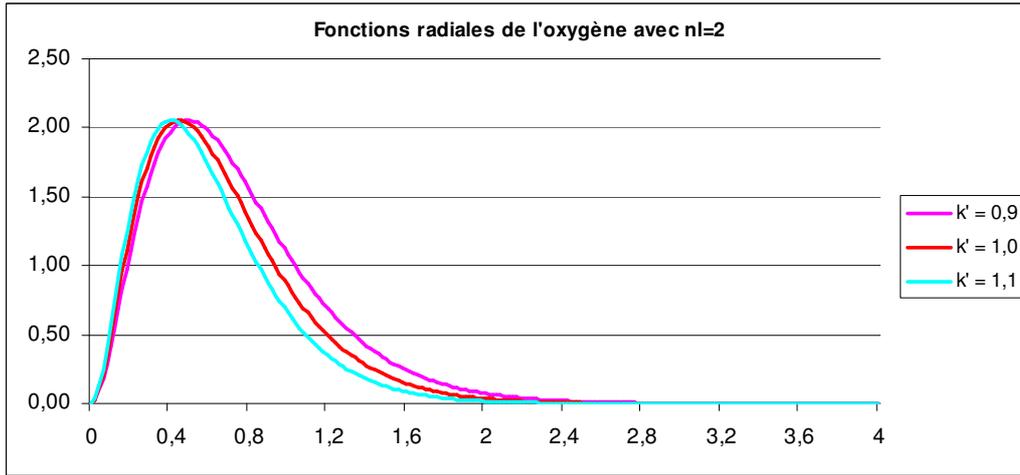


Figure 3 : Courbes des fonctions radiales type Slater de l'atome d'oxygène ($\zeta = 4,5$) avec $n_l = 2$ pour différentes valeurs de k' .

Les termes sphériques d'électrons de cœur et de valence peuvent être calculés à partir des fonctions d'onde de Clementi [7].

Finalement, le troisième terme $\rho_{val,mul}(\mathbf{r})$ est un terme de valence multipolaire et permet de prendre en compte la nature asphérique de la distribution d'électrons de valence.

$$\rho_{val,mul}(\mathbf{r}) = \sum_{l=0}^{l_{max}} k'^3 R_l(k'r) \sum_{m=0}^l P_{lm\pm} y_{lm\pm}(\theta, \varphi) \quad (1.20)$$

$y_{lm\pm}(\theta, \varphi)$ sont des fonctions angulaires harmoniques sphériques réelles normalisées et orientées selon le repère local de l'atome (**Figure 1**). Pour chaque fonction angulaire harmonique sphérique réelle normalisée, un facteur multiplicatif P_{lm} , dit de population électronique de valence multipolaire, peut être affiné. Plus ce terme diffère de zéro, plus la contribution de sa fonction angulaire à l'asphéricité de la distribution électronique est importante. Le terme k' apparaissant dans l'expression est un second coefficient d'expansion-contraction.

$R_l(k'r)$ est une fonction radiale en général de type Slater. Si l'on introduit n_l et ζ des paramètres orbitaux définis pour chaque type atomique qui affectent le profil de la fonction radiale et tout particulièrement la localisation de son maximum (**Figure 2**), et k' le coefficient d'expansion-contraction coefficient (**Figure 3**), on obtient :

$$R_l(k'r) = \frac{\zeta^{n_l+3}}{(n_l+2)!} k'^{n_l} r^{n_l} \exp(-\zeta k'r) \quad (1.21)$$

La distribution moléculaire de charges $\rho_{molécule}$ est dans le formalisme de Hansen & Coppens une décomposition de type « fuzzy » ou floue car en tous points de l'espace, la densité de charges est la somme de plusieurs contributions atomiques.

$$\rho_{molécule}(\mathbf{r}) = \sum_{\text{atomes}} \rho_{i,multipolaire}(\mathbf{r} - \mathbf{R}_i) \quad (1.22)$$

avec \mathbf{R}_i , le vecteur position du noyau du $i^{\text{ème}}$ atome de la molécule.

On parle généralement de décomposition de charges en pseudo-atomes afin d'éviter tout malentendu et confusion avec d'autres formalismes de décompositions de charges. Chaque distribution de charges pseudo-atomique est ainsi décrite comme deux termes sphériques et un terme asphérique qui est la somme de plusieurs termes multipolaires. L'ordre de développement de ce troisième terme est fini et dépend de l'espèce chimique de l'atome considéré, de son environnement et de la qualité des données.

A chaque ordre correspondent plusieurs fonctions multipolaires qui sont définies dans un repère local centré sur le noyau de l'atome :

- L'ordre 1 est représenté par trois fonctions dipôlares,
- L'ordre 2 par cinq quadripôles,
- L'ordre 3 par sept octupôles,
- L'ordre 4 par neuf hexadécapôles,
- L'ordre 5 par onze fonctions 32-pôles.

On utilise généralement un développement jusqu'à l'ordre 3 pour les atomes de carbone, azote, oxygène (électrons 2s, 2p), pour les atomes dits lourds tels que le soufre, le phosphore ou les métaux de transition, un développement jusqu'à l'ordre 4, voire supérieur (électrons f). L'ordre de développement utilisé pour les atomes d'hydrogène est fonction de la qualité des données et des outils utilisés. Ce point sera discuté ultérieurement.

Le modèle le plus simple utilisé par défaut pour les pseudo-atomes d'hydrogène est un unique dipôle orienté (développement d'ordre 1) selon la direction de la liaison covalente H–X qui les lie à leurs atomes X porteurs également appelé : atomes donneurs.

Le niveau minimum de développement pour les atomes d'hydrogène peut s'expliquer par l'incapacité de la diffraction des rayons X de fournir suffisamment d'information précise sur leur localisation dans la structure.

Comme cela a été décrit précédemment, les fonctions harmoniques sphériques réelles sont définies dans le repère local de l'atome. Un choix judicieux de repère atomique peut améliorer l'ajustement du modèle contre les données expérimentales et ainsi améliorer le résultat final de l'affinement [11]. Les repères locaux ne sont généralement pas définis de façon absolue dans le repère de la maille mais en fonction d'atomes voisins.

L'idéal est de définir des axes locaux atomiques qui respectent les symétries naturelles de la distribution de charges (la stéréochimie). De bons axes locaux permettent de garder certaines populations multipolaires nulles pendant les premières étapes de l'affinement ce qui réduit le nombre de variables à affiner et par la même les corrélations entre elles.

De plus, si, compte tenu l'environnement, une distribution électronique symétrique en première approximation est attendue, des axes locaux bien choisis permettent de savoir quantitativement à quel point cette symétrie est brisée. Une symétrie imparfaite ou « brisée » va induire des valeurs non nulles pour des populations multipolaires qui ne devraient pas idéalement contribuer à la déformation de valence.

I.2.3. Stratégies d'affinement cristallographique

I.2.3.A. L'objectif d'un affinement

Comme cela a déjà été expliqué, l'objectif d'un affinement est de réduire au maximum les différences entre les amplitudes des facteurs de structure observés et des facteurs de structure calculés correspondants.

Dans le cas d'une minimisation par moindres carrés, quel que soit le type de modèle utilisé, on cherche à minimiser la fonction suivante :

$$\chi^2 = \sum_{\mathbf{H}} w_{\mathbf{H}} \left[\frac{1}{K} |F_{obs}(\mathbf{H})| - |F_{calc}(\mathbf{H})| \right]^2 + w_{restriction} \sum_i [d_{calculée,i} - d_{référence,i}]^2 + \dots \quad (1.23)$$

Bien que le problème d'un affinement de structure consiste en la minimisation d'une fonction, il ne se résume pas à un challenge mathématique. La difficulté d'un affinement est certes de minimiser la fonction χ^2 (1.23) mais surtout de guider le modèle vers un état final qui sera réaliste d'un point de vue physico-chimique.

Certaines variables des modèles sont plus ou moins corrélées entre elles. Par exemple, dans un affinement multipolaire, les paramètres d'agitation thermique et les populations des fonctions quadripolaires sont corrélés.

Affiner correctement un modèle nécessite une bonne stratégie afin d'éviter d'aboutir à un modèle irréaliste ou un faux minimum pour la fonction χ^2 .

Au début de l'affinement, l'état initial du modèle est éloigné de l'état final attendu. L'utilisation de contraintes et de restrictions décroît le nombre de variables affinées et en réduit la gamme de valeurs. Par exemple, une distance interatomique $d_{calculée}$ peut être restreinte à une valeur de référence $d_{référence}$. Le terme de restriction géométrique sur les distances est affecté d'un poids $w_{restriction}$ optimisé.

Le recours aux contraintes et restrictions simplifie le « paysage » de la fonction χ^2 qui dépend de tous les paramètres du modèle (via les facteurs $F_{calc}(\mathbf{H})$) et facilite la convergence vers le minimum global de χ^2 .

I.2.3.B. Contraintes et restrictions

Une grande diversité de contraintes existe et peut être employée autant dans les affinements de petites molécules que de systèmes macromoléculaires. Les atomes d'hydrogène sont toujours contraints en raison des limites expérimentales imposées par la diffraction des rayons X.

Dans le cas des affinements de grands systèmes comme les protéines, le nombre de restrictions utilisées est généralement très important. En raison de la qualité des jeux de données, d'une amplitude plus importante des déplacements atomiques due à l'agitation thermique et de la taille des systèmes, un affinement sans restrictions sur les atomes lourds est très rare. La minimisation de la fonction χ^2 est réalisée récursivement par la résolution d'un

système matriciel d'équations qui nécessite un nombre d'observations nettement supérieur au nombre de paramètres de l'équation.

Pour pallier ce manque d'information, on utilise usuellement du début à la fin de l'affinement des restrictions sur les distances interatomiques, sur les angles valenciels (angle que forment trois atomes liés $X - Y - Z$), sur la planéité des cycles, sur les paramètres d'agitation thermique afin de guider le mieux possible l'affinement du modèle.

Lors d'un affinement multipolaire, la qualité des données permet de placer les atomes lourds sans restrictions sur les distances ou les angles. Néanmoins, des contraintes sont utilisées pour réduire le nombre de paramètres au début de l'affinement. L'on peut avoir recours à des contraintes d'équivalence entre des atomes présentant des environnements covalents similaires. Les contraintes d'équivalence chimique forcent des atomes de nature et d'environnement identiques à partager les mêmes paramètres de densité de charges (populations de valence, populations multipolaires et paramètres kappa). Cela n'est envisageable que si les repères locaux atomiques sont également équivalents.

Il est possible aussi d'utiliser des contraintes cristallographiques ainsi que de symétrie locale comme exposé précédemment.

I.2.3.C. Affinement d'un modèle multipolaire

Un affinement multipolaire n'est envisageable qu'à la condition nécessaire (mais pas toujours suffisante) de disposer d'un jeu de très bonnes données de résolution meilleure que 0,7 Å.

Pour comprendre cette condition, il faut garder à l'esprit qu'un modèle multipolaire dispose d'un grand nombre de variables pour reproduire l'asphéricité de la distribution électronique de valence.

D'un point de vue algorithmique, un critère empirique permet de s'assurer qu'un modèle est adapté au jeu de données utilisé. Ce critère consiste à vérifier qu'à chaque étape de l'affinement le rapport du nombre de réflexions utilisées et du nombre de variables affinées est supérieur à 10. Pour permettre l'affinement d'un modèle multipolaire, une meilleure résolution est nécessaire afin de disposer de plus d'observations.

Cette condition a également une justification physique. Les données obtenues par diffraction des rayons X, sont proportionnelles aux carrés des amplitudes des facteurs de structure dynamiques qui sont affectés par l'agitation thermique atomique. Même à très basse

température (hélium), une agitation thermique d'amplitude moindre persistera. Pour permettre l'affinement du modèle multipolaire de distribution de charges, les déplacements atomiques liés à l'agitation thermique (modélisé à l'aide des paramètres de déplacements anisotropiques U_{ij}) et la distribution électronique de valence doivent être décorrélés. Les illustrations des facteurs de diffusions atomiques en fonction de la résolution (**Figure 4**) montrent que, pour un atome non hydrogène, l'amplitude des facteurs de diffusion de valence diminue jusqu'à devenir négligeable, au-delà de $0,7 \text{ \AA}^{-1}$ dans l'espace réciproque (soit moins de $0,7 \text{ \AA}$ dans l'espace direct), par rapport aux amplitudes des facteurs de diffusion de cœur.

Pour cette raison, ce n'est qu'à haute résolution que les paramètres de positions atomiques et d'agitations thermiques pourront être affinées indépendamment des variables de la distribution de charges.

Il n'existe pas de stratégie idéale d'affinement cristallographique à haute résolution, une stratégie relève avant toute chose de l'expertise et du bon sens du cristallographe. Les différentes étapes définies pour l'affinement ainsi que leur ordre influenceront le modèle final. L'on peut cependant dégager plusieurs grandes étapes-clés d'un affinement.

Une étape d'affinement consiste à affiner cycliquement un sous-ensemble de variables du modèle jusqu'à convergence. On considère qu'il y a convergence lorsque les variations (ou « shift ») des variables ne sont plus significatives par rapport aux incertitudes correspondantes. Généralement, on considère un rapport variation sur écart-type (Δ/σ) inférieur à 0,1 comme critère satisfaisant.

Un modèle multipolaire de distribution de charges peut être vu comme deux sous-modèles complémentaires : un modèle atomique qui localise les atomes dans l'espace et reproduit leurs agitations thermiques, et un modèle de distribution de charges qui s'appuie sur le modèle atomique et qui, comme son nom l'indique, reproduit la distribution de charges asphérique de chaque atome.

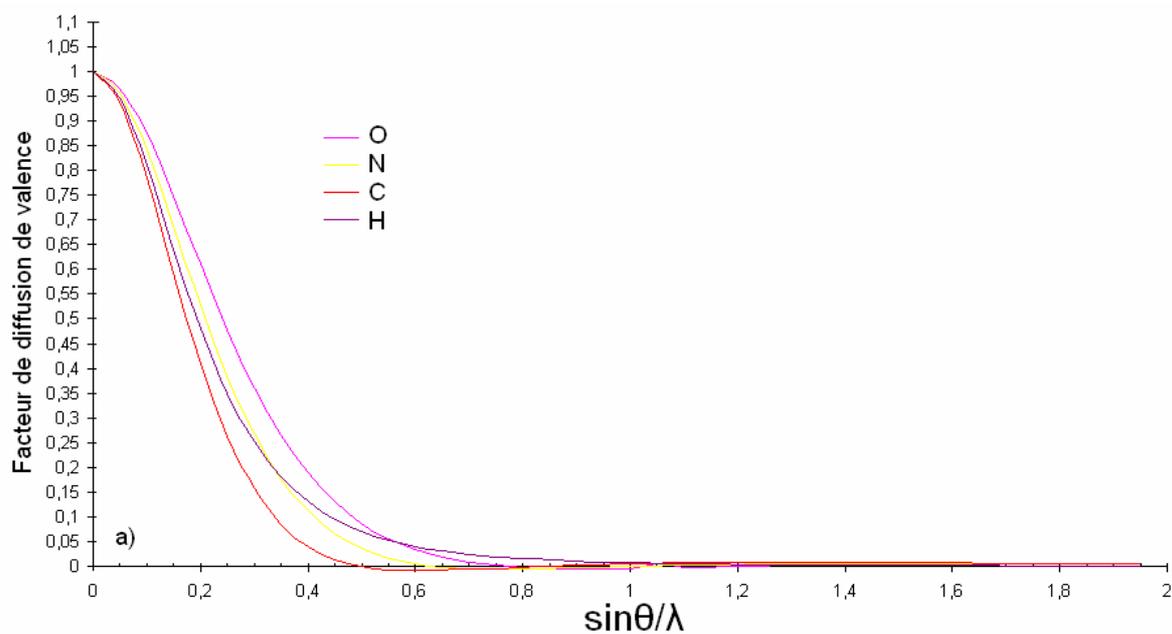
La toute première étape indispensable à tout affinement est la mise à l'échelle des données expérimentales par rapport aux facteurs de structure calculés. En effet, les intensités mesurées après corrections sont proportionnelles aux carrés des amplitudes des facteurs de structure dynamiques expérimentaux. On ajuste le paramètre K appelé facteur de mise à l'échelle (déjà défini dans la partie I.2.1 de ce chapitre) de manière à mettre à la même échelle les amplitudes observées et calculées des facteurs de structure.

L'étape suivante d'un affinement est de décorrélérer ces deux sous-modèles afin de modéliser au mieux l'agitation thermique et éviter ainsi qu'elle ne nuise à la représentation de la distribution de charges. Pour cela, on affine les variables de positions et de déplacements atomiques dus à l'agitation thermique des atomes lourds non hydrogène à haute résolution ($\sin\theta/\lambda > 0,7 \text{ \AA}^{-1}$) ou « haut angle » (**Figure 4**). Cette étape est d'ailleurs appelée affinement « haut angle » ou « grand angle ». Quant aux paramètres des atomes d'hydrogène, ils sont affinés à « bas angle » ($\sin\theta/\lambda < 0,7 \text{ \AA}^{-1}$).

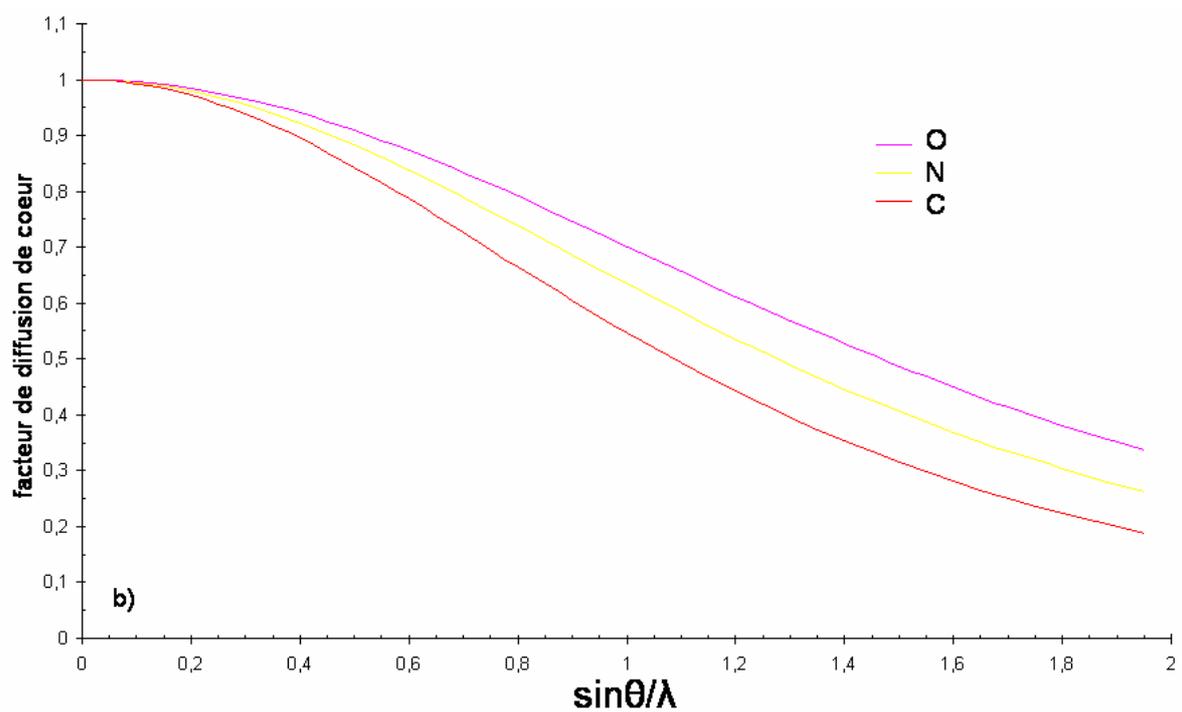
Par la suite, plusieurs étapes consistent à affiner les variables du sous-modèle de distribution de charges contre toute la gamme de résolution.

Avant tout affinement des paramètres de distribution électronique de valence, il est important de choisir judicieusement les repères locaux des atomes, ainsi que l'ordre de développement multipolaire de chacun d'entre eux (ce point a déjà été exposé dans la partie I.2.2.B) et leurs paramètres radiaux.

L'on commence par affiner ces paramètres en imposant un maximum de contraintes d'équivalence et de symétrie, afin de minimiser le nombre de variables et se rapprocher ainsi d'un modèle final satisfaisant. Une fois les différentes étapes de l'affinement du sous-modèle de distribution électronique terminées, l'on réaffine selon ces mêmes étapes mais en supprimant successivement certaines contraintes d'équivalence et de symétrie.



(a)



(b)

Figure 4 : Facteurs de diffusion (en électrons) normalisés à un électron, des électrons de valence (a) et de coeur (b) pour les espèces chimiques C, O, N et H en fonction de la résolution réciproque (\AA^{-1}).

L'on peut lever ces contraintes selon l'ordre suivant :

- Supprimer les équivalences sur les atomes lourds afin de prendre en compte leurs différences d'environnement covalent qui avaient été considérées comme négligeables précédemment.
- Supprimer les contraintes de symétrie sur ces mêmes atomes, afin de finaliser l'affinement en prenant en compte les effets de l'environnement non covalent.
- Les équivalences entre atomes d'hydrogène peuvent être supprimées à la fin, ces atomes étant plus délicats à affiner.

Lorsqu'un maximum de contraintes a été supprimé sans nuire à la qualité du modèle, l'on affine les positions atomiques xyz et les paramètres d'agitation thermique contre toute la gamme de résolution. Cette étape est nécessaire car si la distribution électronique de valence ne contribue pas aux facteurs de diffusion atomique à « haut angle », l'agitation thermique affecte les valeurs des facteurs de diffusion atomiques sur toute la gamme de résolution (Figure 5).

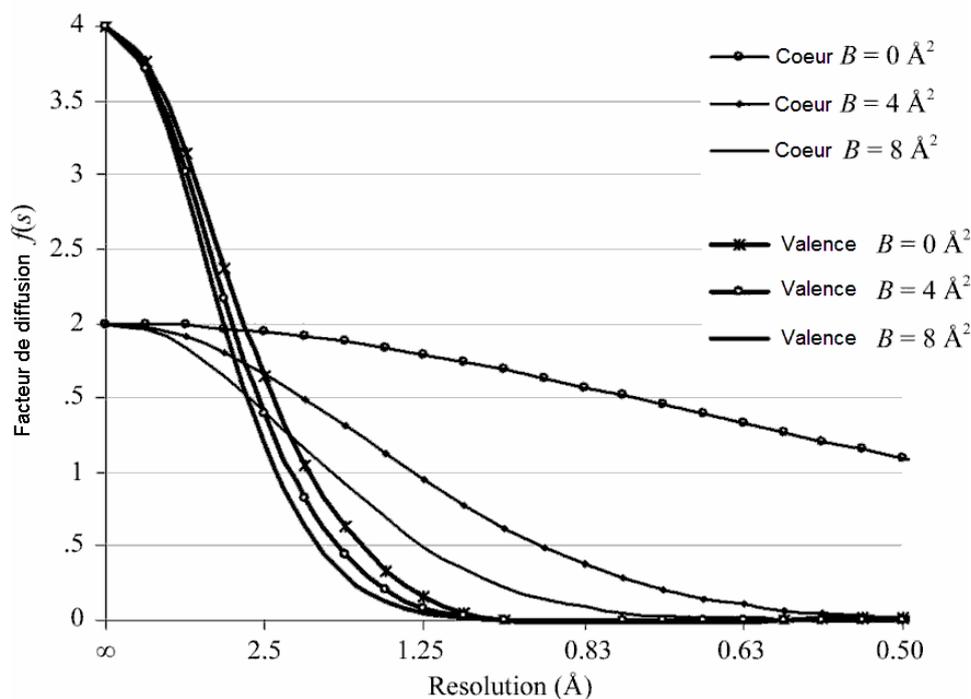


Figure 5 : Facteurs de diffusion des électrons de cœur et de valence en fonction de la résolution pour l'atome de carbone avec un facteur de Debye-Waller isotrope B_{eq} de 0, 4 et 8 \AA^2 .

La toute dernière étape de l'affinement consistera à affiner les paramètres atomiques (xyz et U_{ij}) et les paramètres de la distribution électronique de valence (P_{val} , P_{lmp} , k et k') ensemble jusqu'à convergence afin de profiter de tous les degrés de liberté du modèle et obtenir ainsi un ajustement optimal contre les données expérimentales.

I.2.3.D. Modélisation des atomes d'hydrogène

Comme cela a été expliqué, la clé de la diffraction des rayons X est l'interaction rayon X - électron. L'atome d'hydrogène est l'atome le plus léger avec seulement un électron de valence et aucun électron de cœur.

La présence d'électrons de cœur est essentielle pour localiser les atomes : les électrons de cœur restent au voisinage du noyau de l'atome selon une distribution assumée sphérique ce qui permet de confondre le centre de cette distribution avec la position de ce dernier. Même avec des données de diffraction X d'excellente qualité, les atomes d'hydrogène ne peuvent être localisés par affinement. Si un affinement des positions des atomes d'hydrogène d'un modèle I.A.M. est réalisé, ils seront supposés localisés au maximum de densité électronique le long de la liaison covalente avec l'atome qui les porte (leur atome « Donneur ») ce qui conduira à une liaison covalente irréaliste (longueur de liaison sous-estimée).

Afin de conduire l'affinement vers un modèle final réaliste, les positions des atomes d'hydrogène sont, en général, contraintes de manière à imposer la longueur de leur liaison covalente.

Les valeurs de longueur de liaisons X—H utilisées sont généralement issues de modèles atomiques obtenus par diffraction des neutrons. Les neutrons, à la différence des rayons X, interagissent avec les noyaux des atomes ce qui permet de localiser avec précision les atomes d'hydrogène comme les atomes lourds. Les longueurs standard des liaisons covalentes des atomes d'hydrogène sont tabulées dans les tables internationales de cristallographie [1].

De plus, la distribution électronique de l'hydrogène étant non sphérique, les paramètres d'agitation thermique des atomes hydrogènes ne peuvent être affinés et sont, par conséquent, assumés isotropes et contraints à une relation de proportionnalité avec les paramètres d'agitation thermique isotropes moyens de leurs atomes « donneur ». Cette contrainte sur le paramètre d'agitation thermique de l'hydrogène est appelée « Riding Model ».

Ces dernières années, le modèle utilisé pour modéliser les atomes d'hydrogène fût sujet à débat [23]. Le rôle des atomes d'hydrogène dans les interactions intermoléculaires ne doit en aucun cas être mésestimé, et ce tout particulièrement dans le cas des molécules biologiques et de leurs mécanismes/fonctions dans les voies métaboliques. Ces molécules dans un environnement autant cristallin que *in vivo* sont impliquées dans un réseau complexe de liaisons hydrogène. Les affinements à haute résolution de modèles multipolaires permettent de déduire de la distribution de charges des informations sur leurs propriétés électrostatiques. La fiabilité et la précision de toute information peuvent être biaisées par un modèle des atomes d'hydrogène erroné [12;13;14;28;29;6;33].

Lorsque la qualité des cristaux et les moyens techniques le permettent, une étude peut consister à réaliser une expérience de diffraction des rayons X ainsi qu'une expérience de diffraction neutronique. Les positions des atomes d'hydrogène obtenus à partir du jeu de données neutroniques servent alors à la définition des contraintes sur les atomes d'hydrogène utilisées pour l'affinement du modèle de distribution de charges contre les données RX.

Malheureusement, une telle approche est exigeante (utilisation de grands instruments et cristaux de qualité et de taille appropriées) et ne peut donc être employée que rarement. De plus, disposer de jeux de données rayons X et neutroniques ne constitue en rien une garantie de succès.

Différentes méthodes récentes proposent d'estimer des paramètres d'agitation thermique anisotrope pour les atomes hydrogènes. Ces méthodes peuvent reposer sur les calculs *in silico* (ONIOM) [34] ou sur des données expérimentales tabulées (SHADE) [20;21]. Ici, je me contenterai de présenter SHADE qui a été utilisé dans l'étude du composé fidarestat. SHADE [21] n'est pas à proprement parlé un programme mais un serveur Web qui offre la possibilité d'estimer les paramètres d'agitation thermique anisotrope des atomes d'hydrogène d'un composé à partir des positions atomiques et des paramètres d'agitation thermique anisotrope des atomes lourds.

Cette méthode repose sur l'hypothèse que les déplacements quadratiques d'agitation thermique des atomes d'hydrogène peuvent être décomposés en deux contributions supposées décorrélées : une contribution dite « externe » U_{Rigide}^{ij} liée au mouvement de corps rigide (rigid-body motion) du groupement chimique rigide considéré et une contribution dite « interne » $U_{Interne}^{ij}$ due aux vibrations propres de ses atomes d'hydrogène.

$$U^{ij} = U_{Rigide}^{ij} + U_{Interne}^{ij} \quad (1.24)$$

La contribution « externe » correspond à l'agitation thermique d'ensemble du groupement chimique dans le repère global auquel appartient l'atome d'hydrogène. Les déplacements quadratiques d'agitation thermique de l'atome d'hydrogène appartenant à ce groupement chimique supposé rigide sont obtenus par analyse T.L.S. (Translation-Libration-Screw) réalisée à l'aide du programme THMA11 [25].

La contribution « interne » correspond à l'agitation thermique de l'atome d'hydrogène dans le repère local du groupement chimique. Cette contribution « interne » est la somme des contributions des différents modes vibrationnels internes du groupement chimique.

Elle est estimée à partir de déplacements quadratiques internes d'atomes d'hydrogène appartenant à différents groupements chimiques usuels. Ces données ont été dérivées d'études par diffraction neutronique de différents composés organiques et ont été tabulées.

I.2.4. Critères de validation des affinements cristallographiques

Tout au long d'un affinement, différentes quantités peuvent être calculées pour apprécier la qualité et l'avancement de l'affinement d'un modèle. Il est possible de déterminer les valeurs de différents indicateurs de qualité mais aussi de vérifier localement à l'aide de représentations de la densité électronique résiduelle le bon déroulement de l'affinement.

I.2.4.A. Indicateurs globaux de qualité ou facteurs d'accord

Ces indicateurs sont des fonctions qui fournissent une appréciation globale de l'affinement. Plus petites sont leurs valeurs, plus le modèle est apte à reproduire les amplitudes des facteurs de structure mesurés.

Les facteurs d'accord courants sont :

- Le facteur R_1 qui permet de juger quantitativement l'ajustement du modèle contre les données expérimentales,

$$R_1 = \frac{\sum_{\mathbf{H}} \left| \frac{1}{K} |F_{obs}(\mathbf{H})| - |F_{calc}(\mathbf{H})| \right|}{\sum_{\mathbf{H}} \frac{1}{K} |F_{obs}(\mathbf{H})|} \quad (1.25)$$

Certains programmes tels que SHELXL proposent de calculer le facteur R_2 défini comme :

$$R_2 = \frac{\sum_{\mathbf{H}} \left| \frac{1}{K} |F_{obs}(\mathbf{H})|^2 - |F_{calc}(\mathbf{H})|^2 \right|}{\sum_{\mathbf{H}} \frac{1}{K} |F_{obs}(\mathbf{H})|^2} \quad (1.26)$$

- Le facteur R_w est un indicateur pondéré qui peut être vu comme une normalisation de la fonction χ^2 minimisée au cours de l'affinement. Ainsi, cette quantité permet de suivre la minimisation de χ^2 .

$$R_w = \sqrt{\frac{\sum_{\mathbf{H}} w_{\mathbf{H}} \left| \frac{1}{K} |F_{obs}(\mathbf{H})| - |F_{calc}(\mathbf{H})| \right|^2}{\sum_{\mathbf{H}} \frac{1}{K^2} w_{\mathbf{H}} |F_{obs}(\mathbf{H})|^2}} = \frac{\chi^2}{\sqrt{\sum_{\mathbf{H}} \frac{1}{K^2} w_{\mathbf{H}} |F_{obs}(\mathbf{H})|^2}} \quad (1.27)$$

Remarque: Un désaccord au cours de l'affinement entre les facteurs R_1 et R_w (l'un augmente significativement pendant que l'autre diminue) peut être révélateur : soit la présence de réflexions erronées ou « outliers » qui vont nuire à la convergence du modèle vers un état satisfaisant, soit une mésestimation des erreurs sur les intensités ou un schéma de pondération inapproprié.

- L'indicateur *gof* pour « goodness of fit » (ou « test de validité de l'ajustement ») permet d'apprécier la qualité de l'ajustement du modèle contre les observations expérimentales.

Une valeur élevée du gof implique que le modèle reproduit mal les observations expérimentales. Dans l'idéal, l'indicateur gof est égal à 1 ce qui signifie que les écarts entre les observations et les estimations du modèle sont en accord avec les incertitudes expérimentales sur les observations. Une valeur de gof inférieure à 1 peut signifier qu'il y a suraffinement (ou surparamétrisation) du modèle, ou encore que les incertitudes expérimentales sur les observations ont été surestimées.

$$gof = \sqrt{\frac{\sum_{\mathbf{H}} w_{\mathbf{H}} \left| \frac{1}{K} |F_{obs}(\mathbf{H})| - |F_{calc}(\mathbf{H})| \right|^2}{N_{obs} - N_{var}}} \quad (1.28)$$

avec N_{obs} , le nombre d'observations utilisées et N_{var} , le nombre de variables du modèle affinées.

Pour les affinements de système large comme les protéines, il est également possible de définir les facteurs d'accord libres [3]. Cette méthode consiste à exclure délibérément un petit pourcentage (de l'ordre de 5 %) de données choisi aléatoirement et de le garder pour calculer des indicateurs de qualité R_1 et R_w supplémentaires appelés « indicateurs de qualité libres » ou « facteurs d'accord libres », et communément notés R_{1free} et R_{wfree}

Les données obtenues par diffraction des rayons X pour les systèmes macromoléculaires sont souvent de résolution moyenne et de moindre qualité en raison de la taille du système, d'une agitation thermique et d'un bruit expérimental plus élevés. En conséquence, pendant les dernières étapes de l'affinement, l'algorithme peut tenter d'améliorer le modèle alors que l'écart entre ce dernier et les données expérimentales est essentiellement lié à la contribution du bruit. On parle alors de « suraffinement ». Contrôler l'évolution des facteurs d'accord libres permet de détecter une surestimation de la qualité des données.

Ces facteurs d'accord ou indices de qualité libres ne renseignent pas sur l'efficacité de la stratégie mais sur l'aptitude du modèle à reproduire des données expérimentales pour lesquelles il n'a pas de connaissances au préalable. En résumé, suivre ces facteurs libres au cours de l'affinement permet de réaliser un « quiz cristallographique » à l'aide du modèle (le programme tentant de retrouver à partir du modèle affiné les valeurs de facteurs de structure mesurés mais non utilisés pendant l'affinement).

Pendant que l'algorithme ajuste le modèle contre des informations « utiles » sur la distribution électronique, les facteurs de qualité libres doivent diminuer ou rester stables. Si les valeurs des facteurs de qualité libres augmentent significativement, cela indique que les modifications ont tendance à dégrader le modèle en raison du bruit en nuisant au phasage. Cependant, ces facteurs de qualité présentent des limites en tant qu'indicateurs globaux : même si leurs valeurs augmentent, certaines améliorations locales sont toujours possibles. De plus, le choix aléatoire d'un petit nombre de réflexions pour le calcul des indices de qualité libres induit une grande incertitude sur ces indices. Il est possible de tenter de compenser cette incertitude en réalisant plusieurs affinements avec pour chacun d'entre eux un sous-ensemble « libre » de réflexions différents et de considérer la moyenne des différentes valeurs des facteurs de qualité libres [3;4].

I.2.4.B. Vérification locale de la qualité de l'affinement

Les facteurs d'accord permettent d'apprécier globalement l'évolution d'un modèle à chaque étape de son affinement. Ils ne peuvent donc pas suffire à la validation d'un modèle composé de nombreuses variables comme un modèle multipolaire.

En raison des possibles corrélations entre les paramètres du modèle, les valeurs acceptables d'indicateurs de qualité peuvent correspondre à des modèles irréalistes. Pour suivre l'amélioration d'un modèle asphérique de densité de charges tout au long de son affinement, les variables du modèle peuvent être contrôlées directement afin d'identifier toute valeur étrange voire invraisemblable. Le modèle restant un outil de modélisation de la distribution de charges, il est fondamental de calculer différentes cartes de densité électronique en 2 ou 3 dimensions au voisinage des groupements chimiques afin de se concentrer sur l'information qui nous intéresse : la distribution de la densité électronique.

Les différents types de cartes sont :

- Les cartes de densité électronique résiduelle

Les cartes résiduelles sont un critère de validation classique qui permet de savoir si le modèle reproduit les données expérimentales.

Elles sont la série des différences Fourier entre les facteurs de structure observés et calculés.

$$\Delta\rho_{res}(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{H}} \left(\frac{1}{K} |F_{obs}(\mathbf{H})| - |F_{calc}(\mathbf{H})| \right) e^{i\varphi_{calc}(\mathbf{H})} e^{-2\pi\mathbf{H}\cdot\mathbf{r}} \quad (1.29)$$

avec V , le volume de la maille et K , le facteur de mise à l'échelle.

Les phases des facteurs de structure expérimentaux étant inconnues, elles sont supposées égales aux phases des facteurs de structures calculés correspondants $\varphi_{calc}(\mathbf{H})$.

Ces cartes montrent des régions d'accumulations ou de déplétions de densité électronique qui mettent en évidence les désaccords entre les données expérimentales et le modèle affiné. Une accumulation de densité électronique correspond à de l'information sur la distribution de densité électronique que le modèle n'est pas en mesure de reproduire correctement. A contrario, une déplétion de densité électronique correspond à de l'information présente dans le modèle de distribution de charges mais qui ne correspond pas aux données expérimentales.

Dans un cas idéal d'affinement, les cartes résiduelles doivent être vierges de tous pics résiduels à la fin de l'affinement. Dans la pratique, en raison du bruit et de l'incertitude sur les amplitudes des facteurs de structure expérimentaux, et de la troncature de la série des différences Fourier, des pics résiduels persistent dans les cartes résiduelles. Si ces pics sont inférieurs à $0,05 \text{ e}\cdot\text{\AA}^{-3}$, ils sont considérés comme négligeables car associés au bruit.

Ces cartes permettent d'observer comment le modèle reproduit les données expérimentales sans prendre en compte la contribution véritable de chacune d'entre elles à l'affinement du modèle puis que le schéma de pondération des réflexions n'apparaît pas dans la formule. C'est pour cela que les réflexions de rapport I/σ petit sont souvent exclues de la série de différences Fourier de manière à supprimer les pics résiduels engendrés par les réflexions de faible poids dans la fonction χ^2 .

- Les cartes de déformation de densité électronique expérimentale

Ces cartes sont la série des différences Fourier entre les facteurs de structures expérimentaux et calculés à l'aide du modèle I.A.M, soit :

$$\Delta\rho_{\text{exp}}(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{H}} \left(\frac{1}{K} |F_{\text{obs}}(\mathbf{H})| e^{i\varphi_{\text{mul,calc}}(\mathbf{H})} - |F_{\text{sph,calc}}(\mathbf{H})| e^{i\varphi_{\text{sph,calc}}(\mathbf{H})} \right) e^{-2\pi\mathbf{H}\cdot\mathbf{r}} \quad (1.30)$$

Comme pour les cartes résiduelles, les phases des facteurs de structure observés sont considérées égales aux phases des facteurs calculés à l'aide du modèle multipolaire $\varphi_{\text{mul,calc}}(\mathbf{H})$.

Dans les affinements à haute résolution, on cherche à ajuster un modèle multipolaire afin qu'il reproduise l'asphéricité de la distribution électronique de valence. Il peut être intéressant à la fin de l'affinement « haut angle » de calculer des cartes de déformation de densité électronique de valence expérimentale qui mettent en évidence l'information qui relève de la distribution. Cela permet d'apprécier la qualité du jeu de données utilisé.

- Les cartes de déformation de densité électronique calculée

Leur expression est similaire à celle précédemment posée. Ces cartes sont la série des différences Fourier entre les facteurs de structures calculés à l'aide du modèle multipolaire et ceux calculés à l'aide du modèle I.A.M, et exprimées comme :

$$\Delta\rho_{\text{dyn}}(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{H}} \left(\frac{1}{K} |F_{\text{mul,calc}}(\mathbf{H})| e^{i\varphi_{\text{mul,calc}}(\mathbf{H})} - |F_{\text{sph,calc}}(\mathbf{H})| e^{i\varphi_{\text{sph,calc}}(\mathbf{H})} \right) e^{-2\pi\mathbf{H}\cdot\mathbf{r}} \quad (1.31)$$

Par définition, elles ne dépendent que des variables du modèle affiné et non des données expérimentales.

Ces cartes permettent d'observer la densité électronique dynamique telle qu'elle est reproduite par le modèle dans son ensemble (agitation thermique + distribution de charges).

- Les cartes de déformation de densité électronique statique

Le modèle multipolaire de Hansen & Coppens décrit précédemment reproduit l'asphéricité de la distribution électronique de valence. L'agitation thermique est un facteur handicapant et que l'on tente de réduire le plus possible au cours de l'expérience. La distribution de charges est la partie intéressante du modèle. Pour étudier l'asphéricité de la distribution électronique de valence, on s'intéresse aux cartes de déformation de densité électronique de valence statique.

Ne prenant pas en compte les effets de l'agitation thermique, ces cartes sont obtenues dans l'espace direct. Elles sont définies comme la différence entre la distribution électronique de valence obtenue à l'aide du modèle multipolaire et celle obtenue à l'aide du modèle I.A.M.

$$\Delta\rho_{stat}(\mathbf{r}) = \sum_{i=1}^{N_a} \left(k^3 P_{val} \rho_{val, sph}(kr) - N_{val} \rho_{val, sph}(r) + \sum_{l=0}^{l_{max}} k^{l+3} R_l(k'r) \sum_{m=0}^l P_{lm\pm} Y_{lm\pm}(\theta, \varphi) \right) \quad (1.32)$$

On pose N_{val} , la population de valence neutre de l'atome considéré et N_a , le nombre d'atomes contribuant à la différence de densité électronique représentée.

I.2.4.C. Test de la liaison rigide

Comme nous l'avons vu précédemment, au sein d'une molécule, les déplacements quadratiques d'agitation thermique des atomes peuvent être décomposés en deux contributions supposées décorréelées : une contribution dite « externe » U_{Rigide}^{ij} et une contribution dite « interne » $U_{Interne}^{ij}$.

$$U^{ij} = U_{Rigide}^{ij} + U_{Interne}^{ij} \quad (1.33)$$

La contribution interne est essentiellement due aux oscillations de torsion et d'angles qui n'affectent pas les amplitudes relatives des déplacements thermiques des atomes liés ensemble le long de leur liaison, par conséquent leur liaison est rigide. C'est sur cette observation que repose le test de la liaison rigide.

Vérifier la rigidité des liaisons atomiques permet de s'assurer de la réalité physique des paramètres d'agitations thermiques affinés, et est donc une condition de validation du modèle multipolaire.

Par exemple, la molécule (A—B) est considérée rigide si les deux atomes A et B ont des amplitudes moyennes de vibration égales le long de la liaison, c'est-à-dire que $\Delta_{A,B}$ est nulle :

$$\Delta_{A,B} = |Z_A^2 - Z_B^2| \quad (1.34)$$

Z_A^2 et Z_B^2 étant les amplitudes moyennes des déplacements respectivement des atomes A et B dans la direction de la liaison A—B.

Les amplitudes moyennes de vibration le long de la liaison ne sont pas strictement égales. Hirshfeld [18] suggère que pour les atomes lourds au-delà de carbone, la différence d'amplitudes $\Delta_{A,B}$ doit être inférieure de 10^{-3} \AA^2 .

Si cette condition est observée pour l'ensemble des liaisons de la molécule, on considère que la déconvolution entre la distribution électronique et les facteurs d'agitation thermique est excellente.

Ce premier chapitre consistait en une présentation des concepts et méthodes généraux nécessaires à la compréhension des travaux exposés dans ce manuscrit. Le chapitre suivant est consacré aux méthodes et outils utilisés dans l'étude des distributions de charges, et notamment le calcul des propriétés électrostatiques dérivées de ces dernières.

- [1] F. H. Allen, O. Kennard, D. Watson, L. Brammer, A. Orpen, R. Taylor. *International Tables for Crystallography, Volume C*, **chapter 9.5**, 685–706. Dordrecht: Kluwer Academic Publishers, 1992.
- [2] G. Bricogne. *Acta Crystallographica*, **A32**, 832–847, 1976.
- [3] A. T. Brünger. *Nature (London)*, **355**, 472–475, 1992.
- [4] A. T. Brünger. *Acta Crystallographica*, **D49**, 24–36, 1993.
- [5] M. C. Burla, M. Carmalli, G. Cascarano, G. Giaccovazzo, C. Polidori, R. Spgna, D. Vittorbo. *Journal of Applied Crystallography*, **22**, 389–393, 1989.
- [6] L. Chen & B. Craven. *Acta Crystallographica*, **B51**, 1081–1097, 1995.
- [7] E. Clementi & C. Roetti. *Jouranl of Chemical Physics*, **38**, 2686–2689, 1963.
- [8] P. Coppens, T. N. Guru Row, P. Leung, E. D. Stevens, P. J. Becker, Y. W. Yang. *Acta Crystallographica*, **A35**, 63–72, 1979.
- [9] B. Dawson. *Proceedings of the Royal Society A*, **298**, 255–288, 1967.
- [10] J. J. DeMarco & R. J. Weiss. *Physical Review Section A*, **137**, 1869–1871, 1965.
- [11] S. Domagala & C. Jelsch. *Journal of Applied Crystallography*, **41**, 1140–1149, 2008.
- [12] E. Espinosa, C. Lecomte, E. Molins. *Journal of Chemical Physics*, **300**, 745–748, 1999.
- [13] E. Espinosa, E. Molins, C. Lecomte. *Journal of Chemical Physics*, **285**, 170–173, 1998.
- [14] E. Espinosa, M. Souhassou, H. Lachekar, C. Lecomte. *Acta Crystallographica*, **B55**, 563–572, 1999.
- [15] N. K. Hansen & P. Coppens. *Acta Crystallographica*, **A34**, 909–921, 1978.
- [16] M. Harel & F. L. Hirshfeld. *Acta Crystallographica*, **B31**, 162–172, 1975.
- [17] F. L. Hirshfeld. *Acta Crystallographica*, **B27**, 769–781, 1971.
- [18] F. L. Hirshfeld. *Acta Crystallographica*, **A32**, 239–244, 1976.
- [19] K. Kurki-Suonio. *Acta Crystallographica*, **A24**, 379–390, 1968.
- [20] A. O. Madsen. *Journal of Applied Crystallography*, **39**, 757–758, 2006.
- [21] A. O. Madsen, H. O. Sorensen, C. Flensburg, R. F. Stewart, S. Larsen. *Acta Crystallographica*, **A60**, 550–561, 2004.
- [22] J. F. McConnell & P. L. Sanger. *Acta Crystallographica*, **A26**, 83–93, 1970.
- [23] P. Munshi, A. O. Madsen, M. A. Spackman, S. Larsen, R. Destro. *Acta Crystallographica*, **A64**, 465–475, 2008.
- [24] M. G. Rossmann & D. M. Blow. *Acta Crystallographica*, **15**, 24–31, 1962.
- [25] V. Schomaker & K. N. Trueblood. *Acta Crystallographica*, **B24**, 63–76, 1968.
- [26] G. M. Sheldrick. *Acta Crystallographica*, **A46**, 467–473, 1990.
- [27] G. M. Sheldrick. *Acta Crystallographica*, **D64**, 112–122, 2008.
- [28] M. A. Spackman. *Chemical Physics Letter*, **301**, 425–429, 1999.
- [29] M. A. Spackman. *Acta Crystallographica*, **A55**, 30–47, 1999.

- [30] R. F. Stewart. *Journal of Chemical Physics*, **51**, 4569–4577, 1969.
- [31] R. F. Stewart. *Journal Of Chemical Physics*, **57** (4), 1664–1668, 1972.
- [32] R. F. Stewart. *Journal Of Chemical Physics*, **58** (4):1668–1676, 1973.
- [33] R. F. Stewart. *The application of Charge Density Research to Chemistry and Drug Design*, 63–101. New York: Plenum Press, 1991.
- [34] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, K. Morokuma. *Journal of Physical Chemistry*, **100**, 19357–19363, 1996.

Chapitre II

Caractérisation de la distribution de charges d'un composé et propriétés électrostatiques dérivées

La diffraction des rayons X à haute résolution donne accès à des informations subatomiques sur les composés. Le développement de modèles de distribution de charges asphérique dont le plus populaire et utilisé est le modèle de Hansen & Coppens [22], permet d'exploiter au mieux ces données subatomiques et d'obtenir une connaissance de la distribution de charges de ces composés.

Comme cela a été discuté dans la partie I.2.3. de ce manuscrit, un affinement multipolaire est un travail délicat qui nécessite d'être mené avec expertise et suivi précautionneusement à chacune de ses étapes.

Les modèles multipolaires obtenus, qu'ils soient intermédiaires ou finaux, sont usuellement examinés à l'aide de différentes cartes de densité électronique décrites dans le chapitre 1 de ce rapport. L'examen des paramètres d'un modèle ou des cartes de densité électronique ne permet qu'une appréciation limitée de la distribution de charges du composé. Des informations supplémentaires peuvent être dérivées du modèle et offrent un autre niveau d'interprétation de la distribution de charges.

II.1. Propriétés électrostatiques

II.1.1. Le potentiel électrostatique

La connaissance de la distribution de charges d'un composé permet d'obtenir ses propriétés électrostatiques analytiquement ou par calcul numérique.

Le potentiel électrostatique engendré par une distribution de charges $\rho(\mathbf{r})$ en un point M de vecteur position \mathbf{r} de l'espace est défini par l'expression :

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}' \quad (2.1)$$

Le potentiel électrostatique nous renseigne sur l'influence d'une distribution de charges sur son environnement et plus particulièrement sur la portée et l'orientation de cette influence.

En effet, en raison de son asphéricité, la distribution de charges d'un atome va engendrer un champ électrique qui sera plus intense dans certaines régions de l'espace et qui donc interagira plus favorablement selon certaines directions avec d'autres atomes.

Dans le cas d'une distribution de charges modélisée à l'aide d'un modèle multipolaire de Hansen & Coppens [22], le potentiel électrostatique engendré par un pseudo-atome peut être décomposé en plusieurs contributions, chacune correspondant à un terme du modèle.

$$V(\mathbf{r}) = V_{\text{coeur},sph}(\mathbf{r}) + V_{\text{val},sph}(\mathbf{r}) + V_{\text{val},def}(\mathbf{r}) \quad (2.2)$$

avec le potentiel électrostatique sphérique de cœur,

$$V_{\text{coeur},sph}(\mathbf{r}) = \left[\frac{1}{4\pi\epsilon_0} \left(\frac{Z}{|\mathbf{r} - \mathbf{R}|} - \int \frac{\rho_{\text{coeur},sph}(r')}{|(\mathbf{r} - \mathbf{R}) - \mathbf{r}'|} d^3\mathbf{r}' \right) \right] \quad (2.3)$$

le potentiel électrostatique sphérique de valence,

$$V_{\text{val},sph}(\mathbf{r}) = - \left[\frac{1}{4\pi\epsilon_0} \int k^3 P_{\text{val}} \frac{\rho_{\text{val},sph}(kr')}{|(\mathbf{r} - \mathbf{R}) - \mathbf{r}'|} d^3r' \right] \quad (2.4)$$

et le potentiel électrostatique asphérique de valence,

$$V_{val,def}(\mathbf{r}) = - \left[\frac{1}{4\pi\epsilon_0} \sum_{l=0}^{l_{max}} k^{l3} \int \frac{R_l(k'r')}{|\mathbf{r}-\mathbf{R}-\mathbf{r}'|} \sum_{m=0}^l P_{lm\pm} y_{lm\pm}(\theta, \varphi) d^3\mathbf{r}' \right] \quad (2.5)$$

Dans ces expressions, le vecteur position \mathbf{r} est défini dans le repère global du système étudié (**Figure 1**).

On considère \mathbf{R} comme le vecteur position dans le repère global du noyau de l'atome considéré et \mathbf{r}' vecteur position, du volume élémentaire $d^3\mathbf{r}'$, défini dans le repère local de l'atome centré dans l'expression des intégrales.

On notera que les paramètres de population multipolaire apparaissant dans l'expression du potentiel électrostatique ne sont pas les paramètres multipolaires définis dans le repère local optimal de l'atome (défini selon la stéréochimie de ce dernier). Ces paramètres de populations multipolaires sont définis dans un repère local colinéaire au repère global du système, et sont déduits des paramètres multipolaires originaux par opération de changement de repère.

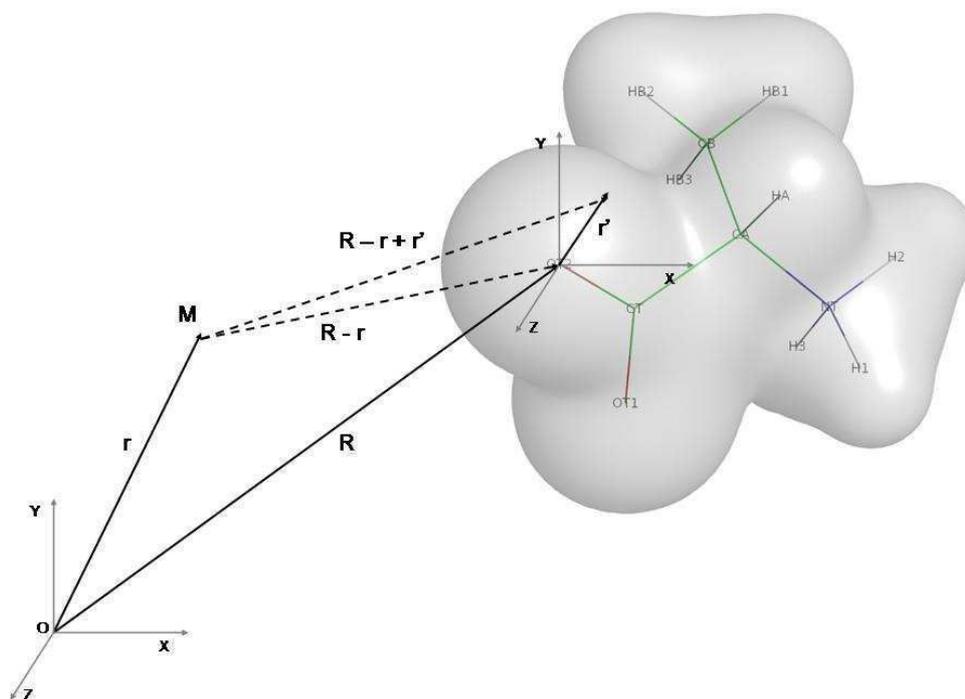


Figure 1 : Illustration des différents vecteurs positions impliqués dans l'expression du potentiel électrostatique engendré par la distribution de charges d'un pseudo-atome. L'alanine a été choisie à titre d'exemple. La surface d'isodensité de l'alanine est représentée pour un isoniveau de $0,05 \text{ e} \cdot \text{Å}^{-3}$. Cette illustration a été réalisée avec le logiciel Pymol [13].

Le potentiel électrostatique engendré par une distribution de charges représentée par un modèle multipolaire peut être calculé analytiquement [18]. Dans ces travaux, le logiciel VMoPro [23] a été utilisé pour calculer le potentiel électrostatique basé sur les travaux de N. E. Ghermani *et al.*[18].

II.1.2. Le Laplacien de densité électronique

Il est également possible de dériver de la connaissance d'une distribution électronique, les valeurs de son Laplacien. Le Laplacien $\nabla^2 \rho(\mathbf{r})$ d'une distribution électronique $\rho(\mathbf{r})$ en un point M de vecteur position \mathbf{r} est défini comme :

$$\nabla^2 \rho(\mathbf{r}) = \frac{\partial^2 \rho(\mathbf{r})}{\partial^2 x} + \frac{\partial^2 \rho(\mathbf{r})}{\partial^2 y} + \frac{\partial^2 \rho(\mathbf{r})}{\partial^2 z} \quad (2.6)$$

La connaissance du Laplacien de ρ permet de faire le lien entre une représentation précise de la distribution électronique et des outils plus simples de représentation du cortège électronique comme le modèle de Lewis [29] ou de prédiction de la stéréochimie comme la méthode VSEPR pour « Valence Shell Electron Pair Repulsion » ou, en français, Répulsion des Paires d'Electrons des Couches de Valence [19].

Bien que l'étude d'un modèle de densité électronique continue et asphérique permette, notamment par le biais de sa topologie (partie II.1.3.), de comprendre la nature des liaisons chimiques dans le cadre de la mécanique quantique, cette précision l'éloigne indéniablement de considérations plus intuitives développées dans le formalisme électronique de Lewis. Le Laplacien de la densité électronique offre la possibilité de réconcilier ces deux niveaux d'abstraction. En effet, le Laplacien nous renseigne sur les zones de concentration/déplétion électronique. Une valeur de Laplacien négative révèle une concentration locale d'électrons alors qu'une valeur positive correspondra à une déplétion locale d'électrons. Il nous est d'ailleurs possible sur des cartes de Laplacien de densité électronique de retrouver la structure en couches électroniques de valence de l'atome par une alternance de couches de concentration et de déplétion de charges. De plus, les doublets électroniques non liants prédits par la méthode VSEPR sont également identifiables et localisables sur ces cartes (**Figure 2**).

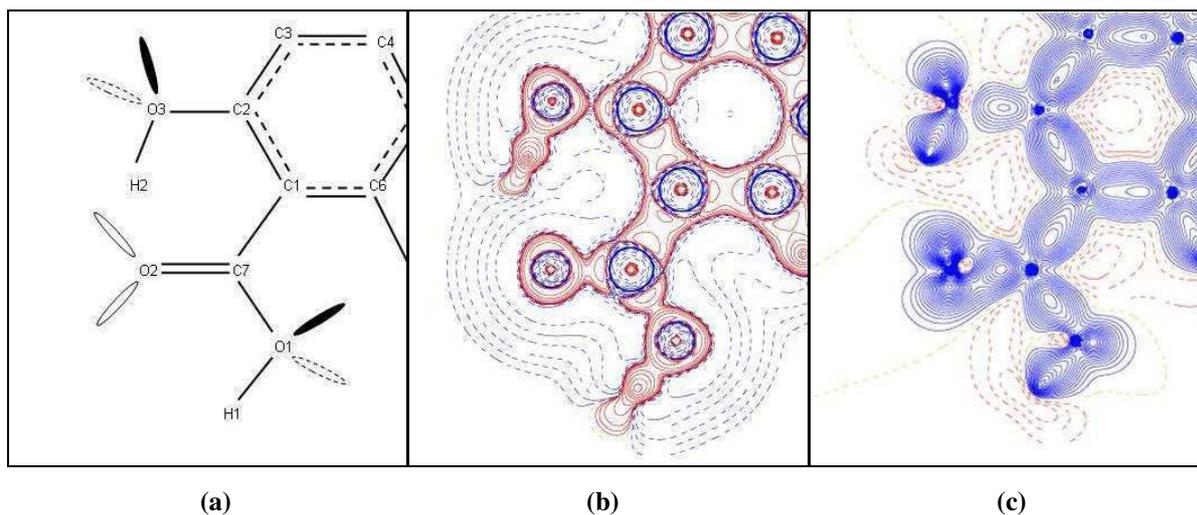


Figure 2 : Illustration comparative du groupement carboxyle de l'acide salicylique avec le modèle VSEPR du groupement (a), les cartes de Laplacien (b) et la carte de déformation de sa densité électronique statique. Sur les cartes (b) et (c), les isocontours positifs sont en bleu, et les isocontours négatifs en rouge. Ces illustrations ont été réalisées à l'aide du logiciel VMoPro [23].

Dans le cas d'un modèle multipolaire de Hansen & Coppens, les cartes de déformation de densité électronique statique permettent également d'observer les doublets électroniques non liants ainsi que les concentrations locales de densité électronique sur les liaisons covalentes. Cependant, les cartes de Laplacien de densité électronique peuvent être utilisées quelque soit l'origine de la distribution.

II.1.3. L'analyse topologique de la distribution de charges

En complément des cartes de densité, de potentiel et de Laplacien, il est possible de réaliser une analyse de la topologie de la densité électronique appelée communément : la méthode A.I.M. (Atomes In Molecules) [4]

Une molécule est perçue simplement comme un ensemble d'atomes liés par des liaisons chimiques. Cette vision de la chimie moléculaire permet d'appréhender et de décrire une molécule en déterminant la longueur de ses liaisons et les valeurs de ses angles valencielles.

Cependant, l'étude d'une molécule en tant que distribution continue de charges nécessite d'avoir recours à un outil d'analyse fondé sur la mécanique quantique : la physique qui gouverne le comportement des électrons.

Cette analyse, qui repose sur des considérations quantiques, a été développée initialement pour l'étude des distributions de charges théoriques [5;6;3] puis c'est vue étendue aux distributions de charges obtenues par affinement cristallographique à partir de jeux de données expérimentales.

Elle repose sur la décomposition de la distribution de charges. Dans cette décomposition, chaque petit élément de la distribution de charges est associé à un seul et unique atome. On parle alors de « distribution discrète » à la différence des distributions dites « floues » dans lesquelles un volume élémentaire peut appartenir à plusieurs atomes comme cela est le cas dans le formalisme de Hansen & Coppens [22].

Le lieu des volumes élémentaires associés à un atome est appelé « bassin atomique ». Le bassin atomique d'un atome i est défini comme étant le volume contenant cet atome i et dont le flux du gradient de densité de charges ρ est nul en tous points de la surface S qui le délimite, ce qui se pose comme :

$$\forall M \in S, \mathbf{grad}(\rho(M)) \cdot \mathbf{n} = 0 \quad (2.7)$$

avec \mathbf{n} vecteur normal à la surface S au point M .

De cette définition découle une propriété-clé des bassins atomiques. En effet, d'après le théorème de flux-divergence ou théorème de Green-Ostrogradski, si l'on considère \mathbf{F} un vecteur flux, V un volume de l'espace et S la surface de ce volume, l'intégrale du flux \mathbf{F} sur la surface S du volume est égale à l'intégrale sur le volume de la divergence de \mathbf{F} :

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iiint_V \text{div}(\mathbf{F}) dV \quad (2.8)$$

Dans notre cas, nous obtenons :

$$\iint_S \mathbf{grad}(\rho(M)) \cdot d\mathbf{S} = \iiint_V \text{div}(\mathbf{grad}(\rho(M))) dV = \iiint_V \Delta(\rho(M)) dV \quad (2.9)$$

On sait, par définition des bassins, que $\forall M \in S, \mathbf{grad}(\rho(M)) \cdot \mathbf{n} = 0$ donc

$$\iiint_V \Delta(\rho(M)) dV = 0 \quad (2.10)$$

Cette propriété sert de critère de validation de la définition des bassins atomiques.

La topologie de Bader s'intéresse à des propriétés dites locales qui caractérisent la distribution de charges en un point de l'espace, et à des propriétés dites intégrées qui caractérisent les bassins atomiques.

II.1.3.A. Propriétés locales

Les propriétés dites locales sont des caractéristiques de points de l'espace où la valeur du gradient de la distribution de charges est nulle. Ces points sont appelés « points critiques » ou « critical points » en anglais.

Ces points critiques sont classés en différentes catégories en fonction des valeurs des dérivées secondes partielles de la densité de charges, et renseignent sur le comportement de la distribution de charges en ces points. On définit la matrice Hessienne H en \mathbf{r} qui est la matrice symétrique des dérivées secondes de la densité électronique au point M de vecteur position \mathbf{r} .

$$H_{ij}(\mathbf{r}) = \frac{\partial^2 \rho(\mathbf{r})}{\partial x_i \partial x_j} \quad (2.11)$$

La diagonalisation de la matrice Hessienne H en \mathbf{r} permet d'obtenir les vecteurs propres $(\mathbf{u}_i)_{i=1..3}$ (axes de quadrature principaux au voisinage du point M) et les valeurs propres correspondantes (valeurs de quadrature principales). Ces valeurs propres sont notées $(\lambda_i)_{i=1..3}$ et numérotées dans l'ordre croissant.

$$\lambda_i = \frac{\partial^2 \rho(\mathbf{r})}{\partial^2 u_i} \text{ avec } i = 1..3 \quad (2.12)$$

Pour identifier le type des points critiques, on les distingue à l'aide du couple (p, n) avec n , le nombre de valeurs propres non nulles appelé « le rang », et s , la somme algébrique des signes des valeurs propres de la matrice Hessienne, que l'on appelle « la signature ».

Leurs noms viennent de leurs origines matricielles : n est, en effet, le rang de la matrice Hessienne au point critique, et p est la trace de la matrice signature correspondante à la matrice Hessienne diagonalisée. (la matrice signature M d'une matrice A est la matrice dont chaque terme M_{ij} est le signe du terme A_{ij}).

Les points critiques ont, dans un espace à 3 dimensions, toujours un rang de 3.

On distingue 3 types de points critiques :

- (3, -3) : Trois courbures négatives ce qui caractérise les maximums de densité électronique. On ne trouve ces maximums qu'aux positions atomiques.
- (3, -1) : Deux courbures négatives (λ_1 et λ_2) et une positive (λ_3). On parle de point critique de selle ou « saddle point » en anglais. Ce type de point critique est rencontré entre chaque paire de noyaux atomiques liés par une liaison chimique qu'elle soit covalente ou non.
- (3,+1) : Une courbure est négative (λ_1) et deux positives (λ_2 et λ_3). Le point critique est dit « de cycle » ou « ring critical point ». On trouve ce type de point critique dans les cycles comme les cycles aromatiques.
- (3, +3) : Trois courbures positives caractérisent un minimum de densité électronique que l'on trouve au sein de cages atomiques formées par exemple par deux cycles chimiques. On les appelle pour cette raison : point critique de cage ou « cage critical point ».

Les points critiques définis précédemment sont intimement liés aux lignes de gradient de la densité électronique. On définit une ligne de gradient comme une ligne dans l'espace qui en chacun de ses points M est tangente au vecteur gradient défini en ce point. Le vecteur gradient de ρ indique la direction de la plus grande augmentation de ρ en partant de M.

Par conséquent, quelque soit le point M de départ, la ligne de gradient de ρ se terminera en un point critique de signature supérieure ou égale à -1 en d'autre terme en un lieu où la densité électronique est maximum selon au moins une direction.

Dans le cas d'un atome isolé, toutes les lignes de gradient naissent à l'infini et convergent au noyau de l'atome.

Dans le cas d'un système composé de plusieurs atomes, deux atomes seront considérés liés si leurs points critiques (3,-3) peuvent être reliés par une ligne de gradient le long de laquelle la densité électronique est maximale. Cette ligne est appelée « chemin de liaison ». Le long de ce chemin de liaison, on trouvera un point critique de type « selle » ou (3,-1). Les propriétés de la densité électronique au point critique permettent de caractériser la nature de la liaison.

Le Laplacien de la densité électronique renseigne sur le comportement de la fonction densité électronique au voisinage du point critique.

Il est défini, comme dit précédemment, par :

$$\Delta\rho(\mathbf{r}) = \frac{\partial^2 \rho(\mathbf{r})}{\partial^2 x} + \frac{\partial^2 \rho(\mathbf{r})}{\partial^2 y} + \frac{\partial^2 \rho(\mathbf{r})}{\partial^2 z} \quad (2.13)$$

En effet, un Laplacien négatif au point critique signifie que la densité électronique se concentre en ce point, ce qui révèle une interaction covalente à couche ouverte. La liaison covalente est obtenue par une mise en commun des électrons appartenant aux atomes en interaction. Par ailleurs, plus la valeur de la densité électronique est élevée et celle du Laplacien faible, plus la liaison est considérée forte. Un Laplacien positif montre une déplétion électronique locale ce qui caractérise une interaction à couche fermée, comme une liaison hydrogène ou ionique. La densité électronique y est nettement plus faible que dans le cas d'une interaction à couche ouverte.

II.1.3.B. Propriétés intégrées

La décomposition de la densité électronique selon le formalisme A.I.M. [4] permet de définir un atome comme l'association d'un noyau atomique et d'un bassin de densité électronique contenant ce noyau.

La valeur de l'intégration de la distribution électronique au sein d'un bassin atomique correspond à la charge électronique caractéristique de l'atome correspondant.

$$N = \int_{V_{\text{BASSIN}}} \rho(\mathbf{r}) dr \quad (2.14)$$

L'analyse de la topologie de la densité électronique permet de déterminer les charges électroniques de chaque atome selon une décomposition unique de la densité électronique en bassins atomiques.

Pour obtenir la charge formelle q de l'atome, il suffit de soustraire sa charge électronique intégrée N à la charge électronique de cet atome lorsqu'il est supposé « neutre » N_{neut} .

$$q = N_{\text{neut}} - N \quad (2.15)$$

La topologie A.I.M. a l'élégance d'offrir un outil de décomposition des distributions électroniques parfaitement indépendant des modèles dont dérivent ces distributions. Elle peut donc être utilisée sur des distributions de charges d'origine diverse (théorique ou expérimentale). Les analyses topologiques réalisées dans le cadre des projets exposés dans ce rapport ont été réalisées à l'aide des logiciels VMoPro [23] et Integrity [26].

Remarque : La charge électronique N de l'atome correspond au moment d'ordre 0 de la distribution électronique du bassin atomique. On peut définir et calculer le moment d'ordre n avec n , un entier positif comme :

$$M^n = \int_{V_{BASSIN}} r^n \rho(\mathbf{r}) d\mathbf{r} \quad (2.16)$$

Le volume de bassin atomique étant fini (ou supposé fini dès lors que la densité électronique devient négligeable) et la densité électronique étant continue, le moment d'ordre n existe.

II.2. Energies d'interaction intermoléculaire

La caractérisation mais surtout l'estimation quantitative des énergies d'interaction d'une molécule avec son environnement sont d'un intérêt majeur pour comprendre les propriétés physico-chimiques des composés ou groupements chimiques.

Les propriétés locales obtenues par analyse topologique de la densité électronique sont utiles pour estimer la force des liaisons chimiques non covalentes tout particulièrement les liaisons hydrogène. Ces propriétés locales apportent une information sur les contributions des interactions atome-atome (ou entre deux groupements chimiques) localisées aux interactions intermoléculaires entre deux molécules. Ces contributions locales bien que corrélées aux interactions intermoléculaires globales ne les représentent pas strictement.

II.2.1. Expression de l'énergie d'interaction totale entre deux composés

L'énergie d'interaction intermoléculaire entre deux composés dépend bien sûr du modèle retenu pour le système étudié.

L'énergie d'interaction intermoléculaire se décompose en plusieurs contributions dont le nombre peut varier selon la précision de la décomposition. Différentes expressions ainsi que paramétrisations sont possibles pour chaque contribution. Il n'existe pas de décomposition de l'énergie d'interaction que l'on pourrait qualifier d'« exacte » ou de « meilleure ». Chaque décomposition de l'énergie d'interaction intermoléculaire constitue un modèle de l'énergie d'interaction à un niveau d'abstraction choisi pour répondre à un besoin donné.

On trouve généralement quatre termes dans les expressions de décomposition des énergies d'interaction intermoléculaire E_{int} :

$$E_{int} = E_{\text{electro}} + E_{\text{polar}} + E_{\text{dispersion}} + E_{\text{repulsion}} \quad (2.17)$$

- La contribution électrostatique $E_{electro}$

La contribution électrostatique $E_{electro}$ correspond à l'énergie d'interaction coulombienne entre deux distributions de charges non perturbées par le champ électrique engendré par l'environnement.

Les distributions de charges sont modélisées à l'aide de modèles plus ou moins précis selon la décomposition considérée. Ils peuvent n'être qu'une collection de charges ponctuelles placées aux positions atomiques comme être une description continue de la distribution électronique.

La formule générale de $E_{electro}$ est :

$$E_{electro} = \frac{1}{4\pi\epsilon_0} \iint \frac{\rho_A(\mathbf{r}_A) \times \rho_B(\mathbf{r}_B)}{|\mathbf{r}_A - \mathbf{r}_B|} d\mathbf{r}_A d\mathbf{r}_B \quad (2.18)$$

- La contribution de polarisation E_{polar}

L'énergie d'interaction de polarisation entre deux distributions de charges notées A et B correspond aux interactions entre les moments permanents de la distribution de charges A et les moments induits de la distribution de charges B sous l'effet du champ électrique engendré par la distribution de charges A.

On considère usuellement que les dipôles induits et permanents des distributions de charges dans l'expression de l'énergie de polarisation E_{polar} , les contributions des moments multipolaires d'ordre supérieurs étant considérées comme négligeables.

Qu'est-ce qu'un dipôle induit ?

Une molécule polaire ou apolaire (dipôle permanent nul) peut, sous l'effet du champ électrique \mathbf{E} créé par son environnement de charges, voir la géométrie de son cortège électronique altérée ce qui induit une modification de ses moments électroniques dont, au premier ordre, de son dipôle. On introduit généralement un dipôle induit $\boldsymbol{\mu}_{induit}$ qui s'ajoute au dipôle permanent $\boldsymbol{\mu}_{permanent}$ de la distribution de charges.

L'amplitude de ce dipôle induit $\boldsymbol{\mu}_{\text{induit}}$ est fonction du champ électrique \mathbf{E} de l'environnement mais aussi de la propension qu'a la distribution de charges à répondre à une perturbation environnementale. Cela est pris en compte par le coefficient de polarisabilité α .

Le dipôle induit $\boldsymbol{\mu}_{\text{induit}}$ s'écrit alors : $\boldsymbol{\mu}_{\text{induit}} = \alpha \mathbf{E}$

L'énergie de polarisation entre deux distributions de charges s'écrit :

$$E_{\text{polar}} = - \left(\frac{|\boldsymbol{\mu}_A|^2 \alpha_B}{4\pi^2 \epsilon_0^2 r^6} + \frac{|\boldsymbol{\mu}_B|^2 \alpha_A}{4\pi^2 \epsilon_0^2 r^6} \right) \quad (2.19)$$

avec, α_A et α_B les coefficients de polarisabilité respectifs des distributions de charges A et B, et, $\boldsymbol{\mu}_A$ et $\boldsymbol{\mu}_B$ les moments dipolaires permanents respectifs des distributions de charges A et B.

- L'énergie de dispersion $E_{\text{dispersion}}$

Elle représente l'énergie d'interaction entre deux distributions de charges A et B due à une force attractive faible à longue portée résultant de l'interaction entre les moments induits de la distribution de charges A et ceux de la distribution de charges B dus au champ électrique engendré par les moments instantanés de la distribution de charges A. Elle dépend de la polarisabilité des distributions de charges en interaction.

Usuellement, on ne considère, en première approximation, que les interactions dipôles induits – dipôles induits.

Cette interaction permet à deux distributions de charges apolaires d'interagir. Cela vient du fait que, bien qu'une distribution de charges puisse être apolaire, ce caractère apolaire ne reflète qu'une tendance moyenne dans le temps. A un instant donné, les électrons étant en mouvement permanent, la distribution de charges peut présenter un dipôle de durée de vie courte. Ce dipôle dit instantané va créer dans l'environnement de la distribution de charges un champ électrique \mathbf{E} qui, comme dans le cas de la polarisation, va engendrer un dipôle induit $\boldsymbol{\mu}_{\text{induit}}$.

Ce phénomène explique pourquoi des gaz non-polaires tels qu'un gaz rare peuvent condenser.

Un modèle pour cette énergie peut s'écrire :

$$E_{dispersion} = - \sum_{i \in A} \sum_{j \in B} \frac{C_{6,ij}}{r_{ij}^6} \quad (2.20)$$

Ils existent de nombreux jeux de coefficients C_6 .

L'expression la plus courante fait intervenir des coefficients a_i monoatomiques :

$$E_{dispersion} = - \sum_{i \in A} \sum_{j \in B} \frac{a_i a_j}{r_{ij}^6} \quad (2.21)$$

Les coefficients a_i sont des paramètres dépendant du type d'atome considéré, et r_{ij} est la distance entre les atomes d'indices i et j et appartenant respectivement aux molécules notées A et B .

- L'énergie de répulsion $E_{repulsion}$

Ce terme d'énergie prend en compte le principe d'exclusion de Pauli qui induit une interaction répulsive à courte portée entre deux distributions, et qui s'oppose à leur interpénétration.

Un modèle couramment utilisé est :

$$E_{repulsion} = \sum_{i \in A} \sum_{j \in B} b_i b_j \exp[-(c_i + c_j)r_{ij}] \quad (2.22)$$

Les coefficients monoatomiques b_j et c_j sont, ici aussi, des paramètres dépendants du type d'atome j considéré.

