



**HAL**  
open science

# Étude de complexes à forte diffusion anormale pour la détermination rapide de la structure de protéines par la méthode MAD

Meike Stelter

► **To cite this version:**

Meike Stelter. Étude de complexes à forte diffusion anormale pour la détermination rapide de la structure de protéines par la méthode MAD. Biophysique [physics.bio-ph]. Université Joseph-Fourier - Grenoble I, 2005. Français. NNT : . tel-00011238

**HAL Id: tel-00011238**

**<https://theses.hal.science/tel-00011238>**

Submitted on 19 Dec 2005

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

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

*UNIVERSITE JOSEPH FOURIER-GRENOBLE 1*  
*SCIENCES ET GEOGRAPHIE*

THESE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITE JOSEPH FOURIER**

**Discipline : Physique**

Présentée et soutenue publiquement par

**Meike STELTER**

Le 12 décembre 2005

**Étude de complexes à forte diffusion anormale  
pour la détermination rapide de la structure  
de protéines par la méthode MAD**

COMPOSITION DU JURY

|                     |  |
|---------------------|--|
| Rapporteurs         | <b>M. Marc SCHILTZ</b><br><b>M. Jorge NAVAZA</b> |
| Examinatrice        | <b>Mme Eva PEBAY-PEYROULA</b>                    |
| Directeurs de thèse | <b>M. Richard KAHN</b><br><b>M. Jean VICAT</b>   |

Thèse préparée au Laboratoire de Cristallographie Macromoléculaire  
de l'Institut de Biologie Structurale Jean-Pierre ÉBEL



## Remerciements

Ce travail de thèse a été effectué au Laboratoire de Cristallographie des Macromolécules, dirigé par Otto Dideberg. Je tiens à le remercier de m'avoir accueillie au sein de son laboratoire et de son soutien durant les mois de transition. Je le remercie également de ses conseils sur la cristallographie pour ma thèse ou à l'occasion de présentations.

Je tiens à remercier les membres du jury de thèse : Marc Schiltz, Jorge Navaza et Éva Pebay-Peyroula d'avoir accepté de juger ce travail. Merci à Marc de m'avoir appris beaucoup de choses à travers ses corrections au manuscrit et ses questions lors de la soutenance. Merci à Éva également pour sa lettre de recommandation.

Je remercie mes directeurs de thèse Richard Kahn et Jean Vicat pour leur disponibilité, leur aide et leur confiance. Je vous souhaite tout le mieux pour l'avenir.

Merci à Eric Girard qui m'a appris les bases pratiques de la cristallographie et de qui j'ai hérité le sujet.

Bonne continuation à Guillaume Pompidor pour la thèse et après. J'ai beaucoup apprécié les discussions que nous avons pu avoir.

Je remercie Juan Hermoso et Rafael Molina Monterrubio pour leur collaboration. Juan, merci beaucoup également pour ton aide et ta lettre de recommandation. Muchas gracias Rafa. Et je suis désolée de ne pas avoir pu être plus disponible au cours de ma dernière année de thèse.

Merci à Pier Lucio Anelli et Jean-François LeBas de m'avoir fourni les complexes de gadolinium.

Je remercie Sandra Jeudy et Chantal Abergel pour la collaboration sur la protéine yeaZ et les conseils qu'elles m'ont donnés. Merci également pour les cristaux de YGGV.

Je remercie les gens qui m'ont aidé avec les différentes méthodes physico-chimiques : Merci beaucoup à Rémy Sounier pour toutes les expériences de RMN effectuées et également pour avoir essayé de m'expliquer les résultats.

Merci à Antoine Royant pour les expériences de luminescence au cryobench. C'était un plaisir de travailler avec toi et de nager avec toi (même si j'ai abandonné les cours). Bonne continuation en longueurs.

*Je remercie Viviana pour son aide avec les gels natifs et plus généralement pour tous ses efforts grâce auxquelles le LCM fonctionne. Plein de bonheur à la future Maman.  
Je remercie Corinne Houles et Pierre LeParloüer pour leur aide avec le biacore et la calorimétrie respectivement.*

*Je remercie tous les membres du LCM et du LPM de leur aide et de la bonne compagnie.*

*Merci Laurence pour les conseils en cristallographie.*

*Merci Pauline de m'avoir supporté comme voisine de bureau et bonne dernière année de thèse.*

*Merci Philippe pour ta compréhension et ton humour.*

*Merci aux collègues (moniteurs ou non) de mon année : Hélène, Grégory, Jacques, Mickaël pour votre présence agréable que ce soit en préparant la fête de la science ou à d'autres occasions, j'ai beaucoup apprécié votre compagnie.*

# Partie I. Introduction

|   |           |
|---|-----------|
| <b>I.0. Introduction</b>  | <b>1</b>  |
| <b>I.1. La détermination de la structure de macromolécules par diffraction des rayons X</b>   | <b>2</b>  |
| I.1.1. Le problème de la phase en cristallographie des protéines  | 2         |
| I.1.2. La méthode du remplacement moléculaire   | 2         |
| I.1.3. Les méthodes directes  | 3         |
| I.1.4. Les méthodes de substitution isomorphe   | 3         |
| I.1.5. Les méthodes utilisant la diffraction anormale   | 4         |
| I.1.6. Comparaison des méthodes de phasage expérimental   | 5         |
| <b>I.2. La diffusion anormale</b>   | <b>6</b>  |
| I.2.1. Présentation du phénomène  | 6         |
| I.2.2. La brisure de la loi de Friedel  | 8         |
| I.2.3. Les différences dispersives  | 9         |
| <b>I.3. Détermination de la position des atomes lourds et phasage</b>   | <b>10</b> |
| I.3.1. Phasage avec la méthode SAD  | 10        |
| I.3.2. Procédures d'amélioration de phase   | 11        |
| I.3.3. Le signal anormal  | 12        |
| <b>I.4. Obtention de cristaux dérivés pour le phasage par des méthodes de remplacement isomorphe ou par des méthodes de diffusion anormale</b>    | <b>13</b> |
| I.4.1. Choix de l'atome lourd et de la longueur d'onde du rayonnement X incident  | 13        |
| I.4.2. Accessibilité du seuil, grandes longueurs d'onde   | 14        |
| I.4.3. Seuils d'absorption et raies blanches  | 15        |
| I.4.4. Revue de méthodes habituelles pour obtenir des dérivés pour le phasage par des méthodes de remplacement isomorphe ou de diffusion anormale | 16        |
| I.4.5. Dégâts d'irradiation   | 18        |
| I.4.6. Comparaison des méthodes d'obtention de dérivés anomaux  | 20        |
| <b>I.5. Utilisation de complexes de lanthanides pour obtenir des dérivés</b>  | <b>22</b> |
| I.5.1. Présentation des complexes de gadolinium utilisés  | 23        |
| I.5.2. Chimie de coordination de Gd(III)  | 24        |
| I.5.3. La diffusion anormale des lanthanides  | 25        |
| I.5.4. Obtention de cristaux dérivés à l'aide de complexes de gadolinium.   | 25        |
| I.5.5. Dégâts d'irradiation II  | 27        |
| I.5.6. Conclusions  | 28        |

# Partie II. Étude cristallographique

|  |           |
|--|-----------|
| <b>II.0. Introduction</b>  | <b>29</b> |
| <b>II.1. Cristallisation des protéines</b>   | <b>29</b> |
| II.1.1. Choix des protéines tests  | 29        |
| II.1.2. Cristallisation des protéines  | 30        |
| II.1.3. Conditions de cristallisation des différentes protéines  | 32        |
| II.1.4. Obtention de cristaux dérivés à une concentration de complexe de 300mM                           | 38        |
| <b>II.2. Acquisition et traitement des données de diffraction</b>  | <b>39</b> |
| II.2.1. Acquisition des données de diffraction   | 39        |
| II.2.2. Stratégie de l'acquisition des données   | 40        |
| II.2.3. Utilisation du robot de montage de cristaux sur BM30A  | 40        |
| II.2.4. Traitement des données de diffraction  | 41        |
| II.2.5. Conditions d'enregistrement et statistiques d'intégration de données pour les différents dérivés | 41        |
| <b>II.3. Obtention des phases expérimentales</b>   | <b>46</b> |
| II.3.1. Estimation de la fixation et détermination des sites de fixation                                 | 46        |
| II.3.2. Détermination des phases   | 47        |
| II.3.3. Procédé de phasage des données obtenues sur BM30A  | 47        |
| II.3.4. Évaluation du succès du phasage  | 48        |
| II.3.5. Résultats de phasage avec les différents dérivés   | 49        |
| II.3.6. Exemples de cartes de densité électronique expérimentale   | 52        |
| II.3.7. Exemple de résolution de structure d'une protéine de structure jusque-là inconnue                | 54        |
| II.3.8. Synthèse des résultats des essais de phasage   | 56        |
| II.3.9. Conclusions  | 59        |
| <b>II.4. Caractérisation du mode de fixation des complexes dans les protéines</b>                        | <b>60</b> |
| II.4.1. Affinement initial   | 60        |
| II.4.2. Les modèles des complexes utilisés   | 61        |
| II.4.3. Affinement avec complexes  | 61        |
| II.4.4. Urate oxydase  | 63        |
| II.4.5. YGGV   | 73        |
| II.4.6. YeaZ   | 75        |
| II.4.7. Glucose isomérase  | 77        |
| II.4.8. Lysozyme   | 79        |
| II.4.9. Thaumatine   | 83        |
| II.4.10. Protéine X  | 88        |
| II.4.11. Résultats sur la fixation des complexes   | 95        |
| <b>II.5. Conclusions</b>   | <b>97</b> |

## **Partie III. Méthodes physico-chimiques pour détecter la fixation d'un complexe sur une protéine**

|   |            |
|---|------------|
| <b>III.0. Méthodes physico-chimiques pour caractériser la fixation des complexes sur des protéines - Introduction</b>       | <b>99</b>  |
| III.0.1. Difficultés liées aux caractéristiques des complexes utilisés  | 100        |
| III.0.2. Considérations expérimentales  | 100        |
| <b>III.1. Calorimétrie de titrage isotherme</b>   | <b>101</b> |
| III.1.1. Introduction   | 101        |
| III.1.2. Étude de l'association d'une solution de lysozyme avec le complexe Gd-HPDO3A par calorimétrie de titrage isotherme | 103        |
| III.1.3. Conditions expérimentales  | 103        |
| III.1.4. Résultats  | 105        |
| III.1.5. Conclusion   | 108        |
| <b>III.2. Expériences de résonance plasmonique de surface</b>   | <b>108</b> |
| <b>III.3. Électrophorèse de gels d'acrylamide natifs</b>  | <b>109</b> |
| III.3.1. Préparation des gels   | 109        |
| III.3.2. Préparation du gel pour la migration du lysozyme de blanc d'œuf de poule   | 116        |
| III.3.3. Résultats obtenus avec le lysozyme de blanc d'œuf de poule   | 110        |
| III.3.4. Essais d'électrophorèse de gels natifs avec d'autres protéines   | 111        |
| III.3.5. Conclusions  | 112        |
| <b>III.4. Expériences de résonance magnétique nucléaire</b>   | <b>113</b> |
| III.4.1. Introduction   | 113        |
| III.4.2. Détection de la fixation d'un complexe sur une protéine  | 113        |
| III.4.3. Expériences de RMN 1D avec le lysozyme de blanc d'œuf de poule et les complexes Gd-HPDO3A et Lu-HPDO3A             | 114        |
| III.4.4. Préparation des échantillons   | 114        |
| III.4.5. Spectres 1D <sup>1</sup> H   | 115        |
| III.4.6. Protéine avec le complexe Gd-HPDO3A  | 117        |
| III.4.7. Protéine avec le complexe Lu-HPDO3A  | 119        |
| III.4.8. Interprétation des spectres  | 121        |
| III.4.9. Spectres 2D <sup>1</sup> H- <sup>1</sup> H NOESY   | 122        |
| III.4.10. Spectres 2D <sup>1</sup> H- <sup>1</sup> H NOESY de lysozyme et de lysozyme avec 100 mM de Lu-HPDO3A              | 123        |
| III.4.11. Interprétation des spectres   | 126        |
| III.4.12. Conclusions sur les expériences RMN 1D et 2D  | 126        |

|   |            |
|---|------------|
| <b>III.5. La luminescence des lanthanides</b>   | <b>127</b> |
| III.5.1. Propriétés photophysiques des ions de lanthanides trivalents   | 127        |
| III.5.2. But : détecter la fixation de nos complexes sur les protéines  | 129        |
| III.5.3. Expériences de luminescence avec des complexes et des ions de lanthanides avec des cristaux et des solutions de protéine                             | 131        |
| III.5.4. Conditions expérimentales  | 132        |
| III.5.5. Expériences de luminescence  | 134        |
| III.5.6. Spectre de fluorescence d'un cristal et d'une solution de protéine native  | 134        |
| III.5.7. Résultats d'expériences préliminaires avec des ions de lanthanides   | 136        |
| III.5.8. Expériences avec les complexes de lanthanides  | 139        |
| III.5.9. Conclusions et perspectives pour les expériences de luminescence des lanthanides   | 145        |
| <br>  |            |
| <b>III.6. Conclusions sur les différentes méthodes physico-chimiques testées</b>  | <b>147</b> |
| <br>  |            |
| <b>IV. Conclusions et perspectives</b>  | <b>149</b> |
| <br>  |            |
| <b>Références</b>   | <b>150</b> |
| <br>  |            |
| <b>Annexe 1 La carte de Patterson des atomes lourds</b>   | <b>I</b>   |
| <br>  |            |
| <b>Annexe 2 Comparaison de la fixation des complexes pour des cristaux dérivés de glucose isomérase obtenus en présence de différents agents précipitants</b> | <b>IV</b>  |

## Notes explicatives

Le mot **complexe** désigne la molécule composée de l'atome lourd (ion  $Gd^{3+}$ ) et du ligand qui chélate l'atome lourd. Le **ligand** est la molécule qui chélate l'atome lourd dans les complexes et le **dérivé** correspond à l'ensemble formé par le complexe lié à la protéine.

## Abréviations

**MIR** : Multiple Isomorphous Replacement

**SIR** : Single Isomorphous Replacement

**MAD** : Multiple-wavelength Anomalous Diffraction

**SAD** : Single-wavelength Anomalous Diffraction

**MIRAS** : Multiple Isomorphous Replacement with Anomalous Scattering

**SIRAS** : Single Isomorphous Replacement with Anomalous Scattering

**TF** : transformée de Fourier

**DO3A** : acide 1,4,7,10-tétraazacyclododecane 1,4,7-triacétique

**DOTA** : acide 1,4,7,10-tétraazacyclododecane 1,4,7,10-tétraacétique

**DOTA-BOM**: acide (phénylméthoxy)méthyl-1,4,7,10-tétrazacyclododecane-1,4,7,10-tétraacétique

**DTPA** : acide di-éthylène-tri-amine pentaacétique

**DTPA-BMA** : acide 1,3-bis-méthylamide di-éthylène-tri-amine pentaacétique

**DOTMA** : acide 1,4,7,10-tétraméthyl-1,4,7,10-tétraazacyclododecane 1,4,7,10-tétraacétique

**HPDO3A** : acide 10-(2-hydroxypropyl)- 1,4,7,10-tétraazacyclododecane 1,4,7-triacétique

**HPSA-DO3A** : acide 10-(2-((hydroxy-1-(hydroxyméthyl)éthyl)amino)-1-(hydroxyméthyl)-2-oxoéthyl)- 1,4,7,10-tétraazacyclododecane 1,4,7-triacétique

**EDTA** : acide éthylène diamine tétraacétique

**PEG** : polyéthylène glycol

## Définition des indicateurs statistiques calculés par les programmes utilisés :

**XDS** (Kabsch, 1993)

**S\_norm/S\_ano** : le rapport des écarts type  $\sigma(I)$  des réflexions générales calculés pour :

**S\_norm** : la loi de Friedel respectée et **S\_ano** : la loi de Friedel brisée.

Des valeurs supérieures à 1,0 indiquent une fixation du diffuseur anomal.

**SCALA** (CCP4, Collaborative Computational Project, 1994)

$$R_{fac} = \frac{\sum_{hkl} \sum_j |I_j - \langle I \rangle|}{\sum_{hkl} \sum_j |I_j|}$$

avec :  $\langle I \rangle$  : valeur moyenne des intensités des réflexions équivalentes.

$$R_{ano} = \frac{\sum_{hkl} |\langle I^+ \rangle - \langle I^- \rangle|}{\sum_{hkl} \langle I^+ \rangle + \langle I^- \rangle}$$

avec :  $\langle I^+ \rangle$  et  $\langle I^- \rangle$  : valeurs moyennes, sur l'ensemble des mesures, de chacune des réflexions  $I^+$  et  $I^-$ .

**SHARP** (de La Fortelle & Bricogne, 1997; Bricogne *et al.*, 2003)

Le facteur de qualité (FOM) correspond à l'inverse de l'étendu de la distribution de probabilité du facteur de structure complexe. Une valeur élevée indique la réussite du phasage.

Pour le programme SHARP, le facteur de qualité s'obtient par (M. Schiltz, EPF-Lausanne, communication) :

$$FOM = \frac{\langle |F(\mathbf{h})| \rangle}{\langle |F(\mathbf{h})| \rangle} = \frac{\iint F(\mathbf{h}) P[F(\mathbf{h})] dXdY}{\iint |F(\mathbf{h})| P[F(\mathbf{h})] dXdY}$$

où  $F(\mathbf{h})$  désigne le facteur de structure complexe et  $P[F(\mathbf{h})]$  sa distribution de probabilité.

**SOLOMON** (Abrahams, 1997), **DM** (Cowtan & Main, 1996)

La distribution de probabilité de facteur de structure complexe se résume à la distribution de probabilité de la phase  $P(\alpha)$ , le module du facteur de structure étant constant.

Le facteur de qualité s'écrit alors :

$$FOM = \left| \frac{\int P(\alpha) \exp(i2\pi\alpha) d\alpha}{\int P(\alpha) d\alpha} \right|$$

**DM**

$$R_{libre} = \frac{\sum |\rho_{obs} - \rho_{calc}|}{\sum |\rho_{obs} + \rho_{calc}|}$$

Facteur résiduel dans l'espace direct, calculé à partir de régions de protéine et de solvant qui ne sont pas incluses dans la modification de densité.

**CNS** (Brünger *et al.*, 1998)

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

Résiduel qui permet de suivre l'évolution de l'affinement.

**R<sub>free</sub>** : même expression que R mais calculé avec les réflexions qui ne sont pas incluses dans le processus de l'affinement (typiquement 5% des réflexions). Il constitue un indicateur non-biaisé de l'avancement de l'affinement.

**Partie I.**

**Introduction**



## I.0. Introduction

Ce travail se situe dans le cadre de la résolution de structures de protéines par diffraction des rayons X.

Après la production de protéine pure et stable, des différentes étapes qui mènent de la protéine à la structure, les étapes critiques sont la cristallisation et l'obtention de cristaux dérivés permettant de déterminer des phases expérimentales. C'est à cette dernière étape du processus que nous nous intéressons et plus précisément à l'obtention de cristaux dérivés pour les méthodes de phasage utilisant la diffusion anormale.

Le but de mon travail est d'apporter un outil pour obtenir des cristaux dérivés comme alternative à d'autres méthodes bien établies, comme l'incorporation de la sélénométhionine dans les protéines. En effet, la production et la cristallisation de protéines sélénées peuvent être difficiles voire impossibles. Dans ce cas, obtenir des cristaux dérivés s'avère un processus souvent long et aléatoire, nécessitant de nombreux essais qui demandent de pouvoir disposer de cristaux en nombre suffisant pour obtenir un dérivé adéquat.

L'outil utilisé ici est une classe de complexes de lanthanides qui présentent un fort potentiel de fixation dans les cristaux de protéines, ce qui permet de produire des dérivés avec un fort pouvoir de phasage. Les complexes se distinguent par leurs propriétés physiques de diffusion des rayons X et par leurs propriétés chimiques qui déterminent leur comportement vis-à-vis des protéines. À certaines longueurs d'onde particulières du rayonnement X, ils présentent en effet une forte diffusion anormale caractéristique des lanthanides. Malgré leur faible affinité pour les macromolécules biologiques, ils s'avèrent un outil puissant pour insérer et fixer des diffuseurs anormaux dans les cristaux de protéines, lorsqu'on les utilise à forte concentration, tout en étant d'utilisation facile.

Une première étape du travail consiste en une étude cristallographique systématique de l'interaction des différents complexes avec différentes protéines. L'étude va de l'obtention des cristaux dérivés au phasage. Les résultats de l'étude permettent de préciser le potentiel des complexes pour obtenir des cristaux dérivés, de vérifier que leur utilisation est adaptée à tout type de protéines, et de déterminer les conditions d'utilisation idéales des complexes pour obtenir des cristaux dérivés.

Dans un deuxième temps, l'affinement de la structure des complexes liés à différentes protéines a pour but d'identifier et de comprendre leur mode de fixation. Ceci devrait permettre d'optimiser l'utilisation des complexes et d'aider à choisir le complexe susceptible de se fixer en fonction des conditions de cristallisation ou des propriétés connues de la protéine. Un tel choix permettrait de limiter le nombre de cristaux à tester pour obtenir un dérivé approprié.

Un deuxième aspect du travail est la recherche d'une méthode physico-chimique, alternative à la cristallographie, qui permettrait de caractériser l'interaction des complexes avec des protéines, en solution ou dans le cristal. Une telle méthode éviterait de tester la fixation du complexe par cristallographie, processus relativement long et nécessitant du temps de faisceau, et pourrait permettre d'identifier, parmi les différents complexes, celui qui se fixe le mieux pour préparer de manière ciblée un cristal dérivé pour l'expérience de diffraction.

## **I.1. La détermination de la structure de macromolécules par diffraction des rayons X**

### **I.1.1. Le problème des phases en cristallographie des protéines**

Résoudre la structure d'une macromolécule par diffraction des rayons X consiste, du point de vue du cristallographe, à obtenir une image de la densité électronique de la macromolécule, image dans laquelle il pourra construire un modèle de la macromolécule d'intérêt à l'échelle atomique.

Expérimentalement, on expose le cristal de la molécule à étudier à un rayonnement dont la longueur d'onde est du même ordre de grandeur que les détails structuraux à déterminer, de l'ordre de l'Ångström (Å). Le rayonnement est diffracté dans des directions bien précises, les réflexions, qui sont repérées à l'aide de 3 nombres entiers, les indices de Miller ( $hkl$ ). La mesure de l'intensité  $I(hkl)$  de chacune de ces réflexions permet de déterminer l'amplitude de l'onde diffractée dans la direction correspondante, puisque l'intensité est proportionnelle au carré de l'amplitude. Or, pour reconstruire la densité électronique à partir de ces mesures, il faut connaître non seulement l'amplitude de l'ensemble des ondes diffractées, mais aussi leur déphasage par rapport au rayonnement incident. Cela revient à connaître, en module et en phase, l'amplitude complexe de chacune des ondes diffractées. Cette amplitude complexe, transformée de Fourier de la densité électronique, est appelée facteur de structure,  $F(hkl)$ , pour la réflexion considérée. Le module de chacun des facteurs de structure, étant déterminé expérimentalement, le problème central de la cristallographie est d'obtenir la phase de chacun de ces facteurs de structure. Parmi les méthodes permettant d'obtenir ces phases, nous nous intéresserons plus particulièrement aux méthodes de phasage *de novo* qui ne se servent que des données obtenues expérimentalement, sans utiliser d'information supplémentaire extérieure sur la structure à étudier.

### **I.1.2. La méthode du remplacement moléculaire**

Le remplacement moléculaire (Rossmann, 1972 ; Machin, 1985 ; Navaza, 1994) ne fait pas partie des méthodes de phasage *de novo*, car l'information de phase recherchée est obtenue à partir d'une structure déjà connue que l'on suppose proche de la structure recherchée. Comme le nombre de structures connues ne cesse d'augmenter, il est de plus en plus vraisemblable de pouvoir trouver une structure homologue à la structure recherchée.

Le choix de la structure de référence est guidé par la constatation suivante : l'homologie de séquence entre deux protéines implique très fréquemment une similarité structurale. La structure choisie peut modéliser tout ou partie de la structure d'intérêt. On recherche l'orientation et la position dans la maille de la structure connue pour lesquelles les intensités calculées à partir du modèle sont en bon accord avec les intensités observées, tout en évitant le chevauchement de molécules symétriques. Des phases de départ appliquées aux données expérimentales sont alors calculées pour les structures qui donnent les meilleurs accords.

La méthode ne nécessite que des données de diffraction enregistrées avec le cristal natif, ce qui constitue un avantage net par rapport aux méthodes de phasage expérimental décrites par la suite. Cependant, malgré le nombre très élevé de structures résolues par remplacement moléculaire et alors que l'on dispose d'un modèle raisonnable de structure connue, la méthode peut être mise en défaut. L'un des problèmes consiste en ce que la solution peut être biaisée par le modèle de départ et qu'une solution initiale apparemment correcte ne se laisse

pas ultérieurement affiner. Comme dans le cas où l'on ne dispose pas d'un modèle de structure connue, on a alors recours aux méthodes de phasage *de novo* décrites par la suite.

### I.1.3. Les méthodes directes

On regroupe sous le terme de « méthodes directes » un ensemble de méthodes qui s'attachent à résoudre le problème des phases en utilisant des *relations entre phases* qui découlent des valeurs observées pour les intensités.

Ces relations mettent en œuvre les facteurs de structure normalisés,  $E(hkl)$ , nombres complexes obtenus en divisant les facteurs de structure  $F(hkl)$  par la racine carrée de l'intensité moyenne qui serait obtenue à une résolution donnée si les atomes étaient distribués au hasard dans la maille. Les relations de phases les plus fortes sont obtenues entre réflexions avec des valeurs élevées du module du facteur de structure normalisé,  $|E(hkl)|$  (Sayre, 1952 ; Karle & Hauptman, 1956 ; Bricogne 1988).

La valeur maximale théorique que peut prendre  $|E(hkl)|$  est proportionnelle à la racine carrée du nombre d'atomes dans la maille. Cette valeur serait obtenue si les contributions au facteur de structure des atomes de la maille étaient toutes en phase pour la réflexion ( $hkl$ ) considérée. La probabilité d'un tel événement décroît très rapidement avec le nombre d'atomes dans la maille. De même, les relations entre phases s'affaiblissent lorsque le nombre d'atomes dans la maille augmente.

C'est pourquoi, alors que les méthodes directes sont quasi universellement utilisées pour résoudre la structure de petites molécules constituées de quelques centaines d'atomes au maximum, leur utilisation pour déterminer la structure de macromolécules biologiques ne comprenant pas d'atome lourd reste encore limitée à des structures constituées de quelque mille atomes et pour lesquelles on dispose de données de diffraction à très haute résolution

$$(d_{rés} = \frac{\lambda}{2\sin\theta_{max}} < 1,2 \text{ \AA}) \text{ (Usón \& Sheldrick, 1999).}$$

Dans le domaine de la détermination de structures de macromolécules biologiques, les méthodes directes sont cependant très utilisées pour rechercher la position d'atomes lourds qui constituent une sous-structure servant de référence pour résoudre la structure de la macromolécule d'intérêt par les méthodes de détermination expérimentale des phases – méthodes de substitution isomorphe et méthodes utilisant la diffraction anormale - décrites ci-dessous.

### I.1.4. Les méthodes de substitution isomorphe

Les méthodes de phasage par remplacement isomorphe consistent à comparer les données de diffraction  $I(hkl)$  d'un cristal de la protéine native aux données de cristaux dans lesquels des atomes lourds sont fixés à la molécule de la protéine. Ces atomes lourds doivent posséder un nombre d'électrons nettement plus élevé que ceux des atomes constituant la molécule biologique pour que les intensités des pics de diffraction du cristal dérivé diffèrent significativement des intensités de diffraction du cristal natif.

Une condition essentielle pour la réussite de ces méthodes est l'isomorphisme des structures cristallines de la protéine native et de la protéine dérivée. La fixation des atomes lourds sur la molécule ne doit pas modifier sa structure, son orientation et sa position dans le cristal ; les paramètres la maille cristalline doivent rester identiques. La densité électronique  $\rho_{PH}$  du

cristal dérivé peut alors être considérée comme la somme de la densité électronique  $\rho_P$  de la macromolécule seule, identique à celle du cristal natif, et de la densité électronique  $\rho_H$  de la sous-structure des atomes lourds. La transformée de Fourier de la densité électronique transfère cette propriété sur les facteurs de structure. L'isomorphisme se traduit donc par l'égalité :

$$\rho_{PH} = \rho_P + \rho_H \xrightarrow{TF} F_{PH} = F_P + F_H$$

où  $F_{PH}$ ,  $F_P$  et  $F_H$  représentent les facteurs de structure respectifs du cristal dérivé, du cristal natif et de la sous-structure des atomes lourds.

Pour déterminer la phase des facteurs de structure, on détermine tout d'abord la sous-structure des atomes lourds à partir des différences entre les intensités des réflexions du cristal dérivé et des réflexions correspondantes du cristal natif. Les méthodes utilisées à cet effet sont exposées dans l'annexe 1.

Une fois la sous-structure des atomes lourds déterminée, on peut calculer  $F_H$ , en module et en phase. Le facteur de structure recherché,  $F_P$ , est donné par la relation  $F_P = F_{PH} - F_H$  où les modules  $|F_P|$  et  $|F_{PH}|$  sont connus expérimentalement. Le membre de gauche de l'équation indique que, dans le plan complexe, le point d'affixe  $F_P$  se trouve sur un cercle de rayon  $|F_P|$  centré à l'origine, alors que le membre de droite indique qu'il se trouve sur un cercle de rayon  $|F_{PH}|$  centré au point d'affixe  $-F_H$ . Le point d'affixe  $F_P$  se trouve donc à l'intersection de ces deux cercles.

La méthode, dénommée méthode SIR (Single Isomorphous Replacement), qui utilise le cristal natif et un seul cristal dérivé, mène à deux solutions pour chaque phase résultante.

On voit que pour déterminer la phase des facteurs de structure sans ambiguïté, il faut au moins deux cristaux dérivés isomorphes différents : c'est la méthode MIR (Multiple Isomorphous Replacement), qui a été la première méthode de phasage utilisée pour résoudre la structure de macromolécules (Green *et al.*, 1954). Elle est toujours largement employée.

Les méthodes SIRAS/MIRAS (Single/Multiple Isomorphous Replacement with Anomalous Scattering) tiennent compte des différences anormales des données de diffraction comme source d'information de phase supplémentaire.

### **I.1.5. Les méthodes utilisant la diffraction anormale**

Les méthodes utilisant la diffraction anormale se servent du fait que le facteur de diffusion atomique d'un atome donné varie fortement, en module et en phase, dans la région d'un seuil d'absorption de cet atome : c'est le phénomène de diffusion anormale. Le facteur de diffusion atomique du diffuseur anormal est alors représenté par un nombre complexe.

Ces méthodes s'affranchissent du problème du non-isomorphisme en utilisant les données de diffraction d'un cristal dérivé unique. C'est la structure de ce cristal dérivé qui est déterminée et il n'est pas nécessaire d'enregistrer les données de diffraction d'un cristal de protéine native.

La méthode MAD (Multiple-wavelength Anomalous Diffraction) utilise des données de diffraction enregistrées à plusieurs longueurs d'onde, dans la région du seuil d'absorption du diffuseur anormal. Son utilisation est devenue possible grâce à l'emploi du rayonnement synchrotron en tant que rayonnement à longueur d'onde ajustable. Elle est aujourd'hui la méthode de phasage *de novo* la plus utilisée (Hendrickson, 1999).

Pour la méthode SAD (Single-wavelength Anomalous Diffraction), les données ne sont enregistrées qu'à une seule longueur d'onde choisie dans une région où le facteur de diffusion atomique présente une partie imaginaire élevée. Comme pour la méthode SIR, la solution de phase dans le cas de la méthode SAD est ambiguë, mais, contrairement à la méthode SIR, les

deux solutions ne sont pas équiprobables et l'ambiguïté de phase peut être partiellement levée avant même l'utilisation de méthodes d'amélioration ultérieure des phases comme les méthodes mettant en œuvre la modification de densité et la symétrie non-cristallographique.

### **I.1.6. Comparaison des méthodes de phasage expérimental**

Les méthodes de remplacement isomorphe présentent l'avantage sur les méthodes utilisant la diffraction anormale que la différence isomorphe des intensités de diffraction sur lesquelles repose le phasage est généralement bien plus grande que les différences anormales sur lesquelles reposent les méthodes MAD et SAD. De plus, elle n'oblige pas d'avoir recours à une source de rayonnement X à longueur d'onde variable.

L'inconvénient majeur de la méthode est que la condition nécessaire d'isomorphisme entre les différents cristaux utilisés n'est, le plus souvent, qu'imparfaitement remplie, ce qui entraîne que les différences entre les intensités de diffraction ne peuvent pas être interprétées en termes de pure substitution isomorphe, ce qui, finalement, peut empêcher le phasage. Par ailleurs, la méthode MIR nécessite des données de diffraction enregistrées avec au moins deux cristaux dérivés, cristaux qui sont le plus souvent difficiles à obtenir.

Les méthodes utilisant la diffusion anormale permettent de s'affranchir du problème du non-isomorphisme, puisque les données de diffraction peuvent, en principe, être enregistrées avec un seul cristal. Par contre, elles demandent des mesures des intensités de diffraction très précises en raison de la faiblesse relative des différences anormales.

La méthode MAD ne peut être utilisée qu'avec des données enregistrées avec du rayonnement de longueur d'onde ajustable. Par rapport à la méthode SAD elle présente l'avantage de donner des phases de bien meilleure qualité, puisqu'elle lève expérimentalement l'ambiguïté de la solution.

La méthode SAD, quant à elle, présente l'avantage de ne nécessiter que des données de diffraction enregistrées à une seule longueur d'onde ; si l'atome lourd s'y prête, ces données peuvent, par exemple, être enregistrées avec un générateur de rayons X de laboratoire. La méthode permet le phasage avec des données anormales enregistrées hors d'un seuil d'absorption du diffuseur anormal considéré, dans des conditions où la partie imaginaire du facteur de diffusion atomique est non négligeable, sans cependant être maximale. Ceci est particulièrement intéressant pour des éléments dont les seuils d'absorption se trouvent dans des domaines de longueur d'onde difficilement accessibles pour des expériences de diffraction (S, Ca, Xe, U...). Par ailleurs, puisque la méthode SAD ne nécessite qu'un jeu unique de données, la durée d'exposition du cristal aux rayons X est réduite, diminuant ainsi les risques de dommage dus à l'irradiation. On sait en effet que ce problème est devenu particulièrement important avec l'utilisation des cristaux de taille de plus en plus petite, utilisation rendue possible avec les sources de rayonnement synchrotron de troisième génération, très intense. Outre la réduction du pouvoir de diffraction du cristal qui se traduit par la perte des réflexions à haute résolution, les dommages d'irradiation sont une source de non-isomorphisme au sein même d'enregistrements de données sur un seul cristal, effet qui, s'il n'est pas bien pris en compte, peut empêcher le phasage. En ce qui concerne la précision des données, la méthode SAD nécessite une précision d'autant plus élevée que les différences anormales dues à des éléments excités loin d'un seuil d'absorption peuvent être relativement faibles.

## I.2. La diffusion anormale

### I.2.1. Présentation du phénomène

Dans le modèle de diffusion des rayons X par des électrons libres, le rayonnement diffusé est déphasé de  $180^\circ$  par rapport au rayonnement incident : le facteur de diffusion pour un électron est réel.

Dans un atome, les électrons des couches internes (K et L) sont liés fortement au noyau, et ces électrons ne peuvent plus être considérés comme libres.

Avec le modèle de l'oscillateur harmonique amorti, de fréquence de résonance  $\omega_0$ , correspondant à l'une des fréquences d'absorption de l'atome, le facteur de diffusion  $f$  pour un électron s'écrit :

$$f = \omega^2 / (\omega^2 - i\alpha\omega_0\omega - \omega_0^2)$$

où  $\omega$  est la pulsation du rayonnement incident et  $\alpha$  le facteur d'atténuation.

La comparaison du modèle (figure 1.2.1) aux valeurs expérimentales (figure 1.2.2), montre que, bien qu'approximatif, le modèle représente bien certaines caractéristiques des variations des facteurs de diffusion au seuil d'absorption.

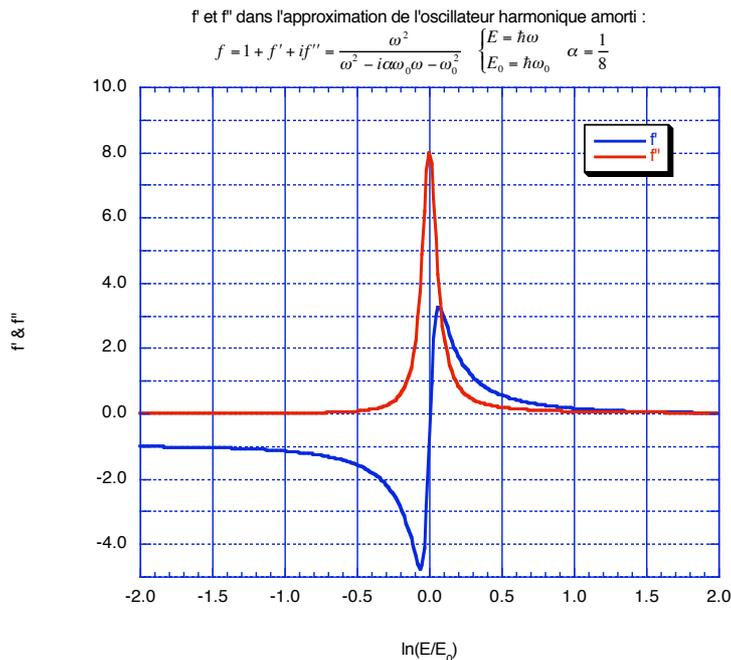


Figure 1.2.1 : Variation avec l'énergie d'excitation des parties réelles et imaginaires  $f'$  et  $f''$  du facteur de diffusion avec le modèle de l'oscillateur harmonique amorti.

De manière générale, le facteur de diffusion atomique  $f$  est un nombre complexe qui dépend de la longueur d'onde du rayonnement incident. On l'écrit :

$$f = f^o + f'(\lambda) + if''(\lambda) \quad (\text{avec } f^o, f' \text{ et } f'' \text{ réels}).$$

Lorsqu'il y a absorption, le rayonnement incident est absorbé par les électrons des couches inférieures qui sont excités vers des états d'énergie plus élevée. Il en résulte un changement de phase entre le rayonnement incident et le rayonnement diffracté. La composante imaginaire du facteur de diffusion,  $f''$ , est liée au coefficient linéaire d'absorption  $\mu$  par le théorème optique (Mie, 1908), qui peut s'écrire :

$$f'' = \frac{1}{2r_0\lambda} \frac{M}{N_A} \frac{\mu}{\rho},$$

où  $N_A$  est le nombre d'Avogadro,  $r_0$  le rayon classique de l'électron,  $\lambda$  la longueur d'onde du rayonnement,  $\mu/\rho$  le coefficient d'atténuation massique et  $M$  la masse atomique de l'élément considéré.

Cette relation permet d'obtenir les variations de  $f''$  avec l'énergie du rayonnement à partir des variations du coefficient d'absorption, obtenues expérimentalement, par exemple, à partir d'un spectre d'absorption (figure 1.2.2).

Les variations de  $f''$  avec l'énergie étant connues on obtient  $f'$  à partir de la relation de dispersion de Kramers-Kronig (de Kronig, 1926 ; Kramers, 1927) :

$$f'(E) = \frac{2}{\pi} \int_0^\infty \frac{E' f''(E') dE'}{E^2 - E'^2}.$$

Les valeurs calculées par la mécanique quantique pour  $f'$  et  $f''$  ne sont correctes qu'en dehors des seuils d'absorption. Au voisinage de ces seuils, elles doivent être déterminées de manière expérimentale.

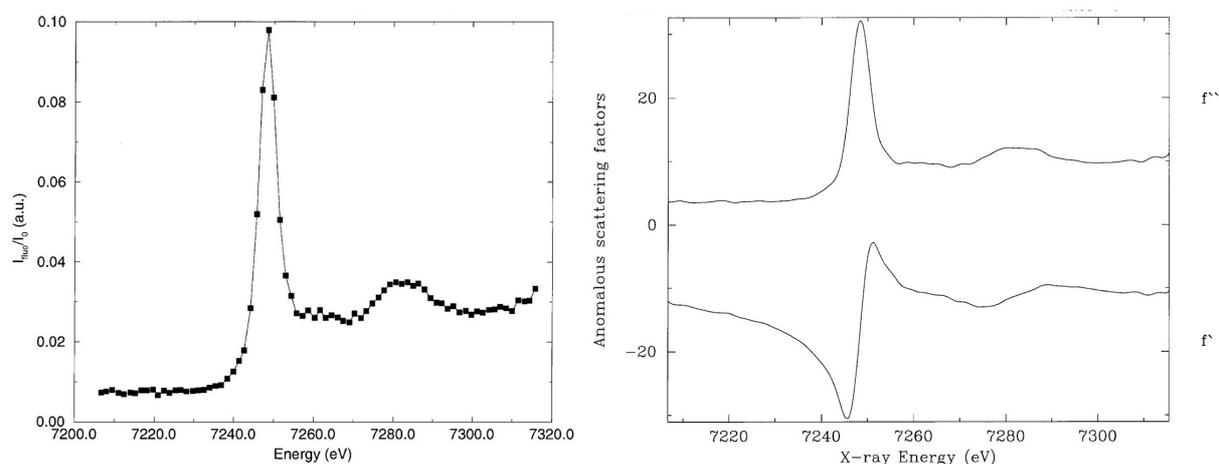


Figure 1.2.2 : Spectre d'absorption, enregistré en mode fluorescence à l'ESRF sur une solution contenant un complexe de gadolinium au seuil  $L_{III}$  du Gd (gauche) et les facteurs de diffusion anomaux  $f'$  et  $f''$  du gadolinium en fonction de l'énergie du rayonnement incident obtenus à partir de ce spectre (droite).

Dans le cas du seuil  $L_{III}$  des lanthanides, les électrons de la couche L sont excités vers un état dans la région d'énergie continue (continuum). Comme on peut le voir sur la figure 1.2.2, le coefficient d'absorption des lanthanides présente un pic important dans ce seuil d'absorption, tout comme  $f'$  et  $f''$ . Ce pic est appelé " raie blanche ". Les raies blanches sont dues à une densité élevée d'états finaux de la transition. En règle générale, l'allure de la raie blanche est fortement influencée par l'environnement moléculaire de l'atome considéré (Brown *et al.*, 1997). Cette influence est cependant peu importante pour le seuil  $L_{III}$  des lanthanides. Les variations de  $f''$  avec l'énergie du rayonnement X incident sont obtenues à partir du spectre d'absorption de l'élément considéré enregistré en mode fluorescence autour du seuil d'absorption (figure 1.2.2). Ce spectre d'excitation de fluorescence est équivalent à un spectre d'absorption, car la fluorescence est directement proportionnelle à l'absorption : les électrons de l'élément considéré, excités par l'absorption du rayonnement incident laissent des lacunes dans les couches inférieures qui sont remplies par des électrons de couches de plus haute énergie, qui émettent alors du rayonnement de fluorescence. Les facteurs de diffusion anomaux  $f'$  et  $f''$  sont obtenus à partir du spectre d'absorption à l'aide du programme CHOOCH (Evans & Pettifer, 2001). Le résultat est représenté dans la partie droite de la

figure 1.2.2. On peut remarquer que le maximum de  $f''$  est décalé d'environ 4 eV par rapport au minimum de  $f'$ .

Dans le cas des lanthanides, du fait que la position et l'allure de la raie blanche sont peu sensibles à l'environnement local de l'ion de lanthanide, il suffit de déterminer les valeurs de  $f'$  et  $f''$  en fonction de l'énergie du rayonnement incident pour l'élément d'intérêt indépendamment de l'état de complexation de l'ion. Cette détermination peut donc se faire soit avec une solution contenant l'élément considéré soit avec le cristal dérivé contenant l'atome lourd.

## I.2.2. La brisure de la loi de Friedel

La loi de Friedel, qui exprime que, dans les conditions normales, les réflexions d'indices de Miller  $(hkl)$  et  $(\bar{h}\bar{k}\bar{l})$  sont d'intensités égales, ne s'applique plus lorsque la diffusion anormale est prise en compte.

Pour des molécules ne contenant que des atomes pour lesquels la contribution anormale est faible, on peut négliger la partie imaginaire du facteur de diffusion atomique. Dans ces conditions, le facteur de diffusion  $f_j$  de l'atome  $j$  et son complexe conjugué  $f_j^*$  sont égaux ( $f_j^* = f_j$ ).

On considère les réflexions  $(hkl) \equiv \mathbf{h}$  et  $(\bar{h}\bar{k}\bar{l}) \equiv -\mathbf{h}$ . On a :

$$F(\mathbf{h}) = F(+)= \sum_j f_j \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j),$$

$$F(-\mathbf{h}) = F(-)= \sum_j f_j \exp(-i2\pi\mathbf{h} \cdot \mathbf{r}_j) \text{ et}$$

$$F^*(-\mathbf{h}) = F^*(-)= \sum_j f_j^* \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j), \text{ où } F^*(-\mathbf{h}) \text{ désigne le complexe conjugué de } F(-\mathbf{h}).$$

Si, pour tous les atomes,  $f_j^* = f_j$ , on voit que  $F^*(-) = F(+)$ . Il en résulte  $I(+)= F(+)\cdot F^*(+)= F^*(-)\cdot F(-)= I(-)$  : les intensités des réflexions  $(hkl)$  et  $(\bar{h}\bar{k}\bar{l})$  (paire de Friedel) sont égales ; elles obéissent à la loi de Friedel.

Si  $f_j$  est complexe,  $F^*(-)$  n'est plus égal à  $F(+)$ , et, dans le cas général,  $I(+)$  n'est plus égale à  $I(-)$  : la loi de Friedel est rompue (figure 1.2.3).

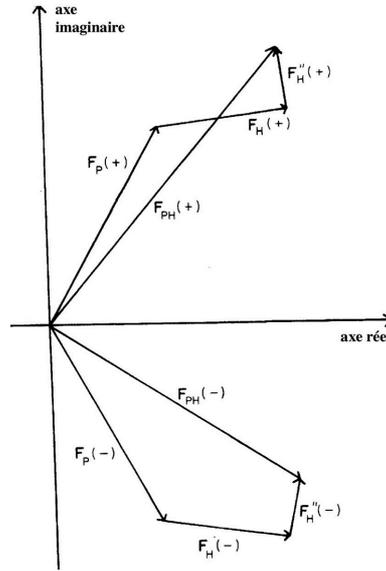


Figure 1.2.3 : Diagramme démontrant la brisure de la loi de Friedel, où "P" est la contribution des atomes qui diffractent normalement et "H" la contribution des atomes qui diffractent anormalement.

$$\text{avec } F_H = \sum_j [f_j^0 + f_j'(\lambda)] \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j) \text{ et } F_H'' = i \sum_j f_j''(\lambda) \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j)$$

On appelle "différence anormale" la quantité  $D_{ANO} = |F(+)| - |F(-)|$ . Pour des réflexions à phase contrainte (centric reflections), la différence anormale est nulle.

L'intensité des effets anomaux à une longueur d'onde donnée peut être estimée à l'aide du rapport de Bijvoet  $\langle |D_{ANO}| \rangle / \langle |F| \rangle$ , où  $|F| = \frac{1}{2} (|F(+)| + |F(-)|)$  est la valeur moyenne de l'amplitude de la paire de Friedel et où la notation  $\langle x \rangle$  représente la racine carrée de la valeur quadratique moyenne de la grandeur  $x$  :  $\langle x \rangle \equiv \sqrt{\overline{x^2}}$ .

Lorsque l'effet anormal est dû pour l'essentiel à un seul type de diffuseur, le rapport de Bijvoet est proportionnel à la partie imaginaire,  $f''$ , du facteur de diffusion de cet atome. Pour obtenir des différences anormales importantes, il faut choisir une longueur d'onde de manière à obtenir la valeur de  $f''$  la plus élevée possible.

### I.2.3. Les différences dispersives

Le phasage par la méthode MAD repose sur les différences entre paires de Friedel mesurées à une seule longueur d'onde et les différences entre paires de facteurs de structure mesurés à deux longueurs d'ondes différentes dues aux variations de la partie réelle et de la partie imaginaire du facteur de diffusion atomique avec la longueur d'onde.

Les différences dispersives des facteurs de structure mesurés à différentes longueurs d'onde découlent de la variation de la partie réelle du facteur de diffusion atomique avec la longueur

$$\text{d'onde : } \frac{D_{dis}}{|F|} = \frac{|\lambda F| - |\lambda' F|}{|F|} \propto \frac{|\lambda f' - \lambda' f'|}{Z_{eff}} \sqrt{\frac{N_H}{2N_P}} \text{ (où les } |\lambda F| \text{ sont moyennés sur les réflexions +)}$$

et -). Afin d'obtenir des rapports dispersifs les plus élevés possible, pour une expérience MAD on est amené à enregistrer des données au minimum de  $f'$  en plus de celles enregistrées au maximum de  $f''$  (pour rendre maximales les différences anormales). Généralement, un troisième jeu de données est enregistré à une longueur d'onde loin du seuil d'absorption.

### I.3. Détermination de la position des atomes lourds et phasage

La détermination des sites de fixation des diffuseurs anormaux peut se faire soit par méthode directe soit par méthode de Patterson. Le principe de l'utilisation de la fonction de Patterson anormale pour déterminer la position des diffuseurs anormaux est donnée dans l'annexe I.

#### I.3.1 Phasage avec la méthode SAD

Dans le cadre de la méthode SAD, le calcul de la phase des facteurs de structure à partir des positions des diffuseurs anormaux connues et des données de diffraction ( $|F(+)|$  et  $|F(-)|$ ) se fait de la manière suivante : connaissant les positions  $\mathbf{r}_j$  des atomes lourds dans la maille et connaissant le facteur de diffusion atomique  $f = f^o + f' + if''$ , on calcule les contributions des atomes lourds au facteur de structure

$$F_H = \sum_j (f_j^o + f_j') \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j) \text{ et}$$

$$F_H'' = i \sum_j f_j'' \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j)$$

Comme la contribution des atomes diffusant normalement peut s'écrire :

$$F_P = \sum_j f_j \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j) = |F_P| \exp(i\alpha_P)$$

où la somme est étendue aux seuls atomes diffusant normalement, on a :

$$F(+)=F_P+F_H+F_H'' \text{ et}$$

$$F^*(-)=F_P+F_H-F_H''$$

où  $F^*(-)$  représente le complexe conjugué de  $F(-)$

On en déduit :

$$F_P = F(+)-(F_H+F_H'') = F^*(-)-(F_H-F_H'')$$

D'où la méthode graphique pour trouver la phase de  $F_P$ , décrite par le diagramme de Harker de la figure 1.3.1 :

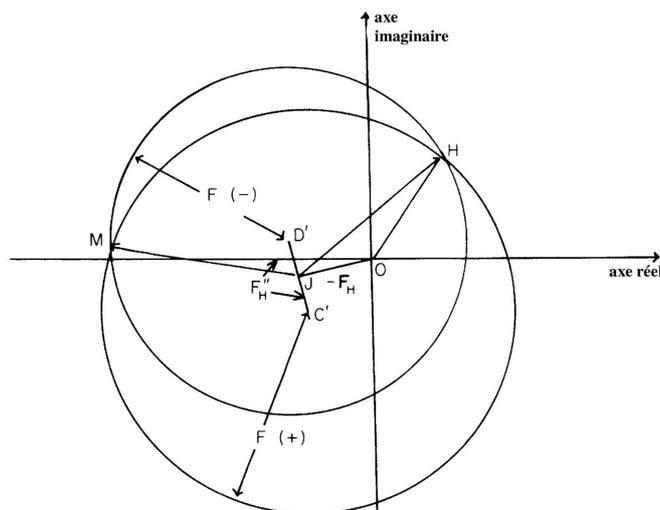


Figure 1.3.1 : Diagramme de Harker, illustrant la détermination des phases dans le cas de la diffusion anormale.

Les cercles des rayons  $|F(+)|$  et  $|F(-)|$  et de centres respectifs  $C'$ , d'affixe  $-(F_H + F_H'')$  et  $D'$ , d'affixe  $-(F_H - F_H'')$  sont tracés. Les deux points d'intersection M et H qui en résultent indiquent les deux solutions possibles, OH et OM, pour  $F_P$  ce qui correspond à deux phases  $\alpha_P$  possibles.

Dans le cas de la méthode MAD, l'ambiguïté dans la solution de phase est levée en utilisant des données enregistrées à des longueurs d'onde différentes, pour lesquelles les contributions  $F_H$  et  $F_H''$  sont suffisamment différentes : les quatre à six cercles ainsi obtenus ne se recoupent qu'en un point unique (aux erreurs expérimentales près).

Nous avons utilisé le programme SHARP (de La Fortelle & Bricogne, 1997; Bricogne et al., 2003) pour obtenir l'information de phase à partir des données expérimentales. Ce programme calcule, pour chaque réflexion, la distribution de probabilité dans le plan complexe du facteur de structure. Pour le calcul des cartes de densité électronique, on en déduit l'information nécessaire pour calculer les coefficients de Fourier "optimaux" (Blow & Crick, 1959). Pour inclure les phases expérimentales dans l'affinement par CNS (Brünger *et al.*, 1998), on exprime l'information de phase sous forme de coefficients de Hendrickson & Lattman (Hendrickson & Lattman, 1970).

### I.3.2. Procédures d'amélioration de phase

Avant de calculer la carte de densité électronique expérimentale, les phases initiales obtenues après le phasage sont améliorées par des procédures complémentaires mettant en œuvre principalement des modifications de densité électronique dans l'espace direct. Ceci est particulièrement important pour le phasage par les méthodes SAD ou SIR où l'information de phase initiale est fortement bimodale. Les modifications de densité dans l'espace direct permettent de lever l'ambiguïté en faisant ressortir la phase correspondant à la densité modifiée. L'enveloppe de la protéine, limitant la zone de solvant, est déterminée à partir des phases à basse résolution, en tenant compte la proportion de solvant dans la maille. Le nivellement de solvant consiste à fixer à une valeur constante, positive, la densité électronique correspondant au solvant. La correspondance d'histogramme consiste à ajuster la distribution des valeurs de la densité électronique observée à la distribution des valeurs qui est attendue

pour la densité électronique à une résolution donnée. Enfin, la densité électronique peut être grandement améliorée en tenant compte d'éventuelles symétries non-cristallographiques. À partir de la densité modifiée, on calcule de nouvelles phases que l'on combine avec les phases expérimentales.

### I.3.3 Le signal anomal

Pour la réussite d'un phasage reposant sur les différences anormales, il est primordial que le signal anomal soit supérieur à l'incertitude des mesures.

L'intensité des effets anomaux à une longueur d'onde donnée peut être estimée à l'aide du rapport de Bijvoet calculé à partir des amplitudes des réflexions  $(hkl)$  et  $(\bar{h} \bar{k} \bar{l})$  (Hendrickson & Ogata, 1997; Dauter *et al.*, 2002) :

$$(I) \quad \frac{\langle \Delta F^\pm \rangle}{\langle |\tilde{F}| \rangle} = \frac{\langle ||F(+)| - |F(-)|| \rangle}{\langle |\tilde{F}| \rangle} = \frac{1}{f^0(\theta)_{eff}} \sqrt{\frac{2 \sum q_j^2 f_j''^2}{N_P}} \quad \text{avec } |\tilde{F}| = \frac{|F(+)| + |F(-)|}{2} \text{ soit}$$

$$(II) \quad \frac{\langle \Delta F^\pm \rangle}{\langle |\tilde{F}| \rangle} = \frac{1}{Z_{eff}} \left( \frac{2 \sum q_j^2 f_j''^2}{N_P} \right)^{1/2} \quad \text{pour } \theta = 0 \text{ et}$$

$$(III) \quad \frac{\langle \Delta F^\pm \rangle}{\langle |\tilde{F}| \rangle} = \frac{f_j''}{f^0(\theta)_{eff}} \sqrt{\frac{2N_j}{N_P}} \quad \text{pour un seul type de diffuseurs anomaux avec des sites de fixation pleinement occupés.}$$

avec :

$f^0(\theta)_{eff}$  : Valeur moyenne du facteur de diffusion atomique de la protéine native en fonction de l'angle de Bragg  $\theta$ ,

$N_j$  : Nombre de diffuseurs anomaux par molécule de protéine,

$N_P$  : Nombre d'atomes dans la protéine,

$q_j$  : Occupation du site de fixation  $j$  par le diffuseur anomal,

$Z_{eff}$  : Valeur moyenne du facteur de diffusion atomique de la protéine native pour  $\theta = 0$ .  $Z_{eff} \approx 6,6e^-$

#### I.3.3.a. L'importance du taux d'occupation des sites et du diffuseur anomal choisi

Selon (II), afin d'obtenir un signal anomal maximal, il est important que les sites de fixation des diffuseurs anomaux soient fortement occupés, et que le diffuseur anomal et la longueur d'onde soient choisis de manière à obtenir la valeur de  $f''$  la plus élevée possible. Ceci est d'autant plus important que la structure à résoudre est de grande taille.

#### I.3.3.b. L'importance de la résolution des données

La résolution des données de diffraction, donnée par l'angle de Bragg maximal des taches de diffraction  $d_{reso} = \frac{\lambda}{2 \sin \theta_{max}}$ , est imposée par la qualité cristalline ou, éventuellement, par la géométrie de l'expérience.

Étant donné que des différentes contributions au facteur de diffusion atomique  $f(\theta, \lambda) = f^0(\theta) + f'(\lambda) + if''(\lambda)$  seule  $f^0(\theta)$  décroît fortement avec l'angle de diffraction, le

signal anomal devrait être, selon la relation (III), relativement plus élevé à plus haute résolution. Cet effet est cependant limité par l'agitation thermique de l'ensemble des atomes et, en particulier, des diffuseurs anomaux, qui fait qu'à haute résolution les réflexions sont beaucoup moins intenses et que leur intensité est mesurée avec une précision moindre que celle des réflexions à basse résolution.

C'est pourquoi, pour l'identification des sites de fixation des atomes lourds, les meilleurs résultats sont souvent obtenus en limitant la résolution des données utilisées à une résolution plus basse que la résolution maximale des données (Dauter & Nagem, 2002). Ceci est particulièrement vrai lorsque l'on utilise les méthodes directes qui utilisent des facteurs de structure normalisés, plus sensibles aux erreurs à haute résolution.

D'un autre côté, une haute résolution des données est très importante pour interpréter la carte de densité électronique expérimentale et pour la construction du modèle.

### **I.3.3.c. L'importance de la redondance des données**

Le signal anomal étant faible par rapport aux intensités de diffraction, il est primordial, pour réussir le phasage, que le signal anomal soit significatif devant l'inexactitude des mesures. Au même titre qu'il faut maximiser le signal anomal, il faut donc maximiser l'exactitude des mesures, exactitude qui peut se mesurer par le rapport signal sur bruit,  $\frac{I}{\sigma}$ , des mesures. Le rapport signal sur bruit dépend, bien sûr, de la qualité cristalline, mais il dépend aussi des conditions d'enregistrement : un temps de pose élevé, et donc un nombre élevé de coups  $N$  par réflexion, améliore la précision statistique des données avec  $\frac{I}{\sigma}$  approximativement proportionnel à  $\sqrt{N}$ .

Une redondance élevée des données (c'est à dire la mesure indépendante, répétée, de l'intensité de réflexions dont l'équivalence est imposée par symétrie cristalline) permet d'améliorer non seulement la précision statistique des mesures, mais améliore surtout leur exactitude, en éliminant les erreurs systématiques dues aux artefacts de l'expérience (Dauter & Nagem, 2002). C'est pourquoi, pour un phasage utilisant le signal anomal, il est essentiel d'enregistrer des données de diffraction avec une redondance élevée.

## **I.4. Obtention de cristaux dérivés pour le phasage par des méthodes de remplacement isomorphe et par des méthodes de diffusion anormale**

### **I.4.1. Choix de l'atome lourd et de la longueur d'onde du rayonnement X incident**

#### **I.4.1.a. Critères de diffusion**

Selon la méthode de phasage employée, différents critères s'imposent pour le choix de l'atome lourd et de la longueur d'onde du rayonnement incident.

Pour le remplacement isomorphe, l'atome avec le numéro atomique  $Z$  le plus élevé maximise les différences isomorphes, et ceci indépendamment de la longueur d'onde.

Pour des méthodes utilisant la diffusion anormale, on choisit l'atome, avec une valeur élevée de  $f''$  à la longueur d'onde disponible. Pour la méthode MAD, ceci nécessite donc que l'un des seuils d'absorption du diffuseur anomal soit accessible à l'expérience, pour que la variation des facteurs anomaux  $f'$  et  $f''$  soit importante. Pour la méthode SAD, dans le cas où on dispose d'une source de rayons X de longueur d'onde ajustable qui permette d'atteindre le seuil d'absorption du diffuseur anomal, on se placera dans le seuil, au maximum de  $f''$ . Si le seuil d'absorption du diffuseur anomal est inaccessible, et dans la mesure où la valeur de  $f''$  est suffisante, il est cependant possible de travailler hors seuil, avec la méthode SAD, en utilisant une source de rayons X à longueur d'onde fixe. Dans le cas où on utiliserait des diffuseurs anomaux intrinsèques des cristaux natifs (métalloprotéines, phasage au soufre) on essaie de choisir la longueur d'onde qui permet d'obtenir la valeur de  $f''$  la plus élevée possible.

#### **I.4.1.b. Critères chimiques**

Outre ses propriétés physiques de diffusion des rayons X, il est important que l'atome lourd choisi se laisse facilement insérer dans le cristal de protéine. Il faut donc trouver un composé incluant l'atome lourd et qui se lie avec une occupation élevée dans des sites spécifiques de la macromolécule biologique, sans détruire l'ordre cristallin de l'échantillon. Dans le cas des méthodes de remplacement isomorphe, il faut en outre que le cristal dérivé soit isomorphe au cristal natif. L'avantage des diffuseurs anomaux intrinsèques est évident : aucun effort supplémentaire n'est nécessaire pour la préparation du dérivé et les sites des diffuseurs anomaux sont, le plus souvent, fortement occupés.

#### **I.4.2. Accessibilité du seuil, grandes longueurs d'onde**

Sur la plupart des lignes de lumière dédiées à la méthode MAD les longueurs d'onde des rayons X accessibles se situent entre 0,7 Å et 2,0 Å. Pour certains éléments d'intérêt, les seuils d'absorption se trouvent dans le domaine des rayons X mous :

- les seuils K du phosphore P ( $\lambda = 5,78$  Å), du soufre S ( $\lambda = 5,02$  Å), du chlore Cl ( $\lambda = 4,40$  Å) et du calcium Ca ( $\lambda = 3,07$  Å),
- le seuil  $L_I$  du xénon Xe ( $\lambda = 2,27$  Å),
- le seuil  $M_V$  de l'uranium U ( $\lambda = 3,49$  Å).

Le problème majeur lorsqu'on utilise de grandes longueurs d'onde est l'absorption. En effet, le terme dominant du coefficient linéaire d'absorption est proportionnel à  $\lambda^3$  (Victoreen, 1949). Ainsi pour des longueurs d'onde supérieures à 2,0 Å, le rayonnement est très fortement absorbé par l'échantillon, mais aussi par l'air : l'intensité diffractée est fortement atténuée, ce qui dégrade la qualité des mesures et nécessite des corrections d'absorption. Le rayonnement absorbé par l'échantillon induit en outre des dégâts d'irradiation qui font évoluer les intensités diffractées avec le temps. Un autre inconvénient des grandes longueurs d'ondes est, qu'à résolution égale, les angles de Bragg sont plus grands : la résolution maximale imposée par les bords du détecteur est alors souvent inférieure à la limite de diffraction du cristal. Ainsi, en se plaçant au seuil d'absorption  $L_{III}$  du Gd ( $\lambda = 1,711$  Å) sur la ligne BM30A de l'ESRF, avec un détecteur de 165 mm de diamètre placé à la distance minimale de 110 m de l'échantillon, la résolution maximale atteinte n'est que de 2,7 Å.

Pour atteindre les seuils K d'éléments lourds, les rayons X très durs peuvent être utilisés (Schiltz *et al.*, 1997). Cependant, un des inconvénients des rayons X très durs est la faible efficacité quantique de détection des détecteurs usuels à ces énergies. Un autre inconvénient est l'importance que prend, à des énergies supérieures à 30 keV, la diffusion Compton par

rapport à la diffusion élastique des rayons X, ce qui augmente le bruit de fond de manière significative.

### I.4.3. Seuils d'absorption et raies blanches

La figure 1.4.1 montre l'évolution de la longueur d'onde des seuils d'absorption K, L et M en fonction du numéro atomique  $Z$  des éléments du tableau périodique. On voit que les seuils K des éléments légers situés entre le manganèse Mn ( $Z = 25$ ) et l'yttrium Y ( $Z = 39$ ) sont facilement accessibles ( $\lambda = 1,89-0,73 \text{ \AA}$ ). Les seuils  $L_{III}$  des éléments situés entre le samarium Sm ( $Z = 62$ ) et l'uranium U ( $Z = 92$ ) sont également facilement accessibles ( $\lambda = 1,85 -0,72 \text{ \AA}$ ).

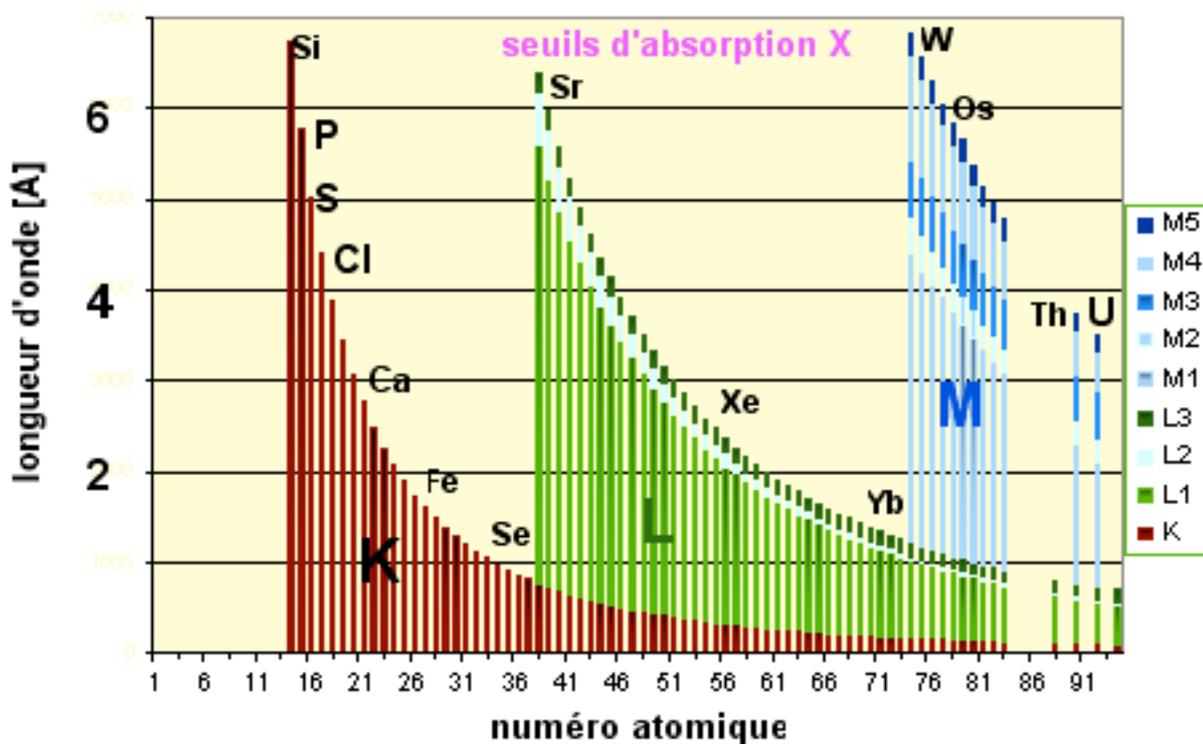


Figure 1.4.1 : Seuils d'absorption K, L, et M pour les éléments stables du tableau périodique.

La partie imaginaire  $f''$  du facteur de diffusion atomique est de l'ordre de  $4 e^-$  pour la plupart des seuils K et de l'ordre de  $10 e^-$  pour les seuils  $L_{III}$  (Müller & Heinemann, 2005), à l'exception de quelques éléments présentant une raie blanche dans leur seuil. C'est le cas pour les seuils K du Mn ( $f'' \sim 10 e^-$ ), du Se ( $f''$  entre  $6 e^-$  et  $10 e^-$ , selon l'état d'oxydation) et les seuils  $L_{III}$  des lanthanides avec des valeurs de  $f''$  pouvant atteindre  $30 e^-$ . Pour des expériences hors seuil, il est plus judicieux de travailler à une longueur d'onde inférieure à celle du seuil, puisque  $f''$  ne décroît que lentement pour  $\lambda < \lambda_{\text{seuil}}$ . Ainsi pour le rayonnement  $K\alpha$  du cuivre, on obtient des valeurs de  $f''$  de  $2 e^-$  à  $4 e^-$  pour les éléments situés entre le vanadium, V, et le cobalt, Co, des valeurs de  $f''$  variant entre  $9,6 e^-$  et  $13,3 e^-$  pour les lanthanides qui vont du lanthane, La, au dysprosium, Dy, et une valeur de  $f'' = 16 e^-$  pour l'uranium, U.

## **I.4.4. Revue de méthodes habituelles pour obtenir des dérivés pour le phasage par des méthodes de remplacement isomorphe ou de diffusion anormale**

### **I.4.4.a. Utilisation de protéines sélénées ou d'acides nucléiques bromés**

La méthode de phasage *de novo* la plus utilisée avec le rayonnement synchrotron est la méthode MAD appliquée à des cristaux de protéines sélénées (Hendrickson *et al.*, 1990 ; Dauter & Nagem, 2002). La méthode consiste à exprimer la protéine d'intérêt en incorporant de la sélénométhionine au lieu des méthionines, l'atome de soufre de cet acide aminé étant remplacé par un atome de sélénium. Les avantages de la méthode sont l'occupation complète des sites de sélénium et le signal anormal relativement élevé, allant de  $6 e^-$  à  $10 e^-$  selon l'état d'oxydation du sélénium, au seuil K du Se. Ce seuil, dont la longueur d'onde est  $\lambda = 0,98 \text{ \AA}$ , est très facilement accessible avec les lignes de lumière habituelles, pour lesquelles, par ailleurs, l'intensité a, le plus souvent, été optimisée à cette longueur d'onde. L'inconvénient majeur de cette méthode est la mise en pratique relativement laborieuse. Il faut exprimer la protéine en milieu sélénié, ce qui conduit souvent à de faibles taux d'expression. La protéine sélénée cristallise souvent avec des conditions modifiées par rapport à la protéine native. Certaines protéines eucaryotes ne peuvent pas être exprimées dans des systèmes bactériens et d'autres systèmes d'expression incorporent généralement mal la sélénométhionine. Enfin, certaines protéines ne contiennent pas de méthionines.

Au même titre que l'utilisation des sélénométhionines pour résoudre la structure de protéines, on utilise des nucléotides bromés ou iodés pour résoudre la structure d'acides nucléiques ou de leurs complexes avec des protéines.

### **I.4.4.b. La diffusion anormale des atomes intrinsèques**

Une méthode consiste à se servir des diffuseurs anormaux naturellement présents dans les protéines, tels les métaux de transition dans les métalloprotéines, ou le soufre présent dans toutes les protéines qui contiennent des méthionines ou des cystéines. Ceci présente l'avantage de sites fortement, sinon totalement, occupés et ne nécessite pas de préparation de dérivé.

#### **Métalloprotéines**

Plus de 20% des protéines contiennent intrinsèquement des métaux comme le Cu, le Fe, le Mn, le Ni et le Zn (Leslie *et al.*, 2004). Les seuils d'absorption K d'un grand nombre de ces métaux sont situés entre 0,7 et 2,0  $\text{\AA}$  avec des valeurs de  $f''$  de l'ordre de  $4 e^-$ . Ainsi, ces protéines se prêtent à des expériences MAD ou SAD au seuil K, en utilisant le rayonnement synchrotron. Certains des éléments présentent une diffusion anormale non négligeable pour le rayonnement  $K\alpha$  du Cu et se prêtent donc à des expériences SAD avec une source de laboratoire. Avec une source de rayons X de longueur d'onde variable, un spectre de fluorescence d'émission enregistré avec une énergie d'excitation supérieure à l'énergie des seuils d'absorption des éléments couramment présents, permet de détecter et d'identifier d'éventuels métaux présents dans la protéine d'intérêt.

#### **Méthode SAD utilisée avec le signal anormal du soufre**

Toutes les protéines contenant des cystéines ou des méthionines contiennent du soufre. Ces deux acides aminés représentent en moyenne 4% des acides aminés. Par la méthode SAD, on

peut utiliser la diffusion anormale hors seuil (le seuil K du soufre étant à 5,02 Å). C'est en 1981 que la première structure a été résolue avec le signal anormal du soufre (Hendrickson & Teeter, 1981). Depuis 1999, de nombreuses structures ont été résolues par la méthode SAD en se servant uniquement du signal anormal du soufre (Ramagopal *et al.*, 2003b). Toutefois, cette méthode ne s'applique, pour le moment, qu'à des cristaux qui diffractent très bien. En effet la diffusion anormale du soufre à des longueurs d'onde habituelles est très faible (pour  $\lambda_{Cu} = 1,5 \text{ \AA}$  et  $\lambda_{Cr} = 2,3 \text{ \AA}$  les valeurs respectives de  $f''$  sont  $0,56 e^-$  et  $1,14 e^-$ ). Comme  $f''$  augmente avec la longueur d'onde utilisée, il faut trouver un compromis entre une valeur plus élevée de  $f''$  et les problèmes liés à l'utilisation de rayons X mous. Une longueur d'onde optimale avec une ligne de lumière non-optimisée pour les grandes longueurs d'onde serait de  $\lambda \sim 1,8 \text{ \AA}$  (Leslie *et al.*, 2004). Ces conditions conduisent à des rapports de Bijvoet très faibles, autour de 1% (Leslie *et al.*, 2004). Utiliser ce faible signal anormal nécessite des données d'une très grande exactitude, les données doivent donc être enregistrées avec une redondance très élevée. Ceci demande des cristaux qui diffractent bien et qui résistent aux dégâts d'irradiation. Le développement de la méthode va de pair avec le développement auprès des sources de rayonnement synchrotron de lignes de lumière adaptées à l'utilisation de grandes longueurs d'onde, associé au développement des traitements de données appropriés.

### **Méthode SAD utilisée avec le signal anormal du phosphore**

Les acides nucléiques contiennent une quantité constante de phosphore (1 atome de P par base). Au même titre que le phasage au soufre s'applique aux protéines, le signal anormal du phosphore peut servir pour résoudre la structure d'acides nucléiques ou de leurs complexes avec des protéines. Ceci conduit à un rapport de Bijvoet de 2% pour un polynucléotide avec le rayonnement  $K\alpha$  du cuivre (Dauter, 2002).

#### **I.4.4.c. Introduction d'atomes lourds par trempage ou cocrystallisation**

Pour obtenir des dérivés pour les méthodes MAD ou MIR, on utilise classiquement les sels des éléments lourds Pt, Au, Hg, U ou Pb, tels les sels suivants ("magic seven") :  $K_2PtCl_4$ ,  $KAu(CN)_2$ ,  $K_2HgI_4$ ,  $UO_2(C_2H_3O_2)_2$ ,  $HgCl_2$ ,  $K_3UO_2F_5$  ou le sulfate para-chloromercurybenzoïque (PCMBs) (Garman & Murray, 2003). Le seuil  $L_{III}$  de chacun de ces éléments est très accessible avec une valeur de  $f''$  d'environ  $10 e^-$ , sauf pour le platine qui présente une raie blanche dans ce seuil avec une valeur de  $f''$  supérieure à  $20 e^-$ . Si on utilise une source de rayons X de laboratoire avec une anode de cuivre les valeurs de  $f''$  sont comprises entre  $7 e^-$  et  $16 e^-$ , cette dernière valeur étant obtenue avec l'uranium. Les diffuseurs lourds sont introduits classiquement dans les cristaux par des trempages relativement longs (de quelques heures à plusieurs jours) des cristaux natifs dans des solutions peu concentrées du sel (0,1-10mM). L'inconvénient principal de la méthode est le mode de fixation souvent aléatoire des différents composés, ce qui rend nécessaires des nombreux essais répétés avec différents composés pour obtenir des cristaux dérivés. Cependant, la connaissance de la séquence de la protéine peut aider à orienter le choix des sels d'atomes lourds à tester. Ainsi les composés du mercure se lient de préférence à des cystéines libres ou à l'azote des histidines et les composés du platine aux cystéines, aux histidines ou aux méthionines (Garman & Murray, 2003). Cependant, l'utilisation de ces sels d'atomes lourds n'est pas toujours aisée. Les sels sont généralement très faiblement solubles et instables dans des solutions aqueuses. Ils sont incompatibles avec de nombreuses conditions chimiques de cristallisation. Ils sont souvent mal tolérés par les cristaux de protéine. Ils sont hautement toxiques. Depuis peu, des trempages beaucoup plus courts (10min-2h) sont pratiqués avec ces mêmes composés à des concentrations plus élevées (> 10mM), ce qui semble donner des

dérivés de meilleure qualité cristalline, avec un meilleur isomorphisme, qu'avec des trempages longs (Dauter, 2002). Des trempages courts sont également pratiqués avec des ions qui ne forment pas de liaison spécifique avec la protéine, comme des ions métalliques non complexés, Ru ou Cs, ou des halogénures. Ainsi, des trempages rapides (10-30s) dans des solutions d'iodure ou d'iode dans l'iodure I<sub>2</sub>/KI peuvent produire des dérivés incorporant les ions I, I<sub>3</sub><sup>-</sup> ou I<sub>5</sub><sup>-</sup> (Evans & Bricogne, 2002). Les solutions d'iodures peuvent éventuellement être introduites dans la solution cryoprotectrice. Les dérivés d'iodures sont utiles pour des expériences SAD ou SIRAS (la valeur de f'' est de 7 e<sup>-</sup> pour l'iode avec le rayonnement K $\alpha$  du cuivre). De même, des dérivés de bromures peuvent servir pour des expériences MAD (seuil K :  $\lambda = 0,92 \text{ \AA}$ ). Les halogénures se fixent dans des sites multiples, occupés avec des taux variables, qui sont situés à la surface protéique, dans les régions de solvant ordonné, sur des sites hydrophobes ou par liaison H avec des groupes polaires ou des molécules d'eau (Evans & Bricogne, 2002). Les ions de lanthanides sont très utilisés pour des protéines à Ca. En effet, leurs sphères de coordination et leur environnement sont très similaires et les ions de lanthanide remplacent facilement le calcium dans ses sites de fixation à la protéine. Dans un tel cas, on peut tirer profit de la diffusion anormale très élevée des lanthanides soit dans leur seuil d'absorption L<sub>III</sub>, en utilisant le rayonnement synchrotron, soit, pour nombre d'entre eux, avec une source de laboratoire (Kahn *et al.*, 1985).

#### **I.4.4.d. Utilisation des gaz nobles xénon et krypton**

On peut obtenir des cristaux dérivés pour le phasage par les méthodes SIRAS ou MIRAS, en introduisant des gaz nobles, xénon ou krypton, dans des cristaux natifs (Schiltz *et al.*, 2003). Pour cela, on soumet les cristaux à une atmosphère du gaz sous pression (1-100 bar) dans une cellule dédiée. Ces gaz étant relativement bien soluble dans l'eau, ils diffusent rapidement dans les canaux de solvant du cristal. La liaison des atomes de Xe ou de Kr met en jeu des forces faibles dans des cavités hydrophobes, des sites enzymatiques ou des cavités intermoléculaires. Les avantages de cette méthode d'obtention de cristaux dérivés sont, l'isomorphisme élevé des cristaux dérivés résultant des liaisons à faible affinité, le nombre élevé de sites potentiels de fixation et leur complémentarité par rapport aux sites de fixation de sels classiques, de nature plutôt polaire. Enfin on peut faire varier le nombre et l'occupation des sites en faisant varier la pression du gaz. La probabilité de fixation suffisante est élevée, > 30-40% (Schiltz *et al.*, 2003), et, lorsqu'il n'y a pas de fixation, des sites potentiels peuvent être introduits par des opérations de mutagenèse dirigée simple. La réversibilité de la fixation permet d'obtenir différents dérivés à partir d'un cristal unique. L'enregistrement des données peut se faire et à température cryogénique et à température ambiante. La diffusion anormale du Xe est relativement élevée pour des expériences au laboratoire ( $f''_{\text{CuK}\alpha} = 7,4 e^-$ ).

#### **I.4.5. Dégâts d'irradiation**

Avec l'utilisation de sources de rayonnement synchrotron de 3<sup>e</sup> génération qui fournissent des faisceaux très intenses et permettent d'utiliser des cristaux de plus en plus petits, les dégâts dus à l'irradiation redeviennent fortement problématiques, surtout pour des données à redondance élevée, et ce malgré la congélation des échantillons (Garman & Nave, 2002 ; Nave & Garman, 2005). Le phénomène des dégâts d'irradiation peut être abordé de différentes manières. Classiquement, lorsque la taille du cristal s'y prête, on essaie de les éviter en déplaçant le cristal dans le faisceau pour irradier un nouveau volume à chaque déplacement du cristal. De nouvelles approches proposent de tenir compte des dégâts d'irradiation lors du

traitement des données et ainsi de réduire leurs conséquences négatives, voire d'en tirer profit comme source d'information de phase.

Les effets du rayonnement ionisant peuvent être divisés en deux classes, les effets généraux, affectant globalement le cristal et les effets spécifiques, localisés. Les premiers provoquent la perte d'ordre à longue distance qui entraîne la réduction du pouvoir de diffraction du cristal. Ils sont accompagnés de l'augmentation du volume de la maille, de l'augmentation de la mosaïcité et du facteur d'agitation thermique, et de la perte des réflexions à haute résolution. Les effets localisés sont généralement spécifiques à certains sites, tels que, le clivage des ponts disulfure ou la perte de définition de groupements carboxyles. Les effets localisés concernent aussi particulièrement les sites d'atomes lourds ou anomaux du fait de leur absorption élevée ( $\propto f''$ ). Les deux classes de dégâts conduisent à une perte d'isomorphisme au cours de l'enregistrement, ce qui peut empêcher la réussite du phasage. Ce sont en premier lieu les dégâts localisés sur certains sites qui surviennent (Weik *et al.*, 2000 ; Burmeister, 2000 ; Ravelli & McSweeney, 2000) avant que le réseau cristallin ne soit affecté par les effets à longue distance. Ce sont les effets localisés qui, lorsqu'ils affectent les sites d'atomes lourds sur lesquels repose le phasage lui-même, ont une influence fatale sur la réussite du phasage. De tels cas ont été observés pour des dérivés contenant des nucléotides bromés, où la débromation, assez rapide au cours de l'irradiation, a empêché la réussite du phasage (Ennifar *et al.*, 2002 ; Schiltz *et al.*, 2004). Pour tenir compte des non-isomorphismes induits par les dégâts d'irradiation et s'en servir comme information de phase supplémentaire il existe plusieurs approches.

Ravelli *et al.* (2003) proposent une méthode de phasage reposant sur les dégâts d'irradiation (RIP : Radiation-damage Induced Phasing). Elle consiste à enregistrer deux jeux de données sur un cristal, dont un jeu complet et redondant avec le cristal non-endommagé et un deuxième jeu, après exposition du cristal à une dose élevée de rayonnement X servant à créer des dégâts d'irradiation localisés. Pour le phasage, les données sont traitées de manière pseudo-SIR(AS). Ceci nécessite l'identification et l'affinement de nombreux changements structuraux spécifiques, modélisés par une baisse d'occupation et une augmentation du facteur d'agitation thermique, B, des groupements d'atomes concernés. Pour des cristaux où les dégâts d'irradiation surviennent rapidement et apparaissent déjà dans le premier jeu, il est recommandé de diviser les données en plusieurs jeux et de les traiter tous comme des dérivés isomorphes indépendants.

Schiltz *et al.* (2004) proposent une modélisation continue de la progression des dégâts d'irradiation en utilisant le programme SHARP avec un déclin exponentiel de l'occupation  $q(d)$  des sites concernés en fonction de la dose  $d$  reçue :  $q(d) = q_0 \exp(-\beta_d d)$  (Ennifar *et al.*, 2002). Pour cette méthode, les mesures d'intensité équivalentes par symétrie ne sont pas combinées pour obtenir une valeur moyenne. À chaque observation est attribuée une valeur de dose reçue et, au lieu d'affiner un seul paramètre pour l'occupation de chaque site, deux paramètres sont affinés,  $q_0$  et  $\beta_d$ . Il est envisageable d'implémenter des modélisations plus complexes des dégâts d'irradiation sous forme de fonctions paramétrisées dans SHARP.

Le développement de procédures de phasage spécifiques et du traitement informatique correspondant pourra permettre d'utiliser les dégâts d'irradiation comme source principale d'information de phase ou comme source complémentaire associée à d'autres méthodes comme le phasage avec des dérivés bromés ou le phasage au soufre (Weiss *et al.*, 2004) ou, du moins, de limiter les effets négatifs des dégâts d'irradiation sur le phasage.

## I.4.6. Comparaison des méthodes d'obtention de dérivés anomaux

Les deux tableaux suivants résument quelques valeurs de grandeurs caractéristiques, afin de permettre de comparer les méthodes pour obtenir des dérivés anomaux.