Certains champs de force proposent une expression différente pour l'énergie de répulsion :

$$E_{repulsion} = \sum_{i \in A} \sum_{j \in B} \frac{b_i b_j}{r_{ij}^{12}} \quad (2.23)$$

Cette expression en puissance inverse de r_{ij} tient son origine d'une considération purement pratique. Lors du développement des premiers champs de force, et donc des premières tentatives de décomposition des énergies d'interaction, les ressources informatiques ne permettaient pas de réaliser certaines opérations coûteuses. Le nombre d'opérations devait être minimisé afin de limiter la durée des calculs. Substituer un terme en exponentielle par

l'inverse de r_{ij} à la puissance 12. (2 fois le terme r_{ij}^6) permettait de réduire significativement le temps de calcul. D'autres expressions peuvent être utilisées pour ces deux contributions.

Les expressions usuelles pour les énergies de dispersion et de répulsion dépendent des paramètres a_i , b_i et c_i pour lesquels plusieurs jeux de valeurs sont envisageables. Les deux jeux de paramètres les plus connus dans le domaine des densités de charges pour ces expressions sont celui proposé par Spackman *et al.* [38] ainsi que celui proposé par Williams & Cox [45;12]. Bien que les décompositions des énergies d'interaction soient similaires dans ces deux cas, les paramètres des termes d'énergie diffèrent. Cela est dû aux méthodes utilisées pour les obtenir.

Il n'y a pas de définition unique pour ces contributions. Leurs expressions peuvent changer selon l'expression de l'énergie d'interaction et la précision souhaitée. Elles sont considérées comme des interactions à deux corps, les contributions d'ordres supérieures étant supposées négligeables.

Il est important de garder à l'esprit que le but d'une décomposition de l'énergie d'interaction totale est de prendre en compte explicitement ou implicitement tous les phénomènes qui vont contribuer significativement à l'interaction.

De plus, il n'y a pas de frontières strictes entre chaque contribution à l'énergie d'interaction totale. Que ces termes soient paramétrés à l'aide de calculs théoriques ou empiriquement de sorte à reproduire certains résultants expérimentaux, ces contributions peuvent rester corrélées les unes aux autres dans leur définition. Pour cela, ils ne sont pas aisément interchangeables les uns avec les autres. Chaque décomposition correspond à un modèle d'interactions intermoléculaires valide selon certaines approximations et dans certaines conditions d'utilisation.

II.2.2. Calcul des énergies d'interaction électrostatique

II.2.2.A. Intégration par méthode de quadrature en coordonnées sphériques

Comme discuté précédemment, il existe plusieurs façons de modéliser et de calculer l'énergie d'interaction électrostatique entre deux molécules. L'expression de l'énergie d'interaction électrostatique est définie en fonction de modèles plus ou moins complexes de distributions de charges.

Le modèle le plus ancien et le plus rudimentaire est sans aucun doute le modèle atomique de charges ponctuelles. Comme son nom l'indique, la distribution de charges d'une molécule est définie comme une distribution discrète de charges ponctuelles centrées sur les positions des noyaux atomiques. L'expression de l'énergie qui en résulte est une simple application de la loi de Coulomb. Elle est pour cela employée dans les champs de force standard qualifiés de « Classe I » comme AMBER ff99 [43], CHARMM22 [30] ou OPLS-AA [25]. Il est aisé de dériver cette expression pour les besoins de la minimisation de la fonction d'énergie de la structure et d'estimer les valeurs de cette énergie ainsi que ses dérivées premières.

Ce modèle ne permet cependant pas, par définition, de prendre en compte l'asymétrie de la distribution de charges au voisinage d'un atome, par exemple, engendrée par les doublets électroniques d'un atome ou la présence d'un nuage électronique π de part et d'autre d'un cycle. Des champs de force tentent de pallier cette faiblesse en introduisant, par exemple, des charges satellites sur les liaisons covalentes des cycles aromatiques, ou localisées sur les doublets électroniques, par exemple : le champ de force XED [10]. De plus, même avec des charges supplémentaires excentrées des noyaux, leurs distributions de charges demeurent fixes, et ne permettent donc pas de prendre en compte les effets du champ électrique de l'environnement sur les distributions de charges. Pour répondre à ce besoin des versions polarisables des champs de force AMBER [44] et CHARMM [34] ont été développées. Une autre méthode consiste à accroître la flexibilité de la distribution de charges et ainsi le réalisme du potentiel électrostatique en ayant recours à des moments atomiques multipolaires d'ordre supérieur, comme cela est le cas, par exemple, dans les champs de force AMOEBA [36] et SIBFA [20]

Dans le domaine de la cristallographie, le modèle multipolaire utilisé dans les affinements cristallographiques à très haute résolution considère la distribution de charges d'une molécule comme une fonction continue et reproduit l'asphéricité des distributions atomiques.

L'énergie d'interaction électrostatique peut être définie à partir des modèles multipolaires de distribution de charges de deux molécules notées A et B en interaction comme une double somme sur les pseudo-atomes discrétisant les distributions de charges. L'énergie d'interaction électrostatique devient une somme de contributions entre paires de pseudo-atomes :

$$E_{\text{electro}} = \frac{1}{4\pi\epsilon_0} \sum_{i \in A} \sum_{j \in B} \iint \frac{\rho_{A,i}(\mathbf{r}) \times \rho_{B,j}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \quad (2.24)$$

$$E_{\text{electro}} = \sum_{i \in A} \sum_{j \in B} \int \rho_{A,i}(\mathbf{r}) \times V_{B,j}(\mathbf{r}) d\mathbf{r} = \sum_{i \in A} \sum_{j \in B} \int V_{A,i}(\mathbf{r}') \times \rho_{B,j}(\mathbf{r}') d\mathbf{r}' \quad (2.25)$$

Le potentiel électrostatique engendré par un pseudo-atome peut être obtenu analytiquement. Cependant, la solution analytique de l'intégration du produit d'une densité de charges d'un pseudo-atome et du potentiel électrostatique engendré par la distribution de charges d'un autre pseudo-atome ne peut être obtenue à ma connaissance.

Une méthode d'intégration numérique est, par conséquent, nécessaire pour estimer l'énergie d'interaction électrostatique. Le recours à une telle méthode numérique implique la discrétisation de la formule analytique. La précision de l'estimation – la différence entre l'estimation obtenue et le résultat idéal attendu – va dépendre de la méthode d'intégration retenue et du schéma de discrétisation utilisé.

L'expression intégrale (2.25) est le produit de la densité de charges d'un pseudo-atome avec le potentiel électrostatique engendré par le second pseudo-atome, soit une densité volumique d'énergie électrostatique.

Une première méthode d'intégration numérique qui fut implémentée dans le logiciel VMOPro [23] reposait sur la stratégie suivante :

- Définition d'un volume 3D parallélépipédique englobant la distribution de charges considérée,
- Discrétisation régulière de l'espace délimité par ce volume en éléments finis selon un pas de discrétisation constant,
- Intégration numérique de la densité volumique d'énergie dans chaque élément fini à l'aide d'un développement de Taylor d'ordre 5.

Cette méthode s'est malheureusement révélée coûteuse en temps avec des durées de calcul de l'ordre de 20-30 minutes pour deux molécules de 30 atomes en interaction (Processeur Xeon 3 GHz).

L'inefficacité de cette stratégie peut s'expliquer par l'intégration des éléments finis sans distinction préalable de leurs localisations par rapport aux atomes. L'algorithme considère les différentes régions du volume intégré, en traitant avec un pas de discrétisation plus fin, et donc une meilleure précision, les régions de forte variation de densité de charges.

Deux informations ont été négligées dans cette première solution et doivent être prise en compte pour un temps de calcul optimal :

- Les positions des atomes qui permettent de focaliser sur les régions de la molécule où la densité est concentrée,
- La symétrie de la densité électronique atomique.

Afin d'éviter d'intégrer sur un espace large présentant un pourcentage significatif de zones intermoléculaires où la densité d'énergie électrostatique est quasi nulle, une décomposition de l'intégrale en contributions de pseudo-atome pseudo-atome est le choix le plus simple.

L'intégration d'une telle contribution consistera à estimer l'intégrale de la densité d'énergie sur un volume centré sur l'un des deux atomes. Si l'on considère l'énergie d'interaction électrostatique $E_{electro,ij}$ entre les pseudo-atomes d'indices i et j , $E_{electro,ij}$ est donnée par :

$$E_{electro,ij} = \int \rho_i(\mathbf{r}) \times V_j(\mathbf{r}) d\mathbf{r} = \int e_{i,j}(\mathbf{r}) d\mathbf{r} \quad (2.26)$$

En première approximation, la distribution électronique d'un atome est assimilable à une sphère. Le potentiel électrostatique engendré par une distribution de charges dans son environnement étant plus régulier, l'on peut considérer que la distribution de densités volumiques d'énergies $e_{i,j}$ présentera également en première approximation une distribution quasi-sphérique. Proposer une méthode d'intégration de cette distribution en coordonnées cartésiennes dans un cube, n'est donc pas le choix le plus judicieux. Une méthode d'intégration en coordonnées sphériques est le choix optimal qui respectera la symétrie naturelle de la distribution d'énergies.

Pour s'en convaincre, on constatera que si l'on considère une fonction de distribution sphérique, le rapport des volumes de la sphère minimale contenant la distribution de la fonction et du cube minimal (qui n'est autre que le cube circonscrit à la sphère) est d'environ 52 %. Ainsi presque la moitié du volume du cube minimal contenant la distribution est prise en compte dans le calcul inutilement.

Fort de ces observations, une nouvelle approche a été retenue pour l'intégration de la densité d'énergies volumiques. On considère la densité volumique d'énergies comme une fonction en coordonnées sphériques définie dans le repère local de l'atome dont on considère la distribution de charges dans l'expression (2.26). Cette méthode d'intégration en coordonnées sphériques repose sur les utilisations complémentaires de deux méthodes de quadrature comme cela a été proposé par Treutler *et al.*[7;40]. Cette méthode a été initialement définie pour l'intégration de la densité électronique dans les calculs *ab initio* de type DFT (théorie de la fonctionnelle de densité) et implémentée dans le programme TURBOMOLE [2].

L'énergie d'interaction électrostatique est obtenue par intégration de la distribution de la densité d'énergie électrostatique volumique correspondante. L'intégration est réalisée en coordonnées sphériques $\mathbf{r} = (R, \Omega)$ avec R , la distance à l'origine et Ω , l'angle solide. Deux méthodes quadratiques complémentaires sont utilisées : l'une radiale définie par les couples coordonnée-pondération (R_i, ω_i) définis sur l'intervalle $[0; +\infty[$, et l'autre sphérique (Ω_j, ϖ_j) sur la sphère unitaire. L'expression de l'énergie d'interaction électrostatique entre deux pseudo-atomes n et m devient :

$$E_{\text{electro},nm} = \int e(\mathbf{r})d\mathbf{r} = \int_{R=0}^{+\infty} \iint_{\text{SphèreUnitaire}} e(R, \Omega)R^2 dRd^2\Omega = 4\pi \sum_{i=1}^I \sum_{j=1}^J \omega_i \varpi_j R_i^2 e(R_i, \Omega_j) \quad (2.27)$$

avec I , le nombre de points radiaux dans l'intervalle et J , le nombre de points angulaires.

L'intégration est réalisée en considérant I points radiaux et J points angulaires ce qui nécessite donc l'évaluation de la densité volumique d'énergies électrostatiques en $I*J$ points.

La discrétisation de la partie radiale est un problème à une dimension résolu à l'aide de la méthode de quadrature de Gauss-Chebyshev avec changement de variable.

La quadrature de Gauss-Chebyshev est employée pour discrétiser l'intégration d'une fonction sur $]-1,+1[$ avec pour fonction de pondération $W(x) = (1-x^2)^{-\frac{1}{2}}$. Pour une discrétisation en I points soit d'ordre $2I-1$, les couples coordonnée-pondération (x_i, w_i) sont données par :

$$x_i = \cos\left(\frac{(2i-1)\pi}{2I}\right) \text{ et } w_i = \frac{\pi}{I} \quad (2.28)$$

Afin de pouvoir employer la méthode de quadrature de Gauss-Chebyshev pour la partie radiale, fonction définie sur $[0; +\infty[$, un changement de variable est nécessaire.

Ce dernier est réalisé selon la formule (M4) proposée par Treutler et Ahlrichs [40] :

$$M_4 :]-1;1[\mapsto]0;+\infty[$$

$$M_4 : x \rightarrow \frac{\xi}{\ln(2)}(1+x)^\alpha \ln\left(\frac{2}{1-x}\right) \text{ avec } \alpha = 0,6 \quad (2.29)$$

La valeur du paramètre ξ est définie pour chaque élément chimique. Ces valeurs ont également été proposées par Treutler et Ahlrichs [40].

L'expression de l'énergie d'interaction électrostatique $E_{\text{electro},nm}$ entre deux pseudo-atomes n et m devient alors :

$$E_{\text{electro},nm} = 4\pi \sum_{i=1}^I w_i M_4(x_i)^2 \frac{dM_4}{dx}(x_i) \sum_{j=1}^J \varpi_j \frac{e(M_4(x_i), \Omega_j)}{W(x_i)} \quad (2.30)$$

avec I , le nombre de points radiaux et J , le nombre de points angulaires.

On constate que pour tout entier $i \in [1; I]$: $W^{-1}(x_i) = \sqrt{1-x_i^2} = \sin\left(\frac{(2i-1)\pi}{2I}\right)$

L'expression de l'énergie d'interaction électrostatique $E_{\text{electro},nm}$ peut être réécrite :

$$E_{\text{electro},nm} = 4\pi \sum_{i=1}^I w_i M_4(x_i)^2 \sin\left(\frac{(2i-1)\pi}{2I}\right) \frac{dM_4}{dx}(x_i) \sum_{j=1}^J \varpi_j e(M_4(x_i), \Omega_j) \quad (2.31)$$

La partie angulaire est discrétisée selon la méthode de quadrature sphérique de Gauss-Lebedev [28]. Les couples de coordonnées angle solide - pondération (Ω_j, ϖ_j) de la quadrature de Gauss-Lebedev sont disponibles en code FORTRAN pour différents ordres de discrétisation [1]. On peut donner comme exemples d'ordres disponibles : 11 (50 points), 17 (110), 23 (194), 31 (350), 41 (590), 47 (770), 59 (1202).

On notera que, comme pour les méthodes de quadrature de Gauss à une dimension, une méthode de quadrature sphérique est d'ordre J lorsqu'elle permet le calcul exact de l'intégrale sur la sphère unitaire de tout les monômes de types $x^i y^j z^k$ avec $i + j + k \leq J$.

J'ai donc implémenté dans le logiciel VMoPro, appartenant à la suite MoPro [23], l'approche présentée ci-dessus.

Pour déterminer le nombre de points nécessaires à une estimation satisfaisante de l'énergie, l'on considère comme systèmes tests les modèles multipolaires expérimentaux de quatre structures cristallographiques : l'acide salicylique [32], l'alanine [15], le 1-formyl-3-thiosemicarbazide [33] et la sarcosine [17]. Ces modèles multipolaires ont été obtenus et fournis par Dr. Slawomir Domagala dans le cadre de son travail au laboratoire sur la généralisation de la banque de paramètres multipolaires expérimentaux ELMAM [46].

Dans l'ensemble de ces modèles, les atomes lourds C, N, O sont décrits jusqu'au niveau octupolaire, les atomes de soufre au niveau hexadécapolaire. Les atomes d'hydrogène ont pu être affinés jusqu'aux quadripôles en ayant recours à une estimation de l'anisotropie du mouvement thermique de ces derniers à l'aide du serveur SHADE [31] présenté dans la partie I.2.3.D de ce rapport. De plus, les symétries locales des pseudo-atomes ont été maintenues.

On notera bien ici que l'objectif est de tester les performances de l'implémentation de la méthode d'intégration en ayant recours à des modèles multipolaires réalistes de distribution de charges mais pas nécessairement optimaux ce qui ne nuit en rien à cette démarche.

Pour chaque modèle multipolaire cristallographique, nous allons calculer les énergies d'interaction électrostatique des principales interactions entre molécules au sein de l'empilement cristallin.

Disposer d'une méthode, qui permet d'estimer les énergies d'interaction électrostatique avec une précision meilleure que 1 kJ.mol^{-1} , serait intéressant compte tenu qu'une interaction électrostatique entre deux molécules polaires (interaction dipôle-dipôle connue sous le nom de

« force de Keesom ») peut présenter une énergie de l'ordre de 2 kJ.mol⁻¹ ou même parfois moins. De plus, certaines liaisons hydrogènes faibles ayant des énergies de cet ordre ont été recensées dans la littérature. On peut citer comme exemples de liaisons hydrogène faibles : CH₄...OH₂ avec une énergie de -2,5 kJ.mol⁻¹, CH₄...SH₂ avec une énergie de -1,7 kJ.mol⁻¹ et CH₄...FCH₃ avec une énergie de -0,8 kJ.mol⁻¹[14]. Les valeurs analytiques des énergies des interactions électrostatiques ne pouvant pas être déterminées, nous prendrons comme valeurs de référence les valeurs obtenues par intégration numérique à l'aide d'une méthode simple de discrétisation. Cette méthode consiste à discrétiser le volume englobant la distribution d'énergie volumique en une grille régulière de pas fin 0,01 Å et de sommer les énergies des différents volumes élémentaires centrés aux points de la grille sans pondération.

Si l'on considère par exemple le $i^{\text{ème}}$ pseudo-atome de la molécule A et le $j^{\text{ème}}$ pseudo-atome de la molécule B, la discrétisation du volume sur lequel est définie la distribution d'énergies volumiques donne :

$$E_{i,j} = \int \rho_{A,i}(\mathbf{r}) \times V_{B,j}(\mathbf{r}) d\mathbf{r} = v_{\text{element}} \sum_{\mathbf{r} \in \text{grille}} \rho_{A,i}(\mathbf{r}) \times V_{B,j}(\mathbf{r}) \quad (2.32)$$

Cette discrétisation est la plus évidente, et par la même, la moins performante. Cependant, pour un pas très fin (ici 0,01 Å), sa simplicité de démonstration et d'implémentation garantit sa validité au prix de ressources en temps de calculs conséquentes. Cette étape est cependant nécessaire pour optimiser l'intégration sphérique qui sera utilisée plus tard.

Différentes combinaisons de couples coordonnées-pondérations ont été testées : (50,194), (100,590), (100,770), (100,1202), sur les quatre systèmes sélectionnés. Dans ces couples, le premier nombre correspond au nombre I de points radiaux, alors que le second correspond au nombre J de points angulaires. Ces combinaisons de couples coordonnées-pondérations seront par la suite notées « **RISJ** » : **R50S194**, **R100S590**, **R100S770**, **R100S1202**.

Chaque système sera dans un premier temps présenté et les contacts significatifs au sein du cristal seront identifiés avant que les résultats ne soient communiqués. Les illustrations des interactions intermoléculaires au sein des différents systèmes ont été réalisées à l'aide du logiciel Pymol [13] (**Figures 3, 5, 7 et 9**).

L'acide salicylique :

L'acide salicylique (acide 2-hydroxybenzoïque) est un acide carboxylique organique cristallin. Son intérêt pharmaceutique est connu puisqu'il est utilisé dans de nombreux produits tels que les crèmes dermatologiques. Les données expérimentales utilisées ont été obtenues par Munshi *et al.* [32].

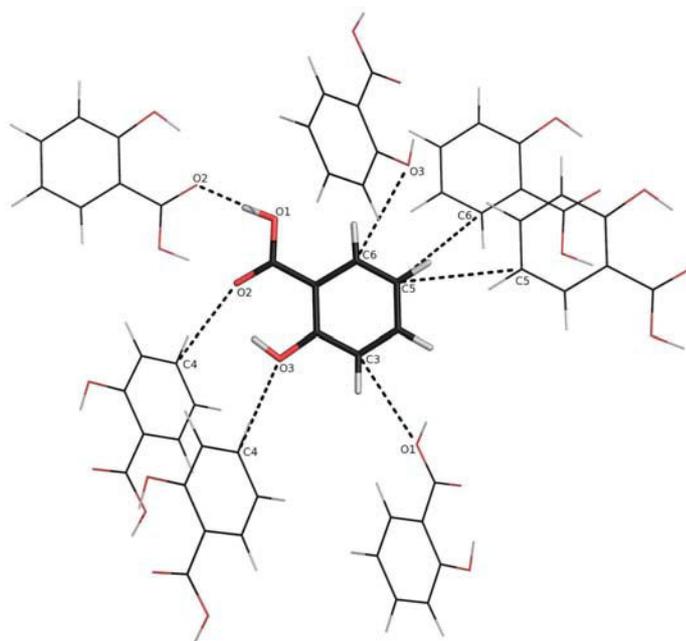


Figure 3 : Diagramme des interactions intermoléculaires au sein dans l'empilement cristallin d'acide salicylique.

Opération de symétrie	Contact le plus court observé entre atomes non hydrogène	Distance interatomique (Å)	Energie (kJ.mol ⁻¹)	Code ORTEP
X-1 ; -Y+1/2 ; Z+1/2	C3-H3...O1	3,299	-5,6	45552
-X ; Y-1/2 ; -Z+1/2	O3...H4-C4	3,622	-12,3	54502
X ; -Y+1/2 ; Z-1/2	C6...O3	3,372	-14,5	55452
-X ; -Y+1 ; -Z	C5...C5	3,863	13,9	56551
-X+1 ; Y-1/2 ; -Z+1/2	O2...C4	3,316	-9,79	64502
-X+1 ; -Y+1 ; -Z	C6...C5	3,964	2,7	66551
-X+2 ; -Y ; -Z	O1-H1...O2	2,648	-169,0	75551

Tableau 1 : Liste des contacts intermoléculaires significatifs au sein du cristal d'acide salicylique. Chaque contact est décrit par l'opérateur de symétrie qui permet de définir le monomère en interaction avec le monomère de référence (X, Y, Z ; code ortep 55501). Les atomes lourds les plus proches les uns et les autres, et impliqués dans l'interaction sont indiqués ainsi que l'atome d'hydrogène s'il s'agit d'une liaison hydrogène. Les distances interatomiques correspondantes sont précisées. L'énergie

d'interaction électrostatique calculée à l'aide de la méthode O1 et prise comme référence est donnée pour chaque contact. Le code ORTEP correspondant est également indiqué. Le groupe d'espace du système est $P2_1/c$.

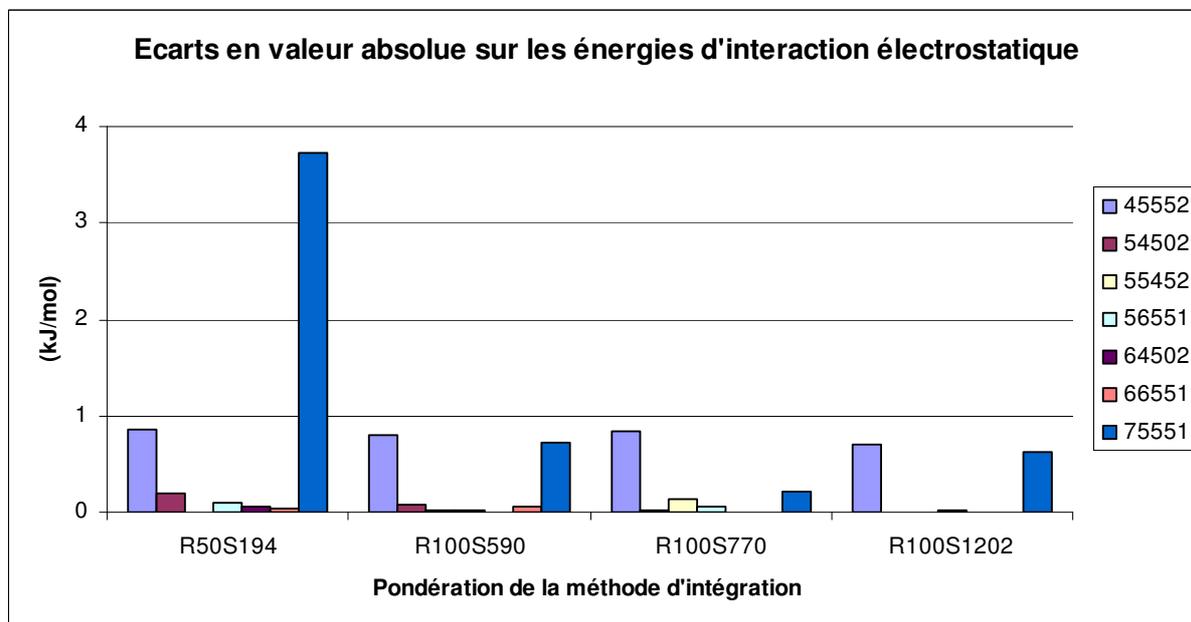


Figure 4 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées selon différents schémas de pondération pour chaque dimère d'acide salicylique. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

L'alanine :

L'alanine est un des 20 acides aminés les plus courants. Il n'est pas essentiel c'est-à-dire qu'il peut être synthétisé par l'organisme. Les données expérimentales utilisées ont été obtenues à haute résolution par Destro *et al.* [15].

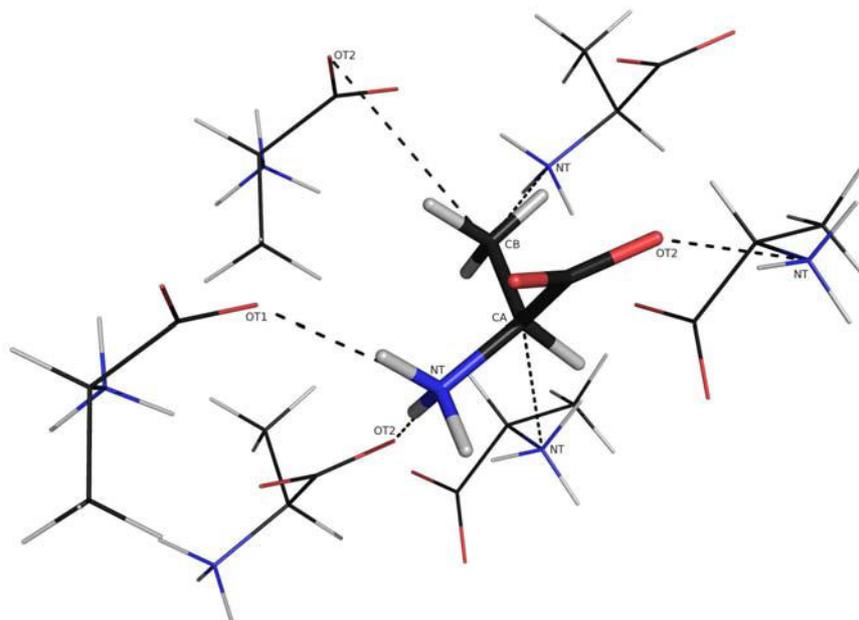


Figure 5 : Diagramme des interactions intermoléculaires au sein dans l'empilement cristallin d'alanine.

Opération de symétrie	Contact le plus court observé entre atomes non hydrogène	Distance interatomique (Å)	Energie (kJ.mol ⁻¹)	Code ORTEP
X-1 ; Y ; Z	CB...NT	3,743	27,7	45501
X-1/2 ; -Y+1/2 ; -Z	CA...NT	4,401	19,3	45503
X-1/2 ; -Y+1/2 ; -Z+1	OT2...H1-NT	2,814	-130,8	45603
X ; Y ; Z-1	NT-H2...OT2	2,792	-121,1	55401
-X+1/2 ; -Y ; Z-1/2	CB...OT2	3,580	8,9	55402
-X+3/2 ; -Y ; Z-1/2	NT-H3...OT1	2,835	-34,7	65402

Tableau 2 : Liste des contacts intermoléculaires significatifs au sein du cristal d'alanine. (Plus d'information cf. **Tableau 1**) Le groupe d'espace du système est P2₁2₁2₁.

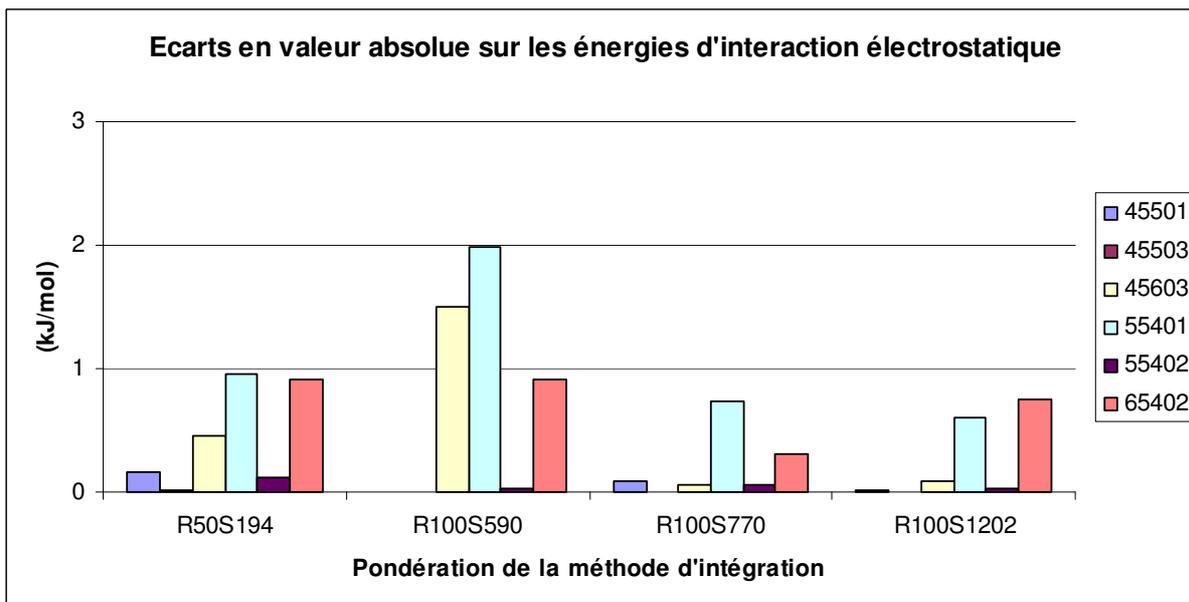


Figure 6 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées selon différents schémas de pondération pour chaque dimère d'alanine. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

Le 1-formyl-3-thiosemicarbazide :

La structure à haute résolution du 1-formyl-3-thiosemicarbazide a été obtenue par Munshi *et al.* [33] dans le cadre de l'étude des liaisons hydrogène.

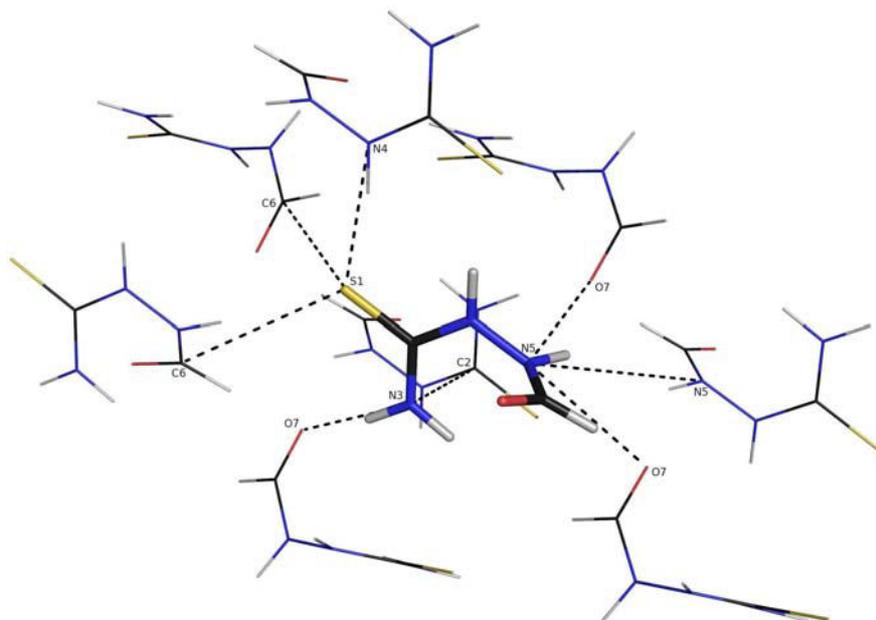


Figure 7 : Diagramme des interactions intermoléculaires au sein de l'empilement cristallin du 1-formyl-3-thiosemicarbazide

Opération de symétrie	Contact le plus court observé entre atomes non hydrogène	Distance interatomique (Å)	Energie (kJ.mol ⁻¹)	Code ORTEP
X-1 ; Y ; Z	S1...C6	3,526	8,1	45501
X ; -Y+3/2 ; Z-1/2	N5-H5...O7	2,906	-8,4	56452
-X+1 ; Y-1/2 ; -Z+3/2	N3-H3A...O7	2,882	-20,3	64602
X-1 ; -Y+3/2 ; Z-1/2	S1...C6	3,545	-78,8	46452
-X+1 ; -Y+1 ; -Z+1	N3...C2	3,226	-5,8	66651
-X+1 ; -Y+2 ; -Z+1	S1...H4-N4	3,277	-33,5	67651
-X+2 ; Y-1/2 ; -Z+3/2	N5...O7	3,361	-15,3	74602
-X+2 ; -Y+1 ; -Z+1	N5...N5	4,836	-0,3	76651

Tableau 3 : Liste des contacts intermoléculaires significatifs au sein du cristal d'1-formyl-3-thiosemicarbazide. (Plus d'information cf. **Tableau 1**) Le groupe d'espace du système est P2₁/c.

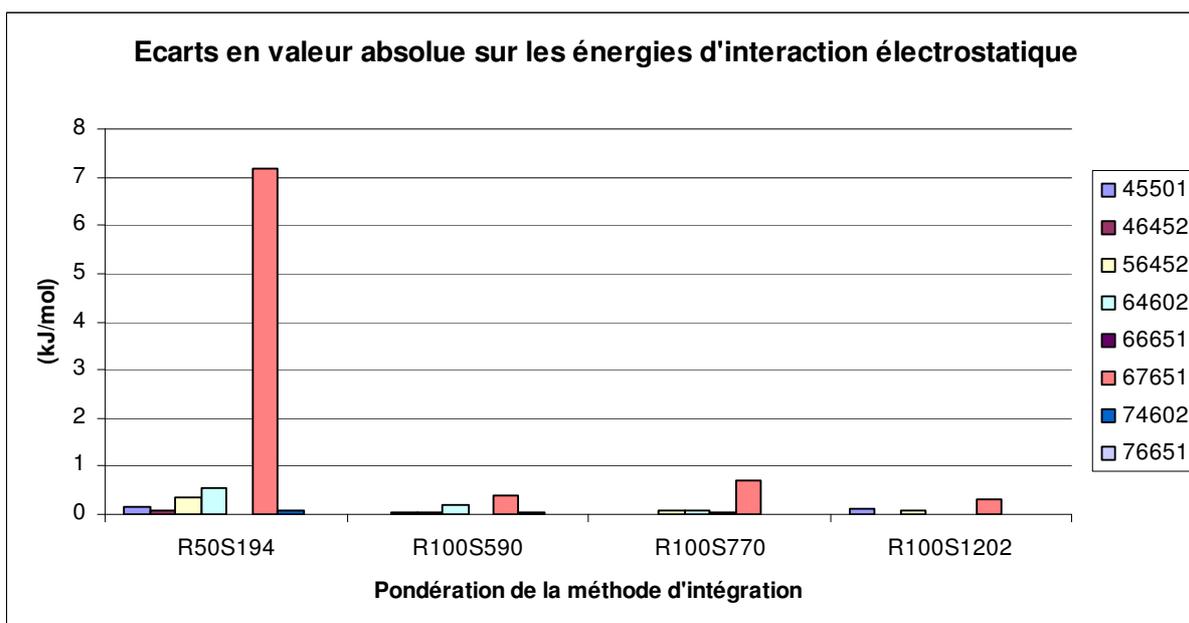


Figure 8 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées selon différents schémas de pondération pour chaque dimère d'1-formyl-3-thiosemicarbazide. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

La sarcosine :

La sarcosine est un acide aminé non-standard qui est le dérivé du N-méthyle de glycine. Elle est présente dans le corps, en particulier dans les muscles. C'est un intermédiaire naturel dans le métabolisme de la glycine à partir de la choline. On la trouve dans les aliments et on l'utilise dans la fabrication de tensioactifs biodégradables. Sa structure à haute résolution a été obtenue par Dittrich *et al.* [17].

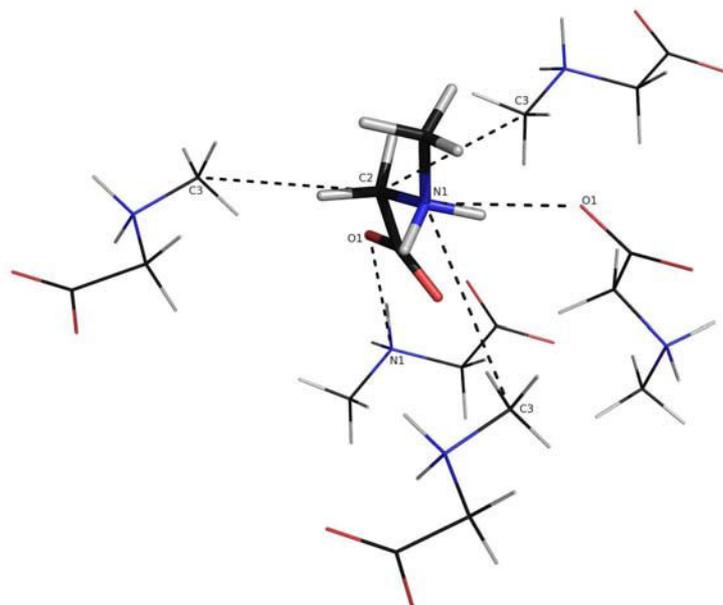


Figure 9 : Diagramme des interactions intermoléculaires au sein de l'empilement cristallin de la sarcosine

Opération de symétrie	Contact le plus court observé entre atomes non hydrogène	Distance interatomique (Å)	Energie (kJ.mol ⁻¹)	Code ORTEP
X-1/2 ; -Y-1/2 ; -Z+1	N1...C3	3,939	-24,1	44604
X-1/2 ; -Y+1/2 ; -Z+1	C2...C3	3,914	-32,3	45604
-X+3/2 ; -Y ; Z-1/2	O1...H2-N1	2,742	-156,8	65402
-X+2 ; Y-1/2 ; -Z+1/2	N1-H1...O1	2,768	-168,7	74503
-X+5/2 ; -Y ; Z-1/2	C1...C3	3,598	2,5	75402

Tableau 4 : Liste des contacts intermoléculaires significatifs au sein du cristal de sarcosine. (Plus d'information cf. **Tableau 1**). Le groupe d'espace du système est P2₁2₁2₁.

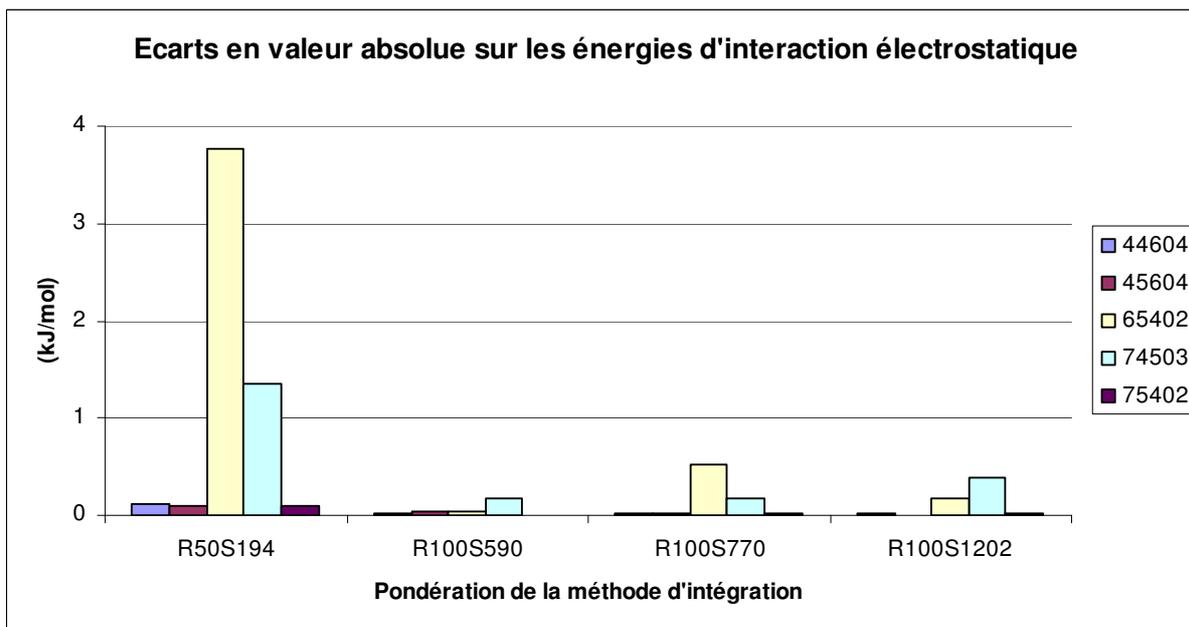


Figure 10 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées selon différents schémas de pondération pour chaque dimère de sarcosine. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

Cette méthode d'intégration en coordonnées sphériques a été employée par Volkov *et al.* pour l'intégration des énergies d'interaction électrostatique en utilisant 50 points radiaux pour 194 points angulaires (**R50S194**). Les tests réalisés après implémentation de la méthode dans le logiciel VMoPro tendent à montrer que ce nombre de points peut ne pas être suffisant pour garantir une précision de 1 kJ.mol^{-1} .

L'examen des histogrammes des écarts en valeur absolue des énergies d'interaction électrostatique calculées par intégration sphérique par rapport aux valeurs de référence, montrent en effet que les écarts peuvent être nettement supérieurs à 1 kJ.mol^{-1} pour les méthodes **R50S194** et **R100S590** (**Figures 4, 6, 8 et 10**). Pour ces méthodes, on trouve pour chacun des systèmes tests, un ou deux dimères de molécules en interaction pour lesquels les écarts d'énergie sont supérieurs à 1 kJ.mol^{-1} . On constate par ailleurs que l'écart sur l'énergie est corrélé à la valeur absolue de l'énergie de référence.

Les méthodes **R100S770** et **R100S1202** donnent des résultats satisfaisants, avec dans tous les cas, des écarts inférieurs au kJ.mol^{-1} .

Le calcul de l'énergie étant décomposé en somme d'énergie d'interaction électrostatique entre deux pseudo-atomes, le temps de calcul est directement proportionnel au nombre d'interactions entre pseudo-atomes.

On peut noter qu'en considérant la méthode (**R50S194**) comme référence, la méthode (**R100S590**) prend 5 fois plus de temps, la méthode (**R100S770**) 6 fois plus et la méthode (**R100S1202**) 10 fois plus. Une telle augmentation du temps de calcul est prévisible, ce dernier dépendant essentiellement du nombre de points où la densité d'énergies est évaluée.

Par exemple, entre les méthodes (**R50S194**) et (**R100S1202**), le nombre de points où la densité d'énergies est estimée, est multiplié par 12 environ.

On retiendra, par précaution, la méthode d'intégration sphérique (**R100S1202**) par la suite. Bien que cette méthode prenne plus de temps, la durée des calculs demeure en moyenne de l'ordre de la minute pour les quatre cas tests considérés.

II.2.2.B. Calcul des énergies par la méthode mixte Quadrature/Buckingham

La méthode d'intégration par quadrature en coordonnées sphériques représente une amélioration significative qui permet une estimation des énergies d'interaction électrostatique en un temps très raisonnable. Il est néanmoins intéressant d'utiliser cette méthode conjointement avec la décomposition des énergies d'interaction électrostatique en somme de produits de moments atomiques appelée communément « la sommation de Buckingham » [9].

Cette méthode permet d'exprimer l'énergie d'interaction électrostatique entre deux atomes comme une somme de produits des moments sans trace de leurs distributions de charges, mais n'est valide qu'à la condition qu'il n'y ait pas recouvrement entre ces distributions de charges.

$$E_{\text{electro}} = \sum_{l=0}^{\infty} \sum_{l'=0}^{\infty} (-1)^{l'} \frac{1}{1.3.5\dots(2l-1)} \frac{1}{1.3.5\dots(2l'-1)} \times T_{\alpha\beta\gamma\dots\nu\alpha'\beta'\gamma'\dots\nu'} M_{\alpha\beta\gamma\dots\nu}^{A(l)} M_{\alpha\beta\gamma\dots\nu\alpha'\beta'\gamma'\dots\nu'}^{B(l')} \quad (2.33)$$

avec $M^{A(l)}$ et $M^{B(l')}$ sont les moments atomiques sans trace l et l' des atomes A et B respectivement, et $\alpha, \beta, \gamma, \dots$ pouvant être x, y ou z .

Cette expression (2.33) étant fonction des moments atomiques des distributions de charges, elle dépend indirectement du formalisme de décomposition de la distribution électronique retenu. On peut, en effet, appliquer l'approximation de Buckingham selon différents schémas de partitionnement de la distribution de charges. Les moments atomiques obtenus selon les formalismes A.I.M. (Atoms In Molecules) [4], et D.M.A. (Distributed Multipole Analysis) [39] ont été utilisés pour le calcul des énergies par Popelier *et al.* [37].

D'après Volkov *et al.* [41], le partitionnement le mieux adapté en terme de convergence mathématique est la partition A.I.M. Cependant, l'objectif ici est de disposer d'une méthode satisfaisant le critère de précision souhaité pour un temps de calcul minimal. En terme de temps de calcul, c'est le partitionnement en pseudo-atomes qui est le plus approprié. En effet, la sommation de Buckingham est définie à l'aide des moments atomiques sans trace. Or, les moments internes d'un pseudo-atome peuvent être aisément obtenus analytiquement à partir des paramètres multipolaires (P_{lmp}, k') de ce même pseudo-atome [11]. De plus, les moments sans trace (ou moments externes) peuvent être déduits analytiquement des moments internes.

Le moment atomique interne Θ_{lmp} correspondant à la contribution d'une harmonique sphérique y_{lmp} défini dans le repère local du pseudo-atome considéré est donné par :

$$\Theta_{lmp} = P_{lmp} \frac{1}{(k'\zeta)^l} \frac{(n(l)+l+2)!}{(n(l)+2)!} \frac{1}{D_{lm}M_{lm}} \quad (2.34)$$

avec D_{lm} et M_{lm} coefficients définis pour chaque couple (l, m) .

Pour les moments d'ordre $l \geq 2$, les moments sans trace ou « externe » de la distribution de charges d'un pseudo-atome peuvent être exprimés comme des combinaisons linéaires des moments atomiques internes.

On notera qu'un moment sans trace d'ordre l est défini comme une combinaison de moments internes d'ordre l .

Les moments atomiques quadripolaires sans trace sont définis en fonction des moments atomiques internes des différents termes multipolaires du pseudo-atome, comme :

$$\Theta_{zz} = \frac{1}{2}\Theta_{20}, \quad \Theta_{xx} = \frac{1}{2}\left(3\Theta_{22+} - \frac{1}{2}\Theta_{20}\right), \quad \Theta_{yy} = \frac{1}{2}\left(-3\Theta_{22+} - \frac{1}{2}\Theta_{20}\right), \quad (2.35)$$

$$\Theta_{xz} = \frac{3}{2}\Theta_{21+}, \quad \Theta_{yz} = \frac{3}{2}\Theta_{21-}, \quad \Theta_{xy} = \frac{3}{2}\Theta_{22-}$$

Les autres moments sans traces sont communiqués en **Annexe B** de ce rapport.

Dans un cas général, la sommation de Buckingham est mathématiquement infinie. Cependant, dans le cas de l'énergie d'interaction entre deux pseudo-atomes, son ordre ne peut dépasser la somme des ordres de développement multipolaire des pseudo-atomes.

De plus, Volkov *et al.* [42] ont montré qu'un développement jusqu'aux termes d'énergie d'ordre 4 de la sommation était suffisant pour obtenir une bonne estimation de l'énergie d'interaction.

L'expression devient alors :

$$\begin{aligned}
E_{electro} = & Tq^A q^B + T_{\alpha} (q^A \mu_{\alpha}^B - q^B \mu_{\alpha}^A) + T_{\alpha\beta} \left(\frac{1}{3} q^A \Theta_{\alpha\beta}^B + \frac{1}{3} q^B \Theta_{\alpha\beta}^A - \mu_{\alpha}^A \mu_{\beta}^B \right) \\
& + T_{\alpha\beta\gamma} \left(\frac{1}{15} q^A \Omega_{\alpha\beta\gamma}^B - \frac{1}{15} q^B \Omega_{\alpha\beta\gamma}^A - \frac{1}{3} \mu_{\alpha}^A \Theta_{\beta\gamma}^B + \frac{1}{3} \mu_{\alpha}^B \Theta_{\beta\gamma}^A \right) \\
& + T_{\alpha\beta\gamma\delta} \left(\frac{1}{105} q^A \Phi_{\alpha\beta\gamma\delta}^B + \frac{1}{105} q^B \Phi_{\alpha\beta\gamma\delta}^A - \frac{1}{15} \mu_{\alpha}^A \Omega_{\beta\gamma\delta}^B - \frac{1}{15} \mu_{\alpha}^B \Omega_{\beta\gamma\delta}^A + \frac{1}{9} \Theta_{\alpha\beta}^A \Theta_{\gamma\delta}^B \right) \quad (2.36)
\end{aligned}$$

Les tenseurs symétriques d'interaction $T_{\alpha\beta\gamma\dots}$ sont définis par :

$$T_{\alpha\beta\gamma\dots} = \frac{1}{4\pi\epsilon_0} \nabla_{\alpha} \nabla_{\beta} \nabla_{\gamma} \dots \mathbf{r}^{-1}$$

avec \mathbf{r} , le vecteur distance défini dans le repère global.

La convention d'Einstein est utilisée pour la sommation sur les indices $\alpha, \beta, \gamma, \dots$

Remarque : On notera que parler de molécules ou groupes chimiques en interaction est un abus de langage puisqu'il faudrait utiliser en toute rigueur les termes « pseudo-molécules » ou « pseudo-groupes chimiques » par extension de la dénomination « pseudo-atome ». Les distributions de charges que l'on considère en interaction sont des ensembles de pseudo-atomes, et sont donc des molécules définies selon le schéma de partitionnement de Hansen & Coppens [22].

L'approximation de Buckingham n'est valable que s'il n'y a pas de recouvrement des distributions de charges des molécules ou groupes chimiques dont on calcule l'énergie d'interaction électrostatique.

Dans le programme XD développé à l'université de Buffalo (Volkov *et al.*) [42], la méthode de calcul des énergies d'interaction implémentée est la méthode EP/MM (Exact Potentiel/Multipole Method) qui utilise conjointement l'approximation de Buckingham pour les interactions à longue portée et l'intégration par méthode de quadrature en coordonnées sphériques pour les interactions à courte portée (recouvrement des distributions de charges en

interaction). Pour deux pseudo-atomes en interaction, le choix de l'une ou l'autre méthode est fait en fonction de la distance entre ces derniers. Si la distance interatomique est supérieure à une certaine distance « limite », l'approximation de Buckingham est utilisée sinon l'intégration par méthode de quadrature est réalisée.

Cette distance « limite » est ajustable par l'utilisateur dans le programme XD en fonction de ses besoins. On notera par ailleurs que dans les différentes études ayant nécessité des calculs d'énergies d'interaction à l'aide du logiciel XD, la « limite » choisie change.

A l'aide des différents systèmes tests précédemment utilisés, l'impact de la distance « limite » a été mis en évidence (**Figures 11, 12, 13, 14**). Selon le système considéré, la convergence est obtenue pour une « limite » plus ou moins élevée.

Pour les systèmes : acide salicylique, l'alanine et la sarcosine (**Figures 12, 14**), la convergence est globalement obtenue pour une limite de 4,0 Å, alors que pour le 1-formyl-3-thiosemicarbazide une limite de 5,0 Å est nécessaire (**Figure 13**).

De plus, pour un système donné, la convergence est obtenue pour une « limite » plus ou moins élevée selon l'interaction électrostatique considérée. Par exemple, pour le 1-formyl-3-thiosemicarbazide, la convergence n'est obtenue que pour une « limite » de 5,0 Å pour les énergies d'interaction électrostatique impliquant les dimères correspondant aux opérateurs de symétrie X-1 ; -Y+3/2 ; Z-1/2 (64602) et -X+1 ; -Y+1 ; -Z+1 (66651), alors qu'une « limite » de 3,0 Å est suffisante pour le dimère d'opérateurs de symétrie -X+2 ; Y-1/2 ; -Z+3/2 (74602) et -X+2 ; -Y+1 ; -Z+1 (76651). On notera à ce sujet que l'interaction entre les monomères d'opérateurs de symétrie X ; Y ; Z (55501) et X-1 ; -Y+3/2 ; Z-1/2 (64602) correspond à la liaison hydrogène N3-H3A...O7 (**Figure 7**) avec une distance interatomique entre les atomes N3 et O7 (distance donneur-accepteur) de 2,882 Å (**Tableau 3**), ce qui correspond au contact intermoléculaire le plus court dans le cristal. Une telle distance intermoléculaire justifie la nécessité d'une distance interatomique élevée. L'interaction entre les monomères d'opérateurs de symétrie X ; Y ; Z (55501) et -X+1 ; -Y+1 ; -Z+1 (66651) correspond quant à elle à un contact court (**Figure 7**) avec une distance interatomique entre les atomes N3 et C2 de 3,226 Å (**Tableau 3**). L'énergie d'interaction électrostatique correspondante est faible avec -5,8 kJ.mol⁻¹. La nécessité d'une distance limite élevée pour obtenir un écart raisonnable peut s'expliquer simplement par sa faible valeur. En effet, l'ensemble des pseudo-atomes des molécules va contribuer à cette énergie contrairement au cas d'une interaction par liaison hydrogène où l'énergie d'interaction électrostatique entre les pseudo-atomes impliqués dans la liaison hydrogène est prédominante.

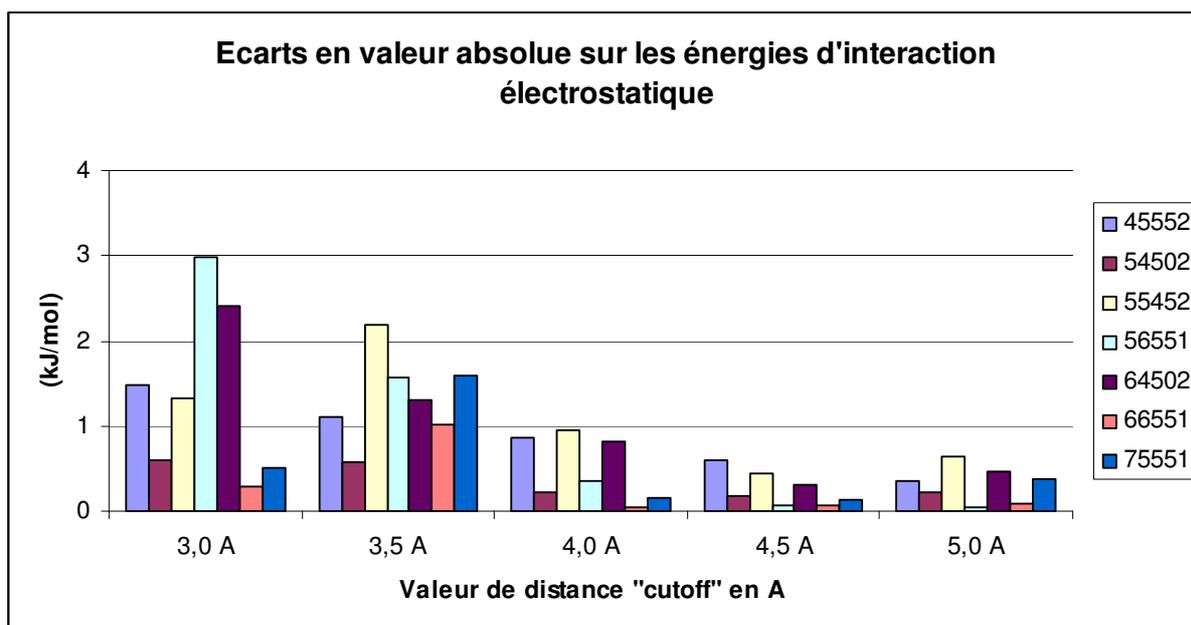


Figure 11 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées avec différentes valeurs de distance « limite » pour l'acide salicylique. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

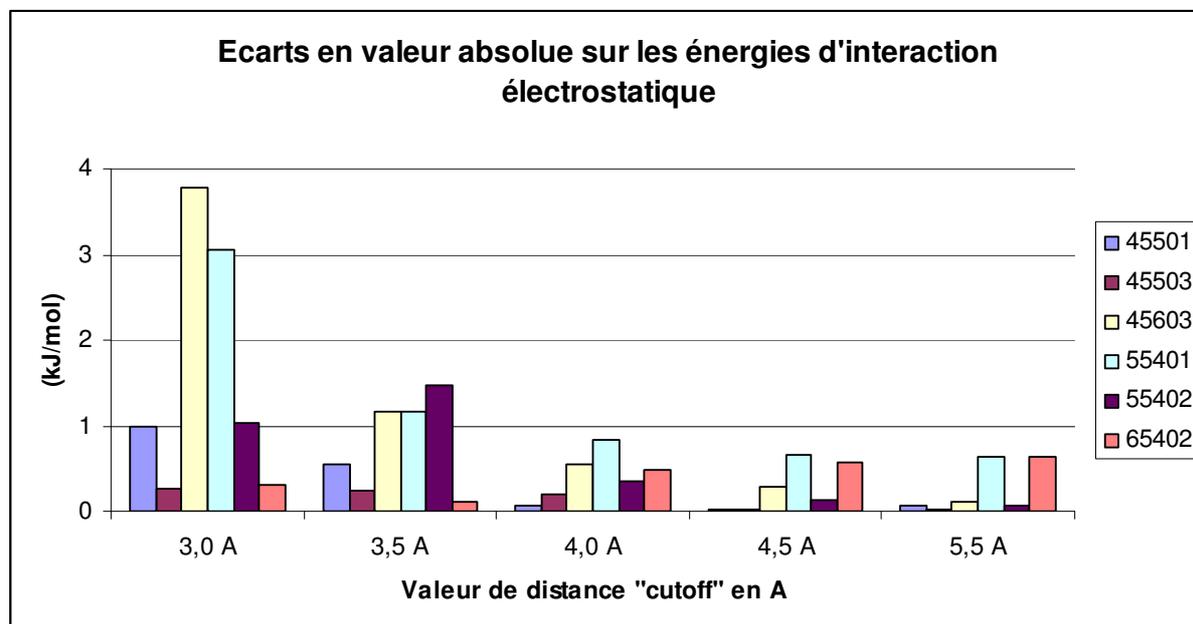


Figure 12 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées avec différentes valeurs de distance « limite » pour l'alanine. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

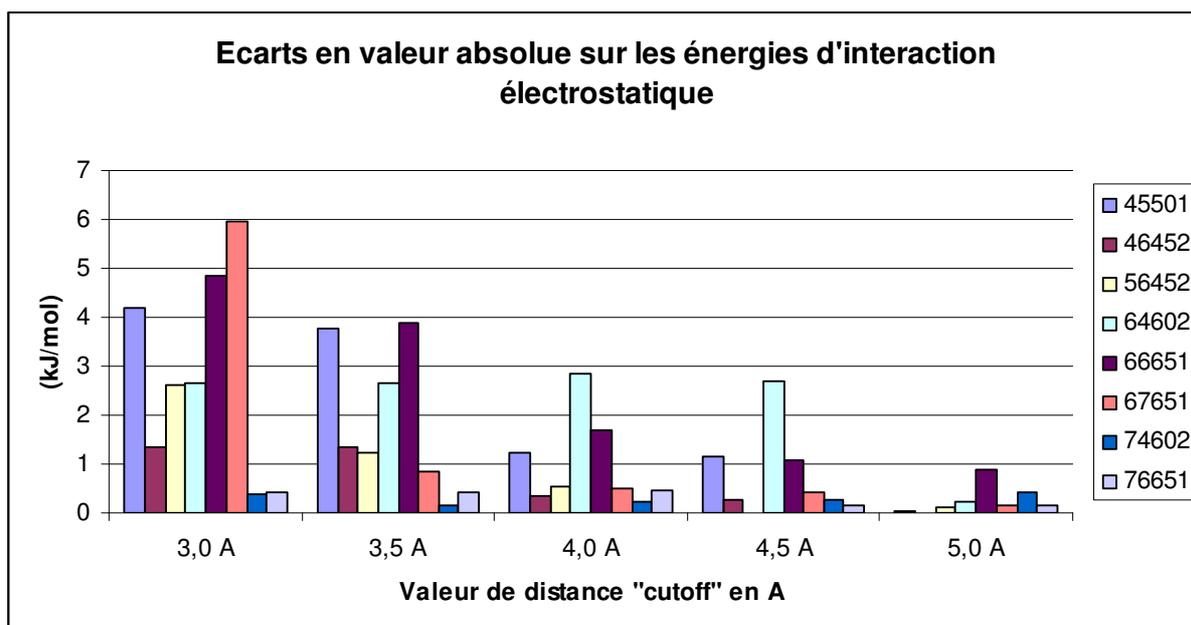


Figure 13 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées avec différentes valeurs de distance « limite » pour le 1-formyl-3-thiosemicarbazide. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

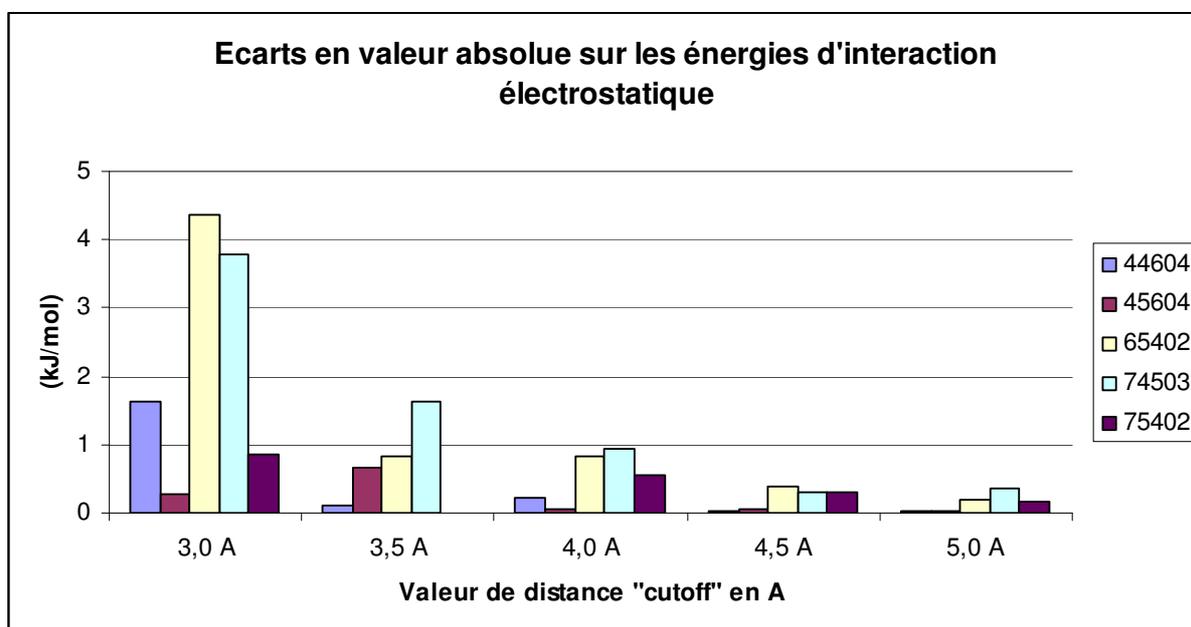


Figure 14 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées avec différentes valeurs de distance « limite » pour la sarcosine. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

II.2.2.C. Méthode mixte d'intégration révisée

Volkov *et al.* suggèrent la tabulation de rayons définis pour chaque élément chimique afin de déterminer une distance « limite » optimisée en fonction du couple de pseudo-atomes en interaction. La distance limite pour le calcul de l'énergie d'interaction entre un pseudo-atome d'azote et un pseudo-atome de carbone serait alors la somme des rayons tabulés R_C et R_N . Cette méthode bien qu'astucieuse présente le désavantage de nécessiter une définition de rayons pour chaque élément chimique ou tout du moins les plus fréquents, et néglige l'environnement du pseudo-atome qui peut présenter une expansion plus ou moins importante de son cortège électronique en fonction de sa charge.

Une autre solution a été retenue lors de l'implémentation de cette méthode d'intégration mixte dans le logiciel VMOPro dans le but d'éviter une modification de la distance « limite » en fonction des systèmes étudiés. Au lieu de déterminer au préalable des rayons « limite » par élément chimique afin d'optimiser la distance « limite » en fonction des pseudo-atomes en interaction, ces rayons « limite » sont calculés préliminairement pour chaque pseudo-atome participant à l'interaction.

Le critère de définition du rayon « limite » d'un pseudo-atome part du critère de validité de l'approximation de Buckingham : Cette approximation n'est valable que si les distributions de charges en interaction ne présentent pas de recouvrement.

Afin de définir des rayons « limite » atomiques permettant de satisfaire cette condition, l'on définit le rayon « limite » d'un atome de manière à garantir la quasi-nullité de sa densité électronique à une distance égale ou supérieure à ce rayon « limite ». La distribution électronique d'un pseudo-atome étant asphérique, le rayon « limite » de chaque pseudo-atome est déterminé en fonction d'une fonction radiale qui majore l'expression de la densité électronique du pseudo-atome.

Rappelons que, selon le formalisme de Hansen & Coppens, la densité électronique d'un pseudo-atome est donnée par :

$$\rho_{electron}(\mathbf{r}) = \rho_{core,sph}(r) + k^3 P_{val} \rho_{val,sph}(kr) + \sum_{l=0}^{l_{max}} k^{l^3} R_l(k'r) \sum_{m=0}^l P_{lm\pm} y_{lm\pm}(\theta, \varphi) \quad (2.37)$$

Cette densité électronique peut donc être majorée par la fonction radiale suivante :

$$\rho_{electron}(\mathbf{r}) \leq \rho_{core,sph}(r) + k^3 P_{val} \rho_{val,sph}(kr) + \sum_{l=0}^{l_{max}} k^{l^3} R_l(k'r) \sum_{m=0}^l |P_{lm\pm}| \max(|y_{lm\pm}(\theta, \varphi)|) \quad (2.38)$$

Cette fonction majorante est une somme de fonctions exponentielles décroissantes et donc, par définition, décroissante.

Les valeurs maximales des fonctions harmoniques sphériques $Y_{lm\pm}$ en valeur absolue $\max|Y_{lm\pm}(\theta, \varphi)|$ dont dépend la fonction majorante, ont été tabulées dans le logiciel VMoPro lors de l'implémentation de la méthode de calcul numérique de l'énergie d'interaction électrostatique.

Ainsi, lors du calcul de l'énergie d'interaction électrostatique entre deux distributions de charges, une étape préliminaire est réalisée. Cette étape consiste en l'évaluation des rayons « limite » de chaque pseudo-atome impliqué dans l'interaction.

Ce rayon est la distance minimale au noyau du pseudo-atome à laquelle la fonction majorante radiale est inférieure à un seuil de densité électronique ρ_{\min} correspondant à une densité électronique négligeable.

L'évaluation de ce rayon pour chaque pseudo-atome peut paraître contraignant. Cependant, il faut garder à l'esprit que cette évaluation n'est réalisée qu'une fois par pseudo-atome et que le problème de détermination du rayon « limite » est un problème à une dimension.

Plusieurs seuils de densité électronique ρ_{\min} ont été testés sur les systèmes tests : $0,05 \text{ e.}\text{\AA}^{-3}$, $0,01 \text{ e.}\text{\AA}^{-3}$, $0,005 \text{ e.}\text{\AA}^{-3}$, $0,001 \text{ e.}\text{\AA}^{-3}$ et $0,0005 \text{ e.}\text{\AA}^{-3}$. Les résultats montrent qu'une densité électronique minimale de $0,001 \text{ e.}\text{\AA}^{-3}$ permet d'avoir la précision désirée (**Figures 15, 16, 17, 18**).

Le temps de calcul moyen pour l'ensemble des énergies d'interaction électrostatique au sein des systèmes tests est d'environ 10 secondes, et le temps maximal observé de 25 secondes.

Il est intéressant de noter que la valeur limite maximale de densité électronique retenue comme satisfaisante de $0,001 \text{ e.}\text{\AA}^{-3}$ est inférieure à la valeur limite de densité électronique de $10^{-3} \text{ e.Bohr}^{-3}$ (soit $6,75 \cdot 10^{-3} \text{ e.}\text{\AA}^{-3}$) utilisée usuellement en chimie quantique dans les calculs ab initio en phase gazeuse pour définir l'isosurface moléculaire de densité électronique servant à identifier les contributions significatives à la distribution de charges du système moléculaire étudié, et donc à en délimiter les frontières.

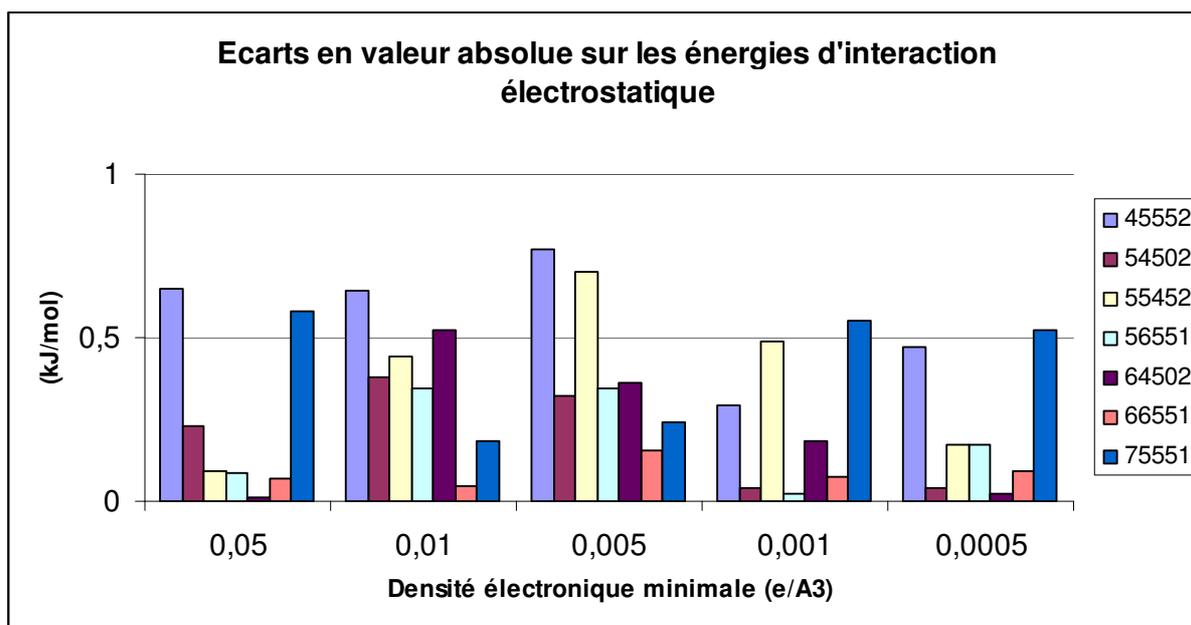


Figure 15 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées avec différentes valeurs de densité électronique minimale pour l'acide salicylique. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

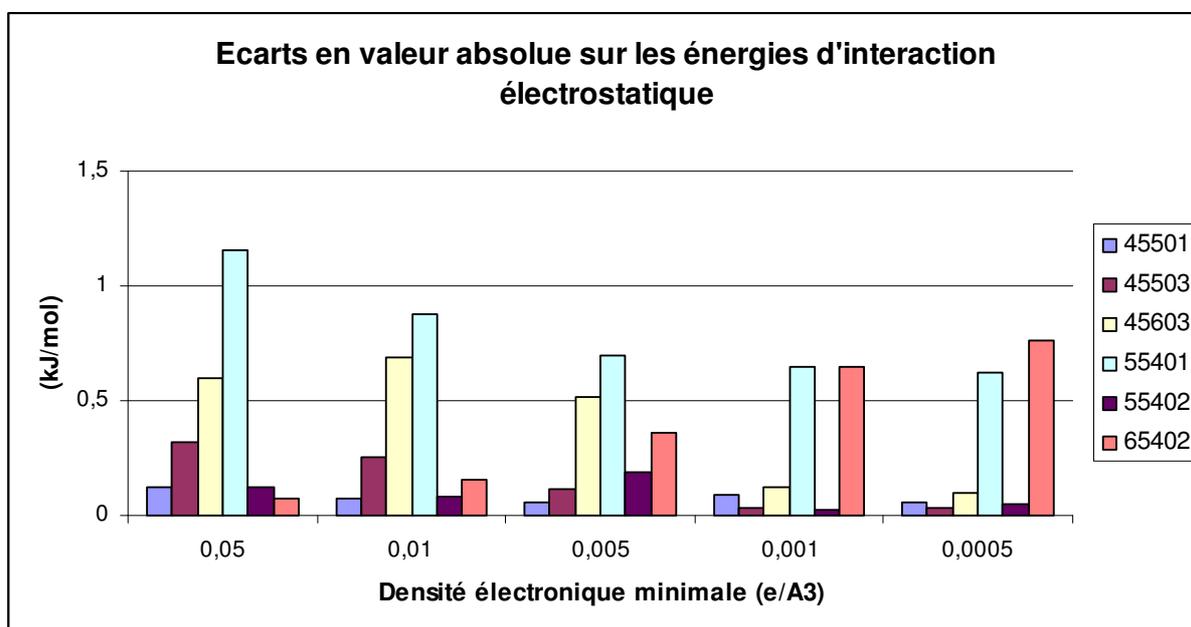


Figure 16 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées avec différentes valeurs de densité électronique minimale pour l'alanine. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

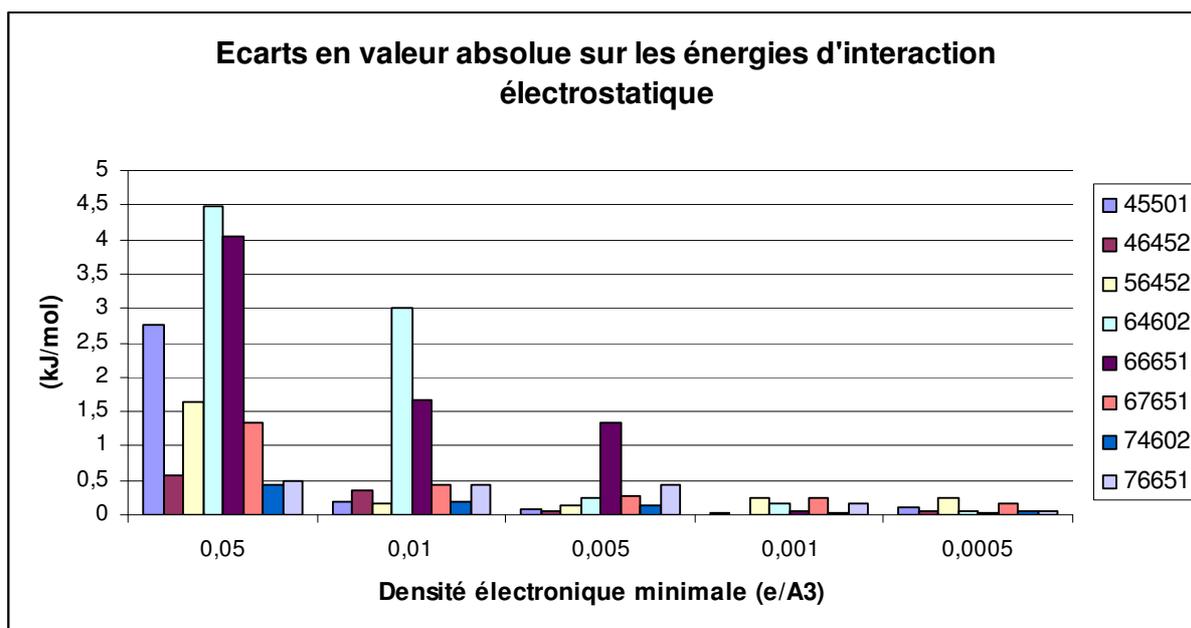


Figure 17 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées avec différentes valeurs de densité électronique minimale pour le 1-formyl-3-thiosemicarbazide. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

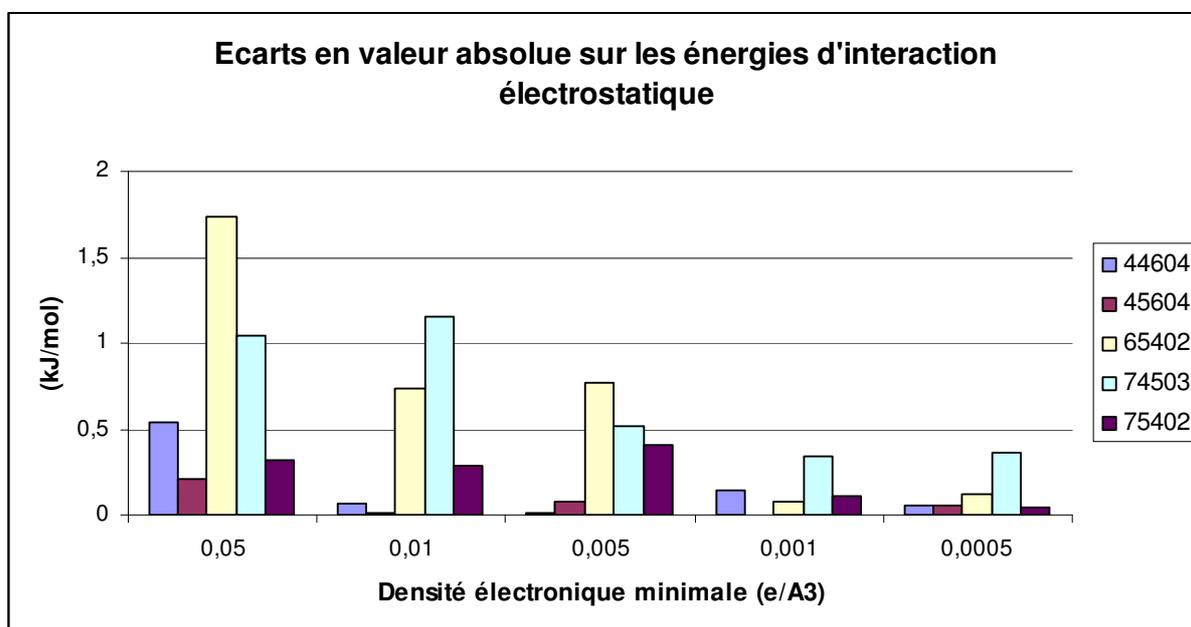


Figure 18 : Histogrammes des écarts absolus aux valeurs de référence des énergies d'interaction électrostatique calculées avec différentes valeurs de densité électronique minimale pour la sarcosine. Chaque contact intermoléculaire est identifié à l'aide du code ORTEP correspondant.