Le premier tableau donne les rapports de Bijvoet théoriques attendues pour des conditions d'expérience types (type de dérivé anormal, longueur d'onde). Plus ce rapport est élevé, plus il sera facile d'obtenir les phases à partir des données de diffraction correspondantes. De plus, le tableau donne une estimation qualitative de la facilité de la préparation des dérivés anomaux, facteur important pour l'expérimentateur. Les rapports de Bijvoet sont calculés selon la formule suivante :

$$\frac{\Delta F^{\pm}}{F} [\%] = \frac{100}{Z_{eff}} \left( \frac{2 \sum q_j^2 f_j^{n_2}}{N_p} \right)^{1/2} = \frac{100}{6,6} \left( \frac{2 N_H q_j^2 f_j^{n_2}}{7,9 * 100} \right)^{1/2} = 0,76 \sqrt{N_H} q_j f_j^{n_2}$$

Les rapports de Bijvoet sont calculés pour  $\theta = 0$ , une occupation  $q_j$  égale de tous les sites et les valeurs suivantes :  $Z_{eff} = 6,6$ ,  $N_p = 7,9 * N_{aa}$ ,  $N_H$  : nombre de diffuseurs anomaux pour 100 acides aminés,  $N_{aa}$  : nombre d'acides aminés.

| Élément  | $N_H$ | $q_j$ | $\lambda$ [Å] | $f''$ [e <sup>-</sup> ] | $\frac{\Delta F^{\pm}}{F}$ [%] | Facilité de préparation des dérivés |
|--|-------|-------|---------------|-------------------------|--------------------------------|-------------------------------------|
| Se   | 1     | 1     | 0,98          | 8                       | 6,1                            | --                                  |
| Se   | 3     | 1     | 0,98          | 8                       | 10,5                           | --                                  |
| Br   | /     | 1     | 0,92          | 3,8                     | /                              | --                                  |
| S  | 4     | 1     | 1,54          | 0,56                    | 0,85                           | ++                                  |
| S  | 4     | 1     | Cr : 2,29     | 1,14                    | 1,7                            | ++                                  |
| P  | /     | 1     | 1,54          | 0,43                    | 2                              | ++                                  |
| Mn   | 1     | 1     | 1,54          | 2,8                     | 2,1                            | ++                                  |
| Mn   | 1     | 1     | seuil : 1,89  | 9,7                     | 7,3                            | ++                                  |
| Fe   | 1     | 1     | seuil : 1,74  | 4,5                     | 3,4                            | ++                                  |
| K <sub>2</sub> HgI <sub>4</sub>  | 1     | 1     | seuil : 1,0   | 10,1                    | 7,7                            | -                                   |
| K <sub>2</sub> HgI <sub>4</sub>  | 1     | 1     | 1,54          | 7,7                     | 5,9                            | -                                   |
| K <sub>2</sub> PtCl <sub>4</sub>   | 1     | 0,5   | seuil : 1,1   | 10,2                    | 3,9                            | -                                   |
| UO <sub>2</sub> (C <sub>2</sub> H <sub>3</sub> O <sub>2</sub> ) <sub>2</sub> | 2     | 0,25  | 1,54          | 16                      | 4,3                            | -                                   |
| Xe   | 3     | 0,3   | 1,54          | 7,4                     | 2,9                            | -                                   |
| Gd <sup>3+</sup>   | 1     | 1     | 1,711         | 28                      | 21,3                           | -                                   |
| Gd-complexe  | 2     | 0,5   | 1,711         | 28                      | 15,0                           | 0                                   |
| Gd-complexe  | 2     | 0,5   | 1,54          | 12                      | 6,4                            | 0                                   |

Tableau 1.4.1 : Rapports de Bijvoet théoriques pour différents types de dérivés anomaux à différentes longueurs d'onde. Pour la facilité de préparation : ++ : atomes intrinsèques, pas de préparation; 0 : préparation relativement simple; - : préparation délicate ou nécessitant un équipement particulier; -- : préparation laborieuse du dérivé.

Le deuxième tableau résume les valeurs de  $f''$  des diffuseurs anomaux habituellement utilisés pour les longueurs d'onde du seuil d'absorption de l'élément considéré et les longueurs d'onde du rayonnement  $K\alpha$  du Cu et du Cr.

| Z  | Élément | Seuil     | $\lambda_{\text{seuil}}[\text{Å}]$ | $f''_{\text{seuil}}[e^-]$ | $f''_{\text{CuK}\alpha} [e^-]$<br>( $\lambda=1,54\text{Å}$ ) | $f''_{\text{CrK}\alpha} [e^-]$<br>( $\lambda=2,29\text{Å}$ ) |
|----|---------|-----------|------------------------------------|---------------------------|--|--|
| 15 | P       | K         | 5,78                               | 4,1                       | 0,43   |  |
| 16 | S       | K         | 5,02                               | 4,1                       | 0,56   | 1,14   |
| 17 | Cl      | K         | 4,39                               | 4,1                       | 0,70   | 1,5  |
| 20 | Ca      | K         | 3,07                               | 4,1                       | 1,3  | 2,7  |
| 25 | Mn      | K         | 1,90                               | 9,7                       | 2,8  |  |
| 26 | Fe      | K         | 1,74                               | 4,5                       | 3,2  |  |
| 27 | Co      | K         | 1,61                               | 3,9                       | 3,6  |  |
| 28 | Ni      | K         | 1,49                               | 4,3                       | 0,51   |  |
| 29 | Cu      | K         | 1,38                               | 4,2                       | 0,59   |  |
| 30 | Zn      | K         | 1,28                               | 4,3                       | 0,68   |  |
| 34 | Se      | K         | 0,98                               | 6-10                      | 1,1  |  |
| 35 | Br      | K         | 0,92                               | 3,8                       | 1,3  |  |
| 36 | Kr      | K         | 0,87                               | 1,6-2,7                   | 1,4  |  |
| 54 | Xe      | $L_I$     | 2,27                               | 13,5                      | 7,4  | 11   |
| 64 | Gd      | $L_{III}$ | 1,71                               | 28                        | 12,0   |  |
| 78 | Pt      | $L_{III}$ | 1,07                               | 10,2                      | 6,9  |  |
| 79 | Au      | $L_{III}$ | 1,04                               | 10,2                      | 7,3  |  |
| 80 | Hg      | $L_{III}$ | 1,01                               | 10,1                      | 7,7  |  |
| 82 | Pb      | $L_{III}$ | 0,95                               | 10,1                      | 8,5  |  |
| 92 | U       | $L_{III}$ | 0,72                               | 10,3                      | 16   | 28   |
| 92 | U       | $M_V$     | 3,49                               | 120                       |  | 28   |

Tableau 1.4.2 : Longueurs d'onde des seuils d'absorption et valeurs de  $f''$  de certains éléments (Müller & Heinemann, 2005). Les valeurs de  $f''$  pour les éléments présentant une raie blanche dans leur seuil d'absorption dépendent généralement de l'environnement du diffuseur anormal, lanthanides exceptés. Les valeurs pour l'uranium proviennent de Chesne (2002).

## **I.5. Utilisation de complexes de lanthanides pour obtenir des dérivés**

Une autre approche pour introduire des atomes lourds dans les cristaux de protéine consiste à utiliser des complexes chélatant l'élément choisi. Cette approche a pour but de séparer les propriétés physiques de l'atome lourd, ici ses propriétés de diffusion des rayons X, des propriétés chimiques du complexe qui imposent son comportement dans le cristal. Les premières sont essentiellement déterminées par l'atome lourd choisi, alors que les deuxièmes sont imposées par le ligand qui chélate l'atome lourd et qui confère les propriétés chimiques voulues au complexe. Si la constante d'association du complexe formé par le ligand et l'atome lourd est élevée, cette approche peut permettre de fixer des atomes lourds en utilisant les liaisons du ligand avec la molécule de protéine. On peut ainsi concevoir une grande variété de ligands. La taille du complexe résultant devra cependant rester limitée au regard de la taille habituelle des canaux de solvant dans les cristaux de protéine. Le complexe résultant devra être compatible avec les conditions physico-chimiques utilisées pour cristalliser les protéines, et suffisamment stable dans ces conditions.

Par exemple, il est possible de concevoir des ligands fonctionnalisés pour former des liaisons covalentes avec certains résidus protéiques, tels des complexes thiol-réactifs se liant aux cystéines libres (Purdy *et al.*, 2002). Il s'est avéré qu'une telle réaction conduit le plus souvent soit à une forte baisse de la solubilité de la protéine soit à des fortes modifications des conditions de cristallisation.

Il est également possible de concevoir des complexes dont les propriétés physico-chimiques permettent de détecter facilement la fixation du complexe dans le cristal : ainsi la coloration de cristaux dérivés obtenus avec des complexes colorés permet d'estimer la fixation du complexe dans le cristal (Chesne, 2002).

Dans le cadre de ce travail, nous nous sommes servis d'une classe particulière de complexes pour des études systématiques sur un grand nombre de dérivés. Le fort pouvoir de phasage de ces complexes dans plusieurs cas avait été démontré antérieurement (Girard *et al.*, 2003a; Girard *et al.*, 2003b; Girard *et al.*, 2002).

## I.5.1. Présentation des complexes de gadolinium utilisés

Tous les complexes présentés ont été utilisés pour préparer des dérivés anomaux dans le but d'étudier leur propriété de fixation dans les cristaux de protéines.

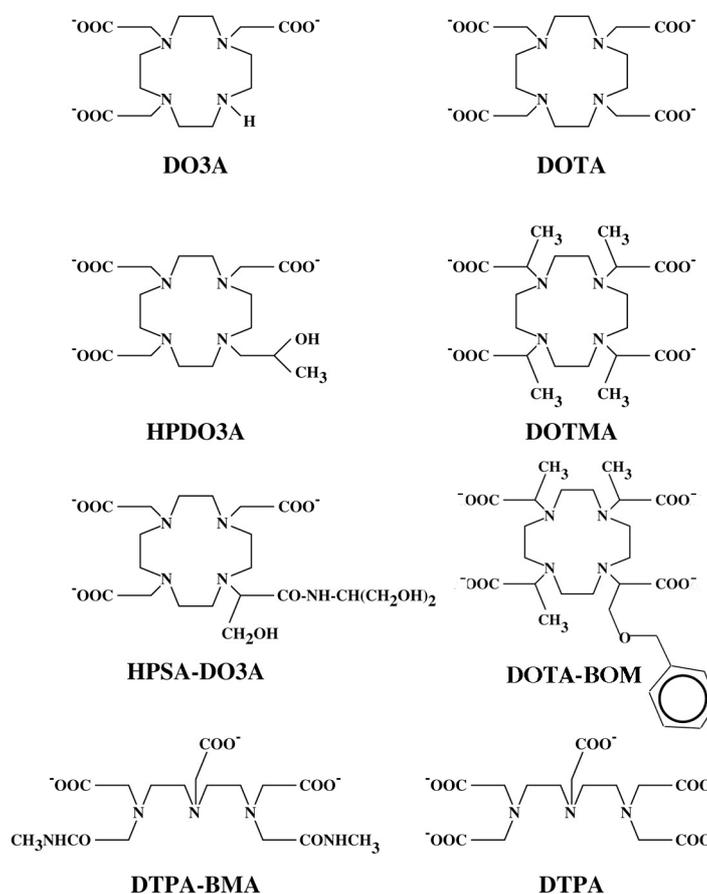


Figure 1.4.1. : Formules développées des ligands des complexes utilisés.

La figure 1.4.1. montre les ligands des complexes utilisés. Tous ces ligands chélatent un ion de Gd<sup>3+</sup> mais peuvent tout aussi bien chélater les autres ions, Ln<sup>3+</sup>, des lanthanides. Les ligands sont dérivés du macrocycle tétra-azacyclododécane (Cyclen) sauf le ligand linéaire DTPA (acide di-éthylène-tri-amine penta-acétique) et son dérivé DTPA-BMA. Les ligands du côté gauche de la figure 1.4.1 forment des complexes électriquement neutres, ceux de droite forment des complexes chargés négativement, 1- pour le Gd-DOTA, le Gd-DOTMA et le Gd-DOTA-BOM et 2- pour le Gd-DTPA. La charge totale du complexe se déduit de sa forme déprotonée. Les valeurs de pK<sub>a</sub> des complexes étant situées autour de 2 (Bianchi *et al.*, 2000) aucune protonation notable n'a lieu dans les conditions physico-chimiques de cristallisation habituelles.

Les complexes sont fortement solubles dans l'eau, leur solubilité étant d'environ 1 M, sauf pour le Gd-HPSA-DO3A dont la solubilité est d'environ 0,3 M, et le Gd-DOTA-BOM qui est soluble à environ 0,6 M.

Avec des constantes de stabilité thermodynamiques des différents complexes très élevées entre logK ~ 22 (Gd-DO3A) et logK ~ 28 [Gd-DOTA]<sup>-</sup> (à titre de comparaison, pour le complexe [Gd-EDTA]<sup>-</sup>, logK = 17,35) (Bianchi *et al.*, 2000), les complexes sont très stables et ne sont pas dissociés dans les conditions habituelles de cristallisation. Seuls des milieux très acides (pH ≤ 2) provoquent une dissociation décelable du complexe.

Les complexes ainsi formés présentent une très faible affinité pour les molécules biologiques.

Ils sont non toxiques.

Du fait des propriétés paramagnétiques du gadolinium et en raison de leurs propriétés chimiques, ces complexes de gadolinium sont très utilisés en imagerie médicale.

## I.5.2. Chimie de coordination de Gd(III)

L'état d'oxydation le plus stable des lanthanides est l'état Ln(III) et l'état ionique correspondant est l'ion  $\text{Ln}^{3+}$ . Pour l'ion  $\text{Gd}^{3+}$ , l'environnement de coordination le plus typique est l'enneacoordination (neuf atomes dans la première sphère de coordination). Lorsque les ligands indiqués sur la figure 1.4.1. complexent les ions  $\text{Ln}^{3+}$  des lanthanides, leurs bras se referment autour de l'ion pour former une cage (figure 1.4.2.), ce qui confère la stabilité très élevée aux complexes. Pour donner un ordre de grandeur de la taille des complexes, la distance entre les deux atomes les plus éloignés du complexe varie entre  $\sim 8,5 \text{ \AA}$  (Gd-DO3A) et  $\sim 10 \text{ \AA}$  (Gd-DOTMA, Gd-DTPA-BMA).



Figure 1.4.2. : a) Le complexe Gd-DOTMA avec comme neuvième ligand une molécule d'eau. b) Le complexe Gd-DTPA, le neuvième ligand étant, dans le cas du dérivé correspondant à la figure, un oxygène carboxylique d'un résidu protéique qui n'est pas montré. Les lignes pointillées montrent les liaisons de complexation dont la longueur varie entre 2,3 et 2,8 Å.

Dans les complexes utilisés, la première sphère de coordination de l'ion de  $\text{Gd}^{3+}$  est occupée, d'une part, par les quatre atomes d'azote du macrocycle, ou les trois atomes d'azote des "pseudo-macrocycle" pour les ligands linéaires, et d'autre part, par les atomes d'oxygène des bras des ligands et qui sont les plus proches de l'ion (figure 1.4.2.). Ainsi, pour aucun des complexes la première sphère de coordination n'est entièrement remplie par les atomes du ligand. En effet, l'ion de  $\text{Gd}^{3+}$  est entouré de 8 atomes chélatant du ligand sauf dans le cas du Gd-DO3A où seuls 7 atomes provenant du ligand assurent la coordination. Pour tous les complexes, il manque donc 1 atome (2 atomes pour Gd-DO3A) pour compléter la première sphère de coordination, ce dont il faut tenir compte lors de l'interprétation du mode de fixation des complexes avec des protéines. Les atomes de coordination manquants peuvent être fournis, soit par des molécules d'eau, soit par des atomes de résidus protéiques, soit encore par un atome d'oxygène appartenant à un bras d'une deuxième molécule de complexe situé à proximité.

Les complexes de gadolinium des ligands HPDO3A, DO3A, DOTMA, HPSA-DO3A et DOTA-BOM nous ont été fournis en forme de poudre par Bracco Imaging spa, Milan, Italie. Pour les complexes Gd-DTPA, Gd-DTPA-BMA et Gd-DOTA nous nous sommes servi des solutions aqueuses des complexes commercialement disponibles (MAGNEVIST™, Schering,

Allemagne, OMNISCAN™, Nycomed Inc., Norvège et DOTAREM™, Guerbet, France respectivement) qui sont fournis à une concentration de 0,5 M.

### Obtention des complexes de gadolinium

Pour le moment, l'accès aux complexes est difficile.

Les trois complexes dont nous utilisons les solutions médicales (Gd-DTPA, Gd-DTPA-BMA et Gd-DOTA) ne s'obtiennent, en France, qu'avec une ordonnance.

Les ligands DOTA et DOTMA sont commercialisés par la compagnie Macrocyclics (Magnetic Resonance Solutions Inc., USA), le ligand DTPA par Sigma-Aldrich. À partir des ligands, il faut synthétiser le complexe en ajoutant le sel de lanthanide.

Le complexe Gd-HPDO3A peut être obtenu en contactant Pier Lucio Anelli de l'entreprise Bracco Imaging. La société fournit des échantillons gratuits du complexe aux groupes de recherche intéressés.

Les complexes Gd-DO3A, Gd-DOTA-BOM et Gd-HPSA-DO3A, fabriqués par la société Bracco Imaging, ne sont pas accessibles sur le marché.

### I.5.3. La diffusion anormale des lanthanides

Les lanthanides (Z=57-71) présentent tous une très forte diffusion anormale de  $f'' \sim 30e^-$  due à une forte résonance (raie blanche) dans leur seuil d'absorption  $L_{III}$ , comme le montre la figure 1.2.2. Les seuils sont situés à des longueurs d'onde inférieures à 2 Å pour les lanthanides allant du Nd au Lu (tableau 1.4.1.). Avec le rayonnement  $CuK\alpha$ , les lanthanides allant du La au Gd ont des valeurs élevées de  $f''$ .

|                             |      |      |      |      |      |      |      |
|-----------------------------|------|------|------|------|------|------|------|
| Éléments                    | La   | Ce   | Pr   | Nd   | Pm   | Sm   | Eu   |
| Z                           | 57   | 58   | 59   | 60   | 61   | 62   | 63   |
| seuil $L_{III}$ [Å]         | 2,26 | 2,16 | 2,08 | 1,99 | 1,92 | 1,84 | 1,78 |
| $f''_{CuK\alpha}$ [ $e^-$ ] | 9,6  | 10,3 | 11,0 | 11,7 | 12,4 | 13,3 | 11,0 |

|                             |      |      |      |      |      |      |      |      |
|-----------------------------|------|------|------|------|------|------|------|------|
| Éléments                    | Gd   | Tb   | Dy   | Ho   | Er   | Tm   | Yb   | Lu   |
| Z                           | 64   | 65   | 66   | 67   | 68   | 69   | 70   | 71   |
| seuil $L_{III}$ [Å]         | 1,71 | 1,65 | 1,58 | 1,54 | 1,48 | 1,43 | 1,39 | 1,34 |
| $f''_{CuK\alpha}$ [ $e^-$ ] | 12,0 | 8    | 8    | 4    | 4    | 5    | 5    | 5    |

Tableau 1.4.1 : Longueurs d'onde des seuils d'absorption  $L_{III}$  (Rieck, 1985) et partie imaginaire  $f''$  du facteur de diffusion atomique à la longueur d'onde  $\lambda_{CuK\alpha} = 1,54$  Å (Ibers *et al.*, 1985) pour les différents lanthanides.

Le gadolinium présente l'avantage de combiner un signal anormal élevé ( $f'' = 12,0 e^-$ ) avec le rayonnement  $K\alpha$  du Cu et un seuil  $L_{III}$  à la longueur d'onde de 1,711 Å, relativement accessible.

### I.5.4. Obtention de cristaux dérivés à l'aide de complexes de gadolinium.

#### I.5.4.a. Insertion du complexe de gadolinium dans les cristaux

Les propriétés des complexes précédemment décrites font qu'ils sont très simples à utiliser pour introduire les atomes lourds dans des cristaux de protéine soit par cocrystallisation soit par trempage. En effet, leur stabilité, leur compatibilité avec la plupart de conditions physico-

chimiques de cristallisation (pH, agent précipitant, tampon), leur solubilité élevée et leur absence de toxicité les rendent aisément utilisables. Sauf rares exceptions, ils modifient peu les conditions de cristallisation des cristaux dérivés par rapport à celles des cristaux natifs. La particularité de ces complexes est leur faible affinité pour les molécules biologiques avec lesquelles ils établissent des liaisons non-covalentes. Pour obtenir une fixation suffisante il faut utiliser des concentrations élevées de complexe, de l'ordre de 100 mM.

On peut, sur un exemple, estimer l'ordre de grandeur de l'affinité d'un des complexes pour un site favorable à sa fixation sur une protéine. Le dérivé de la protéine glucose isomérase avec le complexe Gd-DO3A employé avec une concentration  $C_c = 100\text{mM}$ , présente un site de fixation avec un taux d'occupation  $q = 0,56$ .

Les cristaux dérivés appartiennent au groupe d'espace I222, avec 8 molécules par maille, dont le volume est :

$$V_{\text{maille}} = 93 \times 99 \times 103 \times 10^{-30} \text{ m}^3.$$

La concentration de la protéine dans le cristal est :

$$C_p = \frac{8}{V_{\text{maille}} \times N_A} = 14\text{mM}, \text{ où } N_A \text{ est le nombre d'Avogadro.}$$

On obtient donc une constante de dissociation :

$$K_d = \frac{[(1-q) \times C_p] \times [C_c - q \times C_p]}{[q \times C_p]} = \frac{(0,44)(100\text{mM} - 0,56 \times 14\text{mM})}{0,56} = 72\text{mM}$$

Cette constante de dissociation, relativement élevée, montre bien la faible affinité du complexe pour la protéine.

#### **I.5.4.b. Cocrystallisation**

La cocrystallisation limite les manipulations du cristal. Elle permet au complexe de se fixer au fur et à mesure de la croissance des cristaux, évitant la détérioration des cristaux qui peut se produire lors du trempage avec le même complexe, en particulier pour les cristaux avec une faible proportion de solvant.

Les conditions d'obtention des cristaux dérivés correspondent aux conditions avec lesquelles sont obtenus les cristaux natifs avec un ajout d'environ 100 mM de complexe dans la goutte de cristallisation. L'ajout du complexe peut provoquer de légers changements de solubilité, aussi est-il recommandé de faire varier la concentration de l'agent précipitant. En cas de forte solubilisation de la protéine (ce qui est possible pour l'ajout des complexes Gd-DOTA-BOM et Gd-HPDO3A) il est conseillé de baisser la concentration du complexe à 50mM.

Dans des cas exceptionnels, l'ajout du complexe dans la goutte de cristallisation a empêché toute cristallisation et provoqué la précipitation de la protéine : ceci a été le cas pour deux dérivés de la protéine glucose isomérase obtenus avec Gd-DO3A et avec Gd-DTPA. Mais pour cette protéine, seuls deux des complexes se sont fixés notablement dans le cristal et ce sont ces deux mêmes complexes qui ont une forte influence sur la solubilité de la protéine. On peut supposer qu'un tel changement des conditions de cristallisation peut indiquer une fixation notable du complexe. Dans ce cas, il y a des fortes chances d'obtenir des bons dérivés avec ces mêmes complexes par trempage des cristaux natifs.

#### **I.5.4.c. Trempage**

Si on dispose de cristaux natifs de la protéine et qu'on ne dispose de protéine qu'en petite quantité, le moyen le plus commode pour obtenir des cristaux dérivés consiste à tremper les

cristaux natifs. Pour cela on prépare une solution de trempage correspondant à la liqueur mère avec un ajout d'environ 100 mM de complexe et un ajout de protéine de l'ordre de 2mg/mL. Si on ne peut pas ajouter de protéine dans la goutte de trempage, il convient d'augmenter la quantité de précipitant dans la goutte de trempage afin d'éviter la dissolution du cristal. La durée de trempage typique est relativement courte, variant de quelques dizaines de minutes à quelques heures. Dans quelques cas, il a suffi d'un trempage de quelques dizaines de secondes pour obtenir une fixation suffisante du complexe. Les complexes sont généralement bien tolérés par les cristaux, ce qui peut être dû à leur faible affinité pour les protéines et à leur électroneutralité. Si les cristaux ne supportent pas le trempage, il convient de baisser la concentration de complexe ou de raccourcir la durée de trempage.

Dans le cas du lysozyme de blanc d'œuf de poule dont les cristaux ont une proportion de solvant particulièrement faible, inférieure à 30%, les cristaux natifs ne supportent pas le trempage avec le complexe qui se fixe le mieux : les cristaux dérivés ont dû être préparés par cocrystallisation (Girard, 2001).

Pour des cristaux particulièrement sensibles aux manipulations, on peut procéder au trempage en ajoutant directement de la solution de complexe à la goutte mère du cristal : la solubilité élevée des complexes permet d'ajouter des volumes faibles de solution, ce qui limite la dilution de la goutte.

#### **I.5.4.d. Cryoprotection des cristaux dérivés**

La cryoprotection se fait par trempage dans la solution cryoprotectrice habituelle. Elle ne présente pas de difficultés particulières, sauf cas exceptionnels comme l'a été celui des cristaux dérivés d'urate oxydase avec Gd-DO3A, qui diffractent mal après congélation.

Il est préférable de ne pas ajouter de complexe à la solution cryoprotectrice afin de réduire le plus possible la quantité de complexe présente dans la solution qui entoure le cristal. Ceci permet de diminuer le bruit de fond dû à la fluorescence de l'atome lourd lors de l'acquisition des données, en particulier lorsqu'elle est réalisée dans le seuil d'absorption de l'élément. Le trempage dans la solution cryoprotectrice doit alors être court, de l'ordre d'une dizaine de secondes, pour éviter que les molécules de complexe quittent les sites de fixation.

#### **I.5.4.e. Comparaison des résultats obtenus avec les cristaux dérivés obtenus à la fois par cocrystallisation et par trempage**

Pour les protéines dont les cristaux dérivés ont pu être préparés aussi bien par trempage que par cocrystallisation (urate oxydase, thaumatine, protéine X), aucune différence de fixation du complexe utilisé n'a été constatée pour les cristaux dérivés préparés avec les deux modes, les mêmes sites ont été trouvés, avec le même taux d'occupation.

### **I.5.5. Dégâts d'irradiation II**

Murray *et al.* (2004) estiment que des concentrations d'atomes lourds aussi élevées que 100 mM dans le solvant du cristal conduisent à une forte absorption du rayonnement par le cristal dérivé, pour des rayonnements dont l'énergie est supérieure ou égale à l'énergie du seuil d'absorption du diffuseur anomal. Cette absorption serait alors susceptible de produire très rapidement des dégâts d'irradiation importants. Nos données de diffraction ont été obtenues en utilisant le rayonnement synchrotron provenant le plus souvent d'un aimant de courbure (avec la ligne BM30A à l'ESRF) ou avec le rayonnement d'un générateur à anode tournante de cuivre du laboratoire. Dans ces conditions, nous n'avons jamais rencontré de problèmes

particuliers de dégâts d'irradiation. Nous avons toujours enregistré des données hautement redondantes, avec une rotation du cristal de 180°. Avec le rayonnement synchrotron, nous avons généralement enregistré deux jeux de données hautement redondants en n'utilisant qu'un seul cristal, un jeu avec le rayonnement correspondant au maximum de  $f''$  et un jeu à courte longueur d'onde, afin d'obtenir des données à haute résolution. Pour ces deux jeux de données de diffraction, nous n'avons pas observé de dégradation notable de la qualité de diffraction.

Lorsque nous avons enregistré des données sur des dérivés contenant les complexes de gadolinium sur la ligne ID29 (ESRF) nous avons pu constater, pour certains dérivés, des effets dus aux dégâts d'irradiation, se manifestant par la perte des réflexions à haute résolution, lors des enregistrements au maximum de  $f''$  au seuil d'absorption  $L_{III}$  du gadolinium. Nous n'avons pas procédé à des mesures comparatives avec des cristaux natifs pour voir si cet effet était uniquement dû à la présence des atomes lourds dans les canaux de solvant.

Nous avons également travaillé avec des cristaux dérivés de la protéine urate oxydase, obtenus avec une concentration de complexe de 300 mM. Pour chaque cristal dérivé nous avons enregistré deux jeux de données avec une rotation de 180°. L'acquisition des données a eu lieu sur BM30A et sur ID29 sans qu'on ait détecté de dégâts d'irradiation notables.

En effet, ne considérer pour le calcul de l'effet nocif que la dose théoriquement absorbée (Murray *et al.*, 2004) par les atomes lourds, néglige que la majeure partie des molécules se trouve dans le solvant et que l'atome lourd est entouré d'une cage qui pourrait protéger la protéine des effets de l'absorption du rayonnement au niveau de l'atome de lanthanide.

## I.5.6. Conclusions

Si la faible affinité des complexes pour les protéines fait que nous devons travailler à des concentrations élevées de complexe, elle a en retour, pour conséquence, la bonne tolérance des complexes lors des trempages ou cocrystallisation et fait que les cristaux dérivés sont bien isomorphes des cristaux natifs.

Le signal anomal d'un dérivé typique obtenu avec ces complexes est élevé. Pour une protéine de taille moyenne, comportant 300 acides aminés, avec deux sites de fixation du complexe et une occupation  $q_j = 0,5$  par site, on peut estimer l'ordre de grandeur du signal anomal moyen obtenu avec un cristal dérivé de cette protéine. À cet effet, on calcule le rapport de Bijvoet

moyen attendu à l'aide de la relation :  $\frac{\Delta F^\pm}{F} [\%] = 0,76 \sqrt{N_H} q_j f_j''$  donnée au paragraphe I.4.6.

La valeur de  $N_H$  est alors 2/3. Avec le rayonnement  $K\alpha$  du cuivre, on obtient

$\left(\frac{\Delta F^\pm}{F}\right)_{CuK\alpha} = 3,7\%$  et avec le rayonnement synchrotron, dans le seuil d'absorption  $L_{III}$  de

l'atome de gadolinium, au maximum de  $f''$  ( $f_j'' = 28e^-$ ), on obtient :  $\left(\frac{\Delta F^\pm}{F}\right)_{seuil} = 8,7\%$  qui souligne l'importance du signal anomal attendu.

**Partie II.**

**Étude cristallographique**



## II.0. Introduction

Nous avons entrepris une étude cristallographique systématique et comparative de cristaux dérivés de différentes protéines préparés avec toute la série de complexes afin de déterminer quel est leur potentiel pour obtenir des dérivés à fort pouvoir de phasage. Cette étude devrait permettre de vérifier si ces complexes s'utilisent aisément avec tous types de protéines, et de déterminer leurs conditions d'utilisation optimales. Elle comprend l'obtention des cristaux des protéines, la préparation des cristaux dérivés, l'acquisition et la réduction des données de diffraction, la détermination d'éventuels sites de fixation et le calcul des phases.

Dans un deuxième temps, pour des dérivés avec des sites de fixation suffisamment occupés, l'affinement de la structure du complexe lié à la protéine devrait permettre d'identifier et comprendre le mode de fixation du complexe.

L'analyse des résultats d'une telle étude a pour but d'établir d'éventuelles corrélations entre les conditions physico-chimiques de cristallisation (agent précipitant, pH) ou des caractéristiques de la protéine (pI, séquence, motifs structuraux particuliers) et la fixation des différents complexes à la protéine. Pour une protéine de structure inconnue, cette information pourrait être utilisée pour choisir parmi les différents complexes celui qui serait susceptible de se fixer à la protéine et pour adapter les conditions de son utilisation en fonction des différents paramètres connus.

L'étude systématique a été menée avec six protéines différentes dont la structure et les conditions de cristallisation sont connues. En outre, des études menées dans le cadre de projets en collaboration avec d'autres groupes de cristallographie biologique ont permis de résoudre *de novo* la structure de quatre nouvelles protéines grâce à l'utilisation de ces complexes.

## II.1. Cristallisation des protéines

### II.1.1. Choix des protéines tests

Pour choisir les protéines avec lesquelles, outre celles fournies par des collaborateurs, on pourrait tester l'utilisation des complexes pour obtenir des dérivés lourds, nous avons tenu compte des critères suivants (McPherson, 2001) :

- disposer d'une quantité suffisante de protéine pure,
- des conditions de cristallisation connues, permettant d'obtenir de cristaux diffractant à relativement haute résolution,
- une proportion de solvant élevée dans les cristaux,
- une symétrie cristalline d'ordre élevé,
- si possible, des protéines cristallisant dans différentes conditions.

Des tests réalisés avec une protéine qui cristallise dans différentes conditions devraient permettre de déterminer l'influence des conditions physico-chimiques de cristallisation comme l'agent précipitant utilisé, le pH ou encore l'influence de la forme cristalline sur la fixation des complexes.

Le tableau 2.1.1 montre la liste des protéines systématiquement testées avec la plupart des complexes.

| Protéine  | Nombre d'acides aminés |
|---|------------------------|
| Urate oxydase d' <i>Aspergillus flavus</i>                | 295                    |
| Lysozyme de blanc d'œuf de poule                          | 129                    |
| Thaumatine de <i>Thaumatococcus daniellii</i>             | 207                    |
| Glucose isomérase (GI) de <i>Streptomyces rubiginosus</i> | 388                    |
| YGGV d' <i>Escherichia coli</i>                           | 221                    |
| Protéine X  |                        |

Tableau 2.1.1. : Liste des protéines test utilisées.

Bien que des conditions de cristallisation différentes aient été établies pour trois protéines étudiées, thaumatine (Ko *et al.*, 1994; Charron *et al.*, 2004), urate oxydase (communication de F. Bonneté & D. Vivarès, Université Paris 6-Paris 7) et glucose isomérase (Ramagopal *et al.*, 2003) nous n'avons réussi à obtenir des cristaux dans différentes conditions de cristallisation qu'avec la glucose isomérase.

## II.1.2. Cristallisation des protéines

Toutes les protéines ont été cristallisées par la méthode de la goutte suspendue.

Les cristaux dérivés ont été obtenus par trempage des cristaux natifs ou par cocrystallisation. Comme, dans le cas où les cristaux dérivés pouvaient être obtenus avec les deux méthodes, nous n'avons constaté aucune différence de fixation des complexes pour les dérivés cocrystallisés ou trempés, les cristaux dérivés ont été préparés, de préférence, par cocrystallisation, limitant ainsi la manipulation des cristaux.

### Obtention des dérivés par cocrystallisation

Sauf mention contraire, la concentration du complexe est de 100 mM dans la goutte de cristallisation ayant atteint son volume final. Le volume initial de solution, prélevée dans le réservoir, est  $V_R$ . On y ajoute un volume,  $V_P$ , de la solution protéique et un volume,  $V_C$ , de la solution de complexe, calculé pour obtenir la concentration de complexe,  $C$ , attendue dans la goutte finale. La composition initiale de la goutte de cristallisation peut être, par exemple :  $V_R = 2,5 \mu\text{L}$ ,  $V_P = 5 \mu\text{L}$  et  $V_C = 0,25 \mu\text{L}$  de solution de complexe à la concentration  $C_0 = 1 \text{ M}$ . Le volume attendu de la goutte à l'équilibre est  $V_R$ , et la concentration de complexe attendue dans la goutte finale est  $C = V_C / V_R \times C_0 = 100 \text{ mM}$ .

En observant la taille finale des gouttes au moment de récupérer les cristaux, on s'aperçoit cependant que le volume final de la goutte est le plus souvent largement supérieur à  $V_R$ , soit parce que l'équilibre n'est pas encore atteint, soit parce que l'ajout du complexe et de la protéine dans la goutte influence l'équilibre. Ceci signifie que la concentration de complexe finale est généralement inférieure à la concentration attendue.

Le complexe est introduit dans la goutte en ajoutant un petit volume de solution de complexe à la goutte formée par la solution de réservoir et solution protéique ou, dans le cas des protéines lyophilisées, en dissolvant la protéine dans une solution contenant le complexe. Cette dernière méthode est avantageuse si on veut travailler avec des concentrations de

complexe élevées ou, avec des complexes pour lesquels la concentration de la solution stock est relativement faible, pour ne pas trop augmenter le volume de la goutte de cristallisation.

### **Obtention des dérivés par trempage**

Sauf mention contraire, tous les trempages ont été faits en transférant le cristal natif à l'aide d'une boucle dans la solution de trempage. La solution de trempage est de la liqueur-mère supplémentée par 100 mM de complexe et entre 1 et 5 mg/mL de protéine. La durée de trempage typique varie entre 10 min et 1 h, temps pendant lequel le cristal est observé, afin d'arrêter le trempage en cas de dégradation visible.

### **Congélation du cristal**

Après la durée de trempage souhaitée ou au moment où les cristaux cocrystallisés ont atteint la taille souhaitée, le cristal dérivé est congelé. Sauf mention contraire, le cristal est transféré dans la solution cryoprotectrice à l'aide d'une boucle. Cette solution est composée de la liqueur-mère supplémentée de l'agent cryoprotecteur adapté, à la concentration nécessaire. Elle ne contient pas de complexe. Le cristal est brièvement trempé dans la solution cryoprotectrice. Il est alors congelé, soit sous un flux d'azote gazeux à 100K sur la tête goniométrique du diffractomètre, soit dans de l'azote liquide pour une utilisation ultérieure.

### II.1.3. Conditions de cristallisation des différentes protéines

#### Urate oxydase d'*Aspergillus flavus* (295 acides aminés)

La protéine nous est fournie par la société Sanofi Synthélabo. La solution fournie est à une concentration de protéine,  $C_p$ , de 11,28 mg/mL contenant 0,28 mg/mL de 8-azaxanthine, un inhibiteur de la protéine, et 1,04 mg/mL d'EDTA. Elle est utilisée sans purification supplémentaire. La protéine aliquotée et congelée étant instable après décongélation, les cristaux ne peuvent être préparés qu'avec la solution de protéine fraîchement décongelée.

La protéine cristallise dans les conditions suivantes (conditions adaptées et modifiées d'après Vivarès & Bonneté (2002) et Bonneté *et al.* (2001)) :

6-15% PEG 3350  
75 mM tampon tris pH 8,5  
 $C_p = 11,28$  mg/mL  
Composition des gouttes :  $V_p = 4 \mu\text{L}$ ,  $V_R = 2 \mu\text{L}$  ou  $V_R = 4 \mu\text{L}$ ,  $V_R = 2,5 \mu\text{L}$   
La préparation et le stockage des boîtes de cristallisation se font à  $T = 8^\circ\text{C}$

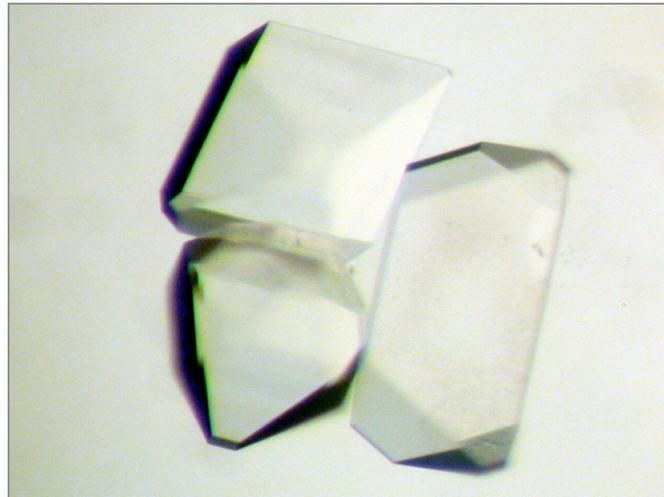


Figure 2.1.1 : Cristaux d'urate oxydase. Taille des cristaux :  $\sim 200$ - $500 \mu\text{m}$ .

Agent cryoprotecteur : 25-30% PEG400

Les cristaux dérivés sont obtenus par cocrystallisation avec 100 mM de complexe. L'ajout de certains complexes dans la goutte de cristallisation engendre une légère modification de la solubilité de la protéine. Il est cependant suffisant de faire varier la concentration de l'agent précipitant pour obtenir les cristaux, à l'exception du complexe Gd-DO3A qui, à la concentration de 100 mM, empêche la formation de cristaux et pour lequel les cristaux dérivés ont été obtenus avec une concentration de Gd-DO3A de 50 mM. Les cristaux dérivés obtenus avec ce même complexe supportent assez mal le trempage dans la solution cryoprotectrice habituelle.

### **Lysozyme de blanc d'œuf de poule** (129 acides aminés)

La protéine lyophilisée provient de Boehringer, lot n° 85438222-14 et elle est utilisée sans purification supplémentaire.

La protéine cristallise dans les conditions suivantes (Girard *et al.*, 2002) :

|   |
|---|
| 0,8-1,2 M NaCl<br>50 mM tampon acétate de sodium pH 4,5<br>$C_p = 40$ mg/mL<br>Composition des gouttes : $V_p = 3$ $\mu$ L, $V_R = 3$ $\mu$ L<br>$T = 20^\circ\text{C}$ |
|---|

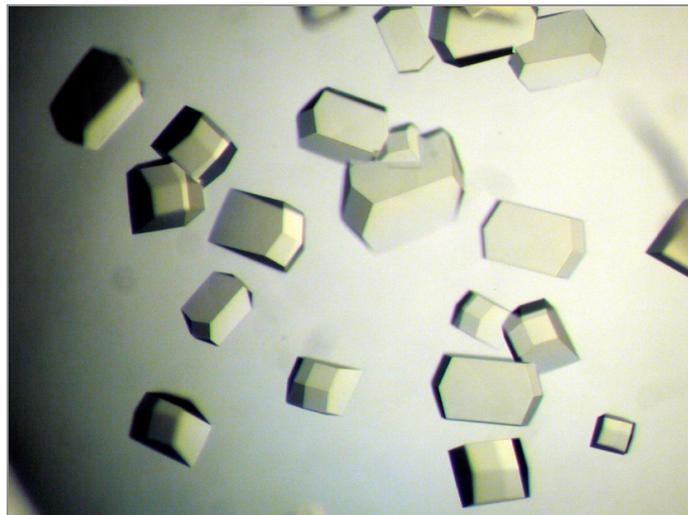


Figure 2.1.2 : Cristaux de lysozyme. Taille des cristaux :  $\sim 300$ - $800$   $\mu\text{m}$ .

Agent cryoprotecteur : 25-30% PEG400

Les cristaux dérivés sont obtenus par cocrystallisation avec 100 mM de complexe. L'ajout de complexe dans les gouttes de cristallisation peut provoquer une légère modification de la solubilité de la protéine. Le composé Gd-DOTA-BOM semble assez fortement augmenter la solubilité de la protéine et pour obtenir des cristaux il est nécessaire d'augmenter la concentration de NaCl, auquel cas on obtient des cristaux peu nombreux et très gros.

### **Thaumatine de *Thaumatococcus daniellii*** (207 acides aminés)

La protéine lyophilisée provient de Sigma (Cat. n° : T-7638, lot n° : 108F0299). Elle est utilisée sans purification supplémentaire.

La protéine cristallise dans les conditions suivantes (conditions adaptées d'après Ko *et al.* (1994) et Charron *et al.*, (2004)) :

|   |
|---|
| 0,6-0,9 M tartrate double de Na-K<br>0 - 5% éthylène glycol<br>100 mM tampon ADA pH 6,6<br>$C_p = 50$ mg/mL<br>Composition des gouttes : $V_p = 6$ $\mu$ L, $V_R = 4$ $\mu$ L<br>$T = 20^\circ\text{C}$ |
|---|

La concentration de protéine indiquée  $C_p$  correspond à la quantité de poudre lyophilisée dissoute et non pas à une concentration mesurée par densité optique.

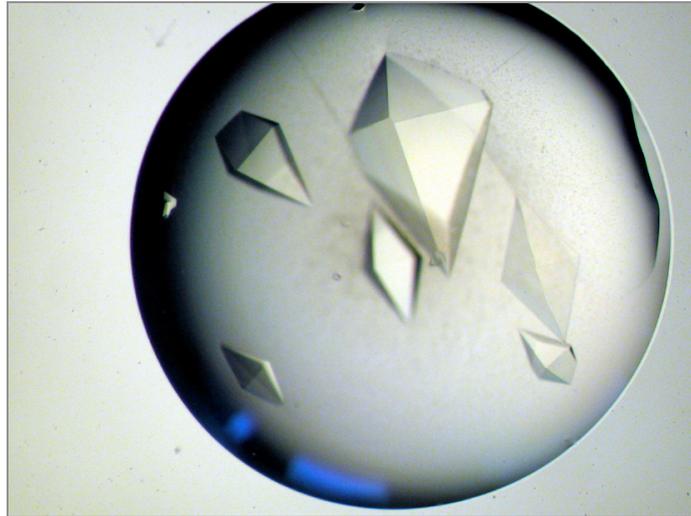


Figure 2.1.3 : Cristaux de thaumatine. Taille des cristaux :  $\sim 200-800 \mu\text{m}$ .

Agent cryoprotecteur : 25% éthylène glycol

Les cristaux dérivés sont obtenus par cocristallisation avec 100 mM de complexe. Seul l'ajout de 100 mM de complexe Gd-DOTA-BOM empêche la cristallisation, les cristaux dérivés de ce complexe sont obtenus à une concentration de complexe de 50 mM.

J'ai essayé d'obtenir des cristaux de thaumatine avec d'autres conditions de cristallisation supposées mener à différentes formes cristallines (Ko *et al.*, 1994; Charron *et al.*, 2004). Aucun de mes essais de cristallisation n'a permis d'obtenir des cristaux.

#### **Glucose isomérase (GI) de *Streptomyces rubiginosus* (388 acides aminés)**

La protéine est commercialisée par la société Hampton research (n° de commande : HR7-102). Elle est fournie sous forme de suspension de microcristaux, la suspension contenant 6 mM Tris HCl pH 7,0  
120 mg/mL sulfate d'ammonium (0,91 M)  
1 mM sulfate de magnésium

Préparation de la solution protéique (selon [www.hamptonresearch.com/support/guides/7100G.pdf](http://www.hamptonresearch.com/support/guides/7100G.pdf))

Afin d'enlever le sulfate d'ammonium de la solution commerciale et ainsi dissoudre les microcristaux, la protéine est dialysée contre un tampon contenant 6 mM Tris pH 7,0 et 1 mM  $\text{MgSO}_4$ . Après dissolution complète des microcristaux, la solution protéique est filtrée avec des filtres de  $0,2 \mu\text{m}$ . La concentration de la protéine est déterminée par densité optique à  $\sim 30 \text{ mg/mL}$ . La protéine est ensuite aliquotée et congelée.

La protéine forme un homotétramère de 173 kDa et cristallise avec un monomère par unité asymétrique dans les conditions suivantes (conditions adaptées d'après Ramagopal *et al.*, (2003)) :

9-14% MPD  
 50 mM MgCl<sub>2</sub>  
 50 mM tampon tris pH7  
 C<sub>p</sub> ~ 30 mg/mL  
 Composition des gouttes : V<sub>p</sub> = 4 μL, V<sub>R</sub> = 2 μL  
 T = 20°C

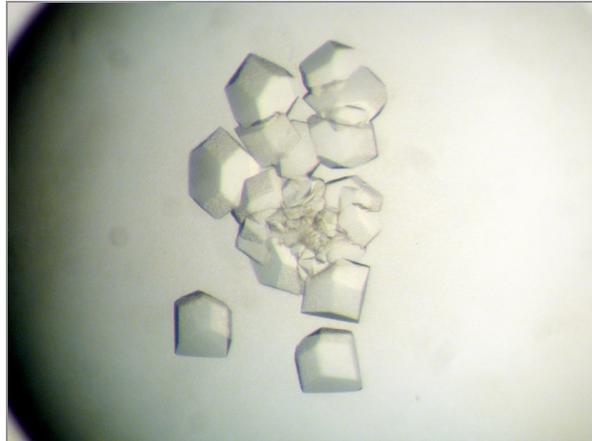


Figure 2.1.4 : Cristaux de GI obtenus en présence de MPD. Taille des cristaux : ~300 μm.

Agent cryoprotecteur : 30% MPD

L'obtention des cristaux par cocrystallisation a été possible pour la plupart des complexes. Cependant, l'ajout des complexes Gd-DTPA et Gd-DO3A dans les gouttes de cristallisation provoque la précipitation de la protéine et empêche toute formation de cristaux. Il se trouve que ce sont ces deux mêmes complexes, et seulement ces complexes, qui se lient fortement à la protéine.

Les dérivés de ces deux complexes ont donc été obtenus par trempage avec des concentrations de complexe de 100 mM et une durée de trempage de 1 heure.

#### **Cristallisation de la protéine GI avec des conditions de cristallisation alternatives** (adaptées d'après [www.hamptonresearch.com/support/guides/7100G.pdf](http://www.hamptonresearch.com/support/guides/7100G.pdf)) :

Les cristaux de glucose isomérase, classiquement obtenus avec du MPD comme agent précipitant, peuvent aussi être obtenus avec d'autres agents précipitants comme le PEG 3350 ou le sulfate d'ammonium. Les conditions de cristallisation sont indiquées ci-après.

|                           |   |
|---------------------------|---|
| "condition PEG" :         | "condition sel" :   |
| 6-8% PEG 3350             | 1,1-1,6 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> |
| 40 mM MgCl <sub>2</sub>   | 50 mM MgCl <sub>2</sub>                                   |
| 50 mM tampon tris pH 7,0  | 50 mM tampon tris pH 7,0                                  |
| C <sub>p</sub> ~ 30 mg/mL | C <sub>p</sub> ~ 30 mg/mL                                 |
| T = 20°C                  | T = 20°C  |

Pour les cristaux obtenus avec les conditions alternatives, les dérivés ont été obtenus par trempage avec des concentrations de complexe de 100 mM et une durée de trempage de 1 heure. Bien que je n'aie pas procédé à des tests de cocrystallisation, j'ai pu voir que dans la "condition PEG", et contrairement aux conditions en MPD, la protéine cristallise en présence du complexe Gd-DO3A : il y a eu formation de cristaux dans la goutte de trempage.

Pour trouver des conditions de cryoprotection adaptées aux cristaux obtenus avec les conditions de cristallisation alternatives, j'ai procédé à de nombreux tests de différents agents cryoprotecteurs. Les meilleures conditions trouvées sont les suivantes :

Agent cryoprotecteur pour la "condition PEG" : 30% de glycérol

Agent cryoprotecteur pour la "condition sel" : 40% de malonate de sodium

La cryoprotection et la congélation des cristaux obtenus dans les deux conditions alternatives sont cependant délicates, les cristaux n'étant pas stables dans les conditions cryoprotectrices respectives trouvées. La cryoprotection des cristaux obtenus en présence de sulfate d'ammonium n'est possible que pour les cristaux de petite taille. Cette petite taille a fait que nous n'avons pas pu acquérir des données sur le générateur de laboratoire, la diffraction des cristaux y étant trop faible.

#### **YGGV d'*Escherichia coli* (221 acides aminés)**

Les cristaux natifs d'YGGV m'ont été fournis par Chantal Abergel (IGS, Marseille)

Conditions de cristallisation :

|   |
|---|
| 33,3% - 38% $(\text{NH}_4)_2\text{SO}_4$<br>0,1 M tampon Bicine pH 9,0<br>5% glycérol<br>T = 20°C |
|---|

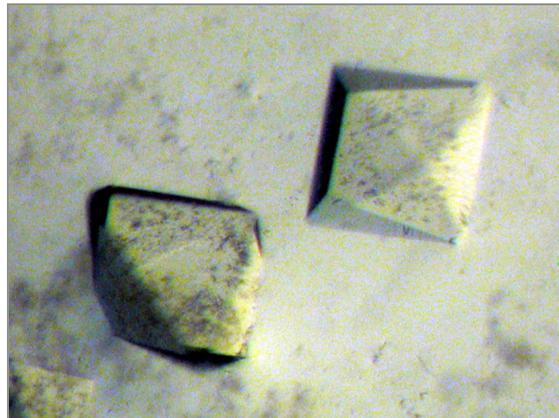


Figure 2.1.5 : Cristaux de YGGV. Taille des cristaux : ~200  $\mu\text{m}$ .

Cryoprotection : la solution mère est cryoprotectrice

Obtention des dérivés par trempage : le complexe a été ajouté à la goutte mère à des concentrations de 50 - 100 mM.

### Conditions de cristallisation des protéines étudiées en collaboration

Les sujets décrits ci-dessous correspondent à des collaborations avec d'autres équipes de cristallographie biologique. Les cristaux dérivés ont été préparés par nos collaborateurs.

#### Collaboration avec Rafael Molina Monterrubio et Juan Hermoso, CSIC, Madrid

##### Protéine X

|                       |  |
|-----------------------|--|
| Obtention des dérivés | Cocrystallisation avec 100 mM ou 50 mM de tous les complexes sauf Gd-HPSA-DO3A ou trempage dans 100 mM de complexe pendant quelques dizaines de secondes |
|-----------------------|--|

La structure de la protéine n'étant pas encore publiée, les conditions de cristallisation et son identité sont pour le moment confidentielles.

##### Phosphorylcholine esterase Pce de *Streptococcus pneumoniae* (547 résidus/ua) (Lagartera *et al.*, 2005)

|                               |   |
|-------------------------------|---|
| Conditions de cristallisation | 17 % PEG 10000<br>0,1 M acétate d'ammonium<br>0,1 M tampon bis tris pH = 5,5                            |
| Composition des gouttes       | $V_p = 4 \mu\text{L}$ , $V_R = 4 \mu\text{L}$<br>+ 1 $\mu\text{L}$ de 1,5 mM N-dodecylphosphorylcholine |
| T                             | 20°C  |
| Obtention des dérivés         | Trempage avec 100 mM de Gd-HPDO3A pendant quelques dizaines de secondes                                 |

#### Collaboration avec Sandra Jeudy et Chantal Abergel, IGS, Marseille

##### YeaZ d'*Escherichia coli* (4 \* 251 résidus/ua) (Jeudy *et al.*, 2005)

|                               |   |
|-------------------------------|---|
| Conditions de cristallisation | 4-8% (w/v) PEG8000<br>0,2 M NaCl<br>10-20% glycérol<br>0,1 M tampon acétate de sodium, pH 4,7-5,5 |
| $C_p$                         | 17 mg/mL  |
| Composition des gouttes       | $V_p = 1 \mu\text{L}$ , $V_R = 0,5 \mu\text{L}$   |
| T                             | 20°C  |
| Obtention des dérivés         | Trempage avec 100 mM de Gd-DOTMA pendant 6 h  |

#### Collaboration avec Stéphanie Gras et Dominique Housset, LCCP, IBS, Grenoble

##### PACE12 native : PABO0955 de *Pyrococcus abyssi* (262 résidus/ua) (Gras *et al.*, 2005)

|                               |  |
|-------------------------------|--|
| Conditions de cristallisation | 15%(w/v) PEG 4000<br>3%(v/v) dioxane<br>0,1 M tampon dihydrate de citrate trisodique pH 5,6<br>20 mM DTT |
| $C_p$                         | 7,8 mg/mL  |
| Composition des gouttes       | $V_p = 1 \mu\text{L}$ , $V_R = 1 \mu\text{L}$  |
| T                             | 18°C   |
| Obtention des dérivés         | Cocrystallisation avec 100 mM de Gd-DTPA-BMA   |

## Protéines étudiées par Eric Girard, SOLEIL, Paris (Girard, 2001)

### Zn- $\beta$ -lactamase de classe B de *Pseudomonas aeruginosa* (230 résidus/ua)

|                               |   |
|-------------------------------|---|
| Conditions de cristallisation | 30% PEG8000<br>100 mM tampon cacodylate de sodium pH 6,5<br>200 mM acétate de sodium  |
| T                             | 20°C  |
| Obtention des dérivés         | Trempage avec 100 mM de complexes et des durées de trempage de Gd-DTPA-BMA : $\Delta t = 7$ h, Gd-DO3A : $\Delta t = 17$ h et Gd-DOTMA : $\Delta t = 5$ h |

### Catalase de *Proteus mirabilis* (484 résidus/ua)

|                               |   |
|-------------------------------|---|
| Conditions de cristallisation | 2,3-2,7 M sulfate d'ammonium<br>100 mM tampon tris maléate pH 6,5 |
| C <sub>p</sub>                | 50 mg/mL  |
| T                             | 6°C   |
| Obtention des dérivés         | Trempage avec 100 mM de Gd-HPDO3A pendant moins de 10 min         |

### OTCase 3630 chimérique (12\*335 résidus/ua)

|                               |  |
|-------------------------------|--|
| Conditions de cristallisation | 2,2 M sulfate d'ammonium<br>100 mM tampon imidazole pH 7,8<br>1 mM DTT |
| C <sub>p</sub>                | 12,5 mg/mL   |
| T                             | 20°C   |
| Obtention des dérivés         | Cocrystallisation avec 100 mM Gd-HPDO3A                                |

## II.1.4. Obtention de cristaux dérivés à une concentration de complexe de 300 mM

On s'attend à ce que l'occupation des sites augmente avec la concentration de complexe utilisée (Girard, 2001, Girard *et al.*, 2002). Pour l'affinement de la structure des complexes liés aux différentes protéines, j'ai donc préparé des dérivés avec une concentration de complexe de 300 mM, ce qui devrait mener à des taux d'occupation des sites de fixation élevés. J'ai préparé des dérivés à haute concentration de complexe pour les complexes qui s'étaient assez fortement fixés lors de la préparation avec 100 mM pour les protéines lysozyme, urate oxydase et thaumatine. Les cristaux dérivés ont été préparés par cocrystallisation avec une concentration de complexe de 300 mM (sauf pour le cas du dérivé de lysozyme et de Gd-DOTA-BOM pour lequel la présence de complexe à haute concentration empêche la cristallisation et pour lequel j'ai cocrystallisé la protéine en présence de 50 mM de Gd-DOTA-BOM). Vu la faible diminution de taille des gouttes de cristallisation au cours du temps, la concentration finale de complexe dans les gouttes était certainement inférieure à 150 mM. C'est pourquoi, dans un deuxième temps, j'ai trempé les cristaux obtenus par cocrystallisation comme décrit ci-dessus dans des solutions de trempage contenant 300 mM du complexe correspondant pendant 45 min.

La préparation par cocrystallisation ou par trempage des dérivés à une concentration de complexe aussi élevée demande des efforts supplémentaires et est moins aisée que l'obtention de dérivés avec 100 mM de complexe : la concentration élevée du complexe risque de

modifier la solubilité de la protéine, le trempage risque d'être moins bien supporté et la préparation des solutions de cocristallisation et de trempage, sans trop les diluer, est plus difficile. Afin d'éviter ces problèmes, il reste recommandé de préparer les dérivés utilisés pour le phasage avec des concentrations de complexe entre 50 et 100 mM, ce qui mène souvent à des taux d'occupation anomaux suffisamment élevés pour un bon phasage.

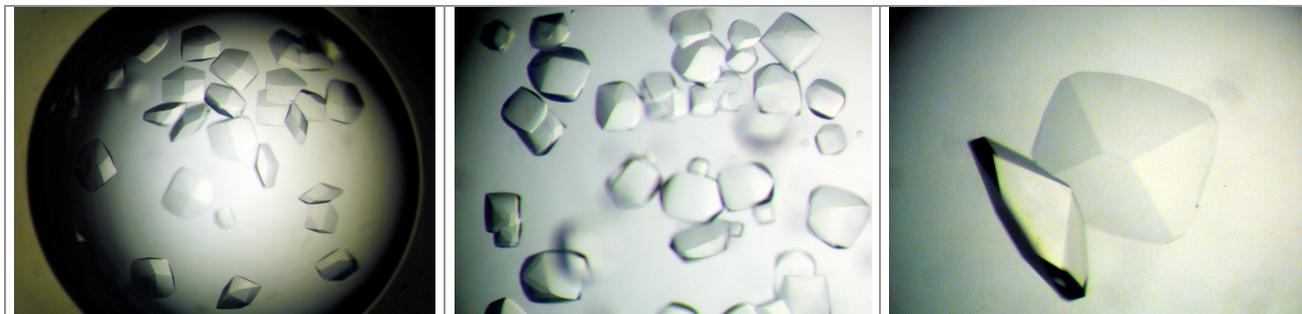


Figure 2.1.6 : Cristaux de lysozyme obtenus en présence de 300 mM de différents complexes. À gauche : Gd-HPDO3A, au milieu : Gd-DOTA, à droite : Gd-DO3A. On remarque l'influence de la présence des différents complexes sur la morphologie des cristaux. L'ajout du complexe mène à des cristaux "arrondis" pour le complexe Gd-DOTA et un aplatissement des cristaux pour le complexe Gd-DO3A. La taille des cristaux est de 150 à 300  $\mu\text{m}$ .

## II.2. Acquisition et traitement des données de diffraction

### II.2.1. Acquisition des données de diffraction

Pour la plupart des cristaux dérivés, des données de diffraction ont d'abord été enregistrées avec le rayonnement  $K\alpha$  du cuivre des générateurs du laboratoire. Ces données présentent l'avantage de combiner une résolution assez élevée, allant jusqu'à 1,64 Å, lorsque la qualité de diffraction le permet, et une diffusion anormale élevée du Gd avec une valeur de  $f''$  de 12  $e^-$ . Ces données, dès lors qu'elles sont enregistrées avec une redondance élevée, permettent d'estimer le taux de fixation des diffuseurs anomaux dans le cristal dérivé, à partir de facteurs statistiques de l'intégration des données et des cartes de Patterson anormales, et, le cas échéant, de déterminer les sites de fixation et de phaser les données par la méthode SAD.

Les dérivés ne présentant pas de fixation de complexe notable et dont le phasage SAD a échoué ne sont pas traités en détail par la suite.

Pour les dérivés où les complexes se fixent de manière significative, des données supplémentaires ont été enregistrées en utilisant du rayonnement synchrotron.

Grâce à son faisceau intense et géométriquement bien défini, le rayonnement synchrotron permet d'obtenir des données à une résolution maximale, ce qui est important pour l'affinement subséquent du mode de fixation. Nous avons enregistré des données MAD pour la majorité des dérivés. En effet, nous voulions tester l'influence de l'information de phase expérimentale sur l'affinement de la structure du complexe lié à la protéine. Cette information de phase est prise en compte au cours de l'affinement sous forme des coefficients de Hendrickson et Lattman (Hendrickson & Lattman, 1970) qui décrivent la distribution de probabilité de phase avant modification de densité. Pour le phasage SAD, cette distribution de probabilité de phases est bimodale et apporte moins d'information à l'affinement que le phasage avec la méthode MAD qui mène à une distribution de probabilité de phases sans cette ambiguïté. Les expériences de diffraction ont le plus souvent eu lieu sur la ligne de lumière

BM30A (FIP) de l'ESRF ou exceptionnellement sur la ligne de lumière ID29. Les résultats présentés par la suite sont le plus souvent issus de données enregistrées avec du rayonnement synchrotron.

## II.2.2. Stratégie de l'acquisition des données

Les données ont été enregistrées avec une redondance élevée pour obtenir des différences anormales estimées avec une bonne précision, car c'est sur elles que repose l'identification des sites de fixation et le phasage. En vue de l'intégration des données avec le programme XDS (Kabsch, 1993) (voir II.2.4. "traitement des données de diffraction") les données sont enregistrées en faisant tourner le cristal par pas angulaire de  $0,5^\circ$ . Sur la ligne BM30A, les contraintes techniques font qu'à la longueur d'onde du seuil de gadolinium ( $1,71 \text{ \AA}$ ) la résolution maximale des données n'est que des  $2,7 \text{ \AA}$ . Le phasage par SAD avec des données enregistrées sur BM30A présente donc l'inconvénient de ne pouvoir combiner des données à haute résolution et un signal anormal élevé. Comme, pour l'affinement des dérivés, la haute résolution des données apporte de l'information précieuse, nous nous sommes placé à des longueurs d'onde courtes, au seuil K du Se ( $\lambda = 0,98 \text{ \AA}$ ) ou du Br ( $\lambda = 0,92 \text{ \AA}$ ), par exemple (sur BM30A la résolution maximale des données à ces longueurs d'onde est de  $1,54 \text{ \AA}$  et de  $1,45 \text{ \AA}$  respectivement). Le coefficient de diffusion anormale du Gd à ces longueurs d'onde est  $f'' = 6 e^-$ . Pour le phasage par la méthode MAD, nous avons enregistré, en outre, des données au maximum de  $f''$  du seuil d'absorption  $L_{III}$  du gadolinium. Pour déterminer la longueur d'onde exacte du pic d'absorption nous avons enregistré un spectre de fluorescence autour du seuil d'absorption en utilisant une goutte de solution de complexe ou un cristal dérivé contenant du complexe de Gd. Les variations des facteurs  $f''$  et  $f'$  ont été déduites du spectre de fluorescence à l'aide du programme CHOOCH (Evans & Pettifer, 2001). Nous avons changé de protocole pour des cristaux ne diffractant qu'à relativement basse résolution et pour des données enregistrées sur la ligne de lumière ID29 où la résolution maximale des données à longue longueur d'onde est plus élevée que sur BM30A, et nous avons enregistré des données SAD au maximum de  $f''$  mais aussi des données MAD avec trois longueurs d'onde proche du seuil d'absorption. Sur les lignes de lumière utilisant un élément d'insertion (ID) nous avons fortement atténué le faisceau, pour éviter les dégâts d'irradiation tout en enregistrant des données avec une redondance élevée.

## II.2.3. Utilisation du robot de montage de cristaux sur BM30A

Depuis 2004, pour les expériences sur la ligne BM30A, nous avons systématiquement utilisé le robot de montage automatique des cristaux congelés (Ohana *et al.*, 2004). Les avantages sont la facilité et la sûreté du montage des échantillons et de leur récupération, résultant en un gain de temps d'expérience, car il n'est plus nécessaire de rentrer dans la cabane d'expérience pour changer l'échantillon. De cette manière, il est aisé de tester la qualité de diffraction de tous les cristaux congelés avec quelques clichés de diffraction et d'en retenir le meilleur. Un gain de temps considérable supplémentaire vient du changement de stratégie d'acquisition de données qui découle de l'utilisation du robot. Lorsqu'on prévoit des acquisitions de données avec différents dérivés et que l'on doit enregistrer des données à une courte longueur d'onde et au seuil d'absorption du Gd, plutôt que de garder un échantillon monté et de changer plusieurs fois la longueur d'onde on va, puisqu'on peut changer facilement d'échantillon, enregistrer d'abord des données avec tous les dérivés à la même longueur d'onde avant de

passer à une autre longueur d'onde. On minimise ainsi les changements de longueur d'onde, qui sont assez longs et souvent source de pannes, pouvant conduire à la perte du faisceau.

## II.2.4. Traitement des données de diffraction

Les intensités diffractées sont intégrées avec le programme XDS (Kabsch, 1993)\*. Ce programme permet une intégration particulièrement précise, car le profil des taches de diffraction est déduit des données expérimentales et ce, avec d'autant plus de précision, que le découpage angulaire de la rotation du cristal est fin. C'est pourquoi nous travaillons avec des pas angulaires de  $0,5^\circ$ . Pour obtenir une intégration des données optimales nous avons appliqué un protocole d'intégration des données qui procède en deux passes : affinement des paramètres cristallins et expérimentaux et détermination du profil des taches à partir du jeu de données entier, avant d'intégrer à nouveau les intensités en prenant comme valeurs de départ les paramètres affinés auparavant.

Les intensités intégrées sont mises à l'échelle avec le programme SCALA de la suite CCP4 (Collaborative Computational Project, 1994).

(\* À l'exception des données du dérivé urate oxydase + Gd-DO3A qui ont été traitées avec DENZO (Otwinowski & Minor, 1997))

## II.2.5. Conditions d'enregistrement et statistiques d'intégration de données pour les différents dérivés

Un résumé des conditions d'enregistrement et des statistiques d'intégration pour les différents dérivés, classés selon les différentes protéines, est donné dans les tableaux suivants. Les procédures de phasage et les affinements décrits par la suite reposent sur ces données.

Les indicateurs statistiques sont calculés par SCALA (CCP4) :

**$I/\sigma$**  : Rapport signal sur bruit

Les indicateurs  $R_{\text{fac}}$  et  $R_{\text{ano}}$  sont définis dans la partie abréviations du début du document

L'indicateur  $\langle I/\sigma \rangle$  est calculé par TRUNCATE (CCP4). Il correspond à la valeur de  $I/\sigma$  moyennée sur les réflexions équivalentes et prend en compte la redondance des données. Pour  $\langle I/\sigma \rangle$ , la largeur de la dernière tranche de résolution est de  $0,01 \text{ \AA}$ .

| Lysozyme                 | Gd-DOTA-BOM<br>300 mM             | Gd-DO3A<br>300 mM                 | Gd-HPDO3A<br>300 mM               | Gd-DOTA<br>300 mM                 |                     |
|--------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|---------------------|
| Groupe d'espace          | P <sub>4</sub> ,2 <sub>1</sub> ,2 |                     |
| Paramètres de maille [Å] | a = b = 77,8;<br>c = 37,4         | a = b = 77,3 ;<br>c = 37,3        | a = b = 77,1 ;<br>c = 38,3        | a = b = 77,2 ;<br>c = 38,4        |                     |
| λ [Å]                    | 0,98                              | 0,98                              | 0,98                              | 0,98                              | 1,54                |
| Source de RX             | FIP                               | FIP                               | FIP                               | FIP                               | labo.               |
| Résolution [Å]           | 1,53<br>(1,53-1,62)               | 1,54<br>(1,54-1,62)               | 1,53<br>(1,53-1,62)               | 1,54<br>(1,54-1,62)               | 1,64<br>(1,64-1,73) |
| I/σ                      | 5,4 (1,1)                         | 6,6 (2,1)                         | 6,3 (5,9)                         | 8,3 (3,6)                         | 6,7 (3,5)           |
| <I/σ>                    | 25,8 (3,0)                        | 31,0 (3,8)                        | 37,7 (8,0)                        | 50,9 (7,1)                        | 40,1 (12,3)         |
| Redondance               | 12,4 (6,4)                        | 12,5 (6,8)                        | 12,7 (6,4)                        | 12,7 (6,6)                        | 11,9 (9,2)          |
| R <sub>fac</sub>         | 0,068 (0,387)                     | 0,061 (0,26)                      | 0,057 (0,079)                     | 0,040<br>(0,15)                   | 0,055<br>(0,086)    |
| R <sub>ano</sub>         | 0,034 (0,138)                     | 0,039 (0,13)                      | 0,050 (0,076)                     | 0,035<br>(0,098)                  | 0,062<br>(0,096)    |
| Domaine angulaire        | 180°                              | 180°                              | 180°                              | 180°                              | 180°                |

| Urate oxydase            | Gd-DOTMA<br>300 mM  |                     | Gd-DTPA-BMA<br>300 mM |                     | Gd-DO3A<br>100 mM (denzo)      |
|--------------------------|---------------------|---------------------|-----------------------|---------------------|--------------------------------|
| Groupe d'espace          | I222                |                     | I222                  |                     | I222                           |
| Paramètres de maille [Å] | a = 79,1            | b = 95,0            | a = 79,6              | b = 95,0            | a = 80,4 b = 95,5<br>c = 104,6 |
| λ [Å]                    | 0,98                | 1,711               | 0,92                  | 1,711               | 0,98                           |
| Source de RX             | ID29                | ID29                | FIP                   | FIP                 | FIP                            |
| Résolution [Å]           | 1,34<br>(1,34-1,41) | 2,16<br>(2,16-2,28) | 1,45<br>(1,45-1,53)   | 2,70<br>(2,70-2,84) | 1,45<br>(1,45-1,53)            |
| I/σ                      | 9,1 (2,2)           | 7,4 (1,8)           | 9,1 (4,2)             | 13,9 (11,6)         | 10,7 (2,9)                     |
| <I/σ>                    | 24,3 (2,8)          | 21,8 (2,2)          | 30,6 (5,9)            | 35,0 (15,0)         | 31,2 (2,1)                     |
| Redondance               | 6,6 (4,0)           | 6,5 (4,0)           | 6,8 (3,8)             | 6,2 (3,6)           | 6,2 (4,2)                      |
| R <sub>fac</sub>         | 0,051<br>(0,32)     | 0,090<br>(0,37)     | 0,040<br>(0,13)       | 0,037<br>(0,051)    | 0,039 (0,19)                   |
| R <sub>ano</sub>         | 0,037<br>(0,23)     | 0,098<br>(0,24)     | 0,044<br>(0,097)      | 0,110<br>(0,148)    | 0,023 (0,13)                   |
| Domaine angulaire        | 180°                | 180°                | 180°                  | 180°                | 180° par pas de 1°             |

| <b>Thaumatine</b>        | Gd-DOTA 300 mM                   |                     | Gd-DOTA-BOM 300 mM               |                     |
|--------------------------|----------------------------------|---------------------|----------------------------------|---------------------|
| Groupe d'espace          | P4 <sub>1</sub> 2 <sub>1</sub> 2 |                     | P4 <sub>1</sub> 2 <sub>1</sub> 2 |                     |
| Paramètres de maille [Å] | a = b = 58,0 c = 150,8           |                     | a = b = 57,8 c = 150,5           |                     |
| λ [Å]                    | 0,92                             | 1,711               | 0,92                             | 1,711               |
| Source de RX             | FIP                              | FIP                 | FIP                              | FIP                 |
| Résolution [Å]           | 1,45<br>(1,45-1,53)              | 2,69<br>(2,69-2,83) | 1,45<br>(1,45-1,53)              | 2,69<br>(2,69-2,84) |
| I/σ                      | 6,7 (1,8)                        | 17,4 (10,0)         | 6,8 (2,6)                        | 8,1 (6,3)           |
| <I/σ>                    | 33,4 (4,0)                       | 37,3 (18,6)         | 32,9 (3,8)                       | 32,3 (9,1)          |
| Redondance               | 12,6 (6,7)                       | 5,4 (3,2)           | 13,0 (7,0)                       | 12,2 (6,8)          |
| R <sub>fac</sub>         | 0,056 (0,24)                     | 0,030 (0,056)       | 0,053 (0,24)                     | 0,066 (0,096)       |
| R <sub>ano</sub>         | 0,019 (0,079)                    | 0,054 (0,066)       | 0,029 (0,11)                     | 0,096 (0,013)       |
| Domaine angulaire        | 180°                             | 80°                 | 180°                             | 180°                |

| <b>Thaumatine</b>        | Gd-DTPA-BMA<br>300 mM            |                     | Gd-DTPA<br>300 mM                |                     | Gd-HPDO3A<br>100 mM              |
|--------------------------|----------------------------------|---------------------|----------------------------------|---------------------|----------------------------------|
| Groupe d'espace          | P4 <sub>1</sub> 2 <sub>1</sub> 2 |                     | P4 <sub>1</sub> 2 <sub>1</sub> 2 |                     | P4 <sub>1</sub> 2 <sub>1</sub> 2 |
| Paramètres de maille [Å] | a = b = 57,9 c = 150,4           |                     | a = b = 57,9 c = 150,1           |                     | a = b = 57,9<br>c = 150,5        |
| λ [Å]                    | 0,92                             | 1,711               | 0,92                             | 1,711               | 1,54                             |
| Source de RX             | FIP                              | FIP                 | FIP                              | FIP                 | labo.                            |
| Résolution [Å]           | 1,45<br>(1,45-1,53)              | 2,66<br>(2,66-2,81) | 1,45<br>(1,45-1,53)              | 2,67<br>(2,67-2,82) | 1,55<br>(1,55-1,63)              |
| I/σ                      | 9,0 (3,5)                        | 11,4 (9,6)          | 10,3 (5,9)                       | 14,8 (11,3)         | 9,2 (8,1)                        |
| <I/σ>                    | 44,9 (7,8)                       | 45,2 (18,0)         | 49,7 (11,0)                      | 57,9 (21,3)         | 57,9 (18,6)                      |
| Redondance               | 13,0 (7,2)                       | 12,0 (6,3)          | 12,8 (6,9)                       | 11,6 (6,0)          | 16,3 (13,9)                      |
| R <sub>fac</sub>         | 0,040<br>(0,15)                  | 0,043<br>(0,063)    | 0,037<br>(0,097)                 | 0,032<br>(0,052)    | 0,039 (0,092)                    |
| R <sub>ano</sub>         | 0,021<br>(0,053)                 | 0,058<br>(0,072)    | 0,025<br>(0,042)                 | 0,049<br>(0,062)    | 0,024 (0,037)                    |
| Domaine angulaire        | 180°                             | 180°                | 180°                             | 180°                | 217°                             |

| <b>Protéine X</b>          | Gd-HPDO3A<br>100 mM              | Gd-DOTA-BOM<br>100 mM            |                     | Gd-DTPA<br>100 mM                |
|----------------------------|----------------------------------|----------------------------------|---------------------|----------------------------------|
| Groupe d'espace            | P2 <sub>1</sub> 2 <sub>1</sub> 2 | P2 <sub>1</sub> 2 <sub>1</sub> 2 |                     | P2 <sub>1</sub> 2 <sub>1</sub> 2 |
| Paramètres de maille [Å]   | a = 48,4 b = 113,4<br>c = 75,1   | a = 49,1 b = 114,7<br>c = 75,9   |                     | a = 48,4 b = 113,5<br>c = 75,4   |
| $\lambda$ [Å]              | 1,711                            | 1,711                            | 1,07                | 1,711                            |
| Source de RX               | ID29                             | ID29                             | ID29                | ID23-1                           |
| Résolution [Å]             | 2,17<br>(2,17-2,29)              | 2,17<br>(2,17-2,29)              | 1,40<br>(1,40-1,47) | 2,03<br>(2,03-2,14)              |
| I/ $\sigma$                | 11,5 (6,5)                       | 10,7 (4,6)                       | 7,3 (1,1)           | 4,3 (2,4)                        |
| $\langle I/\sigma \rangle$ | 28,4 (8,3)                       | 25,4 (3,8)                       | 20,1 (1,3)          | 14,8 (3,2)                       |
| Redondance                 | 6,3 (4,5)                        | 5,7 (2,8)                        | 6,6 (4,7)           | 6,9 (5,9)                        |
| R <sub>fac</sub>           | 0,043 (0,11)                     | 0,048<br>(0,11)                  | 0,057<br>(0,67)     | 0,098 (0,26)                     |
| R <sub>ano</sub>           | 0,102 (0,14)                     | 0,065<br>(0,12)                  | 0,039<br>(0,45)     | 0,109 (0,16)                     |
| Domaine angulaire          | 180°                             | 180°                             | 180°                | 180°                             |

| <b>Protéine X</b>                      | Gd-DOTMA<br>100 mM               | Gd-DTPA-BMA<br>100 mM            | Gd-DO3A<br>100 mM                |
|--|----------------------------------|----------------------------------|----------------------------------|
| Groupe d'espace                        | P2 <sub>1</sub> 2 <sub>1</sub> 2 | P2 <sub>1</sub> 2 <sub>1</sub> 2 | P2 <sub>1</sub> 2 <sub>1</sub> 2 |
| Paramètres de maille [Å]               | a = 49,2 b = 115,1<br>c = 75,9   | a = 48,7 b = 114,0<br>c = 76,3   | a = 49,3 b = 115,7<br>c = 76,7   |
| $\lambda$ [Å]                          | 1,711                            | 1,711                            | 1,711                            |
| Source de RX                           | ID29                             | ID23-1                           | ID29                             |
| Résolution [Å]                         | 2,17<br>(2,17-2,29)              | 2,98<br>(2,98-3,14)              | 2,17<br>(2,17-2,29)              |
| I/ $\sigma$                            | 10,9 (4,7)                       | 7,2 (2,6)                        | 7,2 (2,2)                        |
| $\langle I/\sigma \rangle$             | 25,7 (6,2)                       | 19,8 (3,7)                       | 20,5 (2,6)                       |
| Redondance                             | 6,4 (4,4)                        | 6,5 (5,1)                        | 3,7 (2,7)                        |
| R <sub>fac</sub>                       | 0,053 (0,15)                     | 0,081 (0,24)                     | 0,068 (0,27)                     |
| R <sub>ano</sub>                       | 0,051 (0,102)                    | 0,051 (0,087)                    | 0,0109 (0,20)                    |
| Domaine angulaire                      | 180°                             | 180°                             | 100°                             |
| Nb sites                               | 1                                | 3                                | 6                                |
| Hauteur sites Fourier_ano [ $\sigma$ ] | 59                               | 22 ; 21 ; 11                     | 46 ; 39 ; 29 ; 17 ; 16 ; 15      |

| <b>Glucose isomérase</b>   | Gd-DO3A 100 mM              |                  | Gd-DTPA 100 mM              |                  |
|----------------------------|-----------------------------|------------------|-----------------------------|------------------|
| Groupe d'espace            | I222                        |                  | I222                        |                  |
| Paramètres de maille [Å]   | a = 93,0 b = 98,3 c = 102,6 |                  | a = 93,0 b = 98,2 c = 102,6 |                  |
| $\lambda$ [Å]              | 0,92                        | 1,711            | 0,92                        | 1,711            |
| Source de RX               | FIP                         | FIP              | FIP                         | FIP              |
| Résolution [Å]             | 1,44 (1,44-1,52)            | 2,69 (2,69-2,83) | 1,45 (1,45-1,53)            | 2,68 (2,68-2,83) |
| $I/\sigma$                 | 10,3 (1,4)                  | 17,1 (6,5)       | 8,3 (1,5)                   | 15,4 (8,1)       |
| $\langle I/\sigma \rangle$ | 35,6 (4,9)                  | 39,2 (8,3)       | 33,4 (3,8)                  | 41,8 (10,8)      |
| Redondance                 | 6,8 (3,5)                   | 6,5 (3,5)        | 6,9 (4,2)                   | 6,6 (4,1)        |
| $R_{\text{fac}}$           | 0,036 (0,29)                | 0,034 (0,088)    | 0,043 (0,29)                | 0,034 (0,072)    |
| $R_{\text{ano}}$           | 0,027 (0,13)                | 0,082 (0,102)    | 0,024 (0,112)               | 0,039 (0,060)    |
| Domaine angulaire          | 180°                        | 180°             | 180°                        | 180°             |

| <b>Glucose isomérase</b>   | Gd-DOTA-BOM<br>100 mM |                  | Gd-DOTMA<br>100 mM    |                  | Gd-DTPA-BMA<br>100 mM             |
|----------------------------|-----------------------|------------------|-----------------------|------------------|-----------------------------------|
| Groupe d'espace            | I222                  |                  | I222                  |                  | I222                              |
| Paramètres de maille [Å]   | a = 92,7<br>c = 102,3 | b = 98,3         | a = 92,7<br>c = 102,6 | b = 98,0         | a = 92,7<br>b = 97,9<br>c = 102,6 |
| $\lambda$ [Å]              | 0,92                  | 1,711            | 0,92                  | 1,711            | 1,54                              |
| Source de RX               | FIP                   | FIP              | FIP                   | FIP              | labo.                             |
| Résolution [Å]             | 1,43 (1,43-1,51)      | 2,68 (2,68-2,82) | 1,44 (1,44-1,51)      | 2,68 (2,68-2,82) | 1,75 (1,75-1,84)                  |
| $I/\sigma$                 | 13,0 (1,1)            | 21,0 (8,0)       | 9,2 (1,4)             | 12,2 (5,6)       | 7,8 (4,0)                         |
| $\langle I/\sigma \rangle$ | 30,1 (5,4)            | 44,3 (10,2)      | 23,8 (3,7)            | 25,2 (6,0)       | 20,1 (7,2)                        |
| Redondance                 | 4,7 (2,6)             | 6,6 (3,7)        | 6,8 (3,5)             | 5,3 (3,0)        | 7,3 (6,1)                         |
| $R_{\text{fac}}$           | 0,031 (0,33)          | 0,030 (0,078)    | 0,050 (0,31)          | 0,047 (0,11)     | 0,059 (0,16)                      |
| $R_{\text{ano}}$           | 0,024 (0,20)          | 0,033 (0,056)    | 0,023 (0,18)          | 0,038 (0,092)    | 0,022 (0,082)                     |
| Domaine angulaire          | 180°                  | 180°             | 180°                  | 145°             | 204°                              |

| <b>YGGV</b>                | Gd-DO3A 100 mM                   |                  |                  |
|----------------------------|----------------------------------|------------------|------------------|
| Groupe d'espace            | P4 <sub>3</sub> 2 <sub>1</sub> 2 |                  |                  |
| Paramètres de maille [Å]   | a = b = 78,7<br>c = 77,2         |                  |                  |
| $\lambda$ [Å]              | pic<br>1,711                     | infl1<br>1,712   | infl2<br>1,711   |
| Source de RX               | FIP                              | FIP              | FIP              |
| Résolution [Å]             | 2,68 (2,68-2,82)                 | 2,68 (2,68-2,83) | 2,68 (2,68-2,82) |
| $I/\sigma$                 | 8,9 (1,4)                        | 10,4 (1,6)       | 10,4 (1,6)       |
| $\langle I/\sigma \rangle$ | 26,2 (12,6)                      | 19,5 (10,8)      | 19,3 (10,6)      |
| Redondance                 | 12,2 (5,3)                       | 6,2 (3,4)        | 6,2 (3,3)        |
| $R_{\text{fac}}$           | 0,066 (0,47)                     | 0,050 (0,31)     | 0,051 (0,29)     |
| $R_{\text{ano}}$           | 0,106 (0,24)                     | 0,063 (0,28)     | 0,065 (0,23)     |
| Domaine angulaire          | 180°                             | 90°              | 90°              |

| <b>YGGV</b>              | Gd-DOTA-BOM 100 mM               |                     |                     | Gd-DOTMA100 mM                   |                     |
|--------------------------|----------------------------------|---------------------|---------------------|----------------------------------|---------------------|
| Groupe d'espace          | P4 <sub>3</sub> 2 <sub>1</sub> 2 |                     |                     | P4 <sub>3</sub> 2 <sub>1</sub> 2 |                     |
| Paramètres de maille [Å] | a = b = 79,7 c = 78,23           |                     |                     | a = b = 79,7 c = 78,3            |                     |
| $\lambda$ [Å]            | pic<br>1,712                     | infl1<br>1,712      | infl2<br>1,711      | pic<br>1,712                     | infl1<br>1,712      |
| Source de RX             | FIP                              | FIP                 | FIP                 | FIP                              | FIP                 |
| Résolution [Å]           | 2,68<br>(2,68-2,82)              | 2,68<br>(2,68-2,82) | 2,68<br>(2,68-2,82) | 2,69<br>(2,69-2,84)              | 2,70<br>(2,70-2,84) |
| I/ $\sigma$              | 7,7 (1,7)                        | 11,2 (3,4)          | 11,1 (3,2)          | 8,1 (5,5)                        | 9,7 (7,1)           |
| <I/ $\sigma$ >           | 21,1 (10,2)                      | 28,5 (13,6)         | 27,2 (13,5)         | 23,6 (8,2)                       | 23,1 (9,5)          |
| Redondance               | 8,1 (3,7)                        | 6,2 (3,2)           | 6,2 (3,2)           | 8,3 (4,7)                        | 5,8 (3,3)           |
| R <sub>fac</sub>         | 0,061<br>(0,24)                  | 0,037<br>(0,14)     | 0,039<br>(0,15)     | 0,061 (0,12)                     | 0,053<br>(0,086)    |
| R <sub>ano</sub>         | 0,073<br>(0,16)                  | 0,040<br>(0,11)     | 0,047<br>(0,13)     | 0,079 (0,12)                     | 0,042<br>(0,062)    |
| Domaine angulaire        | 120°                             | 90°                 | 90°                 | 120°                             | 82,5°               |

| <b>YEAZ</b>              | Gd-DOTMA 100 mM                               |              |
|--------------------------|---|--------------|
| Groupe d'espace          | P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> |              |
| Paramètres de maille [Å] | a = 76,3 b = 97,6 c = 141,9                   |              |
| $\lambda$ [Å]            | 1,711   | 0,98         |
| Source de RX             | FIP   | FIP          |
| Résolution [Å]           | 2,70  | 2,28         |
| I/ $\sigma$              | 13,6 (6,9)                                    | 9,3 (1,7)    |
| Redondance               | 4,7 (1,3)                                     | 5,0 (2,6)    |
| R <sub>fac</sub>         | 0,034<br>(0,096)                              | 0,054 (0,23) |
| R <sub>ano</sub>         | 0,043<br>(0,094)                              | 0,028 (0,13) |

## II.3. Obtention des phases expérimentales

### II.3.1. Estimation de la fixation et détermination des sites de fixation

On peut estimer l'éventuelle fixation des diffuseurs anomaux dans un cristal dérivé dès les premières étapes d'intégration et de mise à l'échelle des données de diffraction, à partir des intensités intégrées et des mises à l'échelle de données d'un jeu essentiellement complet. Nous avons utilisé les programmes XDS et SCALA qui calculent des indicateurs permettant d'estimer si le signal anormal est significatif devant l'imprécision des mesures ( $\sigma_{\text{norm}}/\sigma_{\text{ano}}$  et R<sub>ano</sub>, voir la partie abréviations). L'inspection de la carte de Patterson anormale (Patterson, 1935) (voir Annexe I) permet d'estimer le nombre des sites de fixation et leur taux d'occupation. Au cours de l'acquisition des données ces indicateurs aideront à juger du pouvoir de phasage du dérivé et à décider de la suite des enregistrements.