II.2.3. La banque de paramètres multipolaires, ELMAM

II.2.3.A. La transférabilité des paramètres multipolaires

Les progrès techniques et méthodologiques réalisés en biocristallographie ont permis d'atteindre pour des systèmes macromoléculaires des résolutions atomiques et même subatomiques qui étaient jusqu'alors réservées aux petites molécules. Une résolution subatomique semble la promesse d'une appréhension plus précise de l'affinement des macrosystèmes à l'aide d'un modèle plus sophistiqué que le modèle d'atomes indépendants (I.A.M.), tel que le modèle multipolaire (partie I.2.2. de ce manuscrit). Cependant, même avec des données de qualité, les systèmes macromoléculaires souffrent de certains handicaps. Même si une haute résolution peut être obtenue, elle demeure limitée en comparaison aux résolutions atteignables dans les études des petites molécules, l'agitation thermique et le désordre sont plus importants, et la méthode d'affinement par moindres carrés en raison de son rayon de convergence limité peut rencontrer des difficultés à guider le modèle de distribution de charges non sphérique vers un état final satisfaisant. Tous ces facteurs nuisent à l'affinement d'un modèle de distribution de charges non-sphérique. Afin de faciliter un tel affinement, il convient de débiter ce dernier à partir d'un modèle de distribution de charges le plus proche possible de l'état final attendu [21].

Il est possible de répondre à ce besoin, en se fondant sur le principe de transférabilité. Ce principe postule que la déformation de la densité de charges dans un groupement chimique résulte majoritairement des liaisons chimiques covalentes, et est, par conséquent, préservée quelque soit le système considéré et son environnement chimique. En termes d'atomes ou plus précisément de pseudo-atomes dans le formalisme de Hansen & Coppens, deux pseudo-atomes de même nature présentant des environnements covalents similaires ainsi que la même stéréochimie, auront des distributions de charges proches.

Fort de ce principe, plusieurs banques de paramètres multipolaires de distribution de charges ont été conçues. La première d'entre elles fût la banque ELMAM développée dans notre laboratoire [46;35]. Cette banque est composée des paramètres moyennés de distribution de charges, définis selon le formalisme de Hansen & Coppens, de l'ensemble des atomes rencontrés dans les acides amines constituant les protéines. Elle fût définie à partir de modèles multipolaires de densité de charges de petits peptides ou molécules organiques.

Puis, la première banque théorique de paramètres multipolaires (UBCD) fût proposée à partir de calculs quantiques [27]. Une seconde banque théorique pour les affinements structuraux d'acides aminés, d'oligopeptides et protéines fût publiée par Dittrich *et al.*[16]. Développées pour répondre aux mêmes besoins que la banque ELMAM, ces banques théoriques contiennent également des paramètres multipolaires (moyennés dans le cas de la banque UBCD), mais obtenus par affinement contre des facteurs de structures théoriques. Dans les travaux présentés dans ce manuscrit, seule la banque expérimentale ELMAM a été employée.

II.2.3.B. La banque ELMAM (Experimental Library of Multipolar Atom Model)

La banque ELMAM originale est composée des paramètres multipolaires moyennés décrivant les distributions de charges des 20 acides aminés rencontrés dans les protéines [46;35]. Ces paramètres moyens furent calculés à partir des modèles de distribution de charges de plusieurs petits peptides et acides aminés, obtenus par affinement contre des données rayons X à très haute résolution.

Le transfert des paramètres depuis la librairie ELMAM est pris en charge par le programme IMoPro appartenant à la suite MoPro, pour l'obtention de modèle de distribution de charges de structures mesurées à résolution atomique.

Garder les paramètres multipolaires transférés fixes pendant que les paramètres atomiques de positions et d'agitation sont affinés, permet une meilleure déconvolution des déplacements thermiques et de distribution de charges, même quand les données de diffraction n'atteignent pas la résolution subatomique [24]. Après transfert, les statistiques cristallographiques d'affinement sont améliorées, en particulier les facteurs d'accord libres [8], et de plus les cartes de densité électronique résiduelle sont plus propres. (définitions données dans la partie I.2.4.B) La première tentative d'affinement multipolaire sur un composé polypeptidique avec définition d'un état initial par transfert de paramètres, fut réalisée avec succès sur l'octapeptide LBZ [24].

II.2.3.C. Généralisation de banque ELMAM

La version originale de la banque ELMAM est composée des différents acides aminés ainsi que de quelques groupements chimiques usuellement rencontrés dans les macromolécules

organiques. Afin d'élargir le champ d'application de la banque expérimentale des protéines aux groupements chimique usuels, une version généralisée de cette dernière définie à partir des modèles multipolaires expérimentaux de 50 petites molécules, est en construction. Cette banque a pour ambition de permettre le transfert des paramètres multipolaires de distribution de charges à partir de la structure d'une molécule par identification automatique des types atomiques des atomes la constituant.

Cette nouvelle version de la banque ELMAM est développée en utilisant des repères locaux optimisés pour chaque pseudo-atome de manière à respecter au mieux la symétrie naturelle de sa distribution de charges (partie I.2.2.B).

C'est cette version de la banque expérimentale ELMAM qui a été utilisée dans les travaux présentés dans ce manuscrit.

Les différents outils et méthodes d'analyse présentés dans ce chapitre ont été utilisés dans les travaux exposés dans les chapitres 3 et 4 de ce manuscrit.

- [1] <http://www.ccl.net/cca/software/sources/fortran/lebedev-laikov-grids/>.
- [2] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel. *Chemical Physics Letters*, **162**(3), 165–169, 1989.
- [3] R. F. W. Bader. *Journal of Chemical Physics*, **73**, 2871–2883, 1980.
- [4] R. F. W. Bader. *Atoms in Molecules: a Quantum Theory*. 1990.
- [5] R. F. W. Bader, Anderson S. G., Duke A. J. *Journal of American Chemistry Society*, **101**, 1389–1395, 1979.
- [6] R. F. W. Bader, Nguyen-Dang T. T., Tal Y. *Journal of Chemical Physics*, **70**, 4316–4329, 1979.
- [7] A. D. Becke. *Journal of Chemical Physics*, **88**, 2547, 1988.
- [8] A. T. Brünger. *Nature (London)*, **355**, 472–475, 1992.
- [9] A. D. Buckingham. *Advances in Chemical Physics*, **12**, 107, 1967.
- [10] G. Chessari, C. A. Hunter, C. M. R. Low, M. J. Packer, J. G. Vinter, C. Zonta. *Chemistry: A European Journal*, **8**, 2860–2867, 2002.
- [11] P. Coppens. *X-ray Charge Densities and Chemical Bonding*. New York: Oxford University Press, 1997.
- [12] S. R. Cox, L. Y. Hsu, D. E. Williams. *Acta Crystallographica*, **A37**, 293–301, 1981.
- [13] W. L. DeLano. The pymol molecular graphics system, 2002 on world wide web <http://www.pymol.org>. Technical report.
- [14] G. R. Desiraju & T. Steiner. *The weak hydrogen bond in structural chemistry and biology*. Oxford University Press, 1999.
- [15] R. Destro, R. E. Marsh, R. Bianchi. *Journal of Physical Chemistry*, **92**, 966–973, 1988.
- [16] B. Dittrich, C. B. Hübschle, P. Luger, M. A. Spackman. *Acta Crystallographica*, **D62**, 1325–1335, 2006.
- [17] B. Dittrich & M. A. Spackman. *Acta Crystallographica*, **A63**, 426–436, 2007.
- [18] N. E. Ghermani, N. Bouhmaida, C. Lecomte. *Internal report UPRES A CNRS 7036, Université Henri Poincaré, Nancy I, France*, 1992-1998.
- [19] R. J. Gillespie & R. S. Nyholm. *Quarterly Reviews, Chemical Society*, **11**(4), 339–380, 1957.
- [20] N. Gresh, G. A. Cisneros, T. A. Darden, J.-P. Piquemal. *Journal of Chemical Theory and Computation*, **3**, 1960–1986, 2007.
- [21] B. Guillot, C. Jelsch, A. Podjarny, C. Lecomte. *Acta Crystallographica*, **D64**, 567–588, 2008.
- [22] N. K. Hansen & P. Coppens. *Acta Crystallographica*, **A34**, 909–921, 1978.
- [23] C. Jelsch, B. Guillot, A. Lagoutte, C. Lecomte. *Journal of Applied Crystallography*, **38**, 38–54, 2005.
- [24] C. Jelsch, V. Pichon-Pesme, C. Lecomte, A. Aubry. *Acta Crystallographica*, **D54**, 1306–1318, 1998.
- [25] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives. *Journal of the American Chemical Society*, **118**, 11225–11236, 1996.

- [26] C. Katan, P. Rabiller, C. Lecomte, M. Guezo, V. Oison, M. Souhassou. *Journal of Applied Crystallography*, **36**, 65–73, 2003.
- [27] T. S. Koritsanszky, A. Volkov, P. Coppens. *Acta Crystallographica*, **A58**, 464–472, 2002.
- [28] D. N. Lebedev & D. N. Laikov. *Doklady Mathematics*, **59**, 477, 1999.
- [29] G. N. Lewis. *Journal of American Chemistry Society*, **38**, 762, 1916.
- [30] A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus. *Journal of Physical Chemistry B*, **102**, 3586–3616, 1998.
- [31] A. O. Madsen. *Journal of Applied Crystallography*, **39**, 757–758, 2006.
- [32] P. Munshi & T. N. Guru Row *Acta Crystallographica*, **B62**, 612–626, 2006.
- [33] P. Munshi, T. S. Thakur, T. N. Guru Row, and G. R. Desiraju *Acta Crystallographica*, **B62**, 118–127, 2006.
- [34] S. Patel, A. D. MacKerell Jr., C. L. Brooks III *Journal of Computational Chemistry*, **25**(12), 1504–1514, 2004.
- [35] V. Pichon-Pesme, C. Lecomte, H. Lachekar. *Journal of Physical Chemistry*, **99**, 6242–6250, 1995.
- [36] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, Jr. R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, T. Johnson, M. E. Head-Gordon. *Journal of Physical Chemistry B*, **114**, 2549–2564, 2010.
- [37] P. L. A. Popelier, L. Joubert, D. S. Kosov. *Journal of Physical Chemistry A*, **105**, 8254–8261, 2001.
- [38] M. Spackman. *Journal of Chemical Physics*, **85**, 6579–6586, 1986.
- [39] A. J. Stone. *Chemical Physics Letters*, **83** (2), 233–239, 1981.
- [40] O. Treutler & R. Ahlrichs. *Journal of Chemical Physics*, **102**, 346–354, 1995.
- [41] A. Volkov & P. Coppens. *Journal of Computational Chemistry*, **25**, 921–934, 2004.
- [42] A. Volkov, T. Koritsanszky, P. Coppens. *Chemical Physics Letters*, **391**, 170–175, 2004.
- [43] J. M. Wang, P. Cieplak, P. A. Kollman. *Journal of Computational Chemistry*, **21**, 1049–1074, 2000.
- [44] Z. X. Wang, W. Zhang, C. Wu, H. Lei, P. Cieplak, Y. Duan. *Journal of Computational Chemistry*, **27** (6), 781–790, 2006.
- [45] D. E. Williams & S. R. Cox. *Acta Crystallographica*, **B40**, 404–417, 1984.
- [46] B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte, C. Jelsch. *Acta Crystallographica*, **A63**, 108–125, 2007.

Chapitre III

Etude de la densité de charges du fidarestat, un inhibiteur de l'aldose réductase humaine

III.1. Introduction

Le fidarestat ((2*S*,4*S*)-6-fluoro-2',5'-dioxospiro[chroman-4,4'-imidazolidine]-2-carboxamide) (**Figure 1**) est un inhibiteur de l'aldose réductase humaine (hAR), une enzyme de poids moléculaire 36-kDa, membre de la super-famille des aldo-keto réductases. Elle catalyse la réduction de différents composés tels que les aldéhydes, les aldéhydes xenobiotiques, les ketones et les trioses, avec pour cofacteur le NADPH.

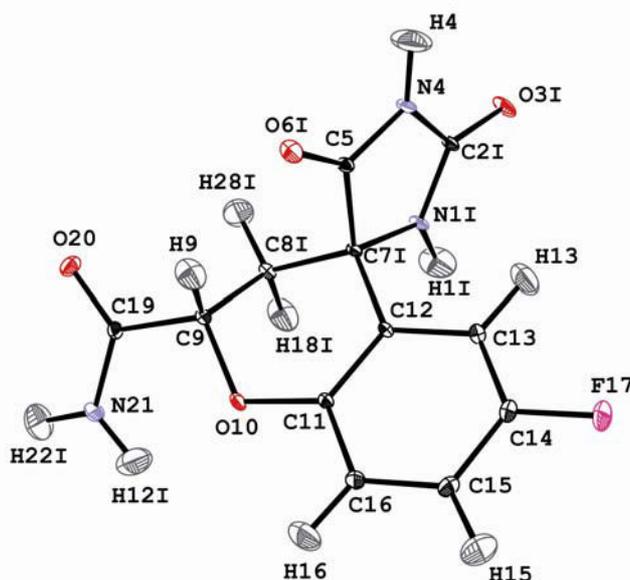


Figure 1 : Représentation OrtepIII [10] du fidarestat avec ses ellipsoïdes d'agitation thermique tracées à 50% de probabilité.

La hAR est une enzyme déterminant l'efficacité de la première étape du mécanisme polyol convertissant le glucose en sorbitol lequel est, dans une seconde étape, converti en fructose par l'enzyme sorbitol déshydrogénase. C'est l'accumulation de sorbitol dans les cellules qui provoque les complications liées au diabète. Pour cette raison, l'inhibition de la hAR est une voie thérapeutique potentielle pour traiter les pathologies débilitantes liées à l'hyperglycémie chronique.

Le fidarestat (SNK-860) a été testé sur des patients humains [29]. Il existait un doute quant à sa configuration absolue, doute qui a été récemment levé par cristallographie des rayons X, en utilisant un dérivé bromé [43].

De nombreux inhibiteurs de l'aldose réductase ont été étudiés au cours de ces dernières années [12;19;30]. Malheureusement, la plupart tels que le tolrestat, zopolrestat, ponarestat et

zenarestat, présentent des effets secondaires inacceptables, ou manquent d'efficacité [45;15 ;55]. Ces effets secondaires sont dus à l'inhibition par ces molécules d'une autre protéine, membre également de la super-famille des aldo-keto réductases, l'aldéhyde réductase, qui partage avec la hAR un haut niveau d'homologie structurale [18].

Le fidarestat présente un groupement imide cyclique auquel est substitué un groupement carbamoyl [44;43] (**Figure 2**). Il montre une plus grande activité ainsi qu'une meilleure sélectivité que les autres inhibiteurs. Le fidarestat a été récemment testé sur des patients souffrant de neuropathie et a donné des résultats encourageants dans la réduction du taux de sorbitol erythrocytique, sans effets secondaires remarquables [4].

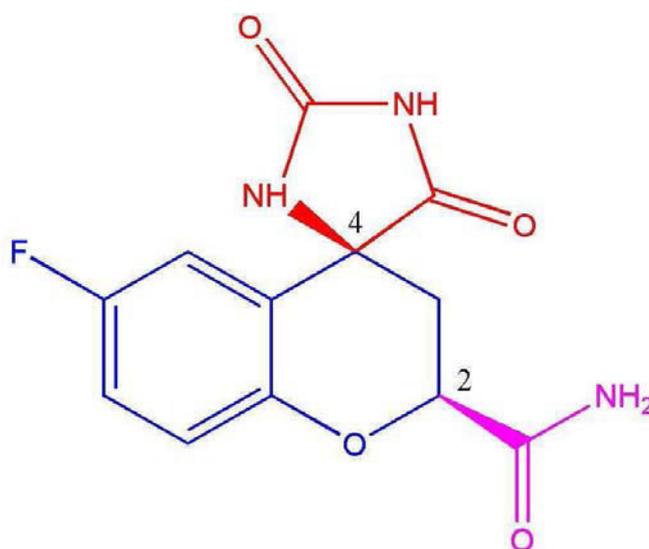


Figure 2 : Diagramme chimique du fidarestat avec le cycle chroman en bleu, l'imide cycle (ou groupement hydantoin) en rouge et le groupement carbamoyl en violet

Compte tenu de l'enjeu pharmaceutique, l'enzyme aldose réductase complexée avec le fidarestat a fait l'objet de nombreuses études à partir de l'analyse de données cristallographiques à haute résolution [19]. L'objet commun de ces études est la compréhension de l'affinité et de la spécificité de cet inhibiteur pour l'aldose réductase. La stéréospécificité de cet inhibiteur dans la fixation à l'aldose réductase par rapport à son stéréoisomère (*2R,4S*) a été étudiée par le biais de l'analyse des structures cristallographiques [42] mais aussi par simulations de dynamique moléculaire (estimations des énergies libres) [34]. Ces deux stéréoisomères diffèrent principalement par l'orientation du groupement carbamoyl dans le site actif de la protéine.

III.2. Matériels et méthodes

III.2.1. Cristallisation

La poudre de fidarestat a été dissoute dans l'eau. Des monocristaux transparents en forme d'aiguille se sont formés à partir de cette solution par évaporation lente à température ambiante sur une période de plusieurs semaines. L'expérience de diffraction des rayons X a été réalisée à l'aide d'un petit spécimen de dimensions $0,2 \times 0,15 \times 0,8 \text{ mm}^3$.

III.2.2. Collecte des données

L'expérience de diffraction des rayons X a été réalisée à une température de 100 K sur la ligne D3 du synchrotron Hasylab/DESY (Hamburg, Allemagne), avec une radiation monochromatique de longueur d'onde $\lambda = 0,450 \text{ \AA}$. Elle fut menée sur un diffractomètre quatre-cercle Huber équipé d'une unité de refroidissement Oxford Cryosystem et d'un détecteur plan Bruker MAR CCD (165 mm). L'espace réciproque a été exploré par une combinaison de différents φ et ω scans avec 2θ fixé à 6° pour des raisons techniques. Les basse et haute résolutions ont été mesurées avec des temps d'exposition de 30 s et 90 s respectivement. Le détecteur a été placé à 60 mm du monocristal. Au total, 1690 images ont été collectées avec un pas de scan de 1° . Les données extraites de ces images ont permis d'atteindre une résolution maximale de $\sin\theta_{\max}/\lambda = 1,01 \text{ \AA}^{-1}$, avec une complétude supérieure à 97% (**Tableau 1**).

Formule chimique	C ₁₂ H ₁₀ N ₃ O ₄ F	
Groupe d'espace	P2 ₁ 2 ₁ 2 ₁ (Z = 4)	
Dimensions de la maille (Å)	a = 6,0690(4), b = 12,5940(15), c = 15,4850(7)	
Température (K)	100(2)	
Radiation/ Longueur d'onde (Å)	Synchrotron / 0.450	
Diffractomètre	Four-circle Huber, D3 beam-line, HasyLab	
Résolution maximale	1,01 Å ⁻¹	
Nb de réflexions mesurées	317 438	
Nb de réflexions uniques (I/σ(I) ≥ 0)	9 991	
Complétude jusqu'à d _{max} = 0.495 Å	97,9 %	
R _{int} (I)*	8,8 %	
Redondance moyenne	31	
Correction d'absorption	Indexation des faces	
Coefficients de transmission T _{max} ; T _{min}	0,995 ; 0,983	
Modèle	Sphérique	Multipolaire
	H_iso	H_aniso
Nb d'observations / de restrictions	9991 / 20	9991 / 81
Nb de variables	221	671
Goodness-of-fit (g.o.f.)	0,801	0,399
Indices de qualité finaux (I/σ(I)>0)	wR ² (F) ^c 4,022	2,003
Déplétion/accumulation maximales de densité électronique résiduelle (e.Å ⁻³) avec I/σ(I)>3 et sinλ /σ ≤ 0,9 Å ⁻¹	-0,733 ; 0,855	-0,267 ; 0,224

* Facteur d'accord interne des données est défini par : $R_{int}(I) = \sum (I_{obs} - \langle I_{obs} \rangle) / \sum I_{obs}$.

Tableau 1 : Détails sur la collecte du fidarestat et statistiques sur les affinements cristallographiques.

III.2.3. Réduction des données

Les intensités diffractées ont été intégrées à l'aide du logiciel XDS [32]. La correction d'absorption a été réalisée par méthode numérique de quadrature gaussienne en utilisant le programme ABSORB [17]. Les coefficients de transmission maximale et minimale calculés par cette méthode sont $T_{max} = 0,995$ et $T_{min} = 0,983$. Les différents jeux de réflexions mesurées ont été mis à l'échelle et moyennés avec le programme SORTAV [8]. 317438 intensités ont été collectées ; en les moyennant selon une symétrie 222, 10 020 réflexions uniques ont été obtenues dont 9 991 satisfont le critère $I_{obs}/\sigma(I_{obs}) \geq 0$. Le facteur interne sur l'ensemble des données est $R_{int}(I) = 0,088$ avec une redondance moyenne de 31. Les détails sur les conditions de l'expérience sont communiqués dans le **Tableau 1**.

III.2.4. Affinement du modèle I.A.M.

La structure du cristal de fidarestat a été résolue dans le groupe d'espace $P2_12_12_1$ par méthode directe avec le programme SHELXS97 [48]. L'affinement de la structure fut réalisé par la méthode de minimisation des moindres carrés implémentée dans le logiciel MoPro [25;31].

Au cours de l'affinement, les facteurs de diffusion atomiques ont été calculés à partir des valeurs tabulées par Clementi & Raimondi [14]. De plus, les coefficients de dispersion réels et imaginaires ont été déterminés à partir de Kissel [35]. L'affinement de la structure a été mené en utilisant l'ensemble des réflexions satisfaisant le critère $I_{\text{obs}}/\sigma(I_{\text{obs}}) \geq 0$. Le schéma de pondération employé fut $1/\sigma^2(F_{\text{obs}})$.

La diffusion anormale n'étant pas significative, le calcul du paramètre de Flack [22] n'a pas permis de déterminer la conformation absolue du fidarestat, autrement dit sa chiralité [23]. La conformation absolue a, par conséquent, été identifiée en prenant comme référence la structure du dérivé bromé du fidarestat [43] ainsi que celle de l'inhibiteur dans le complexe hAR – fidarestat [19;30].

III.2.5. Affinement multipolaire

La qualité et la résolution des données obtenues pour cette étude ont permis de réaliser un affinement à très haute résolution à l'aide d'un modèle asphérique suivant le formalisme de Hansen & Coppens présenté dans la partie I.2.2.B, en utilisant le logiciel MoPro [25;31]. Les pseudo-atomes de carbone, azote, oxygène et fluor ont été modélisés jusqu'à l'ordre 4 (octupôles), alors que l'asphéricité des atomes d'hydrogène a été modélisée à l'aide d'un unique dipôle.

III.2.6. Stratégie d'affinement

La structure résultant de l'affinement avec le modèle d'atomes sphériques a été utilisée comme état initial de l'affinement multipolaire. L'affinement multipolaire a été mené en deux étapes.

Un premier affinement multipolaire standard a été réalisé en utilisant des paramètres d'agitation thermique isotrope restreints pour les atomes d'hydrogène. La longueur des

liaisons covalentes des atomes d'hydrogène a été restreinte aux distances standard obtenues par diffraction des neutrons [3], avec une déviation standard autorisée de 0,002 Å. Les coefficients kappa κ des atomes d'hydrogène ont été restreints aux valeurs 1,160(1) [52]. Les coefficients kappa κ' pour les hydrogènes portés par des atomes C ou N l'ont été aux valeurs communiquées par Volkov *et al.* [53], respectivement 1,180(1) et 1,400(1). Les coefficients κ et κ' de l'atome de fluor F17 sont également restreintes à la valeur 0,988 selon Sorensen *et al.* [49], avec comme déviations standards 0,001 et 0,01, respectivement.

A la fin de ce premier affinement, des paramètres d'agitation thermique anisotrope pour les atomes d'hydrogène ont été déterminés à l'aide du serveur SHADE précédemment introduit dans la partie I.2.3.D de ce rapport [39;40]. Une analyse TLS a dû pour cela être réalisée sur des fragments rigides. La molécule de fidarestat a donc été divisée en quatre groupes: le groupement hydantoin, le cycle chroman, le groupement carbamoyl (**Figure 2**) avec trois atomes supplémentaires, et pour finir, un dernier groupe chimique composé des atomes C19, O10, C9, C8I, C7I, H9, H28I et H18I (**Figure 1**). Les paramètres d'agitation thermique des atomes d'hydrogène ont été affinés avec des restrictions fortes aux valeurs obtenues à l'aide de SHADE avec des déviations autorisées de 0,0003 Å². Un nouvel affinement multipolaire avec atomes d'hydrogène anisotropes a alors été réalisé. La modélisation des atomes d'hydrogène jusqu'au quadripôle n'a pas été jugée nécessaire, les cartes de densité résiduelle (**Figure 3**) ne mettant pas en évidence de densité électronique résiduelle significative au voisinage des atomes d'hydrogène.

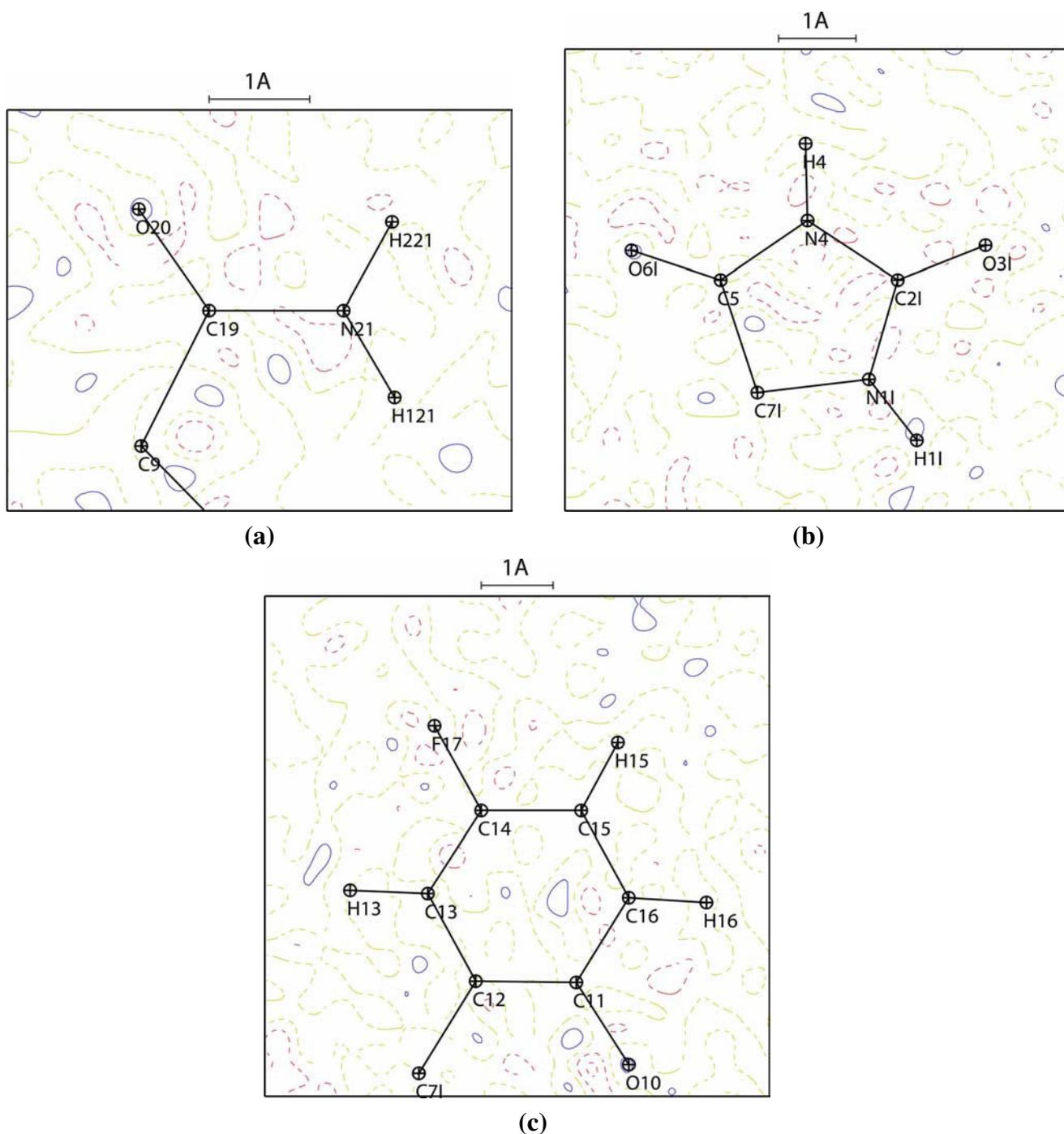


Figure 3 : Cartes de densité électronique résiduelle dans les plans des groupements chimiques :

(a) carbamoyl, (b) cycle hydantoin et (c) la partie pseudo-fluorophenyl du cycle chroman.

Les cartes ont été calculées en utilisant les réflexions avec $I/\sigma_I \geq 3$ et $0 \leq \sin\theta/\lambda \leq 0,9 \text{ \AA}^{-1}$. Les intervalles des contours sont de $0,05 \text{ e/\AA}^3$, les positives et les négatifs en lignes bleues continues et en lignes rouges en pointillés, respectivement.

III.3. Densité électronique du fidarestat

III.3.1. Affinement cristallographique

Les valeurs finales des indicateurs de qualité obtenues sont données dans le **Tableau 1**. Leurs valeurs faibles témoignent de la qualité des données ainsi que celle du modèle multipolaire obtenu.

Le test de la liaison rigide, ou test de Hirshfeld, présenté dans la partie I-2 de ce rapport a été utilisé pour vérifier la bonne décorrélation de la densité électronique et des paramètres atomiques d'agitation thermique [28]. La différence maximale observée entre atomes lourds est de $\Delta Z^2 = 8(3)10^{-4} \text{ \AA}^2$ pour la liaison C9—C8I. Aucun pic résiduel significatif de densité électronique n'a été observé dans les cartes résiduelles. Les valeurs de « *goodness of fit* » sont inférieures à l'unité à la fin de l'affinement A.I.M. et de l'affinement multipolaire. Cela montre que les clichés à basse et haute résolutions sont intégrés de la même manière et que les incertitudes expérimentales standard sont uniformément surestimées après la réduction des données par le programme SORTAV [8]. On remarquera que l'utilisation de paramètres d'agitation thermique anisotrope pour les atomes d'hydrogène conduit à une diminution faible, mais systématique des facteurs d'accord (**Tableau 1**).

III.3.2. Structure et empilement cristallins

La structure moléculaire du fidarestat incluant les atomes d'hydrogène est donnée **Figure 1**. Une représentation de l'empilement cristallin est donnée **Figure 4**. Une molécule de fidarestat est en contact avec 14 molécules voisines. A chaque dimère de molécules voisines correspond un dimère jumeau qui représente la même interaction intermoléculaire. Dans les dimères jumeaux, la molécule de référence (X, Y, Z ; code ortep= 55501) et les molécules voisines sont liées par des opérateurs de symétrie inverses. Ainsi, il y a au sein de l'empilement cristallin seulement sept dimères uniques.

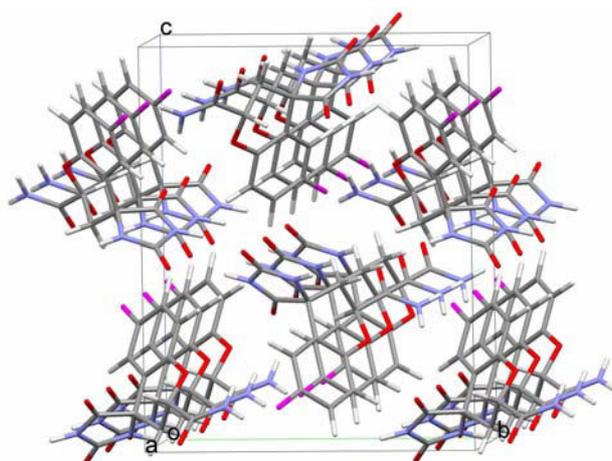


Figure 4 : Vue selon l'axe « a » de l'empilement cristallin du fidarestat. Cette vue a été réalisée avec le logiciel Mercury [38].

La molécule de fidarestat est impliquée dans sept liaisons hydrogène remarquables (**Tableau 2**). Quatre des sept dimères dans l'empilement cristallin sont concernés par au moins une de ces liaisons hydrogène (**Tableau 2, Figure 5**). Tous les atomes d'oxygène sont impliqués en tant qu'accepteurs dans ces liaisons hydrogène : O6I est accepteur pour trois liaisons, O3I pour deux, et les atomes d'oxygène O20 et éther O10 pour seulement une. Deux des liaisons hydrogène : N21—H121...O10 et C9—H9...O6I sont des interactions intramoléculaires. Chaque molécule de fidarestat est impliquée dans cinq interactions H...O comme accepteur ou donneur au sein de liaisons hydrogène avec des molécules voisines.

Tableau 2 : Caractéristiques géométriques des liaisons hydrogène intra- et inter-moléculaires avec le donneur D, l'accepteur A, l'hydrogène H.

D—H...A ^a	Opération de symétrie	D—H (Å)	H...A (Å)	D...A (Å)	D—H...A (degrés)
N21—H121...O10	x, y, z	1,0090(8)	2,15(1)	2,5946(9)	104,3(6)
C9—H9...O6I	x, y, z	1,0989(8)	2,390(5)	3,1386(8)	123,9(3)
C13—H13...O6I	$x-1, y, z$	1,0829(8)	2,457(5)	3,3159(8)	135,4(4)
N1I—H1I...O20	$x-1, y, z$	1,0089(8)	1,957(5)	2,8469(8)	145,6(5)
N4—H4...O3I	$\frac{1}{2}+x, \frac{3}{2}-y, 2-z$	1,0089(8)	1,807(2)	2,8011(8)	167,9(7)
C15—H15...O3I	$-x-\frac{1}{2}, -y+1, z-\frac{1}{2}$	1,0829(8)	2,340(6)	3,1557(8)	130,8(5)
N21—H121...O6I	$1-x, y-\frac{1}{2}, \frac{3}{2}-z$	1,0090(8)	2,483(4)	3,374(1)	146,9(4)

^a La liaison intramoléculaire N21—H121...O10 n'est pas caractérisée par un point critique.

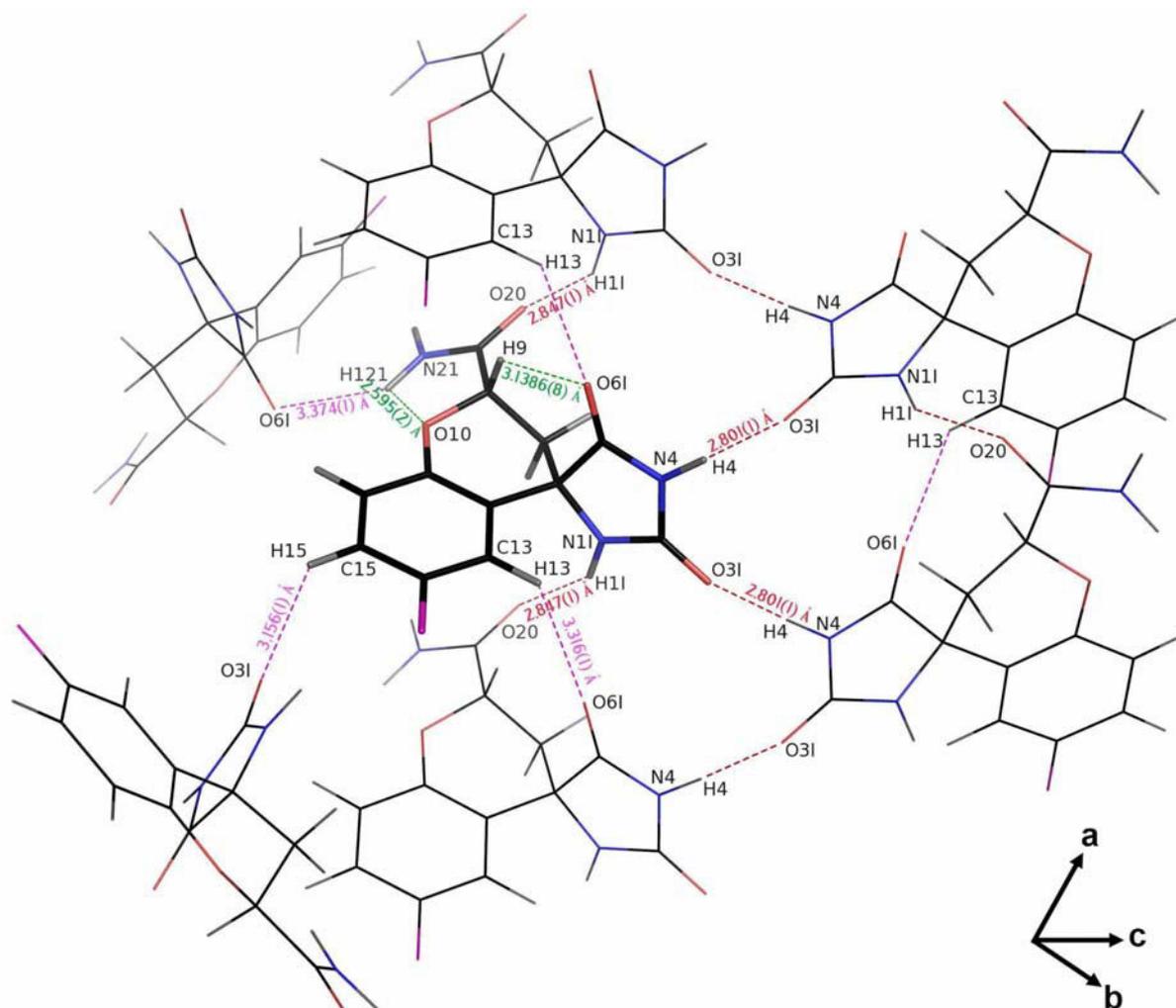


Figure 5 : Diagramme de l'empilement cristallin du fidarestat, montrant le réseau des liaisons hydrogène en lignes pointillées. La vue a été réalisée avec le programme Pymol [16].

Les configurations des dimères de molécules sont illustrées **Figure 6**. Les caractéristiques géométriques des cinq liaisons inter-moléculaires et des deux liaisons intra-moléculaires sont rapportées dans le **Tableau 2**.

D'un point de vue géométrique, on distingue deux catégories de liaisons hydrogène : deux fortes (distance (H...A) < 2,1 Å et un angle D—H...A > 140°) et trois faibles (H...A > 2,1 Å ou D—H...A < 140°). Parmi les quatre dimères de molécules de fidarestat, un seul est impliqué dans deux liaisons hydrogène.

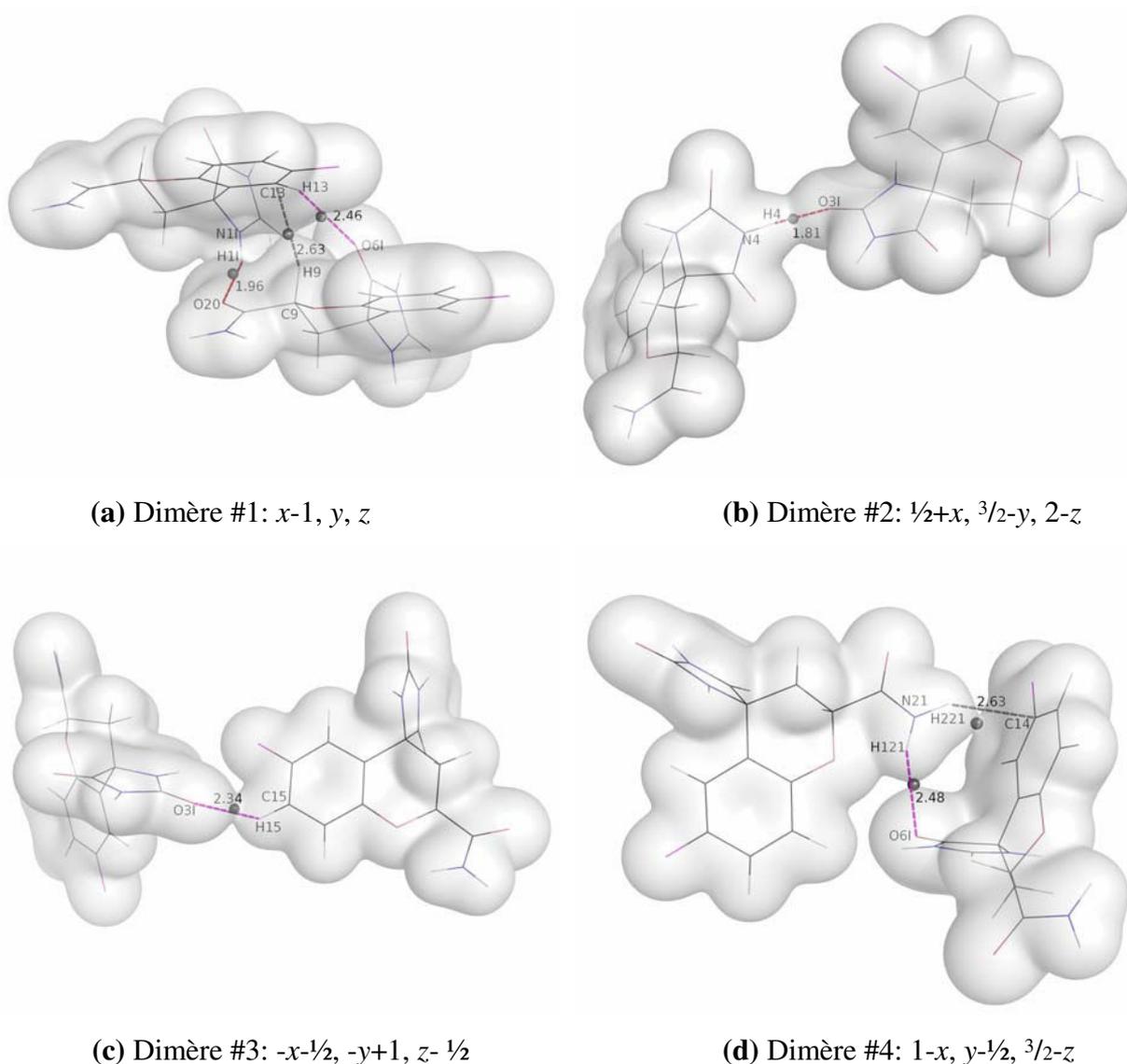


Figure 6 : Vue des quatre dimères de molécules de fidarestat liées par liaisons hydrogène trouvés dans l'empilement cristallin avec les opérateurs de symétrie correspondants. Les surfaces en gris représentent les isocontours de densité électronique expérimentale avec un seuil de $+0,05 e/\text{\AA}^3$. Les lignes pointillées rouge, violette et grises représentent les liaisons hydrogène, les contacts faibles et les liaisons H... π -système, respectivement. Les sphères grises représentent les points critiques intermoléculaires. La vue a été réalisée avec le programme Pymol [16].

Les molécules de fidarestat forment une chaîne en zigzag selon l'axe « a » de la maille cristalline par le biais de deux interactions intermoléculaires distinctes (**Figure 5**). La première est la liaison hydrogène N1I—H1I...O20 entre le groupement imine du cycle hydantoin et l'atome d'oxygène O20 du groupement carbamoyl. La seconde est une liaison hydrogène faible C13—H13...O6I (**Figure 6a**).

Les chaînes en zigzag parallèles à la direction « a » sont interconnectées via les liaisons hydrogène N4—H4...O3I entre les groupements carbonyle et imine de deux cycles hydantoin pour former ainsi un réseau à deux dimensions (**Figure 6b**). Les chaînes sont interconnectées grâce à deux autres liaisons hydrogène C15—H15...O3I (**Figure 6c**) et N21—H121...O6I (**Figure 6d**). De plus, les atomes O10 et O6I interagissent avec leurs groupements adjacents amine N21—H121 et “C_{sp3}—H” C9—H9 respectivement via les liaisons hydrogène intramoléculaires N21—H121...O10 et C9—H9...O6I (**Figure 5**), constituant ainsi respectivement des cycles de cinq et six atomes $R_1^1(5)$ et $R_1^1(6)$ [21;7].

L'atome de fluor F17 est impliqué dans un unique contact intermoléculaire avec l'atome d'hydrogène H121—N21. Cette interaction peut être classifiée comme contact de type van der Waals au lieu de liaison hydrogène compte tenu de sa géométrie. La distance hydrogène – fluor est de 2,55(1) Å et est égale à la somme des rayons de van der Waals (1,47Å et 1,09Å pour le fluor F [9] et l'hydrogène H [47] respectivement). L'angle N—H...F est de 113(2)°, et est donc défavorable pour une liaison hydrogène. Une combinaison d'approches théoriques et statistiques⁴¹ a montré que les contacts C—H...F ne peuvent être considérés comme des liaisons hydrogène compte tenu du fait qu'ils sont faibles avec des énergies d'interaction de l'ordre des interactions van der Waals.

La quasi-perpendicularité des deux groupes cycliques du fidarestat explique l'absence des interactions $\pi\cdots\pi$ entre les cycles aromatiques dans l'empilement cristallin. La structure moléculaire peut être vue comme une alternance donneur-accepteur, dans laquelle les groupements carbonyle et imine jouent un rôle important dans le réseau de liaisons hydrogène.

III.3.3. Densité électronique de déformation

La **Figure 7** montre les cartes de densité électronique de déformation statique ($\Delta\rho$) dans les plans du groupement carbamoyl, de l'hydantoin et du groupement pseudo-fluorophenyl du cycle chroman. Toutes les caractéristiques attendues liées aux liaisons covalentes ainsi que les doublets non liants des atomes d'oxygène sont clairement visibles sur ces cartes $\Delta\rho$. Elles illustrent les accumulations de charges sur les liaisons covalentes non polaires (C—C), sur les liaisons polaires (C—F, C—O) et dans les régions de la couche de valence non liée autour des atomes d'oxygène et de fluor. Il y a une petite déplétion électronique sur les liaisons polaires C=O et C—F près des atomes électro-négatifs. Le pic de densité électronique sur la liaison C—F est faible, comme cela a déjà été observé dans la littérature [5;27;13]. Les doublets électroniques non liants des oxygènes carbonyle O6I et O3I sont dans le plan du cycle hydantoin.

Une vue de la densité électronique de déformation dans le plan perpendiculaire à la liaison C—F (**Figure 7d**) révèle la présence de trois doublets électroniques.

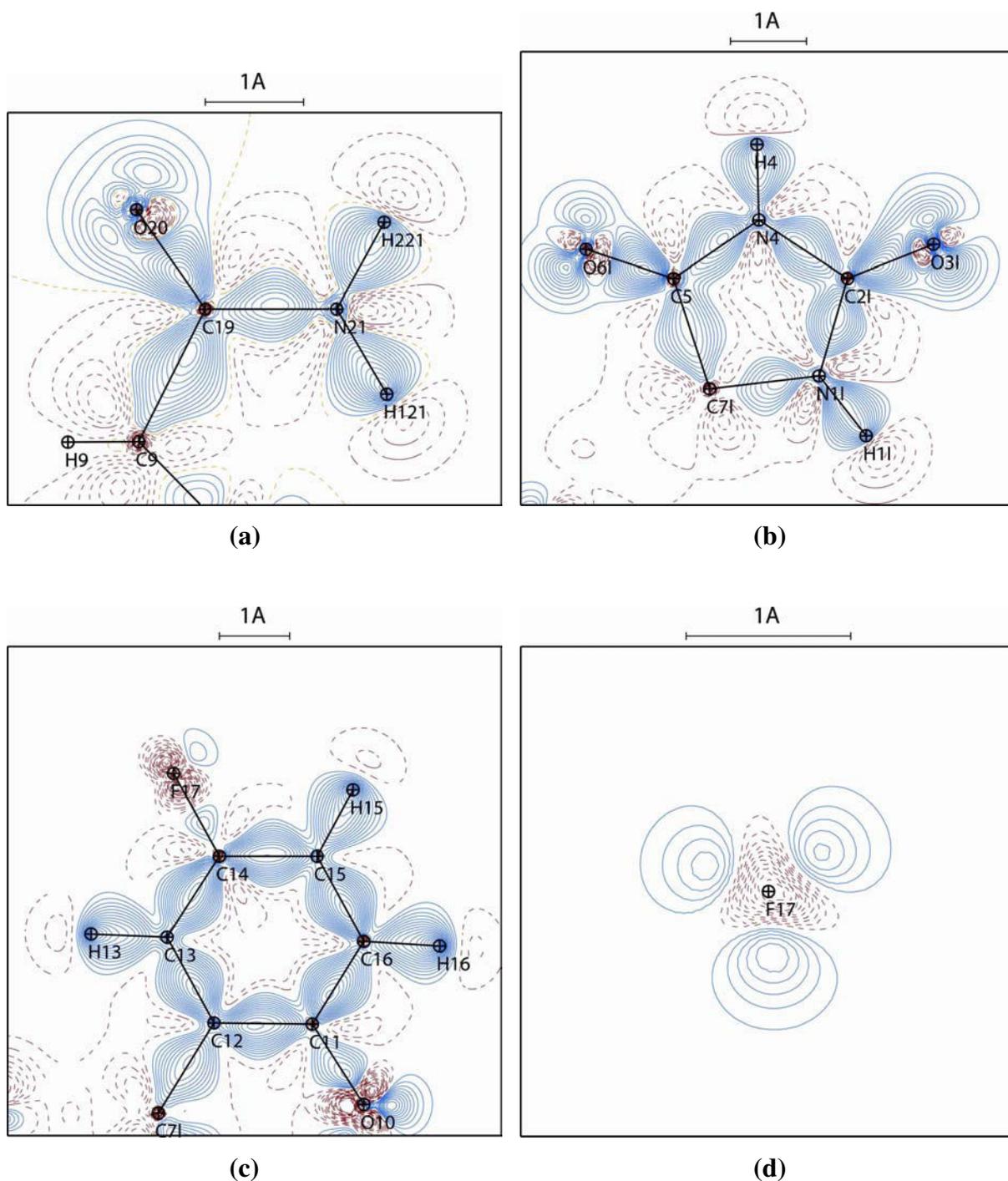


Figure 7 : Cartes de densité de déformation statique expérimentale dans les plans : (a) du groupement carbamoyl, (b) du cycle hydantoin, (c) du groupement pseudo-fluorophenyl du cycle chroman. Pour finir, la carte (d) est dans le plan perpendiculaire à la liaison covalent F17—C14 (plan à l'extérieure de la liaison F17—C14 à une distance de 0,1 Å de l'atome de fluor. La direction verticale est perpendiculaire au plan du cycle pseudo-fluorophenyl. Contours d'intervalles de $\pm 0,05 \text{ e}/\text{Å}^3$, positifs en lignes continues bleues et négatifs en lignes discontinues rouges.

III.3.4. Topologie des liaisons covalentes

L'analyse topologique de la densité électronique totale $\rho(\mathbf{r})$ et la localisation des points critiques de liaisons (BCP) ont été réalisées à l'aide de VMOPro [31]. Les propriétés topologiques aux points critiques (CP) sont résumées en annexe (**Annexe C**). Les incertitudes standard sur la position et la densité au BCP sont estimées à 10^{-3} Å et $0,05 e/\text{Å}^3$, respectivement. L'incertitude standard relative sur les valeurs du Laplacien est estimée à 10% [20]. Toutes les liaisons chimiques du cristal de fidarestat sont caractérisées par des points critiques de liaison (3,-1). Deux points critiques de cycle (3,+1) sont localisés aux centres du groupement pseudo-fluorophenyl du groupement chroman et du cycle hydantoin.

Les propriétés topologiques des groupements pseudo-fluorophenyl et hydantoin (**Annexe C**) sont également corrélées avec les caractéristiques structurales, avec des valeurs de la densité électronique aux points critiques, du Laplacien et des longueurs de liaisons correspondantes cohérentes. En effet, la densité électronique au point critique décroît clairement avec la longueur de la liaison (**Annexe C**).

L'analyse topologique souligne le caractère covalent de la liaison C14—F17. La valeur du Laplacien au point critique de la liaison C14—F17 est $-15,1 e/\text{Å}^5$, ce qui est égal à la valeur $-15(4) e/\text{Å}^5$ obtenue pour la liaison C_{sp^2} —F de l'acide *p*-fluoromandelique [37]. La valeur de la courbure λ_3 avec $14,9 e/\text{Å}^5$ dans la liaison C14—F17 est proche, en valeur absolue, des valeurs de λ_1 et λ_2 avec respectivement $-15,5 e/\text{Å}^5$ et $-14,5 e/\text{Å}^5$ [27]. Il est évident que la valeur du Laplacien sur la liaison C14—F17 diffère des autres liaisons covalentes: on remarque un changement de signe le long du chemin de liaison entre l'atome de fluor et le point critique, qui n'apparaît pas avec les autres liaisons impliquant les atomes d'oxygène et d'azote (**Figure 8**).

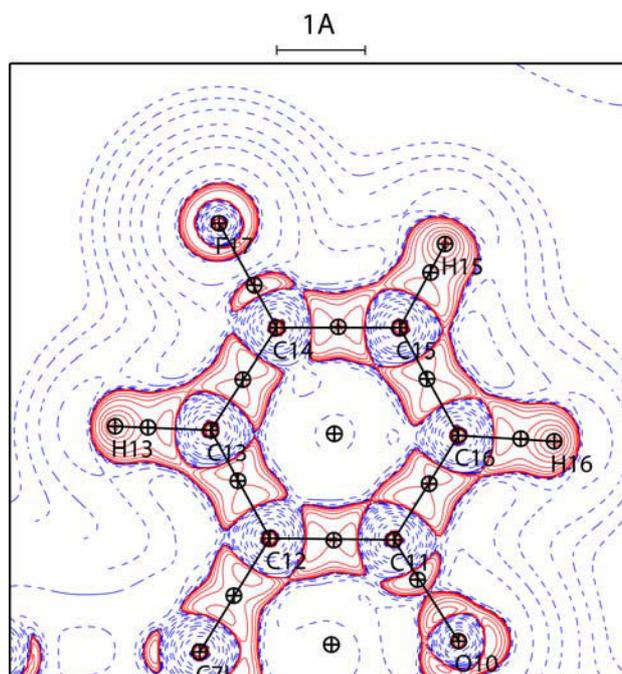


Figure 8 : Carte de distribution du Laplacien expérimentale $\nabla^2\rho$ dans le plan du groupement pseudo-fluorophenyl du cycle chroman. Les lignes pleines rouges montrent les contours négatifs, et les lignes pointillées bleues les contours positifs. Les intervalles de contours sont semilogarithmiques $\pm 0,2, 0,4, 0,8, 2,0 \dots e/\text{\AA}^5$. Les positions des points critiques de liaisons et de cycle sont également montrées.

L'analyse topologique permet l'élucidation du schéma de résonance et de la délocalisation électronique du cycle hydantoin. En effet, les différences des propriétés topologiques entre les liaisons C—N mettent en évidence les variations de distribution de charges autour de chaque atome d'azote. La liaison covalente C7I—N1I est plus longue que les quatre autres liaisons C—N (**Annexe C**). Cela est dû à la nature de l'atome C7I qui est le seul qui ne soit pas impliqué dans une liaison double C=O parmi les atomes de carbone C—N. Ainsi, les liaisons N1I—C2I, N4—C5 et N4—C2I présentent un plus fort caractère de liaison double que la liaison N1I—C7I, ce qui est confirmé par les ellipticités de liaison (0,15, 0,16, 0,12 et 0,05 respectivement). Ce comportement est révélateur de la résonance et de la délocalisation de la densité électronique dans le fragment N1I—C2I—N4—C5. Ceci peut également être observé dans les cartes de densité électronique de déformation (**Figure 7b**), qui montrent clairement une densité électronique de liaison plus faible sur la liaison N1I—C7I, comparée aux autres liaisons C—N du groupement hydantoin.

En dépit des différences dans les environnements électroniques autour des atomes d'oxygène, les propriétés topologiques des trois liaisons C=O (C2I=O3I, C5=O6I et C19=O20) sont proches, leurs ellipticités varient par exemple de 0,14 à 0,18.

III.3.5. Charges atomiques nettes

Les charges et les volumes atomiques ont été obtenus selon le formalisme A.I.M. (« Atom In Molecules ») par intégration numérique à l'aide du programme InteGriTy [33]. La somme des volumes des atomes dans la maille est de 1179,09 Å³, pour un volume de maille unitaire de 1183,56(17) Å³, soit une différence relative de 0,38 %. Plusieurs études récentes [24;1;54;11] suggèrent une conservation d'environ ± 5% pour la précision des propriétés atomiques intégrées, bien que les populations électroniques soient moins sensibles aux erreurs sur l'intégration. La somme des charges topologiques est de +0,004 lel pour une molécule de fidarestat, soit presque l'électronneutralité, ce qui est un indicateur de fiabilité.

Les charges atomiques $q_{PV} = N_{val} - P_{val}$ dérivées des populations de valence du modèle multipolaire, et q_{AIM} intégrées sur les bassins atomiques, sont données dans le **Tableau 3**. Compte tenu la définition des différents schémas de partitionnement, il n'est pas surprenant que certaines différences soient observées.

Les atomes d'azote, d'oxygène et de fluor portent tous des charges négatives, ce qui est en accord avec leur nature électro-négative. Tous les atomes d'hydrogène C_{sp2}—H sont positifs dans les deux formalismes. De même, les atomes d'hydrogène C_{sp3}—H sont positivement chargés comme cela pouvait être attendu.

La méthode A.I.M. conduit le plus souvent à des charges significativement plus importantes, avec des charges négatives non négligeables sur l'atome de fluor (-0,540 lel), sur les atomes d'azote ($\langle q_{AIM} \rangle = -1,03 \pm 0,07$ lel), et sur les atomes d'oxygène ($\langle q_{AIM} \rangle = -1,14 \pm 0,07$ lel). Les charges nettes négatives des atomes d'oxygène du carbonyle sont légèrement plus importantes que celle de l'oxygène éther, ce qui est en accord avec leur tendance à la formation de liaisons hydrogène plus fortes. Dans l'empilement cristallin du fidarestat, tous les atomes d'oxygène sont en effet des accepteurs de liaisons hydrogène.

Atome	V_{AIM}	q_{pv}	q_{AIM}^a	Type atomique
F17	19,03	-0,049	-0,540	
O10	15,69	-0,314	-1,077	C-O-C
O20	19,19	-0,333	-1,142	
O3I	18,33	-0,386	-1,213	O=C
O6I	18,65	-0,276	-1,125	
N1I	13,79	-0,162	-0,975	>NH
N4	14,74	-0,185	-1,033	
N2I	18,46	-0,257	-1,089	-NH ₂
C8I	8,46	-0,248	-0,079	>CH ₂
C16	13,54	0,051	0,033	
C13	10,84	-0,220	0,018	>C-H
C15	12,54	-0,104	0,021	
C14	9,20	0,138	0,481	>C-F
C12	10,13	-0,164	-0,227	>C-
C7I	5,69	0,202	0,416	>C<
C9	5,93	0,353	0,673	H>C-O
C11	9,91	0,025	0,455	>C-O
C2I	5,19	0,139	1,600	
C5	5,39	0,265	1,444	>C=O
C19	6,25	0,109	1,315	
H22I	4,48	0,238	0,444	H ₂ N-
H12I	3,96	0,164	0,370	
H4	3,30	0,202	0,413	H-N<
H1I	3,31	0,245	0,483	
H18I	6,48	0,139	0,107	H ₂ C-
H28I	7,80	0,076	0,004	
H9	5,73	0,050	0,011	H-C _{sp3}
H13	6,64	0,120	0,096	
H15	5,70	0,114	0,104	H-C<
H16	6,42	0,064	0,047	

^a La somme des charges atomiques est $\sum q_{AIM\ ANISO} = 0,004$ lel.

Tableau 3 : Charges (A.I.M.) (lel) et volumes (\AA^3) topologiques, et les charges des pseudo-atomes correspondants $q_{pv}=N_{val}-P_{val}$ pour chaque atome dans le fidarestat.

Les charges q_{pv} des atomes de carbone montrent une large gamme de valeurs (allant de -0,24 lel à +0,35 lel) (**Tableau 3**), et semblent être reliées à la nature des atomes auxquels ils sont liés. Néanmoins, les charges q_{AIM} des atomes de carbone sont plus contrastées. Le signe et l'amplitude des charges q_{AIM} est généralement en accord avec le transfert de charges attendu, compte tenu l'électronégativité des atomes voisins. Les atomes de carbone carbonyle $>C=O$ ont une importante charge q_{AIM} positive, supérieures à 1,30 lel. Les atomes de carbone de la partie pseudo-fluorophenyl du chroman, C14 et C11, respectivement liés aux F17 et O10,

mais aussi les atomes de carbone sp^3 C9 et C7I liés aux O10 et N1I, ont des charges q_{AIM} positives d'environ 0,50 lel. Tous les atomes de carbone liés à des hydrogènes sont proches de la neutralité.

III.4. Propriétés électrostatiques du fidarestat

III.4.1. Potentiel électrostatique du fidarestat

La représentation du potentiel électrostatique sur les surfaces moléculaires renseigne sur la topologie de la distribution de charges. Le potentiel électrostatique généré par une molécule de fidarestat isolée, c'est-à-dire extraite de son environnement cristallin, a été calculé à l'aide du programme VMoPro [31] à partir du modèle multipolaire de la densité de charges expérimentale (**Figure 9**).

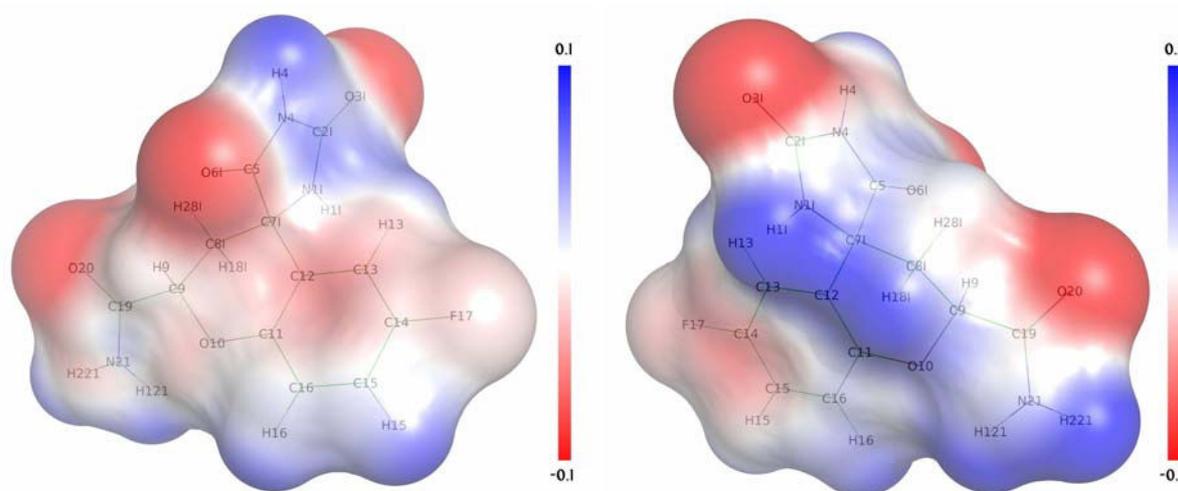


Figure 9 : Potentiel électrostatique de déformation $\Delta\phi$ ($e/\text{\AA}$) généré pour une molécule de fidarestat isolée, et représenté sur la surface d'exclusion solvant avec un rayon de sonde de 1,4 \AA . Le potentiel $\Delta\phi$ est dérivé de la densité électronique de déformation $\Delta\rho$. La vue a été réalisée avec le programme Pymol [16].

Généralement, la liaison C—F, aussi appelée “fluor organique”, est un aspect intéressant de l'analyse de la distribution de charges des composés fluorés [27;13]. Une large surface électro négative est engendrée par les atomes de fluor et les sites oxygènes O6I et O20. Ce résultat est cohérent avec leurs charges négatives. Si les charges $q_{pv} = N_{val} - P_{val}$ des atomes

d'oxygène sont seulement légèrement négatives, une partie du transfert de charges entre les atomes est modélisée par la partie multipolaire du modèle de Hansen & Coppens. Il est intéressant de noter que le potentiel négatif généré par les groupements carbonyl est plus intense qu'au voisinage de l'atome de fluor. Les forts potentiels négatifs au voisinage des groupements carbonyl sont corrélés avec leur aptitude à être de forts accepteurs dans les liaisons hydrogène H...O.

Les surfaces électropositives sont toutes bien distinctes autour du groupement H221—N21—H121 et des deux groupements imine de l'hydantoin (N11—H11 et N4—H4). Cela est corrélé avec les charges fortement positives q_{AIM} des atomes d'hydrogène H—N liés aux atomes d'azote, en comparaison avec les hydrogènes H—C. Ainsi, les sites de fixation polaire de la molécule de fidarestat sont clairement visibles sur la **Figure 9** : le potentiel électropositif généré par le groupement N4—H4 est flanqué de deux zones électronégatives engendrées par les parties carbonyl.

III.4.2. Topologie des liaisons hydrogène H...O

D'après la théorie A.I.M. [6], un contact H...O peut être considéré comme une liaison hydrogène si, parmi les autres conditions définies par Koch & Popellier [36], il y a un point critique le long du chemin de liaison H...O. La recherche des points critiques (3,-1) pour les 8 interactions H...O révèle six points critiques uniques.

L'unique interaction intramoléculaire, N21—H121 ... O10, est faible avec un angle N—H...O de 104(1)° seulement. Cette interaction polaire ferme un cycle intramoléculaire et ne présente pas de point critique. L'interaction H121...O10 est localisée sur le bord de la molécule qui génère un gradient significatif de densité électronique dans toute la région entre ces deux atomes. En raison de la géométrie des cycles de 5 atomes (H121—N21—C19—C9—O10...), il n'y a pas de CP dans le cycle ni sur le chemin d'interaction. Malgré tout, la norme du gradient présente un minimum dans le cycle, qui est non nul.

Les propriétés topologiques locales aux CPs des liaisons hydrogène sont résumées dans le **Tableau 4**. Les deux liaisons hydrogène selon des critères géométriques présentent des densités électroniques aux CP plus élevées : $\rho_{cp} = 0,146 \text{ e}/\text{Å}^3$ pour N11—H11...O20 et $0,196 \text{ e}/\text{Å}^3$ pour la liaison hydrogène N4—H4...O31. Les valeurs du Laplacien sont positives, avec $2,32 \text{ e}/\text{Å}^5$ et $3,38 \text{ e}/\text{Å}^5$ respectivement, ce qui caractérise les interactions à couche fermée.⁴⁷ Pour l'interaction longue N21—H121...O61, le point critique est proche de l'atome H121.

Cette interaction longue est aussi caractérisée par une densité électronique au CP faible ($0,047 \text{ e}/\text{Å}^3$). Il y a trois liaisons hydrogène $\text{C}_{\text{sp}2}\text{—H}\dots\text{O}=\text{C}$ faibles dans l'empilement cristallin. Plus courte est la distance donneur-accepteur, plus élevés sont la densité électronique et le Laplacien au CP de l'interaction (**Tableau 4**).

Symétrie	Liaison hydrogène	d_{Hcp}^a	d_{Acp}^a	ρ_{cp}	$\nabla^2\rho_{\text{cp}}$	$E_{\text{elec,tot}}$
	D—H...A	(Å)	(Å)	($\text{e}/\text{Å}^3$)	($\text{e}/\text{Å}^5$)	($\text{kJ}\cdot\text{mol}^{-1}$)
$x-1, y, z$	N1I—H1I...O2O	0,73	1,25	0,146	2,32	-56
	C13—H13...O6I	1,05	1,44	0,055	0,88	
	C9—H9... π C13	1,05	1,59	0,051	0,70	
$\frac{1}{2}+x, \frac{3}{2}-y, 2-z$	N4—H4...O3I	0,63	1,18	0,196	3,38	-44
$-x-\frac{1}{2}, -y+1, z-\frac{1}{2}$	C15—H15...O3I	0,98	1,38	0,078	1,20	-16
$1-x, y-\frac{1}{2}, \frac{3}{2}-z$	N21—H121...O6I	1,01	1,51	0,047	0,71	-24
	N21—H221... π C14	1,05	1,70	0,056	0,65	
intramoléculaire	C9—H9...O6I	1,03	1,40	0,085	1,21	

^a d_{Hcp} et d_{Acp} sont les distances entre le point critique de liaison et, l'atome d'hydrogène et l'atome accepteur note A, respectivement.

Tableau 4 : Propriétés topologiques aux CPs des liaisons hydrogène entre les molécules dans les dimères de molécules voisines trouvés dans le cristal du fidarestat, et énergies d'interaction électrostatique totale correspondantes.

III.4.3. Les liaisons hydrogène $\pi\dots\text{H—X}$

La recherche des points critiques permet la mise en évidence de deux autres interactions intermoléculaires entre des atomes d'hydrogène et la partie pseudo-fluorophenyl du groupement chroman (**Figure 10**).

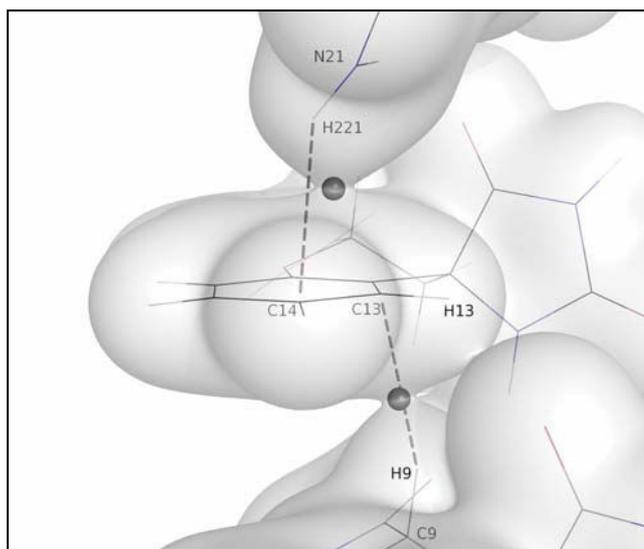
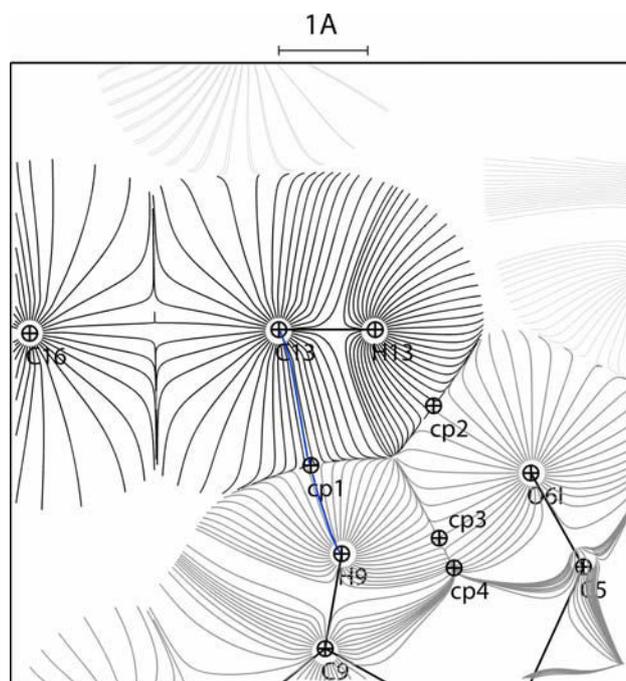


Figure 10 : Vue des liaisons hydrogène avec accepteur π dans un trimère de molécules de fidarestat. Le cycle aromatique contenant les atomes C13—C14 est impliqué des deux côtés du cycle dans les liaisons hydrogène H... π représentées par des lignes grises en pointillés. Le seuil de densité électronique pour les isosurfaces est de $+0,05 \text{ e}/\text{Å}^3$. La vue a été réalisée avec le programme Pymol [16].

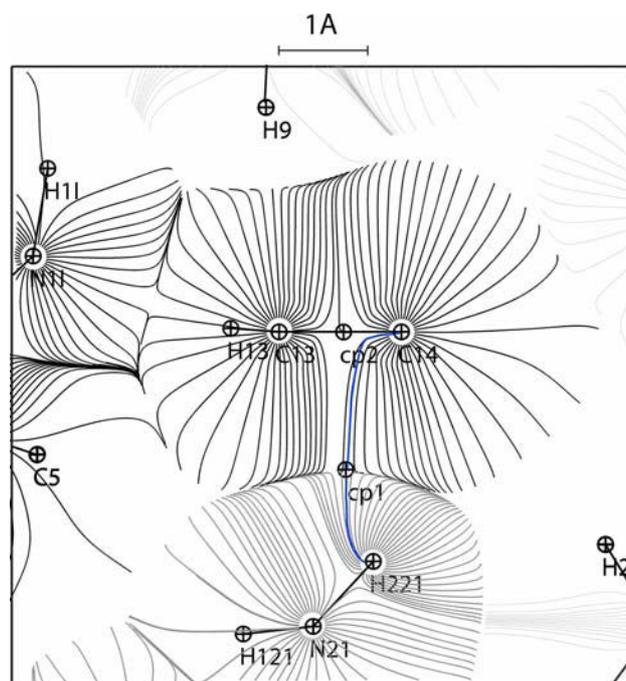
Ces interactions atypiques peuvent être considérées comme des liaisons hydrogène faibles où l'accepteur est un groupement présentant des électrons π , typiquement un cycle aromatique. Ce type d'interaction intermoléculaire X—H... π (avec X = O, N, C) est souvent observé dans les structures de protéines [51] et, plus généralement, dans les structures cristallines où il y a déficit d'accepteurs standard de liaison hydrogène [26]. Ce type d'interaction est souvent schématiquement représenté comme un contact en T où l'hydrogène est proche du centre du cycle. Cependant, une étude menée sur les liaisons hydrogène impliquant des cycles phenyl, par Malone *et al.*[41] à partir des structures de la « Cambridge Structural Database » [2], révèle que le schéma d'interaction est souvent plus diversifié : l'atome d'hydrogène peut en effet interagir directement avec les atomes de carbone du cycle aromatique. La partie pseudo-fluorophenyl du groupement chroman du fidarestat n'est pas strictement un cycle phenyl dans la mesure où la liaison C11—C12 intervient aussi dans le cycle adjacent, et un atome de fluor est lié à son carbone C14. Comparé au cycle phenyl standard, ces particularités perturbent la symétrie axiale du nuage électronique du cycle et favorise ainsi la possibilité d'une liaison hydrogène directe avec des parties spécifiques du cycle.

La première des deux interactions X—H... π dans le cristal de fidarestat implique le groupement C13—C14 du cycle pseudo-fluorophenyl et le groupement H9—C9. La **Figure 11a** montre les lignes de gradient de la densité électronique statique dans le plan (C13—H13, H9) des deux liaisons hydrogène H... π . On remarque qu'un chemin de liaison entre les noyaux H9 et C13 existe pour l'interaction C9—H9... π . Le point critique (abrégé CP) est localisé dans le plan défini par les atomes C13, H13 et H9 (**Figure 11a**). L'atome C sp^2 qui est le plus proche de l'hydrogène H9 est C13, avec une distance de 2,628(9) Å. La seconde distance la plus courte est $d=3,04(1)$ Å avec l'atome C14. L'angle C9—H9 ... C13 a une valeur de 153,4(8)° alors que la liaison C9—H9 est presque perpendiculaire au plan du cycle (avec un angle par rapport au vecteur normal du plan moyen de 11°).