Pour des données avec un signal anormal suffisant, des programmes basés sur les méthodes directes ou les méthodes de Patterson permettent de déterminer les sites de fixation. Nous

avons utilisé les programmes SHELXD (Schneider & Sheldrick, 2002; Sheldrick, 1998) et SnB (Weeks & Miller, 1999)).

En vue de l'affinement subséquent pour des protéines de structure connue nous avons calculé les coordonnées des diffuseurs anomaux par synthèse de Fourier anormale afin que les sites soient placés dans la maille avec la même origine que le modèle de la protéine utilisé pour l'affinement. À cet effet, le modèle initial de protéine est replacé dans la maille à partir des données expérimentales en effectuant un affinement de corps rigide, le modèle résultant servant de modèle pour calculer les phases de la synthèse de Fourier.

### II.3.2. Détermination des phases

En utilisant la position des diffuseurs anomaux maintenant connue, les phases sont calculées par la méthode MAD ou SAD à l'aide du programme SHARP (de La Fortelle & Bricogne, 1997; Bricogne *et al.*, 2003). Le programme utilise la méthode du maximum de vraisemblance (maximum likelihood) appliquée à la distribution de probabilité des facteurs de structure. Celle-ci est exprimée comme fonction implicite des paramètres qui sont affinés. Les paramètres des atomes lourds qui sont affinés sont leurs coordonnées, leur facteur d'agitation thermique (isotrope dans un premier temps, anisotrope pour des sites fortement occupés dans un deuxième temps) et leur occupation. Nous choisissons de ne pas affiner les parties réelles et imaginaires du facteur de diffusion atomique, mais nous introduisons les valeurs expérimentales déduites du spectre de fluorescence ou les valeurs tabulées pour les longueurs d'onde hors seuil. D'autres paramètres globaux affinés sont des facteurs d'échelle et le défaut d'isomorphisme (l.o.i. : lack of isomorphism), paramètre qui tient compte des erreurs de modélisation, en particulier celle provenant du modèle d'atomes lourds.

Le programme SHARP permet l'analyse de cartes résiduelles obtenues après phasage afin de détecter d'éventuels sites anomaux supplémentaires et de mettre en évidence l'anisotropie des sites affinés. Il donne une estimation de la fiabilité des valeurs de  $f'$  et  $f''$  utilisées.

Les phases calculées par SHARP sont améliorées par nivellement de solvant et correspondance d'histogramme avec les programmes SOLOMON (Abrahams, 1997) et DM (Cowtan & Main, 1996) accessibles directement depuis l'interface de SHARP. La valeur de la proportion de solvant utilisée est estimée par le programme TRUNCATE (CCP4).

### II.3.3. Procédé de phasage des données obtenues sur BM30A

Typiquement, nous avons entrepris le phasage avec deux jeux de données du même dérivé : un jeu à haute résolution ( $\sim 1,5$  Å) avec un signal anomal relativement faible ( $f'' \sim 6 e^-$ ) et un jeu à basse résolution ( $\sim 2,7$  Å) enregistré au maximum de  $f''$  ( $f'' \sim 28 e^-$ ). Nous n'avons pas mis à l'échelle les deux jeux avant leur utilisation par le programme SHARP, qui effectue cette mise à l'échelle. Avec de telles données, pourvu qu'elles proviennent d'un dérivé adéquat, on obtient généralement des bonnes phases expérimentales jusqu'à 2,7 Å de résolution. Pour la haute résolution, où le phasage (SAD) repose sur des différences anormales relativement faibles, les valeurs des facteurs de qualité (FOM, voir abréviations pour la définition) sont relativement basses, indiquant l'imprécision des phases à cette résolution. L'étape de modification de densité permet l'extension des phases expérimentales à haute résolution. Après nivellement de solvant, les valeurs de FOM sont élevées dans toutes les gammes de résolution, témoignant de la bonne qualité des phases. Le tableau 2.3.1 montre les valeurs du facteur de qualité en fonction de la résolution pour le dérivé de la thaumatine avec le Gd-DOTA-BOM, pour lequel le phasage a été effectué selon les conditions décrites ci-

dessus. On remarque la variation brusque autour de 2,7 Å du facteur de qualité avant aplatissement de solvant pour les réflexions générales et à phase contrainte. Les valeurs très élevées du facteur de qualité après aplatissement de solvant, pour toute la gamme de résolution, témoignent de l'efficacité du nivellement de solvant et de l'extension des phases.

| Dmin    | Dmax | FOMacen | FOMcen  | RMIN   | RMAX | MEAN FOM (obs.) | MEAN FOM (calc) | MEAN FOM (comb) | MEAN DPHI lobs-call | MEAN DPHI lobs-coml | MEAN DPHI lcal-coml |
|---------|------|---------|---------|--------|------|-----------------|-----------------|-----------------|---------------------|---------------------|---------------------|
| 40.87   | 6.41 | 0.86906 | 0.60719 |        |      |                 |                 |                 |                     |                     |                     |
| 6.41    | 4.56 | 0.82563 | 0.52655 |        |      |                 |                 |                 |                     |                     |                     |
| 4.56    | 3.73 | 0.78766 | 0.50793 |        |      |                 |                 |                 |                     |                     |                     |
| 3.73    | 3.23 | 0.78263 | 0.46814 |        |      |                 |                 |                 |                     |                     |                     |
| 3.23    | 2.89 | 0.79107 | 0.46924 |        |      |                 |                 |                 |                     |                     |                     |
| 2.89    | 2.64 | 0.71328 | 0.39773 |        |      |                 |                 |                 |                     |                     |                     |
| 2.64    | 2.45 | 0.54838 | 0.14333 |        |      |                 |                 |                 |                     |                     |                     |
| 2.45    | 2.29 | 0.52836 | 0.13979 |        |      |                 |                 |                 |                     |                     |                     |
| 2.29    | 2.16 | 0.50771 | 0.12580 |        |      |                 |                 |                 |                     |                     |                     |
| 2.16    | 2.05 | 0.48754 | 0.13804 |        |      |                 |                 |                 |                     |                     |                     |
| 2.05    | 1.95 | 0.46776 | 0.13695 |        |      |                 |                 |                 |                     |                     |                     |
| 1.95    | 1.87 | 0.45226 | 0.13653 |        |      |                 |                 |                 |                     |                     |                     |
| 1.87    | 1.80 | 0.43228 | 0.14412 |        |      |                 |                 |                 |                     |                     |                     |
| 1.80    | 1.73 | 0.40998 | 0.14089 |        |      |                 |                 |                 |                     |                     |                     |
| 1.73    | 1.67 | 0.37179 | 0.14667 |        |      |                 |                 |                 |                     |                     |                     |
| 1.67    | 1.62 | 0.34329 | 0.14235 |        |      |                 |                 |                 |                     |                     |                     |
| 1.62    | 1.57 | 0.31792 | 0.15528 |        |      |                 |                 |                 |                     |                     |                     |
| 1.57    | 1.53 | 0.29974 | 0.14728 |        |      |                 |                 |                 |                     |                     |                     |
| 1.53    | 1.49 | 0.24998 | 0.13686 |        |      |                 |                 |                 |                     |                     |                     |
| 1.49    | 1.45 | 0.18886 | 0.12027 |        |      |                 |                 |                 |                     |                     |                     |
| OVERALL |      | 0.44604 | 0.24266 |        |      |                 |                 |                 |                     |                     |                     |
|         |      |         |         | 100.00 | 6.99 | 0.772           | 0.000           | 0.757           | 63.6                | 32.8                | 58.1                |
|         |      |         |         | 6.99   | 5.30 | 0.748           | 0.777           | 0.888           | 48.0                | 39.6                | 24.5                |
|         |      |         |         | 5.30   | 4.44 | 0.729           | 0.883           | 0.921           | 41.6                | 37.1                | 13.2                |
|         |      |         |         | 4.44   | 3.90 | 0.724           | 0.888           | 0.931           | 42.2                | 38.6                | 12.6                |
|         |      |         |         | 3.90   | 3.51 | 0.734           | 0.862           | 0.927           | 44.3                | 35.9                | 15.8                |
|         |      |         |         | 3.51   | 3.23 | 0.727           | 0.871           | 0.923           | 43.9                | 36.7                | 15.6                |
|         |      |         |         | 3.23   | 3.00 | 0.740           | 0.867           | 0.924           | 41.5                | 35.5                | 15.5                |
|         |      |         |         | 3.00   | 2.81 | 0.756           | 0.863           | 0.920           | 41.0                | 33.4                | 16.3                |
|         |      |         |         | 2.81   | 2.66 | 0.652           | 0.870           | 0.910           | 49.4                | 43.6                | 13.4                |
|         |      |         |         | 2.66   | 2.53 | 0.489           | 0.886           | 0.909           | 56.2                | 52.6                | 8.3                 |
|         |      |         |         | 2.53   | 2.41 | 0.483           | 0.892           | 0.909           | 53.4                | 50.6                | 7.1                 |
|         |      |         |         | 2.41   | 2.31 | 0.472           | 0.893           | 0.909           | 56.2                | 53.2                | 6.9                 |
|         |      |         |         | 2.31   | 2.23 | 0.459           | 0.899           | 0.913           | 56.4                | 53.3                | 6.5                 |
|         |      |         |         | 2.23   | 2.15 | 0.453           | 0.905           | 0.919           | 56.7                | 53.8                | 5.9                 |
|         |      |         |         | 2.15   | 2.08 | 0.443           | 0.907           | 0.916           | 55.1                | 52.9                | 5.3                 |
|         |      |         |         | 2.08   | 2.01 | 0.428           | 0.897           | 0.908           | 55.5                | 52.8                | 5.4                 |
|         |      |         |         | 2.01   | 1.95 | 0.427           | 0.903           | 0.911           | 59.4                | 57.1                | 4.9                 |
|         |      |         |         | 1.95   | 1.90 | 0.414           | 0.889           | 0.897           | 58.0                | 55.6                | 5.5                 |
|         |      |         |         | 1.90   | 1.85 | 0.396           | 0.880           | 0.891           | 57.9                | 55.6                | 5.4                 |
|         |      |         |         | 1.85   | 1.80 | 0.394           | 0.883           | 0.893           | 58.7                | 56.5                | 5.0                 |
|         |      |         |         | 1.80   | 1.76 | 0.393           | 0.882           | 0.888           | 59.0                | 57.3                | 4.9                 |
|         |      |         |         | 1.76   | 1.72 | 0.365           | 0.882           | 0.887           | 59.9                | 57.8                | 4.8                 |
|         |      |         |         | 1.72   | 1.68 | 0.335           | 0.849           | 0.854           | 60.6                | 58.8                | 5.7                 |
|         |      |         |         | 1.68   | 1.65 | 0.335           | 0.861           | 0.866           | 63.3                | 61.8                | 5.3                 |
|         |      |         |         | 1.65   | 1.61 | 0.309           | 0.859           | 0.863           | 65.5                | 64.4                | 4.6                 |
|         |      |         |         | 1.61   | 1.58 | 0.304           | 0.847           | 0.851           | 64.1                | 62.9                | 4.7                 |
|         |      |         |         | 1.58   | 1.55 | 0.290           | 0.840           | 0.842           | 68.1                | 66.6                | 4.6                 |
|         |      |         |         | 1.55   | 1.53 | 0.279           | 0.829           | 0.831           | 67.3                | 66.2                | 5.2                 |
|         |      |         |         | 1.53   | 1.50 | 0.246           | 0.753           | 0.758           | 70.8                | 69.3                | 8.4                 |

Tableau 2.3.1 : Facteurs de qualité pour le phasage du dérivé de la thaumatine avec Gd-DOTA-BOM en fonction de la résolution. **À gauche** : Les valeurs des facteurs de qualité des réflexions générales (FOMacen) et des réflexions à phase contrainte (FOMcen) du phasage par le programme SHARP. On remarque la brusque variation autour d'une résolution de 2,7 Å et des valeurs de FOM relativement basses pour la haute résolution. **À droite** : Tableau des facteurs de qualité après amélioration (et extension) des phases avec le programme DM. On remarque le saut autour de 2,7 Å pour le facteur de qualité moyen avant nivellement de solvant, alors qu'après DM les facteurs de qualité sont très élevés pour la haute et la basse résolution.

### II.3.4. Évaluation du succès du phasage

Le facteur de qualité donne une première indication de la qualité du phasage, mais un critère plus décisif est l'interprétabilité de la carte de densité électronique expérimentale (Ramagopal *et al.*, 2003). La qualité de la carte est déterminée par inspection visuelle ou en essayant de construire automatiquement le modèle de la protéine, en utilisant, par exemple, les programmes ARP/wARP (Lamzin *et al.*, 2001) ou RESOLVE (Terwilliger, 2002).

## II.3.5. Résultats de phasage avec les différents dérivés

Les statistiques de détermination des phases avec le programme SHARP et de modification de densité avec les programmes SOLOMON et DM, obtenues pour les différents dérivés, sont résumées dans les tableaux ci-dessous.

Sauf mention contraire :

- Le nombre de cycles d'amélioration des phases effectués par le programme DM est de 20.
- La longueur d'onde 1,71 Å correspond au maximum de  $f''$  au seuil d'absorption du gadolinium.

- **FOM, FOMshasol, FOMdm** : valeurs respectives du facteur de qualité calculées par le programme SHARP pour les phases initiales, pour les phases après modification de densité par SOLOMON et pour les phases après DM.

- **FOM (2,7)** désigne la valeur du facteur de qualité des réflexions générales, calculée par SHARP pour la gamme de résolution jusqu'à 2,7 Å, qui correspond à la limite de résolution pour des données enregistrées au seuil sur BM30A. Il en va de même pour FOM (2,2) associé à la ligne ID29.

- **Qsha** et **Bsha** désignent le taux d'occupation et le facteur d'agitation thermique des atomes lourds affinés par SHARP. La mise à une échelle absolue des données dépend de leur résolution. Pour des données à basse résolution, les valeurs affinées des occupations anormales n'ont pas de signification physique absolue (Bricogne *et al.*, 2003).

- **R<sub>libre</sub>** : facteur résiduel dans l'espace direct calculé par DM (voir abréviations).

- **pic, inf11, inf12** désignent les longueurs d'onde du seuil d'absorption correspondant au maximum de  $f''$ , au minimum de  $f'$  et à une troisième longueur d'onde, plus courte.

| Lysozyme                             | Gd-DOTA-BOM<br>300 mM | Gd-DO3A<br>300 mM         | Gd-HPDO3A<br>300 mM | Gd-DOTA<br>300 mM |
|--------------------------------------|-----------------------|---------------------------|---------------------|-------------------|
| Méthode                              | SAD                   | SAD                       | SAD                 | MAD               |
| $\lambda$ [Å]                        | 0,98                  | 0,98                      | 0,98                | 0,98; 1,54        |
| Source de RX                         | FIP                   | FIP                       | FIP                 | FIP; labo         |
| Résolution [Å]                       | 55,1-1,53             | 54,6-1,54                 | 54,5- 1,53          | 54,6- 1,54        |
| Nb sites                             | 1                     | 4                         | 2                   | 2                 |
| Qsha                                 | 0,78                  | 0,75; 1,02; 0,35;<br>0,16 | 1,06; 0,85          | 0,82; 0,52        |
| Bsha [Å <sup>2</sup> ]               | 27                    | 27; 45; 23; 15            | 15; 12              | 16; 14            |
| FOM                                  | 0,246                 | 0,303                     | 0,581               | 0,579             |
| FOM (2,7)                            |                       |                           |                     | 0,697             |
| FOMshasol                            | 0,749                 | 0,754                     | 0,906               | 0,884             |
| FOMdm                                | 0,710                 | 0,698                     | 0,648               | 0,773             |
| R <sub>libre</sub> (nb cycles<br>dm) | 0,216 (2)             | 0,180 (3)<br>0,185 (2)    | 0,138(1)            | 0,172 (3)         |

| <b>Urate oxydase</b>              | Gd-DOTMA<br>300 mM | Gd-DTPA-BMA<br>300 mM | Gd-DO3A<br>100 mM |
|-----------------------------------|--------------------|-----------------------|-------------------|
| Méthode                           | MAD                | MAD                   | SAD               |
| $\lambda$ [Å]                     | 0,98; 1,71         | 0,92; 1,71            | 0,98              |
| Source de RX                      | ID29               | FIP                   | FIP               |
| Résolution [Å]                    | 70,2 - 1,34        | 70,2 - 1,45           | 70,5 - 1,45       |
| Nb sites                          | 1                  | 3                     | 3                 |
| Qsha                              | 1,08; 0,31; 0,22   | 0,87; 0,66; 0,82      | 0,03; 0,02; 0,17  |
| Bsha [Å <sup>2</sup> ]            | 25; 30; 30         | 14; 15; 14            | 30; 30; 30        |
| FOM                               | 0,232              | 0,526                 | 0,071             |
| FOM (2,7)                         | 0,565              | 0,668                 |                   |
| FOMshasol                         | 0,855              | 0,915                 | 0,767             |
| FOMdm                             | 0,873              | 0,838                 | 0,667             |
| R <sub>libre</sub> (nb cycles dm) | 0,207 (3)          | 0,174 (3))            | 0,203 (2)         |

| <b>Thaumatine</b>                   | Gd-DOTA<br>300 mM       | Gd-DOTA-BOM<br>300 mM   | Gd-DTPA-BMA<br>300 mM | Gd-DTPA<br>300 mM     | Gd-HPDO3A<br>100 mM  |
|-------------------------------------|-------------------------|-------------------------|-----------------------|-----------------------|----------------------|
| Méthode                             | MAD                     | MAD                     | MAD                   | MAD                   | SAD                  |
| $\lambda$ [Å]                       | 0,92<br>1,71            | 0,92<br>1,71            | 0,92<br>1,71          | 0,92<br>1,71          | 1,54                 |
| Source de RX                        | FIP                     | FIP                     | FIP                   | FIP                   | labo                 |
| Résolution [Å]                      | 54,2 - 1,45             | 53,9 - 1,45             | 54,1 - 1,45           | 54,0- 1,45            | 54,2 - 1,55          |
| Nb sites                            | 1                       | 2                       | 1                     | 1                     | 3                    |
| Qsha                                | 0,61                    | 0,93; 0,48              | 0,70                  | 0,48                  | 0,15; 0,29 ;<br>0,14 |
| Bsha [Å <sup>2</sup> ]              | 29                      | 22; 17                  | 28                    | 16                    | 15 ; 30 ; 23         |
| FOM                                 | 0,260                   | 0,420                   | 0,353                 | 0,319                 | 0,210                |
| FOM (2,7)                           | 0,491                   | 0,713                   | 0,567                 | 0,506                 |                      |
| FOMshasol                           | 0,876                   | 0,912                   | 0,899                 | 0,913                 | 0,875                |
| FOMdm                               | 0,916                   | 0,882                   | 0,891                 | 0,894                 | 0,825                |
| Nb cycles dm                        |                         |                         |                       |                       | 100                  |
| R <sub>libre</sub> ( n b cycles dm) | 0,108 (16)<br>0,135 (3) | 0,090 (5)<br>0,094 (3)) | 0,128 (3)             | 0,101(3)<br>0,096 (5) | 0,113 (3)            |

| <b>Protéine X</b>                 | Gd-HPDO3A<br>100 mM             | Gd-DOTA-BOM<br>100 mM  | Gd-DTPA<br>100 mM |
|-----------------------------------|---------------------------------|------------------------|-------------------|
| Méthode                           | SAD                             | MAD                    | SAD               |
| $\lambda$ [Å]                     | 1,71                            | 1,71; 1,07             | 1,71              |
| Source de RX                      | ID29                            | ID29                   | ID23-1            |
| Résolution [Å]                    | 75,2 - 2,17                     | 75,8 - 1,40            | 75,378 - 2,030    |
| Nb sites                          | 5                               | 1                      | 1                 |
| Qsha                              | 0,23; 0,20; 0,20; 0,35;<br>0,07 | 1,06                   | 0,89              |
| Bsha [Å <sup>2</sup> ]            | 40; 42; 45; 61; 47              | 41                     | 39                |
| FOM                               | 0,426                           | 0,169                  | 0,259             |
| FOM (2,2)                         |                                 | 0,349                  |                   |
| FOMshasol                         | 0,847                           | 0,780                  | 0,807             |
| FOMdm                             | 0,762                           | 0,842                  | 0,760             |
| R <sub>libre</sub> (nb cycles dm) | 0,103 (3)<br>0,099 (4)          | 0,195 (3)<br>0,191 (5) | 0,113 (2)         |

| <b>Glucose isomérase</b>          | Gd-DO3A<br>100 mM    | Gd-DTPA<br>100 mM | Gd-DOTA-BOM<br>100 mM | Gd-DOTMA<br>100 mM | Gd-DTPA-BMA<br>100 mM |
|-----------------------------------|----------------------|-------------------|-----------------------|--------------------|-----------------------|
| Méthode                           | MAD                  | MAD               | MAD                   | SAD                | SAD                   |
| $\lambda$ [Å]                     | 0,92<br>1,71         | 0,92; 1,71        | 0,92<br>1,71          | 1,54               | 1,54                  |
| Source de RX                      | FIP                  | FIP               | FIP                   | labo               | labo                  |
| Résolution [Å]                    | 70,7 - 1,44          | 70,7 - 1,45       | 70,7 - 1,43           | 70,7 - 1,76        | 70,7 - 1,76           |
| Nb sites                          | 3                    | 2                 | 2                     | 1                  | 3                     |
| Qsha                              | 0,78<br>0,75<br>0,46 | 0,47<br>0,12      | 0,21<br>0,22          | 0,40               | 0,13<br>0,08<br>0,25  |
| Bsha [Å <sup>2</sup> ]            | aniso<br>aniso<br>20 | 22,05<br>1,00     | 16<br>13              | 28                 | 43<br>4<br>25         |
| FOM                               | 0,363                | 0,219             | 0,150                 | 0,120              | 0,190                 |
| FOM (2,7)                         | 0,573                | 0,439             | 0,314                 |                    |                       |
| FOMshasol                         | 0,858                | 0,878             | 0,818                 | 0,790              | 0,850                 |
| FOMdm                             | 0,844                | 0,877             | 0,853                 | 0,750              | 0,782                 |
| Nb cycles dm                      | 100                  | 100               | 100                   | 100                | 100                   |
| R <sub>libre</sub> (nb cycles dm) | 0,174 (3)            | 0,123 (3)         | 0,199 (3)             | 0,147 (2)          | 0,191 (3)             |

| <b>YGGV</b>                       | Gd-DO3A<br>100 mM      | Gd-DOTA-BOM<br>100 mM  | Gd-DOTMA<br>100 mM     |
|-----------------------------------|------------------------|------------------------|------------------------|
| Méthode                           | MAD                    | MAD                    | MAD                    |
| $\lambda$ [Å]                     | pic; infl1; infl2      | pic; infl1; infl2      | pic; infl1             |
| Source de RX                      | FIP                    | FIP                    | FIP                    |
| Résolution [Å]                    | 55,9 - 2,68            | 55,9 - 2,68            | 55,9 - 2,70            |
| Nb sites                          | 4                      | 4                      | 3                      |
| Qsha                              | 0,71; 0,37; 0,13; 0,18 | 0,28; 0,26; 0,21; 0,21 | 0,20; 0,13; 0,13       |
| Bsha [Å <sup>2</sup> ]            | 64; 92; 33; 45         | -----                  | 33; 66; 63             |
| FOM                               | 0,667                  | 0,484                  | 0,677                  |
| FOMshasol                         | 0,828                  | 0,789                  | 0,902                  |
| FOMdm                             | 0,771                  | 0,648                  | 0,847                  |
| Nb cycles dm                      | 25                     | 30                     | 22                     |
| R <sub>libre</sub> (nb cycles dm) | 0,111 (3)              | 0,161 (1)              | 0,100 (4)<br>0,107 (3) |

### II.3.6. Exemples de cartes de densité électronique expérimentale

Nous ne pouvons présenter ici l'ensemble des cartes de densité électronique expérimentale pour chacun des dérivés étudiés. Nous avons choisi de ne présenter que trois cartes, deux d'entre-elles correspondant à des dérivés de protéines test et l'autre à une protéine de structure initialement inconnue.

Les phases du premier dérivé ont été obtenues par phasage SAD avec des données enregistrées au seuil K du Se où  $f''$  prend la valeur relativement faible de 6 e<sup>-</sup>. Les phases des deux dérivés suivants ont été obtenues par phasage MAD avec un jeu au pic de  $f''$  et un jeu à courte longueur d'onde et haute résolution (statistiques de phasage pour les trois dérivés voir tableaux ci-dessus). Les cartes de densité électronique correspondent aux phases obtenues après SOLOMON.

La figure 2.3.1 montre la carte expérimentale pour le dérivé du lysozyme avec le Gd-HPDO3A (300 mM). Malgré le signal anomal relativement faible, des phases d'excellente qualité ont été obtenues et la carte expérimentale est très bien définie.

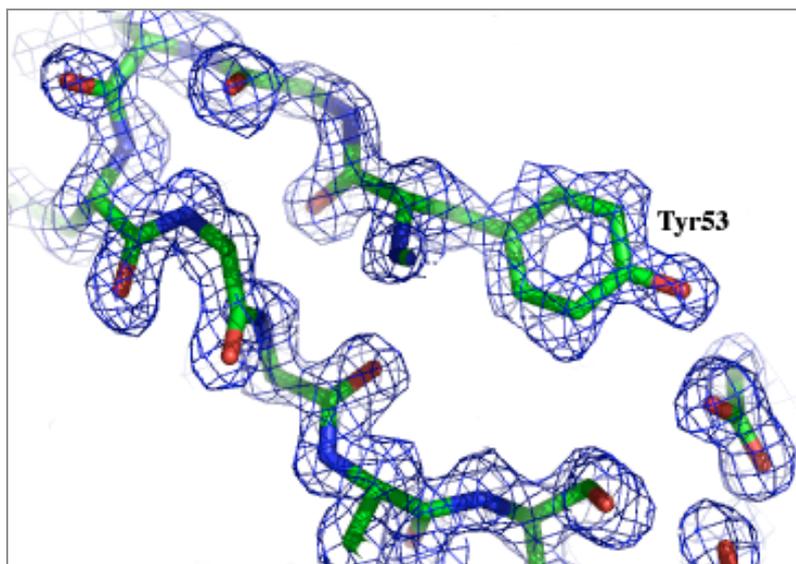


Figure 2.3.1 : Carte de densité électronique (contour  $1\sigma$ ) après SOLOMON pour le dérivé du lysozyme avec le Gd-HPDO3A calculée avec des données à une résolution de 1,53 Å.

La figure 2.3.2 montre la carte expérimentale pour le dérivé de l'urate oxydase avec le Gd-DTPA-BMA (300 mM). La figure montre l'inhibiteur de la protéine, 8-azaxanthine, et les résidus avec lesquels celui-ci interagit. La qualité de la carte est telle qu'on reconnaît facilement les résidus protéiques et l'inhibiteur. Les molécules d'eau sont parfaitement dessinées.

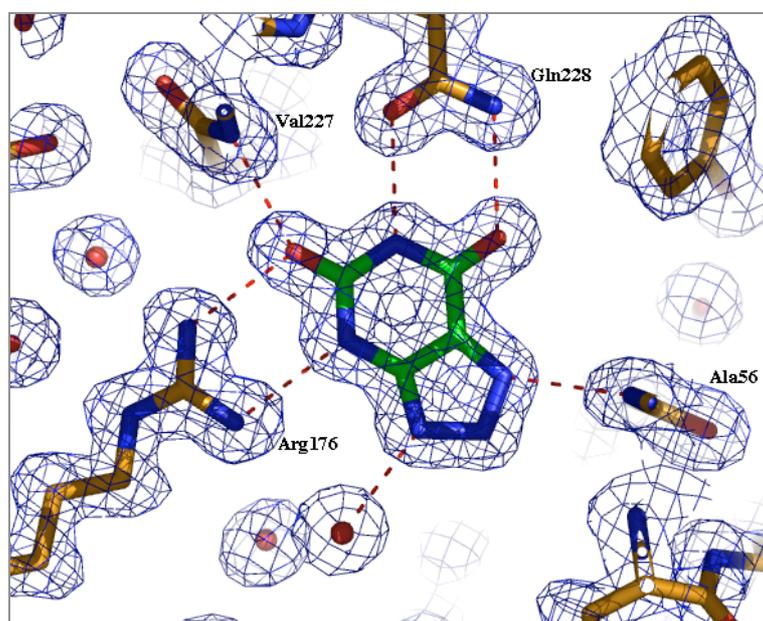


Figure 2.3.2 : Carte de densité électronique (contour  $1\sigma$ ) après SOLOMON pour le dérivé de l'urate oxydase avec le Gd-DTPA-BMA calculée avec des données à une résolution de 1,45 Å. La carte montre l'inhibiteur 8-azaxanthine et ses interactions avec les résidus voisins.

La figure 2.3.3 montre la carte de densité électronique expérimentale du dérivé de la protéine X avec le Gd-DOTA-BOM. La structure de la protéine a été résolue récemment avec un dérivé obtenu avec le complexe Gd-HPDO3A. La densité des résidus protéiques et celle des molécules d'eau sont très bien définies, ce qui permet une construction de modèle sans ambiguïté.

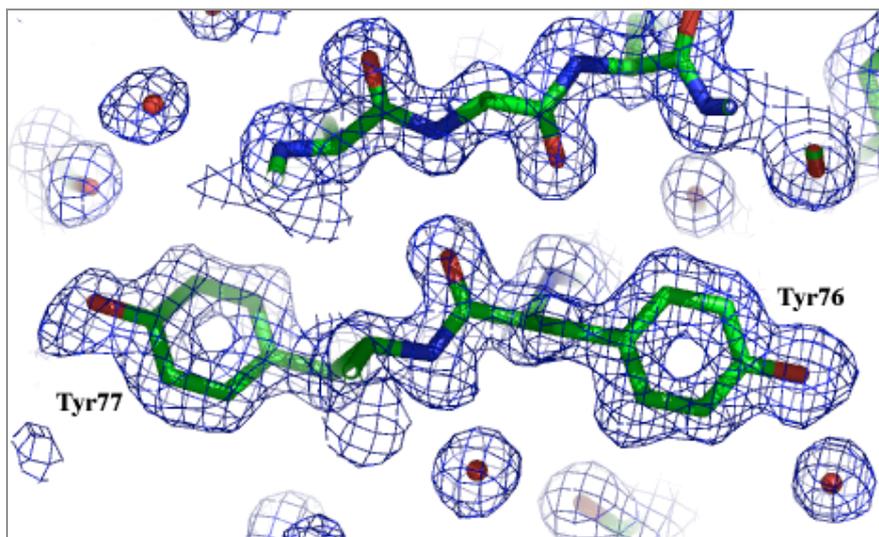


Figure 2.3.3 : Carte de densité électronique (contour 1 $\sigma$ ) après SOLOMON pour le dérivé protéine X + Gd-DOTA-BOM calculée avec des données à une résolution de 1,40 Å.

### II.3.7. Exemple de résolution de structure d'une protéine de structure jusque-là inconnue

#### Résolution de la structure de la protéine yeaZ d'*E. coli* (4\*217 résidus/ua)

(Jeudy *et al.*, 2005) (en collaboration avec Chantal Abergel et Sandra Jeudy, IGS-Marseille)

La préparation du dérivé yeaZ par trempage avec le complexe Gd-DOTMA est décrite dans le chapitre II.1.3. Les statistiques d'acquisition et de traitement des données sont données dans le chapitre II.2.5. Trois jeux de données redondants ont été enregistrés sur la ligne BM30A au maximum de  $f''$ , au minimum de  $f'$  et à courte longueur d'onde respectivement. Six sites de fixation ont été trouvés par le programme SOLVE (Terwilliger & Berendzen, 1999). Le phasage, effectué avec le programme SHARP, a permis l'identification de deux sites de fixation supplémentaires. Les phases obtenues sont d'excellente qualité comme l'indiquent les valeurs élevées des facteurs de qualité avant et après DM.

|                                |  |
|--------------------------------|--|
| <b>YeaZ</b>                    | Gd-DOTMA<br>100 mM                             |
| Méthode                        | MAD  |
| $\lambda$ [Å]                  | pic; infl; 0,98                                |
| Source de RX                   | FIP  |
| Résolution [Å]                 | 80,32 - 2,28                                   |
| Nb sites                       | 8  |
| Qsha                           | 0,55; 0,35; 0,32; 0,10; 0,21; 0,22; 0,20; 0,20 |
| Bsha [Å <sup>2</sup> ]         | 20; 62; 103; 1; 49; 49; 49; 49                 |
| FOM                            | 0,436  |
| FOMshasol                      | 0,821  |
| FOMdm                          | 0,791  |
| R <sub>libre</sub> (nb cycles) | 0,123 (4); 0,129 (3)                           |

Tableau 2.3.2 : Statistiques de phasage du dérivé de yeaZ avec Gd-DOTMA.

Le tableau 2.3.3 donne les valeurs des facteurs de qualité en fonction de la résolution.

| Dmin    | Dmax | FOMacen | FOMcen  | RMIN   | RMAX  | MEAN FOM (obs.) | MEAN FOM (calc) | MEAN FOM (comb) | MEAN DPHI (lobs-call) | MEAN DPHI (lobs-coml) | MEAN DPHI (lcal-coml) |
|---------|------|---------|---------|--------|-------|-----------------|-----------------|-----------------|-----------------------|-----------------------|-----------------------|
| 31.04   | 9.72 | 0.85177 | 0.59945 |        |       |                 |                 |                 |                       |                       |                       |
| 9.72    | 7.05 | 0.82322 | 0.56110 |        |       |                 |                 |                 |                       |                       |                       |
| 7.05    | 5.81 | 0.80542 | 0.54292 | 100.00 | 10.76 | 0.767           | 0.394           | 0.830           | 65.3                  | 29.8                  | 46.4                  |
| 5.81    | 5.05 | 0.76213 | 0.52444 | 10.76  | 8.12  | 0.750           | 0.697           | 0.875           | 57.5                  | 33.4                  | 29.8                  |
| 5.05    | 4.53 | 0.72454 | 0.50317 | 8.12   | 6.79  | 0.754           | 0.748           | 0.894           | 55.8                  | 31.4                  | 28.2                  |
| 4.53    | 4.14 | 0.69863 | 0.50122 | 6.79   | 5.95  | 0.766           | 0.699           | 0.881           | 52.4                  | 26.7                  | 31.2                  |
| 4.14    | 3.84 | 0.67300 | 0.42343 | 5.95   | 5.36  | 0.741           | 0.746           | 0.893           | 52.3                  | 31.5                  | 26.0                  |
| 3.84    | 3.60 | 0.64592 | 0.38825 | 5.36   | 4.92  | 0.717           | 0.791           | 0.900           | 48.9                  | 31.6                  | 20.1                  |
| 3.60    | 3.39 | 0.59438 | 0.36984 | 4.92   | 4.57  | 0.686           | 0.838           | 0.912           | 53.4                  | 37.9                  | 16.8                  |
| 3.39    | 3.22 | 0.57999 | 0.33273 | 4.57   | 4.29  | 0.674           | 0.853           | 0.918           | 54.7                  | 41.6                  | 14.0                  |
| 3.22    | 3.07 | 0.55936 | 0.31744 | 4.29   | 4.05  | 0.660           | 0.854           | 0.906           | 50.7                  | 38.4                  | 13.7                  |
| 3.07    | 2.94 | 0.55729 | 0.36515 | 4.05   | 3.85  | 0.638           | 0.833           | 0.896           | 54.7                  | 41.0                  | 15.5                  |
| 2.94    | 2.83 | 0.53442 | 0.31326 | 3.85   | 3.68  | 0.625           | 0.833           | 0.890           | 53.3                  | 41.1                  | 13.5                  |
| 2.83    | 2.73 | 0.42253 | 0.28765 | 3.68   | 3.53  | 0.584           | 0.834           | 0.882           | 57.9                  | 46.5                  | 13.3                  |
| 2.73    | 2.63 | 0.23252 | 0.12487 | 3.53   | 3.39  | 0.569           | 0.825           | 0.871           | 56.5                  | 45.1                  | 13.1                  |
| 2.63    | 2.55 | 0.19144 | 0.09452 | 3.39   | 3.27  | 0.552           | 0.792           | 0.844           | 56.2                  | 43.5                  | 15.0                  |
| 2.55    | 2.48 | 0.18928 | 0.10692 | 3.27   | 3.16  | 0.548           | 0.785           | 0.838           | 58.8                  | 45.8                  | 15.1                  |
| 2.48    | 2.41 | 0.17824 | 0.10153 | 3.16   | 3.06  | 0.534           | 0.789           | 0.833           | 60.6                  | 49.2                  | 14.2                  |
| 2.41    | 2.34 | 0.16364 | 0.10999 | 3.06   | 2.97  | 0.550           | 0.770           | 0.821           | 58.3                  | 45.1                  | 15.4                  |
| 2.34    | 2.28 | 0.15432 | 0.10079 | 2.97   | 2.89  | 0.527           | 0.765           | 0.807           | 61.6                  | 50.1                  | 14.6                  |
| overall |      | 0.44464 | 0.33461 | 2.89   | 2.82  | 0.497           | 0.758           | 0.796           | 63.5                  | 52.7                  | 13.8                  |
|         |      |         |         | 2.82   | 2.75  | 0.421           | 0.768           | 0.789           | 66.6                  | 58.3                  | 11.1                  |
|         |      |         |         | 2.75   | 2.68  | 0.297           | 0.772           | 0.783           | 73.3                  | 70.1                  | 7.1                   |
|         |      |         |         | 2.68   | 2.62  | 0.194           | 0.769           | 0.770           | 76.4                  | 76.3                  | 4.5                   |
|         |      |         |         | 2.62   | 2.56  | 0.190           | 0.783           | 0.786           | 81.0                  | 80.6                  | 3.8                   |
|         |      |         |         | 2.56   | 2.51  | 0.181           | 0.757           | 0.757           | 79.7                  | 80.0                  | 4.5                   |
|         |      |         |         | 2.51   | 2.46  | 0.187           | 0.773           | 0.775           | 80.5                  | 81.1                  | 3.9                   |
|         |      |         |         | 2.46   | 2.41  | 0.172           | 0.735           | 0.735           | 81.7                  | 83.1                  | 5.2                   |
|         |      |         |         | 2.41   | 2.37  | 0.194           | 0.762           | 0.764           | 79.6                  | 80.6                  | 4.5                   |
|         |      |         |         | 2.37   | 2.33  | 0.151           | 0.732           | 0.735           | 82.9                  | 85.3                  | 4.8                   |
|         |      |         |         | 2.33   | 2.28  | 0.156           | 0.687           | 0.690           | 84.2                  | 87.1                  | 5.7                   |

Tableau 2.3.3 : Facteurs de qualité pour le phasage du dérivé de la protéine yeaz avec Gd-DOTMA en fonction de la résolution. **À gauche** : Les valeurs des facteurs de qualité des réflexions générales (FOMacen) et des réflexions à phase contrainte (FOMcen) du phasage par le programme SHARP. **À droite** : Tableau des facteurs de qualité après amélioration (et extension) des phases avec le programme DM.

La réussite du phasage est confirmée par la qualité de la carte expérimentale (figure 2.3.3) qui est aisément interprétable et qui a permis la construction automatique d'une grande partie du modèle avec le programme ARP/wARP, suivie de la construction manuelle des parties manquantes.

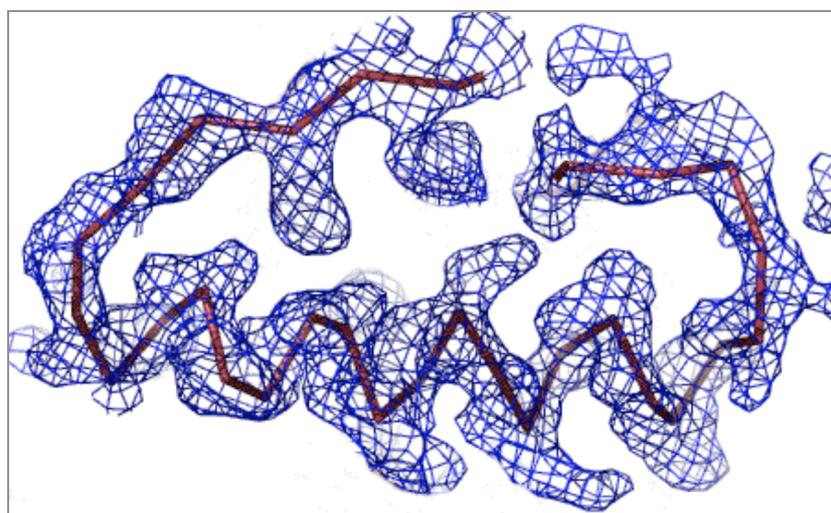


Figure 2.3.3 : Extrait de carte de densité expérimentale après nivellement de solvant pour le dérivé yeaz + Gd-DOTMA (contour  $1\sigma$ ). La carte est calculée avec des données à une résolution de 2,28 Å.

### II.3.8.Synthèse des résultats des essais de phasage

Je résume ici les résultats des essais de phasage avec 53 dérivés de 12 protéines différentes. Pour cinq d'entre-elles, la structure a été résolue récemment avec des cristaux dérivés obtenus avec un des complexes de gadolinium. Sauf pour l'OTCase3630, les projets sur les nouvelles protéines se font en collaboration avec d'autres équipes de cristallographie biologique (yeaZ : Sandra Jeudy et Chantal Abergel, IGS-Marseille; protéine X et Pce : Rafael Molina-Monterrubbio et Juan Hermoso CSIC, Madrid; PACE12 : Stéphanie Gras et Dominique Housset, LCCP, IBS, Grenoble).

Pour certains dérivés, le phasage auquel je me réfère n'est pas le même que celui qui est décrit dans les tableaux précédents. C'est le cas pour les phasages par la méthode SAD avec des données enregistrées à une longueur d'onde proche de 1 Å. Pour ces dérivés, un phasage par SAD avec des données enregistrées au laboratoire a été effectué auparavant et sert de référence pour le tableau 2.3.4. Les données enregistrées à courte longueur d'onde servent pour l'affinement du mode de fixation des complexes.

Le tableau indique le "succès" (+) ou "l'échec" (-) de l'obtention d'un dérivé avec fixation suffisante du diffuseur anomal pour une combinaison de protéine et de complexe donnée. Le "succès" correspond à des dérivés qui permettent d'obtenir de bonnes phases, car ils présentent des taux de fixation suffisamment élevés (typiquement ~50%) d'un nombre suffisamment élevé de sites de fixation (typiquement 1 site/150 résidus). La qualité des phases obtenues est jugée à l'aide des indicateurs statistiques et de la qualité de la carte de densité électronique. Seuls les phasages menant à des cartes de densité électroniques aisément interprétables sont qualifiés de "succès". Ainsi pour la protéine urate oxydase quatre des huit dérivés seulement sont comptés pour "succès" alors que selon Girard *et al.* (2003a), sept d'entre eux permettraient le phasage.

|             | UO      | GI  | thaumatine    | lysozyme | Prot X* | YGGV  | succès | succès |
|-------------|---------|-----|---------------|----------|---------|---|--------|--------|
| DO3A        | +       | +A  | -             | +A       | +A      | +A  | 5/6    | 83%    |
| HPDO3A      | -       | -   | -             | +A       | +A      | -   | 2/6    | 33%    |
| DTPA-BMA    | +A      | -   | +A            | -        | -       | -   | 2/6    | 33%    |
| DOTA        | -       | -   | +A            | +A       | +       | -   | 3/6    | 50%    |
| DOTMA       | +A      | -   | -             | -        | +A      | +A  | 3/6    | 50%    |
| DTPA        | -       | +A  | +A            | -        | +A      | -   | 3/6    | 50%    |
| DOTA-BOM    | -       |     | +             | +A       | +       | +   | 4/5    | 80%    |
| HPSA-DO3A   | +       |     |               | -        |         |   | 1/2    | 50%    |
| précipitant | PEG3350 | MPD | tartrate Na-K | NaCl     | **      | (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> |        |        |
| pH          | 8,5     | 7,0 | 6,6           | 4,5      | **      | 9,0   |        |        |
| pI          | 7,3     | 5,0 | 8,5           | 9,3      |         | 5,9   |        |        |
| % solvant   | 52      | 48  | 48            | 29       |         | 49  |        |        |
| M [kDa]     | 34      | 43  | 22            | 14       | 40      | 24  |        |        |

|             | yeaZ*   | Pce*   | PACE12* | lactamase | catalase  | OTCase*   | succès total | succès total |
|-------------|---------|--------|---------|-----------|---|---|--------------|--------------|
| DO3A        | +A      |        |         | +         |   |   | 7/8          | 88%          |
| HPDO3A      |         | +A     |         |           | +   | +   | 5/9          | 56%          |
| DTPA-BMA    | -       |        | +       | +         |   |   | 4/9          | 44%          |
| DOTA        |         |        |         |           |   |   | 3/6          | 50%          |
| DOTMA       | +A      |        |         | +         |   |   | 5/8          | 63%          |
| DTPA        |         |        |         |           |   |   | 3/6          | 50%          |
| DOTA-BOM    |         |        |         |           |   |   | 4/5          | 80%          |
| HPSA-DO3A   |         |        |         |           |   |   | 1/2          | 50%          |
| précipitant | PEG8k   | PEG10k | PEG4k   | PEG8k     | (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> | (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> |              |              |
| pH          | 4,7-5,5 | 5,5    | 5,6     | 6,5       | 6,5   | 7,8   |              |              |
| pI          | 5,7     | 5,2    | 5,1     | 4,9       | 6,1   | 6,5   |              |              |
| % solvant   | 48      | 49     |         |           | 63  |   |              |              |
| M [kDa]     | 28      | 63     | 30      | 24        | 56  |   |              |              |

Tableau 2.3.4 : Résumé des résultats de phasage obtenus avec 53 dérivés différents. En haut : Protéines testées systématiquement avec la majorité des complexes. En bas : Protéines pour lesquelles des dérivés n'ont été préparés qu'avec certains des complexes. + : Le dérivé a permis la réussite du phasage. - : Le complexe ne présente pas de fixation significative. A : La structure du complexe lié à la protéine est affinée. \* Protéines dont la structure a été résolue pour la première fois en utilisant des dérivés obtenus avec un des complexes de gadolinium. \*\* information confidentielle. Le pourcentage de solvant du cristal (% solvant) est calculé par le programme TRUNCATE (CCP4). Le précipitant et pH se réfèrent aux conditions de cristallisation. M : masse du monomère.

Abréviations utilisées dans le tableau :

Tous les complexes sont abrégés par le nom de leur ligand.

Protéines testées : **UO** : urate oxydase d'*Aspergillus flavus* (Vivarès & Bonneté, 2002; Bonneté *et al.*, 2001), **GI** : glucose isomérase de *Streptomyces rubiginosus* (Ramagopal *et al.*, 2003), **thaumatine** : thaumatine de *Thaumatococcus daniellii* (Ko *et al.*, 1994), **lysozyme** : lysozyme (de blanc d'œuf de poule (Girard, 2001), **prot X** : protéine X, structure non-publiée, **yggv** : protéine hypothétique Yggv d'*Escherichia coli*. Les cristaux d'YGGV nous ont été fournis par Chantal Abergel, IGS, Marseille, **yeaz** : yeaZ d'*Escherichia coli* (Jeudy *et al.*, 2005), **Pce** : phosphorylcholine estérase Pce de *Streptococcus pneumoniae* (Hermoso *et al.*, 2005), **PACE12** : PAB0955 de *Pyrococcus abyssi* (Gras *et al.*, 2005), **lactamase** :  $\beta$ -lactamase à zinc de classe B de *Pseudomonas aeruginosa* (Girard, 2001), **catalase** : catalase de *Proteus mirabilis* (Girard, 2001), **OTCase** : OTCase3630 chimérique (Girard, 2001).

Le taux de succès  $\frac{N_{succès}}{N_{test}}$  pour les différents complexes résultant du tableau 2.3.3 est

représenté dans l'histogramme de la figure 2.3.4. Pour déterminer les chances de succès d'obtention d'un bon dérivé anomal en fonction du complexe utilisé, il convient de se référer aux colonnes bleues (série 1), correspondant uniquement aux protéines testées systématiquement, comprenant donc et échecs et réussites. Les colonnes violettes comprennent, en plus, les protéines qui n'ont pas été testées systématiquement, d'éventuels échecs ne sont pas mentionnés. En ne considérant, en conséquence, que la série 1 pour

calculer le taux de succès total, on obtient :  $\frac{N_{succès}}{N_{test}} = \frac{23}{43} = 53\%$ .

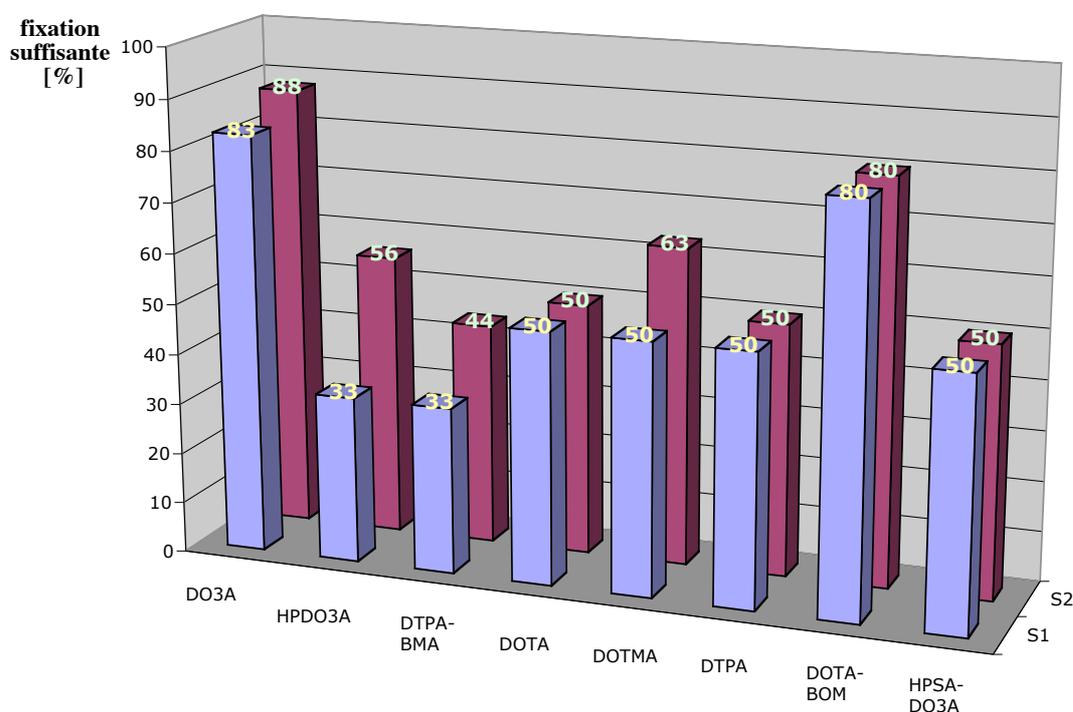


Figure 2.3.4 : Histogramme représentant le taux de succès pour l'obtention de dérivés avec une fixation suffisante du complexe comme fonction du complexe utilisé. Colonnes bleues : S1 : Résultats obtenus avec les protéines testées systématiquement. Colonnes violettes : S2 : Résultats comprenant tous les dérivés testés.

### II.3.9. Conclusions

Le fort potentiel des complexes pour insérer et fixer les ions de gadolinium dans les cristaux de protéines a été indiqué auparavant (Girard *et al.*, 2002; Girard *et al.*, 2003a; Girard *et al.*, 2003b). Cette étude, plus étendue et sur un nombre élevé de protéines différentes, allant de l'obtention des dérivés jusqu'au phasage, a permis de confirmer et de quantifier ce potentiel très élevé, qui correspond à un taux de succès supérieur à 50% pour obtenir un bon dérivé. Elle permet de tirer des conclusions statistiques en fonction du complexe utilisé.

Le tableau 2.3.3 indique les conditions physico-chimiques dans lesquelles les dérivés ont été obtenus. Cependant, aucune relation directe n'apparaît entre ces conditions physico-chimiques (pH, précipitant) ou le pI de la protéine et une éventuelle fixation privilégiée d'un complexe particulier.

Afin de combiner les résultats obtenus ici, avec une compréhension plus approfondie de la fixation des différents complexes, nous avons essayé de préciser le mode de fixation des complexes dans les cristaux de protéine pour un nombre maximum de dérivés et pour lesquels nous avons pu constater une fixation significative, en affinant leur structure.

## II.4. Caractérisation du mode de fixation des complexes dans les protéines

Afin de connaître le mode de fixation des différents complexes avec différentes protéines, nous avons essayé d'affiner la structure du complexe lié à la protéine pour les dérivés avec des sites de fixation suffisamment occupés. Le problème auquel je me suis heurté lors de l'affinement de la structure du ligand entourant l'ion, est que les taux de fixation typiques pour un "bon" dérivé sont de l'ordre de 30-60%, ce qui est suffisant pour fournir un signal anormal permettant un bon phasage, mais ne permet pas obligatoirement d'affiner la structure du ligand. En effet, pour des sites occupés à moins de 50%, la densité électronique correspondant au ligand est très faible et ne permet généralement pas d'identifier l'orientation du complexe et ses éventuelles interactions avec la protéine, ni d'éventuels changements de conformations de résidus protéiques, ni des molécules d'eau liées au ligand. La faible visibilité de la densité électronique du ligand m'a incitée à préparer des dérivés avec une concentration de complexe de 300 mM pour augmenter le taux de fixation des sites anomaux. J'ai préparé ces dérivés pour des combinaisons de protéine et de complexe dont le taux de fixation est déjà élevé pour une concentration de complexe de 100 mM. Ainsi, j'ai pu obtenir des dérivés avec des taux de fixation supérieurs à 70% pour lesquels la densité électronique correspondant au ligand est bien définie.

### II.4.1. Affinement initial

L'affinement de structure des dérivés a été effectué avec le programme CNS (Brünger *et al.*, 1998). J'ai utilisé les jeux de données correspondant au dérivé avec, *a priori*, les taux d'occupation les plus élevés possible (dérivés obtenus avec 300 mM de complexe) et avec la résolution la plus élevée possible.

Pour des dérivés pour lesquels nous disposons de jeux de données enregistrés à plusieurs longueurs d'onde, j'ai effectué le phasage MAD avec SHARP et utilisé les données phasées pour l'affinement. J'ai alors utilisé le mode mlhl de CNS en prenant FPsha pour amplitude et l'information de phase expérimentale sous forme des coefficients d'Hendrickson et Lattman (Hendrickson & Lattman, 1970). Pour des données à longueur d'onde unique, nous avons utilisé les données non phasées, tout en incluant l'information anormale dans l'affinement en utilisant les amplitudes  $F^+$  et  $F^-$  en mode mlf. Que le signal anormal des données à longueur d'onde unique (seuil  $K$  du Se ou du Br) soit relativement faible ne pose pas de problème pour l'affinement, le modèle à affiner étant déjà essentiellement complet. Les codes PDB des modèles atomiques initiaux de protéine utilisés pour l'affinement sont donnés dans le tableau 2.5.0. Les coordonnées des sites de fixation des ions de gadolinium obtenues auparavant sont introduites dans le fichier pdb de la structure. Afin de suivre l'évolution du résiduel libre  $R_{\text{free}}$  (Brünger, 1992), une fraction représentant 5% des réflexions est exclue de l'affinement. Un affinement par corps rigide permet d'orienter le modèle dans la maille et d'affiner la position des ions de gadolinium. On effectue un recuit simulé dans le cas d'un éventuel non-isomorphisme entre la structure de la protéine modèle et celle de notre dérivé. Cette étape est suivie par l'affinement du facteur d'agitation thermique des atomes de la protéine et des ions de gadolinium, puis par l'affinement du taux d'occupation des ions. Le taux d'occupation initial des ions de gadolinium correspond au taux d'occupation affiné par SHARP lors du phasage. Dans le cas où on ne dispose pas d'une estimation de départ fiable du taux

d'occupation (Qcns), car l'affinement de Qcns et Bcns est corrélé. Ensuite, des molécules d'eau sont introduites de manière automatique dans le modèle. L'inspection du modèle et des cartes de densité électronique expérimentales, des cartes 2Fo-Fc et des cartes 1Fo-Fc, à l'aide du programme O (Jones *et al.*, 1991), permet de placer d'éventuelles petites molécules liées à la protéine dans le cristal dérivé. De telles petites molécules sont, par exemple, des molécules d'agent précipitant comme le MPD ou des ions tartrate pour les protéines glucose isomérase et thaumatine, l'inhibiteur 8-azaxanthine de l'urate oxydase ou des ions Ni<sup>2+</sup> pour la protéine X.

| Protéine | Lysozyme    | Urate oxydase | Thaumatine  | Glucose isomérase | YeaZ        | YGGV        | Protéine X |
|----------|-------------|---------------|-------------|-------------------|-------------|-------------|------------|
| Code PDB | 193L<br>(1) | 1WS3<br>(2)   | 1THW<br>(3) | 1XIB<br>(4)       | 1OKJ<br>(5) | 1K7K<br>(6) | *          |

Tableau 2.4.0 : Codes PDB des modèles de protéine initiaux utilisés pour les affinements. \* La structure de la protéine n'étant pas encore publiée, son modèle nous a été fourni par R. Molina et J. Hermoso (CSIC, Madrid, Espagne). 1 : Vaney *et al.*, 1996, 2 : Retailleau *et al.*, 2005, 3 : Ko *et al.*, 1994, 4 : Carrell *et al.*, 1994, 7 : Jeudy *et al.*, 2005, 6 : Sanishvili *et al.*

## II.4.2. Les modèles des complexes utilisés

Pour la majeure partie des dérivés, il aurait été impossible de construire *de novo* les modèles atomiques des complexes dans la carte de densité, car la densité électronique du ligand, fixé sur des sites occupés à moins de 100%, n'est pas suffisamment bien définie. Pour les complexes dont la structure cristallographique a été résolue auparavant, nous avons utilisé ce modèle pour l'affinement. Tel est le cas pour les complexes Gd-DTPA-BMA (Ehnebom & Pedersen, 1992), Gd-HPDO3A (Kumar *et al.*, 1994), Gd-DOTA (Dubost *et al.*, 1991) et Gd-DO3A (Chang *et al.*, 1993). Pour modéliser le complexe Gd-DOTMA, j'ai obtenu son modèle à partir de la structure du complexe Gd-DOTA, en ajoutant des groupements méthyle aux bras du ligand.

Le complexe Gd-DOTA-BOM a été modélisé par le modèle du complexe Gd-DOTMA, car le groupement (phénylméthoxy)méthyle qui distingue ce premier de ce dernier n'est visible dans aucune des cartes de densité électronique obtenues. Le dérivé de protéine X avec le Gd-DTPA a permis de déduire la structure du complexe Gd-DTPA de la structure connue de Gd-DTPA-BMA. Le calcul *de novo* de la structure des complexes à partir de la formule chimique connue n'est pas possible, car les programmes habituels utilisés pour calculer la structure 3D de petites molécules (tel le serveur ProdrG (Schuettelkopf & van Aalten, 2004)) ne gèrent pas les interactions chélatantes dans un complexe de métal.

## II.4.3. Affinement avec complexes

Après inspection de la carte et du modèle obtenus après l'affinement initial, les molécules d'eau situées dans un rayon de 5 Å autour des sites des ions de gadolinium sont enlevées. On calcule la carte 1Fo-Fc correspondant à ce modèle. Le modèle du ligand est alors placé de manière à obtenir le meilleur accord possible entre le modèle et la densité électronique (on utilise la carte 1Fo-Fc et la carte de densité électronique expérimentale, cette dernière étant non-biaisée par le modèle initial). Lorsque la densité correspondant au ligand est faible, il peut être difficile de trouver la bonne orientation du modèle. La première étape consiste alors à identifier la position du macrocycle. Ensuite, on essaie d'identifier des parties de densité qui correspondent aux groupements carboxyle ou méthylamide des extrémités des bras du ligand.

Le complexe existe sous deux formes énantiomères qui se déduisent l'une de l'autre par inversion des coordonnées. Pour placer le bon modèle de complexe il faut déterminer laquelle des deux structures énantiomères se lie au site de fixation (les deux énantiomères peuvent être liés sur différents sites de fixation d'un même dérivé). Le modèle du ligand peut être ajusté dans la densité par translation et rotation de la molécule autour de l'ion  $Gd^{3+}$ , tout en gardant une distance correcte entre les atomes d'azote et d'oxygène chélatants et l'ion  $Gd^{3+}$ . Des parties du ligand peuvent être mieux placées dans la densité en leur appliquant des rotations chimiquement permises. À partir de ce dernier modèle, on calcule des fichiers de topologie et de paramètres énergétiques pour l'affinement automatique du ligand par CNS. Pour ce calcul, nous nous sommes servi du serveur HIC-Up (Hetero-compound Information Center - Uppsala (Kleywegt & Jones, 1998)) qui engendre les fichiers utilisés par CNS à partir du fichier de structure en format pdb. La signification physique des paramètres énergétiques ainsi produits est faible, car le serveur ne gère pas l'interaction des atomes chélatants avec l'ion lourd. Les résultats d'affinement automatique doivent donc être vérifiés avec précaution. Un affinement par minimisation d'énergie ("minimize") est suivi de la détermination du facteur d'agitation thermique du ligand et la détermination du taux d'occupation de l'ensemble ion + ligand. L'inspection des cartes 2Fo-Fc et 1Fo-Fc (niveaux négatif et positif) permet de vérifier l'orientation du modèle, d'ajouter à la main d'éventuelles molécules d'eau supplémentaires au voisinage du site de fixation, de déplacer des résidus protéiques et de définir d'éventuelles double conformations.

Dans le dérivé de l'YGGV avec le Gd-DOTMA le site de fixation principal se trouve sur un axe de symétrie cristallographique d'ordre 2. La molécule de complexe n'obéissant pas à cette symétrie, son orientation dans le site est statistiquement distribuée entre deux orientations possibles. Pour l'affinement automatique du modèle du complexe nous avons dû annuler les interactions de Van der Waals entre deux molécules de complexe.

Par la suite, je décris les résultats de l'affinement de la structure du complexe lié à la protéine pour les différents dérivés étudiés. Les différents dérivés sont regroupés selon la protéine utilisée.

### Notes explicatives pour la suite :

La charge du complexe **isolé** est indiquée entre parenthèses.

Il est rappelé que l'ion  $Gd^{3+}$  nécessite neuf atomes de coordination dans sa première sphère de coordination. De ces neuf atomes, huit sont fournis par le ligand entourant l'ion, sauf pour le ligand DO3A, qui ne fournit que sept atomes de coordination à l'ion.

### Grandeurs calculées par CNS :

**R** et **Rfree** : facteurs résiduel et résiduel libre (voir abréviations)

**Bcns** : facteur d'agitation thermique des ions de gadolinium

**Qcns** : taux d'occupation du site de fixation

### Figures :

Toutes les figures sont préparées avec le programme PYMOL (DeLano, 2002). Ce programme permet de générer des surfaces colorées en fonction de la charge électrostatique. Toutes les surfaces électrostatiques sont à prendre avec précaution, la charge de la surface étant calculée sans prendre en compte d'effet de solvant. Sauf mention contraire, les contours des cartes de densité électronique correspondent au niveau  $1\sigma$ .

Couleurs : sauf mention contraire, les atomes de carbone des modèles de ligands sont représentés en vert, ceux des résidus protéiques en or ou/et en gris (utilisation de deux

couleurs différentes si la fixation implique plusieurs molécules de protéine symétriques). L'ion  $Gd^{3+}$  est représenté en magenta.

#### II.4.4. Urate oxydase

|                    | Nb complexes affinés | Qcns  | Bcns [ $\text{\AA}^2$ ]                   | R    | Rfree | Résolution [ $\text{\AA}$ ] |
|--------------------|----------------------|---|---|------|-------|-----------------------------|
| Gd-DOTMA 300 mM    | 1                    | 1,01;0,29;0,21  | 19;23;21                                  | 0,20 | 0,21  | 1,35                        |
| Gd-DTPA-BMA 300 mM | 3                    | 0,92;0,48;0,70  | 13;8;9                                    | 0,19 | 0,20  | 1,45                        |
| Gd-DO3A 100 mM     | 0                    | 0,19; 0,14; 0,08;<br>0,26; 0,21; 0,14;<br>0,07; 0,11; 0,17;<br>0,05 | 55;54;59;<br>50;45;<br>40;34;36;<br>42;29 | 0,22 | 0,23  | 1,45                        |

|          | Résidus voisins aux sites de fixation   |
|----------|---|
| Gd-DOTMA | site 1 : Trp174, Lys189<br>site 2 : Tyr30 (5 $\text{\AA}$ ), His19 (7 $\text{\AA}$ )  |
| Gd-DO3A  | site 1 : Glu41; site 2 : Glu213; site 3 : Glu136; site 4 : Glu249; site 5 : Lys289; site 6 : Glu139; site 7 : Asp141; site 8 : Glu196 |

Tableau 2.4.1 : Statistiques d'affinement des dérivés d'urate oxydase.

Avec la protéine urate oxydase, trois des complexes ont mené à des dérivés avec une très bonne fixation et permettant un phasage aisé : Gd-DTPA-BMA, Gd-DOTMA et Gd-DO3A. Nous avons procédé à l'affinement de la structure des dérivés correspondant. Pour les deux premiers, les occupations des sites anomaux étaient suffisamment élevées (voir tableau 2.4.1) pour permettre de placer les modèles des ligands dans la densité électronique et d'affiner le mode de fixation. Le dérivé obtenu avec le complexe Gd-DO3A présente de nombreux sites de fixation avec des occupations de l'ordre de 20%. La faible occupation des sites fait que la densité électronique des ligands des complexes est trop peu visible pour placer et affiner la structure des ligands. Le dimère de la protéine forme un tunnel long de 50  $\text{\AA}$  et de 12  $\text{\AA}$  de diamètre où se situent deux des sites de fixation du complexe Gd-DTPA-BMA (voir figure 2.4.1 pour l'emplacement des différents sites de fixation).

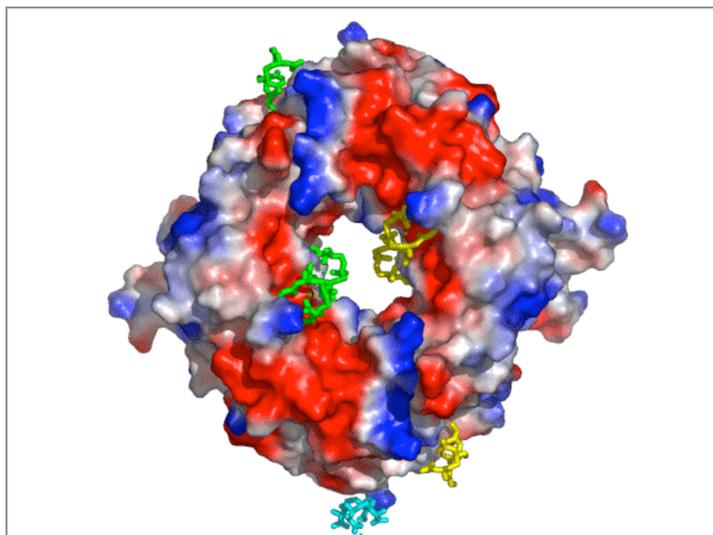


Figure 2.4.1 : Dimère d'urate oxydase formant un tunnel et l'emplacement des sites de fixation des molécules de complexe : vert Gd-DTPA-BMA (molécules symétriques par symétrie cristallographique : jaune), en cyan : Gd-DOTMA.

### Description du mode de fixation des différents complexes

#### Gd-DOTMA (charge : 1-)

Dans le dérivé obtenu à 300 mM de Gd-DOTMA, le complexe se fixe sur trois sites, dont un avec une occupation très élevée, et les deux autres avec des occupations affinées inférieures à 30% (tableau 2.4.1). La bonne qualité de la densité électronique autour de l'ion de Gd du premier site a permis de placer le ligand et d'affiner la structure du complexe lié à la protéine.

#### Site principal (taux d'occupation : 1,0)

Comme le montre la figure 2.4.2, le complexe, chargé négativement, se lie via le côté de son macrocycle à une partie peu chargée de la surface, le site de fixation étant une légère cavité formée à l'interface avec une molécule de protéine symétrique. Le côté  $Gd^{3+}$  pointe vers le solvant.

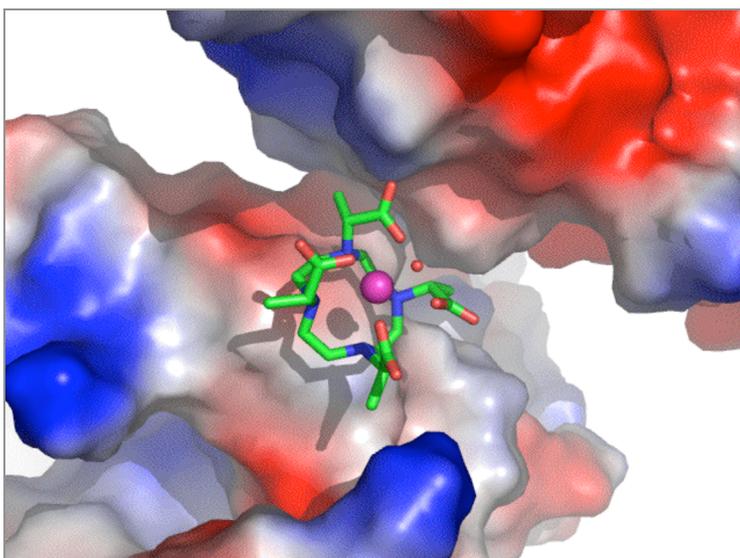


Figure 2.4.2 : Site de fixation principal du complexe Gd-DOTMA formé par deux molécules de la protéine urate oxydase. La surface voisine au complexe est peu chargée.

Comme le montre la figure 2.4.3.a, la liaison du complexe fait intervenir une interaction hydrophobe entre le macrocycle et le cycle aromatique du Trp174 (distance  $\sim 3,3\text{\AA}$ ), les plans des deux cycles étant parallèles, un pont salin entre un atome d'oxygène carboxylique du ligand et l'atome d'azote NZ de la Lys189 ( $2,9\text{\AA}$ ) et la coordination via les atomes d'oxygène du ligand de trois molécules d'eau. Ces molécules d'eau sont liées à d'autres molécules d'eau ordonnées entre le complexe et la surface protéique, qui ne sont pas montrées ici. Le neuvième ligand de la première sphère de coordination de l'ion  $\text{Gd}^{3+}$  est présent sous forme d'une molécule d'eau liée à l'ion ( $\text{O-Gd} : 2,9\text{\AA}$ ) La figure 2.4.3.b montre la densité  $2\text{fo-fc}$  de la structure affinée qui est très bien définie.

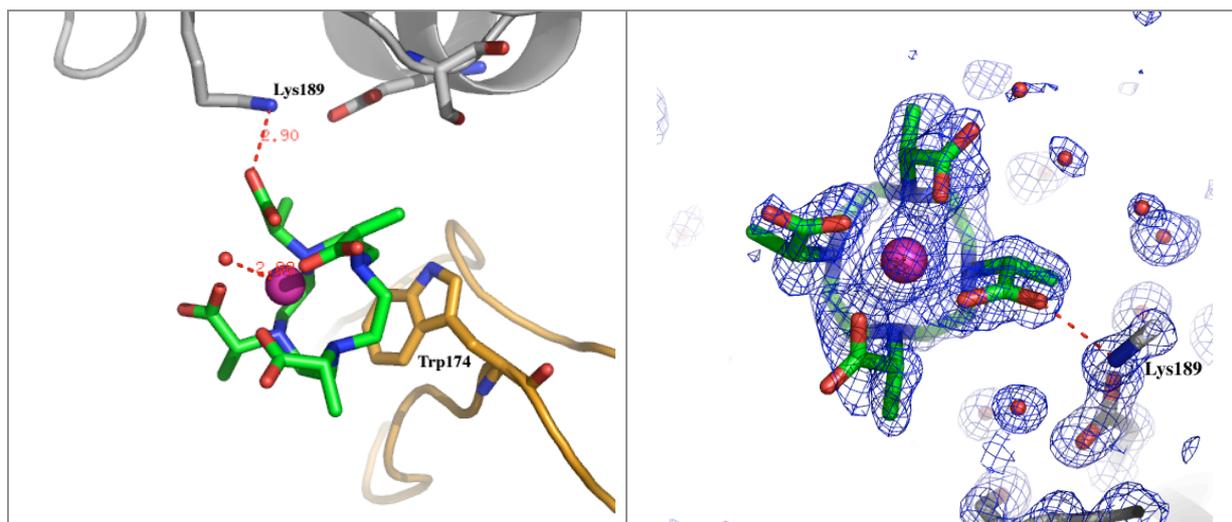


Figure 2.4.3 : Site de fixation principal du complexe Gd-DOTMA formé par deux molécules de la protéine urate oxydase. En gris, les résidus protéiques de la molécule de protéine symétrique. a) Les traits pointillés indiquent un pont salin et la coordination d'une molécule d'eau par l'ion de Gd. b) Densité  $2\text{Fo-Fc}$  du modèle de complexe affiné.

### **Gd-DTPA-BMA (charge : 0)**

Le complexe Gd-DTPA-BMA se lie assez fortement sur trois sites. Les occupations affinées avec CNS sont de 0,92, 0,48 et 0,70 respectivement pour les sites 1, 2 et 3 du dérivé obtenu avec 300 mM de complexe. Grâce à une densité électronique bien définie, l'affinement de la structure du complexe et l'interprétation du mode de fixation sont possibles pour les trois sites.

#### **Site 1 et 2**

Les sites 1 et 2 sont proches l'un de l'autre (distance  $\text{Gd1-Gd2} : 8,4\text{\AA}$ ) dans le tunnel formé par le dimère de la protéine (figure 2.4.1). Comme le montre la figure 2.5.4.a, les molécules de complexe se lient à la surface de la protéine via le côté de l'ion. La surface, au site 1, est chargée négativement, sauf aux endroits où pointent les atomes d'oxygène carboxyliques extérieurs du ligand.

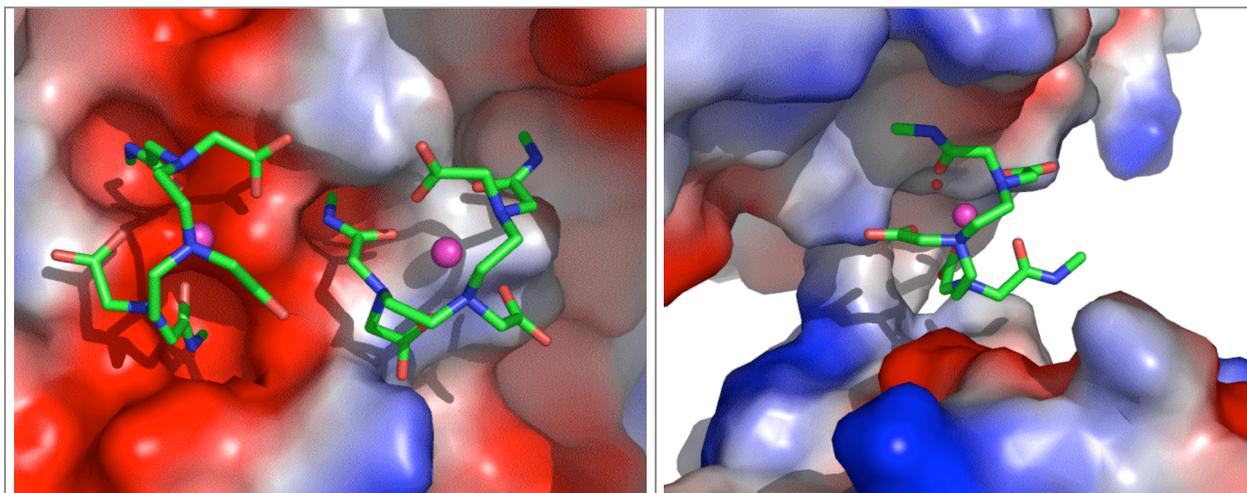


Figure 2.4.4 : De gauche à droite : a) les sites de fixation 1 et 2 et b) le site3 du dérivé urate oxydase + Gd-DTPA-BMA.

### Site 1 (taux d'occupation : 0,92)

La figure 2.4.5.a montre les interactions du complexe lié au site 1 avec les résidus protéiques de la surface. La conformation des acides aminés dans la protéine native est montrée en bleu. L'ion est coordonné par un atome d'oxygène de la chaîne latérale de l'Asp133 (OD-Gd<sup>3+</sup> : 2,5 Å), cet atome agissant comme 9<sup>e</sup> ligand de l'ion. Les atomes du ligand interagissent avec la protéine : un atome d'azote du bras du ligand forme une liaison H avec l'atome d'oxygène carboxylique du Glu39 (N-OE : 2,7 Å) et un atome d'oxygène du ligand forme un pont salin avec l'atome d'azote NZ de la Lys143 (O-NZ : 2,7 Å). Le complexe est entouré d'un réseau de molécules d'eau ordonnées, dont six molécules d'eau liées aux atomes du ligand, à des distances inférieures à 3 Å, non montré ici. Pour le rôle des molécules d'eau ordonnées pour la fixation des complexes voir plus loin paragraphe II.4.5.a.

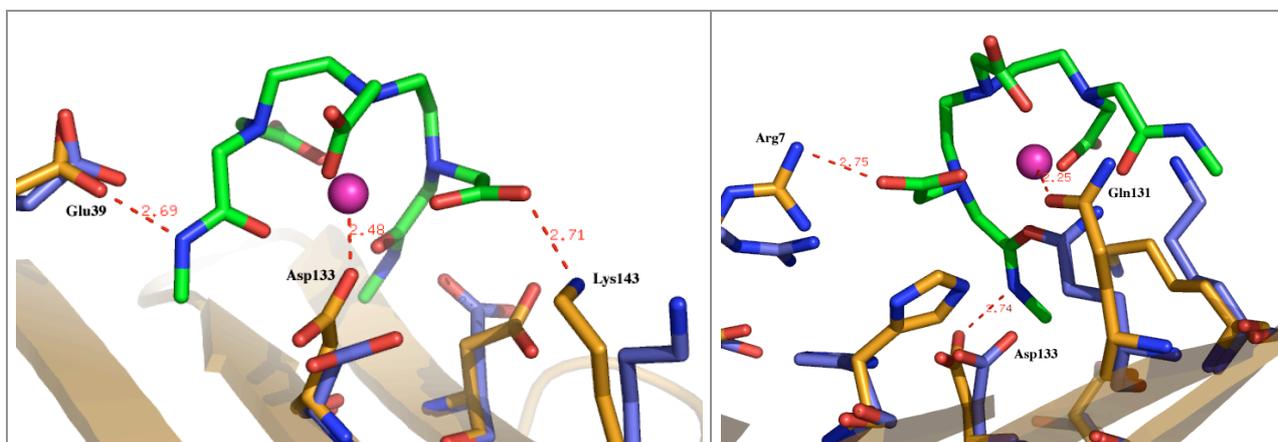


Figure 2.4.5 : Les molécules de complexe Gd-DTPA-BMA liées à la protéine urate oxydase a) sur le site de fixation 1, b) sur le site 2. En or : orientation des résidus protéiques dans le dérivé, en bleu : orientation des résidus protéiques dans la protéine native.

### Site 2 (taux d'occupation : 0,48)

La figure 2.4.5.b montre le mode de fixation du complexe Gd-DTPA-BMA et la conformation des résidus dans le cristal dérivé et natif au site 2. Le site se trouve à proximité du site 1 mais son taux d'occupation est sensiblement plus bas. L'ion Gd<sup>3+</sup> est coordonné par l'atome d'oxygène OE1 de Gln131 (Gd-OE1 : 2,3 Å). Une liaison H est formée entre l'atome d'azote d'un bras du ligand et un oxygène OD de l'Asp133 (N-OD : 2,7 Å). C'est ce même atome d'oxygène qui coordonne l'ion Gd<sup>3+</sup> du site 1. Un pont salin existe entre un oxygène du ligand

et l'azote NZ de la Lys7 (O-NZ : 2,8 Å). Le complexe est entouré d'un réseau de molécules d'eau ordonnées, non montré ici, dont quatre molécules d'eau liées directement aux atomes du ligand, à des distances < 3 Å.

### Site 1 et 2

La figure 2.4.6.a montre le modèle des complexes liés aux sites 1 et 2 et leurs interactions avec les résidus protéiques, dont celles avec l'Asp133 qui interagit avec les deux molécules de complexe. La figure 2.4.6.b montre la carte 2Fo-Fc, très bien définie, de la structure affinée du complexe lié dans le site 1 et des résidus protéiques impliqués dans les interactions. La densité correspondant à la chaîne du ligand du site 2 présente quelques lacunes.

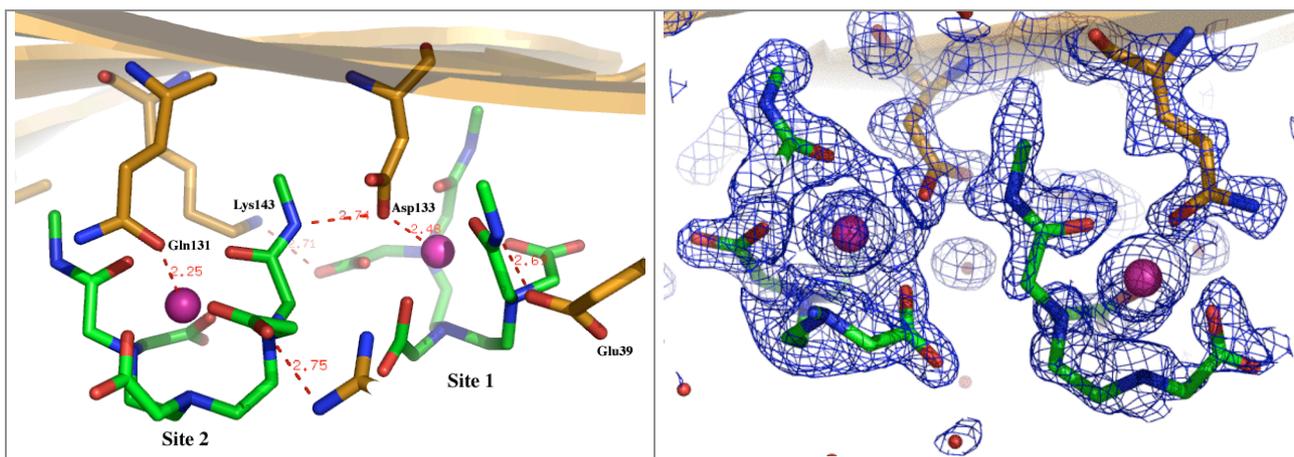


Figure 2.4.6 : a) Modèles des molécules de Gd-DTPA-BMA liés aux sites 1 et 2 et leurs interactions avec les résidus protéiques. b) La carte 2Fo-Fc des modèles affinés de la protéine et des molécules de complexe.