La seconde liaison hydrogène X—H... π implique N21—H221 comme accepteur, lequel interagit avec le groupe C13—C14 par l'autre côté du cycle aromatique (**Figure 11b**). Le CP de liaison de l'interaction N21—H221... π est localisé dans le plan défini par les atomes C13, C14 et H221 comme cela peut être vu sur la **Figure 11b**. La liaison N21—H221 forme un angle de 43° avec le vecteur normal du plan moyen du cycle pseudo-phenyl. Les caractéristiques géométriques de l'interaction N21—H221... π montre que les atomes qui sont proches de H221 sont les atomes de carbone C14 et C13, respectivement à une distance de 2,62(1) Å et 2,81(1) Å. L'interaction N21—H221... π est inhabituelle selon les résultats de l'étude des liaisons X—H... π menées par Malone *et al.* [41], puisque le CP de la liaison H... π est localisé à égale distance des atomes C13 et C14 (**Figure 11b**). L'angle formé par N21—H221...C14 est de 144(1)°, ce qui est plus favorable que l'angle N21—H221...C13 de 115(1)°. La position de l'hydrogène H221 par rapport au cycle pseudo-fluorophenyl et l'orientation de la liaison N21—H221 favorisent une interaction essentiellement avec l'atome C14. En effet, les lignes de gradient de densité électronique révèlent un chemin de liaison entre les atomes H221 et C14 (**Figure 11b**). Cependant, compte tenu de l'incertitude sur la densité expérimentale et sur le critère géométrique, le chemin de liaison pourrait également relier l'atome H221 et le point critique de la liaison covalente C13—C14. Des chemins de liaison entre un atome et un CP de liaison ont déjà été observés, par exemple, dans les complexes carbides impliqués dans des liaisons Sc... π (C=C) [46].



(a)



(b)

Figure 11 : Représentation des lignes de gradient de la densité électronique totale dans les régions des liaisons hydrogène avec accepteur π . Dans les plans définis par les atomes C13, H13 et H9 (a), et par les atomes C13, C14 et H221 (b). Les points critiques les plus intéressants sont indiqués. Les vues sont faites avec le logiciel VMOPro [31]. Les chemins de liaison intermoléculaire sont ajoutés en lignes noires.

Les deux interactions C—H... π sont très faibles compte tenu de leurs valeurs de Laplacien plus faibles que $1 \text{ e}/\text{\AA}^5$, ce qui est lié au faible recouvrement des couches électroniques des atomes. Néanmoins, ces interactions faibles présentent des points critiques de liaisons (3,-1) avec des valeurs de densité électronique aux CPs comparables à celles usuellement observées pour les liaisons hydrogène C—H...O, qui sont maintenant communément admises comme des liaisons hydrogène valides, spécialement dans les structures biologiques [50]. Ces interactions H... π contribuent aux énergies d'interaction électrostatique des dimères de molécules de fidarestat correspondants, comme cela sera discuté par la suite.

III.4.4. L'empilement cristallin et les énergies d'interaction électrostatique

L'énergie d'interaction électrostatique est calculée (**Tableau 4**) pour chaque dimère de molécules de fidarestat voisines impliquées dans une liaison hydrogène.

Les calculs d'énergie ont été réalisés à l'aide du programme VMOPro selon la méthode décrite dans le chapitre 2 de ce rapport. Une telle approche est pertinente dans le cas du fidarestat : les interactions intermoléculaires étant assurées par des liaisons hydrogène, la contribution électrostatique est attendue prédominante comparée à la contribution de la dispersion comme c'est le cas des interactions π ... π , qui sont faibles mais coopérativement additives. Le fidarestat est une molécule fortement polaire ce qui est confirmé par un moment dipolaire de 7,7 Debyes dans le cristal. En comparaison, le moment dipolaire généré par la chaîne principale d'un dipéptide dans une conformation d'hélice α idéale est d'environ 10,6 Debyes. Une autre caractéristique du fidarestat qui indique que l'énergie de dispersion ne peut être prédominante est la structure en forme de « T » de la molécule qui empêche l'empilement parallèle des systèmes π des différentes molécules dans le cristal.

Dans le dimère n°1, il y a trois liaisons hydrogène contribuant à l'interaction intermoléculaire (**Figure 6a**). La plus forte est la liaison hydrogène N1I—H1I...O20 entre le groupement imino N1I—H1I du cycle hydantoin et l'atome d'oxygène O20 du carbamoyl. La seconde liaison hydrogène C13—H13...O6I, entre un groupement pseudo-fluorophenyl et un cycle hydantoin, est faible. La troisième liaison hydrogène C9—H9... π C13 est également faible comme décrit dans le paragraphe précédent.

Dans le dimère n°2, les cycles hydantoin des deux molécules interagissent seulement par la liaison hydrogène N4—H4...O3I (**Figure 6b**). Si les distances donneur-accepteur des deux liaisons hydrogène N—H...O sont comparables, l'interaction N4—H4...O3I est plus courte ($d_{N...O} = 2,801(1) \text{ \AA}$) que N1I—H1I...O20 ($d_{N...O} = 2,847(1) \text{ \AA}$). De plus, N4, H4 et O3I sont presque alignés (avec un angle N—H...O de $168(1)^\circ$), alors que la liaison hydrogène N1I—H1I...O20 forme un angle moins favorable de $146(1)^\circ$ (**Tableau 2**). Ainsi, d'un point de vue géométrique (distance et angle), la liaison hydrogène N4—H4...O3I devrait être plus forte que la liaison N1I—H1I...O20. Cela est également confirmé par les valeurs de la densité électronique aux points critiques qui sont de $0,196 \text{ e/\AA}^3$ et $0,146 \text{ e/\AA}^3$ respectivement (**Tableau 4**). L'énergie d'interaction électrostatique totale pour le dimère n°1 est la plus forte avec -56 kJ.mol^{-1} comparée à -44 kJ.mol^{-1} pour dimère n°2. C'est une conséquence des deux contacts supplémentaires C13—H13...O6I et C9—H9... π C13, qui contribuent à la stabilisation du premier dimère.

Avec une énergie de -16 kJ.mol^{-1} , le dimère n°3 a l'interaction électrostatique totale la plus faible. Cette observation est en accord avec la longue distance donneur-accepteur de $3,156(1) \text{ \AA}$ dans la liaison hydrogène C15—H15...O3I (**Figure 6c**). Les liaisons hydrogène où C—H est le groupement donneur, sont également connues comme étant plus faibles que celles impliquant les groupements N—H ou O—H [50]. La charge positive q_{AIM} des atomes d'hydrogène H—C est en effet de l'ordre du quart de celle des atomes d'hydrogène H—N (**Tableau 3**).

L'énergie d'interaction électrostatique totale calculée pour le dimère n°4 est de -24 kJ.mol^{-1} (**Tableau 4**). Ce dimère a une liaison hydrogène faible N21—H121...O6I avec une longue distance donneur-accepteur $d_{NO} = 3,374(1) \text{ \AA}$ (**Figure 6d**). C'est la liaison hydrogène la plus faible (en excluant les interactions H... π) dans le cristal de fidarestat d'après la densité électronique au CP de $0,047 \text{ e/\AA}^3$ (**Tableau 4**). Le dimère n°4 est aussi caractérisé par une liaison hydrogène avec un accepteur π entre l'atome d'hydrogène H221—N21 du groupement carbamoyl et l'atome sp^2 C14 du cycle pseudo-fluorophenyl. L'addition de ces deux liaisons hydrogène faibles résulte en une énergie d'interaction électrostatique totale pour ce dimère significativement plus favorable que pour le dimère n°3.

En conclusion, ce type d'interactions faibles H... π contribue à l'énergie de l'empilement cristallin et ne doit donc pas être négligée.

III.5. Conclusion

Le réseau de liaisons hydrogène dans l'empilement cristallin du fidarestat a été analysé. Ce dernier présente notamment des interactions entre les orbitales π de cycle pseudo-fluorophenyl et deux atomes d'hydrogène.

Disposer de données à très haute résolution pour cet inhibiteur fidarestat a contribué également au développement et à l'extension de la nouvelle version généralisée de la banque expérimentale de paramètres multipolaires ELMAM. (cf partie II.2.3) Les nouveaux types atomiques ainsi définis ont permis en particulier d'avoir recours à la transférabilité des paramètres multipolaires de distribution de charges pour définir des modèles de charges d'enzymes aldose réductase (ou aldéhyde réductase) complexées avec le fidarestat ou d'autres inhibiteurs présentant un groupement hydantoin (**Figure 2**). L'étude de ces modèles de distribution de charges est l'objet du quatrième et dernier chapitre de ce manuscrit.

- [1] F. M. Aicken & P. L. A. Popelier. *Canadian Journal of Chemistry*., **78**, 415–426, 2000.
- [2] F. H. Allen. *Acta Crystallographica*, **B58**, 380–388, 2002.
- [3] F. H. Allen, O. Kennard, D. Watson, L. Brammer, A. Orpen, R. Taylor. *International Tables for Crystallography, Vol. C*, chapter Chapter 9.5, 685–706. Dordrecht: Kluwer Academic Publishers, 1992.
- [4] T. Asano, Y. Saito, M. Kawakami, N. Yamada. *Journal of Diabetes and its Complications*, **16**, 133–138, 2002.
- [5] A. Bach, D. Lentz, P Luger. *Journal of Physical Chemistry A*, **105**, 7405– 7412, 2001.
- [6] R. F. W. Bader. *Atoms in Molecules: a Quantum Theory*. The International Series: Monographs in Chemistry Oxford: Clarendon Press., 1990.
- [7] J. Bernstein, R. E. Davis, L. Shimoni, N.-L. Chang. *Angewandte Chemie International Edition*, **34**, 1555–1573, 1995.
- [8] R. H. Blessing. *Acta Crystallographica*, **A51**, 33–38, 1995.
- [9] A. Bondi. *Journal of Physical Chemistry*, **68**, 441–452, 1964.
- [10] M. N. Burnett & C. K. Johnson. *ORTEP III 1996, Report ORNL-6895, Oak Ridge National Laboratory, Tennessee, USA*.
- [11] I. Bytheway, D. J. Grimwood, D. Jayatilaka. *Acta Crystallographica*, **A58**, 232–243, 2002.
- [12] R. Cachau, E. Howard, P. Barth, A. Mitschler, B. Chevrier, V. Lamour, A. Joachimiak, R. Sanishvili, M. Van Zandt, E. Sibley, D. Moras, A. Podjarny. *Journal de Physique*, **10**, 3–13, 2000.
- [13] D. Chopra, T. S. Cameron, J. D. Ferrara, T. N. Guru Row. *Journal of Physical Chemistry A*, **110**, 10465–10477, 2006.
- [14] E. Clementi & D.L. Raimondi. *Journal of Chemical Physics*, **38**, 2686, 1963.
- [15] L. Costantino, G. Rastelli, P. Vianello, G. Cignarella, D. Barlocco. *Medical Research Reviews*, **19**, 3–23, 1999.
- [16] W. L. DeLano. The pymol molecular graphics system, 2002 on world wide web <http://www.pymol.org>. Technical report.
- [17] G. T. DeTitta. *Journal of Applied Crystallography*, **18**, 75–79, 1985.
- [18] O. El-Kabbani, V. Carbone, C. Darmanin, M. Oka, A. Mitschler, A. Podjarny, C. Schulze-Briese, R. P. Chung. *Journal of Medicinal Chemistry*, **48**, 5536–5542, 2005.
- [19] O. El-Kabbani, C. Darmanin, T. R. Schneider, I. Hazemann, F. Ruiz, A. Oka, M. Joachimiak, C. Schulze-Briese, T. Tomizaki, A. Mitschler, A. Podjarny. *Protein*, **55**, 805–813, 2004.
- [20] E. Espinosa, E. Molins, C. Lecomte. *Chemical Physics Letters*, **285**, 170–173, 1998.
- [21] M. C. Etter, J. C. MacDonald, J. Bernstein. *Acta Crystallographica*, **B46**, 256–262, 1990.
- [22] H. D. Flack. *Acta Crystallographica*, **A39**, 876–881, 1983.
- [23] H. D. Flack & G. Bernardinelli. *Journal of Applied Crystallography*, **33**, 1143–1148, 2000.

- [24] C. Flensburg & D. Madsen. *Acta Crystallographica*, **A56**, 24–28, 2000.
- [25] B. Guillot, Viry L., R. Guillot, C. Lecomte, C. Jelsch. *Journal of Applied Crystallography*, **34**, 214–223, 2001.
- [26] L. R. Hanton, C. A. Hunter, D. H. Purvis. *Journal of the Chemical Society, Chemical Communications*, 1134–1136, 1992.
- [27] D. E. Hibbs, J. Overgaard, J. A. Platts, M. P. Waller, M. B. Hursthouse. *Chemical Physics Letters B*, **108**, 3663–3672, 2004.
- [28] F. L. Hirshfeld. *Acta Crystallographica*, **A32**, 239–244, 1976.
- [29] N. Hotta, T. Toyota, Y. Matsuoka, K. and Shigeta, R. Kikkawa, T. Kaneko, A. Takahashi, K. Sugimura, Y. Koike, J. Ishii, N. Sakamoto. *Diabetes Care*, **24**, 1776–1782, 2001.
- [30] E. R. Howard, R. Sanishvili, R. E. Cachau, A. Mitschler, B. Chevrier, P. Barth, V. Lamour, M. Van Zandt, E. Sibley, C. Bon, D. Moras, T. R. Schneider, A. Joachimiak, A. Podjarny. *Proteins*, **55**, 792–804, 2004.
- [31] C. Jelsch, B. Guillot, A. Lagoutte, C. Lecomte. *Journal of Applied Crystallography*, **38**, 38–54, 2005.
- [32] W. Kabsch. *Journal of Applied Crystallography*, **26**, 795–800, 1993.
- [33] C. Katan, P. Rabiller, C. Lecomte, M. Guezo, V. Oison, M. Souhassou. *Journal of Applied Crystallography*, **36**, 65–73, 2003.
- [34] D. Kim, S.I. Hong, D.S. Lee. *International Journal of Molecular Sciences*, **7**, 519–536, 2006.
- [35] L. Kissel, B. Zhou, S. C. Roy, S. K. Sen Gupta, R. H. Pratt. *Acta Crystallographica*, **A51**, 271–288, 1995.
- [36] U. Koch & P.L. Popelier. *Journal of Physical Chemistry*, **99**, 9747–9754, 1995.
- [37] S. Larsen, C. Flensburg, H. S. Bengacted, H. O. Sørensen. *Acta Crystallographica*, **S55**, 38, 1999.
- [38] C.F. Macrae, P.R. Edgington, P. McCabe, E. Pidcock, G.P. Shields, R. Taylor, Towler M., J. van de Streek. *Journal of Applied Crystallography*, **39**, 453–457, 2006.
- [39] A.Ø. Madsen. *Journal of Applied Crystallography*, **39**, 757–758, 2006.
- [40] A.Ø. Madsen, H.O. Sørensen, C. Flensburg, R. F. Stewart, S. Larsen. *Acta Crystallographica*, **A60**, 550–561, 2004.
- [41] J. F. Malone, C. M. Murray, M. H. Charlton, R. Docherty, A. J. Lavery. *Journal of the Chemical Society, Faraday Transactions* **93**, 3429–3436, 1997.
- [42] El-Kabbani O., C. Darmanin, M. Oka, C. Schulze-Briese, T. Tomizaki, I. Hazemann, A. Mitschler, A. Podjarny. *Journal of Medicinal Chemistry*, **47**, 4530–4537, 2004.
- [43] M. Oka, Y. Matsumoto, S. Sugiyama, N. Tsuruta, M. Matsushima. *Journal of Medicinal Chemistry*, **43**, 2479–2483, 2000.
- [44] T. Petrova, H. Steuber, I. Hazemann, A. Cousido-Siah, A. Mitschler, R. Chung, M. Oka, G. Klebe, O. El-Kabbani, A. Joachimiak, A. Podjarny. *Journal of Medicinal Chemistry*, **48**, 5659–5665, 2005.
- [45] M. A. Pfeifer, M. P. Schumer, D. A. Gelber. *Diabetes*, **46**, S82–S89, 1997.

- [46] B. Rohmoser, G. Eickerling, M. Presnitz, W. Scherer, V. Eyert, R. Hoffmann, U Rodewald, C. Vogt, R. Pöttgen. *Journal of the American Chemical Society*, **129**, 9356–9365, 2007.
- [47] R. S. Rowland & R. Taylor. *Journal of Physical Chemistry*, **100**, 7384–7391, 1996.
- [48] G.M. Sheldrick & T.R. Schneider. *Methods in Enzymology*, **277**, 319–343, 1997.
- [49] H.O. Sørensen, R.F. Stewart, G.J. McIntyre, S. Larsen. *Acta Crystallographica*, **A59** S540–550, 2003.
- [50] T. Steiner. *Crystallography Reviews*, **6**, 1–51, 1996.
- [51] T. Steiner & G. Koellner. *Journal of Molecular Biology.*, **305**, 535–57, 2001.
- [52] R. F. Stewart, E. R. Davidson, W. T. Simpson. *Journal of Chemical Physics*, **42**, 3175–3187, 1965.
- [53] A. Volkov, Y. Abramov, P. Coppens. *Acta Crystallographica*, **A57**, 272–282, 2001.
- [54] A. Volkov, C. Gatti, Yu. Abramov, P. Coppens. *Acta Crystallographica*, **A56**, 252–258, 2000.
- [55] H. Yasuda, M. Terada, K. Maeda, S. Kogawa, M. Sanada, M. Haneda, A. Kashiwagi, R. Kikkawa. *Progress in Neurobiology*, **69**, 229–285, 2003.

Chapitre IV

Estimation de l'incertitude statistique sur les énergies d'interaction électrostatique au sein des systèmes macromoléculaires

IV.1. Motivations et objectifs

Comme cela a été expliqué dans le chapitre 2 de ce manuscrit, les propriétés électrostatiques de molécules ou groupes chimiques peuvent être dérivées de leurs modèles multipolaires de distributions de charges obtenus par affinement cristallographique à haute résolution ou par transfert de paramètres multipolaires issus de banques expérimentales (ELMAM) [30;39] ou théoriques (UBDB [26], Invariom Databank [7]).

Une estimation précise de ces propriétés électrostatiques (potentiel électrostatique ou énergies d'interaction électrostatique) au sein d'un système moléculaire peut fournir des informations utiles sur l'origine de ses propriétés physico-chimiques.

Le modèle multipolaire de Hansen & Coppens, décrit dans le chapitre 1 de ce rapport, a été développé afin de prendre en compte l'asphéricité de la distribution électronique d'un atome. Ce modèle permet ainsi une modélisation précise de la distribution de charges et, par la même, des propriétés électrostatiques dérivées.

La transférabilité des paramètres multipolaires selon le formalisme de Hansen & Coppens [19] permet de construire des modèles non-sphériques de distribution de charges pour des systèmes de grande taille, comme cela a été montré dans plusieurs études pour des systèmes protéiques [22;17;9;25;16;28]. Ces modèles ont servi à la caractérisation du potentiel électrostatique, ou à une estimation des énergies d'interaction électrostatique au sein de ces systèmes.

Dans le cas d'une estimation des énergies d'interactions électrostatiques, les valeurs sont en général communiquées à l'unité près, la précision espérée est donc, implicitement, de l'ordre du $\text{kJ}\cdot\text{mol}^{-1}$. Mais l'erreur expérimentale est-elle de cet ordre ?

Prétendre à une incertitude de l'ordre du $\text{kJ}\cdot\text{mol}^{-1}$ est purement arbitraire. Elle reste discutable compte tenu de la précision numérique de la méthode de calcul des énergies utilisée (partie II.2.2) mais surtout de l'incertitude sur les différents paramètres d'un modèle affiné, et tout particulièrement dans le cas des systèmes macromoléculaires. Aucune étude n'a été menée à ma connaissance afin d'éclaircir ce point.

Dans le cadre d'études pharmacologiques portant sur l'inhibition de protéines, identifier les spécificités des inhibiteurs connus leur conférant leur efficacité permet de concevoir et

d'éprouver *in silico* de nouveaux inhibiteurs avant toute tentative de synthèse en laboratoire. Les méthodes *in silico* usuellement employées sont des optimisations structurales à l'aide de champs de force (minimisation d'énergie de la structure) et des dynamiques moléculaires qui prennent en compte l'entropie du système. Les structures cristallographiques de systèmes macromoléculaires servent de conformations initiales à ces simulations.

Les banques de paramètres multipolaires ont été envisagées comme outil supplémentaire permettant à moindre coût d'obtenir une estimation des énergies d'interactions électrostatiques entre différents composés (molécules, ions,...) ou groupements chimiques (acides aminés) au sein du site actif des protéines, via la reconstruction de leurs distributions de charges par le principe de transférabilité.

Au cours de ces dernières années, plusieurs travaux ont démontré la faisabilité de calculs d'énergie à partir de modèles asphériques obtenus par transfert de paramètres [31;23;24;37;27;16;9]. L'idée forte de la transférabilité est de pouvoir obtenir des modèles asphériques pour des systèmes de grande taille (plus de 1000 atomes) à partir d'un modèle structural affiné suivant le formalisme I.A.M. (ou Modèle d'Atomes Indépendants). En d'autres termes, on accède en terme de distribution de charges à un niveau de détail élevé nettement supérieur à la résolution expérimentale.

Par exemple, dans l'étude du fidarestat [16], inhibiteur de l'aldose réductase, que nous avons menée en utilisant la banque de paramètres multipolaires expérimentaux ELMAM [30;39] (présentée dans le partie II.2.3.), les énergies d'interactions électrostatiques entre le site actif de l'aldose réductase et ses inhibiteurs sont interprétées afin d'expliquer la stéréospécificité du fidarestat avec l'aldose réductase. De même, l'étude menée par Dominiak *et al.* [9] traite des complexes influenza neuraminidase en utilisant la banque de paramètres multipolaires théorique UBDB [26], et tente de mettre en évidence des corrélations entre les énergies d'interactions électrostatiques, et les pKa et les CI50 obtenus expérimentalement.

Bien que la faisabilité de la méthode ait été montrée, sa légitimité peut être discutée. En effet, même si nous considérons les banques de paramètres utilisées comme d'excellente qualité et donc aptes à reproduire la distribution de charges non perturbée d'une protéine ou d'autres molécules organiques, le « soubassement » ou « squelette » du modèle de distribution de charges souffre toujours d'une imprécision liée à son origine expérimentale. Le modèle I.A.M servant de support au transfert a été obtenu par affinement cristallographique contre un jeu de données de diffraction des rayons X de plus ou moins bonne qualité, et allant jusqu'à une

résolution minimale donnée. En dépit d'une description de la distribution de charges considérablement améliorée par un transfert, l'imprécision sur les positions atomiques demeure et ne peut être, en aucun cas, compensée par le transfert de paramètres de charges.

On peut alors s'interroger au sujet des répercussions des incertitudes sur les coordonnées atomiques sur la précision des énergies d'interaction électrostatique. Qu'en est-il en particulier des énergies d'interaction électrostatique obtenues à partir de modèles affinés à des résolutions modérées de l'ordre de 2 Å ou plus. Trouver une réponse analytique à une telle question est délicat. Si l'on considère l'énergie d'interaction électrostatique entre deux modèles asphériques de distribution de charges, déterminer l'expression de son erreur est un problème difficile compte tenu des nombreuses corrélations entre les paramètres en particulier dans les cas d'affinements de macromolécules. Comme je l'ai expliqué dans le chapitre 1 de ce manuscrit, les restrictions et contraintes fondées sur une connaissance *a priori* de la stéréochimie des groupements chimiques sont utilisées dans les affinements de macromolécules afin de faciliter la convergence du modèle vers un état final réaliste. Tous ces termes introduits dans l'expression de la fonction χ^2 sont autant de termes de corrélation entre variables du modèle qui vont affecter les incertitudes sur les positions et, par conséquent, celles sur les énergies.

De plus, indépendamment de la stratégie d'affinement, la complexité de l'expression dépend des paramètres de charges transférés. Dans le cas d'un modèle de charges ponctuelles, le modèle de charges étant rudimentaire il n'introduit pas de corrélations supplémentaires. Par contre, les paramètres multipolaires modélisant la distribution de charges d'un pseudo-atome dans le formalisme de Hansen & Coppens, sont orientés selon un repère local propre à ce pseudo-atome dont la définition dépend des atomes qui lui sont liés de façon covalente (partie I.2.2.B). Cette dépendance du repère local à son environnement covalent proche serait à prendre en considération.

Il faut, pour finir, garder à l'esprit un autre aspect de ce problème : toute expression analytique de l'incertitude sur une grandeur est obtenue en faisant certaines hypothèses sur les variables dont dépend cette grandeur. Les incertitudes sur les variables communiquées par les programmes d'affinement ne sont valables que si l'état final du modèle affiné correspond à un minimum local ou global du paysage énergétique de la fonction χ^2 . Elle prête également à la zone au voisinage de ce minimum un profil quadratique, et attribue au χ^2 , d'un point de vue statistique, une distribution normale multivariée. Ce type d'hypothèse est une approximation souvent appliquée dans les algorithmes mathématiques. Elle permet de simplifier la mise en

équation des problèmes complexes (sous forme matricielle) et donc leurs résolutions. Elle repose sur un constat selon lequel les distributions des variables, ici des mesures expérimentales, suivent généralement une loi normale (distribution gaussienne) ou quasi-normale.

La stratégie utilisée durant un affinement tend à faire évoluer le modèle de manière à minimiser la fonction χ^2 vers son minimum global ou tout du moins un minimum local satisfaisant tout en garantissant la réalité stéréochimique de la structure. Cependant, en pratique, en raison de la qualité des données expérimentales et de la taille des systèmes macromoléculaires étudiés, l'affinement optimise le modèle jusqu'à un état que l'on assume être un minimum satisfaisant. Le minimum optimal n'est pas aisément atteignable et l'état final peut correspondre à un minimum local malgré des valeurs d'indices de qualité et des cartes de densité électronique résiduelle tout à fait satisfaisantes.

Afin de prendre en compte ces différents aspects, je propose d'accéder à une estimation des incertitudes sur les paramètres du modèle obtenu par affinement, ainsi que sur les énergies d'interaction par une étude statistique.

Quelle procédure utiliser ?

L'étude proposée ici consiste à réaliser un large nombre d'affinements cristallographiques d'un même système macromoléculaire afin d'obtenir autant de conformations voisines de ce système. Chacune de ces conformations obtenues par affinement cristallographique jusqu'à convergence, correspond à un état final acceptable d'un point de vue stéréochimique. Un nombre suffisant de conformations constituera un échantillonnage du paysage énergétique de la fonction χ^2 au voisinage du minimum global ou local jugé satisfaisant. Une étude statistique menée à partir des conformations obtenues en fin d'affinement du système étudié peut nous permettre d'obtenir les distributions statistiques des coordonnées atomiques mais aussi des propriétés dérivées de ces conformations. Pour résumer, au lieu de tenter d'apprécier les incertitudes sur les énergies d'interaction électrostatique dans le site actif d'une protéine à partir d'une unique conformation supposée optimale et d'estimer analytiquement l'incertitude sur ces énergies par des approximations mathématiques, je propose de déterminer ces énergies

dans plusieurs milliers de conformations voisines et d'en déduire les incertitudes sur les coordonnées atomiques, ainsi que sur les énergies d'interaction par la statistique.

Une telle approche, bien que complexe de prime abord, présente l'avantage de prendre en compte implicitement l'effet de toutes les restrictions et contraintes utilisées au cours de l'affinement sans faire la moindre hypothèse sur le profil du paysage énergétique au voisinage du minimum.

L'objectif de cette étude sera donc de déterminer les incertitudes statistiques sur les énergies d'interaction électrostatique entre les ligands et les principaux résidus présents dans le site actif des protéines auxquels ils sont complexés. La motivation ici est de pouvoir discuter la légitimité de la construction d'un modèle asphérique de charges par transfert de paramètres multipolaires dans le cas de structures obtenues à des résolutions modérées qui ne permettent pas en principe la détermination expérimentale d'une distribution de charges multipolaire.

Les incertitudes statistiques sur les énergies d'interaction électrostatique vont dépendre de nombreux facteurs plus ou moins influents : la qualité du jeu de données, la résolution limite atteinte, le désordre dans le site actif, l'agitation thermique des groupements chimiques en interaction,... Discuter de l'effet de tous ces facteurs nécessiterait de tenter de décorrélérer leurs contributions et, pour cela, de réaliser un nombre conséquent de tests, ce qui en terme de temps et de ressources n'est pas envisageable. En première approximation, la résolution limite atteinte est le facteur le plus évident et sera, par conséquent, considéré ici.

IV.2. Matériels et méthodes

L'approche statistique suivie dans cette étude nécessite de réaliser un nombre d'affinements suffisant pour en déduire les distributions des coordonnées atomiques ainsi que des propriétés dépendantes de ces variables.

Afin d'éviter de biaiser la statistique, il est essentiel de ne pas partir du même état initial pour les différents affinements. Pour cela, avant chaque affinement, la structure initiale du système macromoléculaire doit être « perturbée » de manière aléatoire. La méthode retenue pour cela sera discutée ultérieurement.

Cette approche s'apparente au premier abord à la méthode « Simulated Annealing » ou « recuit simulé » qui fait partie des métaheuristiques qui sont utilisées pour résoudre des problèmes d'optimisation de modèles par minimisation de fonction. Lorsque la fonction présente un « paysage énergétique » accidenté, l'algorithme d'optimisation risque de converger vers un minimum local. Cette méthode consiste à perturber les paramètres du système (« chauffer » le système) de manière à ressortir d'un minimum local d'énergie, quitte à accepter une augmentation ΔE de la fonction du système (assimilée à une énergie) selon une

certaine probabilité $\min\left(1; e^{\frac{-\Delta E}{T}}\right)$, et d'affiner le modèle progressivement (« refroidir »

lentement le modèle). Ainsi, cela permet de se rapprocher du minimum global. Sous certaines conditions, l'algorithme du recuit simulé converge vers un optimum global [2].

Dans cette étude, l'objectif n'est pas de localiser le minimum global (tout du moins pas seulement) mais d'obtenir le profil des distributions d'énergies d'interaction électrostatique au voisinage du minimum global ou supposé global. Pour cela, la méthode utilisée pourrait être qualifiée de méthode itérative de criblage aléatoire au voisinage d'un minimum, plutôt que de méthode de recuit qui consiste à perturber/minimiser le même système récursivement.

Du point de vue des mathématiques statistiques, l'ensemble des états finaux d'affinement possibles peut être perçu comme l'univers des événements. Seuls les événements qui correspondent à des conformations stéréo-chimiquement satisfaisantes sont considérés, et vont former un sous-ensemble de l'univers global. Pour le système macromoléculaire étudié, l'ensemble des résultats d'affinement va constituer un échantillon de l'univers des événements.

Différentes questions doivent être abordées afin de mener dans les meilleures conditions ce travail :

- a) Les conformations d'un même système macromoléculaire étant obtenues par affinement, elles dépendront des outils d'affinement cristallographique utilisés. Ces outils devront être préalablement décrits.
- b) Les distributions de populations des paramètres des modèles et des énergies d'interaction électrostatique correspondantes devront être analysées. Quels coefficients ou indicateurs statistiques utiliser dans ce but ?
- c) Pour qu'une étude statistique soit satisfaisante, il faut que le cardinal de l'échantillon soit de taille suffisante pour que ce dernier soit représentatif de l'univers des événements. Il faudra donc déterminer le nombre de structures suffisant pour disposer d'un échantillon acceptable.
- d) Comme expliqué précédemment, pour permettre cette étude, chaque affinement doit être réalisé en partant d'un état initial différent. On perturbera donc les paramètres du modèle de manière aléatoire. Pour cela, quel type de perturbation devons-nous utiliser ? Mais aussi, quelle amplitude maximale donner à ces perturbations ?

Pour répondre à ces questions, j'ai réalisé des tests préliminaires en choisissant comme complexe protéine-inhibiteur le complexe aldose réductase porcine – sorbinil obtenu à une résolution de 2,3 Å. Ce système sera présenté ultérieurement.

IV.2.1. Outils d'affinement cristallographique

Les conformations obtenues par affinement vont dépendre du logiciel d'affinement cristallographique.

Il existe différents algorithmes d'affinement. Le minimum atteint à la fin de l'affinement ainsi que son profil, dépendent de l'expression de la fonction χ^2 minimisée. Comme expliqué dans

le chapitre 1 de ce rapport, les logiciels les plus courants utilisent un algorithme de minimisation par la méthode des moindres carrés.

De plus, différentes restrictions et contraintes peuvent être employées dans un affinement de macromolécules afin de réduire la complexité du modèle, et ainsi augmenter le rapport nombre d'observations sur nombre de variables comme cela a été expliqué dans la partie I.2.3 de ce manuscrit. Toutes ces restrictions et contraintes sont autant de termes supplémentaires dans l'expression de la fonction χ^2 qui modifient son profil. Les valeurs utilisées pour paramétrer les restrictions/contraintes sont des informations *a priori* que l'on considère acceptables mais qui ne sont pas pour autant uniques. Les valeurs retenues auront une influence sur le résultat de l'affinement.

On notera également que les restrictions sont des termes supplémentaires dans la fonction χ^2 qui introduisent une corrélation entre des variables du modèle ou qui imposent à la valeur d'une variable de rester au voisinage d'une valeur de référence. Ces restrictions sont plus ou moins strictes en fonction de leurs pondérations $w_{restriction}$. On définit généralement la pondération $w_{restriction}$ en fonction d'un paramètre sigma noté σ tel que $w_{restriction} = \frac{1}{\sigma^2}$. Plus la valeur du σ est petite, plus la restriction est stricte (lorsque σ tend vers 0, la restriction peut être assimilée à une contrainte). Les paramètres σ utilisés auront donc également une influence sur les incertitudes des variables du modèle identifié comme optimal par le programme d'affinement.

Tous les affinements dans cette étude seront réalisés avec le même logiciel, et les mêmes types de restrictions et contraintes. C'est le logiciel d'affinement SHELXL [35] utilisant le modèle I.A.M. (voir partie I.2.2.A) qui est ici employé. Ce logiciel affine les structures contre les intensités mesurées F^2 . De plus, au cours de chaque étude de structure de complexe enzyme – inhibiteur, 5 % des observations choisies aléatoirement ne seront pas prises en compte au cours des affinements afin de pouvoir calculer des indices de qualité libres (voir partie I.2.4.A.).

Les restrictions sur les distances, angles, planéité de groupements chimiques, volumes chiraux et agitations thermiques pour les acides aminés sont générées automatiquement par le logiciel SHELXPRO, appartenant à la suite SHELX [35]. Le logiciel n'étant pas en mesure de fournir automatiquement des restrictions pour les inhibiteurs ou autres molécules organiques

présentes dans les structures macromoléculaires, leurs restrictions seront obtenues via le serveur WEB Dundee PRODRG2 (<http://davapc1.bioch.dundee.ac.uk/prodrg/>)[34]. Ce serveur détermine à partir d'une structure de petite molécule des restrictions standard adaptées. Le logiciel SHELXPRO propose également des pondérations par défaut pour les différents types de restrictions. Ce sont ces valeurs de pondération qui seront utilisées, l'objectif étant de se placer dans les conditions d'un affinement standard.

IV.2.2. Outils d'analyse statistique

Les distributions statistiques des variables du modèle au voisinage d'un minimum seront des distributions à maximum unique. Autrement dit, des distributions croissantes jusqu'à une valeur maximale puis décroissantes. Les distributions sont généralement supposées gaussiennes. Pour cette raison, la loi normale est considérée comme référence dans les études statistiques. Cependant, les distributions de grandeurs obtenues par statistiques peuvent présenter des profils gaussiens imparfaits. Les caractéristiques principales d'une distribution sont sa valeur principale (valeur la plus représentative), sa largeur ou dispersion, mais également sa dissymétrie et son aplatissement qui permettent de quantifier cette imperfection.

Afin d'apprécier les différences entre une distribution statistique et un profil idéal gaussien, plusieurs coefficients ou indicateurs mathématiques peuvent être utilisés [18]. Si l'on considère une variable x dont on dispose d'une population de n valeurs, et x_i , la $i^{\text{ème}}$ valeur dans cette variable avec $1 \leq i \leq n$, on peut définir :

- La moyenne m :

La moyenne arithmétique m est le plus connu et le plus rudimentaire des indicateurs.

$$m = \frac{1}{n} \sum_{i=1}^n x_i \quad (4.1)$$

- La médiane M :

La médiane M est le deuxième quartile : si l'on trie les données par ordre croissant, elle correspond à la valeur de l'observation du milieu (ou à la moyenne des deux valeurs du milieu si le nombre d'observations est pair). Comme la moyenne, la médiane est un indicateur de position centrale mais beaucoup plus robuste aux erreurs.

- Le mode ω :

Le mode ω d'une série qualitative ou discrète est la modalité ou la valeur qui enregistre le plus grand effectif. En d'autres termes, le mode d'une loi de probabilité est l'éventualité qui a la probabilité la plus élevée d'être réalisée.

- Les extremums et la largeur :

Les extremums (minimum *Min* et maximum *Max*) de la distribution renseignent sur l'intervalle de valeurs représentées. La largeur d'une distribution correspond à la largeur de l'intervalle de ses valeurs.

- Ecart-type :

L'écart-type σ est une quantité réelle positive utilisée pour caractériser la répartition d'une variable aléatoire réelle autour de sa moyenne. On fait également référence à son carré, appelé « variance » σ^2 .

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - m)^2}{n}} \quad (4.2)$$

Cette expression se révèle être biaisée dans le cas où le nombre de valeurs utilisées est petit. On lui préfère une expression corrigée dite « sans biais » :

$$\sigma^* = \sqrt{\frac{\sum_{i=1}^n (x_i - m)^2}{n-1}} = \sqrt{\frac{n}{n-1}} \sigma \quad (4.3)$$

La moyenne m et l'écart-type σ sont des indicateurs suffisant pour décrire des distributions gaussiennes (ils sont d'ailleurs les uniques paramètres d'une loi normale).

- Le coefficient d'asymétrie de Pearson (Skewness) appelé « γ_1 »

Il s'agit du moment centré d'ordre 3, normalisé par le cube de l'écart-type σ^* , c'est-à-dire :

$$\gamma_1 = \frac{1}{n\sigma^{*3}} \sum_{i=1}^n (x_i - m)^3 \quad (4.4)$$

Il permet d'apprécier la dissymétrie d'une distribution. Le coefficient γ_1 étant normalisé, il est sans dimension et permet donc de comparer des distributions, même d'échelles différentes. Lorsque l'étalement est à gauche (moyenne arithmétique m inférieure à la médiane M), le coefficient d'asymétrie est négatif. Inversement, lorsque l'étalement est à droite, le coefficient est positif. Une distribution suivant une loi normale a un coefficient γ_1 de 0.

De même que l'écart-type, le coefficient γ_1 est biaisé. Une formule corrigée est proposée :

$$\gamma_1^* = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - m}{\sigma^*} \right)^3 \quad (4.5)$$

L'erreur standard sur le coefficient Skewness de Pearson $std(\gamma_1)$ est obtenue suivant la formule : $\sqrt{\frac{6}{n}}$.

- Le coefficient d'aplatissement de Pearson (Kurtosis) appelé « β_2 »

Le coefficient d'aplatissement est le moment centré d'ordre 4 divisé par le carré de la variance σ^{*2} . Comme le coefficient d'asymétrie, le coefficient Kurtosis est sans dimension ce qui permet de comparer des distributions quelques soient leurs échelles.

$$\beta_2 = \frac{1}{n\sigma^{*4}} \sum_{i=1}^n (x_i - m)^4 \quad (4.6)$$

Le coefficient de Kurtosis d'une loi normale est égal à 3. La loi normale étant considérée comme référence, on préfère au coefficient de Kurtosis un coefficient corrigé de manière à être nul pour une loi gaussienne.

Ce coefficient est appelé coefficient « Kurtosis excess » de Pearson. De même que l'écart-type σ et le coefficient γ_1 , il existe une formule pour ce coefficient non biaisée :

$$KurtosisExcess = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_{i=1}^n (x_i - m)^4}{\sigma^{*4}} - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (4.7)$$

Par la suite, on se référera au coefficient « Kurtosis excess » comme étant le coefficient de Kurtosis de Pearson, noté β_2^* .

Un coefficient de Kurtosis de Pearson positif traduit une distribution leptokurtique c'est-à-dire que cette distribution est très piquée (en bleu) (**Figure 1**). A l'inverse, un coefficient négatif correspond à une distribution à « queues épaisses » dite platikurtique (en jaune). Les distributions gaussiennes (en rose) sont mésokurtiques (coefficient de Kurtosis de Pearson nul).

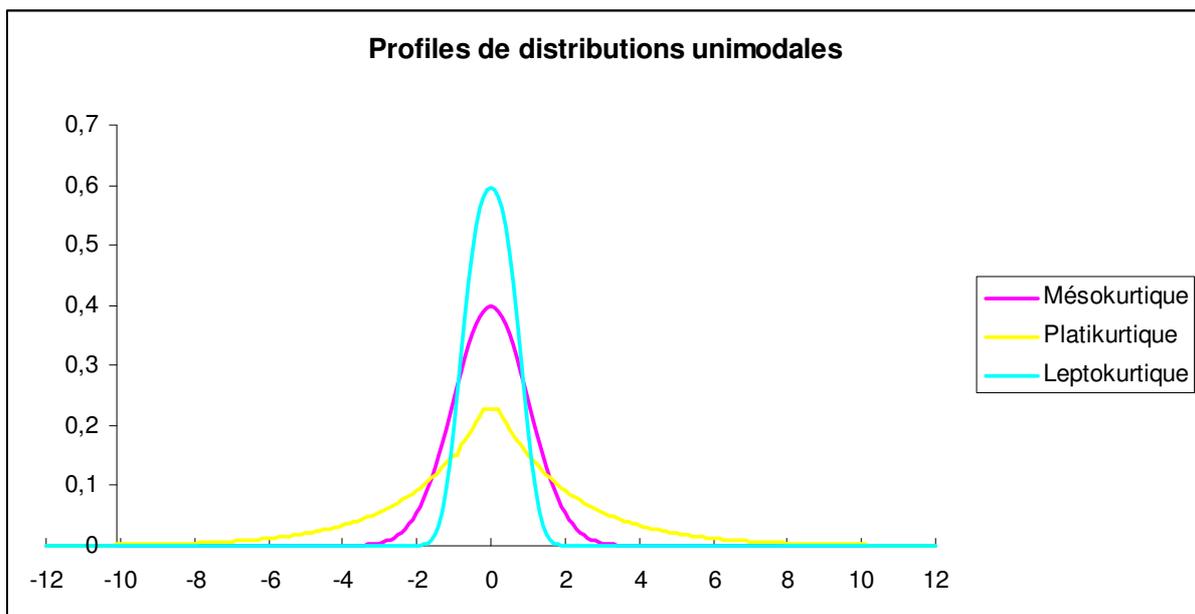


Figure 1 : Profils des distributions unimodales symétriques.

L'erreur standard sur le coefficient Kurtosis de Pearson $std(\beta_2^*)$ est obtenue suivant la

formule : $\sqrt{\frac{24}{n}}$

- Test de la normalité d'une distribution :

Les coefficients de Skewness et de Kurtosis de Pearson fournissent une information sur la dissymétrie et l'aplatissement d'une distribution. Cependant, ces coefficients sont continus et ne donnent donc pas de réponse claire (booléenne) quant à la nature « normale » de la distribution.

Un critère permet de satisfaire à cela : il consiste à calculer les rapports des coefficients de Pearson par les erreurs standards respectives.

$$\text{rapport}(\gamma_1^*) = \frac{\gamma_1^*}{\text{std}(\gamma_1^*)} \quad (4.8)$$

$$\text{et } \text{rapport}(\beta_2^*) = \frac{\beta_2^*}{\text{std}(\beta_2^*)} \quad (4.9)$$

Si ces rapports sont compris entre -2 et 2, les distributions sont considérées comme suivant une loi normale.

Remarque : Ces critères permettent de répondre à la question : « la distribution est-elle non gaussienne ? » en privilégiant les « vrai vrais ». En effet, plus le nombre de structures est élevé, plus les tests sur la nature des distributions sont stricts (Les erreurs standards sur les coefficients de Pearson sont inversement proportionnelles à \sqrt{n}). Une distribution sera d'autant plus facilement considérée comme « imparfaite » que le nombre de structures utilisées pour la statistique sera grand.

IV.2.3. Fonction de perturbation

Dans cette étude, je souhaite focaliser sur le site actif des protéines complexées avec des inhibiteurs. Mais, faut-il pour autant perturber et affiner uniquement ce sous-système ?

Dans un modèle I.A.M., les corrélations significatives sont entre atomes voisins. Cette constatation est d'ailleurs utilisée pour faciliter les affinements de systèmes macromoléculaires par la méthode du gradient conjugué : La méthode « sparse matrix » ou « matrice creuse » consiste à négliger dans la matrice de variance/covariance les termes de

variance entre les paramètres d'atomes éloignés d'une distance supérieure à une certaine limite (usuellement 5 Å) [20]. Cependant, si l'on considère qu'un affinement cristallographique est réalisé dans l'espace réciproque et consiste à minimiser la fonction χ^2 qui n'inclut pas les phases des facteurs de structure, une corrélation persiste entre toutes les parties du système macromoléculaire.

Par conséquent, pour éviter toutes mauvaises estimations dans l'étude statistique, mais aussi pour se placer dans le cadre d'un affinement cristallographique de macrosystème standard dans lequel l'intégralité du système est affinée, il est plus raisonnable de perturber l'ensemble du système.

Quel type et quelle amplitude maximale de perturbation faut-il utiliser ?

Pour répondre à cette question, il faut bien comprendre que les distributions statistiques obtenues ne renseigneront pas strictement sur la quadrature de la région au minimum de la fonction χ^2 , mais sur la capacité de l'algorithme d'affinement utilisé à distinguer deux conformations voisines, ce que l'on pourrait présenter, par analogie avec les instruments optiques, comme son « pouvoir de résolution ». Pour une variable, les différentes valeurs représentées dans sa distribution statistique sont extraites de conformations retenues comme acceptables du point de vue algorithmique. Les distributions vont donc traduire l'efficacité de l'algorithme à faire converger le modèle vers une solution unique.

Perturber les variables d'un système consiste à leur ajouter une petite valeur choisie aléatoirement. Obtenir une série aléatoire de nombre est mathématiquement impossible ; en développement logiciel, on a recours à des fonctions génératrices de nombres pseudo-aléatoires. Ces fonctions sont des formules mathématiques qui permettent d'obtenir des séries de nombres réels suivant une distribution uniforme sur [0 ; 1]. Ces séries de nombre sont de bonnes approximations du hasard sur de larges périodes. A partir de ces générateurs de nombres pseudo-aléatoires, il est possible d'obtenir par simple opération mathématique des distributions de nombres suivant tous les types de distributions usuelles (loi normale, loi gamma, loi de student,...).

Au premier abord, le choix le plus naturel semblait être une loi normale qui est considérée comme la loi de distribution des variables continues aléatoires, mais cela nuirait aux performances de l'étude. En effet :

Premièrement, une loi gaussienne est définie mathématiquement sur $]-\infty ; +\infty[$. Bien que les valeurs nettement éloignées de la moyenne ont une probabilité quasi-nulle, leur apparition dans une série de nombres pseudo-aléatoires peut nuire aux affinements cristallographiques. Après que la structure ait été perturbée, si certains atomes voient leurs coordonnées trop s'éloigner de leurs valeurs initiales, le programme d'affinement sera dans l'incapacité d'appliquer les restrictions et donc de permettre une convergence satisfaisante du modèle.

Deuxièmement, la conformation initiale du système macromoléculaire n'est pas nécessairement le minimum optimal. Or, utiliser une fonction génératrice de nombres pseudo-aléatoires suivant une loi normale favoriserait les faibles perturbations et produirait des conformations perturbées majoritairement voisines de la conformation initiale.

Pour ces raisons, j'ai utilisé au final une fonction génératrice suivant une distribution uniforme centrée en zéro, avec une perturbation maximale ajustable, notée *PerturMax* dans ce rapport. En pratique, j'ai dans ce but développé un programme permettant de perturber les positions atomiques ainsi que les paramètres d'agitation thermique. Ce programme développé en FORTRAN 90 utilise la fonction intrinsèque RANDOM_NUMBER (compilateur Intel Fortran 10.0).

IV.3. Etude préliminaire

IV.3.1. Présentation du complexe aldose réductase – sorbinil

L'enzyme aldose réductase, déjà présentée dans le chapitre III consacré à l'affinement et à l'étude du modèle de distribution de charges de l'un de ces inhibiteurs : le fidarestat, est impliquée dans les complications liées à l'hyperglycémie chronique. Son inhibition est une voie thérapeutique pour traiter ces pathologies.

Parmi les nombreux inhibiteurs testés au cours de ces quinze dernières années, on peut distinguer deux grandes familles d'inhibiteurs : les inhibiteurs ayant un groupement imide cyclique tels que le fidarestat, le sorbinil, le minalrestat, et ceux ayant un groupement carboxylate tels que l'IDD594, le tolrestat, le zopolrestat.

Le mode de fixation de ces inhibiteurs dans le site actif de l'aldose réductase est relativement bien préservé. Les groupements de « tête » : imide cyclique ou groupement carboxylate, interagissent avec le groupement nicotinamide du cofacteur NADP⁺ (**Figure 2**) ainsi qu'avec les résidus His110, Tyr48 et Trp111 (**Figure 3**), qui constituent la partie du site actif appelé communément « poche de fixation anionique ». Ces inhibiteurs présentent un ou plusieurs cycles aromatiques qui interagissent par liaisons hydrogène ou contacts hydrophobes avec les résidus Val47, Phe122 et Leu300 qui constituent la partie du site actif appelée « poche de spécificité ».

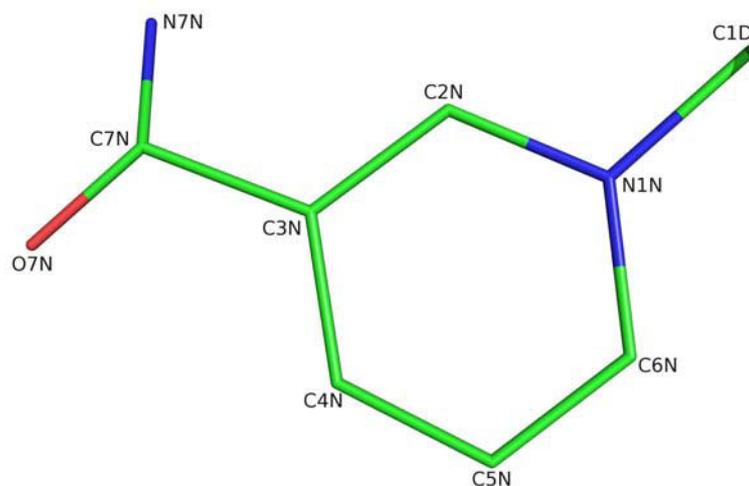


Figure 2 : Illustration du groupement nicotinamide du cofacteur NADP⁺

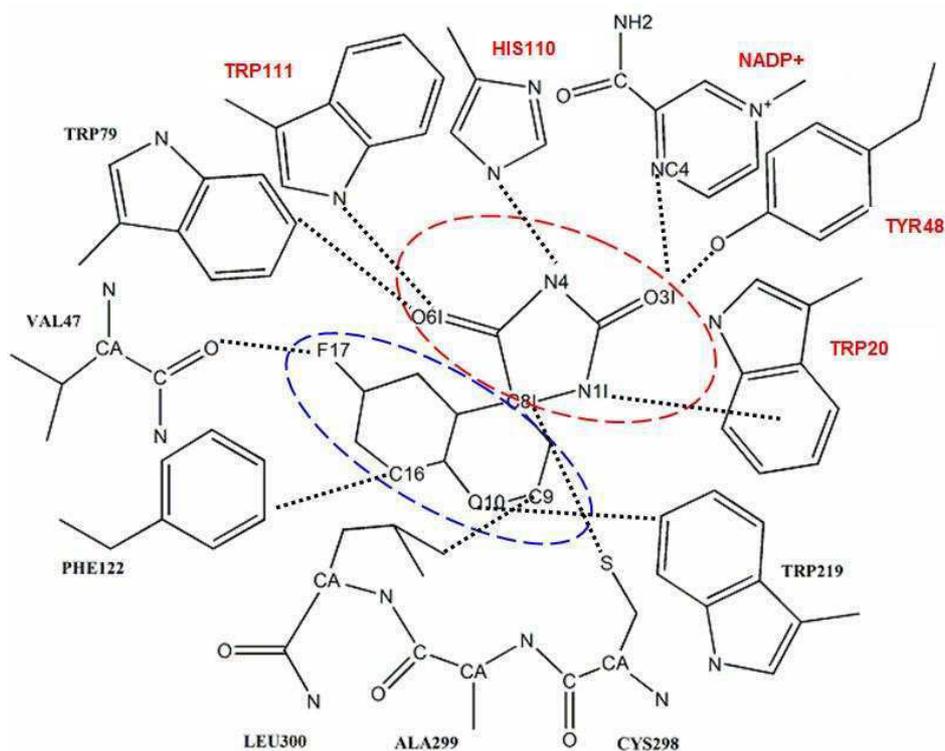


Figure 3 : Diagramme des contacts intermoléculaires impliquant le sorbinil dans le site actif de l'aldose réductase porcine. Les interactions entre les atomes lourds les plus proches sont indiquées par des lignes en pointillés. Les différents groupements chimiques du sorbinil sont précisés par des cercles en pointillés colorés : le groupement hydantoin (imide cyclique) en rouge, le cycle chroman en bleu. Les noms des résidus appartenant à la « poche de fixation anionique » sont indiqués en rouge.

Pour ce test préliminaire, j'ai retenu le complexe aldose réductase porcine – sorbinil. Le sorbinil (**Figure 4**) fait partie des inhibiteurs ayant un imide cyclique comme groupement de « tête ». La structure cristallographique de ce complexe a été obtenue à une résolution de 2,3 Å et étudiée par Urzhumtsev, A. *et al.* [36]. Le groupe d'espace de la structure est $P4_32_12$. Le diagramme de l'inhibiteur sorbinil dans le site actif de l'aldose réductase montre les contacts enzyme-inhibiteur (**Figure 3**). On peut observer sur le diagramme du site actif que le groupement de « tête » hydantoin (un imide cyclique) de l'inhibiteur sorbinil est bien impliqué dans trois liaisons hydrogène avec les résidus His110, Tyr48 et Trp111. Le groupement de « tête » hydantoin forme également une liaison hydrogène supplémentaire avec l'un des cycles de la chaîne latérale du résidu Trp20, ainsi qu'un contact avec le groupement nicotinamide du cofacteur $NADP^+$. Le groupement chroman interagit au sein de la « poche de spécificité » avec plusieurs résidus par contacts courts.

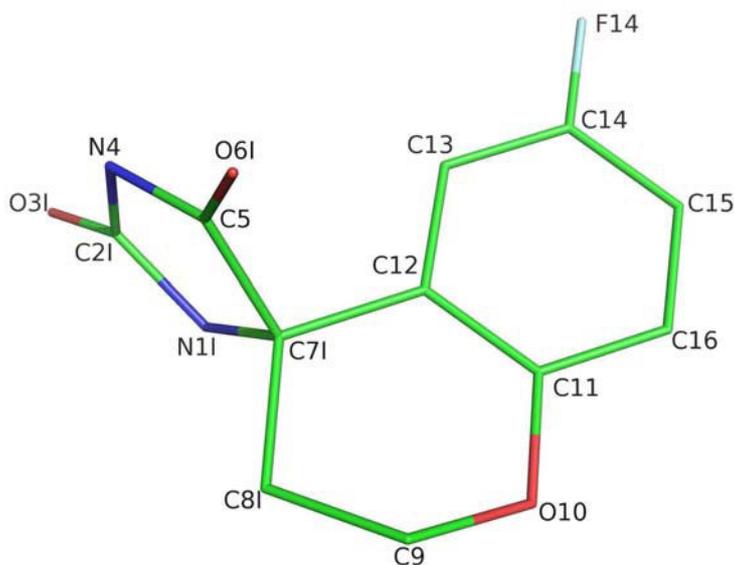


Figure 4 : Représentation de l'inhibiteur sorbinil.

IV.3.2. Préparation « Shake »

IV.3.2.A. Préparation de la structure initiale

Pour l'étude statistique du complexe aldose réductase porcine - sorbinil, la structure initiale a été obtenue à partir du fichier PDB (code : 1AH0) disponible sur le serveur RCSB Protein Databank [4]. Le fichier de réflexions contenant les données expérimentales ($h k l, I(hkl), \sigma(I)$) nécessaires aux affinements a également été téléchargé sur ce même serveur.

Il est important de noter que l'ensemble des affinements a été réalisé sur des structures non protonées quel que soit le système étudié ou la qualité/résolution du jeu de données.

Certaines structures dans la PDB peuvent être partiellement protonées mais, pour l'ensemble des études décrites dans ce chapitre, les atomes d'hydrogène ont systématiquement été supprimés des modèles originaux afin d'étudier les différentes structures de complexes macromoléculaires dans les mêmes conditions d'affinement. Pour la même raison, les paramètres d'agitation thermique sont tous isotropes même si la résolution et la qualité du jeu de données permettent d'avoir recours à des paramètres d'agitation thermique anisotropes pour les atomes non hydrogène.

IV.3.2.B. Sélection du sous-système perturbé et choix des perturbations

Le solvant présent au sein des cristaux de protéines est modélisé à l'aide d'une fonction mathématique qui attribue une densité électronique moyenne à l'environnement des macromolécules. Si le jeu de réflexions le permet, certains atomes d'oxygène des molécules d'eau du solvant peuvent être explicitement modélisés. En raison de leur forte agitation thermique et du bruit expérimental, les atomes d'oxygène du solvant sont souvent placés et optimisés manuellement par observation des cartes de densité électronique résiduelle.

Pour éviter toute complication, les atomes d'oxygène du solvant présents dans le modèle ne sont pas pris en compte dans l'étude et sont donc conservés fixes. Ce choix est raisonnable compte tenu de leurs facteurs B généralement élevés, ces atomes oxygènes ne contribuent que peu aux facteurs de structure du système. Les paramètres d'occupation sont également conservés afin de réduire le nombre total de variables.

Dans le cas du complexe aldose réductase porcine – sorbinil, l'amplitude maximale des perturbations appliquées aux paramètres d'agitation thermique est fixée à 5% relativement à la valeur du paramètre.

La finalité de cette étude étant les incertitudes statistiques sur les énergies, les coordonnées atomiques sont les paramètres qui nous intéressent ici. L'amplitude maximale *PerturMax* des perturbations appliquées aux coordonnées atomiques sera discutée afin d'en déterminer l'influence. Dans un premier temps, cette amplitude maximale *PerturMax* est choisie égale à 0,1 Å compte tenu la moyenne de 0,071 Å obtenue pour les incertitudes expérimentales des coordonnées des atomes des résidus du site actif et de l'inhibiteur dans la structure du complexe initialement téléchargée sur le serveur RCSB [4] calculées à l'aide de SHELXL en maintenant les restrictions [35].

Pour ce premier test, 4000 « perturbations / affinements » ont été réalisés. Sur ces 4000 affinements, aucun n'a abouti à une conformation présentant des aberrations structurales menant à des erreurs de connectivité rendant les restrictions inutilisables par SHELXL [35].

IV.3.3. Etudes statistiques des indicateurs de qualité

Les premiers paramètres pouvant faire l'objet d'une analyse statistique sont bien sûr les indicateurs de qualité : R_1 , R_w et gof (goodness of fit) définis dans la partie I.2.4.A de ce rapport. Les indicateurs libres peuvent être également examinés. Le programme d'affinement SHELXL [35] permet d'obtenir différents indicateurs de qualité en fin d'affinement cristallographique. J'utiliserai les deux facteurs R_1 disponibles : l'un calculé avec toutes les données que j'appellerai R_{1_all} , l'autre avec uniquement les réflexions de rapport $F/\sigma(F) > 4$ appelé R_{1_sigc4} , ainsi que les facteurs R_1 libres correspondants que j'appellerai $R_{1_free_all}$ et $R_{1_free_sigc4}$. Le facteur R_w calculé avec toutes les données sera pris aussi en compte.

Pour finir, j'utiliserai également l'indicateur gof (goodness of fit), ainsi qu'un second gof dit « restreint » calculé en considérant dans son expression les restrictions utilisées pendant l'affinement et que je nommerai gof_{rest} .

On remarque que le «goodness of fit » avec restrictions peut s'écrire :

$$gof_{rest} = \sqrt{\frac{\chi^2}{N_{obs} + N_{rest} - N_{var}}} \quad (4.10)$$

avec N_{obs} , le nombre d'observations utilisées, N_{rest} , le nombre de restrictions utilisées, et N_{var} , le nombre de variables du modèle affinées.

Les coefficients statistiques définis précédemment sont utilisés pour caractériser les distributions de ces indicateurs de qualité.

Un premier examen des résultats statistiques, décrits dans le **Tableau 1**, permet de constater que les indicateurs de qualité présentent des distributions très étroites ce qui était attendu compte tenu de l'étude réalisée ici. En effet, l'ensemble des conformations étant obtenues par « perturbation / affinement », elles sont très proches les unes des autres.

L'examen des valeurs des modes ω et des médianes M nous renseigne sur la nature des distributions. Leurs valeurs sont, dans le cas des six premiers indicateurs ($R_{1_free_sigc4}$, $R_{1_free_all}$, R_{1_sigc4} , R_{1_all} , R_w , gof), très proches des valeurs des moyennes m , à l'écart-type σ^* près, ce qui tend à montrer que leurs distributions sont symétriques.

Indicateurs	$R_{1_free_sigc4}$	$R_{1_free_all}$	R_{1_sigc4}	R_{1_all}	R_w	gof	gof_rest
Moyenne m	0,2748	0,3023	0,2006	0,2131	0,5169	21,842	5,738
Ecart-type σ^*	0,0014	0,0013	0,0004	0,0005	0,0012	0,070	0,065
Mode ω	0,2753	0,3027	0,2006	0,2131	0,5172	21,872	5,722
Médiane M	0,2748	0,3023	0,2006	0,2131	0,5169	21,840	5,727
Minimum	0,2696	0,2976	0,1992	0,2114	0,5124	21,595	5,611
Maximum	0,2790	0,3068	0,2024	0,2150	0,5218	22,118	6,699
Coefficient							
Skewness γ_1^*	-0,119	-0,063	0,083	0,152	0,178	0,201	2,692
Coefficient							
Kurtosis β_2^*	-0,007	-0,021	0,027	-0,030	-0,012	-0,025	21,956
Rapport							
skewnew	-3,1	-1,6	2,1	3,9	4,6	5,2	69,5
Rapport							
Kurtosis	-0,1	-0,3	0,4	-0,4	-0,2	-0,3	283,5

Tableau 1 : Statistiques sur les distributions des valeurs des indicateurs de qualité au cours du test n°1 de l'étude du complexe aldose réductase – sorbinil. Pour 4000 événements, les erreurs standard sur les coefficients de Pearson de Skewness et de Kurtosis sont respectivement de 0,039 et 0,078.

Dans le cas de l'indicateur gof_rest , la médiane M et le mode ω ont des valeurs différentes de la moyenne m même en considérant l'écart-type σ^* comme tolérance. De plus, si l'on classe leurs valeurs, on observe la double inégalité suivante : $m > M > \omega$. Cette double inégalité indique que la distribution est légèrement étalée à droite. Elle doit présenter une dissymétrie avec une « queue » plus épaisse à droite.

Bien que la moyenne m , la médiane M et le mode ω nous renseigne sur la dissymétrie, on utilise les coefficients de Pearson pour une meilleure appréciation des profils des distributions, ainsi que leurs rapports - valeur sur erreur standard - comme cela a été exposé précédemment. Les six premiers indicateurs ont des rapports de Kurtosis nuls alors que leurs rapports de Skewness sont pour cinq d'entre eux en dehors de la tolérance. Ces résultats nous informent que les distributions des indicateurs $R_{1_free_sigc4}$, R_{1_sigc4} , R_{1_all} , R_w et gof sont légèrement dissymétriques, ce qui n'avait pas été détecté à l'aide des valeurs des moyennes, médianes et modes. L'indicateur $R_{1_free_all}$, en revanche, présente une distribution gaussienne.

Je rappellerai que le test sur les rapports des coefficients de Pearson est d'autant plus strict que le nombre de structures utilisées pour la statistique est grand. Compte tenu des 4000 conformations utilisées ici, ces rapports de Pearson restent raisonnables. Seul l'indicateur de qualité gof_{rest} présente un profil bien à part avec des valeurs de rapport de coefficients de Pearson très élevées. On remarque, par ailleurs, un rapport pour le coefficient de Skewness positif ce qui indique un étalement sur la droite et confirme l'interprétation précédente. L'observation des histogrammes de populations des différents indicateurs de qualité valide les conclusions faites à l'aide des données statistiques (**Figure 5**).

Les indicateurs statistiques utilisés ici permettent donc d'apprécier le profil des distributions sans avoir à les illustrer. Ils pourront donc être utilisés pour apprécier la nature des distributions des différentes variables.

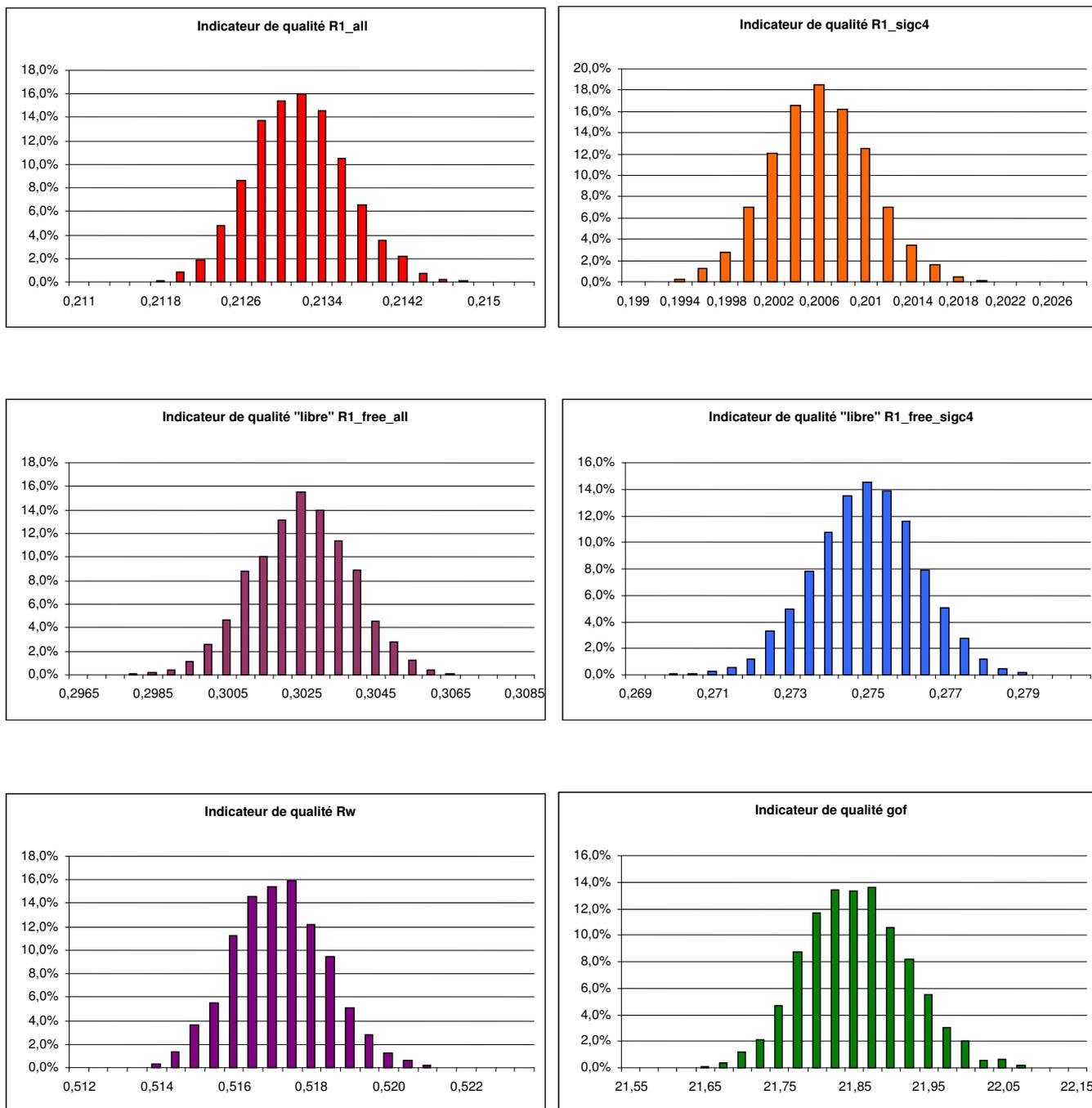
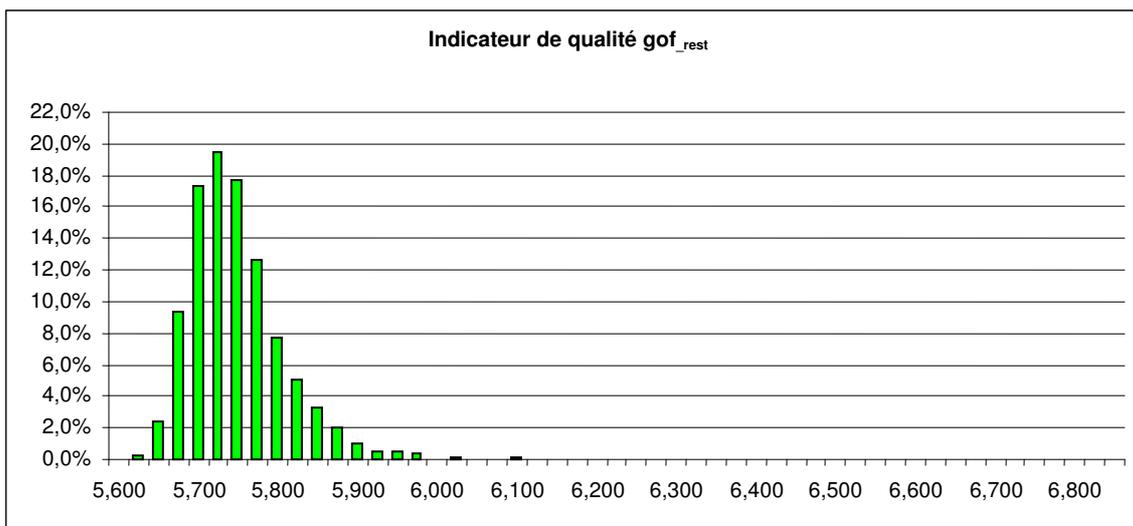
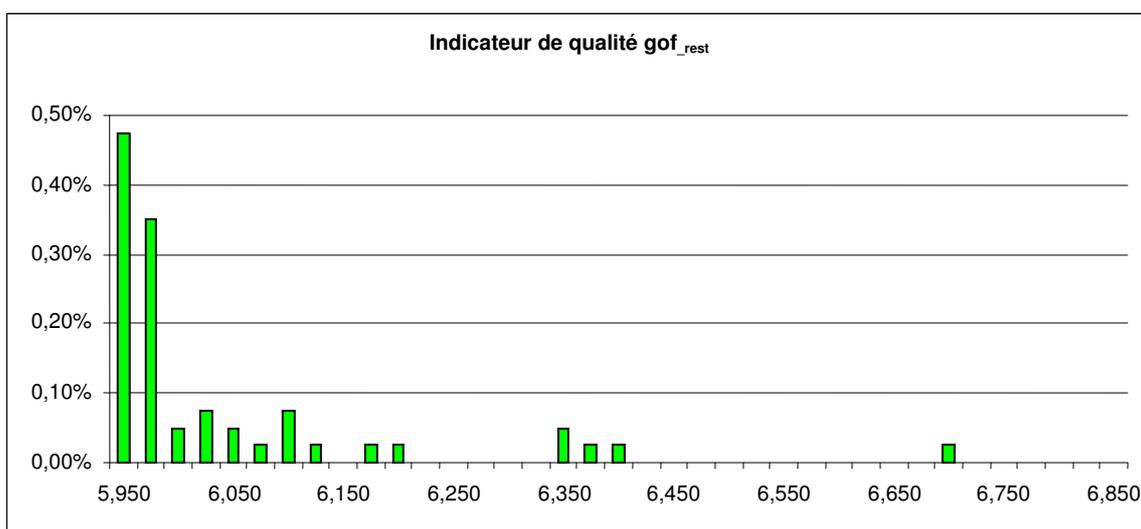


Figure 5 : Histogrammes des populations des valeurs des indicateurs de qualité R_{1_all} , R_{1_sigc4} , $R_{1_free_all}$, $R_{1_free_sigc4}$, R_w et gof obtenues à partir des 4000 conformations du complexe aldose réductase porcine – sorbinil.



(a)



(b)

Figure 6 : Histogrammes des populations des valeurs de l'indicateur de qualité gof_{rest} obtenues à partir des 4000 conformations du complexe aldose réductase porcine – sorbinil. (a) Histogramme défini sur l'ensemble des valeurs de l'indicateur, (b) Histogramme défini sur un sous-intervalle de [5,95 ; 6,85].

Comme déduit des données statistiques, l'indicateur de qualité gof_{rest} présente une dissymétrie nettement appréciable à la vue de ces histogrammes des populations (**Figure 6**). Le second histogramme des populations (**Figure 6b**) réalisé sur un sous-intervalle de la gamme des valeurs du gof_{rest} révèle quelques valeurs éloignées ou « outliers » de la distribution quasi-gaussienne.

Les coefficients de Pearson de Skewness et Kurtosis permettent non seulement d'apprécier le « piqué » et la dissymétrie d'une distribution mais aussi de détecter les « outliers » ou l'existence de modes secondaires (lois multimodales).

L'indicateur de qualité gof_{rest} est un indicateur de « Goodness of Fit » qui prend en compte dans son expression, comme la fonction χ^2 , les termes de restriction utilisés au cours de l'affinement cristallographique de la structure. Les valeurs déviantes aussi appelées « outliers » observés sur l'histogramme des populations du gof_{rest} ne sont pas présentes sur l'histogramme des populations du gof sans restrictions. Ces « outliers » correspondent donc à des structures ne satisfaisant pas les restrictions imposées lors de l'affinement, et qui peuvent donc être rejetées. J'appellerai, par la suite, ces conformations : des « outliers de restrictions ».

Dans le cas-test du complexe aldose réductase porcine – sorbinil, 17 structures peuvent être rejetées si l'on considère 6,0 comme valeur limite de gof_{rest} au-delà de laquelle les structures ne sont pas acceptables. Ces « outliers de restrictions » ne seront plus pris en compte dans la suite de cette étude.

IV.3.4. Etudes statistiques des coordonnées atomiques

Les incertitudes sur les énergies d'interaction électrostatique que je cherche ici à estimer, dépendent directement des incertitudes sur les coordonnées des atomes impliqués. Il convient donc de réaliser une analyse statistique sur ces coordonnées du système macromoléculaire.

Pour des raisons pratiques, je me borne ici aux coordonnées des atomes impliqués dans les interactions au sein du site actif du complexe protéine-inhibiteur, déjà décrites dans la partie IV.3.1 de ce chapitre.

Dans le cas test, ces coordonnées sont celles des atomes de l'inhibiteur Sorbinil, des résidus Trp20, Val47, Tyr48, Trp79, His110, Trp111, Phe122, Trp219, Cys298, Ala299, Leu300, ainsi que du groupement nicotinamide du cofacteur NADP⁺ (**Figure 2**), ce qui constitue un total de 139 atomes non hydrogène (**Figure 3**).

Compte tenu du nombre important de coordonnées atomiques considérées, chaque paramètre ne peut être examiné par le biais d'un histogramme de population. Malgré cela, les indicateurs

statistiques précédemment employés pour l'étude des indicateurs de qualité permettront d'apprécier le profil de leurs distributions.

Remarque : Pour permettre une meilleure appréciation des résultats des analyses statistiques sur les coordonnées atomiques, les valeurs des coordonnées atomiques seront données en Å et non en coordonnées fractionnaires. Ces coordonnées demeureront cependant définies dans le repère de la maille cristallographique qui n'est pas nécessairement orthogonale.

L'analyse statistique des coordonnées atomiques permet de détecter les valeurs éloignées du mode principal. Ces valeurs sont associées à des conformations correspondant à des minima locaux. De telles conformations seront appelées, par la suite, « outliers conformationnels ». Les résultats des analyses statistiques réalisées sur les coordonnées atomiques de l'inhibiteur sorbinil et sur le groupement nicotinamide du cofacteur NADP⁺ sont disponibles en annexe (**Annexe D**) comme exemple.

Parmi l'ensemble des atomes non hydrogène impliqués dans les interactions au sein du site actif, certains présentent des rapports de Kurtosis et/ou de Skewness en dehors de l'intervalle de tolérance [-2 ; 2] ce qui trahit les écarts de leurs distributions par rapport à une loi normale. Pour les coordonnées des atomes de la protéine (site actif et cofacteur NADP⁺), on observe comme valeur de rapport de Kurtosis la plus extrême 11,6 (coordonnée Y de l'atome CG1 de la chaîne latérale du résidu Val47), et de rapport de Skewness 11,1 (coordonnée Z de l'atome C5N du cofacteur). En examinant les histogrammes de populations de ces deux coordonnées atomiques, l'on constate que la coordonnée Y de l'atome CG1 du résidu Val47 n'est pas excessivement « piquée » (**Figure 7**). Cela ne remet pas en question les performances du test sur les rapports de Pearson, mais met en évidence le caractère strict de ces tests pour un nombre élevé d'événements (4000 dans ce cas). En ce qui concerne la coordonnée Z de l'atome C5N du cofacteur, sa distribution devrait présenter un étalement sur la droite compte tenu de la valeur de son coefficient de Skewness. Cette dissymétrie est visible, bien que faible, sur son histogramme des populations. On notera également la présence de quelques valeurs éparses à la droite du mode principal bien qu'invisibles sur la **Figure 8**.

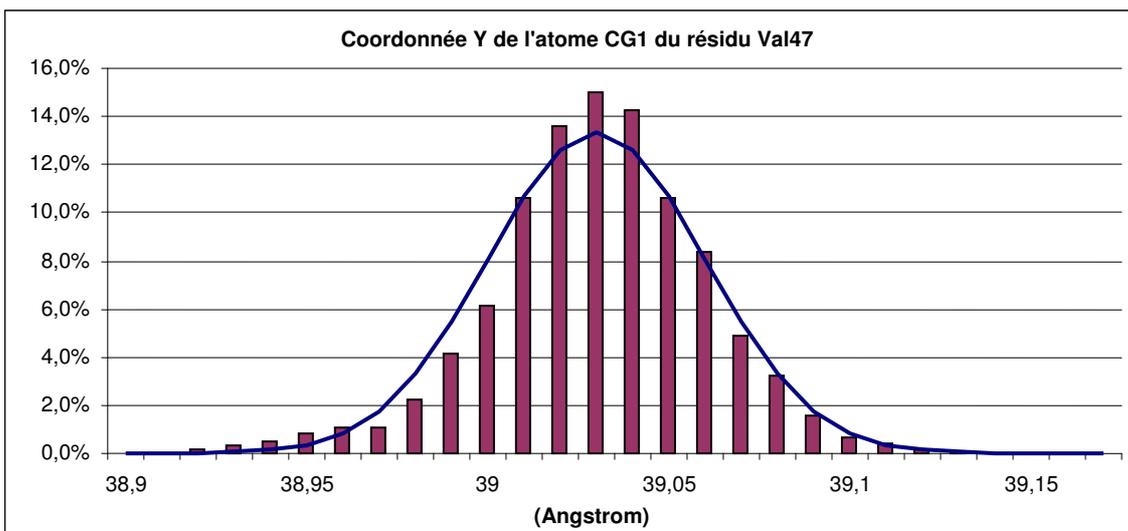


Figure 7 : Histogramme des populations pour la coordonnée Y de l'atome CG1 – résidu 47. La courbe en bleu représente la fonction gaussienne définie avec comme espérance et écart-type, la moyenne et l'écart-type de la distribution.

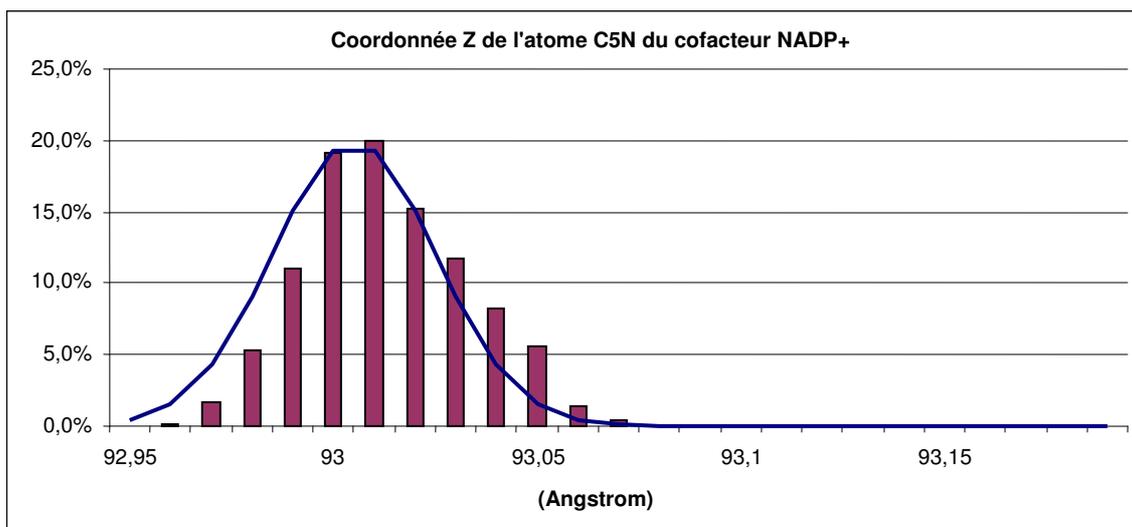


Figure 8 : Histogramme des populations pour la coordonnée Z de l'atome C5N du cofacteur NADP⁺. La courbe en bleu représente la fonction gaussienne définie avec comme espérance et écart-type, la moyenne et l'écart-type de la distribution.

Les coordonnées des atomes de l'inhibiteur sorbinil présentent des valeurs de rapports tout particulièrement aberrantes pour une distribution quasi-gaussienne. Ces valeurs maximales sont considérablement plus élevées avec, par exemple, pour la coordonnée Y de l'atome C9 : un rapport de Kurtosis de plus de 900 et un rapport de Skewness de moins de -100, ou pour la

coordonnée Y de l'atome C13 : un rapport de Kurtosis de plus de 80 et un rapport de Skewness de moins de -40.

Comme cela a déjà été observé dans le cas de l'indicateur gof_{rest} , des valeurs aberrantes de rapports de Pearson peuvent mettre en évidence la présence d'« outliers » ou d'un profil multimodal (la présence de plusieurs pics maximums distincts dans la distribution).

On peut observer sur l'histogramme des populations de la coordonnée Y de l'atome C9 trois « outliers conformationnels » au voisinage de valeur 35,9 Å (**Figure 9**).

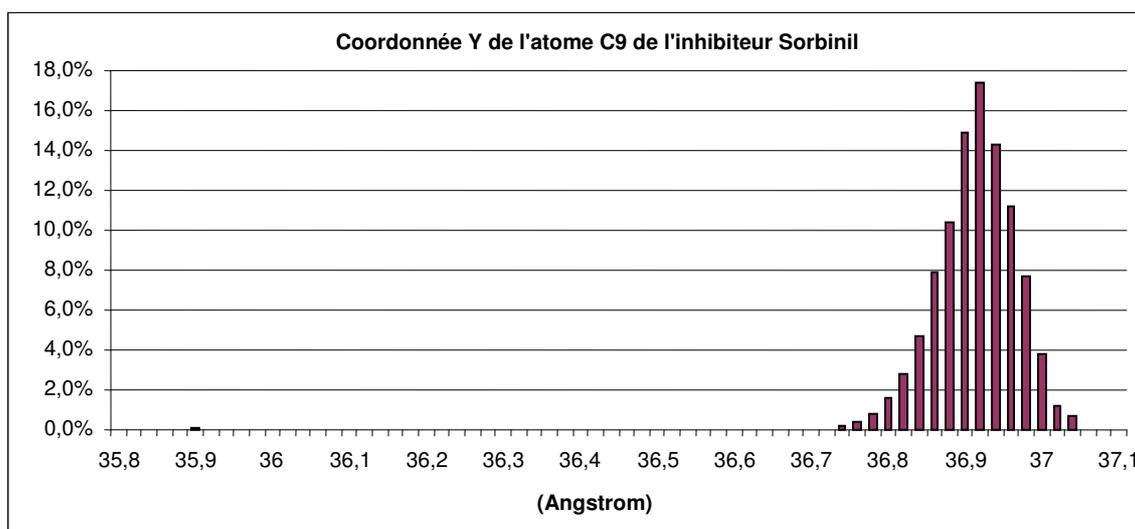


Figure 9 : Histogramme des populations pour la coordonnée Y de l'atome C9 du sorbinil.

Dans le cas de la coordonnée Y de l'atome C13, des valeurs de rapports de Pearson élevés correspondent à un long étalement sur la gauche de faible amplitude (**Figure 10**).

On notera également que l'étalement des distributions sur la gauche est confirmé par la double inégalité : le mode $\omega >$ la médiane $M >$ la moyenne m (**Tableau 2**). Cependant, les différences entre ces trois caractéristiques des distributions sont infimes. Ces exemples des distributions montrent à nouveau la grande sensibilité des coefficients de Pearson de Skewness et de Kurtosis.

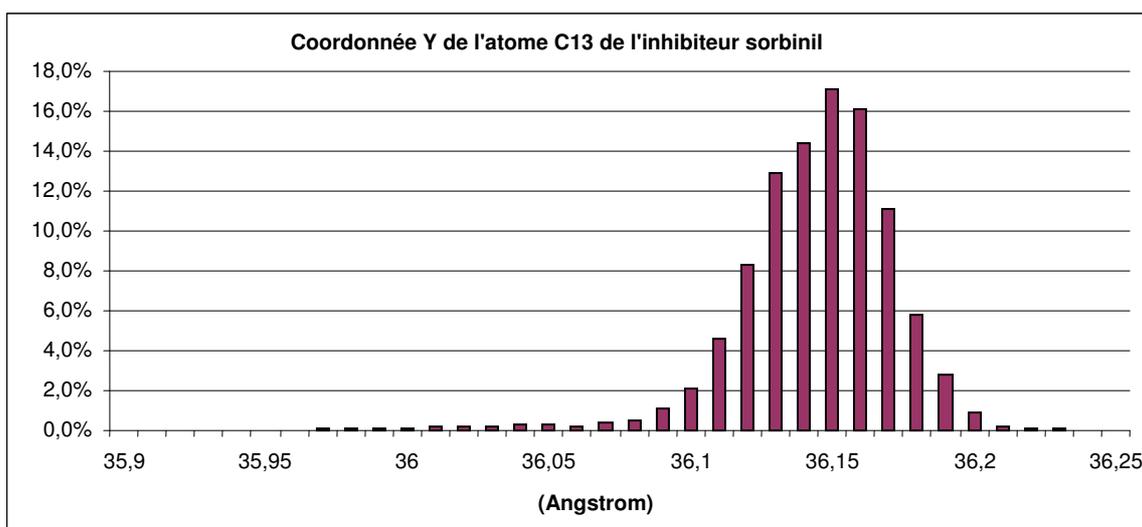


Figure 10 : Histogramme des populations pour la coordonnée Y de l'atome C13 du sorbinil.

	Sorbinil	Sorbinil	Val47	NADP ⁺
Atome	C13	C9	CG1	C5N
Coordonnées	Y	Y	Y	Z
Moyenne m	36,140	36,905	39,026	93,009
Ecart-type σ^*	0,028	0,058	0,030	0,021
Mode ω	36,152	36,916	39,025	92,997
Médiane M	36,142	36,908	39,027	93,007
Minimum	35,939	35,881	38,900	92,953
Maximum	36,228	37,065	39,142	93,140
Coefficient Skewness				
γ_1^*	6,690	71,139	0,900	0,541
Coefficient Kurtosis				
β_2^*	-1,575	-4,283	-0,302	0,432
Rapport de Skewnew	86,2	916,5	11,6	7,0
Rapport de Kurtosis	-40,6	-110,3	-7,8	11,1

Tableau 2 : Analyse statistique réalisée sur les coordonnées atomiques : coordonnée Y des atomes C13 et C9 du sorbinil, coordonnée Y de l'atome CG1 du résidu Val47, et coordonnée Z de l'atome C5N du cofacteur NADP⁺ à partir des données du test n°1.

L'examen des 3 conformations « outliers » observées sur l'histogramme des populations de la coordonnée Y de l'atome C9 du sorbinil révèle qu'elles correspondent à un changement conformationnel du groupement chroman de l'inhibiteur sorbinil. En effet, l'atome C9 a

« basculé » de l'autre côté du cycle (**Figure 11**). Ce sont donc des « outliers conformationnels » et sont pour cette raison rejetés du jeu de structures. Ces outliers n'ont pas été détectés lors de l'analyse statistique des indicateurs de qualité, en particulier du gof_{rest} , car les restrictions sur les distances et les angles valencielles n'imposent pas la position de l'atome C9 par rapport au cycle chroman de l'inhibiteur sorbinil.

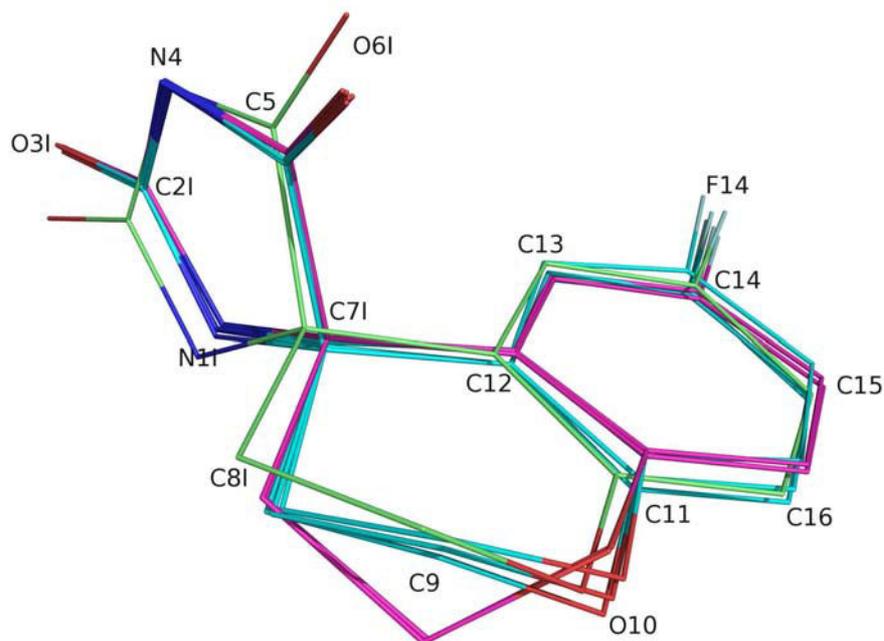


Figure 11 : Illustrations des différents types de conformations obtenues avec : la conformation initiale en vert, trois conformations choisies aléatoirement et correspondant aux modes principaux des distributions des coordonnées atomiques en bleu, et trois conformations « outliers conformationnels » dues à un basculement de l'atome C9 en violet.

Par ailleurs, la suppression de ces trois « outliers conformationnels » fait chuter les rapports de Pearson de Skewness et de Kurtosis de la coordonnée Y de l'atome C9 qui deviennent respectivement de -8 et 2,5.

On constate que les conformations correspondantes aux modes principaux des distributions de coordonnées atomiques sont nettement éloignées de la conformation initiale. En ce qui concerne la coordonnée Y de l'atome C13, la chose est très différente : bien que les rapports de Pearson soient de valeurs significatives, l'étalement sur la gauche de la distribution reste

faible avec une valeur minimale de 92,95 Å pour une valeur de mode de 93,00 Å soit un écart de 0,05 Å.

Au total 20 conformations ont dû être exclues du jeu de structures : 17 « outliers » de qualité (qui ne respectent pas les restrictions utilisées dans l'affinement) et 3 « outliers » conformationnels.

Remarque : Le coefficient de Pearson de Kurtosis s'avère être plus performant que le coefficient de Pearson de Skewness pour la détection d'« outliers » dans une distribution avec des valeurs absolues de rapport supérieures à 100. Cela peut s'expliquer par sa seule définition : le coefficient de Kurtosis correspond au moment centré d'ordre 4 normalisé d'une distribution, les termes sommés dans sa définition sont élevés à la puissance 4 ce qui va exacerber tout écart par rapport à une loi normale.

Malgré l'exclusion de 20 conformations « outliers » du jeu de structures, on note toujours des coordonnées présentant des rapports de coefficients de Pearson significatifs bien que plus faibles que ceux relevés précédemment. D'autres conformations « outliers » sont donc présentes au sein du jeu de structures. Rechercher et exclure manuellement ces « outliers » représente un travail long et fastidieux qui va à l'encontre d'une application de cette démarche à d'autres complexes. La détermination d'une stratégie d'exclusion automatique des « outliers conformationnels » est indispensable à cette étude.

Les distributions observées jusqu'ici présentent un pic principal significatif (mode principal) de profil quasi-gaussien auxquels viennent s'ajouter des pics (ou modes) secondaires de faible amplitude ou quelques événements éparses.

L'objectif est de supprimer ces pics secondaires ou événements isolés dans les distributions des coordonnées des atomes du site actif, et donc de supprimer du jeu de structures, les « outliers conformationnels » correspondants.

Après plusieurs tentatives, deux algorithmes de rejets sont retenus :

- a) Le premier algorithme consiste à identifier pour chaque variable les valeurs appartenant au pic principal (et donc voisines du mode de la distribution) et d'exclure toutes les autres valeurs.

Pour cet algorithme, le principe est donc de :

- déterminer pour chaque distribution la valeur du mode principal,
- rechercher à droite et à gauche les frontières du pic principal (les valeurs de population nulles) qui vont définir l'intervalle des valeurs retenues comme appartenant au pic principal de la distribution,
- et pour finir rejeter les structures correspondant aux valeurs n'appartenant pas à cet intervalle,
- appliquer à nouveau la méthode sur l'ensemble des structures retenues qui ont satisfait le critère.

On applique la méthode récursivement jusqu'à ce qu'il n'y ait plus d'« outliers » dans le jeu de structures, en d'autres mots, jusqu'à convergence du cardinal de l'ensemble des structures retenues.

- b) Le second algorithme consiste à supprimer les valeurs appartenant au pic principal mais nettement excentrées.

La présence d'un pic principal nettement dominant permet d'obtenir une moyenne m de valeur proche du mode principal (ou maximum du pic principal).

On sait que pour une distribution gaussienne de moyenne m et d'écart-type σ^* , plus de 99 % des événements sont compris entre $m - 3 \times \sigma$ et $m + 3 \times \sigma$.

Bien que les pics principaux des distributions ne soient pas strictement gaussiens, on leur prêtera la même propriété. On peut considérer que plus de 99 % des événements d'une distribution quasi-gaussienne seront compris entre $m - 4 \times \sigma$ et $m + 4 \times \sigma$, et définir ainsi un algorithme simple visant à supprimer les structures excentrées de l'échantillon.

Pour cet algorithme, le principe est donc de :

- déterminer pour chaque distribution la moyenne m et l'écart-type σ ,
- rejeter les structures correspondant aux valeurs n'appartenant pas à l'intervalle $[m - 4 \times \sigma; m + 4 \times \sigma]$,
- puis appliquer à nouveau la méthode sur l'ensemble des structures retenues qui ont satisfait le critère.

Comme pour le premier algorithme, on applique la méthode récursivement jusqu'à ce qu'il n'y ait plus d' « outliers » dans le jeu de structures.

En appliquant ces algorithmes récursifs sur le jeu de structures sans les « outliers de restrictions » identifiés et rejetés suite à l'analyse statistique des indicateurs de qualité, on détecte 184 structures « outliers conformationnels ». On trouve parmi ces « outliers » ceux exclus manuellement précédemment. En considérant les 17 « outliers » de qualité, un total de 201 structures a été exclu du jeu des structures initiales soit environ 5% des structures.

Les statistiques sur les coordonnées atomiques à partir des 3799 structures préservées donnent des coefficients de Pearson nettement plus faibles ce qui confirme la bonne exclusion des « outliers ». L'application de ces deux algorithmes permet de supprimer l'étalement sur la gauche de la distribution de la coordonnée Y de l'atome C13 du sorbinil, commenté précédemment (**Figure 12**).

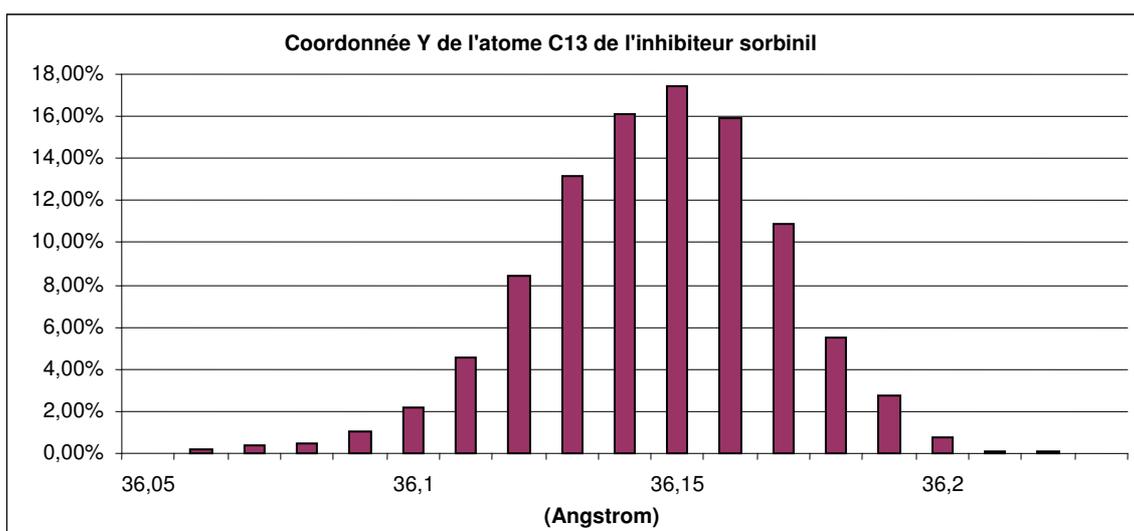


Figure 12 : Histogramme des populations pour la coordonnée Y de l'atome C13 du sorbinil, après exclusion automatique des « outliers conformationnels » au cours du test n°1.

Pour information, le coefficient de Pearson de Kurtosis maximal en valeur absolue est de 1,7 pour la coordonnée X de l'atome CH2 du résidu Trp219, et le coefficient de Pearson de Skewness maximal en valeur absolue est de 0,6 pour la coordonnée Y de l'atome NE2 du résidu His110. Les distributions des populations de ces coordonnées sont également communiquées (Figures 13, 14).

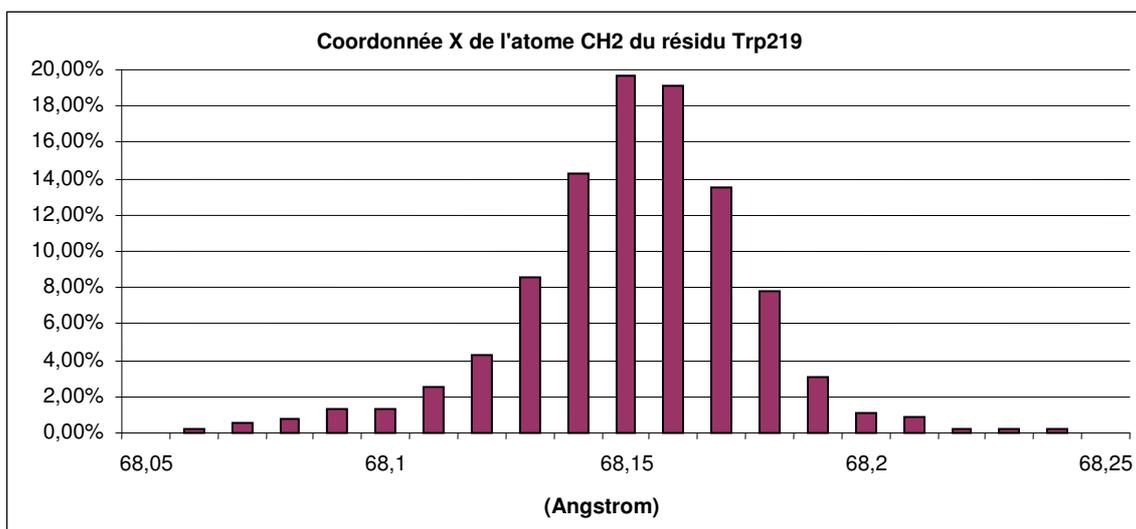


Figure 13 : Histogramme des populations pour la coordonnée X de l'atome CH2 du résidu Trp219, après exclusion automatique des « outliers conformationnels » au cours du test n°1.

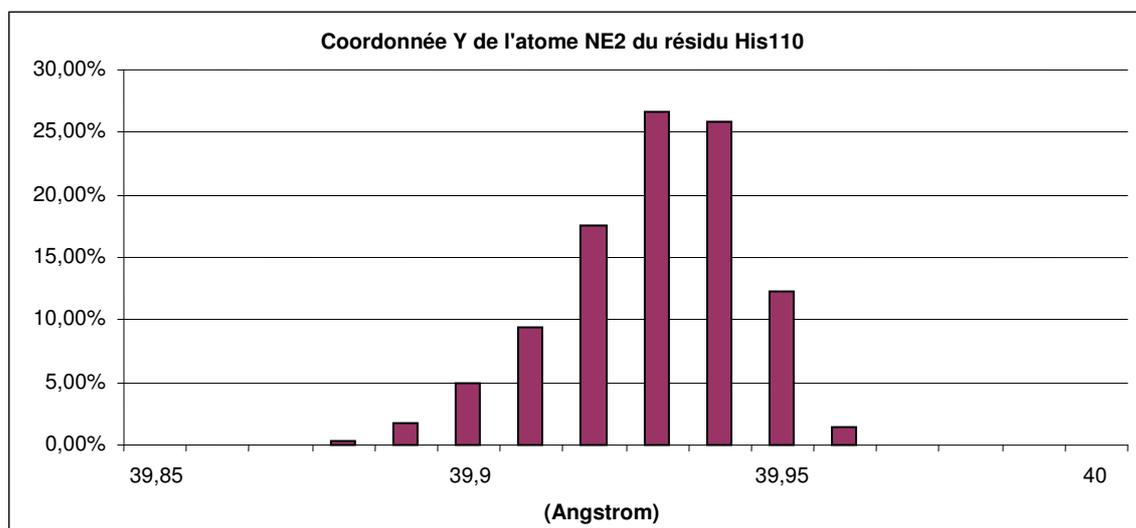


Figure 14 : Histogramme des populations pour la coordonnée Y de l'atome NE2 du résidu His110, après exclusion automatique des « outliers conformationnels » au cours du test n°1.

IV.3.5. Détermination des distributions d'énergies d'interaction électrostatique.

L'objectif demeure, avant toute chose, la détermination des distributions statistiques des énergies d'interaction électrostatique entre l'inhibiteur sorbinil et les principaux résidus du site actif de la protéine.

Les résidus de la protéine en contact avec l'inhibiteur sorbinil ont déjà été identifiés dans la partie IV.3.4 : il s'agit des résidus Trp20, Val47, Tyr48, Trp79, His110, Trp111, Phe122, Trp219, Cys298, Ala299, Leu300, auxquels vient s'ajouter le groupement nicotinamide du cofacteur NADP⁺. L'inhibiteur sorbinil est impliqué dans 12 interactions significatives avec les résidus du site actif de l'aldose réductase porcine (**Figure 3**).

Avant de réaliser un transfert de paramètres de densité de charges à partir de chaque structure retenue, il est nécessaire de protoner ces dernières. La protonation des résidus du site actif ainsi que de l'inhibiteur a été réalisée à l'aide du programme SHELXL [35], qui dispose de plusieurs contraintes géométriques sur les distances atomiques et angles valenciels permettant d'imposer la position d'un atome d'hydrogène dans le repère local du groupe chimique auquel il est lié.

L'orientation des atomes d'hydrogène présentant un degré de liberté est gérée automatiquement par SHELXL. Le programme SHELXL utilise des règles pour orienter l'hydrogène. Par exemple, dans le cas d'un groupement -OH lié à un cycle, seules les deux positions imposant à l'atome d'hydrogène d'être dans le plan du cycle seront envisagées et seront discriminées en fonction de l'environnement de sorte à favoriser la liaison hydrogène la plus forte [1].

L'état de protonation de l'inhibiteur Sorbinil au sein du site actif est imposé : on considère l'inhibiteur sous sa forme neutre et, par la même, le résidu His110 de type HID (protoné une fois sur l'azote ND2).

Une fois les structures protonées disponibles, un modèle de distribution de densité de charges peut être construit par transfert depuis une banque de paramètres. Il existe plusieurs banques de paramètres de distribution de charges, certaines ayant été présentées dans la partie II.2.3. de ce manuscrit. Ici, c'est la banque expérimentale de paramètres multipolaires ELMAM [39;30] qui est actuellement en cours de généralisation au CRM2 qui sera utilisée. L'ensemble

des résidus impliqués dans le site actif ainsi que l'inhibiteur sont représentés dans cette banque. Les modèles de distribution de charges définis à partir des différentes structures atomiques sont obtenus dans le format de la suite de logiciels MoPro [21].

Le logiciel VMoPro, appartenant à cette suite, est employé pour l'ensemble des calculs d'énergie électrostatique. La méthode de calcul appliquée dans cette étude est celle exposée dans la partie II.2.2. Cette méthode permet d'obtenir numériquement l'énergie d'interaction électrostatique avec une précision de l'ordre du $\text{kJ}\cdot\text{mol}^{-1}$. Pour cette raison, les énergies calculées dans cette étude sont toutes arrondies à l'unité près. Les valeurs moyennes sont arrondies au $\text{kJ}\cdot\text{mol}^{-1}$ près et les écarts-types au $\text{kJ}\cdot\text{mol}^{-1}$ supérieur.

Une analyse statistique a été réalisée sur les différentes distributions d'énergie d'interaction électrostatique. Les résultats sont résumés dans le **Tableau 3**. Les rapports des coefficients de Pearson sont communiqués avec un unique chiffre significatif. En effet, les énergies étant définies à l'unité près, il serait déraisonnable d'interpréter ces rapports avec plus de précision. Certaines énergies présentent des écarts-types dont les valeurs sont proches de $0 \text{ kJ}\cdot\text{mol}^{-1}$. Ces écarts-types ne peuvent donc pas être déterminées et sont donc considérés comme égaux à $1 \text{ kJ}\cdot\text{mol}^{-1}$. Les coefficients de Pearson sont, dans ces cas, indisponibles.

On remarquera que l'interaction du sorbinil avec le résidu Ala299 n'est pas renseignée. Cela vient du fait que le contact intermoléculaire n'est pas significatif. Cette interaction est malgré tout présente dans le **Tableau 3** pour préserver la même liste de contacts intermoléculaires que pour les autres complexes protéine – inhibiteur qui seront étudiés par la suite.

Aldose réductase - Sorbinil				Moyenne en kJ.mol ⁻¹	Ecart-type en kJ.mol ⁻¹	Valeur minimale en kJ.mol ⁻¹	Valeur maximale en kJ.mol ⁻¹	Largeur en kJ.mol ⁻¹	Rapport de Skewness	Rapport de Kurtosis
Residu	Atomes non-H les plus proches	Distance (Å)								
His 110	NE2...N4	3,00	-58	4	-75	-48	27	-20	8	
Leu 300	CD1...C9	4,01	-1	1	---	---	1	---	---	
Trp 111	NE...O6I	2,86	-82	3	-93	-70	23	7	8	
Tyr 48	OH...O3I	2,81	-58	5	-81	-39	42	20	20	
Val47	O...F17	3,79	-2	1	---	---	1	---	---	
Cys298	SG...C8I	3,82	10	1	7	11	4	-20	20	
Ala299	N...O20	---	1	1	---	---	1	---	---	
Phe122	CZ...C16	3,63	-9	1	-13	-6	7	-10	3	
Trp20	Ring...N1I	3,2	-87	2	-98	-80	18	-30	40	
Trp219	CH2...O10	3,38	2	1	-1	3	4	---	---	
Trp79	CZ3...O6I	3,38	-6	1	-7	-5	2	---	---	
NADP ⁺	NC4...O3I	3,10	3	1	1	7	6	---	---	

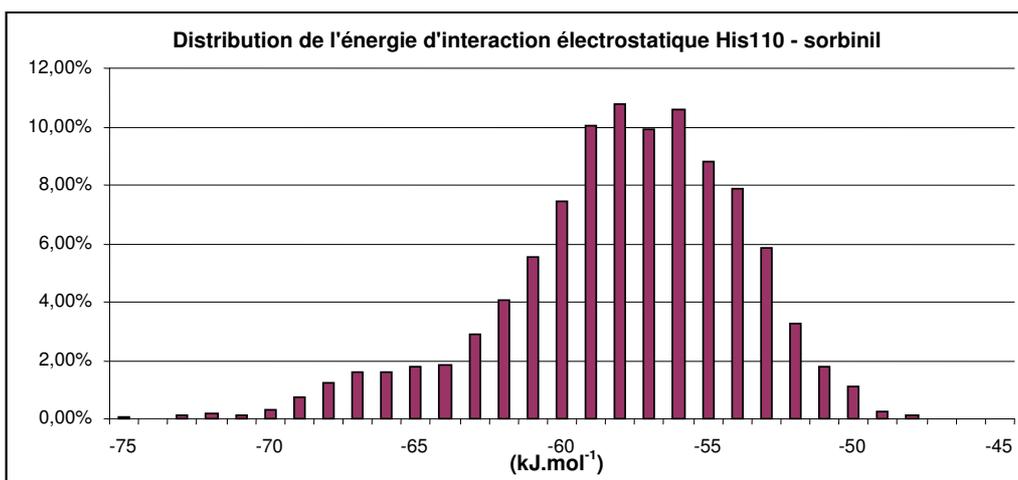
Tableau 3 : Statistiques sur les distributions d'énergie électrostatique obtenues pendant le test n°1 de l'étude du complexe aldose réductase – sorbinil. Toutes les interactions entre l'inhibiteur sorbinil et les résidus du site actif de l'aldose réductase porcine sont décrites. Les atomes non hydrogène les plus proches entre l'inhibiteur sorbinil et les résidus sont donnés avec les distances correspondantes.

Un examen rapide des indicateurs statistiques tabulés nous permet de constater que l'incertitude sur l'énergie est corrélée à la valeur absolue de la moyenne.

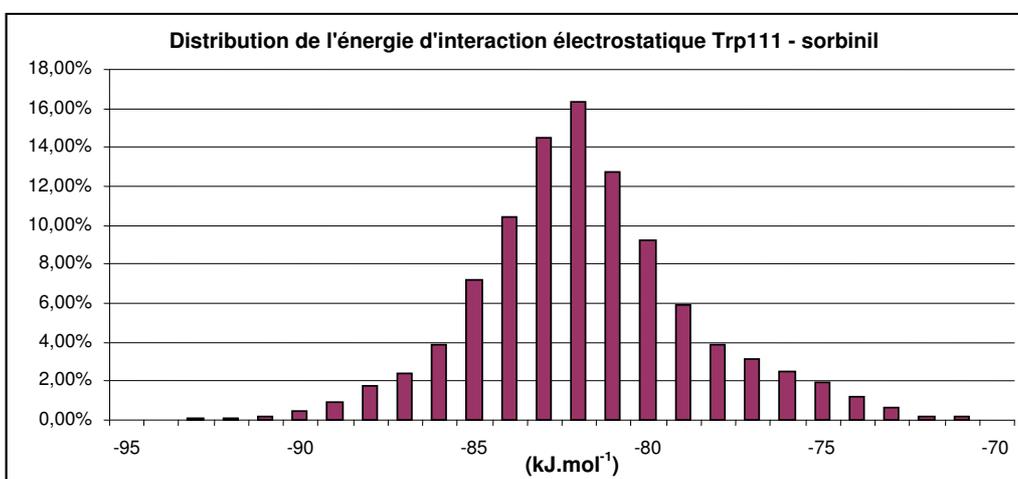
Compte tenu des écarts-types, les énergies d'interaction électrostatique semblent globalement être obtenues avec une bonne précision : seulement quelques $\text{kJ}\cdot\text{mol}^{-1}$ avec un écart-type maximal pour l'interaction Tyr48 – sorbinil de $5 \text{ kJ}\cdot\text{mol}^{-1}$. Néanmoins, si l'on considère chaque conformation du système macromoléculaire obtenue et conservée au cours de cette étude comme satisfaisante d'un point de vue cristallographique. Elle pourrait potentiellement être retenue comme représentative du système macromoléculaire, l'incertitude statistique sur les énergies d'interaction électrostatique correspondra alors à la largeur de leurs distributions (**Tableau 3**). Comme pour les écarts-types, on observe une corrélation entre les largeurs des distributions et leurs moyennes.

De plus, les valeurs des rapports de Kurtosis et de Skewness, bien que données avec un unique chiffre significatif, sont non négligeables. Les distributions des énergies d'interaction électrostatique ne suivent pas une loi normale. Quatre interactions de l'inhibiteur sorbinil présentent des écarts-types et des largeurs de distribution suffisamment importants pour permettre l'examen de leurs histogrammes des populations : celles avec Trp111, His110, Trp20 et Tyr48 (**Figure 15**). Ces interactions correspondent aux trois liaisons hydrogène entre le groupement hydantoin de l'inhibiteur et la « poche de fixation de l'anion » du site actif, ainsi qu'une liaison hydrogène $\text{H}\dots\pi$ entre l'hydrogène H11-N11 de ce même groupement hydantoin et le cycle du résidu Trp20. Ce type de liaison hydrogène ayant un cycle aromatique comme système accepteur, a été rencontré dans l'analyse du modèle de distribution de densité de charges de l'inhibiteur fidarestat. (Chapitre III)

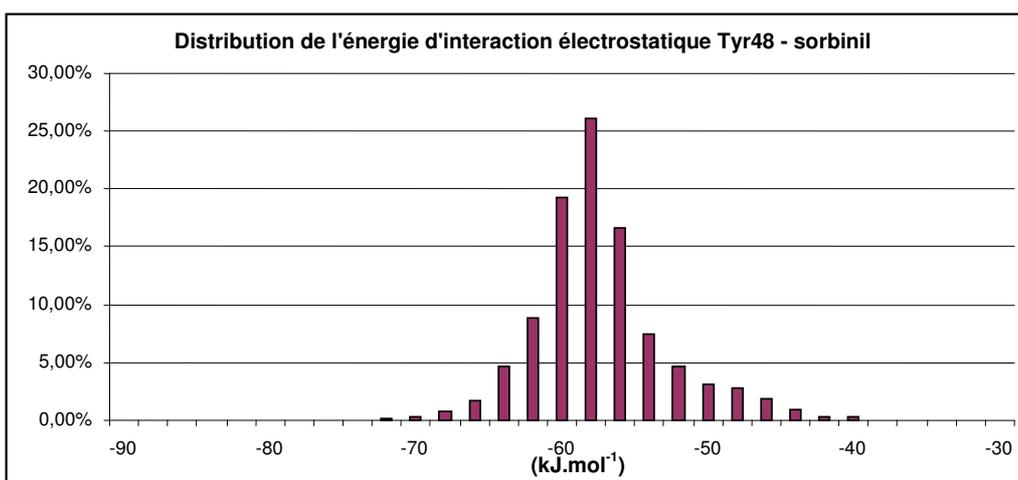
Les histogrammes des populations présentent tous une dissymétrie qui se manifeste par un étalement sur la gauche ou sur la droite selon les cas. Dans le cas, de l'interaction entre le sorbinil et le résidu His110, l'étalement est nettement visible.



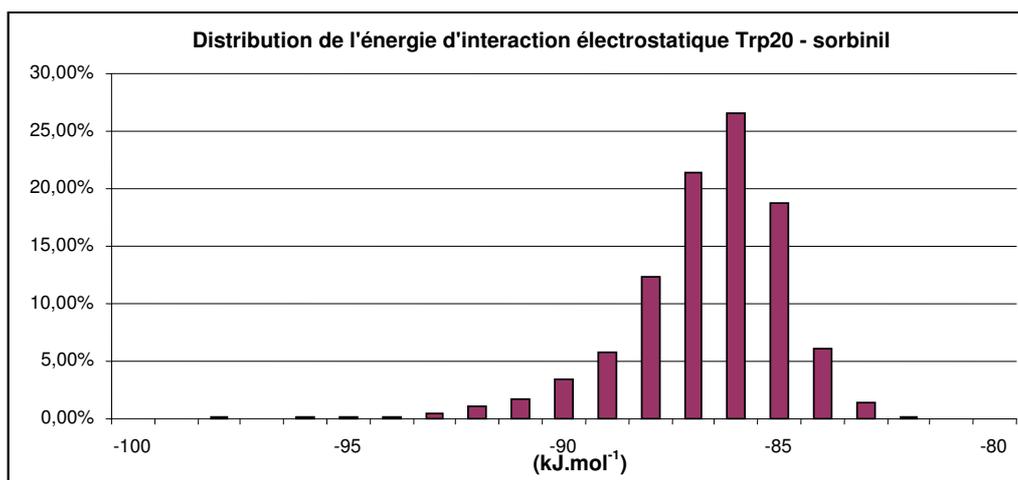
(a)



(b)



(c)



(d)

Figure 15 : Histogrammes des populations des énergies d'interaction électrostatique significatives au sein du site actif du complexe aldose réductase porcine – sorbinil (en kJ.mol⁻¹).

On peut remarquer que l'interaction entre le sorbinil et le résidu Tyr48 est la seule impliquant un atome d'hydrogène présentant un degré de liberté. En effet, l'atome d'hydrogène HH lié à l'oxygène OH de la tyrosine 48 peut être placé, suivant les règles imposées par SHELXL, dans deux positions préférentielles le contraignant à être dans le plan du cycle aromatique de la chaîne latérale du Tyr48. L'orientation des hydrogènes ayant été laissée à la charge du programme SHELXL [35], il convient de vérifier si, dans le cas de cette interaction, la largeur élevée de la distribution de l'énergie d'interaction électrostatique ne résulte pas d'un changement de position de l'atome d'hydrogène HH—OH.

Les résultats de la statistique montrent que les coordonnées de l'atome HH présentent bien un profil unimodal (**Tableau 4**). Les valeurs des moyennes, modes et médianes, ainsi que des rapports de Pearson nous indiquent que les distributions sont quasi-gaussiennes au même titre que celles des atomes lourds. La dissymétrie de la distribution d'énergies n'est donc pas due à un changement de position arbitraire de l'hydrogène HH du résidu Tyr48.

Atome HH du résidu Tyr48			
Coordonnées	X	Y	Z
Moyenne m	59,498	36,979	89,735
Ecart-type σ^*	0,019	0,019	0,013
Mode ω	59,493	36,972	89,739
Médiane M	59,498	36,978	89,734
Minimum	59,432	36,919	89,691
Maximum	59,565	37,053	89,783
Coefficient Skewness γ_1^*	0,03	0,16	0,13
Coefficient Kurtosis β_2^*	0,03	0,08	0,04
Rapport skewnew	0,8	4,1	3,2
Rapport Kurtosis	0,3	1,0	0,5

Tableau 4 : Analyse statistique réalisée sur les coordonnées de l'atome d'hydrogène HH du résidu Tyr48, obtenues au cours du test n°1

En conclusion, les résultats jusqu'ici obtenus montrent que les analyses statistiques réalisées sur les énergies d'interaction électrostatique au sein du site actif présentent des distributions de valeurs larges et cela même en négligeant les queues faibles et étalées comme celle observable sur l'histogramme de la distribution de l'énergie d'interaction électrostatique entre le sorbinil et le résidu Trp20 (**Figure 15d**). Les largeurs des distributions des énergies les plus significatives sont entre 10 et 20 kJ.mol⁻¹, ce qui est comparable à l'énergie d'une liaison hydrogène moyenne-forte (entre 17 et 63 kJ.mol⁻¹) [6] et ne peut donc être négligé dans l'étude des énergies d'interaction au sein d'un complexe enzyme-inhibiteur.

En dépit de cela, les distributions des indicateurs de qualité, après exclusion des « outliers de restrictions », demeurent très étroites (**Tableau 1**). A titre de rappel, les indicateurs R_{1_sigc4} et $R_{1_free_sigc4}$ présentent des valeurs comprises respectivement entre 0,2114 et 0,2150, et 0,2696 et 0,2790. Cela souligne les limites de ces indicateurs en tant qu'indicateurs globaux.

Des dissymétries significatives ont pu être également observées dans les histogrammes des distributions de certaines coordonnées atomiques (**Figures 12, 13**). Il est indispensable de réaliser des tests supplémentaires afin de déterminer si ces dissymétries sont raisonnables, et ainsi de s'assurer que les conditions de l'étude statistique sont acceptables pour obtenir des résultats fiables.

IV.3.6. Discussion des conditions du test

Avant de pousser plus loin l'interprétation de ces résultats, je vais m'efforcer de répondre à plusieurs questions restées jusque là en suspens :

- Quel nombre de conformations est effectivement nécessaire pour obtenir des résultats satisfaisants ?
- Quelle amplitude maximale de perturbation est suffisante dans cette étude ?
- Quelle influence a l'état initial choisi sur les résultats statistiques ?

IV.3.6.A. Quel nombre de conformations est effectivement nécessaire pour obtenir des résultats satisfaisants ?

Dans ce premier test, 4000 affinements ont été initialement réalisés pour aboutir au final à 3799 structures exploitables pour ce premier test. Mais un si grand nombre est-il vraiment nécessaire ? Cette question n'est pas si évidente : le nombre d'événements, ici de structures, nécessaire pour une étude statistique va directement influencer sur la précision ou confiance que l'on pourra accorder aux résultats obtenus.

Quelle « confiance » souhaitons-nous pouvoir accorder à mes résultats ?

Le degré de confiance dépend des objectifs que je me suis fixé : estimer les incertitudes statistiques sur les énergies d'interaction électrostatique. Il est donc nécessaire d'avoir une bonne estimation des moyennes et des écarts-types de chaque distribution.

Cependant, comme cela a été observé précédemment, les profils des distributions d'énergies d'interaction électrostatique ne suivent pas une loi normale et présentent des dissymétries. De plus, dans le cadre d'une étude cristallographique, l'incertitude sur les énergies peut être reliée à la largeur de la distribution plutôt qu'à l'écart-type statistique. Il convient donc de considérer suffisamment d'événements pour disposer de distributions d'énergies dont le profil peut être apprécié. (Les coefficients de Skewness et Kurtosis sont nettement plus sensibles.)

Pour répondre à cette question, des conformations supplémentaires ont été obtenues par « Perturbation / Affinement » de la structure du complexe aldose réductase - sorbinil. Après

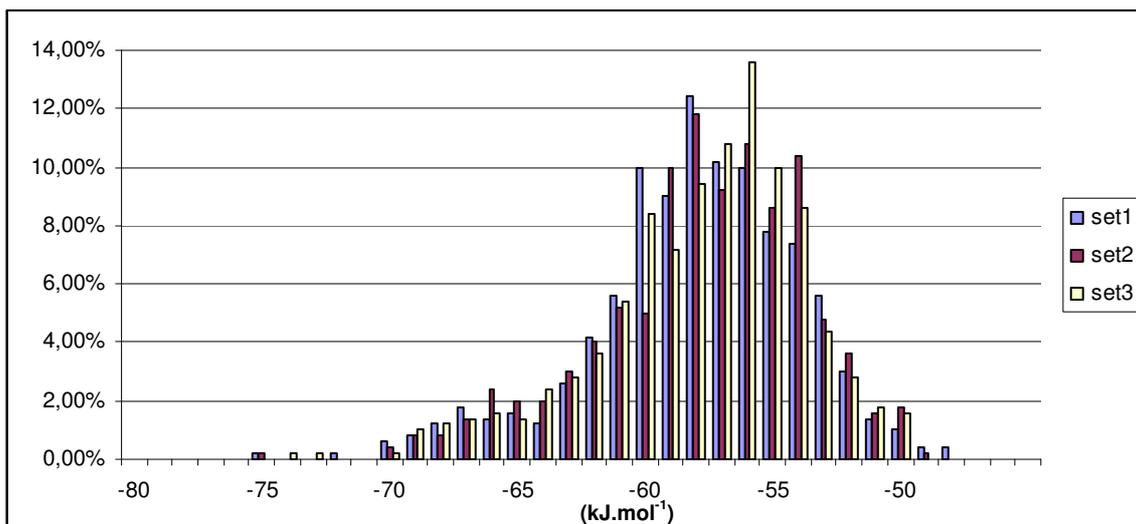
application de la méthode définie précédemment pour supprimer les différents types d'« outliers », près de 10000 conformations du système macromoléculaire étudié ont pu être utilisées.

Afin de déterminer quel cardinal est suffisant pour obtenir des distributions fiables d'énergies d'interaction électrostatique, j'ai testé la reproductibilité de ces distributions.

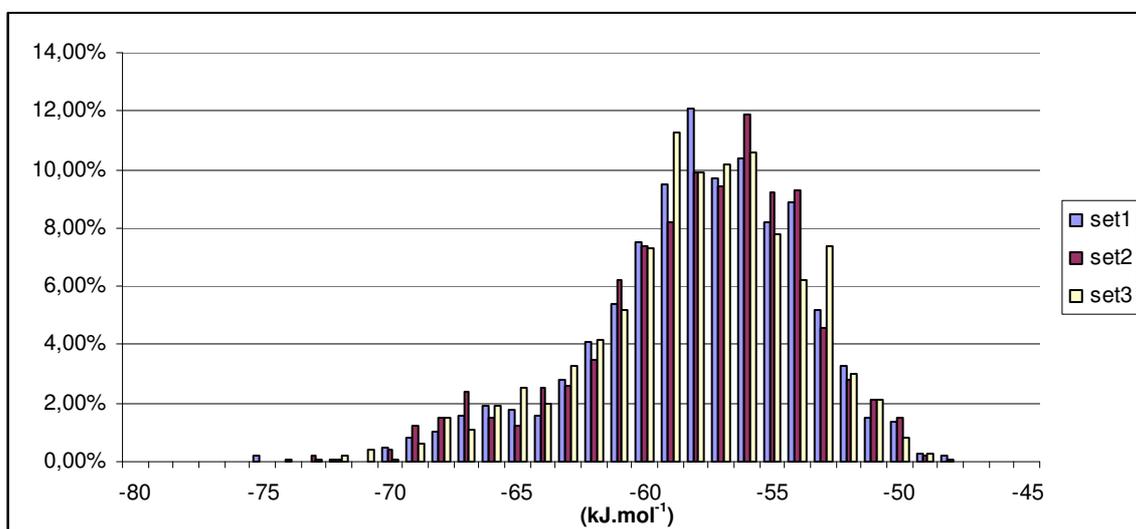
Pour cela, pour différentes tailles N , trois échantillons parfaitement distincts de N conformations ont été définis et leurs distributions d'énergie d'interaction électrostatique représentées.

Je considère, à titre d'exemple, une distribution d'énergie large : celle de l'interaction électrostatique entre l'inhibiteur sorbinil et le résidu Trp111.

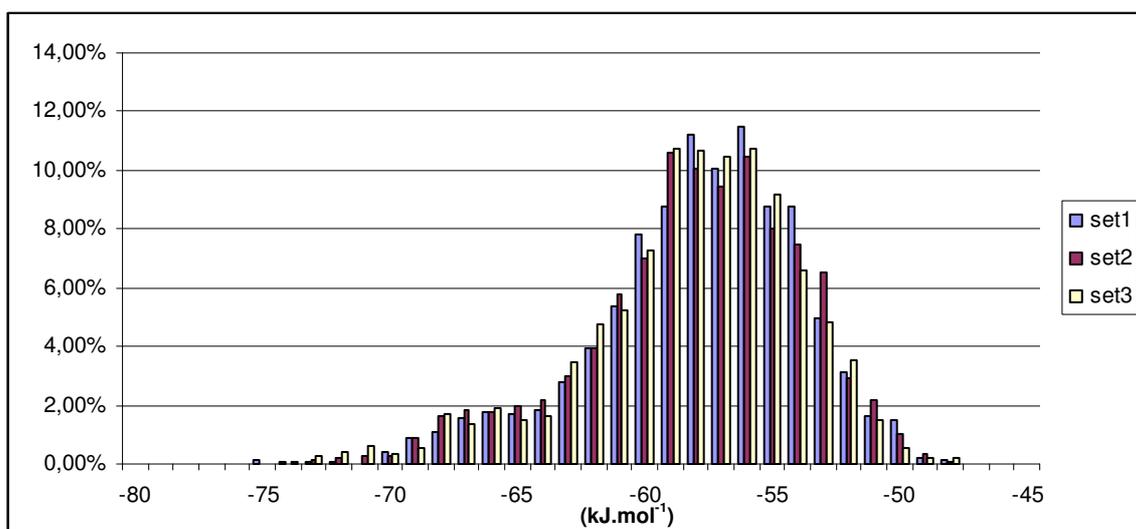
Des jeux de 500 structures sont suffisants pour obtenir les moyennes, les écarts-types, les modes et médianes des distributions des énergies d'interaction électrostatique. Les coefficients de Pearson sont plus difficilement reproductibles ce qui peut s'expliquer par leur définition, mais aussi par les données analysées. Les coefficients de Pearson sont des moments d'ordre 3 et 4, ce qui les rend nettement plus sensibles aux variations de la distribution. De plus, les énergies sont toutes arrondies au kJ.mol^{-1} , la statistique est donc réalisée sur des valeurs entières, ce qui nuit à la reproductibilité de ces coefficients. Le nombre d'évènements requis sera, par conséquent, déterminé par observation des histogrammes des populations.



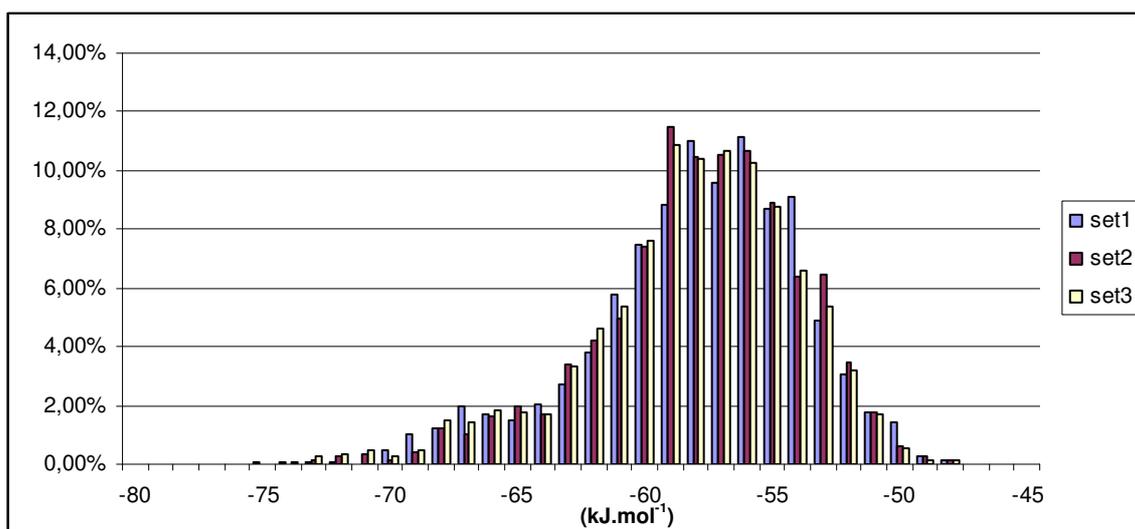
(a) Pour 500 événements



(b) Pour 1000 événements



(c) Pour 1500 événements



(d) Pour 2000 événements

Figure 16 : Histogrammes des populations de l'énergie d'interaction électrostatique entre le sorbinil et le résidu His110 pour différentes tailles d'échantillon (test n°1).

Pour 500 et 1000 événements (**Figures 16a, 16b**), on note la présence de modes secondaires pour certains jeux de données. Pour 500 événements, les distributions des jeux 1 et 3 présentent des maximums secondaires au voisinage de la valeur -60 kJ.mol^{-1} . Pour 1000 événements, la distribution du jeu 3 a un maximum local en -53 kJ.mol^{-1} .

Pour 1500 et 2000 événements (**Figures 16c, 16d**), les trois distributions ont toutes le même profil.

Environ 1500 conformations semblent donc être un nombre approprié d'événements pour obtenir des distributions d'énergies, et en déduire par la statistique les caractéristiques des énergies d'interaction électrostatique. L'examen des distributions des énergies pour les autres interactions électrostatiques confirme ce choix.

Remarque : Chercher à obtenir le profil des distributions d'énergie peut sembler beaucoup trop exigeant, quelques centaines d'événements permettant d'obtenir les principales caractéristiques (moyenne, mode, écart-type et largeur) qui nous intéressent ici. Cependant, ces énergies étant, par définition, des sommes de contributions pseudo-atome - pseudo-atome, il est important de savoir si la présence d'un mode secondaire est légitime ou n'est due qu'à une imprécision des histogrammes liée à un nombre d'événements considérés insuffisant.

IV.3.6.B. Quelle amplitude maximale autoriser aux perturbations de la structure avant affinement ?

Ce test a pour but de déterminer si une perturbation d'amplitude maximale 0,1 Å est suffisante pour obtenir des distributions quasi-gaussiennes représentatives.

Un second test a donc été réalisé en considérant 2500 structures ce qui est, compte tenu des conclusions du test précédent, largement suffisant. Dans ce test, les coordonnées des atomes faisant partie du site actif ont été perturbées avec une amplitude maximale *PerturMax* de 0,2 Å. Le reste du système protéine-ligand est toujours perturbé mais en considérant une amplitude maximale de perturbation de 0,1 Å. A la fin des affinements, 14 structures ont été exclues du jeu de structures car présentant une géométrie qui ne permet pas à SHELXL [35] d'appliquer certaines restrictions.

- Etudes statistiques des indicateurs de qualité

Indicateurs	$R_{1_free_sigc4}$	$R_{1_free_all}$	R_{1_sigc4}	R_{1_all}	R_w	<i>gof</i>	<i>gof_rest</i>
Moyenne m	0,2746	0,3019	0,2006	0,2132	0,5169	21,831	5,744
Ecart-type σ^*	0,0013	0,0013	0,0004	0,0005	0,0013	0,079	0,156
Mode ω	0,2743	0,3020	0,2007	0,2131	0,5170	21,786	5,692
Médiane M	0,2746	0,3019	0,2006	0,2132	0,5169	21,830	5,711
Minimum	0,2688	0,2962	0,1992	0,2117	0,5129	21,589	5,595
Maximum	0,2795	0,3064	0,2021	0,2147	0,5206	22,072	7,017
Coefficient							
Skewness γ_1^*	-0,054	0,017	-0,015	0,097	0,085	0,110	4,459
Coefficient							
Kurtosis β_2^*	0,151	0,107	0,055	0,019	-0,181	-0,256	21,687
Rapport							
Skewnew	-1,11	0,35	-0,31	1,96	1,74	2,23	90,77
Rapport							
Kurtosis	1,53	1,09	0,55	0,19	-1,84	-2,60	220,72

Tableau 5 : Statistiques sur les distributions des valeurs des indicateurs de qualité au cours du test n°2 de l'étude du complexe aldose réductase – sorbinil. Pour 2500 événements, les erreurs standard sur les coefficients de Pearson de Skewness et de Kurtosis sont respectivement de 0,049 et 0,098.

Comme dans le test précédent, l'analyse statistique des indicateurs de qualité communiqués par SHELXL, et en particulier du gof_{rest} a été réalisée (**Tableau 5**).

Les statistiques obtenues pour les indicateurs $R_{1_free_sigc4}$, $R_{1_free_all}$, R_{1_sigc4} , R_{1_all} , R_w et gof sont très proches de ceux obtenus précédemment, les distributions ont un profil gaussien avec des moyennes, écarts-types, médianes et modes similaires. Comme précédemment, seul l'indicateur gof_{rest} fait exception. On constate également que les valeurs de son écart-type et son coefficient de Skewness de Pearson ont doublés.

L'observation de l'histogramme de populations de l'indicateur gof_{rest} révèle la présence d'un second mode au voisinage de 6,6.

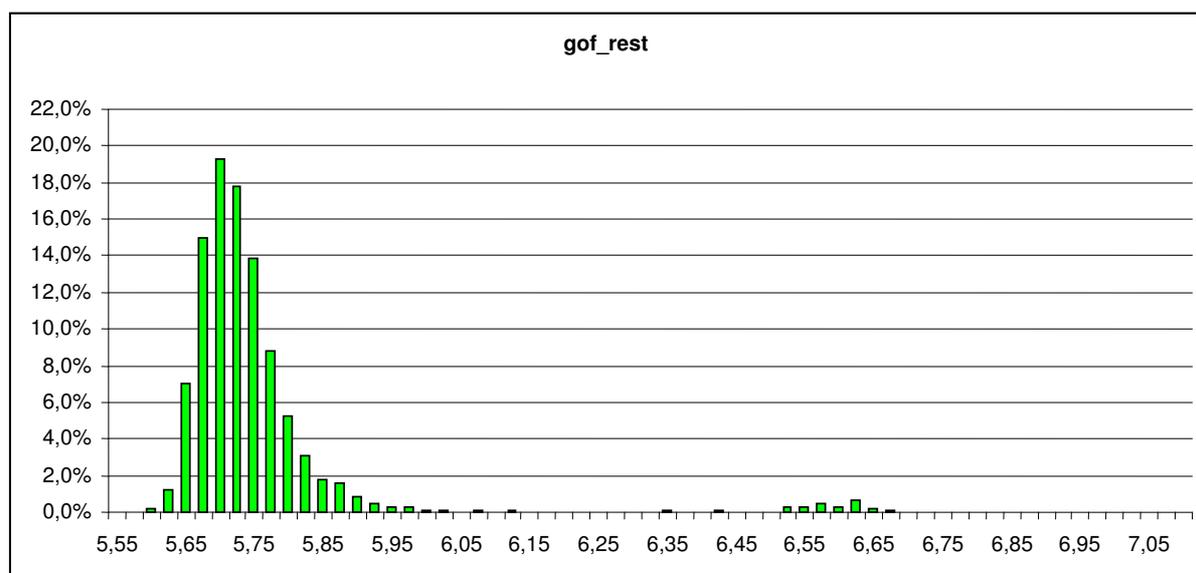


Figure 17 : Histogrammes des populations des valeurs de l'indicateur de qualité gof_{rest} obtenues à partir des 2500 conformations du complexe aldose réductase porcine – sorbinil avec une perturbation maximale de 0,2 Å.

Un examen de la distribution de gof_{rest} permet d'identifier un nombre de « outliers » plus élevé que précédemment. En effet, dans le premier test, 17 structures sur les 4000 avaient été rejetées en raison d'un gof_{rest} trop élevé, soit 4 pour 1000.

Ici, en utilisant le même critère de rejet (valeurs de gof_{rest} supérieure à 6,0), on relève 83 structures « outliers de restrictions » sur 2500 structures initiales soit plus de 3%. Si l'on n'y

ajoute les « outliers » détectés par examen des fichiers log de SHELXL, on obtient 97 structures « outliers de restrictions » au total, soit presque 4 %.

Cette augmentation du nombre de « outliers » peut s'expliquer par les conditions de ce second test : l'amplitude maximale des perturbations introduites étant plus élevée, la stratégie et l'algorithme d'affinement en sont d'avantage éprouvés.

- Etudes statistiques des coordonnées atomiques

En appliquant les algorithmes de détection et de rejet automatique des « outliers conformationnels », 226 structures sont identifiées comme « outliers » ce qui nous laisse au final 2177 structures pour mener cette étude.

Examiner les résultats de l'analyse de la statistique sur les coordonnées des atomes retenus comme contribuant au site actif du complexe aldose réductase – sorbinil, serait complexe et peu explicite. Afin de déterminer l'influence de l'amplitude de perturbation sur les distributions, je propose de comparer les moyennes et écarts-types obtenus lors des deux tests afin d'identifier toutes différences significatives, à l'aide de plusieurs quantités :

- La moyenne arithmétique $diff(m)_{moy}$ des différences relatives des moyennes des coordonnées atomiques :

$$diff(m)_{moy} = \frac{1}{n} \sum_{i \in \text{coordonnées}} \left| \frac{m_{i,1} - m_{i,2}}{\sigma_{i,moy}} \right| \quad (4.11)$$

avec n , le nombre de coordonnées

$m_{i,1}$, la valeur moyenne de la $i^{\text{ème}}$ coordonnée obtenue dans le test 1 (perturbation maximale $PerturMax = 0,1 \text{ \AA}$)

$m_{i,2}$, la valeur moyenne de la $i^{\text{ème}}$ coordonnée obtenue dans le test 2 ($PerturMax = 0,2 \text{ \AA}$)

et $\sigma_{i,moy}$, l'écart-type moyen de la $i^{\text{ème}}$ coordonnée, défini comme $\sigma_{i,moy} = \frac{\sigma_{i,1} + \sigma_{i,2}}{2}$

- La différence relative maximale $diff(m)_{max}$ des moyennes des coordonnées atomiques :

$$diff(m)_{max} = \max_{i \in \text{coordonnées}} \left| \frac{m_{i,1} - m_{i,2}}{\sigma_{i,moy}} \right| \quad (4.12)$$

Ces grandeurs permettent d'apprécier les différences des moyennes en considérant leurs écarts-types moyens comme tolérances.

On peut calculer également :

- La moyenne $diff(\sigma)_{moy}$ des différences relatives des écarts-types des coordonnées atomiques :

$$diff(\sigma)_{moy} = \frac{1}{n} \sum_{i \in \text{coordonnées}} \left| \frac{\sigma_{i,1} - \sigma_{i,2}}{\sigma_{i,moy}} \right| \quad (4.13)$$

avec n , le nombre de coordonnées

$\sigma_{i,1}$, l'écart-type de la $i^{\text{ème}}$ coordonnée obtenue dans le test 1 (perturbation maximale $PerturMax = 0,1 \text{ \AA}$)

$\sigma_{i,2}$, l'écart-type de la $i^{\text{ème}}$ coordonnée obtenue dans le test 2 ($PerturMax = 0,2 \text{ \AA}$)

et $\sigma_{i,moy}$, l'écart-type moyen de la $i^{\text{ème}}$ coordonnée, défini comme $\sigma_{i,moy} = \frac{\sigma_{i,1} + \sigma_{i,2}}{2}$

- La différence relative maximale $diff(\sigma)_{max}$ des écarts-types des coordonnées atomiques :

$$diff(\sigma)_{max} = \max_{i \in \text{coordonnées}} \left| \frac{\sigma_{i,1} - \sigma_{i,2}}{\sigma_{i,moy}} \right| \quad (4.14)$$

En déterminant ces différentes grandeurs pour les coordonnées des 139 atomes retenus comme contribuant aux interactions au sein du site actif, on constate que les écarts-types sont relativement bien préservés avec, au maximum, une différence relative $diff(\sigma)_{max}$ de 27 % et une moyenne arithmétique $diff(\sigma)_{moy}$ de 7 %.

La moyenne arithmétique $diff(m)_{moy}$ des différences relatives des moyennes est nettement plus élevée avec 34 %. Cela signifie qu'en moyenne les distributions des coordonnées atomiques sont superposées, et que les moyennes sont égales à l'écart-type moyen près.

Cependant, les moyennes s'écartent l'une de l'autre pour certaines coordonnées. On obtient en effet une différence relative maximale de moyennes $diff(m)_{max}$ de 106 %. Cette valeur correspond à la coordonnée Y de l'atome C5N du groupement nicotinamide du cofacteur NADP⁺.

On observe aussi une valeur supérieure à 100% (104%) pour la coordonnée Y de l'atome C4N de ce même groupement.

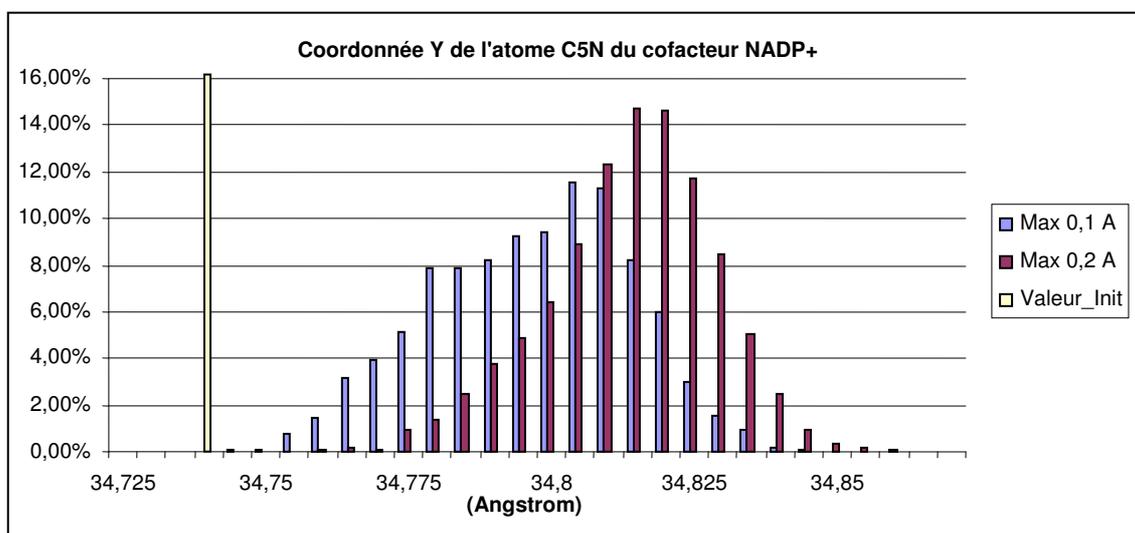


Figure 18 : Histogrammes des populations pour la coordonnée Y de l'atome C5N du cofacteur NADP⁺, obtenus aux cours des deux tests ($PerturMax = 0,1 \text{ \AA}$ et $PerturMax = 0,2 \text{ \AA}$). La valeur initiale de la variable est également indiquée.

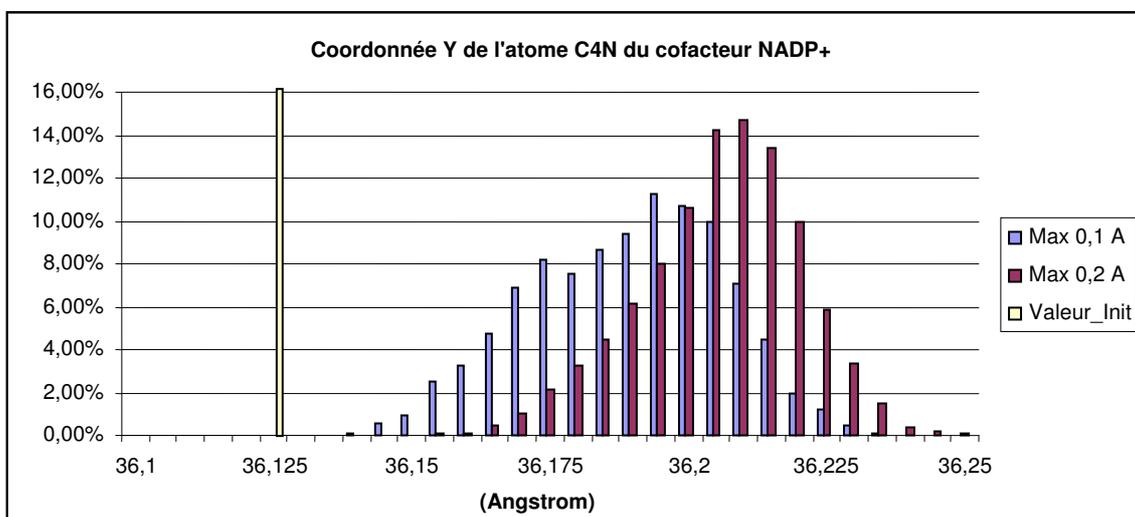


Figure 19 : Histogrammes des populations pour la coordonnée Y de l'atome C4N du cofacteur NADP⁺, obtenus aux cours des deux tests ($PerturMax = 0,1 \text{ \AA}$ et $PerturMax = 0,2 \text{ \AA}$). La valeur initiale de la variable est également indiquée.

L'examen des histogrammes de populations de ces deux atomes (**Figures 18 et 19**) met en évidence le déplacement des modes des distributions. On constate également que les

distributions sont plus étroites dans le cas du second test avec des écarts-types de 0,015 Å alors que ces derniers sont de 0,017 Å dans le premier test.

Les valeurs initiales de ces coordonnées sont indiquées sur leurs histogrammes respectifs. L'étalement des distributions est orienté du côté où sont localisées les valeurs initiales des coordonnées. Ce constat nous amène à nous interroger sur un dernier aspect qui peut potentiellement avoir un impact sur les incertitudes des coordonnées atomiques et, par la même, sur les incertitudes des énergies d'interaction électrostatique : le choix de l'état initial. Ce point sera abordé dans le paragraphe suivant.

- Etudes statistiques des énergies d'interaction électrostatique

Les résultats obtenus suite à l'analyse statistique sur les coordonnées atomiques tendent à montrer des différences plus ou moins significatives de distributions. Qu'en est-il des distributions des énergies d'interaction électrostatique ?

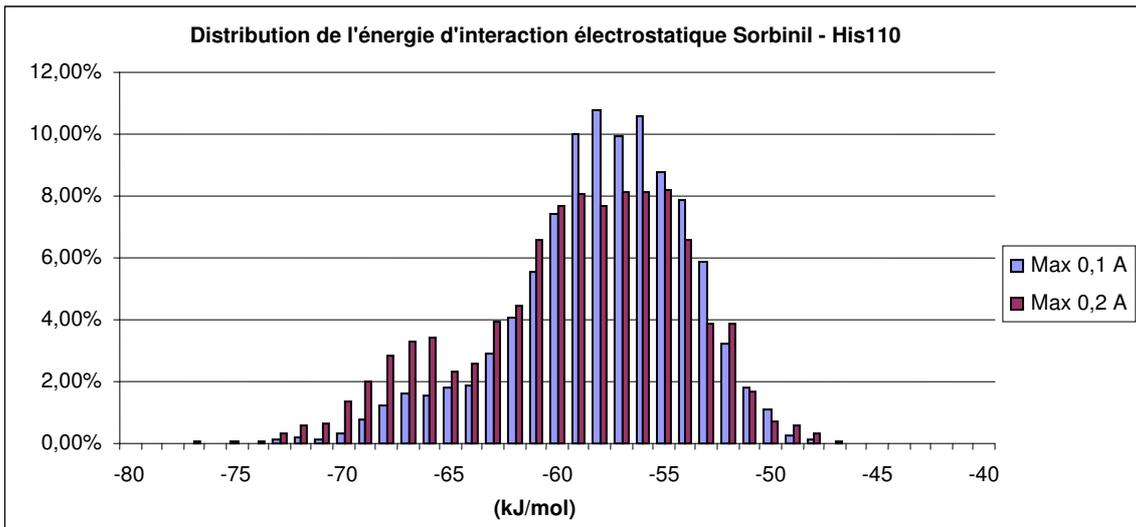
Les valeurs des moyennes et des écarts-types sont différentes, et les profils des distributions sont également affectés. Néanmoins, les largeurs des distributions, que l'on relie à l'incertitude statique, sont préservées (**Tableau 6**).

Bien que les profils des distributions d'énergies soient globalement préservés, on remarque des modes principaux moins intenses. Dans le cas de l'énergie d'interaction entre le sorbinil et le résidu His110, le mode secondaire à -66 kJ.mol^{-1} est plus élevé. De plus, la distribution de l'énergie d'interaction entre le sorbinil et le résidu Trp111 voit son mode principal se déplacer légèrement sur la gauche, et la distribution de l'énergie d'interaction entre le sorbinil et le résidu Tyr48 a un étalement sur la droite plus prononcé (**Figure 20**).

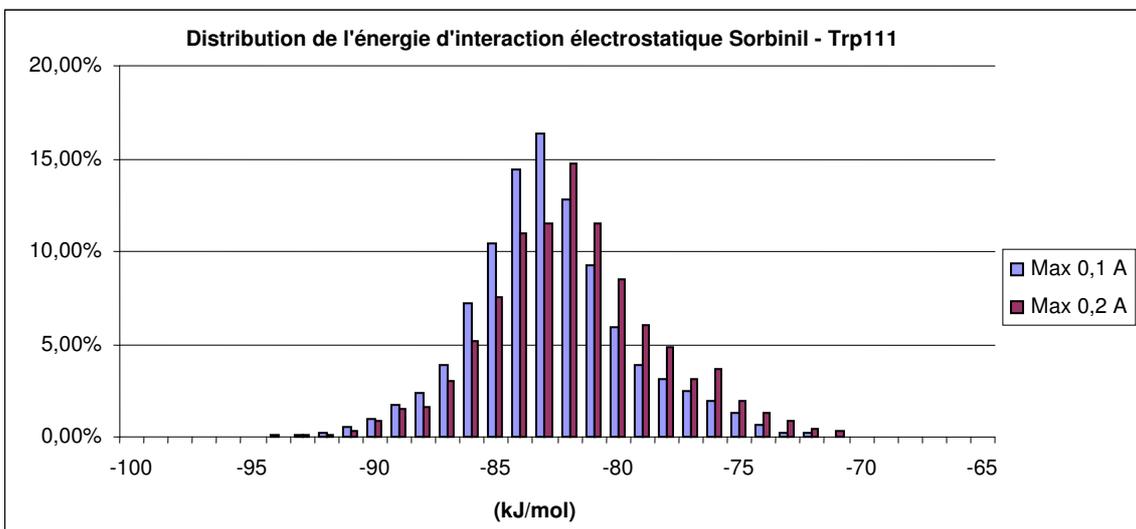
Les valeurs élevées de rapports de Pearson pour l'interaction entre le sorbinil et le résidu 298 trahissent la présence de « outliers » dans la distribution d'énergie. Ces derniers sont dûs à la liberté rotationnelle de l'hydrogène HS—SG de la chaîne latérale du résidu cystéine. Comme expliqué précédemment, l'orientation des atomes d'hydrogène présentant un degré de liberté rotationnel est laissée au soin du logiciel SHELXL [35].