### Site 3 (taux d'occupation : 0,70)

Le site 3 se trouve dans une cavité formée par deux molécules de protéine symétriques (voir figure 2.4.4.b). Le complexe tourne son côté ionique vers la surface protéique, mais, contrairement aux sites 1 et 2, l'ion n'est pas coordonné par un résidu protéique. Comme le montre la figure 2.4.7.a, l'ion  $Gd^{3+}$  coordonne une molécule d'eau comme 9<sup>e</sup> ligand ( $Gd^{3+}$ -O : 2,3 Å). Cette même molécule d'eau forme des liaisons H avec l'atome d'azote de la chaîne principale de l'Asn223 (O-N : 2,9 Å) et l'atome d'oxygène OD de l'Asp222 (O-O : 2,6 Å). Un pont salin entre un atome d'oxygène du ligand et l'atome d'azote NZ de la Lys233 (O-NZ : 2,9 Å) renforce la fixation du complexe. Comme le montre la figure 2.4.7.b, la densité 2Fo-Fc de la structure affinée du complexe lié à la protéine est très bien définie, ainsi que la densité de la molécule d'eau impliquée dans la liaison. Le complexe est entouré par un réseau de molécules d'eau ordonnées, dont 8 molécules d'eau directement liées au ligand.

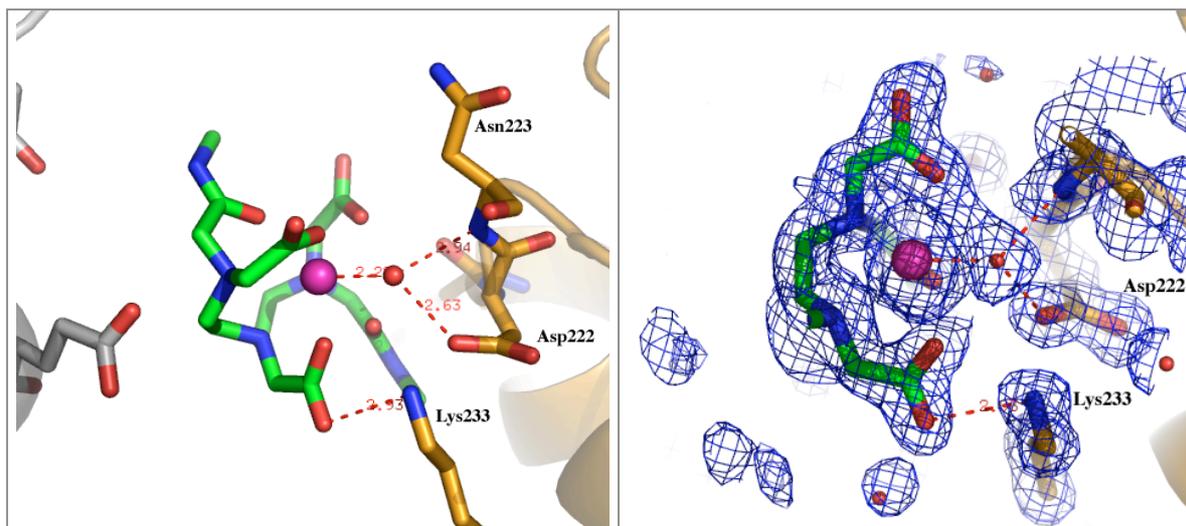


Figure 2.4.7 : a) Le complexe Gd-DTPA-BMA lié sur le site 3 de l'urate oxydase et ses interactions avec la protéine (en gris : résidus de la molécule de protéine symétrique), b) La carte 2Fo-Fc, très bien définie, du complexe affiné.

### Gd-DO3A (charge : 0) :

Le dérivé obtenu avec une concentration de complexe de 100 mM présente 10 sites de fixation dont les taux d'occupation varient de 5 à 26%. Après l'affinement du modèle incluant la protéine et les molécules d'eau et l'affinement des coordonnées et des occupations des ions  $Gd^{3+}$ , la densité 1Fo-Fc correspondant aux ligands des complexes est trop faible pour placer les molécules de complexe et les affiner. L'analyse des résidus protéiques situés à proximité des sites de fixation des ions  $Gd^{3+}$  montre qu'il s'agit majoritairement de glutamates ou d'aspartates. Ceci est en accord avec le mode de fixation principal que j'ai observé pour ce complexe avec d'autres protéines (glucose isomérase, yeaZ, YGGV). Dans ce complexe, il manque deux atomes de coordination à l'ion  $Gd^{3+}$ . Avec plusieurs dérivés de ce complexe, j'ai observé que le complexe se lie de préférence via son côté ionique aux glutamates et aspartates de la surface protéique, l'ion  $Gd^{3+}$  coordonné par les deux atomes d'oxygène carboxyliques du résidu. Les deux atomes d'oxygène saturent ainsi la première sphère de coordination du  $Gd^{3+}$ .

## II.4.4.a. Le rôle des molécules d'eau ordonnées pour la liaison des complexes

### Exemple : Urate oxydase + Gd-DTPA-BMA 300 mM, site 3

Le plus souvent, dans le cas où la sphère de coordination de l'ion  $Gd^{3+}$  n'est pas saturée par des atomes de résidus protéiques, une molécule d'eau liée à l'ion sert de 9<sup>e</sup> ligand. Ici, la molécule d'eau qui agit comme 9<sup>e</sup> ligand pour l'ion  $Gd^{3+}$  forme deux liaisons H avec un atome d'oxygène et un atome d'azote de la surface protéique. Outre les molécules d'eau liées à l'ion  $Gd^{3+}$ , des molécules d'eau sont coordonnées par les atomes d'oxygène carboxyliques extérieurs des ligands (et les atomes d'azote des groupements carbamide de Gd-DTPA-BMA). Le complexe est ainsi intégré dans le réseau de molécules d'eau ordonnées entourant la surface protéique. Parfois, les molécules d'eau coordonnées par le complexe pontent le complexe à des résidus protéiques. Dans l'exemple montré sur la figure 2.4.8.a, huit molécules d'eau interagissent par liaison H avec les atomes d'oxygène et d'azote des bras du ligand, dont deux qui pontent le ligand à des atomes d'oxygène de la surface protéique. Comme le montre

la figure 2.4.8.b, la carte 2Fo-Fc correspondant aux molécules d'eau est généralement bien définie.

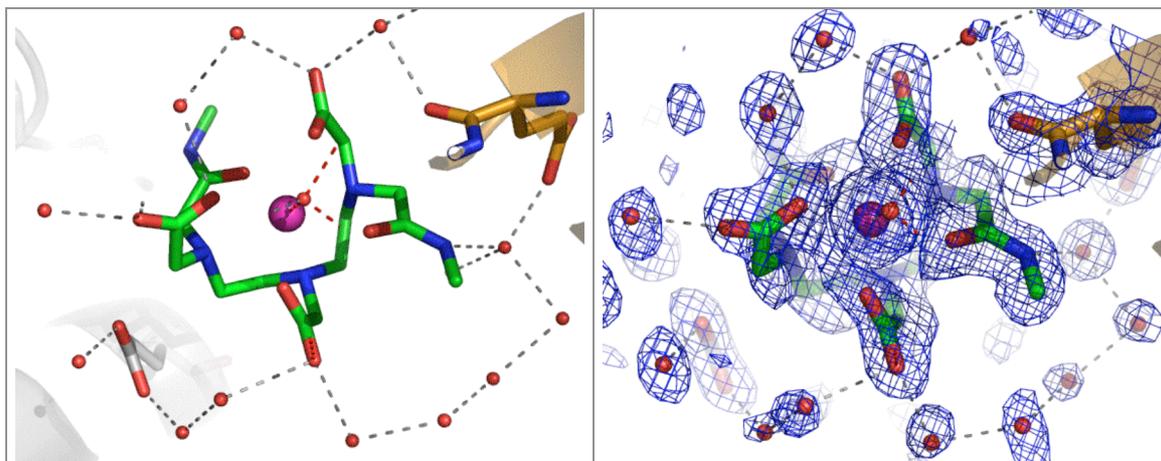


Figure 2.4.8. : a) Le complexe Gd-DTPA-BMA du site 3 du dérivé de l'urate oxydase. Les traits pointillés gris indiquent les liaisons H entre des atomes du ligand et des molécules d'eau ou des liaisons H entre deux molécules d'eau. Toutes les liaisons indiquées correspondent à des distances interatomiques  $< 3 \text{ \AA}$ . b) La densité 2Fo-Fc du complexe et des molécules d'eau coordonnées est très bien définie.

Pour d'autres exemples de description du rôle des molécules d'eau pour la fixation des complexes voir les affinements du dérivé du lysozyme avec le Gd-HPDO3A (chapitre II.4.9.) et de la protéine X avec le Gd-DTPA (chapitre II.4.11.).

#### II.4.4.b. L'apport d'un taux d'occupation élevé des sites de fixation pour identifier le mode de fixation. Comparaison de dérivés obtenus à des concentrations de complexe de 100 mM et de 300 mM.

Le tableau 2.4.2.a-c permet de comparer les valeurs de taux d'occupation des différents sites de fixation des dérivés préparés avec une concentration de complexe de 100 mM et de 300 mM. On remarque les taux d'occupation sensiblement plus élevés pour les dérivés préparés avec 300 mM de complexe.

| Lysozyme + | Site | Qcns   | Bcns [ $\text{\AA}^2$ ] | Qcns   | Bcns [ $\text{\AA}^2$ ] |
|------------|------|--------|-------------------------|--------|-------------------------|
|            |      | 100 mM | 100 mM                  | 300 mM | 300 mM                  |
| Gd-HPDO3A  | 1    | 0,81   | 15                      | 1,00   | 13                      |
|            | 2    | 0,73   | 16                      | 0,87   | 11                      |
| Gd-DOTA    | 1    | 0,69   | 29                      | 0,82   | 17                      |
|            | 2    | 0,50   | 44                      | 0,53   | 16                      |
| Gd-DO3A    | 1    | 0,57   | 23                      | 0,96   | 38                      |
|            | 2    | 0,36   | 22                      | 0,69   | 21                      |
|            | 3    | 0,21   | 18                      | 0,31   | 17                      |

Tableau 2.4.2.a : Valeurs des taux d'occupation et des facteurs d'agitation thermique affinés par SHARP des atomes de gadolinium pour les différents sites de dérivés préparés avec 100 et 300 mM de complexe pour les dérivés de lysozyme

| Urate oxydase + | Site | Qsha<br>100 mM | Bsha [ $\text{\AA}^2$ ]<br>100 mM | Qsha<br>300 mM | Bsha [ $\text{\AA}^2$ ]<br>300 mM |
|-----------------|------|----------------|-----------------------------------|----------------|-----------------------------------|
| Gd-DTPA-BMA     | 1    | 0,47           | 20                                | 0,87           | 14                                |
|                 | 2    | 0,28           | 20                                | 0,66           | 15                                |
|                 | 3    | 0,30           | 21                                | 0,82           | 14                                |

Tableau 2.4.2.b : Valeurs des taux d'occupation et des facteurs d'agitation thermique affinés par SHARP des atomes de gadolinium pour les différents sites de dérivés préparés avec 100 et 300 mM de complexe pour les dérivés d'urate oxydase

| Thaumatine + | Site | Qsha<br>100 mM | Bsha [ $\text{\AA}^2$ ]<br>100 mM | Qsha<br>300 mM | Bsha [ $\text{\AA}^2$ ]<br>300 mM |
|--------------|------|----------------|-----------------------------------|----------------|-----------------------------------|
| Gd-DOTA      | 1    | 0,32           | 24                                | 0,61           | 29                                |
| Gd-DOTA-BOM  | 1    | 0,72           | 29                                | 0,93           | 22                                |
| Gd-DTPA-BMA  | 1    | 0,54           | 25                                | 0,70           | 28                                |
| Gd-DTPA      | 1    | 0,31           | 16                                | 0,48           | 16                                |

Tableau 2.4.2.c : Valeurs des taux d'occupation et des facteurs d'agitation thermique affinés par SHARP des atomes de gadolinium pour les différents sites de dérivés préparés avec 100 et 300 mM de complexe pour les dérivés de thaumatine

### L'exemple des sites de fixation 1 et 2 du dérivé de l'urate oxydase avec Gd-DTPA-BMA

Les valeurs des taux d'occupation des sites 1 et 2 sont de 0,47 et 0,28 pour le dérivé obtenu à 100 mM et sont de 0,87 et 0,66 à pour le dérivé obtenu à 300 mM de complexe. Comme le montre la figure 2.4.9.a, la densité 2Fo-Fc des complexes avec une occupation de l'ordre de 30% est très mal définie. Pour le site 2, il n'y a que quelques fragments de densité pour certains atomes d'azote ou d'oxygène. Pour le site 1, de la densité est visible pour les bras du ligand. La figure 2.4.9.b montre le dérivé à 300 mM. La carte expérimentale du site 1 est très bien définie et permet d'identifier l'emplacement des bras et de la chaîne du ligand. Pour le site 2, elle permet d'identifier l'emplacement des bouts de bras et de distinguer les deux différents types de bras. Les atomes d'azote de la chaîne sont également indiqués par des fragments de densité. La carte 2Fo-Fc du dérivé à 300 mM (figure 2.4.9.c) confirme sans ambiguïté l'orientation du modèle placé pour les deux sites, même si elle présente quelques lacunes pour le complexe du site 2. Les deux résidus coordonnant les ions  $\text{Gd}^{3+}$  (Asp133 et Gln131) sont en double conformation. Pour le dérivé à 100 mM leur conformation "liante" est minoritaire, c'est pourquoi la densité correspondante est mal définie (figure 2.4.9.a).

On peut résumer que, en utilisant un dérivé préparé avec une concentration de complexe de 300 mM au lieu de 100 mM, placer le modèle du ligand dans la densité, auparavant difficile pour le site 1 et impossible pour le site 2, devient aisé (site1) et possible (site 2) avec le dérivé à 300 mM.

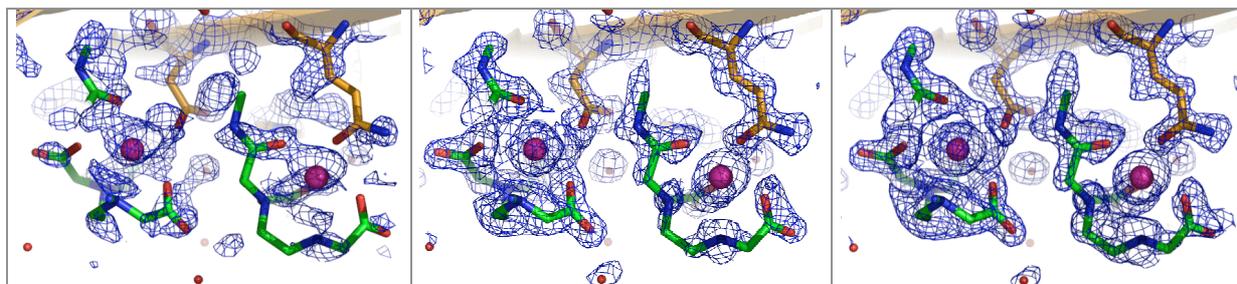


Figure 2.4.9 : Carte de densité électronique correspondant aux molécules de complexe liée à la protéine dans le dérivé urate oxydase + Gd-DTPA-BMA sur le site 1 (gauche) et 2 (droite). a) Carte 2Fo-Fc du modèle affiné pour le dérivé à 100 mM, b) Carte expérimentale pour le dérivé à 300 mM, c) Carte 2Fo-Fc du modèle affiné pour le dérivé à 300 mM.

### II.4.4.c. Comparaison de la qualité des cartes expérimentales calculées avec les phases obtenues après modification de densité

#### Exemple : Urate oxydase + Gd-DTPA-BMA 100 mM, site 1

Les cartes de densité électronique expérimentales peuvent apporter une information importante sur l'orientation des ligands dans les dérivés. Pour des sites fortement occupés et des densités bien définies, une comparaison entre les cartes expérimentales calculées avec les phases obtenues 1) après SOLOMON et 2) après SOLOMON et DM montre qu'il y a peu de différences entre 1) et 2). Pour des sites avec des taux d'occupation faibles, les cartes calculées directement après SOLOMON semblent apporter plus d'information (voir figure 2.4.11). En effet, DM semble niveller plus fortement les zones à l'interface avec le solvant. Ceci se manifeste souvent pour des molécules d'eau présentes dans la carte après SOLOMON et absentes pour la carte après DM.

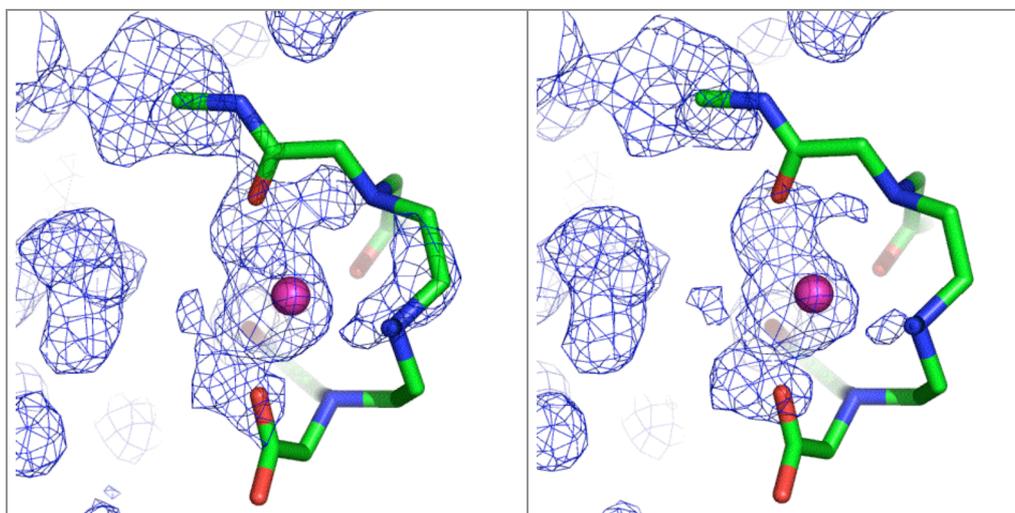


Figure 2.4.10 : Carte expérimentale (MAD) dans la région du site de fixation 1 du dérivé urate oxydase + Gd-DTPA-BMA 100 mM. a) Carte calculée avec les phases après SOLOMON, b) Carte calculée avec les phases après SOLOMON et DM. Bien que les différences soient faibles, on remarque la densité de la chaîne légèrement mieux définie en a) qu'en b).

### II.4.4.d. Utilisation de l'information de phase expérimentale pour l'affinement

#### Exemple : Glucose isomérase + Gd-DO3A 100 mM, site 1

Généralement, les deux affinements, avec des données SAD en mode mlf, utilisant F+ et F-, d'une part et avec des données MAD phasées, en mode mlhl, d'autre part, ont convergé sans problème vers la bonne solution. Ceci n'est pas étonnant puisque nos modèles de départ étaient toujours très proches de la structure finale. L'étape critique, pour nos affinements, est de construire ou de placer le modèle du ligand dans la densité électronique (expérimentale ou 1Fo-Fc) souvent mal définie. J'ai voulu déterminer si l'information de phase expérimentale apporte des améliorations dans la définition de la carte 1Fo-Fc dans la région des parties non modélisées. J'ai donc calculé deux cartes correspondant au même modèle et aux mêmes données : une sans, et l'autre avec, utilisation de l'information de phase expérimentale. Les figures 2.5.11.a et b montrent que les cartes de densité présentent de légères différences.

Néanmoins, la carte calculée avec les phases expérimentales n'est pas mieux définie que la carte sans apport de phases expérimentales.

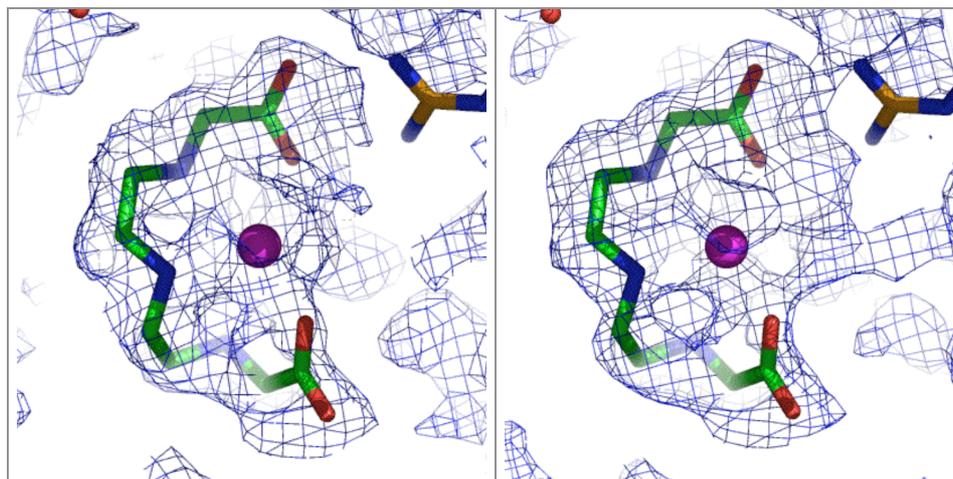


Figure 2.4.11 : Carte de densité 1Fo-Fc calculée avec le modèle affiné de la protéine mais sans placer le modèle du ligand. Pour les deux cartes, le même modèle a été utilisé. a) Carte calculée sans utiliser de l'information de phase expérimentale (mode sigma); b) Carte calculée en incluant de l'information de phase expérimentale (mode combine).

La figure 2.4.12 montre la carte de densité 2Fo-Fc calculée avec le modèle comprenant le complexe, a) sans et b) avec utilisation de l'information de phase expérimentale. Pour cette carte non plus, on ne peut constater de meilleure définition de la carte utilisant les phases expérimentales.

Néanmoins, l'obtention de bonnes phases expérimentales permet de calculer des cartes de densité électronique expérimentales précises. Ces dernières peuvent s'avérer utiles pour construire ou placer le modèle du ligand dans des cas où la carte 1Fo-Fc présente des ambiguïtés. La carte expérimentale peut alors apporter de l'information non biaisée par le modèle de la protéine.

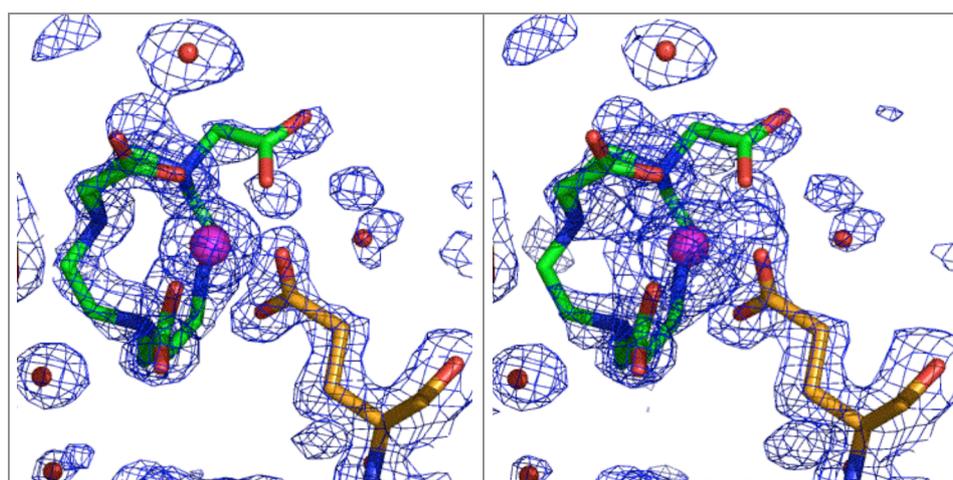


Figure 2.4.12 : Carte de densité 2Fo-Fc calculée avec le modèle affiné de la protéine et du complexe. Pour les deux cartes, le même modèle a été utilisé. a) Carte calculée sans utiliser de l'information de phase expérimentale (mode sigma); b) Carte calculée en incluant de l'information de phase expérimentale (mode combine).

## II.4.5. YGGV

| Complexe    | Nb complexes affinés | Qcns                      | Bcns [ $\text{\AA}^2$ ] | R    | Rfree | Résolution [ $\text{\AA}$ ] |
|-------------|----------------------|---------------------------|-------------------------|------|-------|-----------------------------|
| Gd-DOTMA    | 1                    | 2*0,33 ; 0,25 ; 0,17      | 24; 62; 49              | 0.28 | 0.33  | 2,70                        |
| Gd-DOTA-BOM | 0                    | 0,47 ; 0,39 ; 0,34        | 44 ; 40 ; 45            | 0.31 | 0.38  | 2,68                        |
| Gd-DO3A     | 1                    | 0,51 ; 0,17 ; 0,25 ; 0,20 | 24 ; 45 ; 25 ; 25       | 0.34 | 0.38  | 2,68                        |

| Complexe    | Site | Résidus proches, distances de l'ion de Gd                      |
|-------------|------|--|
| Gd-DOTMA    | 2    | Tyr 84, Tyr 88, Lys101 tous à 6-7 $\text{\AA}$                 |
|             | 3    | His127 ( $\sim 7 \text{\AA}$ )                                 |
| Gd-DOTA-BOM | 1    | His127 ( $\sim 7 \text{\AA}$ ) (= site 3 Gd-DOTMA)             |
|             | 2    | Tyr84 ( $\sim 6-7 \text{\AA}$ ) (= site 2 Gd-DOTMA)            |
|             | 3    | Asn10 ( $\sim 4 \text{\AA}$ )                                  |
| Gd-DO3A     | 2    | Tyr 84, Tyr 88, Lys101 (6-7 $\text{\AA}$ ) (= site 2 Gd-DOTMA) |

Tableau 2.4.3 : Statistiques d'affinement des dérivés d'YGGV.

La résolution des données, limitée à 2,7  $\text{\AA}$ , fait que l'interprétation de la densité électronique, l'identification d'éventuelles interactions entre les complexes et la protéine et enfin l'estimation de distances ne peuvent être effectuées avec précision.

### YGGV + Gd-DOTMA (charge : -1)

#### Site principal (taux d'occupation : 2\*0,33)

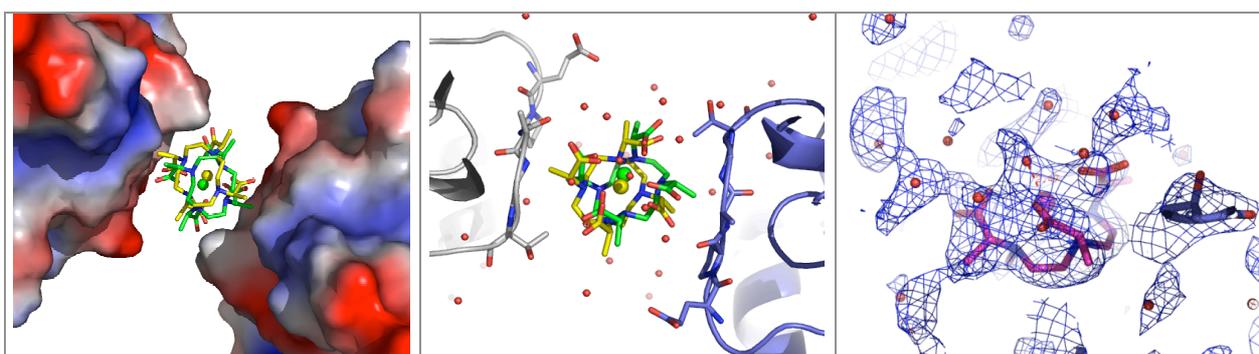


Figure 2.4.13 : Liaison du complexe Gd-DOTMA sur le site de fixation principal du dérivé d'YGGV. Le site se trouve sur un axe de symétrie cristallographique. a) Représentation en mode surface électrostatique, b) L'interaction du complexe avec la molécule, c) La carte 2Fo-Fc du modèle affiné. a) et b) En vert et en jaune : les deux molécules de complexe liées par symétrie (en réalité les deux orientations ne coexistent pas sur un même site, mais sont distribuées statistiquement).

Le site de fixation est situé sur un axe de symétrie cristallographique d'ordre 2. La molécule de complexe n'obéissant pas à cette symétrie, son orientation dans le site est statistiquement distribuée entre deux orientations possibles.

Le complexe est lié entre les chaînes principales de deux molécules de protéine symétriques (voir figure 2.4.13). Le complexe n'interagit pas directement avec la surface protéique. La densité de la molécule d'eau qui agit comme neuvième ligand de l'ion de  $\text{Gd}^{3+}$  est bien définie

(figure 2.4.13.c) (Gd-O : 2,5 Å). Le complexe est entouré de nombreuses molécules d'eau ordonnées. Comme on peut le voir sur la figure 2.4.13.c, la densité 2Fo-Fc correspondant au ligand est relativement forte. L'interprétation de la densité expérimentale et de la carte 1Fo-Fc n'est cependant pas aisée, car la densité des deux orientations du complexe est superposée. Après avoir modélisé le complexe, il reste beaucoup de densité non modélisée autour des molécules.

#### YGGV + Gd-DO3A (charge : 0)

##### Site principal (taux d'occupation : 0,51)

Le complexe est lié dans une relativement étroite cavité formée par deux molécules de protéine symétriques (figure 2.4.14.a). La faible taille de la cavité peut expliquer la fixation exclusive du plus petit complexe à cet endroit. La fixation se fait selon le mode habituellement observé pour ce complexe. L'ion  $Gd^{3+}$  est coordonné par les deux atomes d'oxygène carboxyliques de l'Asp33 (Gd-O ~2,7 Å). Les atomes d'oxygène agissent ainsi comme 8<sup>e</sup> et 9<sup>e</sup> ligand de l'ion de Gd. L'orientation exacte de la molécule de ligand ne peut être déterminée avec certitude, la densité électronique correspondant au ligand étant forte, mais assez mal définie (figure 2.4.14.c).

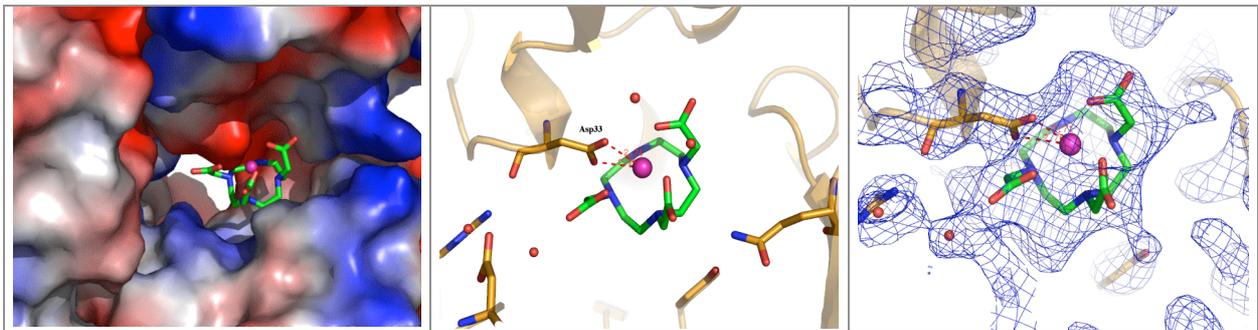


Figure 2.4.14 : Site de fixation principal du complexe Gd-DO3A dans le dérivé d'YGGV. a) La représentation en mode surface électrostatique montre l'étroite cavité de fixation, l'ion de Gd pointant vers une zone de surface fortement négative. b) L'ion est coordonné par l'Asp33. c) La carte de densité 2Fo-Fc du modèle affiné.

## II.4.6. YeaZ

| Complexe | Nb de complexes affinés | Nb de Sites | Qcns                              | Bcns [ $\text{\AA}^2$ ]   | R    | Rfree | Résolution [ $\text{\AA}$ ] |
|----------|-------------------------|-------------|-----------------------------------|---------------------------|------|-------|-----------------------------|
| Gd-DOTMA | 1                       | 8           | 0,66;0,36;0,42;<br>0,20;0,17;0,17 | 48,75,100,<br>94,67,63,68 | 0,22 | 0,25  | 2,3                         |
| Gd-DO3A  | 2                       | 6           | 0,74;0,30;0,33;<br>0,67;0,18;0,29 | 30;41;53;<br>39;48; 45    | 0,23 | 0,29  | 2,7                         |

| Complexe | Site | Résidus proches  | Complexe | Site | Résidus proches  |
|----------|------|------------------|----------|------|------------------|
| Gd-DOTMA | 1    | AspB182, HisC33  | Gd-DO3A  | 1    | AspB182, HisC33  |
|          | 2    | TrpD102, ArgD103 |          | 2    | GluC32           |
|          | 3    | PheD67           |          | 3    | TrpD102, ArgD103 |
|          | 4    | HisB33, AspC182  |          | 4    | HisB33, AspC182  |
|          | 5    | TyrB216          |          | 5    | TyrB216          |
|          | 6    | PheA67           |          | 6    | TrpB102          |

Tableau 2.4.4 : Statistiques d'affinement des dérivés affinés de YeaZ. La protéine cristallise avec quatre molécules dans l'unité asymétrique, certains sites de fixation se retrouvent pour plusieurs molécules. Ces sites structurellement proches sont indiqués dans la même couleur. Les deux complexes ont la plupart des sites de fixation en commun. Les sites correspondants sont également indiqués dans la même couleur.

Les deux complexes Gd-DOTMA et Gd-DO3A ont plusieurs sites de fixation en commun. La protéine cristallise avec 4 molécules par unité asymétrique. Pour certains sites de fixation, le complexe se lie également sur des sites équivalents, structurellement proches des autres monomères. Les taux d'occupation des sites homologues sont cependant différents. Le mode de fixation a pu être affiné pour les sites de fixation principaux. Le taux d'occupation des sites minoritaires est trop faible pour pouvoir identifier le mode de fixation des complexes.

### YeaZ + Gd-DOTMA (charge : -1)

#### Site 1 (taux d'occupation : 0,66)

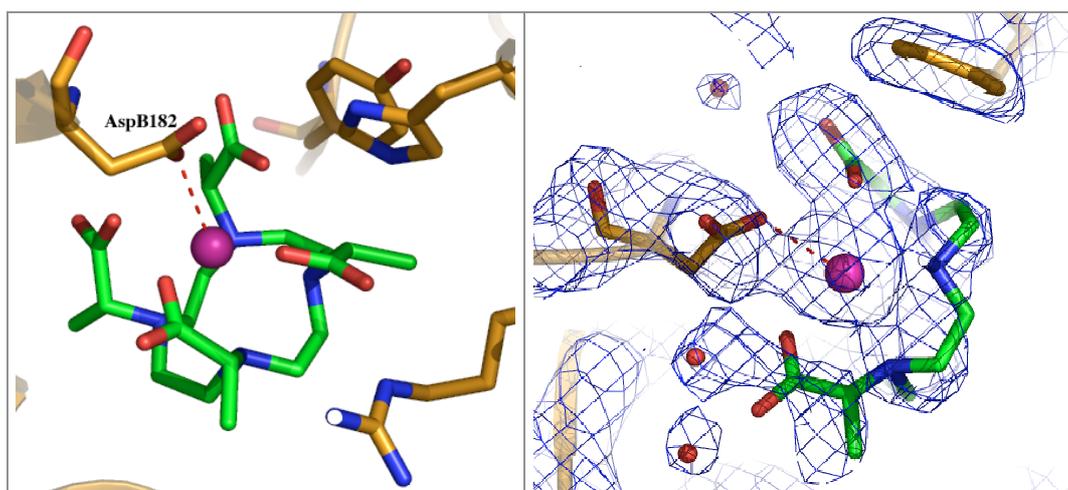


Figure 2.4.15 : Site de fixation principal du dérivé yeaZ avec Gd-DOTMA. a) Le complexe se lie via son côté ionique. Il est entouré de résidus protéiques. b) Densité 2Fo-Fc du modèle affiné.

Le site de fixation se trouve dans une cavité, entouré de résidus appartenant à différents monomères de la protéine (figure 2.4.15.a). Le complexe se lie via son côté ionique, l'ion  $Gd^{3+}$  étant coordonné par un atome d'oxygène carboxylique de l'AspB182 (Gd-OD : 2,7 Å). Le macrocycle est tourné vers le solvant et les côtés du complexe sont entourés de résidus protéiques et d'un réseau de molécules d'eau ordonnées. Comme le montre la figure 2.4.15.b, la densité 2Fo-Fc est bien définie pour certaines parties du ligand alors qu'elle est moins bonne pour d'autres parties. Après affinement du modèle, il reste un peu de densité 1Fo-Fc résiduelle.

### YeaZ + Gd-DO3A (charge : 0)

#### Site 1 (taux d'occupation : 0,74)

Le site de fixation est le même que le site principal du dérivé de yeaZ avec Gd-DOTMA. La fixation du complexe engendre des changements de conformation de résidus du site de fixation (voir figure 2.4.16.a). Ces mouvements permettent l'interaction entre le ligand et deux résidus protéiques et la coordination de l'ion  $Gd^{3+}$  par les deux atomes d'oxygène carboxyliques de l'AspB182 qui fournit ainsi le 8<sup>e</sup> et 9<sup>e</sup> ligand à l'ion  $Gd^{3+}$ . En effet, malgré la relativement basse résolution des données de 2,7 Å, la carte de densité électronique montre assez clairement l'interaction entre l'atome d'oxygène du ligand et l'atome d'azote NE2 de l'HisC33 (O-NE2 : 2,7 Å) d'une part et l'interaction du même atome d'oxygène avec l'atome d'azote NE de l'ArgC31 (O-NE : 2,7 Å) d'autre part (figure 2.4.16.b). Le taux d'occupation, plus élevé pour le complexe Gd-DO3A que pour le complexe Gd-DOTMA, s'explique peut-être par la plus petite taille du complexe Gd-DO3A et par le renforcement de la fixation par les interactions ligand-protéine supplémentaires. Cet effet est particulièrement important pour le site 4 (site équivalent au site 1), pour lequel les taux d'occupation sont de 0,64 (Gd-DO3A) et de 0,20 (Gd-DOTMA) respectivement. La bien meilleure définition de la densité électronique du complexe Gd-DO3A (figure 2.4.16.b) peut être due à une orientation plus figée de ce complexe. L'affinement du modèle du complexe lié dans le site 4 (site équivalent au site principal) montre que la fixation se fait de la même manière, la densité électronique étant cependant moins bien définie.

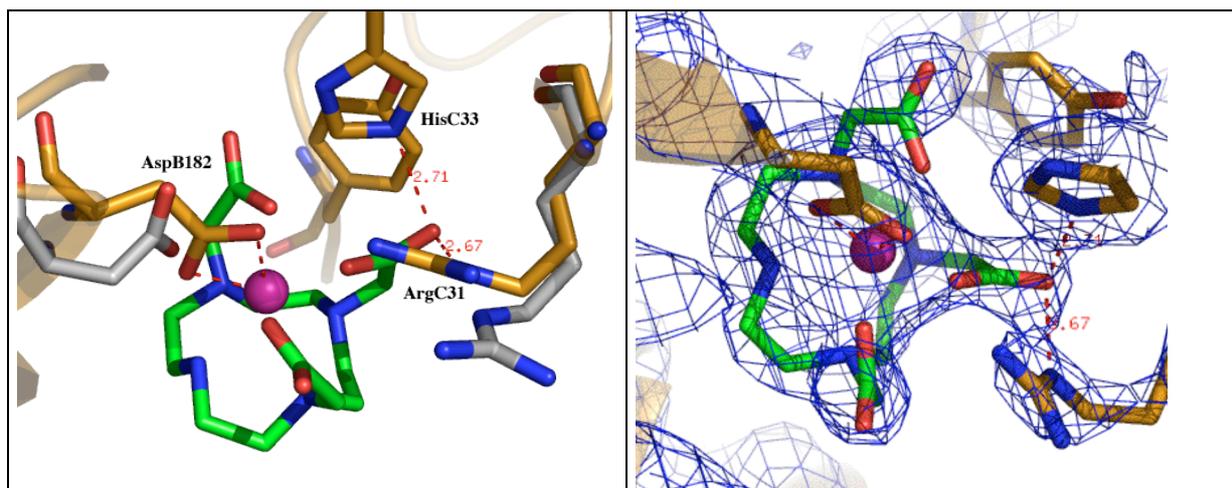


Figure 2.4.16 : Site de fixation principal du dérivé yeaZ + Gd-DO3A. a) en or : l'orientation des résidus protéiques dans le dérivé yeaZ + Gd-DO3A, en gris : leur orientation dans le dérivé yeaZ + Gd-DOTMA. b) Carte 2Fo-Fc du modèle affiné.

## II.4.7. Glucose isomérase

| Complexe | Nb de complexes affinés | Qcns                          | Bcns [ $\text{\AA}^2$ ] | R    | Rfree | Résolution [ $\text{\AA}$ ] |
|----------|-------------------------|-------------------------------|-------------------------|------|-------|-----------------------------|
| Gd-DTPA  | 1                       | 0,45                          | 19                      | 0,19 | 0,20  | 1,45                        |
| Gd-DO3A  | 3                       | 0,56;0,56;0,37;0,24;0,15;0,10 | 22;26;21;19;15;15       | 0,21 | 0,21  | 1,44                        |

| Complexe | Site | Résidus proches                               |
|----------|------|---|
| Gd-DO3A  | 4    | Glu132, Arg10                                 |
|          | 5    | proche du site 3 (distance 5,3 $\text{\AA}$ ) |
|          | 6    | Asp297, Arg42                                 |

Tableau 2.4.5 : Statistiques d'affinement pour les dérivés affinés de glucose isomérase

### Glucose isomérase + Gd-DO3A (charge 0)

Le dérivé glucose isomérase + Gd-DO3A présente un nombre relativement élevé de six sites de fixation du complexe, dont les taux d'occupation varient de 10 à 56%. Pour tous les sites (sauf le site 5), un glutamate ou un aspartate se trouve à proximité de l'ion ainsi qu'une ou plusieurs arginines. Le taux d'occupation élevé des sites 1, 2 et 3 a permis l'affinement de la structure du complexe lié à la protéine. Il est très probable que le mode de fixation aux sites 4 et 6 soit le même que celui qui est observé pour les sites 1, 2 et 3.

#### Site 1 (taux d'occupation : 0,56)

Comme le montre la figure 2.4.17.a, la fixation du complexe se fait grâce à la coordination de l'ion  $\text{Gd}^{3+}$  par les deux atomes d'oxygène carboxyliques du Glu207, le macrocycle étant tourné vers le solvant. La fixation est renforcée par la formation de ponts salins entre deux atomes d'oxygène d'un même bras du ligand et les deux atomes d'azote NH1 et NH2 de l'Arg208 (O-NH: 2,8 et 2,9  $\text{\AA}$  respectivement). Le taux d'occupation du site est proche de 50%, ce qui fait qu'il a été possible d'affiner la double conformation des deux résidus protéiques impliqués dans la fixation du complexe. Comme le montre la figure 2.4.17.b, la densité 2Fo-Fc du modèle affiné est assez bien définie.