Aldose réductase - Sorbinil				Moyenne en kJ.mol ⁻¹	Ecart-type en kJ.mol ⁻¹	Valeur minimale en kJ.mol ⁻¹	Valeur maximale en kJ.mol ⁻¹	Largeur en kJ.mol ⁻¹	Rapport de Skewness	Rapport de Kurtosis
Residu	Atomes non-H les plus proches	Distance (Å)								
His 110	NE2...N4	3,00	-59	5	-77	-47	30	-9	-2	
Leu 300	CD1...C9	4,01	-1	1	---	---	---	---	---	
Trp 111	NE...O6I	2,86	-82	4	-96	-69	27	4	4	
Tyr 48	OH...O3I	2,81	-57	6	-80	-5	75	20	50	
Val47	O...F17	3,79	-2	1	---	---	---	---	---	
Cys298	SG...C8I	3,82	10	1	-5	12	17	-200	1000	
Ala299	N...O20	---	1	1	---	---	---	---	---	
Phe122	CZ...C16	3,63	-9	1	-14	-6	8	-10	5	
Trp20	Ring...N1I	3,2	-86	2	-96	-81	15	-20	20	
Trp219	CH2...O10	3,38	2	1	0	4	4	---	---	
Trp79	CZ3...O6I	3,38	-6	1	-7	-5	2	---	---	
NADP ⁺	NC4...O3I	3,10	3	1	1	7	6	---	---	

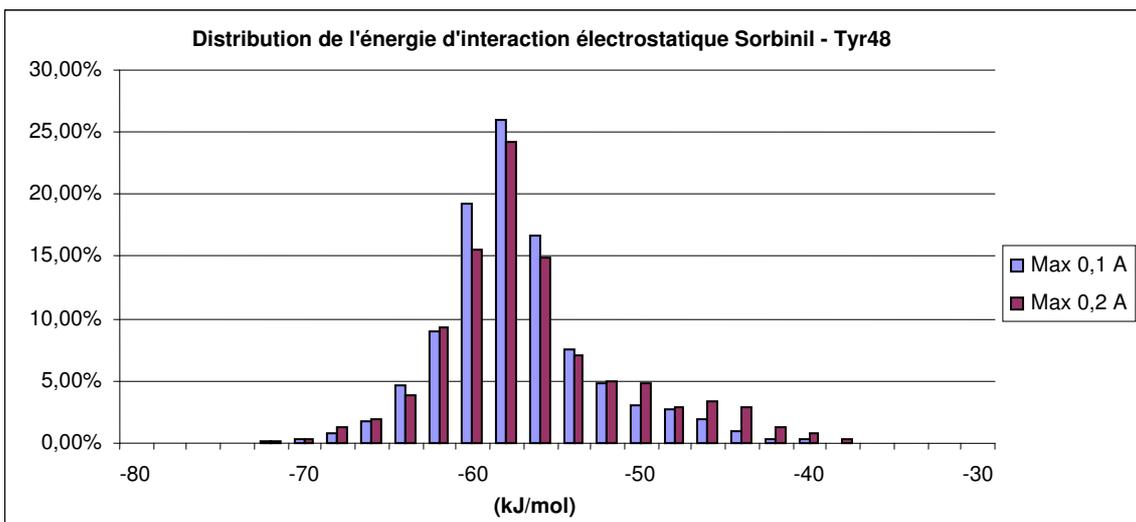
Tableau 6 : Statistiques sur les distributions des énergies électrostatiques obtenues pendant le test n°2 de l'étude du complexe aldose réductase – sorbinil. Toutes les interactions entre l'inhibiteur sorbinil et les résidus du site actif de l'aldose réductase porcine sont décrites. Les atomes non hydrogène les plus proches entre l'inhibiteur sorbinil et les résidus sont donnés avec les distances correspondantes.



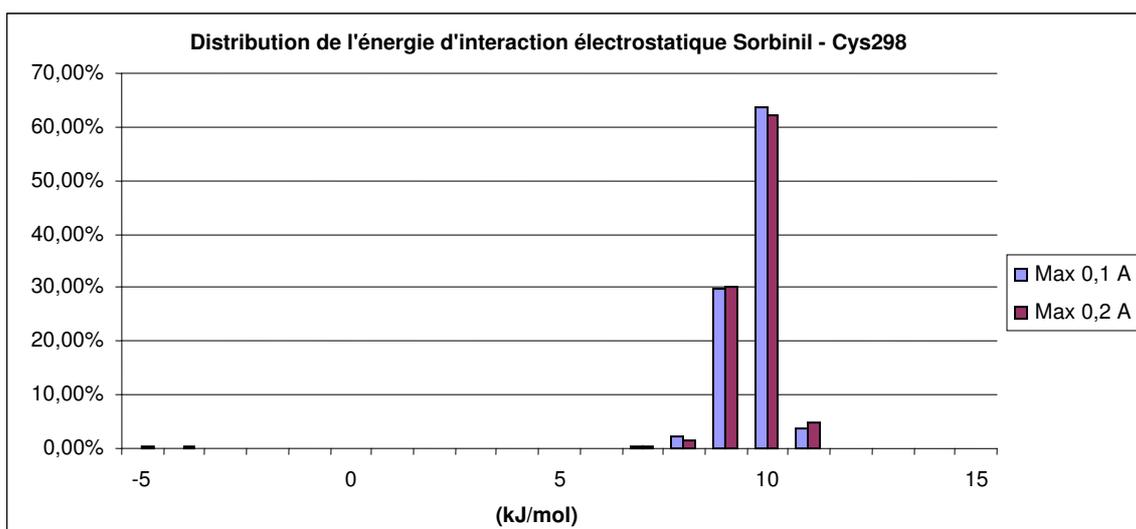
(a)



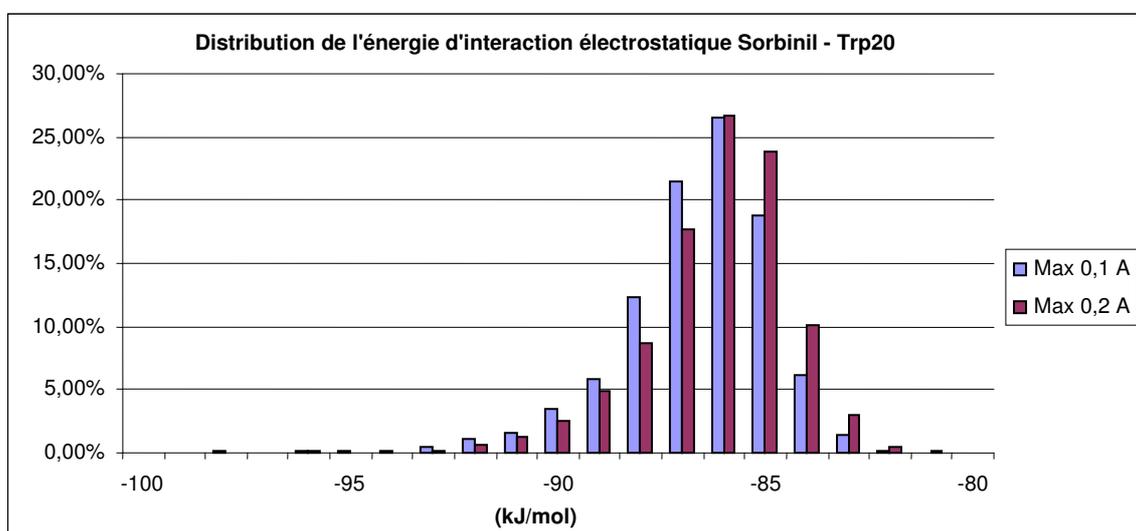
(b)



(c)



(d)



(e)

Figure 20 : Histogrammes des populations des énergies d'interaction électrostatique significatives au sein du site actif du complexe aldose réductase porcine – sorbinil (en $\text{kJ}\cdot\text{mol}^{-1}$) obtenus au cours des tests n°1 et n°2.

IV.3.6.C. Influence de l'état initial sur le résultat de la statistique

Les histogrammes des populations des coordonnées Y des atomes C5N et C4N du cofacteur NADP⁺ commentés précédemment, montrent que l'état initial utilisé peut avoir un impact sur les résultats en particulier lorsque les valeurs initiales des coordonnées sont nettement différentes des valeurs des modes principaux de leurs distributions. En d'autres termes, si la structure initiale au sein du site actif n'est pas voisine de la structure optimale d'un point de vue statistique, cela peut affecter le profil des distributions des coordonnées atomiques et donc, probablement, celui des profils des énergies.

Bien que les informations utiles soient les mêmes dans les deux tests : les moyennes, les écarts-types et l'ordre de grandeur des largeurs des distributions, il est essentiel de réaliser cette étude dans les meilleures conditions permettant de s'affranchir de tout biais.

Les distributions des énergies d'interaction électrostatique peuvent présenter des profils leptokurtiques (piqués) ou platikurtiques (écrasés), et des dissymétries significatives. Discuter leurs profils nécessite de minimiser les répercussions de l'état initial.

Intuitivement, il est légitime d'observer des dissymétries de distribution pour les coordonnées atomiques lorsque leurs valeurs initiales sont nettement différentes des modes principaux. Dans un cas à une dimension (une fonction monovariée), si la structure initiale utilisée pour la méthode Perturbation/Affinement est sur la droite du minimum global, une majorité de conformations vont converger par la droite, ce qui se manifestera par un étalement à droite de la distribution statistique. Les profils des distributions ne seront exploitables que si les perturbations de la structure sont réalisées en partant de la structure optimale d'un point de vue statistique.

Afin de s'en assurer, deux derniers tests préliminaires ont été réalisés en considérant, comme structure initiale, une structure dont les coordonnées atomiques au sein du site actif ont pour valeurs les modes principaux des distributions obtenues lors du second test, et en considérant une amplitude maximale de perturbation *PerturMax* de 0,1 Å pour le premier d'entre eux (test n°3), et 0,2 Å pour le deuxième (test n°4).

Ces deux tests ont été menés de façon à obtenir initialement 1500 structures affinées. Après exclusion des « outliers de restrictions » et « outliers conformationnels », un jeu de 1355 structures a pu être utilisé pour le test n°3, et 1241 structures pour le test n°4.

Dans les tests précédents, les différences en valeur absolue moyennes observées entre les valeurs initiales des coordonnées (structure issue de la Protein Databank) et les valeurs de leurs modes principaux sont dans les deux cas de 0,10 Å, et les différences maximales de 0,67 Å.

La **Figure 11** permet de constater une nette différence entre la conformation initiale obtenue à partir du fichier PDB, et plusieurs conformations dont les coordonnées appartiennent aux modes principaux des distributions des coordonnées atomiques du sous-système étudié.

Dans ces deux nouveaux tests (n°3 et n°4) utilisant une structure initiale définie à partir des modes principaux des distributions obtenues au cours du test n°2, les différences moyennes sont respectivement de 0,020 Å et de 0,023 Å seulement, ce qui est proche de l'écart-type moyen de 0,015 Å. De plus, les différences maximales observées en valeur absolue sont respectivement de 0,16 Å et de 0,165 Å.

Si l'on observe à nouveau les distributions des coordonnées Y des atomes C4N et C5N du cofacteur NADP⁺, on constate que les profils des distributions sont nettement plus symétriques et étroits (**Figures 21, 22**). Obtenir des profils significativement plus proches d'une loi normale est en accord avec l'approximation qui consiste en fin d'affinement à prêter au voisinage du minimum un profil de puit harmonique.

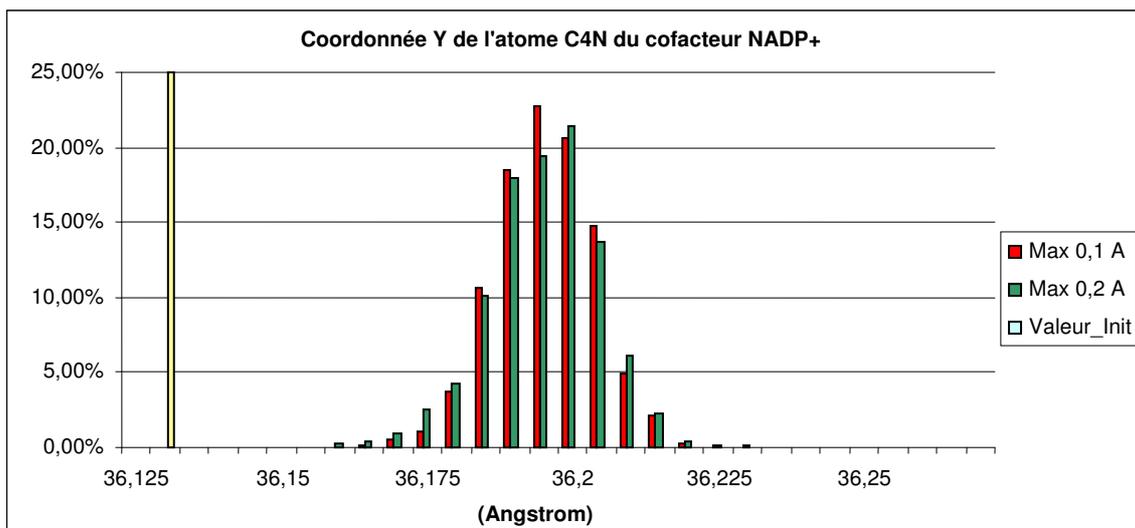


Figure 21 : Histogrammes des populations pour la coordonnée Y de l'atome C4N du cofacteur NADP⁺, obtenus aux cours des tests n°3 et n°4 (*PerturMax* = 0,1 Å et *PerturMax* = 0,2 Å avec un nouvel état initial).

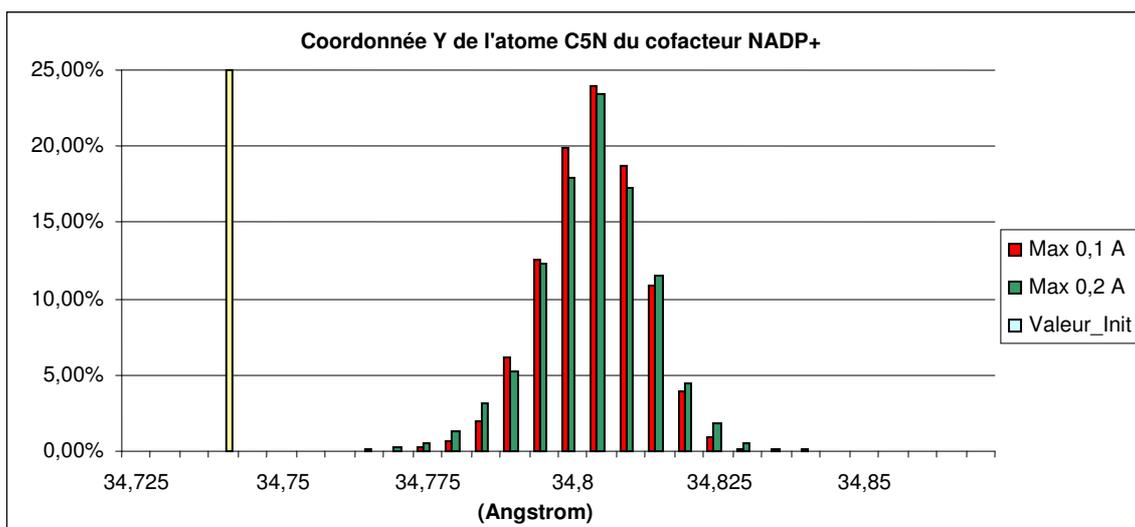


Figure 22 : Histogrammes des populations pour la coordonnée Y de l'atome C5N du cofacteur NADP⁺, obtenus aux cours des tests n°3 et n°4 (*PerturMax* = 0,1 Å et *PerturMax* = 0,2 Å avec un nouvel état initial).

L'inhibiteur sorbinil subit un changement conformationnel important au sein du site actif de l'aldose réductase (**Figure 11**). Ce changement est vérifié quantitativement par comparaison des valeurs des coordonnées de l'inhibiteur dans la conformation initiale issue du serveur RCSB, et les valeurs des modes des distributions de ces mêmes coordonnées. On obtient en

effet une différence moyenne arithmétique de 0,22 Å, et une différence maximale de 0,69 Å pour la coordonnée X de l'atome O6I.

La diminution des écarts-types statistiques des distributions des coordonnées Y des atomes C4N et C5N du cofacteur NADP⁺ observée sur les **Figures 21** et **22**, est généralisable à l'ensemble des distributions des coordonnées des atomes du sous-système considéré dans cette étude. En effet, dans les tests n°1 (structure initiale issue du serveur RCSB et *PerturMax* = 0,1 Å) et n°2 (structure initiale issue du serveur RCSB et *PerturMax* = 0,2 Å), les moyennes arithmétiques des écarts-types statistiques sont respectivement de 0,020 Å et 0,022 Å, et les écarts-types maximaux sont de 0,083 Å et 0,084 Å pour la coordonnée Y de l'atome C16 de l'inhibiteur sorbinil.

Les tests n°3 et n°4 permettent d'obtenir des valeurs moyennes et maximales d'écarts-types plus petites. Les valeurs des moyennes arithmétiques sont de 0,014 Å (test n°3) et 0,015 Å (test n°4). Les valeurs maximales d'écart-type sont obtenues à nouveau pour la coordonnée Y de l'atome C16 avec des valeurs respectives de 0,052 Å et de 0,054 Å.

Considérer de tels ordres de grandeur de longueur peut paraître inutile, mais il faut garder à l'esprit qu'une distribution gaussienne a, pour un écart-type σ , une largeur d'intervalle de valeurs de six fois σ correspondant à 99 % des événements, et que les coordonnées des atomes d'un même groupement chimique sont corrélées. Cette corrélation est pris en compte au cours d'un affinement par l'usage des restrictions structurales et confère aux groupements chimiques une certaine rigidité. L'ensemble des contributions à l'incertitude sur une énergie d'interaction électrostatique pourra être constructive. Les résultats déjà obtenus au cours des tests n°1 et n°2 nous ont montré que les largeurs des distributions des énergies pouvaient être de l'ordre de l'énergie d'une liaison hydrogène moyenne. Qu'en est-il des résultats des tests n°3 et n°4 ?

Les résultats de l'analyse statistique révèlent des différences significatives de ces distributions par rapport à celles obtenues précédemment : les moyennes peuvent changer de plusieurs kJ.mol⁻¹ et les écarts-types sont plus petits (**Tableaux 7, 8**). Par exemple, dans le cas des interactions sorbinil-Tyr48 et sorbinil-His110, les écarts-types sont divisés par deux. De même, leurs largeurs diminuent considérablement. Le résultat est à mettre en parallèle avec la diminution des écarts-types des distributions des coordonnées atomiques précédemment commentée.

Les illustrations des histogrammes des énergies d'interaction électrostatique les plus importantes montrent l'impact significatif de l'état initial (**Figure 23**). Les profils des histogrammes de populations sont quasi-symétriques comparés à ceux précédemment obtenus. Ces résultats sont révélateurs de la nécessité de partir de la conformation de la structure correspondant au minimum global avant toute étude statistique, afin de ne pas surestimer les écarts-types et largeurs des distributions des énergies des interactions électrostatiques.

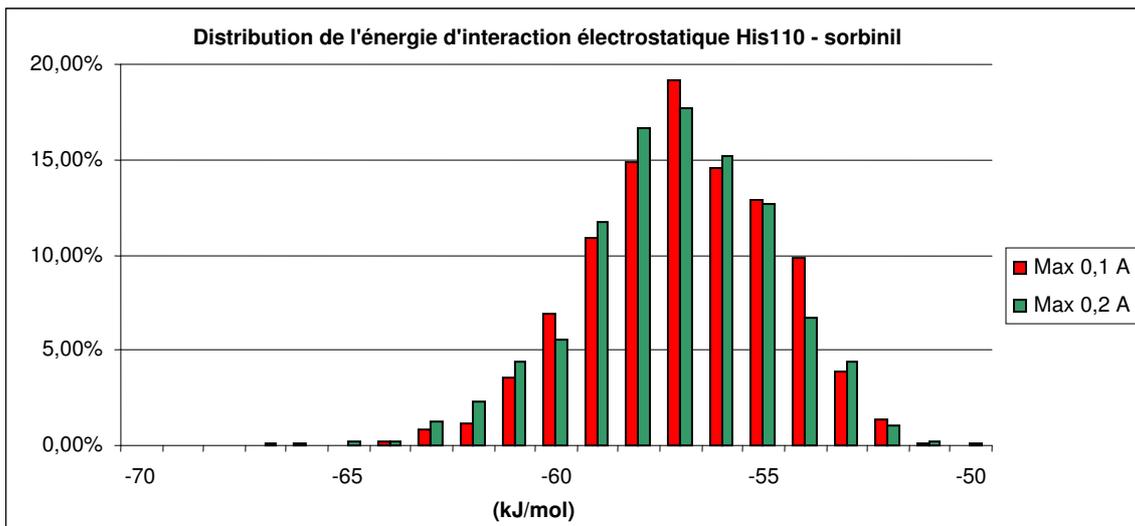
Les distributions des valeurs des coordonnées atomiques, ainsi que des énergies d'interaction électrostatique déterminées dans les tests n°3 et n°4 sont équivalentes, ce qui montre qu'une amplitude maximale de perturbation *PerturMax* de 0,1 Å est suffisante pour une telle étude.

Aldose réductase - Sorbinil				Moyenne en kJ.mol ⁻¹	Ecart-type en kJ.mol ⁻¹	Valeur minimale en kJ.mol ⁻¹	Valeur maximale en kJ.mol ⁻¹	Largeur en kJ.mol ⁻¹	Rapport de Skewness	Rapport de Kurtosis
Residu	Atomes non-H les plus proches	Distance (Å)								
His 110	NE2...N4	3,00	-57	2	-66	-51	15	-4	0	
Leu 300	CD1...C9	4,01	-1	1	---	---	---	---	---	
Trp 111	NE...O6I	2,86	-78	2	-82	-69	13	17	19	
Tyr 48	OH...O3I	2,81	-58	3	-73	-47	26	0	13	
Val47	O...F17	3,79	-2	1	---	---	---	---	---	
Cys298	SG...C8I	3,82	10	1	9	11	2	---	---	
Ala299	N...O20	---	1	1	---	---	---	---	---	
Phe122	CZ...C16	3,63	-9	1	-12	-7	5	-10	7	
Trp20	Ring...N1I	3,2	-87	1	-95	-83	12	-30	40	
Trp219	CH2...O10	3,38	2	1	0	3	3	---	---	
Trp79	CZ3...O6I	3,38	-6	1	---	---	---	---	---	
NADP ⁺	NC4...O3I	3,10	4	1	3	6	3	---	---	

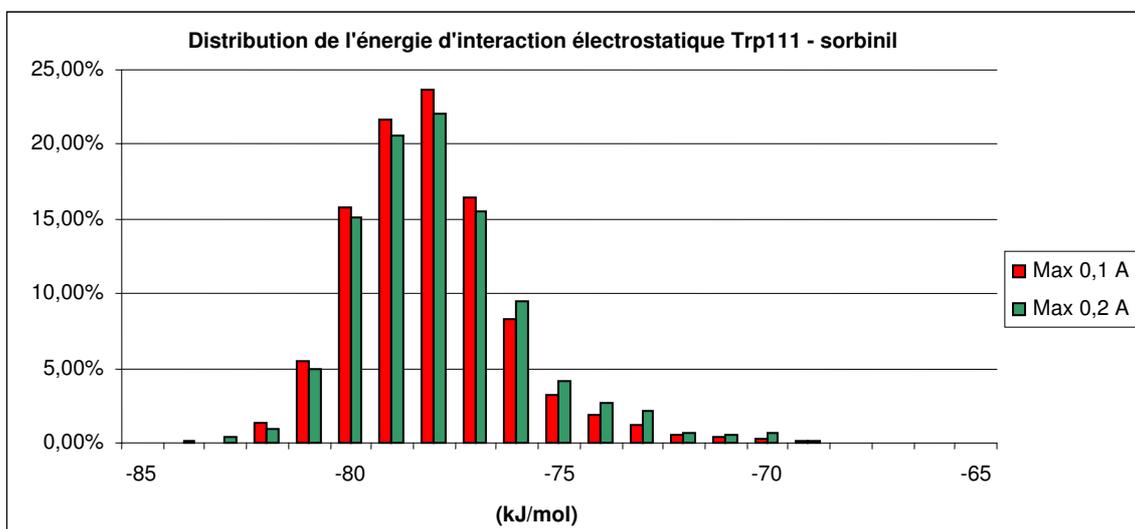
Tableau 7 : Statistiques sur les distributions des énergies électrostatiques obtenues pendant le test n°3 de l'étude du complexe aldose réductase – sorbinil. Toutes les interactions entre l'inhibiteur sorbinil et les résidus du site actif de l'aldose réductase porcine sont décrites. Les atomes non hydrogène les plus proches entre l'inhibiteur sorbinil et les résidus sont donnés avec les distances correspondantes.

Aldose réductase - Sorbinil				Moyenne en kJ.mol ⁻¹	Ecart-type en kJ.mol ⁻¹	Valeur minimale en kJ.mol ⁻¹	Valeur maximale en kJ.mol ⁻¹	Largeur en kJ.mol ⁻¹	Rapport de Skewness	Rapport de Kurtosis
Residu	Atomes non-H les plus proches	Distance (Å)								
His 110	NE2...N4	3,00	-57	2	-67	-50	17	-4	2	
Leu 300	CD1...C9	4,01	-1	1	---	---	---	---	---	
Trp 111	NE...O6I	2,86	-78	2	-84	-69	15	10	10	
Tyr 48	OH...O3I	2,81	-58	3	-71	-47	24	1	7	
Val47	O...F17	3,79	-2	1	---	---	---	---	---	
Cys298	SG...C8I	3,82	10	1	9	11	2	---	---	
Ala299	N...O20	---	1	1	---	---	---	---	---	
Phe122	CZ...C16	3,63	-9	1	-13	-7	6	-10	5	
Trp20	Ring...N1I	3,2	-87	1	-92	-83	9	-6	6	
Trp219	CH2...O10	3,38	2	1	-1	3	4	---	---	
Trp79	CZ3...O6I	3,38	-6	1	-6	-5	1	---	---	
NADP ⁺	NC4...O3I	3,10	4	1	3	6	3	---	---	

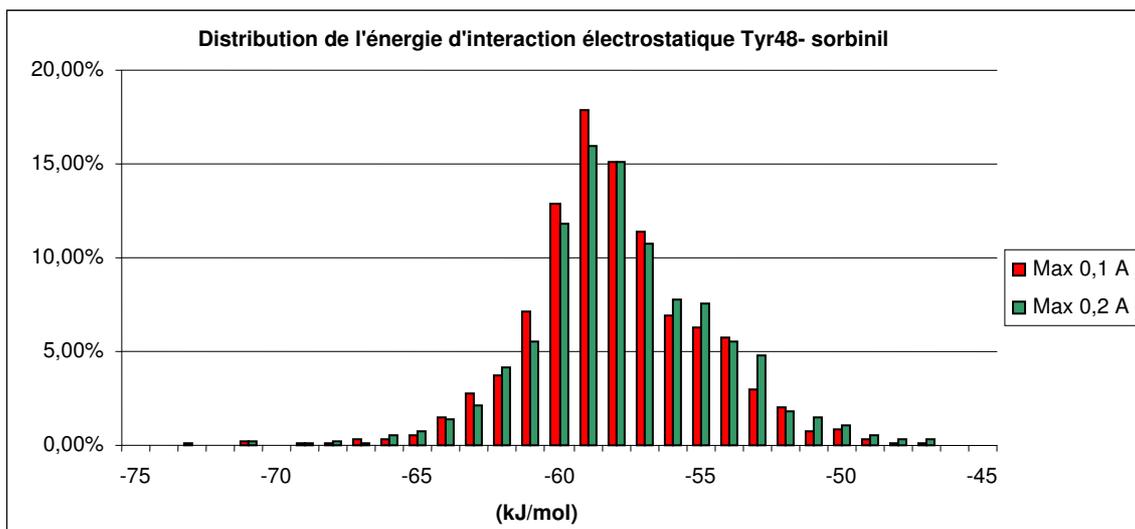
Tableau 8 : Statistiques sur les distributions des énergies électrostatiques obtenues pendant le test n°4 de l'étude du complexe aldose réductase – sorbinil. Toutes les interactions entre l'inhibiteur sorbinil et les résidus du site actif de l'aldose réductase porcine sont décrites. Les atomes non hydrogène les plus proches entre l'inhibiteur sorbinil et les résidus sont donnés avec les distances correspondantes.



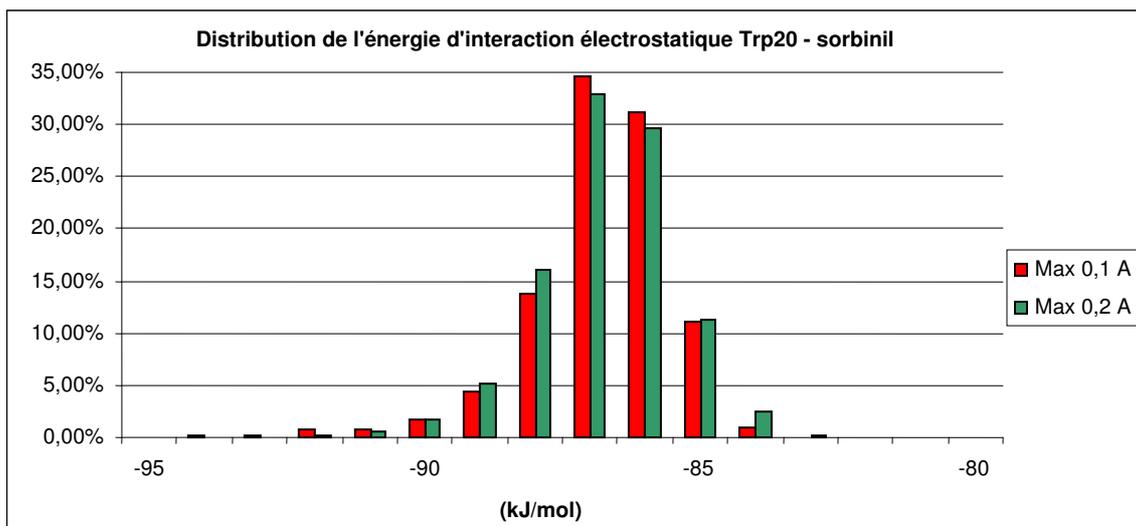
(a)



(b)



(c)



(d)

Figure 23 : Histogrammes des populations des énergies d'interaction électrostatique significatives au sein du site actif du complexe aldose réductase porcine – sorbinil (en $\text{kJ}\cdot\text{mol}^{-1}$) obtenus au cours des tests n°3 et n°4.

sorbiniil		Distance (Å)		Energie (kJ.mol ⁻¹)	
Residu	Atomes non-H les plus proches	Conformation initiale	Modes des distributions	Conformation initiale	Modes des distributions
His 110	NE2...N4	3,00	2,74	-11	-57
Leu 300	CD1...C9	4,01	4,74	0	-1
Trp 111	NE...O6I	2,86	2,52	-43	-78
Tyr 48	OH...O3I	2,81	2,29	7	-59
Val47	O...F17	3,79	3,93	-2	-2
Cys298	SG...C8I	3,82	3,46	-7	10
Ala299	N...O20	---	---	0	1
Phe122	CZ...C16	3,63	3,36	-7	-9
Trp20	Ring...N1I	3,2	3,0	-59	-87
Trp219	CH2...O10	3,38	3,69	1	2
Trp79	CZ3...O6I	3,38	3,84	-10	-6
NADP ⁺	NC4...O3I	3,10	3,30	3	4

Tableau 9 : Comparaison des valeurs des énergies d'interaction électrostatique entre l'inhibiteur sorbinil et les résidus du site actif de l'aldose réductase porcine dans la conformation initiale (PDB) et dans la conformation correspondant aux modes principaux des distributions obtenues dans le test n°3. Les distances interatomiques minimales entre atomes non hydrogène sont également communiquées.

La différence conformationnelle observée sur la **Figure 11** entre la conformation initiale obtenue à partir du fichier PDB, et les conformations correspondant aux modes principaux des distributions, se répercute sur les propriétés électrostatiques du complexe, notamment les énergies d'interaction électrostatique entre l'inhibiteur sorbinil et les principaux résidus du site actif (**Tableau 9**).

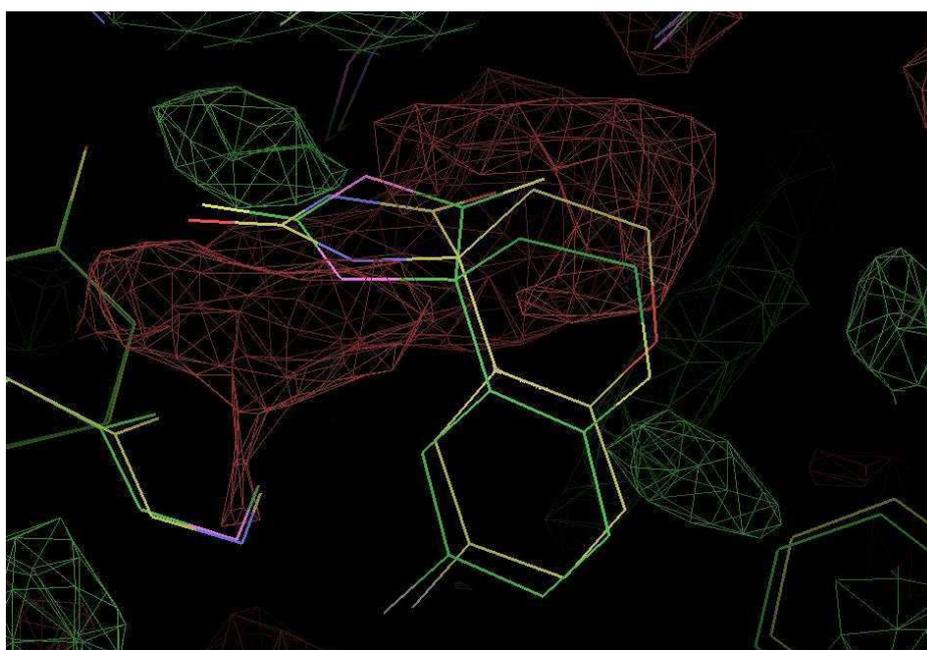
IV.3.7. Discussion des résultats obtenus

Avant même de discuter des distributions d'énergies d'interaction électrostatique, on peut constater en premier lieu que la structure du complexe aldose réductase porcine – sorbinil est nettement éloignée de la conformation optimale identifiée au cours de cette étude. L'examen des cartes de densité résiduelle dans le site actif du complexe enzyme – inhibiteur dans la conformation issue du serveur de structures RCSB et dans l'une des conformations correspondant aux modes principaux des distributions des coordonnées atomiques révèlent

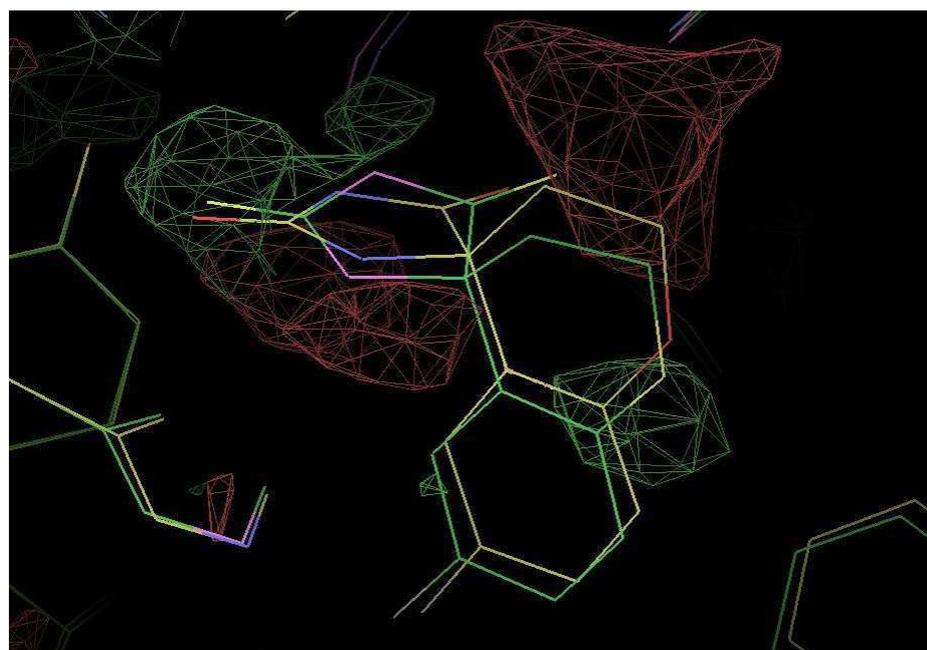
une diminution des pics résiduels dans la conformation optimale obtenue au cours de ce test préliminaire (**Figure 24**). Cette conformation a été choisie au hasard parmi les nombreuses possibles. Ce premier résultat nous révèle que les fichiers de structures issus de serveurs de banques de données, même affinées contre des données *a priori* de qualité et de résolution satisfaisantes, doivent être employés avec précaution.

Le changement conformationnel de l'inhibiteur a modifié considérablement ses interactions avec les résidus du site actif ce qui a des répercussions flagrantes sur les énergies d'interaction électrostatique avec des variations allant jusqu'à 60 kJ.mol^{-1} (interaction entre le sorbinil et le résidu Tyr48) (**Tableau 9**). On constate notamment que le changement d'orientation du groupement de « tête » (groupement hydantoin) qui se fixe dans la « poche de fixation de l'anion », s'accompagne d'une diminution des distances entre les atomes impliqués dans ses liaisons hydrogène avec les résidus Tyr48, Trp111 et His110 ce qui améliore l'interaction d'un point de vue électrostatique (**Tableau 9**).

Réaliser un criblage au voisinage du minimum global a permis d'obtenir des distributions d'écart-type plus petit pour les coordonnées atomiques, ce qui a affecté également les distributions des énergies d'interaction électrostatique.



(a)



(b)

Figure 24 : Illustrations des cartes de densité électronique résiduelle en trois dimensions. La conformation initiale issue du serveur PDB est représentée en jaune, et l'une des conformations optimales en vert. La carte de densité résiduelle (a) correspond à la conformation initiale, la carte (b) à la conformation optimale. Les isomaillages sont définis à $0,10 \text{ e.}\text{\AA}^{-3}$.

L'interprétation des distributions d'énergies d'interaction électrostatique n'est pas chose aisée. Cependant, l'on peut constater que, bien qu'au premier abord, la valeur de l'écart-type ou la largeur de la distribution puissent être corrélées à la valeur absolue de la moyenne correspondante, il existe également une relation entre la distance interatomique minimale entre les groupements chimiques en interaction et la valeur de l'écart-type de la distribution des énergies correspondante. Malgré le fait que l'interaction entre le sorbinil et le résidu Trp20 soit en valeur absolue la plus intense avec -87 kJ.mol^{-1} , l'écart-type correspondant est de l'ordre du kJ.mol^{-1} . La distance interatomique entre atomes non hydrogène dans le cas de cette liaison hydrogène H... π est de $3,2 \text{ \AA}$.

Les dépendances de l'incertitude sur l'énergie à la distance interatomique minimale entre les atomes des groupements chimiques en interaction, et à l'amplitude de cette énergie peuvent s'expliquer mathématiquement en se ramenant à un cas simple d'interaction électrostatique entre deux charges ponctuelles q_A et q_B , localisées respectivement en \mathbf{r}_A et \mathbf{r}_B .

L'expression de leur énergie d'interaction électrostatique est, selon la loi de Coulomb :

$$E_{elec,AB} = \frac{q_A q_B}{4\pi\epsilon_0 d_{AB}} \quad (4.15) \text{ avec } d_{AB} = |\mathbf{r}_A - \mathbf{r}_B| = \sqrt{(\mathbf{r}_A - \mathbf{r}_B) \cdot (\mathbf{r}_A - \mathbf{r}_B)}$$

Considérons $\mathbf{r}_{A,0}$ et $\mathbf{r}_{B,0}$ les positions de référence (ou à l'équilibre) des charges ponctuelles q_A et q_B . Si l'on introduit des perturbations $d\mathbf{r}_A$ et $d\mathbf{r}_B$ sur les vecteurs positions \mathbf{r}_A et \mathbf{r}_B .

Les vecteurs positions deviennent alors : $\mathbf{r}_A = \mathbf{r}_{A,0} + d\mathbf{r}_A$ et $\mathbf{r}_B = \mathbf{r}_{B,0} + d\mathbf{r}_B$

Et la distance interatomique peut s'écrire :

$$d_{AB} = |(\mathbf{r}_{A,0} - \mathbf{r}_{B,0}) + (d\mathbf{r}_A - d\mathbf{r}_B)| \quad (4.16)$$

ou encore, $d_{AB}^2 = ((\mathbf{r}_{A,0} - \mathbf{r}_{B,0}) + (d\mathbf{r}_A - d\mathbf{r}_B)) \cdot ((\mathbf{r}_{A,0} - \mathbf{r}_{B,0}) + (d\mathbf{r}_A - d\mathbf{r}_B))$

Le développement de l'expression de d_{AB}^2 donne :

$$d_{AB}^2 = (\mathbf{r}_{A,0} - \mathbf{r}_{B,0}) \cdot (\mathbf{r}_{A,0} - \mathbf{r}_{B,0}) + 2(\mathbf{r}_{A,0} - \mathbf{r}_{B,0}) \cdot (d\mathbf{r}_A - d\mathbf{r}_B) + (d\mathbf{r}_A - d\mathbf{r}_B) \cdot (d\mathbf{r}_A - d\mathbf{r}_B) \quad (4.17)$$

Si l'on suppose : $|\mathbf{r}_{A,0} - \mathbf{r}_{B,0}| \gg |d\mathbf{r}_A - d\mathbf{r}_B|$, on obtient à l'ordre 1 :

$$d_{AB}^2 = d_{AB,0}^2 + 2(\mathbf{r}_{A,0} - \mathbf{r}_{B,0})(d\mathbf{r}_A - d\mathbf{r}_B) \text{ avec } d_{AB,0} = |\mathbf{r}_{A,0} - \mathbf{r}_{B,0}| \quad (4.18)$$

L'expression de l'inverse de d_{AB} devient :

$$\frac{1}{d_{AB}} = \frac{1}{d_{AB,0}} \left(1 - \frac{(\mathbf{r}_{A,0} - \mathbf{r}_{B,0})(d\mathbf{r}_A - d\mathbf{r}_B)}{d_{AB,0}^2} \right) \quad (4.19)$$

avec toujours comme hypothèse : $d_{A,0} \gg |d\mathbf{r}_A - d\mathbf{r}_B|$

L'expression de l'énergie correspondante peut s'écrire alors :

$$E_{elec,AB} = \frac{q_A q_B}{4\pi\epsilon_0} \frac{1}{d_{AB}} = \frac{q_A q_B}{4\pi\epsilon_0} \frac{1}{d_{AB,0}} \left(1 - \frac{(\mathbf{r}_{A,0} - \mathbf{r}_{B,0})(d\mathbf{r}_A - d\mathbf{r}_B)}{d_{AB,0}^2} \right) \quad (4.20)$$

$$E_{elec,AB} = E_{elec,AB,0} \left(1 - \frac{(\mathbf{r}_{A,0} - \mathbf{r}_{B,0})(d\mathbf{r}_A - d\mathbf{r}_B)}{d_{AB,0}^2} \right) \quad (4.21)$$

$$\text{avec } E_{elec,AB,0} = \frac{q_A q_B}{4\pi\epsilon_0} \frac{1}{d_{AB,0}}$$

On a donc comme expression de la variation d'énergie en valeur absolue en réponse à la perturbation des vecteurs positions :

$$\Delta E_{elec,AB} = |E_{elec,AB} - E_{elec,AB,0}| = |E_{elec,AB,0}| \left| \frac{(\mathbf{r}_{A,0} - \mathbf{r}_{B,0})(d\mathbf{r}_A - d\mathbf{r}_B)}{d_{AB,0}^2} \right| \leq |E_{elec,AB,0}| \frac{|d\mathbf{r}_A - d\mathbf{r}_B|}{d_{AB,0}} \quad (4.22)$$

La variation d'énergie $\Delta E_{elec,AB}$ est proportionnelle à la valeur de l'énergie d'interaction électrostatique des deux charges à l'équilibre. Elle est également inversement proportionnelle à la distance entre ces charges.

On notera également que la différence d'énergie $dE_{elec,AB} = E_{elec,AB} - E_{elec,AB,0}$ n'est pas une combinaison linéaire des perturbations $d\mathbf{r}_A$ et $d\mathbf{r}_B$ mais dépend de leurs produits scalaires avec le vecteur $(\mathbf{r}_{A,0} - \mathbf{r}_{B,0})$.

Cette formule peut également être obtenue aisément en considérant l'incertitude sur l'énergie d'interaction électrostatique entre ces deux charges, comme la différentielle d'ordre 1 de l'énergie en fonction de la distance.

Cette formule n'est valable qu'à l'ordre 1, c'est-à-dire si l'on a $d_{A,0} \gg |d\mathbf{r}_A - d\mathbf{r}_B|$. Compte tenu que dans cette étude, les énergies d'interaction sont calculées à partir de modèles de distribution de charges asphérique et continue, lors d'interactions à courte distance telles que les trois liaisons hydrogène impliquant les résidus His110, Trp111 et Tyr48, il y a recouvrement des distributions de charges des groupements chimiques en interaction. Par conséquent, la condition $d_{A,0} \gg |d\mathbf{r}_A - d\mathbf{r}_B|$ ne peut être respectée. De plus, en raison de la nature des modèles utilisés, les corrélations entre leurs termes sont plus complexes et nombreuses ce qui complique le problème (comme cela a déjà été abordé dans la partie IV.1).

Une stratégie satisfaisante à appliquer pour tout système protéine-ligand, serait :

- De réaliser une statistique sur les coordonnées atomiques du site actif à partir de la structure disponible sur le serveur RCSB en utilisant une amplitude maximale de perturbation suffisamment élevée pour obtenir les modes principaux des distributions des coordonnées ($PerturMax = 0,2 \text{ \AA}$). 500 affinements seraient a priori suffisants, le but n'étant pas de déterminer des profils précis mais seulement les valeurs des modes des distributions.
- Puis, dans un second temps, si les valeurs des modes (ou des moyennes) se révèlent être très différentes des valeurs initiales des coordonnées atomiques, de réaliser une nouvelle statistique en considérant comme état initial une structure ayant pour coordonnées atomiques les modes précédemment obtenus. Une perturbation maximale de $0,1 \text{ \AA}$ sera alors suffisante avec un nombre d'affinements plus important de l'ordre de 1500.
- Il est intéressant de comparer les valeurs des coordonnées atomiques de la conformation initiale aux valeurs des modes principaux ou des moyennes des distributions statistiques correspondantes afin de s'assurer que ces valeurs restent proches et ainsi éviter de biaiser les profils des distributions des énergies comme cela a pu être noté dans les tests n°1 et n°2.

IV.4. Applications à d'autres complexes macromoléculaires

IV.4.1. Choix des complexes

Dans la partie IV.1 de ce chapitre, les motivations et objectifs de cette étude ont été définis et argumentés. La priorité est ici d'estimer par une approche statistique les incertitudes sur les énergies d'interaction électrostatique afin d'être en mesure de discuter la nécessité de la définition d'un modèle de charges continue asphérique par transfert depuis une banque de paramètres multipolaires de distribution de charges. La résolution à laquelle les structures ont été obtenues a été retenue comme le facteur expérimental le plus évident pouvant affecter les incertitudes sur les coordonnées atomiques et, par conséquent, les incertitudes sur les énergies.

Pour cette raison, les systèmes retenus afin d'éprouver la méthode définie au cours du test préliminaire ont été choisis à des résolutions nettement différentes.

Les trois systèmes macromoléculaires sont tous des complexes enzyme-inhibiteur :

- Le complexe aldose réductase – fidarestat, à une résolution de 0,92 Å
- Le complexe aldéhyde réductase – fidarestat, à une résolution de 1,85 Å
- Le complexe aldose réductase – stéréoisomère (2*R*,4*S*) du fidarestat, à une résolution de 1,40 Å

Ces systèmes ont été choisis pour deux raisons :

- Premièrement, comme expliqué dans la conclusion du chapitre III, le modèle multipolaire de distribution de charges du fidarestat sous forme cristalline a été utilisé au cours du développement de la nouvelle banque ELMAM généralisée, les types atomiques nécessaires à la modélisation des distributions de charges des inhibiteurs de ces complexes sont ainsi disponibles.
- Deuxièmement, ne considérer que des complexes aldose réductase (ou aldéhyde réductase) permet de comparer les résultats obtenus d'un point de vue physico-chimique.

IV.4.2. Cas de la structure de l'aldose réductase complexée avec l'inhibiteur fidarestat

IV.4.2.A. Présentation de la structure

La structure du complexe aldose réductase – fidarestat avec comme cofacteur le NADP⁺ a été affinée à une résolution de 0,92 Å par El-Kabbani *et al.* [13]. La maille a pour groupe d'espace P2₁. L'inhibiteur fidarestat présente une très grande similitude avec l'inhibiteur sorbinil. Il ne se distingue de ce dernier que par la présence du groupement carbamoyl. Le fidarestat est impliqué au sein du site actif de l'aldose réductase dans les mêmes contacts intermoléculaires observés précédemment dans le complexe aldose réductase - sorbinil (**Figure 25**).

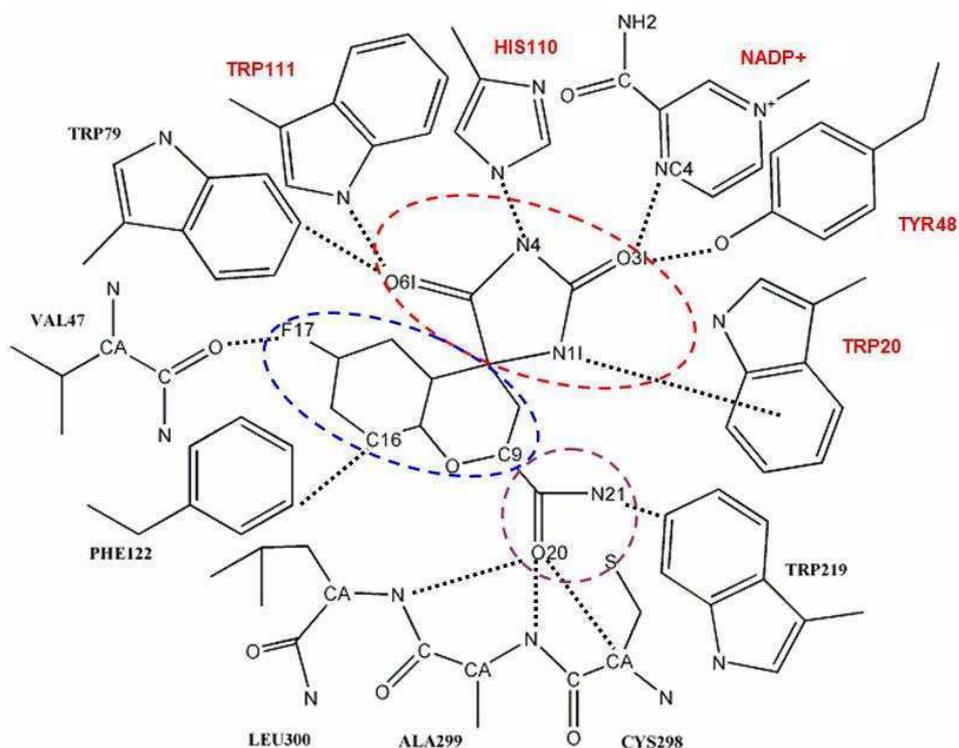


Figure 25 : Diagramme des contacts intermoléculaires impliquant le fidarestat dans le site actif de l'aldose réductase humaine. Les interactions entre les atomes lourds les plus proches sont indiquées par des lignes en pointillés. Les différents groupements chimiques du fidarestat sont précisés : le groupement hydantoin en rouge, le cycle chroman en bleu et le groupement carbamoyl en violet. Les résidus de la « poche de fixation anionique » sont précisés en rouge.

Par ailleurs, on retrouve les trois liaisons hydrogène au sein de la « poche de fixation de l'anion » entre le groupement de « tête » hydantoin et les trois résidus His110, Trp111 et Tyr48, mais également la liaison hydrogène avec comme accepteur le cycle de la chaîne latérale du Trp20. La présence du groupement carbamoyl dans le fidarestat lui permet de former une liaison hydrogène supplémentaire avec le groupement N—H de la chaîne principale du résidu Leu300 (**Figure 25**). C'est cette quatrième liaison hydrogène qui prête au fidarestat une meilleure efficacité dans l'inhibition de l'enzyme aldose réductase [15].

Par conséquent, le sous-système étudié sera similaire à celui retenu lors de l'étude du complexe aldose réductase – sorbinil. On s'intéressera donc aux coordonnées atomiques des atomes de l'inhibiteur fidarestat, des résidus Trp20, Val47, Tyr48, Trp79, His110, Trp111, Phe122, Trp219, Cys298, Ala299, Leu300, ainsi que le groupement nicotinamide du cofacteur NADP⁺ (**Figure 2**).

Le cas du complexe aldose réductase – fidarestat est cependant plus délicat. Contrairement au complexe précédent, certains résidus du site actif de la protéine sont désordonnés et présentent deux conformations : les résidus Val47, Ala299 et Leu300 le sont complètement alors que seule la chaîne latérale du résidu Cys298 l'est. Les atomes de ces conformations doubles sont traités comme les autres atomes, et leurs occupations sont maintenues à leurs valeurs initiales. Le sous-système retenu compte donc 164 atomes non hydrogène avec des occupations totales ou partielles.

Pour réaliser cette étude, 1600 structures ont été obtenues suivant la procédure (Perturbation/Affinement) définie précédemment en utilisant une amplitude maximale de perturbation *PerturMax* de 0,1 Å, cette amplitude maximale s'étant révélée suffisante pour l'étude de la structure du complexe aldose réductase – sorbinil obtenue à une résolution supérieure de 2,30 Å. Aucun pré-test Perturbation/Affinement n'a été réalisé pour identifier les modes principaux compte tenu les distributions étroites obtenues pour ces coordonnées ainsi que pour les énergies (comme cela va être expliqué dans cette partie du chapitre).

IV.4.2.B. Etudes statistiques des indicateurs de qualité

L'analyse statistique des indicateurs de qualité fournis par le programme SHELXL [35] a été réalisée (**Tableau 10**).

Indicateurs	$R_{1_free_sigc4}$	$R_{1_free_all}$	R_{1_sigc4}	R_{1_all}	R_w	gof	gof_rest
Moyenne m	0,1801	0,1880	0,1647	0,1718	0,4398	3,728	3,650
Ecart-type σ^*	0,0002	0,0002	0,0001	0,0001	0,0002	0,002	0,002
Mode ω	0,1801	0,188	0,1647	0,1718	0,4398	3,728	3,65
Médiane M	0,1801	0,188	0,1647	0,1718	0,4398	3,728	3,65
Minimum	0,1796	0,1875	0,1644	0,1716	0,4393	3,723	3,646
Maximum	0,1808	0,1887	0,1649	0,1721	0,4404	3,734	3,656
Coefficient							
Skewness γ_1^*	0,475	0,321	0,168	0,072	0,442	0,552	0,604
Coefficient							
Kurtosis β_2^*	0,313	0,281	0,013	0,079	0,280	0,300	0,343
Rapport							
skewnew	7,8	5,2	2,7	1,2	7,2	9,0	9,9
Rapport							
Kurtosis	2,6	2,3	0,1	0,6	2,3	2,4	2,8

Tableau 10 : Statistiques sur les distributions des valeurs des indicateurs de qualité au cours de l'étude du complexe aldose réductase – fidarestat. Pour 1600 événements, les erreurs standard sur les coefficients de Pearson de Skewness et de Kurtosis sont respectivement de 0,061 et 0,122.

Les indicateurs de qualité présentent des distributions plus étroites que dans le cas du complexe aldose réductase - sorbinil. En effet, les écarts-types diffèrent d'un ordre de grandeur. Par exemple, l'écart-type de l'indicateur $R_{1_free_sigc4}$ était de 0,0014 dans le cas du complexe aldose réductase – sorbinil (**Tableau 1**) alors qu'il n'est plus que de 0,0002 dans le cas du complexe aldose réductase – fidarestat.

De plus, les valeurs des deux indicateurs goodness-of-fit (gof , gof_rest) sont proches et sont égales en moyenne à respectivement 3,728 et 3,650, alors qu'elles différaient nettement dans le cas du complexe aldose réductase – sorbinil avec des moyennes respectives de 21,842 et

5,738. Des valeurs de goodness-of-fit plus proches de l'unité montrent que le modèle reproduit mieux les données expérimentales. De plus, des valeurs des goodness-of-fit sans restrictions gof et avec restrictions gof_{rest} , proches peuvent s'interpréter comme un rapport du nombre d'observations sur le nombre de variables du système suffisamment élevé pour permettre une bonne reproductibilité des données expérimentales sans prise en compte des restrictions. Cela s'explique naturellement par un nombre de réflexions plus élevé et une meilleure qualité du jeu de données compte tenu de la résolution atteinte. La largeur de distribution statistique des valeurs du gof_{rest} est très petite avec seulement 0,011. Aucun « outlier de restriction » n'est identifié.

IV.4.2.C. Etudes statistiques des coordonnées atomiques

Les algorithmes définis au cours du test préliminaire afin de supprimer, d'une part, les modes secondaires distincts des modes principales des distributions des coordonnées atomiques, et d'autre part, les étalements de distribution extrêmes, ont été utilisés sur les coordonnées des 164 atomes du site actif.

Sur les 1600 conformations obtenues pour le système aldose réductase - fidarestat, 191 ont été automatiquement identifiées et exclues comme « outliers de conformation », ce qui réduit le jeu de conformations à 1409.

Les résultats de l'analyse statistique sur les coordonnées atomiques montrent des écarts-types très faibles. Par ailleurs, la majorité des coordonnées atomiques présentent des valeurs d'écart-type de l'ordre de 10^{-4} Å ce qui est inférieur à la précision numérique sur ces dernières (10^{-3} Å). On observe malgré tout des valeurs d'écart-type supérieures pour les parties désordonnées du site actif de l'enzyme (les résidus Val47, Ala299, Leu300, et la chaîne latérale du résidu Cys298).

Pour les coordonnées des atomes des parties ordonnées du site actif, les écarts-types maximaux observés sont de 0,003 Å pour les coordonnées Y et Z de l'atome O de la chaîne principale du résidu Trp219, ce qui demeure extrêmement faible. Définir l'écart-type moyen n'aurait ici aucun sens compte tenu de la précision numérique sur les coordonnées atomiques. Pour les coordonnées des atomes des parties désordonnées, l'écart-type maximal est de 0,34 Å pour la coordonnée Y de l'atome CB conformation B du résidu Ala299, suivi par la

coordonnée Y de l'atome CB conformation A du résidu Ala299 avec 0,15 Å. La valeur moyenne des écarts-types pour les atomes des parties désordonnées est de 0,018 Å.

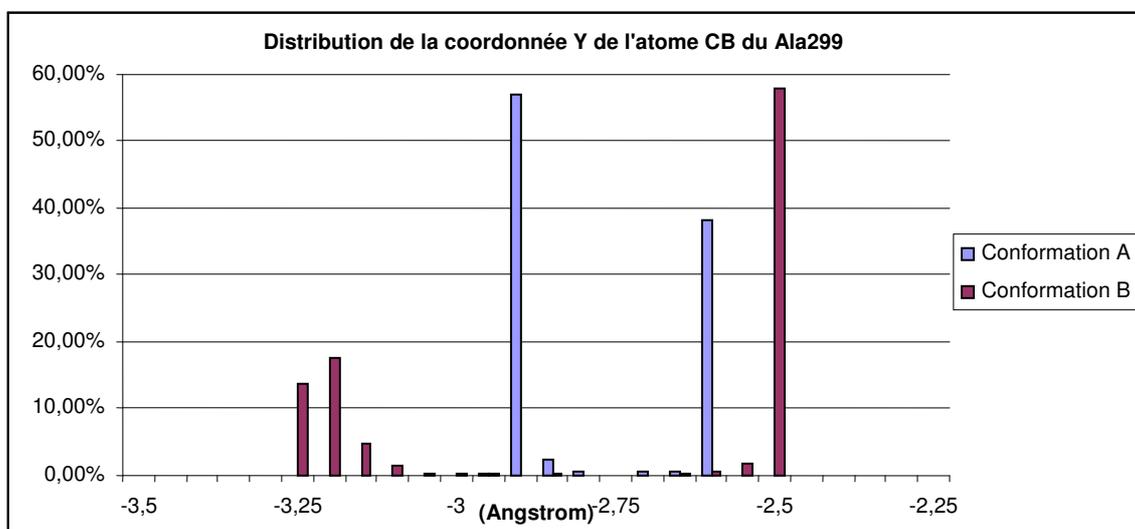


Figure 26 : Histogramme des valeurs de la coordonnée Y de l'atome CB du résidu Ala299 dans ses deux conformations.

L'histogramme des populations de la coordonnée Y de l'atome CB de Ala299 (**Figure 26**) dans ces deux conformations présente, pour chaque conformation, deux modes distincts. L'examen des structures permet également de constater que ces modes sont corrélés : les modes principaux des conformations A et B dont les amplitudes sont d'environ 55 %, situés respectivement à -2,90 Å et à -2,50 Å correspondent aux mêmes structures. Au mode secondaire de la distribution des valeurs de la coordonnée Y dans la conformation A d'amplitude de l'ordre de 40 % au voisinage de -2,60 Å, correspond au second mode plus large de la conformation B localisé à -3,20 Å.

Une telle distribution met en évidence la difficulté représentée par les parties désordonnées dans une telle étude.

Dans la stratégie définie en conclusion du test préliminaire, il a été suggéré de vérifier que les différences de valeurs des coordonnées atomiques dans la structure initiale utilisée pour l'étude ne diffèrent pas des moyennes des distributions obtenues.

Pour cette comparaison, les atomes des parties ordonnées et désordonnées ont été examinés séparément. Pour les parties ordonnées, la différence moyenne en valeur absolue entre les valeurs des coordonnées initiales et leurs moyennes est de 0,009 Å, pour une différence

maximale de 0,041 Å pour la coordonnée X de l'atome CH2 du résidu Trp219. La différence moyenne est trois fois supérieure à l'écart-type maximal observé.

Dans les parties désordonnées, la différence moyenne absolue est de 0,058 Å, et la différence maximale de 0,28 Å pour la coordonnée Z de l'atome O conformation B du résidu Leu300. La différence moyenne est d'un ordre de grandeur supérieur à l'écart-type moyen calculé de 0,018 Å.

Pour l'ensemble du sous-système étudié, Les différences absolues entre les valeurs initiales des coordonnées et leurs moyennes statistiques sont nettement plus élevées que les écarts-types correspondants. Malgré cela, l'algorithme fait converger le modèle vers un minimum très bien défini aux vues des valeurs des écarts-types. De plus, les différences absolues demeurent faibles avec au maximum 0,28 Å. Ces différences peuvent simplement s'expliquer par le fait que la structure initiale issue de la PDB a été obtenue par affinement à l'aide de programmes et stratégies différents.

IV.4.2.D. Détermination des distributions des énergies d'interaction électrostatique.

Comme dans le test préliminaire, pour chaque structure du jeu de données, le site actif de la protéine est protoné à l'aide du programme SHELXL, puis un modèle de distribution de charges est défini par transfert de paramètres multipolaires depuis la banque ELMAM [39].

L'état de protonation de l'inhibiteur fidarestat au sein du site actif est imposé. Ce dernier est considéré sous sa forme neutre, et par là même, le résidu His110 est considéré neutre de type HID (protoné une fois sur l'azote ND2). L'inhibiteur fidarestat est impliqué dans 12 interactions dans le site actif de la protéine, interactions équivalentes à celles du sorbinil.

A partir des énergies d'interaction électrostatique calculées à l'aide du programme VMOPro, une analyse statistique a été réalisée sur les différentes distributions des énergies. Les résultats sont communiqués dans le **Tableau 11**.

L'ensemble des énergies d'interaction électrostatique présente des valeurs d'écart-types inférieur à 1 kJ.mol^{-1} , ainsi que des largeurs de distribution au maximum de 2 kJ.mol^{-1} . L'erreur numérique sur les énergies d'interaction étant estimée à 1 kJ.mol^{-1} , l'incertitude expérimentale sur les énergies d'interaction électrostatique est donc du même ordre. Les énergies dérivées de la conformation initiale issue du serveur RCSB sont les mêmes au kJ.mol^{-1} près. Malgré le désordre des résidus Val47, Ala299, Leu300 et Cys298, les énergies correspondantes ont des incertitudes expérimentales faibles. Cela peut s'expliquer simplement par l'amplitude moyenne ou même faible de ces énergies.

Disposer d'un modèle structural affiné à une résolution inférieure à $1,0 \text{ \AA}$ semble permettre d'obtenir des énergies d'interaction électrostatique sans incertitudes expérimentales significatives.

On notera cependant que les seules énergies présentant des largeurs de distribution de 2 kJ.mol^{-1} sont celles des interactions impliquant des parties désordonnées (Val47, Ala299 et Leu300), ce qui suggère que le désordre et l'agitation thermique sont des facteurs à prendre en compte dans la sélection des structures cristallographiques.

Aldose réductase - Fidarestat				Moyenne en kJ.mol ⁻¹	Ecart-type en kJ.mol ⁻¹	Valeur minimale en kJ.mol ⁻¹	Valeur maximale en kJ.mol ⁻¹	Largeur en kJ.mol ⁻¹	Rapport de Skewness	Rapport de Kurtosis
Residu	Atomes non-H les plus proches	Distance (Å)								
His 110	NE2...N4	2.75	-55	1	-56	-55	1	---	---	
Leu 300	N...O20	2.95, 3.03	-24	1	-26	-24	2	---	---	
Trp 111	NE...O6I	2.79	-51	1	-51	-50	1	---	---	
Tyr 48	OH...O3I	2.63	-28	1	---	---	---	---	---	
Val47	O...F17	3.21, 3.09	-5	1	-6	-4	2	---	---	
Cys298	CA...O20	3.44	-14	1	-14	-13	1	---	---	
Ala299	N...O20	3.46, 3.55	3	1	2	4	2	---	---	
Phe122	CE1...C16	3.83	-3	1	---	---	---	---	---	
Trp20	Ring...N1I	3.16	-42	1	---	---	---	---	---	
Trp219	CH2...N21	3.62	-2	1	-2	-2	1	---	---	
Trp79	CZ3...O6I	3.62	-7	1	-7	-6	1	---	---	
NADP ⁺	NC4...O3I	3.04	10	1	---	---	---	---	---	

Tableau 11 : Statistiques sur les distributions des énergies électrostatiques obtenues lors de l'étude du complexe aldose réductase – fidarestat. Toutes les interactions entre l'inhibiteur fidarestat et les résidus du site actif de l'aldose réductase humaine sont décrites. Les atomes non hydrogène les plus proches entre l'inhibiteur fidarestat et les résidus sont donnés avec les distances correspondantes. En cas de désordre, deux distances interatomiques peuvent être communiquées.

IV.4.3. Cas de la structure de l'aldéhyde réductase complexée avec l'inhibiteur Fidarestat

IV.4.3.A. Présentation de la structure

La structure du complexe aldéhyde réductase porcine – fidarestat avec comme cofacteur le NADP⁺, a été affinée à une résolution de 1,85 Å par El-Kabbani *et al.* [10] et est disponible sur le serveur RCSB. La maille a pour groupe d'espace P6₅22.

L'une des difficultés majeures dans la conception et la synthèse d'un médicament efficace pour traiter les complications liées à l'hyperglycémie chronique par inhibition de l'enzyme aldose réductase, est la grande similitude de séquence (environ 65 %) mais également homologie structurale de cette dernière avec une autre enzyme appartenant à la super-famille des aldo-keto réductase : l'aldéhyde réductase. Cela leur confère des fonctions voisines : ces deux aldo-keto réductases catalysent la réduction des aldéhydes, aldéhydes xénobiotiques, des cétones, trioses et trioses phosphates [33;32;8].

La poche de fixation du groupement hydantoin du fidarestat est préservée dans les deux protéines. Le groupement hydantoin est impliqué dans trois liaisons hydrogène avec les résidus Tyr48, His110 et Trp111 dans le complexe aldose réductase, auxquels correspondent les résidus Tyr50, His113 et Trp114 dans le complexe aldéhyde réductase (**Figure 27**) [36;12;14]. La liaison hydrogène H... π , dans le site actif de l'aldose réductase, entre l'hydrogène H1I-N1I de l'hydantoin et le cycle du tryptophane Trp20 est également présente (Trp22).

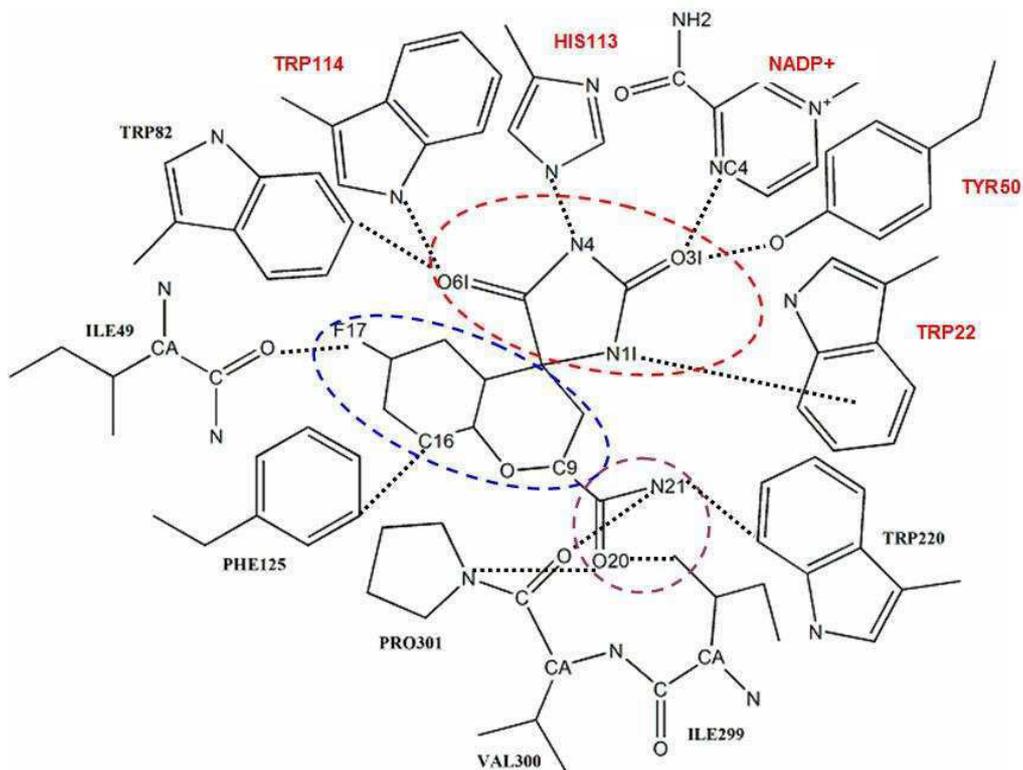


Figure 27 : Diagramme des contacts intermoléculaires impliquant le fidarestat dans le site actif de l'aldéhyde réductase porcine. Les interactions entre les atomes lourds les plus proches sont indiquées par des lignes en pointillés. Les différents groupements chimiques du fidarestat sont précisés par des cercles en pointillés colorés : le groupement hydantoin en rouge, le cycle chroman en bleu et le groupement carbamoyl en violet. Les noms des résidus appartenant à la « poche de fixation anionique » sont indiqués en rouge.

El-Kabbani *et al.* suggère que la spécificité du fidarestat envers l'aldose réductase peut s'expliquer par deux différences significatives entre les sites actifs de ces deux aldo-keto réductases [3]. Dans le site actif de l'aldose réductase, une liaison hydrogène (2,95 Å) se forme entre le groupement carbamoyl (C19—O20) de l'inhibiteur fidarestat et la chaîne principale (N—H) du résidu Leu300. Cette liaison hydrogène semble avoir un rôle majeur dans la grande affinité du fidarestat pour l'aldose réductase comparée à l'aldéhyde réductase [11]. En effet, ce résidu n'est pas conservé dans le site actif de l'aldéhyde réductase, et est remplacé par un résidu proline (Pro301) qui forme un contact court moins favorable avec l'inhibiteur. De plus, dans le site actif de l'aldéhyde réductase, la chaîne secondaire du résidu Trp220 est désordonnée ce qui n'est pas le cas dans le site actif de l'aldose réductase (Trp219). Cela serait à mettre en relation avec une différence d'interaction selon les sites actifs.

IV.4.3.B. Sélection du sous-système étudié

Pour calculer les énergies, on considère un sous-système correspondant au site actif de l'enzyme (l'inhibiteur fidarestat et les principaux résidus), équivalent à celui retenu pour les deux premiers complexes étudiés.

Ce sous-système est composé des atomes non hydrogène de l'inhibiteur fidarestat, du groupement nicotinamide du cofacteur NADP⁺, des résidus constituant la poche de fixation du groupement hydantoin de l'inhibiteur : Tyr50, His113 et Trp114, mais aussi les résidus Pro301, Trp82, Trp22, Phe125, Ile49, Ile299, Trp220, Val300.

Comme dans le cas du complexe aldose réductase humaine – fidarestat, certains résidus du site actif sont désordonnés : le résidu Trp220, comme cela a été dit précédemment, mais aussi l'extrémité de la chaîne latérale du résidu Val300. Les atomes correspondants aux conformations de ces résidus seront considérés comme des atomes d'occupation partielle fixe. Les coordonnées atomiques d'un total de 159 atomes sont donc considérées.

Une première étude a été réalisée en considérant la structure ici du serveur RCSB [4] comme structure initiale. Cette première étude réalisée sur 800 structures a permis d'obtenir les valeurs des modes principaux des distributions des coordonnées qui se sont révélées significativement différents des valeurs des coordonnées dans la structure initiale.

Une étude statistique a finalement été réalisée en considérant une structure initiale dont les coordonnées atomiques du sous-système correspondent aux modes précédemment déterminés. Pour réaliser cette étude, 1700 structures ont été obtenues suivant la procédure (Perturbation/Affinement) en utilisant une amplitude maximale de perturbation *PerturMax* de 0,1 Å.

IV.4.3.C. Etudes statistiques des indicateurs de qualité

L'analyse statistique des indicateurs de qualité fournis par le programme SHELXL [35] a été réalisée (**Tableau 12**).

Indicateurs	$R_{1_free_sigc4}$	$R_{1_free_all}$	R_{1_sigc4}	R_{1_all}	R_w	gof	gof_rest
Moyenne m	0,2632	0,2642	0,2114	0,2114	0,517	7,308	5,655
Ecart-type σ^*	0,0008	0,0008	0,0002	0,0002	0,0005	0,010	0,009
Mode ω	0,2629	0,2643	0,2114	0,2114	0,5169	7,305	5,651
Médiane M	0,2632	0,2642	0,2114	0,2114	0,517	7,308	5,654
Minimum	0,2608	0,2618	0,2109	0,2109	0,5157	7,278	5,631
Maximum	0,2659	0,2669	0,2119	0,2119	0,5188	7,342	5,703
Coefficient							
Skewness γ_1^*	0,051	0,048	-0,027	0,0221	0,179	0,151	0,624
Coefficient							
Kurtosis β_2^*	0,044	0,037	0,064	0,041	-0,125	-0,141	1,043
Rapport							
skewnew	0,9	0,8	-0,4	0,4	3,0	2,5	10,5
Rapport							
Kurtosis	0,4	0,3	0,5	0,3	-1,1	-1,2	8,8

Tableau 12 : Statistiques sur les distributions des valeurs des indicateurs de qualité au cours de l'étude du complexe aldéhyde réductase – fidarestat. Pour 1700 événements, les erreurs standard sur les coefficients de Pearson de Skewness et de Kurtosis sont respectivement de 0,059 et 0,118.

Les indicateurs de qualité présentent des distributions plus étroites que dans le cas du complexe aldose réductase - sorbinil, mais aussi plus larges que pour le complexe aldose réductase – fidarestat, ce qui tend à suggérer une dépendance en résolution de leurs écarts-types. L'écart-type de l'indicateur $R_{1_free_sigc4}$ était de 0,0008 dans le cas du complexe aldéhyde réductase – fidarestat, alors qu'il est respectivement de 0,0014 et 0,0002 pour les structures des complexes aldose réductase – sorbinil (**Tableau 1**) à 2,30 Å, et aldose réductase – fidarestat (**Tableau 10**) à 0,92 Å.

On observe également cet état intermédiaire pour les valeurs des deux indicateurs goodness-of-fit (gof , gof_{rest}). Leurs valeurs moyennes respectives sont de 7,308 et 5,655. Elles sont donc comprises entre celles des complexes précédents (3,728 et 3,650 pour le complexe aldose réductase – fidarestat, et 21,842 et 5,738 pour le complexe aldose réductase – sorbinil). Elles sont également proches mais moins que dans le cas du complexe aldose réductase – fidarestat. Une nouvelle fois, des valeurs proches de l'unité de goodness-of-fit traduisent l'aptitude du modèle à reproduire les résultats expérimentaux. La valeur du goodness-of-fit dépend notamment de la différence entre le nombre d'observations et le nombre de variables du système. Tous ces complexes étant similaires par nature, leurs nombres de variables sont les mêmes. Un nombre de réflexions (ou observations) plus important tendra à faire diminuer la valeur du goodness-of-fit, ce que l'on observe concrètement par une diminution des goodness-of-fit avec la résolution dans l'espace réel.

Seuls trois « outliers de restriction » ont été identifiés par observation de l'histogramme du gof_{rest} , et exclus.

IV.4.3.D. Etudes statistiques des coordonnées atomiques

Les algorithmes d'exclusion des « outliers conformationnels » ont été utilisés sur les distributions des coordonnées des 159 atomes du site actif.

Sur les 1697 conformations obtenues, 71 ont été automatiquement exclues, ce qui réduit le jeu de conformations à 1626.

Comme pour le cas du complexe aldose réductase – fidarestat, l'on peut distinguer les parties désordonnées du reste du site actif.

L'écart-type moyen des coordonnées des atomes appartenant aux parties ordonnées du site actif est de 0,009 Å, pour une valeur maximale de 0,05 Å pour la coordonnée Y de l'atome O du résidu Trp220. On observe des écarts-types supérieurs à la moyenne pour les atomes O, N et C du résidu Trp220, ce qui peut simplement s'expliquer par le désordre de sa chaîne latérale.

Pour les 13 atomes désordonnés (ce qui fait donc 26 atomes d'occupation partielle), les écarts-types moyens de leurs coordonnées sont nettement plus élevés avec 0,05 Å, pour une valeur maximale de 0,2 Å pour la coordonnée Y de l'atome CG1 conformation A du résidu Val300.

Les écarts-types sont suffisamment significatifs pour que l'on puisse examiner les rapports de coefficients de Pearson. Pour les coordonnées des atomes des parties ordonnées, les rapports des coefficients de Pearson demeurent compris entre [-4 ; 4] ce qui reste raisonnable compte tenu du nombre de conformations (ou évènements) prises en compte. La distribution des valeurs de la coordonnée Z de l'atome O20 du fidarestat qui a le rapport de coefficient de Skewness le plus élevé (4,0) est donnée à titre d'exemple (**Figure 28**). L'on peut constater que l'étalement sur la droite reste faible.

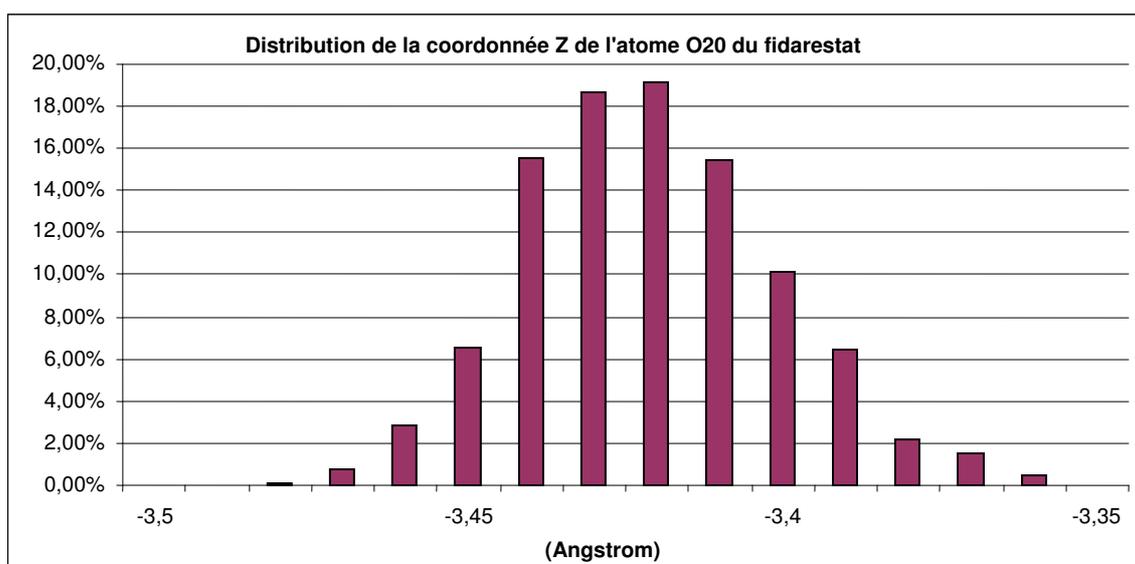


Figure 28 : Histogramme des populations de la coordonnée Z de l'atome O20 du fidarestat.

Pour les coordonnées des atomes désordonnés, les rapports des coefficients de Pearson sont plus élevés. Le rapport de coefficient de Skewness d'amplitude maximale est de -12,0, et correspond à la coordonnée Y de l'atome CG2 conformation B du Val300. La dissymétrie de la distribution est ici beaucoup plus prononcée comme le confirme la **Figure 29**, et cela malgré une valeur initiale de cette coordonnée (indiquée en bleu sur cette figure) proche du mode principal de sa distribution. Cela met une nouvelle fois en évidence la difficulté du traitement des parties désordonnées.

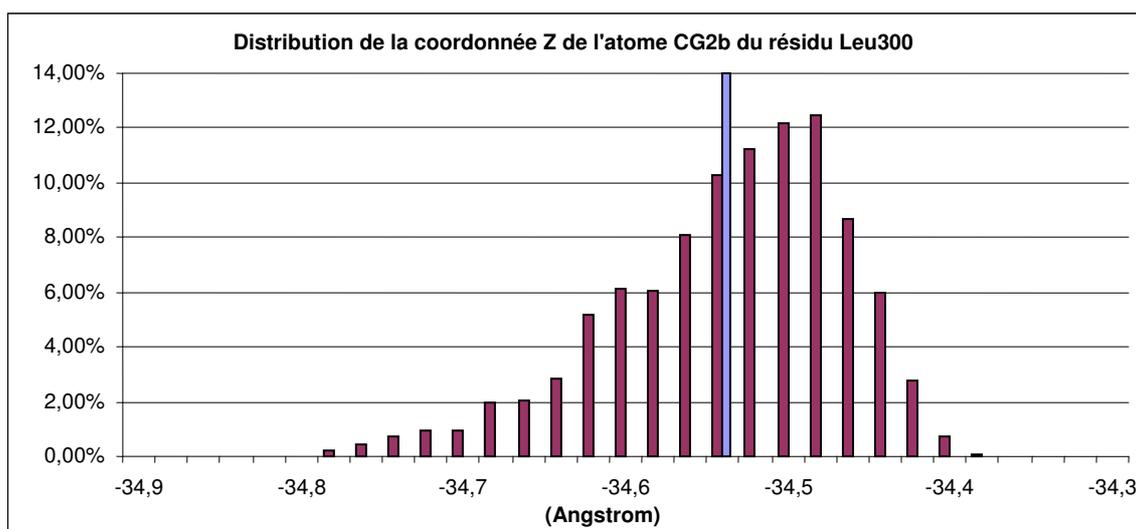
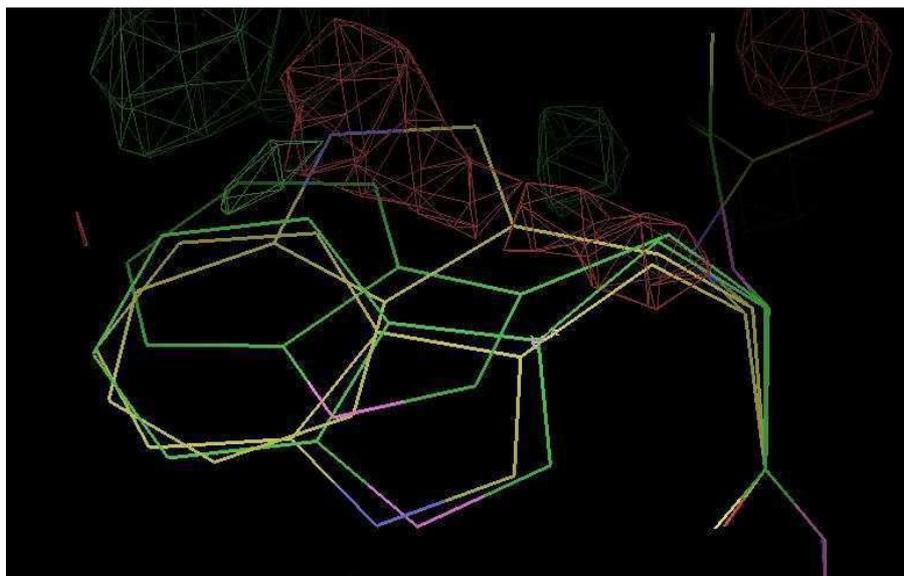


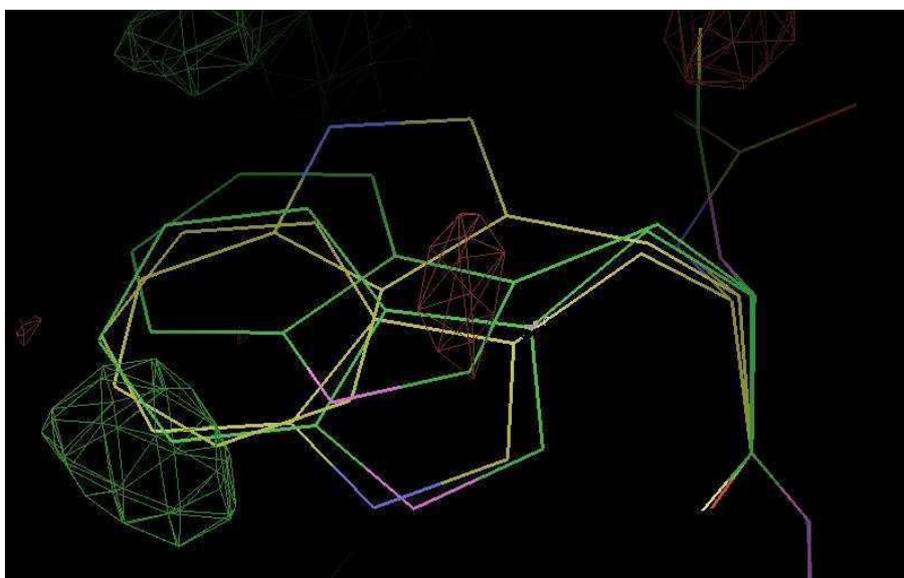
Figure 29 : Histogramme des populations de la coordonnée Z de l'atome CB2 conformation B du résidu Val300. La valeur de la coordonnée dans la structure initiale utilisée pour cette étude est indiquée en bleu.

On constate à nouveau que l'utilisation pour l'étude d'une structure initiale corrigée permet d'obtenir des valeurs moyennes plus proches des valeurs initiales. En effet, si l'on compare les coordonnées atomiques de la structure originale, ici du serveur RCSB, aux valeurs des moyennes correspondantes, on note pour les atomes des parties ordonnées, une différence absolue moyenne de 0,047 Å, et une différence maximale de 0,40 Å pour la coordonnée Y de l'atome N de la chaîne principale du résidu Trp220 qui est partiellement désordonnée. Pour les parties désordonnées, ces différences moyenne et maximale sont nettement supérieures avec 0,50 Å et 3,73 Å (coordonnée X atome NE1b de la chaîne latérale du Trp220).

En comparant les coordonnées atomiques de la structure initiale corrigée suite à l'étude préliminaire, les différences absolues moyenne et maximale pour les parties ordonnées ne sont plus que 0,014 Å et 0,25 Å (coordonnée Y atome O de la chaîne principale du résidu Trp220), et pour les parties désordonnées, 0,16 Å et 0,70 Å (coordonnée Z atome CZ2b de la chaîne latérale du Trp220). Les cartes de densité résiduelle au voisinage de la chaîne latérale du Trp220 obtenue pour la conformation initiale issue du serveur RCSB [4] et pour l'une des conformations dont les valeurs des coordonnées atomiques du sous-système étudié appartiennent aux modes principaux de leurs distributions ont été calculées. Leur comparaison permet de constater que le déplacement sur la gauche de la conformation B de la chaîne latérale du Trp220 supprime certains pics résiduels négatifs, bien que cela ne soit qu'à un niveau d'isomailage relativement faible de 0,20 e.Å⁻³ (**Figure 30**).



(a)



(b)

Figure 30 : Illustrations des cartes de densité électronique résiduelle en trois dimensions au voisinage de la chaîne latérale du résidu Trp202 du complexe aldéhyde réductase - fidarestat. La conformation initiale issue du serveur PDB est représentée en jaune, et l'une des conformations optimales en vert. La carte de densité résiduelle (a) correspond à la conformation initiale, la carte (b) à la conformation optimale. Les isomailles sont définies à $0,20 \text{ e.}\text{\AA}^{-3}$.

La représentation en nuages de points des couples (Uiso ; écart-type_moyen) pour tous les atomes non hydrogène du sous-système retenu (**Figure 31**) met en évidence une corrélation entre les écarts-types statistiques moyens des coordonnées des atomes des parties ordonnées

et les valeurs de leurs paramètres d'agitation thermique. Pour chaque atome, l'écart-type statistique moyen est calculé en réalisant la moyenne arithmétique des écarts-types statistiques de ces trois coordonnées, et la valeur du paramètre d'agitation thermique U_{iso} est issue de la conformation initiale utilisée pour générer le jeu de conformations analysé dans cette étude.

Cette corrélation n'est pas observée chez les atomes des parties désordonnées. Leurs couples (U_{iso} ; écart-type moyen) constituent un nuage de points nettement dispersé.

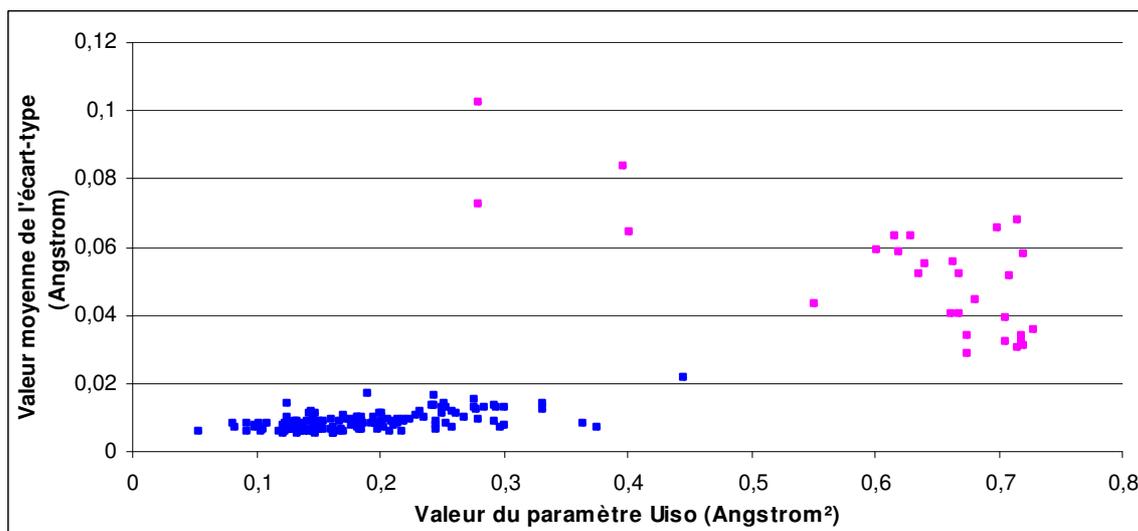


Figure 31 : Représentation en nuage de points des couples (U_{iso} ; écart-type moyen) pour tous les atomes non hydrogène du sous-système retenu dans l'étude du complexe aldéhyde réductase - fidarestat. Les points en violet correspondent aux atomes appartenant aux parties désordonnées, et les points en bleu aux atomes des parties ordonnés dans le site actif du complexe.

IV.4.3.E. Détermination des distributions des énergies d'interaction électrostatique.

A partir de chaque structure, un modèle de distribution de charges asphérique est défini selon la même procédure que précédemment. Comme dans le cas du complexe aldose réductase – fidarestat, l'inhibiteur fidarestat au sein du site actif est considéré sous sa forme neutre, et le résidu His113 est considéré neutre de type HID (protoné une fois sur l'azote ND2). L'inhibiteur fidarestat est impliqué dans 12 interactions dans le site actif de la protéine.

Les résultats de l'analyse statistique sur les énergies d'interaction électrostatique sont communiqués dans le **Tableau 13**.

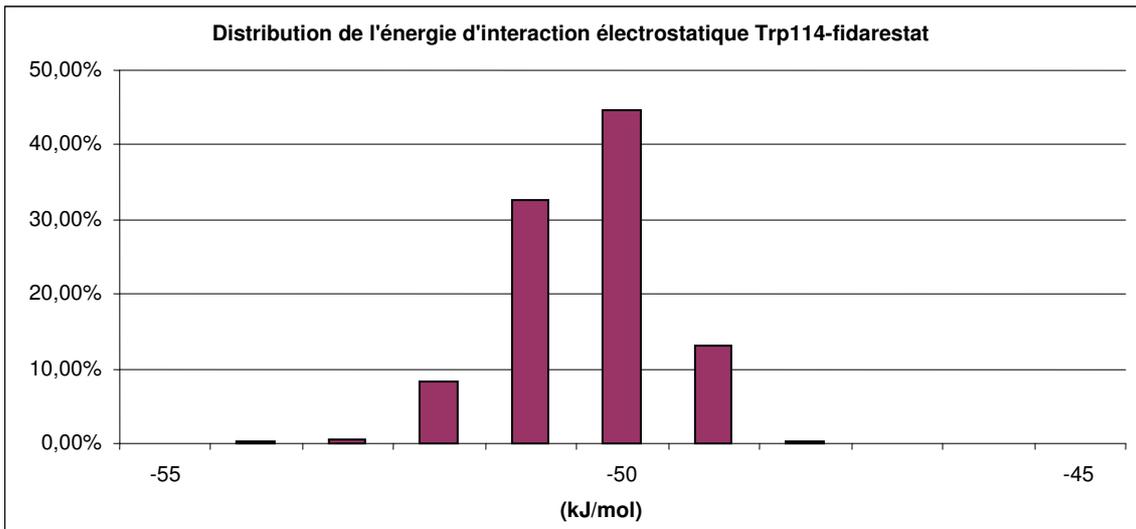
Aldéhyde réductase - Fidarestat				Moyenne en kJ.mol ⁻¹	Ecart-type en kJ.mol ⁻¹	Valeur minimale en kJ.mol ⁻¹	Valeur maximale en kJ.mol ⁻¹	Largeur en kJ.mol ⁻¹	Rapport de Skewness	Rapport de Kurtosis
Residu	Atomes non-H les plus proches	Distance (Å)								
His 113	NE2...N4	2,78	-40	2	-44	-36	8	0	0	
Pro 301	N...O20	3,09	-18	1	-20	-17	3	---	---	
Trp 114	NE...O6I	2,87	-50	2	-54	-48	6	-6	2	
Tyr 50	OH...O3I	2,47	-51	3	-59	-43	16	-2	-3	
Ile49	O...F17	3,27	-5	1	-5	-4	1	---	---	
Ile299	CG2...O20	3,41	8	1	6	9	3	---	---	
Val300	O...N21	3,17	-27	1	-30	-24	6	-6	0	
Phe125	CZ...C15	3,95	-3	1	-4	-3	1	---	---	
Trp22	Ring...N1I	3,17	-50	1	-57	-48	9	-3	11	
Trp220	CH2,CZ2...N21	3,30 ; 3,34	-5	1	-8	-3	5	2	2	
Trp82	CZ3...O6I	4,00	-6	1	-6	-5	1	---	---	
NADP ⁺	NC4...O3I	3,01	20	1	17	23	6	-3	1	

Tableau 13 : Toutes les interactions entre l'inhibiteur fidarestat et les résidus du site actif de l'aldéhyde réductase porcine sont décrites. Les atomes non hydrogène les plus proches entre l'inhibiteur fidarestat et les résidus sont donnés avec les distances correspondantes. Deux distances sont indiquées dans le cas de l'interaction entre le fidarestat et le résidu Trp220 en raison du désordre de ce dernier.

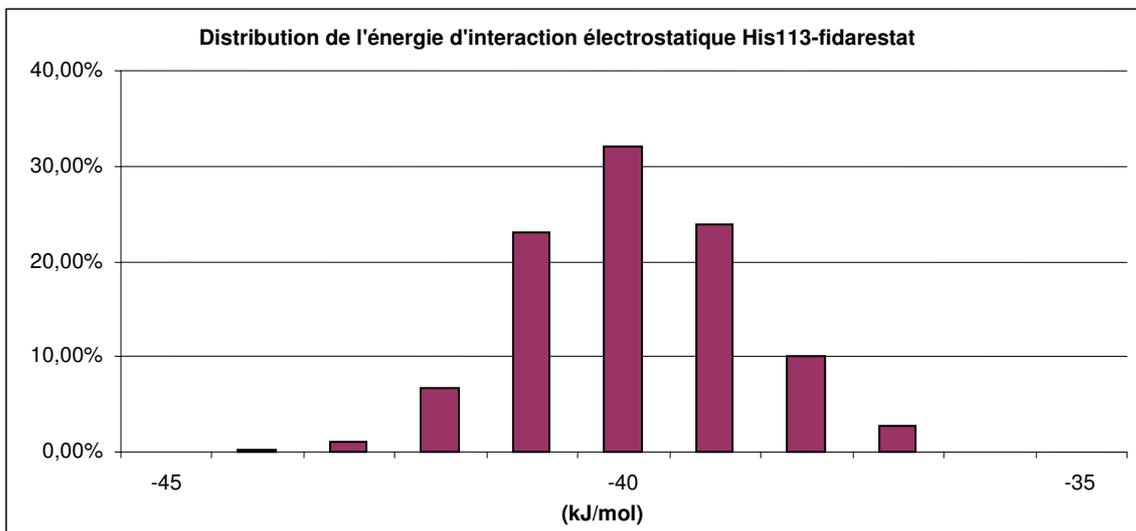
On observe à nouveau des largeurs de distribution supérieures pour les énergies d'interaction électrostatique d'amplitude importante.

Les écarts-types des distributions d'énergies ne sont au maximum que de quelques $\text{kJ}\cdot\text{mol}^{-1}$ comme cela est le cas avec les énergies des interactions au sein du site actif du complexe aldose réductase – sorbinil. L'écart-type maximal obtenu est celui de l'énergie de l'interaction électrostatique entre l'inhibiteur et le résidu Tyr50 avec $3 \text{ kJ}\cdot\text{mol}^{-1}$, alors que dans le cas du complexe aldose réductase – sorbinil, l'écart-type maximal est de $5 \text{ kJ}\cdot\text{mol}^{-1}$ pour l'interaction équivalente sorbinil – Tyr48 (**Tableau 13**). Les largeurs des distributions de par leurs amplitudes permettent plus aisément d'apprécier l'incertitude de chaque énergie.

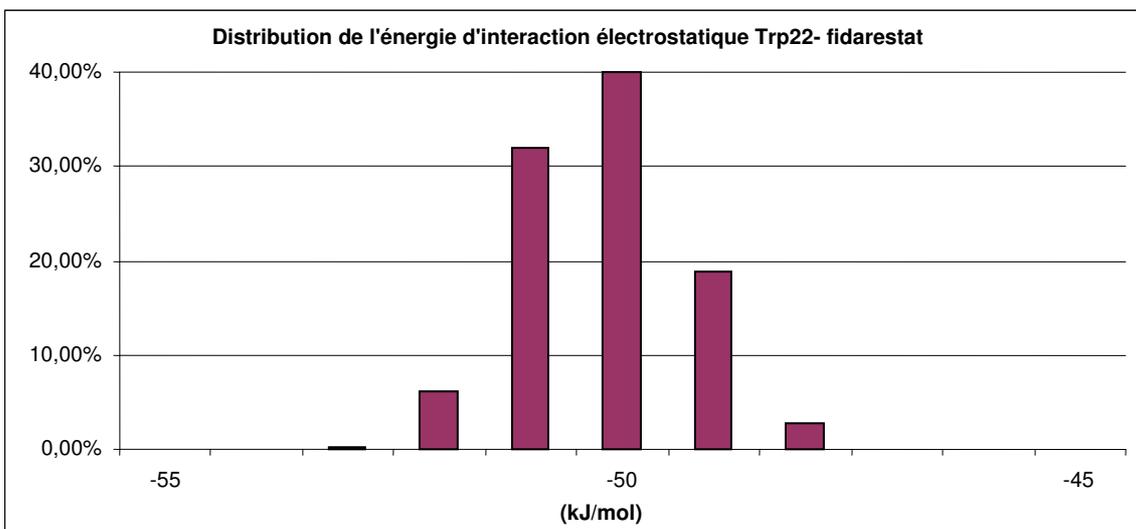
Quatre interactions de l'inhibiteur sorbinil présentent des écarts-types statistiques et des largeurs de distribution suffisamment importants pour permettre l'examen de leurs histogrammes des populations : celles impliquant les résidus Trp114, His113, Trp22 et Tyr50 (**Figure 32**).



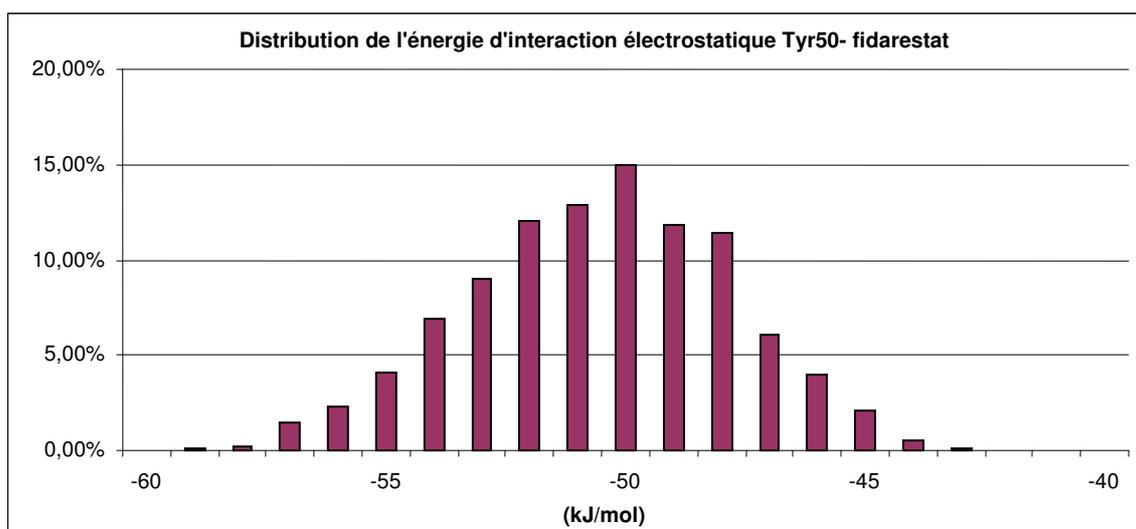
(a)



(b)



(c)



(d)

Figure 32 : Histogrammes des populations des énergies d'interaction électrostatique significatives au sein du site actif du complexe aldéhyde réductase porcine – fidarestat (en $\text{kJ}\cdot\text{mol}^{-1}$).

On notera que la valeur élevée de rapport de Kurtosis (11,0) de la distribution de l'énergie d'interaction fidarestat – Trp22 est due à un unique outlier qui n'apparaît pas sur l'histogramme des populations.

Les distributions ne présentent pas de dissymétries visibles sur les histogrammes de populations. La précision numérique sur les calculs d'énergie d'interaction électrostatique ne permet pas de discuter d'avantage leurs profils.

IV.4.4. Cas de la structure de l'aldose réductase complexée avec le stéréoisomère (2*R*,4*S*) de l'inhibiteur fidarestat

IV.4.4.A. Présentation de la structure

La structure du complexe aldose réductase – stéréoisomère (2*R*,4*S*) de fidarestat avec, comme cofacteur le NADP⁺, a été affinée à une résolution de 1,40 Å par El-Kabbani *et al.* [29]. La maille a pour groupe d'espace P2₁.

La structure cristallographique de ce complexe a été obtenue dans le but d'élucider la stéréospécificité de la fixation du fidarestat.

Dans le complexe aldose réductase humaine - stéréoisomère (2*R*,4*S*), le groupement de "tête" hydantoin est dans la même configuration que le fidarestat (stéréoisomère (2*S*,4*S*)) dans la "poche de fixation de l'anion". Ainsi, le stéréoisomère (2*R*,4*S*) participe aux mêmes liaisons hydrogène avec les résidus Tyr48, His110 et Trp111 avec des distances d'interaction similaires. De la même manière, le stéréoisomère (2*R*,4*S*) interagit également avec le groupement nicotinamide du cofacteur NADP⁺.

Dans le complexe aldose réductase - fidarestat, il y a plusieurs contacts courts entre les chaînes latérales des résidus Trp219 et Cys298 (3,61 Å et 3,29 Å, respectivement) et le groupement carbamoyl du fidarestat, qui sont significativement différents dans le cas du complexe aldose réductase - stéréoisomère (2*R*,4*S*). La liaison hydrogène entre le groupement carbamoyl et le résidu Leu300 est toujours présente (2,80 Å), mais le changement de configuration du carbone chiral C9 dans le site actif induit des contacts proches plus courts entre les chaînes latérales des résidus Trp219 et Cys298, et le groupement carbamoyl (3,00 Å et 2,98 Å, respectivement). Cela peut contribuer à une affinité plus faible du stéréoisomère (2*R*,4*S*) avec l'aldose réductase d'après El-Kabbani *et al.* [29].

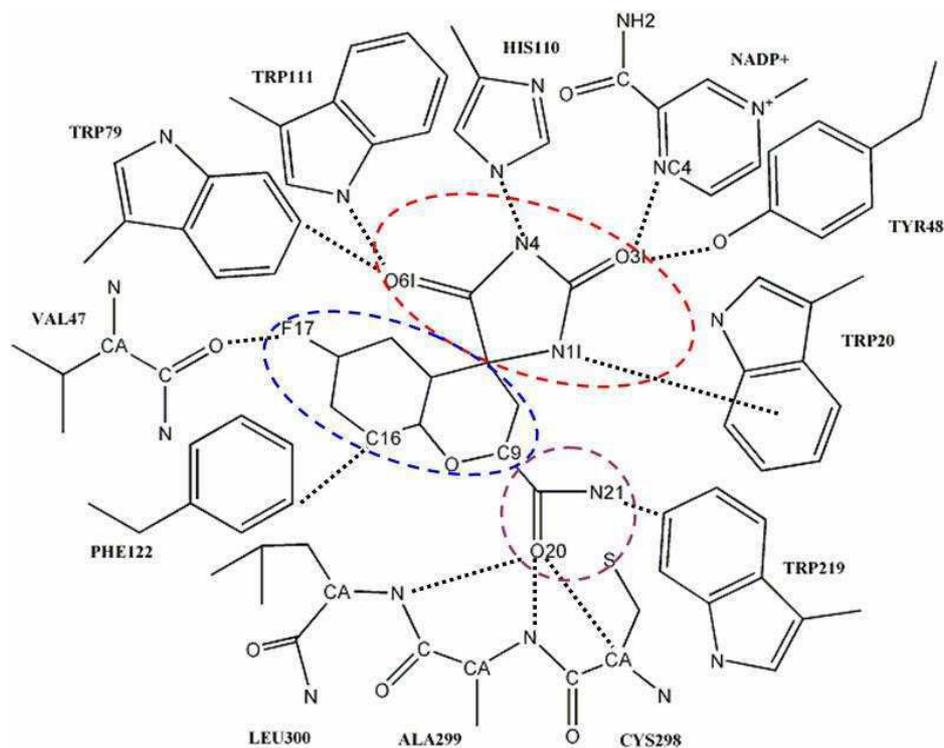


Figure 33 : Diagramme des contacts intermoléculaires impliquant le stéréoisomère ($2R,4S$) du fidarestat dans le site actif de l'aldose réductase humaine. Les interactions entre les atomes lourds les plus proches sont indiquées par des lignes en pointillés. Les différents groupements chimiques du stéréoisomère ($2R,4S$) sont précisés par des cercles en pointillés colorés : le groupement hydantoin en rouge, le cycle chroman en bleu et le groupement carbamoyl en violet. Les noms des résidus appartenant à la « poche de fixation anionique » sont indiqués en rouge.

Le stéréoisomère ($2R,4S$) du fidarestat est impliqué au sein du site actif de l'aldose réductase dans les mêmes contacts intermoléculaires que le fidarestat (stéréoisomère ($2S,4S$)) (**Figure 33**). Par conséquent, le sous-système étudié sera strictement le même que dans le cas du complexe aldose réductase - fidarestat. On s'intéressera donc aux coordonnées atomiques des atomes de l'inhibiteur stéréoisomère ($2R,4S$), des résidus Trp20, Val47, Tyr48, Trp79, His110, Trp111, Phe122, Trp219, Cys298, Ala299, Leu300, ainsi que le groupement nicotinamide du cofacteur $NADP^+$.

L'atome SG de la chaîne terminale du résidu Cys298 est l'unique atome présentant un désordre dans le site actif du complexe aldose réductase – stéréoisomère, ce qui fait donc un total de 143 atomes non hydrogène.

Le cas du complexe aldose réductase - stéréoisomère est, en dépit de sa résolution expérimentale atomique ($< 1,5 \text{ \AA}$), délicat. En effet, contrairement aux inhibiteurs des complexes précédemment examinés, le stéréoisomère (2*R*,4*S*) n'occupe que partiellement le site actif de l'enzyme, ce qui s'accompagne également d'une agitation thermique supérieure. Par ailleurs, il partage ce dernier avec des molécules d'eau qui ne sont pas prises en compte dans le sous-système perturbé et affiné. Cette mesure prise pour des raisons pratiques peut nuire à l'étude statistique.

Comme pour le complexe aldéhyde réductase – fidarestat, un premier test a été réalisé afin d'identifier les modes de distribution des coordonnées atomiques, mais ce dernier n'a pas été concluant compte tenu des écarts-types sur les coordonnées atomiques.

Pour réaliser cette étude, 1600 structures ont été obtenues suivant la procédure (Perturbation/Affinement) définie précédemment en utilisant une amplitude maximale de perturbation *PerturMax* de $0,1 \text{ \AA}$. 87 conformations sont exclues du jeu de données en raison de restrictions de distances non appliquées par de SHELXL. Seuls 1513 conformations sont donc utilisées.

IV.4.4.B. Etudes statistiques des indicateurs de qualité

L'analyse statistique des indicateurs de qualité fournis par le programme SHELXL [35] a été réalisée (**Tableau 14**).

Les indicateurs de qualité présentent les distributions plus larges que pour les autres complexes étudiés. L'écart-type de l'indicateur $R_{1_free_sigc4}$ est de 0,0019 alors que dans le cas du complexe aldose réductase – sorbinil, sa valeur est de 0,0014 (**Tableau 1**). Les valeurs des deux indicateurs goodness-of-fit (gof , gof_{rest}) sont de l'ordre de l'unité avec respectivement 1,483 et 1,401. Les distributions des indicateurs de qualité se distinguent des cas précédemment étudiés par des rapports de Pearson élevés et tout particulièrement des rapports de Skewness positifs ce qui peut se traduire par un étalement sur la droite de toutes ses distributions.

L'indicateur gof_{rest} est celui présentant les rapports de Pearson les plus élevés (**Tableau 14**). L'histogramme des populations de l'indicateur gof_{rest} (**Figure 34**) révèle la présence de « outliers de restrictions ». Une valeur limite de 1,55 a été choisie pour les valeurs de gof_{rest} , ce qui a conduit à l'exclusion de 22 conformations.

Indicateurs	$R_{1_free_sigc4}$	$R_{1_free_all}$	R_{1_sigc4}	R_{1_all}	R_w	gof	gof_rest
Moyenne m	0,2405	0,2477	0,2108	0,2143	0,4749	1,483	1,401
Ecart-type σ^*	0,0019	0,0019	0,0019	0,0019	0,0043	0,022	0,047
Mode ω	0,24	0,2476	0,2097	0,2144	0,4734	1,481	1,395
Médiane M	0,2404	0,2476	0,2106	0,2141	0,4747	1,481	1,392
Minimum	0,2357	0,2431	0,2057	0,2091	0,4633	1,437	1,332
Maximum	0,2522	0,2592	0,222	0,2251	0,5045	1,661	1,908
Coefficient							
Skewness γ_1^*	0,668	0,631	0,783	0,705	1,128	2,047	3,651
Coefficient							
Kurtosis β_2^*	1,697	1,515	2,203	1,810	4,321	10,473	25,536
Rapport							
skewnew	10,6	10,0	12,4	11,1	17,8	32,4	57,7
Rapport							
Kurtosis	13,4	12,0	17,4	14,3	34,6	82,8	201,9

Tableau 14 : Statistiques sur les distributions des valeurs des indicateurs de qualité au cours de l'étude du complexe aldose réductase – stéréoisomère (2R,4S) du fidarestat. Pour 1500 événements, les erreurs standard sur les coefficients de Pearson de Skewness et de Kurtosis sont respectivement de 0,063 et 0,126.

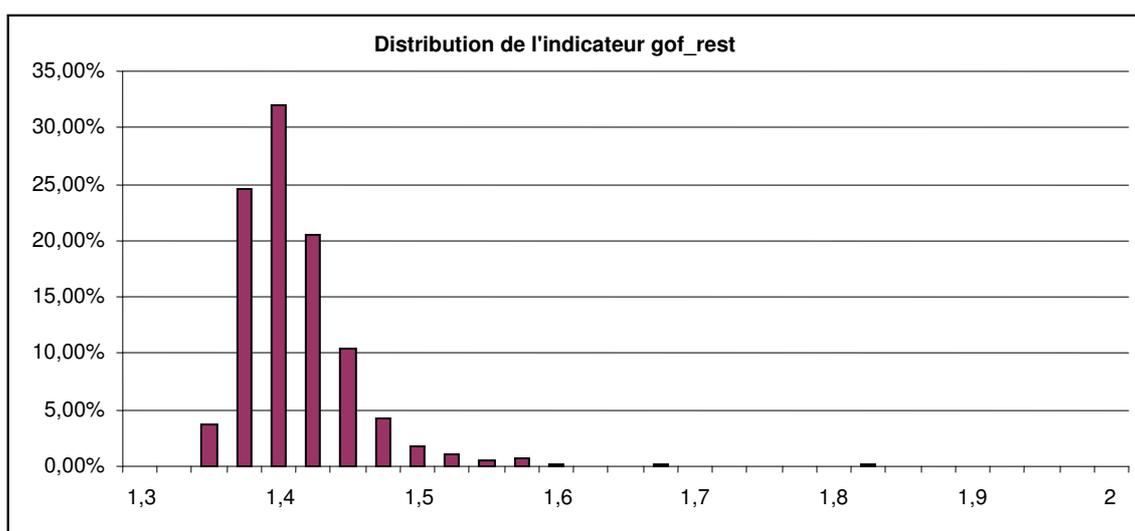


Figure 34 : Histogramme des populations de l'indicateur gof_rest du complexe aldose réductase – stéréoisomère (2R,4S) du fidarestat.

IV.4.4.C. Etudes statistiques des coordonnées atomiques

Les algorithmes d'exclusion des « outliers conformationnels » ont été utilisés sur les distributions des coordonnées des 143 atomes du site actif. Sur les 1491 conformations obtenues, 153 ont été automatiquement exclues, ce qui réduit le jeu à 1338 conformations.

Comme attendu, les coordonnées des atomes de l'inhibiteur se distinguent par des valeurs d'écart-type supérieures. En effet, la valeur moyenne des écarts-types des coordonnées atomiques de l'inhibiteur est de 0,050 Å, pour une valeur maximale de 0,174 Å pour la coordonnée X de l'atome F17. La localisation de l'inhibiteur stéréoisomère (2R,4S) du fidarestat paraît donc imprécise.

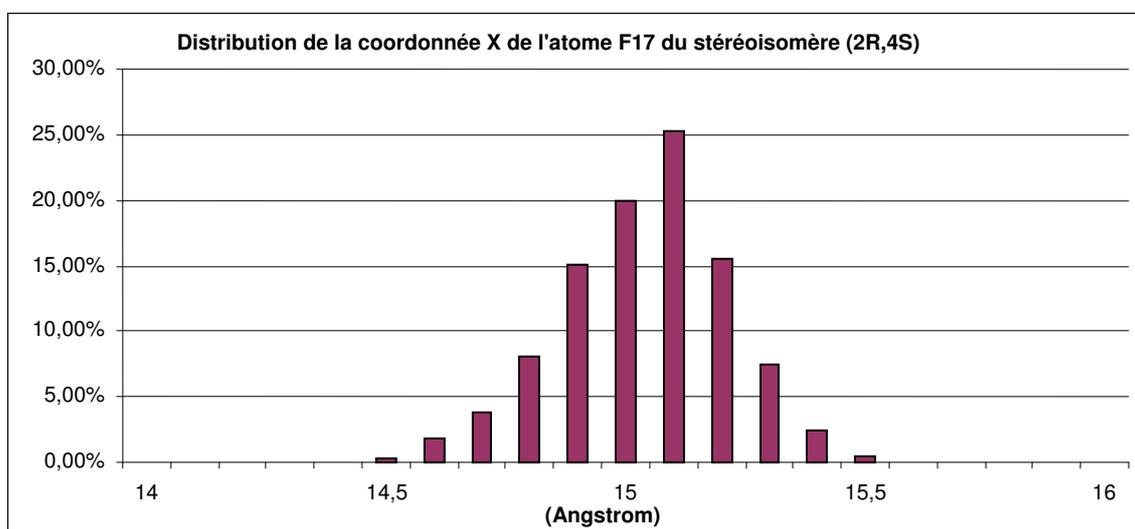


Figure 35 : Histogramme des populations de la coordonnée X de l'atome F17.

L'histogramme des populations de cette coordonnée (**Figure 35**) révèle une largeur de distribution de 1 Å, avec malgré cela un mode unique.

Les coordonnées des autres atomes du site actif présentent une valeur d'écart-type moyenne de 0,011 Å, et une valeur maximale de 0,048 Å pour la coordonnée X de l'atome O du résidu Ala299. Ces valeurs d'écart-type moyen et maximal sont comparables à celles observées pour les parties ordonnées du site actif du complexe aldéhyde réductase – fidarestat.

On notera que, malgré tout, les différences absolues entre les valeurs moyennes des coordonnées et leurs valeurs initiales dans la structure du serveur RCSB [4] restent globalement inférieures aux écarts-types. Pour les atomes du fidarestat, la différence moyenne

est de 0,026 Å, et la différence maximale de 0,13 Å qui correspond justement à la coordonnée X de l'atome F17. Pour les autres atomes du site actif, la différence moyenne est de 0,007 Å, et la différence maximale de 0,064 Å (coordonnée X de l'atome CB du Trp219).

Ces coordonnées atomiques présentent des valeurs de rapports des coefficients de Pearson comprises entre [-6 ; 6]. Cette gamme de valeurs de rapports de Pearson est plus large que celle observée dans le cas du complexe aldéhyde réductase – fidarestat.

Les coordonnées des atomes de l'inhibiteur stéréoisomère ont des valeurs de rapport de Skewness entre [-11 ; 8], et de rapport de Kurtosis entre [-2 ; 21]. Pour exemple, l'histogramme des populations de la coordonnée Z de l'atome de fluor F17 de rapport de Skewness -21 est donné (**Figure 36**). L'étalement sur la gauche de la distribution est nettement visible.

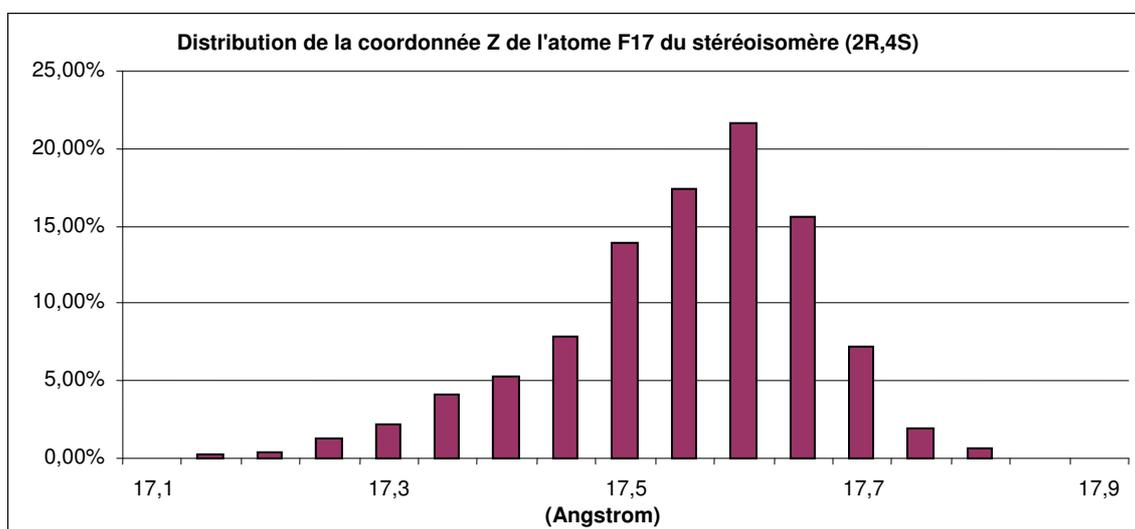


Figure 36 : Histogramme des populations de la coordonnée Z de l'atome F17 du stéréoisomère (2R,4S).

Une corrélation entre les écarts-types des coordonnées des atomes et les paramètres d'agitation thermique correspondants est mise en évidence à l'aide de la **Figure 37**.

L'atome ayant le paramètre d'agitation thermique le plus élevé est le fluor F17 de l'inhibiteur avec $U_{iso} = 0,508 \text{ \AA}^2$ ce qui correspond à un facteur B de 40,1 Å². Son écart-type moyen est le plus élevé avec 0,12 Å. Cet atome est par nature le plus lourd de l'inhibiteur, obtenir un paramètre de déplacement atomique thermique élevé était donc prévisible.

Globalement, les atomes de l'inhibiteur présentent les paramètres d'agitation thermique les plus élevés, comme il est possible de le voir (**Figure 37**). La valeur moyenne $U_{iso,moy}$ des

paramètres d'agitation thermique pour les atomes de l'inhibiteur est de $0,291 \text{ \AA}^2$, soit un facteur B moyen de $23,0 \text{ \AA}^2$. Les autres atomes du site actif présentent une valeur moyenne $U_{\text{iso,moy}}$ de $0,136 \text{ \AA}^2$ (facteur B de $10,7 \text{ \AA}^2$). On observe également une corrélation entre les écarts-types moyens et les paramètres d'agitation thermique pour les atomes n'appartenant pas à l'inhibiteur. Les points correspondant aux atomes de l'inhibiteur qui globalement ont des paramètres d'agitation thermique plus élevés, sont beaucoup plus dispersés comme cela a pu être observé dans l'étude du complexe aldéhyde – fidarestat.

L'agitation thermique n'est pas l'unique facteur responsable des écarts-types élevés des atomes de l'inhibiteur. Dans le test préliminaire, les atomes de l'inhibiteur sorbinil présentaient également des paramètres d'agitation thermique élevés et la structure avait été obtenue à une moins bonne résolution expérimentale ($2,30 \text{ \AA}$), mais malgré cela leurs écarts-types étaient nettement inférieurs. Le complexe aldose réductase – stéréoisomère est au final le système qui présente pour les coordonnées atomiques de son sous-système les incertitudes statistiques les plus élevées. L'explication la plus raisonnable serait l'occupation partielle du site actif du complexe stéréoisomère ($2R,4S$) du fidarestat – aldose réductase, par l'inhibiteur et des molécules d'eau. Rappelons que les molécules d'eau n'ont pas été prises en compte au cours de cette étude pour des raisons pratiques.

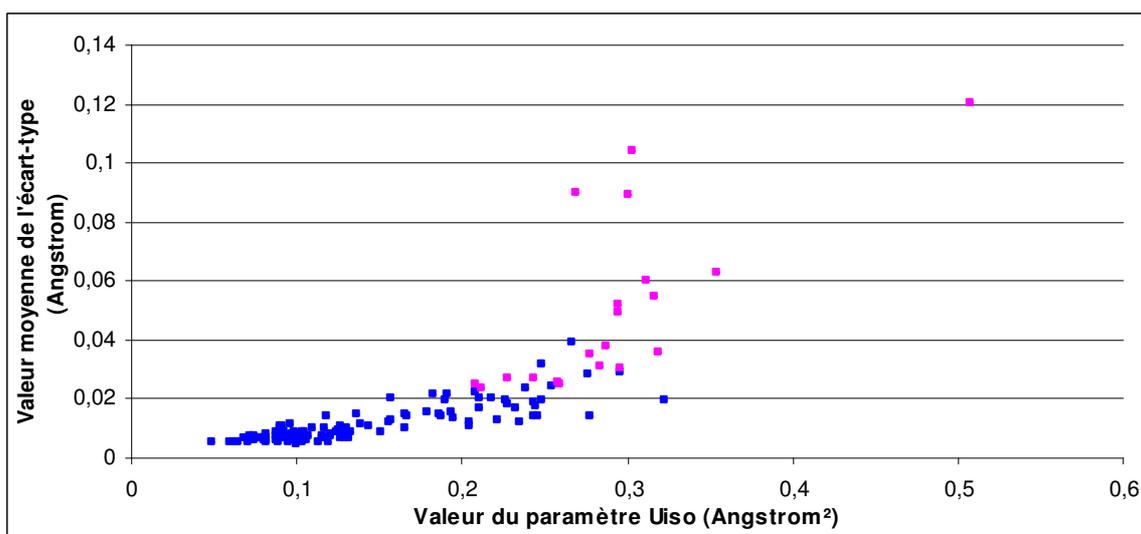


Figure 37 : Représentation en nuage de points des couples (U_{iso} ; écart-type moyen) pour tous les atomes non hydrogène du sous-système du complexe aldose réductase – stéréoisomère (2*R*,4*S*). Les points en violet correspondent aux atomes de l'inhibiteur et les points en bleu aux autres atomes dans le site actif du complexe aldose réductase – stéréoisomère (2*R*,4*S*).

IV.4.4.D. Détermination des distributions des énergies d'interaction électrostatique.

Comme dans les autres études, pour chaque structure du jeu de données, le site actif de la protéine est protoné à l'aide du programme SHELXL [4], puis un modèle de distribution de charges est défini par transfert de paramètres multipolaires depuis la banque ELMAM [30;39]. L'état de protonation du stéréoisomère (2*R*,4*S*) de l'inhibiteur fidarestat au sein du site actif est considéré neutre, et le résidu His110 est considéré neutre de type HID (protoné une fois sur l'azote ND2). Le stéréoisomère (2*R*,4*S*) de l'inhibiteur fidarestat est impliqué dans 12 interactions dans le site actif de la protéine, interactions strictement équivalentes à celles du fidarestat lorsqu'il est complexé au site actif de l'aldose réductase.

A partir des énergies d'interaction électrostatique calculées à l'aide du programme VMoPro, une analyse statistique a été réalisée sur les différentes distributions d'énergie. Les résultats sont communiqués dans le **Tableau 15**.

L'ensemble des énergies d'interaction électrostatique présente des valeurs d'écart-type nettement supérieures aux cas déjà traitées. L'écart-type maximal observé correspond à l'interaction entre le résidu Cys298 et l'inhibiteur avec 11 kJ.mol⁻¹. Néanmoins, cette

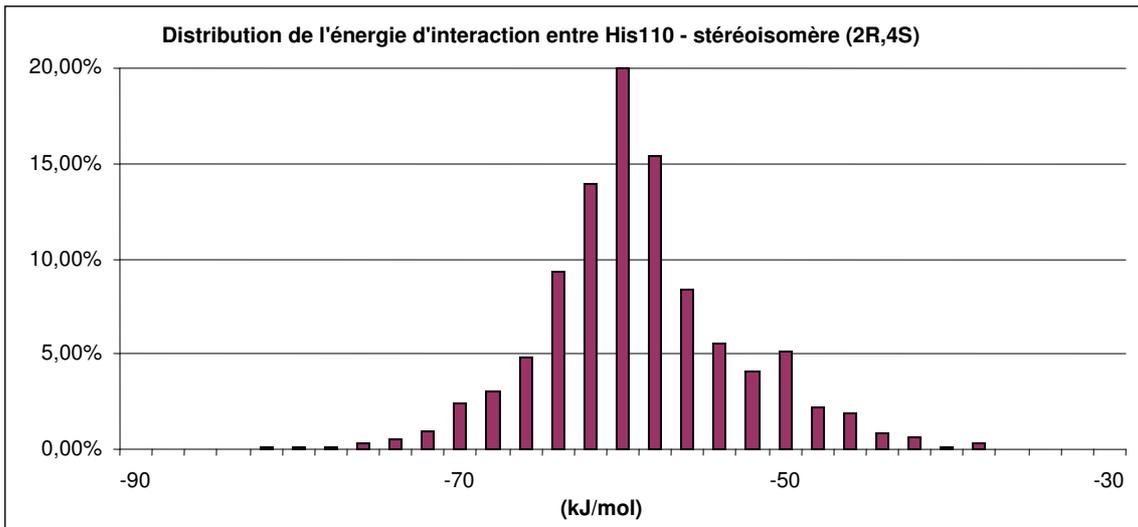
interaction est un cas particulier. En effet, l'atome SG à l'extrémité de la chaîne latérale du Cys298 est l'unique atome désordonné du site actif, et est, de plus, protoné (le résidu Cys298 étant considéré neutre). Son atome d'hydrogène de label HG-SG présente un degré de liberté qui lui permet de s'orienter selon son environnement. L'orientation des hydrogènes étant réalisée automatiquement par le programme SHELXL [35], la distribution bimodale de l'énergie d'interaction inhibiteur-Cys298 est imputable à des différences d'orientation de cet hydrogène (**Figure 38c**). Bien que la distribution de l'énergie de cette interaction puisse être ainsi écartée de l'étude, d'autres interactions présentent des distributions d'énergies unimodales particulièrement larges (**Figures 38a, 38c, 38d, 38e**).

Les énergies des interactions via liaisons hydrogène impliquant les résidus Tyr48, His110 et Trp219 ont des écarts-types supérieurs à 5 kJ.mol^{-1} , et des largeurs supérieures à 25 kJ.mol^{-1} (**Tableau 15**). On notera également que pour l'interaction Trp219 – inhibiteur, la gamme des valeurs d'énergie comprend à la fois des valeurs négatives et positives. D'un point de vue électrostatique, la nature de cette interaction (attractive ou répulsive) change donc selon la conformation considérée. De telles distributions sont une conséquence directe des incertitudes expérimentales élevées sur les coordonnées des atomes de l'inhibiteur stéréoisomère (2*R*,4*S*).

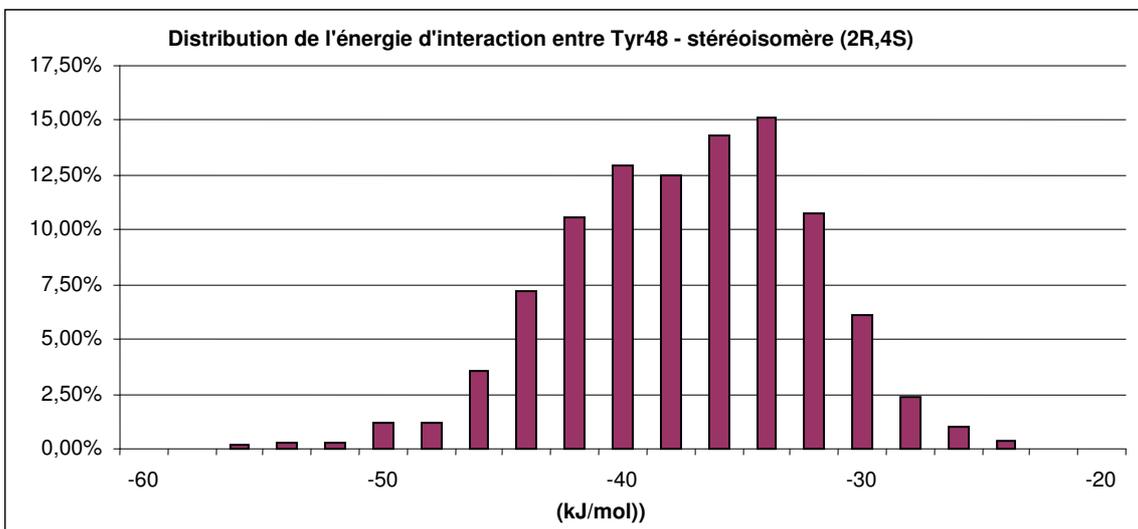
Le complexe aldose réductase – stéréoisomère (2*R*,4*S*) est un cas atypique intéressant. En dépit d'une résolution atteinte de $1,4 \text{ \AA}$, les énergies d'interaction électrostatique présentent des distributions nettement plus larges que pour le complexe aldose réductase – sorbinil. A titre de rappel, dans le cas du complexe aldose réductase – sorbinil, la distribution d'énergie la plus large est observée pour l'interaction Tyr48 – sorbinil avec un écart-type de 3 kJ.mol^{-1} et une largeur de 26 kJ.mol^{-1} (**Tableau 7**). Un tel exemple montre la nécessité d'un sous-système ordonné et de faibles agitations thermiques pour pouvoir prétendre à des énergies d'interaction électrostatique de faibles incertitudes statistiques.

Aldose réductase - Stéréoisomère (2 <i>R</i> ,4 <i>S</i>)										
Residu	Atomes non-H les plus proches	Distance (Å)	Moyenne en kJ.mol ⁻¹	Ecart-type en kJ.mol ⁻¹	Valeur minimale en kJ.mol ⁻¹	Valeur maximale en kJ.mol ⁻¹	Largeur en kJ.mol ⁻¹	Rapport de Skewness	Rapport de Kurtosis	
His 110	NE2...N4	2.70	-60	6	-83	-38	45	4	7	
Leu 300	N...O20	2.80	-26	2	-32	-20	12	-4	4	
Trp 111	NE...O6I	2.79	-45	3	-52	-39	13	-2	-1	
Tyr 48	OH...O3I	2.50	-38	5	-56	-24	32	-5	1	
Val47	O...F17	3.11	-7	2	-17	-5	12	-30	30	
Cys298	CA...O20	3.00	-13	11	-43	15	58	7	-9	
Ala299	N...O20	3.72	19	1	16	21	5	---	---	
Phe122	CE1...C16	3.70	-12	1	-14	-11	3	---	---	
Trp20	Ring...N1I	3.17	-36	2	-45	-30	15	-3	1	
Trp219	CZ2...N2I	2.98	-13	6	-45	8	53	-6	9	
Trp79	CZ3...O6I	3.48	-12	1	-14	-10	4	---	---	
NADP ⁺	NC4...O3I	2.93	14	1	10	17	7	1	1	

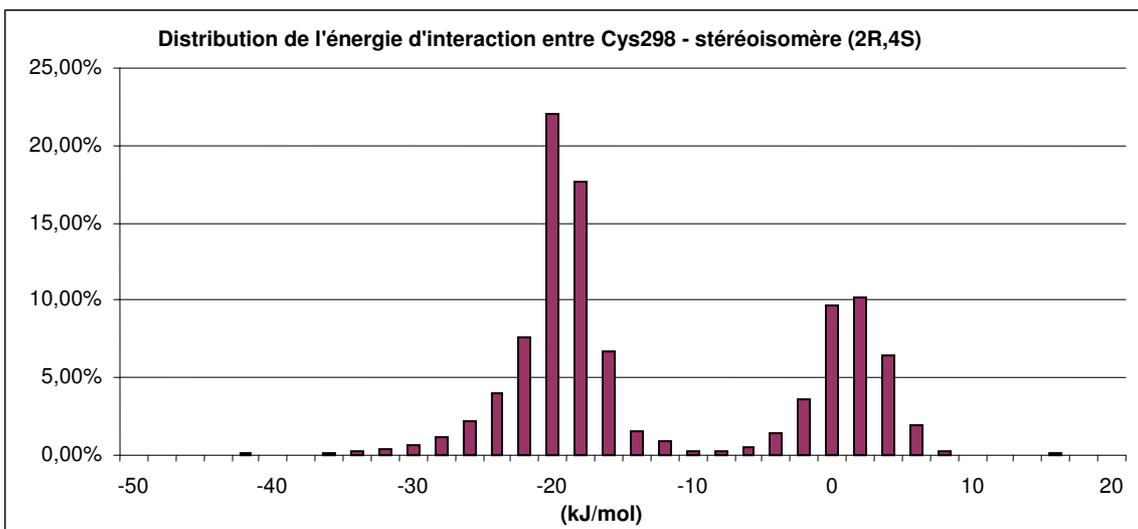
Tableau 15 : Toutes les interactions entre le stéréoisomère (2*R*,4*S*) de l'inhibiteur fidarestat et les résidus du site actif de l'aldose réductase humaine sont décrites. Les atomes non hydrogène les plus proches entre le stéréoisomère (2*R*,4*S*) et les résidus sont donnés avec les distances correspondantes.



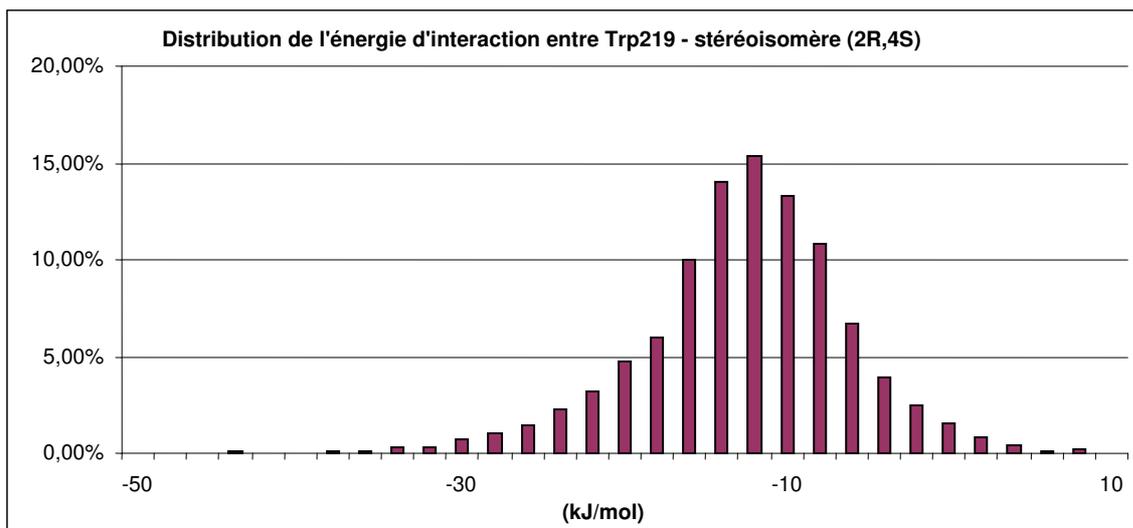
(a)



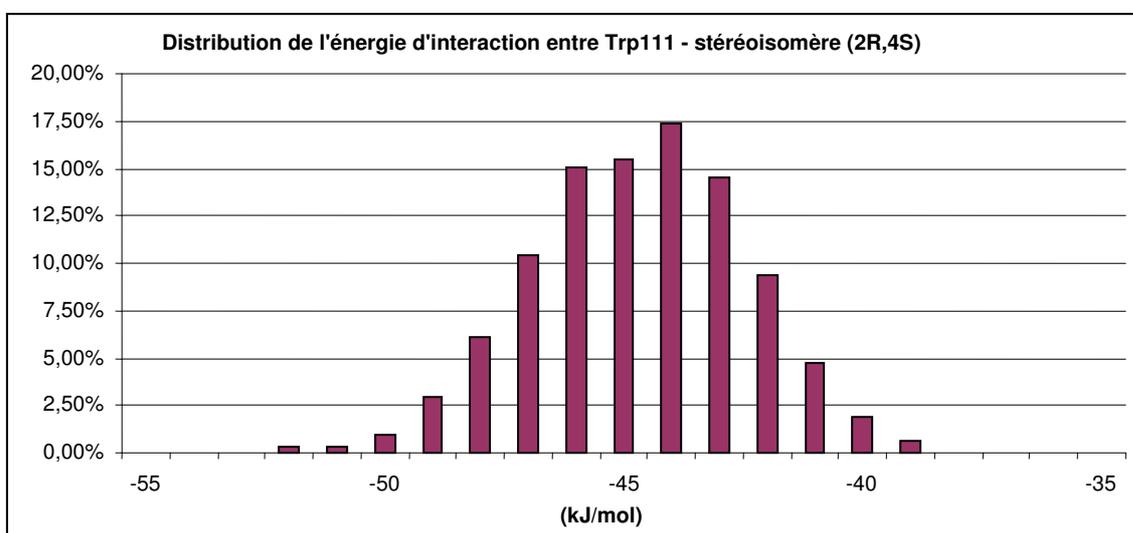
(b)



(c)



(d)



(e)

Figure 38 : Histogrammes des populations des énergies d'interaction électrostatique significatives au sein du site actif du complexe aldose réductase humaine – stéréoisomère (2R,4S) du fidarestat (en $\text{kJ}\cdot\text{mol}^{-1}$).

IV.5. Discussions et perspectives

Les résultats obtenus au cours de cette étude sont les premières tentatives d'estimation de l'incertitude statistique sur les énergies d'interaction électrostatique entre molécules ou groupements chimiques dans des systèmes macromoléculaires.

Les études statistiques menées sur quatre complexes enzyme – inhibiteur ont permis de mettre en évidence la nécessité de faire preuve de prudence dans l'utilisation des structures cristallographiques issues de banques de structures. En effet, dans la structure du complexe aldose réductase porcine – sorbinil, l'inhibiteur s'est révélé mal placé dans le site actif. Cela a affecté les liaisons hydrogène dans lesquelles cet inhibiteur est impliqué au sein du site actif de l'enzyme aldose réductase, et, par la même, les énergies d'interaction électrostatique correspondantes. Cela peut conduire à une mésinterprétation du mode de fixation de l'inhibiteur.

L'étude statistique sur les énergies d'interaction réalisée sur le complexe aldose réductase – fidarestat, structure à 0,98 Å, tend à montrer qu'utiliser des structures obtenues à une résolution atomique légitimise le recours au principe de transférabilité des paramètres multipolaires. En effet, les incertitudes statistiques sur les énergies d'interaction électrostatique demeurent de l'ordre de la précision numérique de la méthode de calcul utilisée (partie II.2.2.). La précision apportée par la définition d'un modèle de distribution de charges continue asphérique pourra, par conséquent, être interprétée.

Les études des complexes : aldose réductase porcine – sorbinil (2,30 Å) et aldéhyde réductase porcine – fidarestat (1,85 Å), révèlent une corrélation entre les amplitudes des énergies d'interaction et leurs incertitudes (écarts-types ou largeurs de distribution). Cette corrélation peut également s'interpréter en termes de distances interatomiques minimales. Plus les groupements chimiques en interaction impliquent des atomes proches, plus l'incertitude statistique sur l'énergie d'interaction électrostatique semble sensible aux incertitudes statistiques de leurs coordonnées.

Les écarts-types des distributions statistiques des coordonnées atomiques ne sont que de l'ordre de 0,01 Å ou 0,02 Å ce qui paraît négligeable comparé à la longueur d'une liaison covalente H-C_{arom} (1,083 Å) par exemple. Néanmoins, les répercussions de ces incertitudes

statistiques sur celles des énergies d'interaction électrostatique peuvent être significatives en particulier dans le cas des liaisons hydrogène. Les écarts-types des distributions d'énergies ne sont que de quelque kJ.mol^{-1} , mais leurs largeurs de distribution peuvent être de 10 à 20 kJ.mol^{-1} .

Les conformations utilisées dans cette étude étant considérées acceptables car proches les unes des autres compte tenu des écarts-types sur leurs coordonnées atomiques, elles peuvent être potentiellement obtenues lors d'un affinement de cette structure. Pour cette raison, l'incertitude statistique sur les énergies d'interaction peut être interprétée en termes de largeurs de distribution au lieu d'écarts-types. Pour les interactions électrostatiques significatives, la largeur de la distribution d'énergies peut correspondre à 20-30 % de la valeur absolue de la moyenne correspondante.

Afin de déterminer dans quelle mesure l'information supplémentaire apportée sur les énergies d'interaction électrostatique par un modèle de distribution de charges continue asphérique est masquée par l'incertitude statistique, il serait intéressant de réaliser une analyse similaire sur les énergies d'interaction électrostatique obtenues à partir de modèles de distribution de charges plus rudimentaires tels que des charges ponctuelles (AMBER) [38] ou des distributions de charges sphériques obtenues par affinement Kappa contre des jeux de données théoriques [5]. Les résultats de cette analyse pourraient être comparés à ceux déjà obtenus.

Le dernier complexe enzyme – inhibiteur étudié : aldose réductase – stéréoisomère (2*R*,4*S*) du fidarestat, est un cas particulier qui a permis de montrer les effets d'une agitation thermique élevée et du désordre sur les incertitudes statistiques des énergies. Bien que la résolution à laquelle la structure cristallographique du complexe a été résolue soit de 1,4 Å, les incertitudes statistiques sur les énergies d'interaction électrostatique sont nettement supérieures à celles obtenues avec la structure du complexe aldose réductase porcine – sorbinil pourtant résolue à une résolution moindre de 2,30 Å. Un tel résultat s'explique par l'incertitude statistique élevée des coordonnées atomiques de l'inhibiteur dans le site actif. Ces derniers présentent des paramètres d'agitation thermique de valeur importante, mais aussi des paramètres d'occupation seulement partiels. Ainsi, bien que la résolution des données de diffraction X ait été retenue comme facteur principal dans cette étude, les résultats montrent aussi que le désordre et l'agitation thermique constituent des facteurs déterminants qui peuvent nuire aux incertitudes statistiques sur les coordonnées atomiques.

La méthode présentée dans ce chapitre peut être améliorée sur de nombreux points.

Pour des raisons pratiques, les études menées sur les complexes macromoléculaires ont consistées à se focaliser sur leurs sites actifs tout en perturbant malgré cela l'ensemble de la structure. La perturbation des coordonnées de l'ensemble de la structure (excepté les atomes d'oxygène des molécules d'eau du solvant modélisées explicitement) avait pour but de prendre en compte l'ensemble des contributions possibles aux incertitudes statistiques sur les coordonnées des atomes du site actif. Par soucis de cohérence, la recherche et l'exclusion des « outliers conformationnels » pourrait être étendue à l'ensemble de la structure ou tout du moins aux atomes ne présentant pas des valeurs de paramètres d'agitation thermique trop élevées ou de désordre. Peu difficile à mettre en œuvre, les programmes d'exclusion des « outliers » ayant été développés au cours de ce travail, cette généralisation nécessiterait cependant un plus large jeu de conformations (le nombre d' « outliers conformationnels » augmentant avec le nombre de coordonnées atomiques prises en compte).

Les paramètres de déplacements atomiques du modèle sont corrélés à l'incertitude statistique sur les coordonnées atomiques. Or, les paramètres de déplacements atomiques isotropes (U_{iso}) des atomes n'ont pas été examinés au cours de cette étude. De plus, malgré le fait qu'ils aient été perturbés au cours de la procédure Perturbation / Affinement, leur amplitude maximale de perturbation n'a pas été discutée. Dans l'ensemble des tests menés ici, l'amplitude maximale de perturbation a été définie relativement à la valeur du paramètre considéré, et fixée à 5 %. L'analyse statistique pourrait être étendue aux paramètres d'agitation thermique et des tests sur l'amplitude maximale de perturbation réalisés.

Une corrélation entre les paramètres d'agitation thermique et les écarts-types sur les coordonnées atomiques est légitime. Des atomes agités voient leurs facteurs de diffusion atomique diminuer. Ils contribuent donc moins aux réflexions et sont, par conséquent, plus difficiles à localiser au cours de l'affinement. Cependant, cela peut également remettre en question le choix d'une perturbation relative pour ces paramètres. Une perturbation absolue pourrait être envisagée.

L'étude menée dans ce dernier chapitre a permis de proposer une première approche permettant d'obtenir une incertitude statistique sur une énergie d'interaction électrostatique. Lorsque la convergence d'une structure est obtenue par affinement, il est d'usage de déterminer les incertitudes expérimentales (ou écarts-types) sur les coordonnées atomiques à partir de la matrice de variance/covariance calculée par le programme d'affinement (ici,

SHELXL) [35]. Les incertitudes statistiques obtenues au cours de ce travail sont-elles équivalentes ou peuvent-elles être mises en parallèle avec les incertitudes expérimentales calculées ?

Les incertitudes statistiques et expérimentales calculées ne sont en aucun cas égales.

Pour exemple, dans le cas du test n°3 ($PerturMax = 0,1 \text{ \AA}$ et structure initiale corrigée) réalisé sur la structure du complexe aldose réductase – sorbinil (partie IV.3.6.C), l'incertitude statistique moyenne obtenue est de $0,015 \text{ \AA}$ alors que l'incertitude expérimentale calculée à partir de la structure initiale corrigée et en prenant en compte les restrictions et contraintes d'affinement, est de $0,071 \text{ \AA}$. L'incertitude statistique moyenne est donc nettement inférieure.

La représentation en nuage de points des incertitudes statistiques des coordonnées atomiques en fonction des incertitudes expérimentales calculées correspondantes suggère une corrélation qui demeure malgré tout grossière (**Figure 39**).

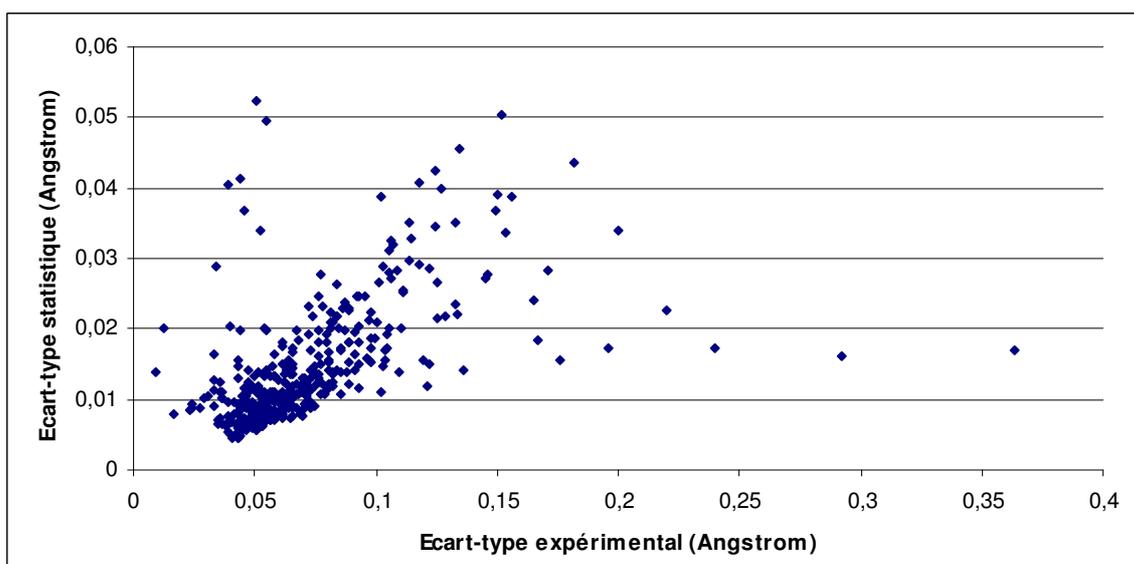


Figure 39 : Représentation en nuage de points des couples d'écart-types (σ_{exp} ; σ_{exp}) pour les atomes non hydrogène

De tels résultats sont tout à fait légitimes. Il est important de comprendre ce que représentent les incertitudes statistiques étudiées dans ce dernier chapitre. Ces incertitudes caractérisent des distributions de valeurs de coordonnées atomiques issues de modèles structuraux de complexes macromoléculaires obtenus par affinement cristallographique par méthode des moindres carrés. Chacune de ces structures a été considérée comme un minimum acceptable

du χ^2 . Pour un système donné, les conformations utilisées pour l'étude statistique constituent un ensemble de minima potentiels d'un point de vue algorithmique. Les incertitudes statistiques représentent donc une incertitude sur le minimum du χ^2 obtenu par le programme. Si l'on fait l'hypothèse qu'au voisinage du minimum de la fonction χ^2 , cette dernière peut être approximée au 2^{ème} ordre par un puit harmonique, et suit alors une loi normale multivariée. Les incertitudes expérimentales calculées par le programme d'affinement correspondent aux écarts-types de la loi normale multivariée, alors que les incertitudes statistiques correspondent aux incertitudes sur la moyenne (ou le mode) de cette même loi. Il est ainsi légitime d'obtenir des incertitudes statistiques plus petites que les incertitudes expérimentales calculées.

Plus concrètement, ces incertitudes statistiques sur les distributions des coordonnées atomiques ainsi que sur les énergies d'interaction électrostatique caractérisent « le pouvoir de résolution » du programme d'affinement (déjà suggéré dans la partie IV.1). Les incertitudes statistiques sur les énergies d'interaction électrostatique sont donc l'erreur qu'on peut commettre sur ces énergies en considérant la structure du système obtenue par affinement comme un minimum global parfait (qui correspond à la moyenne de la distribution normale multivariée), et représentative du système étudié.

- [1] *The SHELX-97 Manual*.
- [2] E. H. L. Aarts & V. Laarhoven. *Philips Journal of Research*, **40**(4), 193–226, 1985.
- [3] O. A. Barski, K. H. Gabbay, K. M. Bohren. *Biochemistry*, **35**, 14276–14280, 1996.
- [4] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. *Nucleic Acids Research*, **28**, 235–242, 2000.
- [5] P. Coppens, T. N. Guru Row, P. Leung, E. D. Stevens, P. J. Becker, Y. W. Yang. *Acta Crystallographica*, **A35**, 63–72, 1979.
- [6] G. R. Desiraju & T. Steiner. *The weak hydrogen bond in structural chemistry and biology*. Oxford University Press, 1999.
- [7] B. Dittrich, C. B. Hübschle, P. Luger, M. A. Spackman. *Acta Crystallographica*, **D62**, 1325–1335, 2006.
- [8] B. L. Dixit, G. K. Balendiran, S. J. Watowich, S. Srivastava, K. V. Ramana, J.M. Petrash, A. Bhatnagar, S. K. Srivastava. *Journal of Biological Chemistry*, **275**, 21587–21595, 2000.
- [9] P. M. Dominiak, A. Volkov, A. P. Dominiak, K. N. Jarzembska, P. Coppens. *Acta Crystallographica*, **D65**, 485–499, 2009.
- [10] O. El-Kabbani, V. Carbone, C. Darmanin, M. Oka, A. Mitschler, A. Podjarny, C. Schulze-Briese, R. P.-T. Chung. *Journal of Medicinal Chemistry*, **48**, 5536–5542, 2005.
- [11] O. El-Kabbani, V. Carbone, C. Darmanin, M. Oka, A. Mitschler, A. Podjarny, C. Schulze-Briese, R. P. Chung. *Journal of Medicinal Chemistry*, **48**, 5536–5542, 2005.
- [12] O. El-Kabbani, M. H. McGowan, D.A. Carper, Y. Devedjiev, K. J. Rees-Milton, T. G. Flynn. *Proteins: Structure, Function, Genetics*, **29**, 186–192, 1997.
- [13] O. El-Kabbani, C. Darmanin, T. R. Schneider, I. Hazemann, F. Ruiz, A. Oka, M. Joachimiak, C. Schulze-Briese, T. Tomizaki, A. Mitschler, A. Podjarny. *Protein*, **55**, 805–813, 2004.
- [14] O. El-Kabbani, S. E. Old, S. L. Ginell, D. A. Carper. *Molecular Vision*, **5**, 20–26, 1999.
- [15] O. El-Kabbani & A. Podjarny. *Cellular and Molecular Life Sciences*, **64**, 1970–1978, 2007.
- [16] B. Fournier, E-E Bendeif, B. Guillot, A. Podjarny, C. Lecomte, C. Jelsch. *Journal of American Chemistry Society*, **131**(31), 10929–10941, 2009.
- [17] B. Guillot, C. Jelsch, A. Podjarny, C. Lecomte. *Acta Crystallographica*, **D64**, 567–588, 2008.
- [18] D. N. Gujarati. *Basic Econometrics*.
- [19] N. K. Hansen & P Coppens. *Acta Crystallographica*, **A34**, 909–921, 1978.
- [20] C. Jelsch. *Acta Crystallographica*, **A57**, 558–570, 2001.
- [21] C. Jelsch, B Guillot, A Lagoutte, C Lecomte. *Journal of Applied Crystallography*, **38**, 38–54, 2005.
- [22] C. Jelsch, M. M. Teeter, V. Lamzin, V. Pichon-Pesme, R. H. Blessing, C. Lecomte. *Proceedings of the National Academy of Sciences (USA)*, **97**, 3171–3176, 2000.
- [23] L. Joubert & P. L. A. Popelier. *Molecular Physics*, **100**, 3357, 2002.

- [24] L. Joubert & P. L. A. Popelier. *Physical Chemistry Chemical Physics*, **4**, 4353, 2002.
- [25] D. Kim, S. I. Hong, D. S. Lee. *International Journal Molecular Sciences*, **7**, 519–536, 2006.
- [26] T. S. Koritsanszky, A. Volkov, P. Coppens. *Acta Crystallographica*, **A58**, 464–472, 2002.
- [27] X. Li, A.V. Volkov, K. Szalewicz, P. Coppens. *Acta Crystallographica*, **D62**, 639–647, 2006.
- [28] D. Liebschner, M. Elias, S. Moniot, B. Fournier, K. Scott, C. Jelsch, B. Guillot, C. Lecomte, E. Chabrière. *Journal of American Chemistry Society*, **131** (22), 7879–7886, 2009.
- [29] O. El-Kabbani, C. Darmanin, M. Oka, C. Schulze-Briese, T. Tomizaki, I. Hazemann, A. Mitschler, A. Podjarny. *Journal of Medicinal Chemistry*, **47**, 4530–4537, 2004.
- [30] V. Pichon-Pesme, C. Lecomte, H. Lachekar. *Journal of Physical Chemistry*, **99**, 6242–6250, 1995.
- [31] P. L. A. Popelier, L. Joubert, D. S. Kosov. *Journal of Physical Chemistry A*, **105**, 8254–8261, 2001.
- [32] H. Sawada, A. Hara, T. Nakayama, M. Hayashibara. *Journal of Biochemistry (Tokyo)*, **92**, 185–191, 1982.
- [33] H. Sawada, A. Hara, T. Nakayama, F. Kato. *Journal of Biochemistry (Tokyo)*, **87**, 1153–1165, 1980.
- [34] A. W. Schuettelkopf & D. M. F. van Aalten. *Acta Crystallographica*, **D60**, 1355–1363, 2004.
- [35] G.M. Sheldrick. *Acta Crystallographica*, **D64**, 112–122, 2008.
- [36] A. Urzhumtsev, F. Tête-Favier, A. Mitschler, J. Barbanton, P. Barth, L. Urzhumtseva, J.-F. Biellmann, A.D. Podjarny, D. Moras. *Structure*, **5**, 601–612, 1997.
- [37] A. Volkov & P. Coppens. *Journal of Computational Chemistry*, **25**, 921–934, 2004.
- [38] J. M. Wang, P. Cieplak, and P. A. Kollman. *Journal of Computational Chemistry*, **21**, 1049–1074, 2000.
- [39] B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte, C. Jelsch. *Acta Crystallographica*, **A63**, 108–125, 2007.

Conclusion & Perspectives

Les travaux présentés dans ce manuscrit s'intègrent dans une volonté d'étendre les outils et méthodes utilisées usuellement dans l'étude des petites molécules aux systèmes macromoléculaires.

Comme cela a été expliqué (chapitre 2) et illustré (chapitre 4), pour prétendre aux mêmes types d'études sur des systèmes macromoléculaires que sur des petites molécules, la précision expérimentale sur des macrosystèmes doit être transcendée afin d'accéder à un niveau de détail justifiant l'utilisation de certaines méthodes/outils (affinement à l'aide d'un modèle de charges non-sphérique, analyse topologique A.I.M., calculs de propriétés électrostatiques à partir de modèle non-sphérique de charges,...)

La transférabilité des paramètres multipolaires dans le formalisme de Hansen & Coppens [9] est une solution efficace pour obtenir un modèle estimé de distribution de charges non-sphérique. Cependant, il est nécessaire de demeurer critique dans l'utilisation de la transférabilité de paramètres de distribution de charges. Bien que ces paramètres constituent des informations supplémentaires sur la nature du complexe modélisé en termes de distribution de charges, elles n'améliorent en rien la véracité de sa structure atomique à une résolution modérée. Les incertitudes sur les paramètres structuraux ne peuvent, par conséquent, être négligées car elles peuvent affecter les propriétés électrostatiques dérivées des modèles de distribution de charges comme cela a été discuté dans le chapitre 4 de ce manuscrit.

Les calculs d'énergies d'interaction électrostatique réalisés au sein des sites actifs des complexes enzyme-inhibiteur restent préliminaires et nécessitent de poursuivre les améliorations techniques et méthodologiques. L'étude préliminaire présentée dans ce manuscrit n'aborde que l'un des aspects à prendre en considération pour éprouver l'estimation de propriétés électrostatiques de systèmes macromoléculaires à partir de modèles estimés de distribution de charges. En effet, ce travail a été mené dans le but de discuter la légitimité du recours à la transférabilité pour le calcul des énergies d'interaction électrostatique au sein des complexes protéine – inhibiteur, en terme de précision. Deux études [6;4] ont récemment été publiées dans lesquelles des modèles de distribution de charges définis par transfert sont utilisés pour l'estimation des énergies d'interaction électrostatique protéine – inhibiteur. Elles tentent de mettre en évidence un accord, voir une corrélation, entre les propriétés physico-chimiques de complexes protéine-inhibiteurs et les énergies correspondantes, mais cela sans aborder le problème de la précision de ces énergies.

Malgré le fait que la méthode proposée dans le chapitre 4 ne renseigne pas sur la précision expérimentale des énergies d'interaction, elle permet d'estimer la précision sur ces énergies lorsque la structure atomique statique obtenue par affinement X est supposée être représentative du système étudié. En d'autres mots, les précisions statiques sur les énergies caractérisent la capacité du programme d'affinement à fournir une conformation finale unique correspondant au minimum global du paysage énergétique, ce que l'on pourrait appeler son « pouvoir de résolution » (par analogie avec les instruments optiques).

Cette approche peut néanmoins faire l'objet d'améliorations méthodologiques, comme cela a déjà été discuté. Le résultat d'un affinement étant conditionné par un ensemble de choix imposés ou non aux utilisateurs, cette étude pourrait être menée avec un autre logiciel d'affinement cristallographique, afin de déterminer dans quelle mesure les choix algorithmiques, d'implémentation et de paramétrisation des modèles affectent les distributions d'énergies d'interaction électrostatique. Les résultats des tests présentés dans le chapitre 4 de ce manuscrit ont été réalisés à l'aide du programme SHELX [13] qui utilise les « moindres carrés » comme méthode d'affinement, en considérant des structures affinées à différentes résolutions en dessous de la limite d'utilisation reconnue par ses développeurs (2,5 Å). On notera malgré cela que, dans le cadre des affinements à moyenne résolution de structures macromoléculaires, la méthode du « maximum de vraisemblance » est couramment employée et reconnue pour ses performances. Les mêmes analyses statistiques pourraient être menées en ayant recourt à un programme utilisant le « maximum de vraisemblance », par exemple REFMAC [15], BUSTER [2] ou PHENIX [1].

D'autres structures obtenues à différentes résolutions pourraient faire l'objet d'analyses statiques supplémentaires afin de mettre en évidence des corrélations entre la précision des énergies d'interaction dérivées et la résolution cristallographique. L'objectif étant d'éprouver la transférabilité des paramètres multipolaires en terme de précision, réaliser le même type d'analyse statistique à l'aide d'autres banques de paramètres de charges plus rudimentaires tels que des modèles de distribution de charges d'atomes sphériques ou ponctuelles permettrait de discuter de l'utilité des paramètres de distributions de charges asphériques compte tenu de la précision des énergies dérivées. Existe-t-il en particulier une résolution limite à laquelle les précisions des énergies d'interaction sont suffisamment pénalisantes pour masquer les détails apportés par un modèle asphérique de charges ?

L'utilisation de la banque ELMAM [11;16] a également été un facteur contraignant dans ce travail. Cette banque est expérimentale et, pour cette raison, difficile à étendre tant son développement dépend des jeux de données rayons X à très haute résolution disponibles. L'étude d'autres complexes macromoléculaires ne peut donc être envisagée sans la définition de nouveaux types atomiques permettant la prise en charge, lors d'un transfert de paramètres, de molécules atypiques. L'étude du fidarestat, un inhibiteur de l'enzyme aldose réductase, présenté dans le chapitre 2 de ce manuscrit, a servi notamment à son amélioration [16]. Cet inhibiteur existe dans l'environnement biologique sous deux formes : l'une neutre, l'autre chargée négativement (par déprotonation du groupement hydantoin). Des études ont montré que l'inhibiteur fidarestat pouvait se complexer avec l'aldose réductase sous ces deux formes chargées [5]. Etudier la distribution de charges de la forme déprotonée du fidarestat permettrait non seulement de comparer ces deux formes mais aussi d'utiliser ce modèle de charges pour définir de nouveaux types pseudo-atomiques indispensables à l'étude des complexes aldo-keto reductase – fidarestat chargé.

Les énergies d'interaction calculées dans les analyses préliminaires ne sont que les contributions électrostatiques. Bien que les interactions intermoléculaires au sein des complexes biomoléculaires impliquent généralement un réseau de liaisons hydrogène, les interactions peuvent reposer sur des contacts courts, par exemple, entre cycles aromatiques. Ces interactions ne peuvent être modélisées correctement sans la prise en compte des autres contributions à l'interaction totale telles que les énergies liées à la dispersion et à l'exclusion de Pauli. Une expression de l'énergie d'interaction totale permettrait une analyse plus précise des interactions et notamment de discuter leur nature. Cependant, disposer des modèles et paramètres pour ces énergies complémentaires au terme d'énergie électrostatique défini à partir du formalisme de Hansen & Coppens nécessite recherche et réflexion. Deux possibilités semblent intéressantes : les paramètres employés dans le champ de force AMOEBA [12] ou le formalisme défini par A. Gavezzotti dans la méthode PIXEL [7;8]. Cependant, dans les deux cas, les distributions de charges ne sont pas modélisées suivant le formalisme de Hansen & Coppens et de plus sont obtenues par calculs théoriques, ce qui laisse présager que l'utilisation de ces modèles d'énergie d'interaction ne sera pas aisée.

Sans débattre la question de la précision d'un calcul d'énergie à partir d'un modèle statique de distribution de charges, la méthode demeure discutable également en terme de justesse ou « reliability » en anglais. Rappelons que cette dernière repose sur l'utilisation d'un modèle I.A.M. statique obtenu par affinement cristallographique RX comme « squelette » pour un

modèle de distribution de charges estimé, ce qui ne permet pas de prendre en compte le comportement thermodynamique du système. Les structures atomiques issues des études par diffraction X sont obtenues en déconvoluant les effets moyens de l'agitation thermique et la distribution de charges dont est déduite la structure atomique dite « statique ». Cette structure atomique statique est supposée être la structure « à l'équilibre » du système macromoléculaire étudié. Ce statut est purement fictif tant les modèles d'agitation thermique employés peuvent être rudimentaires et mis en défaut en particulier dans les parties désordonnées du système à des résolutions modérées. Tant bien même une structure atomique obtenue par affinement RX puisse être considérée comme la structure « à l'équilibre » du système macromoléculaire, les effets de l'agitation thermique ne sont pas pris en compte dans le calcul des énergies d'interaction électrostatique dérivées.

La démonstration de la justesse de tels calculs comme approximation de grandeurs physiques ne peut être réalisée sans la mise en évidence de corrélations avec différents types de mesures expérimentales telles que les valeurs V_{c50} (obtenues par spectrométrie de masse supramoléculaire), CI_{50} (concentration d'inhibiteur requise pour une inhibition à 50% *in vitro* d'un mécanisme enzymatique), pK_a (Constantes d'affinité obtenues par calorimétrie à titrage isotherme),... Ces mesures expérimentales caractérisent l'énergie d'interaction entre la protéine et l'inhibiteur implicitement sous l'effet de l'agitation thermique. Il est d'ailleurs important de noter que, pour des raisons purement mathématiques, une énergie d'interaction calculée en considérant la structure supposée « à l'équilibre » ne correspondra pas nécessairement à l'énergie d'interaction moyennée en temps en considérant les différentes conformations prises par le système au cours du temps sous l'effet de l'agitation thermique. L'omission des effets de l'agitation thermique biaise donc l'estimation des énergies d'interaction. Afin de se rapprocher de la réalité expérimentale, il conviendrait de rétablir les effets de l'agitation thermique. Ce point n'est pas trivial : les modèles d'agitation thermique utilisés dans le modèle I.A.M. renseignent sur l'amplitude des déplacements thermiques des atomes mais en aucun cas ne fournissent d'informations sur leurs corrélations, ce qui ne permet donc pas en se fondant exclusivement sur les données diffraction X d'obtenir les énergies d'interaction électrostatique moyennées en temps. Les structures atomiques statiques issues des affinements RX présentent d'autres limites liées au système qu'elles représentent. En effet, les structures cristallographiques de complexes protéine-inhibiteur correspondent par définition aux conformations dans un environnement cristallin. En raison de l'empilement cristallin, du solvant employé lors de la cristallisation de ces complexes et des conditions expérimentales d'une expérience de diffraction RX sur monocristal (notamment la basse

température), ces structures ne sont pas nécessairement représentatives des conformations prises par ce même complexe *in vivo* ou en solution. Ces contraintes nous encouragent à considérer les énergies d'interaction électrostatique dérivées des modèles estimés de distribution de charges statiques avec prudence. Néanmoins, la possibilité de corrélation entre ces énergies et des mesures expérimentales peut être examinée. Dans le cas de l'étude des complexes aldo-keto réductases – inhibiteurs, les structures d'autres complexes sont justement disponibles ainsi que des données Vc50 de spectrométrie de masse, CI50,...[3;14]

L'utilisation des structures cristallographiques comme support à la conception de modèles de distribution de charges n'est donc pas exempt de faiblesses. D'autres applications pour la transférabilité pourraient être envisagées afin de pallier ces limites. L'analyse des modes normaux tout atomes d'un système macromoléculaire [10] permet de simuler sa flexibilité propre au voisinage de sa conformation d'énergie minimale. La dynamique moléculaire quant à elle offre la possibilité de simuler les fluctuations structurales dues à l'agitation thermique à la température *in vivo*. L'ensemble de conformations générées sur la base des modes normaux principaux de la structure ou correspondant aux différents états du système macromoléculaire au cours de la trajectoire simulée par dynamique moléculaire pourraient être utilisées pour la définition de modèles de distribution de charges, et les énergies d'interaction électrostatique correspondantes être calculées, puis moyennées. Cependant, reproduire les fluctuations structurales dues à l'agitation thermique doit être réalisé judicieusement, l'analyse des modes normaux et la dynamique moléculaire reposant sur d'autres modèles d'énergie d'interaction interatomique : potentiels d'interaction ou champs de force.

En pratique, la transférabilité des paramètres multipolaires pourrait être envisagée comme information supplémentaire lors de dynamiques moléculaires réalisées à l'aide de champs de force utilisant des charges ponctuelles dans le cadre d'études de complexes protéine-inhibiteur. Si ces simulations mettent en évidence plusieurs modes de fixation pour un inhibiteur ou ne permettent pas de départager des ligands potentiels présentant *in silico* des affinités comparables, les énergies d'interaction dérivées de modèles de charges asphériques transférés pourraient permettre de discriminer ces différentes possibilités.

- [1] P.D. Adams, P.V. Afonine, G. Bunkoczi, 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, P.H. Zwart. *Acta Crystallographica*, **D66**, 213–221, 2010.
- [2] E. Blanc, P. Roversi, C. Vornrhein, C. Flensburg, S.M. Lea, G. Bricogne. *Acta Crystallographica*, **D60**, 2210–2221, 2004.
- [3] C. Darmanin, G. Chevreux, N. Potier, A.V. Dorselaer, I. Hazemann, A. Podjarny, O. El-Kabbani. *Bioorganic & Medicinal Chemistry*, **12**, 3797–3806, 2004.
- [4] P. M. Dominiak, A. Volkov, A. P. Dominiak, K. N. Jarzemska, P. Coppens. *Acta Crystallographica*, **D65**, 485–499, 2009.
- [5] O. El-Kabbani, C. Darmanin, T. R. Schneider, I. Hazemann, F. Ruiz, A. Oka, M.;Joachimiak, C. Schulze-Briese, T. Tomizaki, A. Mitschler, A. Podjarny. *Protein*, **55**, 805–813, 2004.
- [6] B. Fournier, E-E Bendeif, B. Guillot, A. Podjarny, C. Lecomte, C. Jelsch. *Journal of American Chemistry Society*, **131**(31), 10929–10941, 2009.
- [7] A. Gavezzotti. *Crystal Engineering Communications*, **5**(76), 429–438, 2003.
- [8] A. Gavezzotti. *Crystal Engineering Communications*, **5**(77), 439–446, 2003.
- [9] N K Hansen & P Coppens. *Acta Crystallographica*, **A34**, 909–921, 1978.
- [10] E. Lindahl, C. Azuara, P. Koehl, M. Delarue. *Nucleic Acids Research*, **34**, 2006.
- [11] V. Pichon-Pesme, C. Lecomte, H. Lachekar. *Journal of Physical Chemistry*, **99**, 6242–6250, 1995.
- [12] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, Jr. R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, T. Johnson, M. E. Head-Gordon. *Journal of Physical Chemistry B*, **114**, 2549–2564, 2010.
- [13] G.M. Sheldrick. *Acta Crystallographica*, **D64**, 112–122, 2008.
- [14] H. Steuber, A. Heine, A. Podjarny, G. Klebe. *Journal of Molecular Biology*, **379**, 991–1016, 2008.
- [15] M.D. Winn, G.N. Murshudov, M.Z. Papiz. *Method in Enzymology*, **374**, 300–321, 2003.
- [16] B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte, C. Jelsch. *Acta Crystallographica*, **A63**, 108–125, 2007.

ANNEXES

Annexe A : L'histoire et la théorie de la diffraction des rayons X par un monocristal

Annexe B : Informations supplémentaires sur l'approximation de Buckingham

Annexe C : Matériels supplémentaires sur l'étude de la distribution de charges de l'inhibiteur fidarestat

Annexe D : Exemple de résultat d'analyse statistique sur les coordonnées atomiques de l'inhibiteur sorbinil et du groupement nicotinamide du facteur NADP⁺ dans le complexe aldose réductase – sorbinil.

Annexe A : L'histoire et la théorie de la diffraction des rayons X par un monocristal

1 – Un peu d'histoire

Les rayons X furent découverts par hasard le 8 novembre 1895 par Wilhelm Conrad Röntgen au cours d'une expérience portant sur les rayons cathodiques. Suite à cela, Röntgen caractérisa au cours des trois semaines qui suivirent les propriétés de ces rayons par plusieurs expériences. Ses résultats furent publiés dans un article intitulé : « Über eine neue Art von Strahlen. Vorläufige Mitteilung. » (Sur une nouvelle sorte de rayons. Communication préliminaire).

Il appela ces rayons : « rayons X » car leur nature était alors inconnue. La propriété la plus marquante pour l'époque était qu'ils puissent traverser toutes sortes de corps.

À la suite de la première radiographie (celle de la main Bertha Röntgen, sa femme), les applications médicales des rayons X apparurent dans les semaines qui suivirent la parution de cet article. Wilhelm Conrad Röntgen reçut le premier prix Nobel de physique, en 1901.

La découverte des rayons X est une des découvertes les plus importantes des sciences. L'utilisation des rayons X pour l'étude de la structure des cristaux ne survint que plus tard avec une expérience réalisée à l'initiative de Max von Laue.

C'est en effet en 1912, alors qu'il enseignait à l'université de Munich et s'intéressait aux phénomènes d'interférences, que Von Laue comprit, à l'occasion de la soutenance de thèse de P. Ewald sur l'étude de l'interaction de la lumière et d'un cristal, que si un faisceau de rayons X rencontrait un cristal il devrait se produire un phénomène analogue à la diffraction de la lumière par un réseau.

La première expérience de diffraction des rayons X fut réalisée sur un cristal de blende ZnS (**Figure A1**). Cette expérience permit de fournir à la fois les preuves de la nature ondulatoire des rayons X et de la structure atomique des cristaux.

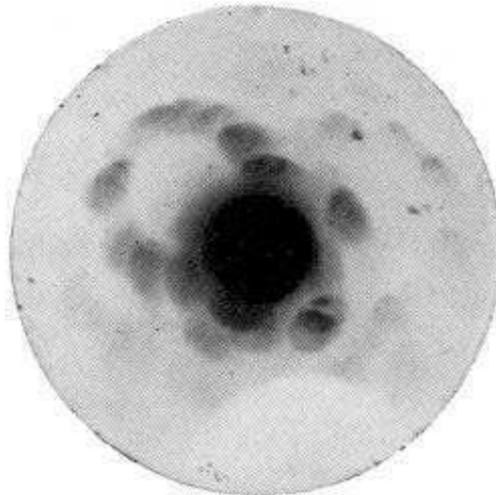


Figure A1 : Premier cliché de diffraction d'un cristal de blende ZnS

La diffraction des rayons X fut développée comme technique d'étude des cristaux par les Bragg père et fils. À Cambridge, constatant que les taches de diffraction se déplaçaient lorsqu'on bougeait le cristal - comme une réflexion sur un miroir, ils eurent l'idée de considérer la diffraction comme une réflexion sur des plans réticulaires. Une telle réflexion n'était possible qu'à la condition que la longueur d'onde, l'angle de diffraction et la distance inter-réticulaire satisfassent une relation devenue célèbre : la relation de Bragg. Cette idée permit d'interpréter les résultats de Laue et de montrer que la blende ZnS était cubique faces centrées. D'autres structures furent par la suite caractérisées : celles des sels KCl et NaCl, puis de ZnS, FeS₂, CaF₂ et CaCO₃.

Max von Laue eut le prix Nobel de physique en 1914, les Bragg père et fils l'année suivante.

2 – Pourquoi la diffraction des rayons X ?

La diffraction des rayons X permet d'obtenir des informations structurales sur des composés moléculaires via la détermination de leur distribution électronique sous forme cristalline. Elle peut être utilisée pour différents types de composés : les minéraux, les molécules organiques, organométalliques, dont la taille peut aller de quelques atomes à plusieurs milliers. Pour exemple, dans la RCSB protein databank, la macromolécule la plus grande déposée est la structure obtenue par diffraction des rayons X de la sous-unité catalytique de la protéine kinase humaine ADN-dépendante (code 3KGV) avec 20320 atomes non hydrogène à une résolution de 6,6 Å.

Comme toute méthode expérimentale, la diffraction des rayons X repose sur des propriétés physiques de la matière ce qui induit des conditions requises et par là même des limites d'applications.

3 – La clé de la diffraction des rayons X sur un monocristal

1) Le phénomène d'interaction particule matière.

La diffraction des rayons X par un monocristal est un phénomène d'interaction particule - matière. L'échantillon est en effet exposé à un faisceau de rayons X qui va interagir avec la distribution électronique du matériau. Le faisceau interagit également avec les protons des noyaux mais leur probabilité d'interaction est négligeable par rapport à celle des électrons (environ 1800 fois plus faible).

Les rayons X sont des ondes électromagnétiques et peuvent, comme toute onde électromagnétique, interagir avec la matière selon deux phénomènes : la diffusion et l'absorption (**Figure A2**).

Le phénomène de diffusion est la clé des expériences par diffraction des rayons X.

Ce phénomène a été interprété dans un premier temps grâce au formalisme de Thomson qui permet de calculer la diffusion par un électron quasi-libre.

A l'échelle subatomique, le champ électrique de l'onde électromagnétique met en oscillation l'électron qui réémet alors une onde électromagnétique en réponse à cette stimulation. Cette onde est dite diffusée par l'électron. Le formalisme de Thomson prévoit que le photon incident et le photon diffusé ont la même énergie, la diffusion de Thomson est dite élastique.

Dans le cas de la diffusion des rayons X par un atome, la perturbation du cortège électronique de l'atome par une onde électromagnétique peut conduire à deux types de diffusion. Lorsque l'onde électromagnétique diffusée a la même fréquence (et donc la même énergie) que l'onde incidente, cela correspond à la diffusion de Thomson.

Lorsque l'interaction entraîne la perte d'un électron faiblement lié et donc une altération du cortège électronique de l'atome, l'onde électromagnétique diffusée est de fréquence inférieure à l'onde incidente et donc d'énergie inférieure. La diffusion est alors dite inélastique appelée également diffusion de Compton.

L'autre phénomène possible est l'absorption qui entraîne une transition électronique. Si l'énergie apportée par le photon est suffisante il peut y avoir ionisation. L'ionisation se traduit par la libération d'un électron appelé photoélectron. C'est l'effet photoélectronique.

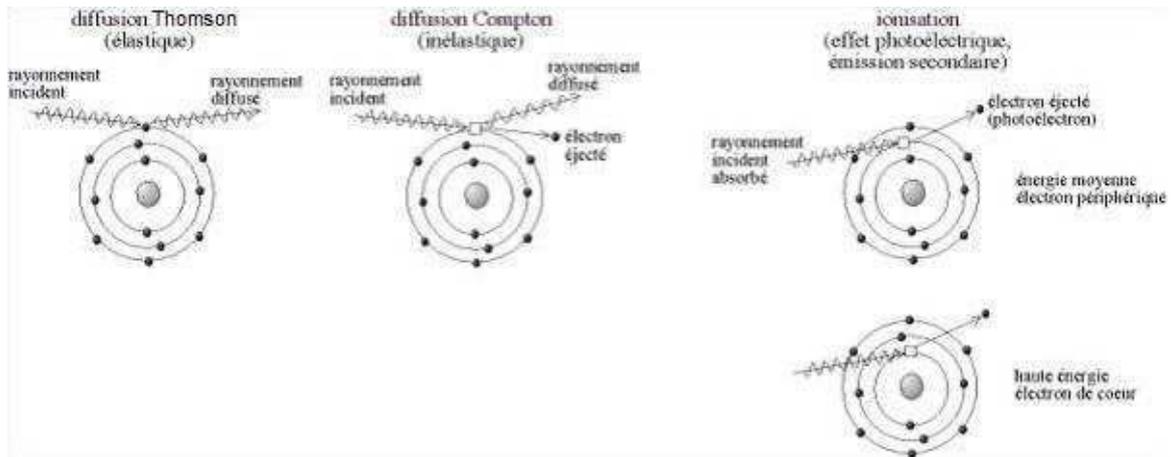


Figure A2 : Schéma des différents phénomènes d'interaction particule-matière.

Dans les expériences de diffraction des rayons X, la diffusion est essentiellement élastique mais peut cependant être inélastique selon les espèces chimiques rencontrées dans le matériau et la longueur d'onde des rayons X utilisés.

La démonstration de ce principe est faite sous certaines approximations connues sous le nom d'approximation cinématique :

- La diffusion est élastique (L'énergie du faisceau incident est préservée).
- Pas de diffraction multiple (un faisceau diffracté n'est pas rediffracté ultérieurement).

On considère également le détecteur ou « observateur » situé très loin par rapport à l'échantillon. La diffraction est alors dite « à l'infini ». Elle est également connue comme la diffraction de Fraunhofer.

Si l'on considère un électron isolé placé à l'origine de notre repère spatial et interagissant avec un faisceau de rayons X incident non polarisé, l'amplitude du faisceau diffusé sera donnée par :

$$E_e = -E_0 \frac{e^2}{4\pi\epsilon_0 m_e c^2} \frac{1}{R} \sqrt{\frac{1 + \cos^2 2\Theta}{2}} e^{i(\omega t - \mathbf{K} \cdot \mathbf{R})}$$

On pose E_0 , l'amplitude du faisceau incident,

m_e et e , respectivement la masse et la charge d'un électron,

c , la célérité de la lumière dans le vide,

ϵ_0 , la constante de permittivité électrique du vide,

\mathbf{R} , le vecteur distance entre l'origine et l'"observateur" (ici le détecteur),

ω , la pulsation et \mathbf{K} , le vecteur d'onde.

Si l'on considère r_e , le rayon d'un électron comme $r_e = \frac{e^2}{4\pi\epsilon_0 m_e c^2}$ et \mathbf{H} , le vecteur de

diffusion comme $\mathbf{H} = \frac{\mathbf{S} - \mathbf{S}_0}{\lambda}$ avec λ , la longueur d'onde et \mathbf{S} et \mathbf{S}_0 , respectivement les

vecteurs normaux de propagation des faisceaux diffusé et incident (**Figure A3**).

Si l'électron est localisé en \mathbf{r} (vecteur position défini dans le repère spatial), l'amplitude du rayon diffusé est définie comme :

$$E_e = -E_0 \frac{r_e}{R} e^{i(\omega t - \mathbf{K} \cdot \mathbf{R})} \sqrt{\frac{1 + \cos^2 2\Theta}{2}} e^{2\pi i(\mathbf{H} \cdot \mathbf{r})}$$

avec $2\pi\mathbf{H} \cdot \mathbf{r}$ la différence de phase entre les faisceaux diffusé à l'origine 0 du repère et diffusé en \mathbf{r} .

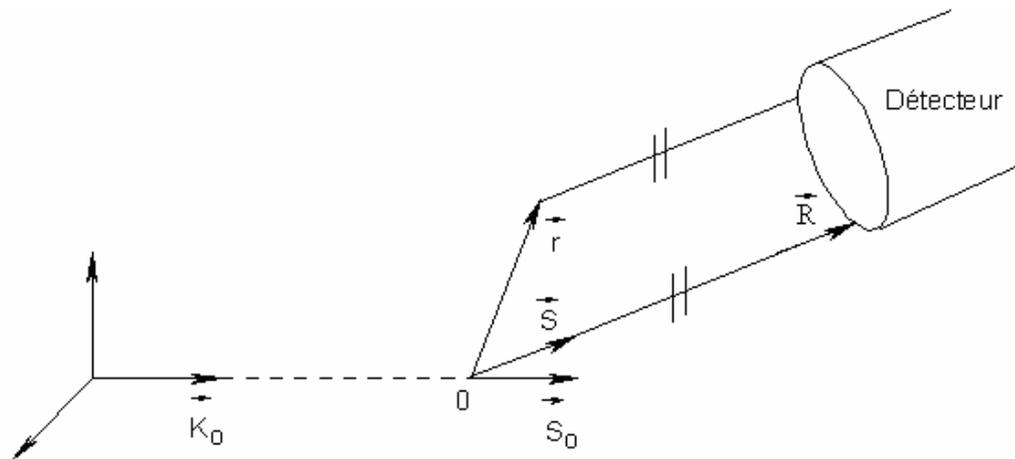


Figure A3 : Représentation schématique des vecteurs normaux de propagation et des vecteurs positions.

On considère maintenant arbitrairement un électron du $j^{\text{ème}}$ atome d'un échantillon, localisé en \mathbf{r}_j avec dans le repère spatial cartésien : $\mathbf{r}_j = x_j \mathbf{a} + y_j \mathbf{b} + z_j \mathbf{c}$ (**Figure A4**).

Loin de l'échantillon à une distance R , le $j^{\text{ème}}$ atome sera perçu par l'observateur comme une onde sphérique dont l'amplitude sera proportionnelle à la densité électronique de l'atome $\rho(\mathbf{r}_j)$.

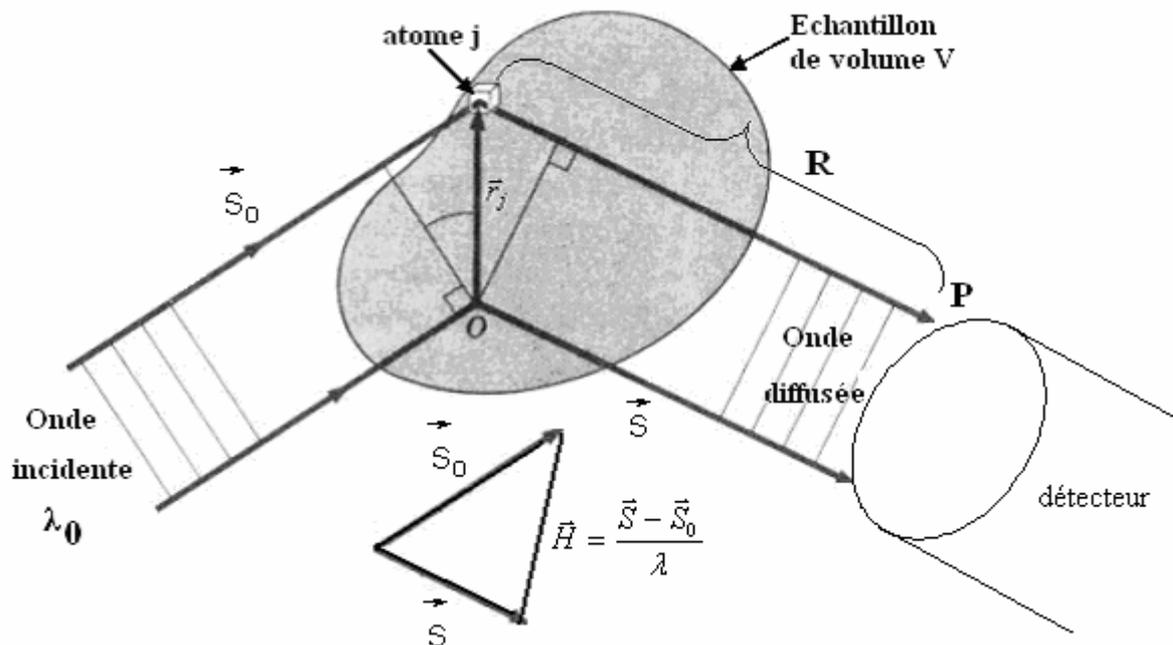


Figure A4 : Représentation schématique de la diffraction des rayons X par un échantillon.

Si l'on considère le cortège électronique de l'atome de volume V , l'amplitude du faisceau diffusé par cet atome dans la direction du vecteur de diffusion sera donnée par :

$$E_{at,j^{th}}(\mathbf{H}) = E_e(\mathbf{H}) \iiint_V \rho(\mathbf{r}) e^{2\pi i(\mathbf{H},\mathbf{r})} dV = E_e f_{j^{th}}(\mathbf{H})$$

(On néglige ici la différence de phase entre les positions des électrons au sein du cortège électronique de l'atome à un instant donné.)

$f_{j^{th}}$ est appelé "facteur de forme" ; il est la transformée de Fourier de la distribution électronique du $j^{ème}$ atome et prend en compte des interférences constructives entre les différents rayons X diffusés par les électrons de la distribution électronique de l'atome.

L'expression générale d'un facteur de forme atomique ou facteur de diffusion pour les rayons X $f(\mathbf{H})$ est : $f(\mathbf{H}, \omega) = f_0(\mathbf{H}) + \Delta f'(\mathbf{H}, \omega) + \Delta f''(\mathbf{H}, \omega)$

$\Delta f'$ et $\Delta f''$ sont appelés facteurs de dispersion anormale. $\Delta f''$ correspond à la portion absorbée de l'énergie de l'onde incidente lorsque l'interaction est inélastique.

Ces facteurs de diffusion et de dispersion anormale atomiques dépendent de l'espèce chimique et de l'état de ionisation de l'atome. Ils sont tabulés dans les tables internationales et peuvent être également calculés à partir des tables Clementi.

2) La périodicité de l'échantillon :

La diffraction des rayons X par un échantillon conduit à la diffusion dans toutes directions de rayons X qui bien que détectables ne peuvent être exploités car constituent sur la surface d'un détecteur un continuum de faisceau diffusés qui ne peuvent être distingués les uns des autres.

La diffraction des rayons X repose également sur le fait que les échantillons utilisés sont des cristaux et présentent par définition une périodicité spatiale (triple périodicité). Cette périodicité de l'échantillon affecte le schéma de diffraction et impose des directions préférentielles de diffusion.

Considérons dans un premier temps un cristal à une dimension. Cet échantillon sera donc composé d'un même motif appelé « maille » répété un grand nombre N de fois.

Dans cet échantillon périodique à une dimension, le $j^{\text{ème}}$ atome sera présent dans chaque maille de l'échantillon.

Le $j^{\text{ème}}$ atome contenu dans la $n^{\text{ème}}$ maille sera localisé en $\mathbf{r}_{nj} = (x_j + n)\mathbf{a}$.

A partir de l'expression de l'amplitude du faisceau diffuse par un atome communiqué précédemment, on peut déterminer l'amplitude du faisceau diffusé par l'ensemble des $j^{\text{ème}}$ atomes dans la direction du vecteur de diffusion comme étant :

$$E_{j^{\text{ème}}}(\mathbf{H}) = E_e(\mathbf{H}) \sum_{n=0}^{N-1} f_j \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_{jn})$$

L'amplitude du faisceau diffusé dans la direction du vecteur de diffusion par l'ensemble des atomes présent dans la maille de l'échantillon périodique sera la somme des contributions propres à chaque atome :

$$E(\mathbf{H}) = E_e(\mathbf{H}) \sum_{n=0}^{N-1} \sum_{j=1}^{n_{at}} f_j \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_{jn}) \text{ avec } n_a, \text{ le nombre d'atomes dans la maille.}$$

ou encore,

$$E(\mathbf{H}) = E_e(\mathbf{H}) \sum_{n=0}^{N-1} \sum_{j=1}^{n_{at}} f_j \exp(2\pi i \mathbf{H} \cdot (x_j + n)\mathbf{a})$$

Pour chaque contribution, on peut factoriser par l'exponentielle de la différence de phases entre l'origine du repère local de la $n^{\text{ème}}$ maille et la position du $j^{\text{ème}}$ atome dans cette même maille.

$$E(\mathbf{H}) = E_e(\mathbf{H}) \sum_{j=1}^{n_{at}} f_j \exp(2\pi i \mathbf{H} \cdot (x_j)\mathbf{a}) [1 + \exp(2\pi i \mathbf{H} \cdot \mathbf{a}) + \exp(2\pi i \mathbf{H} \cdot 2\mathbf{a}) + \dots + \exp(2\pi i \mathbf{H} \cdot (N-1)\mathbf{a})]$$

Le terme somme des exponentielles de différence de phases entre les origines des différentes mailles et l'origine du repère global de l'échantillon est commun à toutes les contributions atomiques et peut donc être factorisé.

$$E(\mathbf{H}) = E_e(\mathbf{H}) \sum_{j=1}^{n_{at}} f_j \exp(2\pi i \mathbf{H} \cdot (x_j)\mathbf{a}) \times \Gamma(\mathbf{H}) = E_e(\mathbf{H}) F(\mathbf{H}) \times \Gamma(\mathbf{H})$$

L'amplitude du faisceau diffusé peut donc s'exprimer comme le produit du facteur de structure global $F(\mathbf{H}) = \sum_{j=1}^{n_{at}} f_j \exp(2\pi i \mathbf{H} \cdot (x_j) \mathbf{a})$ et le terme $\Gamma(\mathbf{H})$ appelé « fonction d'interférence » du cristal.

$F(\mathbf{H})$, appelé facteur de structure, dépend seulement de la nature et des positions des atomes dans la maille. C'est l'amplitude de ce terme qui est mesurée expérimentalement. La connaissance de ce terme permet de connaître le motif.

$\Gamma(\mathbf{H})$ ne dépend pas de la composition de la maille mais uniquement de ces dimensions. Il permet de caractériser le réseau.

$$\Gamma(\mathbf{H}) = 1 + \exp(2\pi i \mathbf{H} \cdot \mathbf{a}) + \exp(2\pi i \mathbf{H} \cdot 2\mathbf{a}) + \dots + \exp(2\pi i \mathbf{H} \cdot (N-1)\mathbf{a}) = \sum_{n=0}^{N-1} \exp(2\pi i \mathbf{H} \cdot \mathbf{a})^n$$

$\Gamma(\mathbf{H})$ est la somme des termes d'une suite géométrique de raison $\exp(2\pi i \mathbf{H} \cdot \mathbf{a})$.

$$\sum_{n=0}^{N-1} \exp(2\pi i \mathbf{H} \cdot \mathbf{a})^n = \frac{1 - \exp(2\pi i \mathbf{H} \cdot \mathbf{a})^N}{1 - \exp(2\pi i \mathbf{H} \cdot \mathbf{a})}$$

Expérimentalement, on acquière l'intensité des faisceaux diffractés, l'intensité du faisceau diffracté devient :

$$I(\mathbf{H}) = E(\mathbf{H}) \times \overline{E(\mathbf{H})} = I_e(\mathbf{H}) \times (F(\mathbf{H}) \times \overline{F(\mathbf{H})}) \times (\Gamma(\mathbf{H}) \times \overline{\Gamma(\mathbf{H})})$$

$$\Gamma(\mathbf{H}) \times \overline{\Gamma(\mathbf{H})} = \left(\frac{1 - \exp(2\pi i \mathbf{H} \cdot \mathbf{a})^N}{1 - \exp(2\pi i \mathbf{H} \cdot \mathbf{a})} \right) \times \left(\frac{1 - \exp(-2\pi i \mathbf{H} \cdot \mathbf{a})^N}{1 - \exp(-2\pi i \mathbf{H} \cdot \mathbf{a})} \right) = \left(\frac{\exp(-\pi i \mathbf{H} \cdot \mathbf{a})^N - \exp(\pi i \mathbf{H} \cdot \mathbf{a})^N}{\exp(-\pi i \mathbf{H} \cdot \mathbf{a}) - \exp(\pi i \mathbf{H} \cdot \mathbf{a})} \right)^2$$

$$\Gamma(\mathbf{H}) \times \overline{\Gamma(\mathbf{H})} = \left(\frac{\sin(\pi N \mathbf{H} \cdot \mathbf{a})}{\sin(\pi \mathbf{H} \cdot \mathbf{a})} \right)^2$$

On en déduit que :

$$I(\vec{\mathbf{H}}) = I_e(\mathbf{H}) \times (F(\mathbf{H}) \times \overline{F(\mathbf{H})}) \times \left(\frac{\sin(\pi N \mathbf{H} \cdot \mathbf{a})}{\sin(\pi \mathbf{H} \cdot \mathbf{a})} \right)^2$$

Cette formule nous montre que la périodicité de l'échantillon implique que les intensités ne seront significatives que lorsque le dénominateur sera nul c'est-à-dire lorsque $\pi \mathbf{H} \cdot \mathbf{a} = k\pi$ avec k un entier. Les rayons diffusés interfèrent de manière constructive que dans ces directions spécifiques de l'espace.

Dans le cas d'un échantillon cristallin réel et donc à 3 dimensions (**a, b, c**), il y aura 3 conditions au lieu d'une seule. Ces conditions sont dites « de Laue » :

Si h, k, l des entiers, $\mathbf{H}\cdot\mathbf{a} = h$, $\mathbf{H}\cdot\mathbf{b} = k$, $\mathbf{H}\cdot\mathbf{c} = l$

Annexe B : Données supplémentaires sur l'approximation de Buckingham

Comme cela est expliqué dans la partie II.2.2.C de ce manuscrit, la sommation ou approximation de Buckingham permet de décomposer l'énergie d'interaction électrostatique en une somme de produits de moments atomiques. Cette approximation est définie en fonction des moments sans trace également appelés « moments externes ». La contrainte de cette méthode est qu'il faut être en mesure de calculer les moments externes des distributions.

Dans le cas de deux distributions de charges décomposées en sommes de pseudo-atomes selon le formalisme de Hansen-Coppens, ses moments internes de la distribution électronique d'un pseudo-atome sont exprimables en fonction de ses populations multipolaires.

Il est possible d'obtenir les moments externes à partir des moments internes, les moments externes étant une combinaison linéaire de ces derniers.

Pour les quadripôles, les formules sont :

$$\Theta_{zz} = \frac{1}{2}\Theta_{20}, \quad \Theta_{xx} = \frac{1}{2}\left(3\Theta_{22+} - \frac{1}{2}\Theta_{20}\right), \quad \Theta_{yy} = \frac{1}{2}\left(-3\Theta_{22+} - \frac{1}{2}\Theta_{20}\right), \quad (2.33)$$

$$\Theta_{xz} = \frac{3}{2}\Theta_{21+}, \quad \Theta_{yz} = \frac{3}{2}\Theta_{21-}, \quad \Theta_{xy} = \frac{3}{2}\Theta_{22-}$$

Pour les octupôles,

$$\begin{aligned} \Omega_{xxx} &= \frac{5}{8}\Omega_{33+} - \frac{3}{8}\Theta_{31+}, \quad \Omega_{yyy} = -\frac{5}{8}\Omega_{33-} - \frac{3}{8}\Theta_{31-}, \quad \Omega_{zzz} = \frac{1}{2}\Omega_{30}, \quad \Omega_{zzy} = \frac{1}{2}\Omega_{31-}, \\ \Omega_{zzx} &= \frac{1}{2}\Omega_{31+}, \quad \Omega_{xyz} = \frac{5}{4}\Omega_{32-}, \quad \Omega_{xxz} = -\frac{1}{4}\Omega_{30} + \frac{5}{4}\Omega_{32+}, \quad \Omega_{yyz} = -\frac{1}{4}\Omega_{30} - \frac{5}{4}\Omega_{32+}, \\ \Omega_{yyx} &= -\frac{5}{8}\Omega_{33+} - \frac{1}{8}\Theta_{31+}, \quad \Omega_{xxy} = \frac{5}{8}\Omega_{33-} - \frac{1}{8}\Theta_{31-} \end{aligned}$$

Pour les hexadécapôles,

$$\Phi_{xxxx} = \frac{35}{64}\Phi_{44+} - \frac{5}{16}\Phi_{42+} + \frac{3}{64}\Phi_{40}, \quad \Phi_{yyyy} = \frac{35}{64}\Phi_{44+} + \frac{5}{16}\Phi_{42+} + \frac{3}{64}\Phi_{40}$$

$$\begin{aligned}
\Phi_{zzzz} &= \frac{1}{8}\Phi_{40}, \Phi_{zzzx} = \frac{5}{8}\Phi_{41+}, \Phi_{zzzy} = \frac{5}{8}\Phi_{41-}, \Phi_{xxyy} = \frac{35}{64}\Phi_{44-} - \frac{5}{32}\Phi_{42-}, \\
\Phi_{yyyx} &= -\frac{35}{64}\Phi_{44-} - \frac{5}{32}\Phi_{42-}, \Phi_{xxxz} = \frac{35}{32}\Phi_{43+} - \frac{15}{32}\Phi_{41+}, \Phi_{yyyz} = -\frac{35}{32}\Phi_{43-} - \frac{15}{32}\Phi_{41-}, \\
\Phi_{xxzz} &= \frac{5}{16}\Phi_{42+} - \frac{1}{16}\Phi_{40}, \Phi_{yyzz} = -\frac{5}{16}\Phi_{42+} - \frac{1}{16}\Phi_{40}, \Phi_{xxzz} = -\frac{35}{64}\Phi_{44+} + \frac{1}{64}\Phi_{40}, \\
\Phi_{xxyy} &= -\frac{35}{64}\Phi_{44+} + \frac{1}{64}\Phi_{40}, \Phi_{xyzz} = \frac{5}{16}\Phi_{42-}, \Phi_{xxyz} = \frac{35}{32}\Phi_{43-} - \frac{5}{32}\Phi_{41-}, \\
\Phi_{xyyz} &= -\frac{35}{32}\Phi_{43-} - \frac{5}{32}\Phi_{41-}
\end{aligned}$$

Annexe C : Matériels supplémentaires sur l'étude de la distribution de charges de l'inhibiteur fidarestat

Tableau C1 : Propriétés topologiques aux points critiques de liaison (BCP) de la molécule fidarestat

Liaison	d_{A1-A2}^a (Å)	d_{A1-BCP}^a (Å)	d_{BCP-A2}^a (Å)	ρ_{cp}^b (e/Å ³)	$\nabla^2\rho_{cp}^c$ (e/Å ⁵)	λ_3^d (e/Å ⁵)	λ_2^d (e/Å ⁵)	λ_1^d (e/Å ⁵)	ε^e
C14—F17	1,3558(8)	0,548	0,808	1,895	-15,1	14,9	-14,5	-15,5	0,06
C19—N21	1,3363(7)	0,520	0,817	2,239	-26,3	8,4	-15,1	-19,7	0,24
C2I—N1I	1,3479(6)	0,586	0,762	2,310	-22,8	13,6	-16,8	-19,7	0,15
C5—N4	1,3646(6)	0,551	0,813	2,151	-21,9	11,4	-15,2	-18,1	0,16
C2I—N4	1,3993(6)	0,594	0,806	2,108	-18,9	13,5	-15,1	-17,2	0,12
C7I—N1I	1,4524(5)	0,603	0,851	1,746	-12,2	11,7	-11,7	-12,3	0,05
C5—O6I	1,2136(7)	0,437	0,777	2,955	-34,3	22,6	-26,4	-30,5	0,14
C2I—O3I	1,2205(7)	0,433	0,788	2,881	-29,5	24,8	-24,5	-29,8	0,18
C19—O20	1,2322(8)	0,449	0,783	2,862	-36,9	17,4	-24,8	-29,5	0,16
C11—O10	1,3666(8)	0,528	0,840	2,028	-18,1	12,7	-14,2	-16,6	0,15
C9—O10	1,4271(8)	0,566	0,862	1,640	-8,8	13,8	-11,1	-11,5	0,04
C13—C14	1,3798(7)	0,680	0,699	2,187	-20,5	10,1	-13,0	-17,6	0,26
C15—C16	1,3884(7)	0,654	0,736	2,183	-21,8	9,6	-14,3	-17,2	0,17
C14—C15	1,3915(7)	0,696	0,696	2,146	-19,7	10,7	-13,3	-17,2	0,23
C11—C16	1,3980(7)	0,755	0,643	2,104	-20,3	9,7	-13,2	-16,8	0,22
C11—C12	1,3991(6)	0,677	0,722	2,125	-19,6	10,3	-13,0	-16,9	0,23
C12—C13	1,4002(6)	0,749	0,652	2,081	-19,2	9,3	-12,4	-16,0	0,22
C9—C19	1,5184(6)	0,713	0,805	1,731	-12,6	11,5	-11,3	-12,8	0,12
C7I—C12	1,5180(6)	0,748	0,770	1,667	-11,3	10,9	-10,5	-11,7	0,10
C9—C8I	1,5211(6)	0,732	0,790	1,651	-11,8	10,3	-10,9	-11,3	0,04
C7I—C8I	1,5320(6)	0,769	0,763	1,572	-10,1	10,2	-10,1	-10,2	0,01
C7I—C5	1,5426(6)	0,764	0,778	1,587	-9,6	11,5	-9,9	-11,1	0,10
N1I—H1I	1,0089(8)	0,759	0,250	2,191	-32,4	26,5	-28,5	-30,4	0,06
N4—H4	1,0089(8)	0,745	0,264	2,161	-27,5	27,4	-27,4	-27,5	0,01
N21—H121	1,0090(8)	0,732	0,277	2,209	-26,8	27,4	-26,0	-28,2	0,08
N21—H221	1,0089(8)	0,749	0,260	2,160	-29,9	25,8	-27,1	-28,6	0,05
C13—H13	1,0829(8)	0,710	0,372	1,805	-17,0	15,0	-15,5	-16,5	0,06
C15—H15	1,0829(8)	0,719	0,364	1,802	-17,2	16,0	-16,2	-17,0	0,04
C16—H16	1,0829(8)	0,702	0,381	1,818	-17,7	15,6	-16,3	-17,0	0,04
C8I—H18I	1,0919(8)	0,729	0,363	1,776	-16,8	15,3	-15,7	-16,4	0,04
C8I—H28I	1,0919(8)	0,722	0,370	1,791	-16,3	15,9	-15,8	-16,3	0,03
C9—H9	1,0989(8)	0,721	0,378	1,770	-15,5	16,6	-15,8	-16,3	0,04

^a d_{A1-A2} , d_{A1-BCP} et d_{BCP-A2} sont respectivement les distances entre les atomes liés, entre le premier atome et le point critique de liaison BCP, et le point critique BCP et le second atome. Les liaisons sont ordonnées en fonction de leur nature et de la longueur de la liaison. ^b ρ_{cp} est la densité électronique totale au point critique CP. ^c $\nabla^2\rho_{cp}$ le Laplacien au CP. ^d λ_3 , λ_2 , λ_1 sont les valeurs propres de la matrice Hessienne $\partial^2\rho/\partial x_i\partial x_j$. ^e $\varepsilon = 1-\lambda_2/\lambda_1$ est l'ellipticité.

Tableau C2 : Paramètres du modèle multipolaire de distribution de charges du fidarestat.

Atome	κ	κ'	P_{val}	$d_{1,+1}$	$d_{1,-1}$	$d_{1,0}$	$q_{2,0}$	$q_{2,+1}$	$q_{2,-1}$	$q_{2,+2}$	$q_{2,-2}$	$o_{3,0}$	$o_{3,+1}$	$o_{3,-1}$	$o_{3,+2}$	$o_{3,-2}$	$o_{3,+3}$	$o_{3,-3}$
F17	0,98828	0,98855	7,049	0,001	-0,009	0,004	-0,080	0,008	-0,014	-0,023	0,	-0,029	-0,004	-0,016	-0,040	0,014	-0,015	0,046
N1I	0,98389	0,92817	5,162	-0,013	0,003	-0,013	0,022	0,026	0,003	0,066	-0,031	0,019	0,024	0,037	-0,005	-0,016	-0,228	-0,005
N4	0,98385	0,91608	5,185	-0,013	0,022	0,064	-0,017	-0,026	0,029	-0,024	-0,016	-0,055	0,006	0,	0,021	0,009	-0,224	0,020
N21	0,97744	0,88722	5,257	0,054	0,001	0,027	0,010	-0,004	0,036	-0,036	-0,006	-0,006	0,048	-0,023	-0,007	-0,007	0,228	-0,033
O6I	0,96775	1,13212	6,276	-0,042	-0,010	0,043	-0,037	-0,022	-0,009	-0,038	-0,024	0,009	0,016	-0,026	0,001	0,012	0,038	-0,002
O3I	0,96739	1,13280	6,386	-0,046	0,001	-0,004	-0,059	-0,007	-0,001	-0,053	-0,020	0,002	-0,044	0,003	0,001	-0,023	0,026	0,008
O10	0,97056	1,16332	6,314	0,004	-0,016	-0,049	-0,040	-0,013	-0,005	-0,053	0,049	-0,102	0,004	0,025	0,058	0,009	-0,003	-0,025
O20	0,96755	1,12979	6,333	-0,055	0,018	0,006	-0,041	0,005	-0,005	-0,029	0,025	0,003	-0,007	-0,002	0,013	0,007	0,002	0,002
C7I	0,99367	0,99475	3,797	0,017	0,009	-0,054	-0,053	-0,009	0,035	-0,005	-0,021	0,227	-0,067	-0,043	0,040	0,025	0,187	-0,015
C2I	0,99857	1,02553	3,861	-0,053	-0,024	0,035	-0,237	0,022	0,055	0,091	-0,026	-0,009	-0,052	0,017	0,001	-0,012	-0,273	-0,008
C5	1,00489	1,00289	3,735	-0,007	-0,079	-0,019	-0,255	0,035	0,018	-0,044	0,129	-0,005	-0,011	0,006	-0,054	0,017	0,258	0,087
C11	0,99124	0,99404	3,974	0,110	0,012	0,008	-0,185	0,060	-0,008	-0,030	0,046	0,010	-0,005	-0,069	0,011	0,004	-0,277	-0,011
C12	0,97659	1,03124	4,164	0,035	-0,011	0,004	-0,166	-0,007	-0,012	-0,045	0,022	-0,056	-0,001	0,011	-0,003	-0,014	-0,273	-0,030
C13	0,97664	0,94447	4,220	-0,012	0,052	-0,016	-0,169	0,007	-0,026	-0,018	0,024	-0,003	0,025	-0,021	-0,007	0,002	-0,273	-0,019
C14	0,99086	1,07104	3,862	0,086	-0,007	-0,010	-0,114	0,021	0,006	-0,063	-0,008	-0,019	0,081	-0,026	0,025	-0,007	-0,246	-0,018
C15	0,98162	0,98807	4,104	0,014	-0,063	-0,024	-0,205	0,023	0,023	0,001	0,	-0,012	0,034	-0,013	-0,022	-0,013	-0,261	-0,034
C16	0,98929	1,01102	3,949	0,012	-0,054	0,028	-0,174	-0,012	0,007	-0,038	-0,054	-0,060	0,054	0,048	0,020	0,017	-0,265	0,013
C9	1,00891	0,96481	3,647	0,028	0,084	0,015	-0,012	0,004	-0,106	0,007	0,059	0,029	-0,002	-0,009	0,339	-0,038	-0,002	0,021
C8I	0,97473	0,96128	4,247	0,020	0,006	-0,076	-0,033	0,038	0,003	-0,019	-0,051	0,010	-0,004	0,015	0,416	-0,059	0,021	-0,029
C19	0,99407	1,03382	3,891	0,011	0,056	-0,024	-0,231	-0,005	-0,003	-0,008	-0,098	0,	-0,051	-0,014	0,024	0,008	0,307	0,034
H1I	1,16000	1,40000	0,755			0,116												
H4	1,16000	1,40000	0,798			0,106												
H13	1,16000	1,18000	0,880			0,136												
H15	1,16000	1,18000	0,886			0,123												
H16	1,16001	1,18000	0,936			0,137												
H9	1,16000	1,18000	0,950			0,103												
H18I	1,16001	1,18000	0,861			0,112												
H28I	1,16000	1,18000	0,924			0,085												
H12I	1,16000	1,40000	0,836			0,119												
H22I	1,16000	1,40000	0,762			0,116												

Tableau C3 : Systèmes d'axes locaux définis spécifiquement pour chaque atome du fidarestat

Atome	orientation	Atome1	Atome2
C7I	ZX	N1I	C5
N1I	bXY	C7I	C2I
N4	bXY	C2I	C5
C2I	bXY	N1I	N4
O3I	XY	C2I	N1I
C5	XY	C7I	N4
O6I	XY	C5	C7I
C11	bXY	C12	C16
C12	bXY	C11	C13
C13	bXY	C12	C14
C14	bXY	C13	C15
F17	ZX	C14	C13
C15	bXY	C14	C16
C16	bXY	C15	C11
O10	bZX	C11	C9
C9	bZX	C8I	C19
C8I	bZX	C7I	C9
C19	XY	C9	O20
O20	XY	C19	C9
N21	XY	C19	H221

Tableau C4 : Coefficients α et exposants n_l des fonctions radiales. Les fonctions radiales de type Slater sont utilisées pour définition des termes multipolaires (dipôle, quadrupôle, octapôle, hexadécapôle) des différents atomes.

$$R_{nl}(\kappa' r) = \zeta^{n_l+3} (\kappa' r)^{n_l} \exp(-\zeta \kappa' r) / (n_l + 2)!$$

Atome	α (Bohr ⁻¹)	n_1	n_2	n_3	n_4
H	2,26	1	/	/	/
C	3,00	2	2	3	4
N	3,80	2	2	3	4
O	4,50	2	2	3	4
		2	2	3	4

Tableau C5 : Paramètres d'agitation thermique des atomes non hydrogène utilisés pour l'estimation des paramètres d'agitation thermique anisotrope des atomes d'hydrogène à l'aide du serveur SHADE, ainsi que les paramètres estimés obtenus pour ces derniers.

Atom	<i>U11</i>	<i>U22</i>	<i>U22</i>	<i>U12</i>	<i>U13</i>	<i>U23</i>
F17	0,0160	0,0194	0,0186	0,0084	-0,0084	-0,0006
N1I	0,0072	0,0045	0,0099	-0,0012	0,0047	-0,0023
N4	0,0086	0,0041	0,0108	-0,0009	0,0021	-0,0025
N2I	0,0142	0,0071	0,0233	0,0048	-0,0066	-0,0049
O6I	0,0085	0,0078	0,0174	-0,0020	0,0057	0,0001
O3I	0,0131	0,0081	0,0149	0,0026	0,0066	-0,0042
O10	0,0110	0,0113	0,0110	0,0069	-0,0042	-0,0061
O20	0,0098	0,0084	0,0144	-0,0009	-0,0050	0,0016
C7I	0,0054	0,0037	0,0064	-0,0002	0,0019	-0,0013
C2I	0,0076	0,0048	0,0086	0,0011	0,0022	-0,0024
C5	0,0057	0,0041	0,0085	-0,0007	0,0016	-0,0007
C11	0,0070	0,0082	0,0073	0,0029	-0,0013	-0,0032
C12	0,0058	0,0050	0,0068	0,0013	0,0003	-0,0014
C13	0,0069	0,0077	0,0100	0,0019	-0,0005	-0,0007
C14	0,0078	0,0109	0,0105	0,0018	-0,0025	0,0003
C15	0,0097	0,0132	0,0092	0,0009	-0,0030	-0,0024
C16	0,0103	0,0127	0,0086	0,0036	-0,0024	-0,0047
C9	0,0065	0,0055	0,0075	0,0012	-0,0002	-0,0007
C8I	0,0076	0,0053	0,0069	0,0009	0,0012	-0,0003
C19	0,0061	0,0056	0,0111	0,0009	-0,0010	-0,0003
H1I	0,0247	0,0202	0,0400	-0,0104	0,0139	-0,0017
H4	0,0334	0,0131	0,0414	-0,0083	0,0040	-0,0092
H13	0,0263	0,0219	0,0267	0,0119	-0,0003	-0,0070
H15	0,0307	0,0356	0,0174	0,0036	-0,0102	-0,0061
H16	0,0322	0,0305	0,0244	0,0139	-0,0043	-0,0147
H9	0,0206	0,0141	0,0257	-0,0009	0,0061	0,0030
H18I	0,0224	0,0135	0,0398	-0,0020	0,0103	0,0042
H28I	0,0276	0,0293	0,0182	0,0057	-0,0078	-0,0080
H12I	0,0383	0,0282	0,0506	0,0060	-0,0255	-0,0151
H22I	0,0307	0,0202	0,0474	0,0132	-0,0099	0,0004

Annexe D : Exemple de résultats d'analyse statistique sur les coordonnées atomiques de l'inhibiteur sorbinil et du groupement nicotinamide du facteur NADP⁺ dans le complexe aldose réductase – sorbinil.

Ces tableaux donnent les résultats de l'analyse statistique réalisée sur les coordonnées atomiques des atomes non hydrogène de l'inhibiteur sorbinil, ainsi que du groupement nicotinamide du cofacteur NADP⁺, dans le cadre de l'étude du complexe aldose réductase – sorbinil (test n°1)

Inhibiteur Sorbinil

Atome	C1			F1			C2			C3			C4			O1		
Coordonnées	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
Moyenne	64,666	35,973	85,793	63,857	35,830	84,725	66,049	35,965	85,648	66,871	36,115	86,758	66,318	36,296	88,027	67,166	36,390	89,080
Ecart-type	0,056	0,043	0,029	0,070	0,056	0,026	0,058	0,068	0,041	0,044	0,084	0,050	0,030	0,066	0,043	0,022	0,085	0,053
Coefficient kurtosis de Pearson	-0,228	0,574	0,043	-0,205	0,955	0,078	-0,225	0,042	-0,031	-0,218	0,249	-0,072	-0,166	0,314	-0,074	1,015	0,599	-0,108
Coefficient skewness de Pearson	0,048	-0,286	-0,039	0,048	-0,315	-0,051	0,050	-0,207	-0,030	0,064	-0,281	-0,017	0,125	-0,377	-0,023	0,029	-0,395	-0,014
Mode	64,672	35,984	85,801	63,853	35,842	84,732	66,054	35,948	85,652	66,876	36,128	86,747	66,301	36,313	88,036	67,164	36,406	89,090
Médiane	64,667	35,974	85,793	63,857	35,831	84,725	66,049	35,968	85,649	66,872	36,120	86,759	66,318	36,301	88,028	67,165	36,397	89,081
Minimum	64,497	35,749	85,691	63,640	35,547	84,632	65,872	35,674	85,503	66,741	35,748	86,584	66,227	36,009	87,876	67,021	36,007	88,899
Maximum	64,871	36,132	85,896	64,104	36,051	84,811	66,271	36,164	85,783	67,048	36,388	86,938	66,435	36,500	88,172	67,262	36,776	89,270
Erreur standard sur kurtosis	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078
Erreur standard sur skewness	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039
Rapport Kurtosis/erreur standard	-2,938	7,398	0,550	-2,644	12,297	1,003	-2,905	0,547	-0,395	-2,807	3,209	-0,925	-2,138	4,051	-0,956	13,070	7,722	-1,386
Rapport skewnew/erreur standard	1,239	-7,374	-1,007	1,230	-8,108	-1,326	1,283	-5,339	-0,767	1,661	-7,231	-0,441	3,232	-9,719	-0,588	0,758	-10,175	-0,365
Atome	C5			C6			C7			C9			O3			N2		
Coordonnées	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
Moyenne	66,571	36,905	90,269	65,295	36,126	90,627	64,275	36,424	89,539	63,624	37,782	89,744	64,225	38,826	89,873	62,327	37,586	89,908
Ecart-type	0,023	0,058	0,041	0,025	0,028	0,038	0,020	0,017	0,028	0,024	0,016	0,026	0,030	0,017	0,028	0,023	0,019	0,025
Coefficient kurtosis de Pearson	0,214	71,139	-0,037	0,602	0,175	-0,076	0,332	0,184	0,012	0,174	0,127	0,176	0,073	0,207	0,138	0,143	0,026	0,006
Coefficient skewness de Pearson	0,203	-4,283	-0,044	0,135	-0,149	-0,015	0,302	-0,170	-0,038	0,117	-0,176	-0,011	0,083	-0,082	-0,095	0,135	-0,247	0,071
Mode	66,564	36,916	90,286	65,296	36,124	90,618	64,275	36,425	89,546	63,614	37,785	89,748	64,217	38,830	89,869	62,334	37,590	89,906
Médiane	66,570	36,908	90,270	65,295	36,126	90,628	64,274	36,425	89,539	63,624	37,782	89,744	64,225	38,826	89,873	62,327	37,587	89,908
Minimum	66,493	35,881	90,126	65,180	36,007	90,506	64,215	36,345	89,441	63,539	37,701	89,639	64,121	38,749	89,767	62,247	37,508	89,827
Maximum	66,670	37,065	90,402	65,396	36,208	90,752	64,354	36,482	89,640	63,710	37,835	89,837	64,335	38,883	89,976	62,411	37,655	89,990
Erreur standard sur kurtosis	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078
Erreur standard sur skewness	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039
Rapport Kurtosis/erreur standard	2,761	916,451	-0,482	7,756	2,248	-0,983	4,273	2,374	0,150	2,242	1,630	2,266	0,942	2,660	1,776	1,846	0,333	0,077
Rapport skewnew/erreur standard	5,229	-110,340	-1,129	3,487	-3,843	-0,387	7,783	-4,378	-0,968	3,025	-4,547	-0,291	2,145	-2,124	-2,443	3,471	-6,369	1,820
Atome	C8			O2			N1			C10			C11					
Coordonnées	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z			
Moyenne	62,034	36,306	89,766	60,896	35,869	89,827	63,094	35,534	89,590	64,936	36,314	88,166	64,114	36,140	87,056			
Ecart-type	0,020	0,021	0,018	0,020	0,029	0,023	0,023	0,017	0,028	0,028	0,026	0,031	0,040	0,028	0,026			
Coefficient kurtosis de Pearson	0,343	0,081	0,151	0,591	0,295	1,715	0,600	-0,028	0,087	-0,191	0,249	-0,025	-0,255	6,690	0,061			
Coefficient skewness de Pearson	0,319	-0,264	-0,043	0,491	-0,273	-0,642	0,469	-0,240	0,023	0,107	-0,340	-0,029	0,058	-1,575	-0,018			
Mode	62,038	36,308	89,763	60,893	35,867	89,831	63,091	35,542	89,595	64,925	36,318	88,145	64,117	36,152	87,057			
Médiane	62,033	36,308	89,766	60,894	35,871	89,828	63,092	35,535	89,590	64,936	36,315	88,166	64,114	36,142	87,056			
Minimum	61,973	36,223	89,703	60,834	35,741	89,721	63,013	35,456	89,494	64,848	36,201	88,061	63,996	35,939	86,960			
Maximum	62,109	36,388	89,827	60,979	35,993	89,906	63,190	35,590	89,687	65,045	36,396	88,278	64,264	36,228	87,150			
Erreur standard sur kurtosis	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078			
Erreur standard sur skewness	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039			
Rapport Kurtosis/erreur standard	4,415	1,044	1,948	7,614	3,804	22,095	7,732	-0,363	1,120	-2,455	3,202	-0,319	-3,280	86,186	0,791			
Rapport skewnew/erreur standard	8,216	-6,799	-1,106	12,638	-7,033	-16,546	12,090	-6,190	0,600	2,754	-8,767	-0,742	1,486	-40,572	-0,469			

Groupement nicotinamide du NADP

Atome	N1N			C2N			C3N			C7N			O7N		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
Moyenne	59,597	34,974	92,900	59,651	36,382	92,919	60,900	37,004	92,932	61,025	38,523	92,943	62,155	39,020	92,965
Ecart-type	0,012	0,012	0,016	0,012	0,012	0,022	0,012	0,013	0,028	0,016	0,014	0,030	0,017	0,017	0,032
Coefficient kurtosis de Pearson	0,108	0,011	-0,431	0,006	0,001	-0,498	-0,059	-0,365	-0,699	-0,130	-0,386	-0,617	-0,090	-0,296	0,034
Coefficient skewness de Pearson	-0,087	0,103	0,186	0,059	0,091	0,062	0,090	-0,075	0,136	0,098	-0,099	0,131	0,114	0,046	0,059
Mode	59,596	34,972	92,896	59,647	36,384	92,925	60,901	37,013	92,931	61,026	38,533	92,944	62,153	39,024	92,962
Médiane	59,597	34,974	92,899	59,651	36,382	92,918	60,900	37,004	92,930	61,025	38,524	92,942	62,155	39,020	92,964
Minimum	59,549	34,937	92,856	59,604	36,346	92,849	60,855	36,964	92,848	60,973	38,483	92,857	62,102	38,970	92,860
Maximum	59,639	35,019	92,946	59,690	36,426	92,983	60,943	37,044	93,008	61,088	38,564	93,026	62,231	39,082	93,075
Erreur standard sur kurtosis	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078
Erreur standard sur skewness	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039
Rapport Kurtosis/erreur standard	1,393	0,139	-5,553	0,079	0,010	-6,411	-0,759	-4,704	-8,999	-1,674	-4,968	-7,951	-1,157	-3,818	0,432
Rapport skewnew/erreur standard	-2,242	2,659	4,800	1,532	2,334	1,610	2,325	-1,933	3,498	2,530	-2,555	3,372	2,948	1,177	1,515

Atome	N7N			C4N			C5N			C6N		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
Moyenne	59,951	39,168	93,401	62,037	36,187	93,040	62,043	34,795	93,009	60,804	34,197	92,980
Ecart-type	0,020	0,014	0,035	0,012	0,017	0,019	0,012	0,017	0,021	0,014	0,014	0,018
Coefficient kurtosis de Pearson	-0,172	-0,111	0,295	0,007	-0,552	-0,142	0,113	-0,540	0,541	0,052	-0,284	-0,188
Coefficient skewness de Pearson	0,117	0,044	0,333	0,053	-0,208	0,187	-0,114	-0,200	0,432	-0,148	-0,120	0,128
Mode	59,946	39,162	93,393	62,037	36,195	93,034	62,041	34,809	92,997	60,801	34,201	92,971
Médiane	59,951	39,168	93,399	62,036	36,189	93,038	62,043	34,797	93,007	60,804	34,198	92,979
Minimum	59,892	39,123	93,282	61,989	36,136	92,984	61,992	34,742	92,953	60,750	34,152	92,920
Maximum	60,035	39,213	93,550	62,076	36,233	93,104	62,084	34,842	93,140	60,851	34,240	93,048
Erreur standard sur kurtosis	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078	0,078
Erreur standard sur skewness	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039	0,039
Rapport Kurtosis/erreur standard	-2,216	-1,424	3,799	0,085	-7,109	-1,830	1,451	-6,953	6,974	0,665	-3,662	-2,422
Rapport skewnew/erreur standard	3,007	1,140	8,574	1,369	-5,362	4,817	-2,940	-5,164	11,121	-3,818	-3,089	3,300