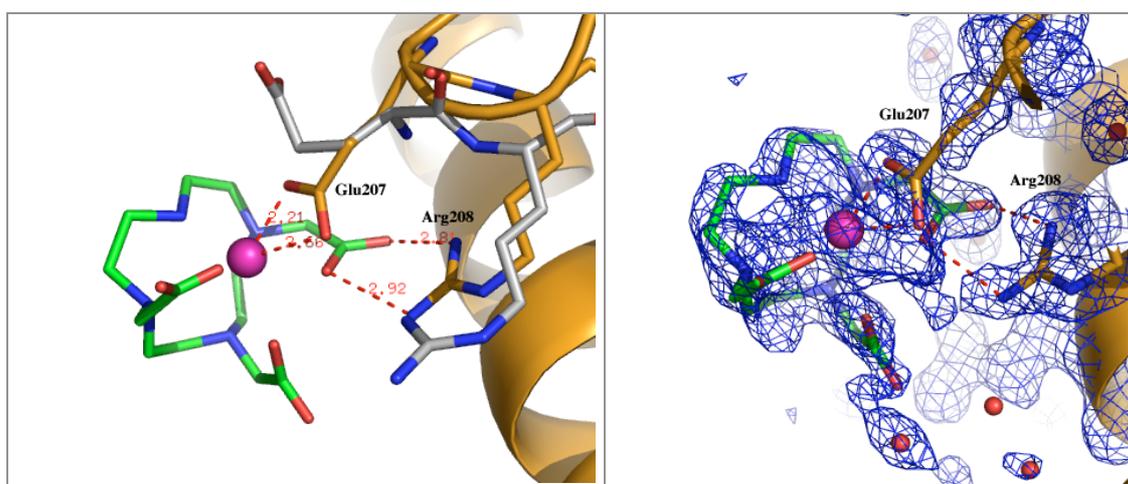


Figure 2.4.17 : Site de fixation 1 du dérivé glucose isomérase + Gd-DO3A. a) Les deux résidus en interaction avec le complexe sont en double conformation. En or : la conformation "liante", en gris : la conformation "non-liante". b) La densité 2Fo-Fc du modèle affiné est bien définie.

### Site 2 (taux d'occupation : 0,56)

La coordination de l'ion se fait, de manière analogue au site 1, par le résidu Glu325 (Gd-O : 2,6 et 2,4 Å respectivement). La double conformation du résidu est montrée dans la figure 2.4.18.a. Les atomes d'oxygène de deux bras différents du ligand (atomes d'oxygène proches de l'ion) forment des ponts salins avec les atomes d'azote NH de deux arginines (O-NH(Arg374) : 2,8 Å et O-NH(Arg387) : 2,9 Å). La densité 2Fo-Fc calculée pour la structure affinée, non montrée ici, est assez bien définie, sauf pour le bras du ligand qui n'interagit pas avec la protéine.

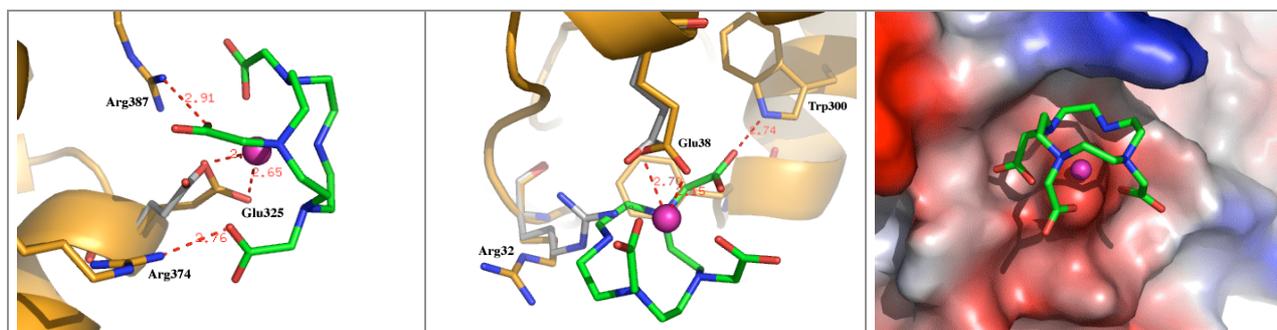


Figure 2.4.18 : a) Site de fixation 2 du dérivé glucose isomérase + Gd-DO3A. En gris, la conformation "non-liante" du résidu Glu325. b) Site de fixation 3 du dérivé glucose isomérase + Gd-DO3A. En gris, les conformations "non-liantes" des résidus voisins au site. c) Représentation en mode surface électrostatique du site 3.

### Site 3 (taux d'occupation : 0,37)

La coordination de l'ion se fait de manière analogue aux sites 1 et 2 par les deux atomes d'oxygène du Glu38, dont la double conformation est indiquée dans la figure 2.4.18.b. L'Arg32 est également en double conformation, sa conformation "native" provoquerait un conflit stérique avec le ligand. Un atome d'oxygène du ligand interagit avec l'atome d'azote NE1 du Trp300 (O-NE1 : 2,7 Å). La figure 2.4.18.c montre la surface électrostatique de la protéine, bien adaptée par sa forme à la fixation du complexe. La zone de surface en contact avec le complexe est chargée négativement.

### Glucose isomérase + Gd-DTPA (charge -2)

Initialement, il m'a manqué le modèle du complexe à utiliser pour cet affinement-ci, car la structure cristallographique du complexe Gd-DTPA n'était pas connue. J'ai pu résoudre ensuite la structure du complexe avec un dérivé d'une autre protéine où la densité électronique du complexe était mieux définie. Cette structure m'a servi pour continuer l'affinement du dérivé de glucose isomérase.

Dans ce dérivé, le complexe se lie sur un site unique, occupé à 45%. Comme le montre la figure 2.4.19.b, la densité électronique correspondant au complexe n'est que partiellement bien définie et présente des lacunes au niveau de la chaîne et du bras qui pointe vers le solvant. Le côté ionique du complexe est tourné vers la surface protéique. Une molécule d'eau, liée à l'ion  $Gd^{3+}$  et dont la densité est bien définie, sert de neuvième ligand (Gd-O : 2,8 Å) et pontre l'ion à la surface protéique et à d'autres molécules d'eau.

À proximité du site de fixation, une partie de densité 1Fo-Fc, relativement étendue, n'est pas occupée par des résidus protéiques. Je l'ai modélisée en plaçant le modèle d'une molécule de diméglumine, molécule qui sert de contre-ion dans la solution médicale du complexe utilisée. Le modèle de cette molécule ne permet cependant pas de remplir entièrement la densité résiduelle.

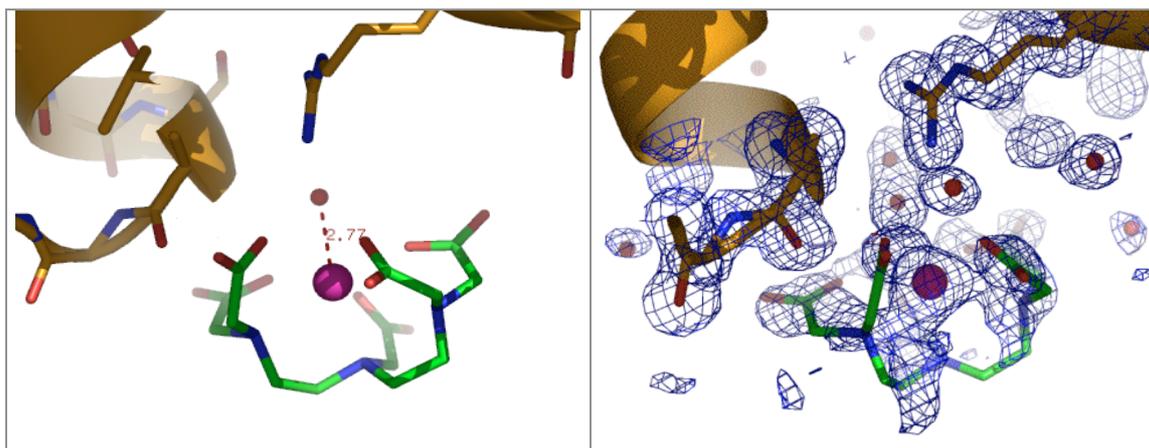


Figure 2.4.19 : Fixation du complexe Gd-DTPA dans la protéine glucose isomérase. a) Interaction du complexe avec la protéine et une molécule d'eau. b) La densité 2Fo-Fc du complexe n'est pas partout bien définie.

## II.4.8. Lysozyme

| Complexe<br>( $C_{\text{complexe}} = 300 \text{ mM}$ ) | Nb de complexes affinis | Qcns                | Bcns [ $\text{\AA}^2$ ] | R    | Rfree | Résolution [ $\text{\AA}$ ] |
|--|-------------------------|---------------------|-------------------------|------|-------|-----------------------------|
| Gd-HPDO3A  | 2                       | 1,00;0,87           | 13;11                   | 0,17 | 0,20  | 1,53                        |
| Gd-DOTA-BOM  | 1                       | 0,74                | 22                      | 0,22 | 0,25  | 1,53                        |
| Gd-DOTA  | 2                       | 0,82;0,53           | 17;16                   | 0,19 | 0,19  | 1,54                        |
| Gd-DO3A  | 2                       | 0,96;0,69;0,31;0,18 | 38;21;17;13             | 0,26 | 0,28  | 1,54                        |

Tableau 2.4.6 : Statistiques d'affinement pour les dérivés affinés de lysozyme

Le site de fixation dont le taux d'occupation est le plus élevé est identique pour les quatre complexes. Le deuxième site est identique pour trois des complexes, il est inoccupé pour le complexe Gd-DOTA-BOM. La particularité de la protéine est sa taille relativement petite et le faible contenu de solvant du cristal de  $\sim 25\%$  seulement, ce qui limite le nombre de sites de fixations potentiels. Les deux sites de fixation principaux sont situés dans une cavité étroite formée par trois molécules symétriques de la protéine (voir figure 2.4.20). Les figures 2.5.21.a et b montrent plus en détail les sites de fixation 1 et 2. On remarque le bon accord entre la taille des complexes et la taille de la cavité. Les zones de surface en contact avec le complexe sont essentiellement chargées positivement.

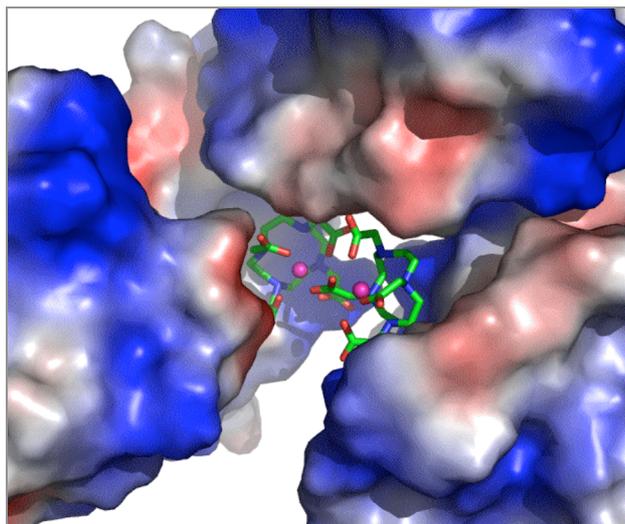


Figure 2.4.20 : Représentation en mode surface électrostatique de la cavité de fixation formée par trois molécules de protéine (avec deux molécules de complexe Gd-DOTA).

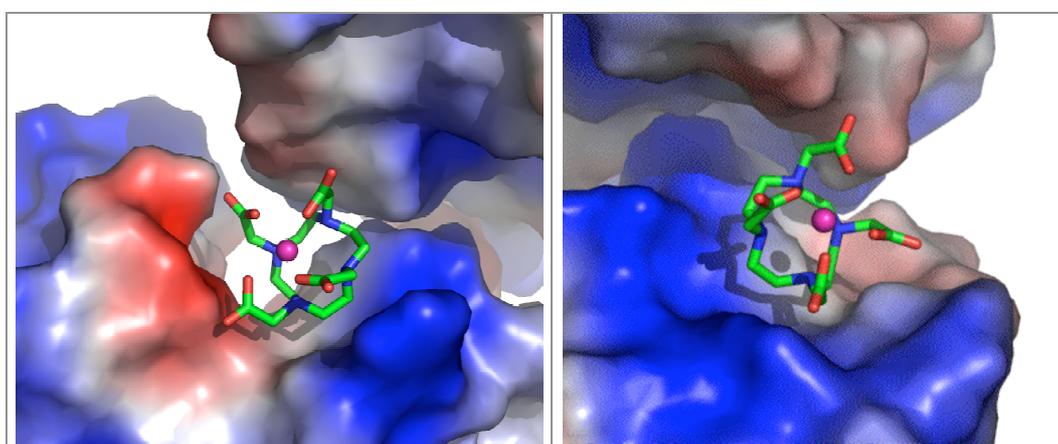


Figure 2.4.21 : a) Site de fixation 1, b) site de fixation 2 (avec des molécules de complexe Gd-DOTA).

Pour le dérivé du complexe Gd-DOTA-BOM, la structure du complexe a été modélisée par la structure du complexe Gd-DOTMA, car il n'y a pas de densité qui correspondrait au groupement (phénylméthoxy)méthyle du complexe et la structure cristallographique du complexe seul n'est pas connue.

### Mode de fixation

La figure 2.4.22 montre la cavité de fixation avec les modèles affinés des différents complexes. Les quatre figures correspondent à la même perspective, ce qui permet de comparer l'orientation des différents complexes. Pour les quatre complexes, le macrocycle de la molécule liée au site 1 est en interaction hydrophobe avec le cycle aromatique du Trp62, les plans formés par les deux cycles sont parallèles. Pour les complexes Gd-HPDO3A et Gd-DO3A, la distance entre les deux plans est inférieure à 3,5 Å, elle est de  $\sim 3,8$  Å pour Gd-DOTA et de  $\sim 4$  Å pour Gd-DOTA-BOM. Dans le site 2, le macrocycle des complexes Gd-DOTA et Gd-DO3A est en interaction hydrophobe avec le Trp123, le macrocycle étant parallèle au cycle aromatique du tryptophane (distance  $\sim 3,5$  Å). Pour le complexe Gd-DO3A, le macrocycle et le cycle aromatique du tryptophane sont cependant légèrement décalés. La molécule du complexe Gd-HPDO3A dans le site 2 est dans une orientation différente par rapport aux deux autres complexes. La molécule oppose son côté ionique au côté ionique du complexe du site 1. C'est un côté formé par deux bras du complexe qui se place parallèlement au Trp123. Alors que les plans formés par le macrocycle des complexes sont parallèles pour

les deux molécules de Gd-HPDO3A, ils ne le sont pas pour les complexes Gd-DO3A et Gd-DOTA. Pour une description plus détaillée du mode de fixation du complexe Gd-HPDO3A voir plus loin, au paragraphe II.4.8.a.

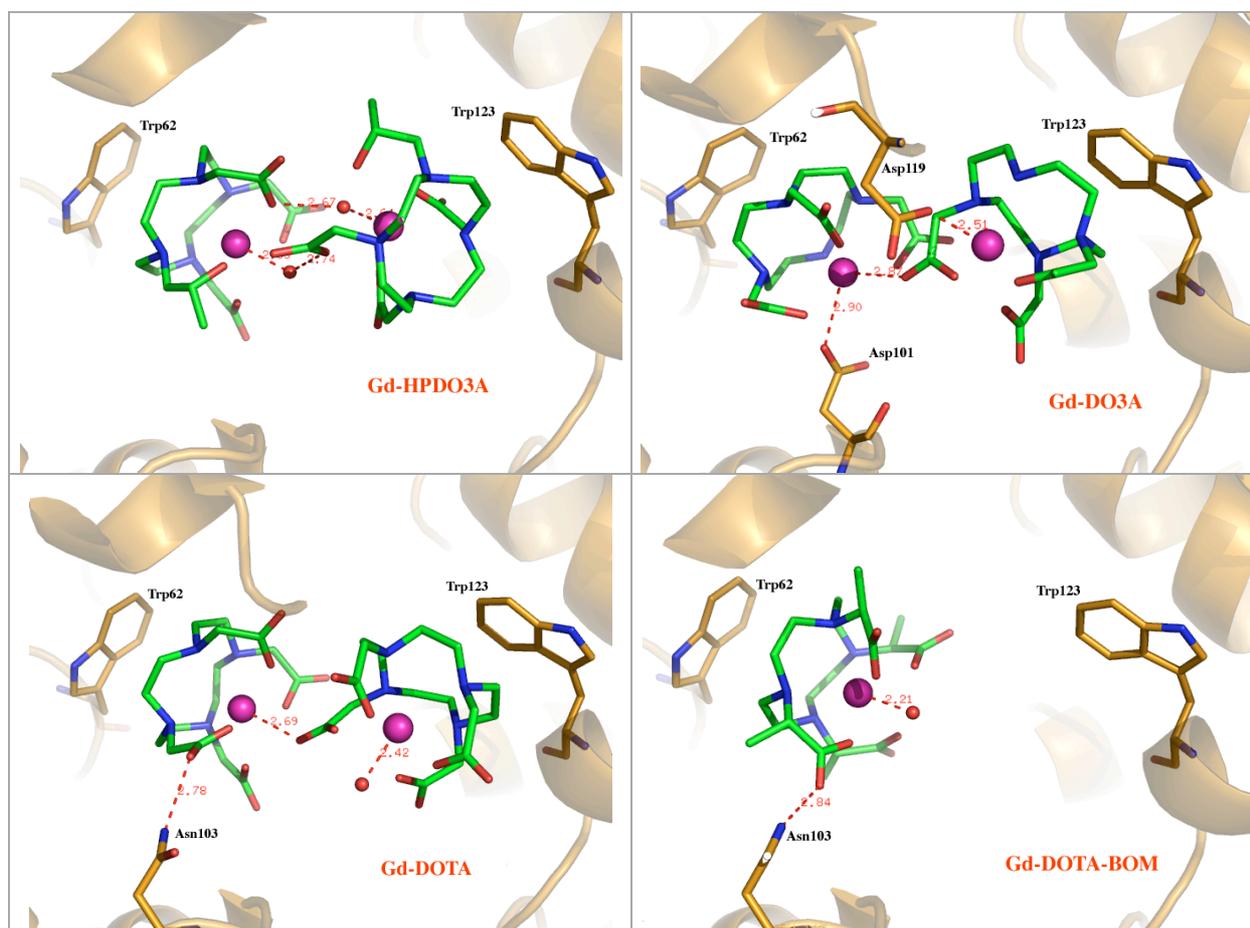


Figure 2.4.22 : Modèles affinés des différents complexes liés dans le site de fixation 1 (gauche) et 2 (droite) des dérivés du lysozyme. Les quatre figures correspondent à la même perspective.

### Gd-DO3A (charge : 0)

#### Site 1 (taux d'occupation : 0,96)

Les deux atomes qui assurent la coordination de l'ion  $Gd^{3+}$  sont l'atome d'oxygène OD de l'Asp101 (Gd-O : 2,9 Å) et un atome d'oxygène appartenant au complexe lié dans le site 2 (O(ligand 2)-Gd(site 1) : 2,9 Å).

#### Site 2 (taux d'occupation : 0,69)

Un des atomes d'oxygène du ligand pointe vers l'ion  $Gd^{3+}$  du site 1. L'ion  $Gd^{3+}$  du site 2 est coordonné par un ou deux atomes d'oxygène carboxyliques de l'Asp119. La densité de la double conformation du résidu Asp119 n'est pas assez bien définie pour déterminer les interactions de manière précise.

### Gd-DOTA (charge : -1)

#### Site 1 (taux d'occupation : 0,82)

Le neuvième ligand de l'ion  $Gd^{3+}$  est présent sous forme d'un atome d'oxygène carboxylique du complexe fixé dans le site 2 (Gd-O : 2,7 Å). Une liaison H entre un atome d'oxygène du

ligand et l'atome d'azote ND2 de l'Asn103 renforce la fixation du complexe (O-ND2 : 2,8 Å). Les molécules de complexe liées sur les sites 1 et 2 sont entourées d'un réseau de molécules d'eau ordonnées, stabilisant la fixation des complexes. La densité 2Fo-Fc du complexe dans le site 1, qui n'est pas montrée ici, est bien définie.

#### Site 2 (taux d'occupation : 0,53)

Le neuvième ligand de l'ion est formé par une molécule d'eau dont la densité est bien définie (Gd-O : 2,4 Å). Un des atomes d'oxygène du ligand pointe vers l'ion Gd<sup>3+</sup> du premier site.

#### Gd-DOTA-BOM (charge : -1)

Le complexe (représenté par le modèle du complexe Gd-DOTMA) ne se lie que dans le site 1. De manière analogue au complexe Gd-DOTA, un atome d'oxygène du ligand forme une liaison H avec l'atome d'azote ND2 de l'Asn103 (O-ND2 : 2,8 Å). Le complexe est entouré d'un réseau de molécules d'eau ordonnées. Le neuvième ligand de l'ion est une molécule d'eau dont la densité est très bien définie.

La figure 2.4.23 résume les mouvements des résidus engendrés par la fixation des différents complexes.

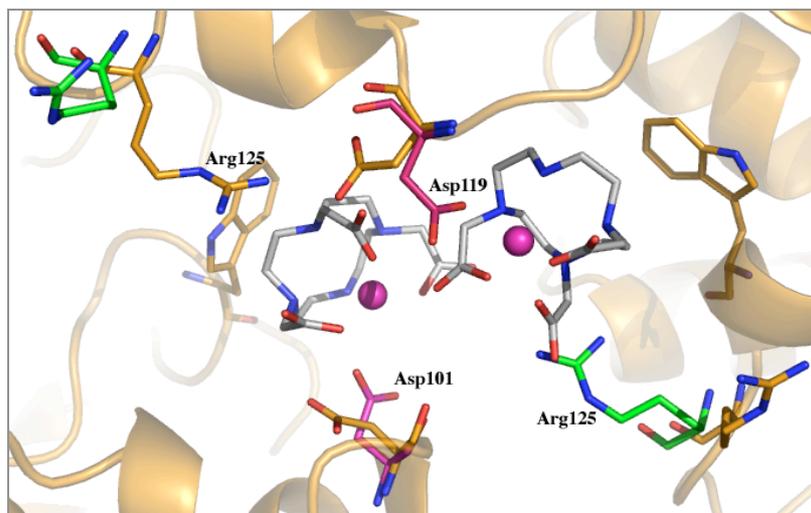


Figure 2.4.23 : Figure indiquant les mouvements de résidus protéiques engendrés par la fixation des complexes. En gris : modèle du complexe Gd-DO3A pour indiquer l'emplacement des complexes liés. En jaune : orientation générale des résidus dans les dérivés sauf mention contraire. En rouge : conformation dans le dérivé lysozyme + Gd-DO3A. En vert : orientation de l'Arg125 dans la structure native.

#### II.4.8.a. Le rôle des molécules d'eau ordonnées pour la fixation coopérative des deux molécules du complexe Gd-HPDO3A (charge : 0)

Les molécules de complexe liées dans le site 1 et le site 2 (taux d'occupation : 1,0 et 0,87 respectivement) ne forment aucune liaison H avec les résidus protéiques. Le neuvième ligand des ions Gd<sup>3+</sup> est présent sous forme de molécules d'eau dont la densité est très bien définie pour les deux sites (O-Gd(site1) : 2,3 Å, O-Gd(site 2) : 2,6 Å). Le taux d'occupation des deux sites, élevé malgré le faible nombre de liaisons directes entre les complexes et la protéine, peut éventuellement être expliqué par le caractère coopératif de la fixation des deux molécules. En effet, les deux molécules légèrement décalées se faisant face, cette orientation permet d'établir un réseau d'interactions symétriques (voir figure 2.4.24). Ainsi, deux liaisons H sont formées entre des atomes d'oxygène des deux ligands. L'atome d'oxygène hydroxyle du bras portant le groupement méthyle interagit avec un atome d'oxygène carboxylique du

ligand situé en face (et réciproquement) (O-O : 2,7 Å). De plus, un réseau symétrique de six molécules d'eau ordonnées entre les deux molécules de complexe renforce l'interaction, toutes les molécules d'eau interagissant avec les deux molécules de complexe. La densité 2Fo-Fc des deux ligands affinés est en effet très bien définie. Ainsi il est possible de distinguer les trois bras carboxyliques des ligands du bras portant un groupement méthyle, ce qui a permis de corriger l'orientation de la molécule décrite dans Girard *et al.* (2002).

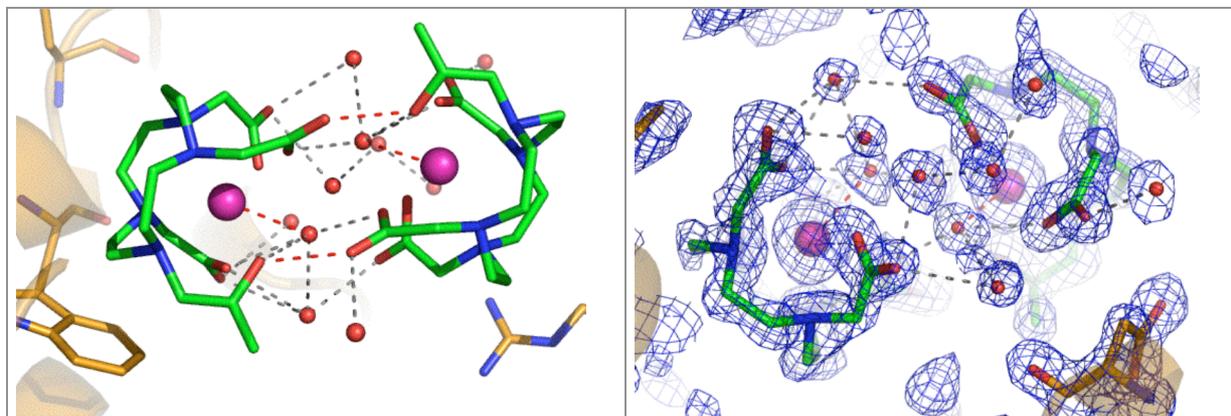


Figure 2.4.24 : a) Un réseau de molécules d'eau assure le contact entre les deux molécules de complexe. Les liaisons H sont indiquées par des traits pointillés gris. b) La densité 2Fo-Fc des molécules de complexe et des molécules d'eau est très bien définie.

#### II.4.8.b. Conclusions sur la sélectivité de la fixation des complexes dans les dérivés de lysozyme

Les deux complexes à ligand linéaire ne sont pas fixés dans les cristaux. Il semble que la présence du macrocycle soit nécessaire pour fixer le complexe, car il permet l'interaction hydrophobe avec le cycle aromatique du tryptophane (ou des tryptophanes). La fixation des complexes ne semble pas dépendre de la charge globale du complexe, car aussi bien des complexes neutres que des complexes chargés négativement se sont fixés.

#### II.4.9. Thaumatine

| Complexe<br>( $C_{\text{complexe}} = 300 \text{ mM}$ ) | Nb de complexes affinés | Qcns       | Bcns [Å <sup>2</sup> ] | R    | R <sub>free</sub> | Résolution [Å] |
|--|-------------------------|------------|------------------------|------|-------------------|----------------|
| Gd-DOTA-BOM  | 2                       | 0,84; 0,48 | 18; 15                 | 0,20 | 0,20              | 1,45           |
| Gd-DTPA-BMA  | 1                       | 0,66       | 22                     | 0,19 | 0,20              | 1,45           |
| Gd-DOTA  | 1                       | 0,54       | 20                     | 0,20 | 0,20              | 1,45           |
| Gd-DTPA  | 1                       | 0,50       | 15                     | 0,19 | 0,20              | 1,45           |

Tableau 2.4.7 : Statistiques d'affinement pour les dérivés affinés de thaumatine

Le site de fixation principal des quatre complexes Gd-DOTA-BOM, Gd-DTPA-BMA, Gd-DOTA et Gd-DTPA est identique. Il s'agit d'une cavité de 14 Å de diamètre et 11 Å de profondeur dont la surface est chargée négativement (figure 2.4.25.a). Tous les complexes se lient à la protéine dans une orientation similaire, leur macrocycle ("pseudomacrocycle" pour les ligands linéaires DTPA et DTPA-BMA) étant en superposition, ainsi que, du moins approximativement, l'orientation des bras des ligands (figure 2.4.25.b). La fixation des complexes n'engendre aucun changement d'orientation de résidus protéiques. Le complexe

Gd-DOTA-BOM est le seul complexe qui se lie à un deuxième endroit de la surface protéique. Les quatre autres complexes ne se lient ni au site de fixation principal ni à aucun autre endroit.

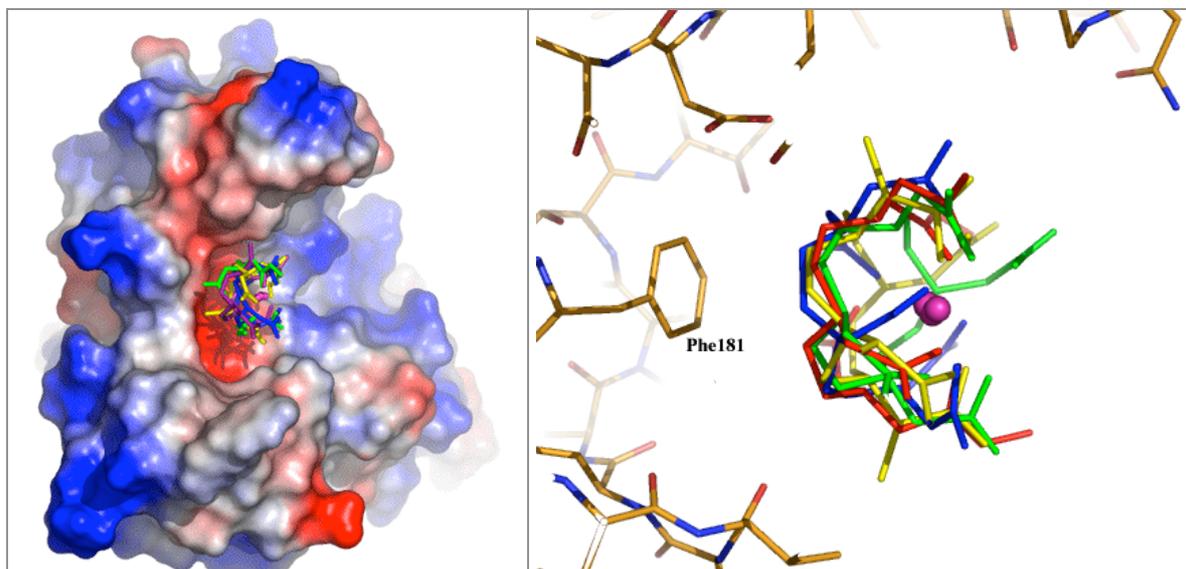


Figure 2.4.25 : a) Représentation en mode surface électrostatique du monomère de la protéine et de la cavité de fixation avec les modèles affinés des quatre complexes. b) Cavité de fixation avec les résidus protéiques voisins et la superposition des quatre complexes qui se fixent à cet endroit. En jaune : le ligand du complexe Gd-DOTA-BOM modélisé par le complexe Gd-DOTMA, en rouge : le ligand de Gd-DTPA-BMA, en bleu : le ligand de Gd-DTPA, en vert : le ligand de Gd-DOTA.

### Mode de fixation - site 1

Tous les complexes se lient via le côté ligand, le macrocycle (ou "pseudomacrocycle") étant parallèle au cycle aromatique de la Phe181 de la cavité de fixation, à une distance d'environ 3,5 Å (figure 2.4.26). On remarque cependant un léger décalage pour le complexe Gd-DOTA. La fixation des complexes Gd-DTPA-BMA et Gd-DOTA met en jeu la formation d'un pont salin entre un atome d'oxygène carboxylique du ligand et l'atome d'azote NZ de la Lys49 (O-NZ : ~2.5 Å). La fixation du complexe Gd-DTPA met en jeu une liaison H entre un atome d'oxygène du ligand et l'atome d'oxygène OG de la Thr190 (O-O : 2,9 Å). Pour les quatre complexes, l'ion  $Gd^{3+}$  pointe vers le solvant. Pour trois des complexes, une molécule d'eau liée à l'ion  $Gd^{3+}$  joue le rôle de neuvième ligand. Pour le complexe Gd-DOTA, il n'y a pas de densité électronique qui indiquerait la présence d'une molécule d'eau liée à l'ion  $Gd^{3+}$ . La fixation des quatre complexes implique un réseau de molécules d'eau ordonnées entre le ligand et la surface protéique. Malgré leurs taux d'occupation élevés, supérieurs à 0,5, la densité électronique des ligands est généralement mal définie. La densité permet de facilement identifier l'emplacement et l'orientation du macrocycle (ou "pseudomacrocycle"), mais elle est généralement assez "globulaire" et il est difficile de distinguer l'emplacement des bras des ligands dans la densité. Le manque de détail de la densité électronique indique probablement un désordre de rotation du ligand. Cependant, la position de l'ion  $Gd^{3+}$  ne semble pas affectée par le désordre, les valeurs du facteur d'agitation thermique étant de l'ordre de 20 Å<sup>2</sup> et le pouvoir de phasage des différents dérivés étant excellent comme le prouvent les statistiques de phasage.

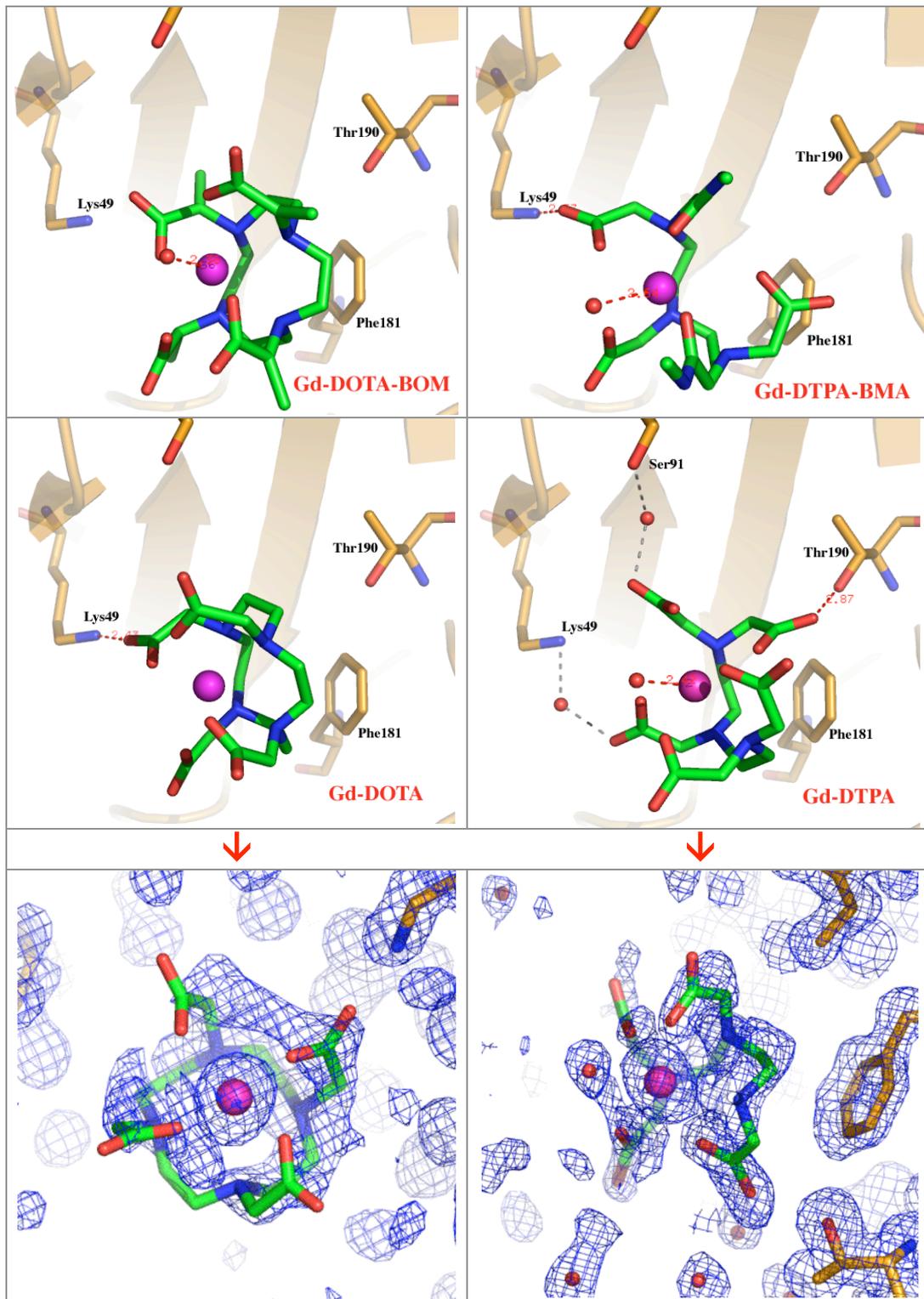


Figure 2.4.26 : a), b), c) et d) Site de fixation principal de la thaumatine et les modèles affinés des quatre complexes, le complexe Gd-DOTA-BOM étant représenté par le complexe Gd-DOTMA. Pour les complexes pour lesquels le désordre de fixation fait qu'on ne peut pas placer le modèle du ligand sans ambiguïté, la conformation qui correspond le mieux à la densité est représentée. e) et f) : Cartes de densité 2Fo-Fc des modèles affinés de e) Gd-DOTA et f) Gd-DTPA. On remarque que la densité du complexe Gd-DOTA est assez mal définie.

Gd-DOTA-BOM, le complexe avec le taux d'occupation le plus élevé présente la densité la moins bien définie. Il n'y a pas de densité indiquant l'emplacement éventuel du groupement (phénylméthoxy)méthyle, le complexe a donc été modélisé par le modèle du complexe Gd-

DOTMA. Comme on peut le voir sur la figure 2.4.26.e, la densité des bras du complexe Gd-DOTA est assez mal définie, sauf pour le bras impliqué dans un pont salin. La densité du complexe Gd-DTPA est la mieux définie de tous les complexes, bien que le complexe se lie avec le taux d'occupation le plus bas (figure 2.4.26.f). Il est possible que le ligand de ce complexe soit plus ordonné à cause de la liaison H entre le ligand et la protéine et les deux interactions par liaison H via des molécules d'eau, indiquées dans la figure 2.4.26.d. Cependant, il n'aurait pas été possible de construire le modèle du complexe à partir de la densité 1Fo-Fc de ce dérivé, et nous avons utilisé le modèle du complexe construit avec le dérivé d'une autre protéine.

### Conclusions concernant le site 1

La fixation exclusive des quatre complexes ne s'explique apparemment ni par leur charge (les complexes liés portent des charges -1, -2 ou sont neutres), ni par la forme cyclique ou linéaire de leur ligand.

### Gd-DOTA-BOM, site de fixation 2 (taux d'occupation : 0,48)

Le complexe Gd-DOTA-BOM se lie sur un deuxième site, d'une manière tout à fait différente du premier site (figure 2.4.27). Le complexe tourne son côté formé par deux bras du ligand vers la surface protéique, l'ion  $Gd^{3+}$  et le macrocycle étant exposés au solvant. La liaison du complexe se fait par formation de liaisons H entre deux atomes d'oxygène du ligand et les atomes d'oxygène des chaînes latérales d'une sérine et d'une tyrosine (O(ligand)-O(Ser10) : 2,8 Å, O(ligand)-O(Tyr11) : 2,5 Å). La fixation du complexe implique également un réseau de molécules d'eau ordonnées. La densité électronique qui correspond à la partie du complexe impliquée dans les liaisons H vers les résidus protéiques est relativement bien définie, alors que la densité de la partie tournée vers le solvant est très mal définie (figure 2.4.27.b). La densité 2fo-Fc de la molécule d'eau agissant comme neuvième ligand est très bien définie (Gd-O : 2,7 Å).

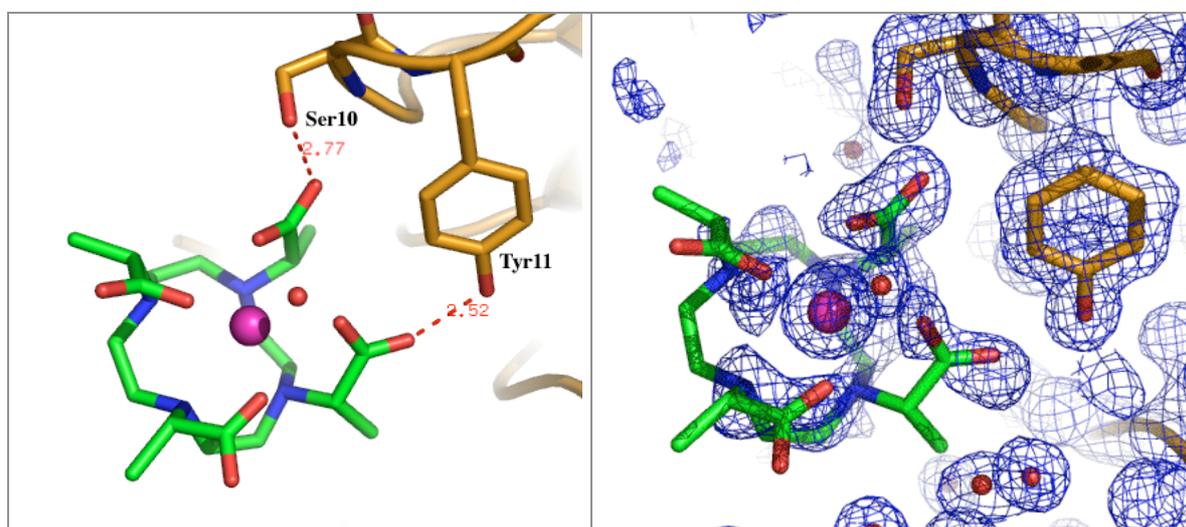


Figure 2.4.27 : a) Deuxième site de fixation du complexe Gd-DOTA-BOM. Comme il n'y a pas de densité électronique qui indiquerait l'emplacement du groupement (phénylméthoxy)méthyle du ligand, le complexe est modélisé par la structure du complexe Gd-DOTMA. b) Carte 2Fo-Fc du modèle affiné.

### Conclusions

De manière analogue au site de fixation dans le lysozyme, le site 1 de ce dérivé sert de site de fixation pour plusieurs complexes à la fois. Les sites de fixation de ces deux protéines correspondent à des cavités prononcées dans la structure cristalline de la protéine.

Dans le cas de la thaumatine, on voit que ce n'est pas uniquement le taux d'occupation d'un site de fixation qui détermine l'interprétabilité de la carte de densité électronique correspondant au ligand, mais aussi la fixation du complexe dans une orientation privilégiée. Ainsi, la densité électronique des parties de ligand impliquées dans des liaisons directionnelles (liaisons H) est généralement plutôt bien définie.

## II.4.10. Protéine X

|             | Nb de sites de fixation | Nb de complexes affinés | Qcns                              | Bcns [ $\text{\AA}^2$ ] | R    | Rfree | Résolution [ $\text{\AA}$ ] |
|-------------|-------------------------|-------------------------|-----------------------------------|-------------------------|------|-------|-----------------------------|
| Gd-DTPA     | 1                       | 1                       | 0,99                              | 37                      | 0,26 | 0,30  | 2,03                        |
| Gd-DOTMA    | 1                       | 1                       | 0,65                              | 44                      | 0,21 | 0,25  | 2,17                        |
| Gd-HPDO3A   | 5                       | 4                       | 0,87;0,66;0,60;<br>0,72;0,13      | 33;36;31;<br>46;27      | 0,21 | 0,24  | 2,17                        |
| Gd-DO3A     | 6                       | 3                       | 0,83;0,66;0,63;<br>0,30;0,28;0,29 | 41;41;50;<br>44;37;41   | 0,22 | 0,25  | 2,17                        |
| Gd-DOTA-BOM | 4                       | 0                       | 0,95;0,17;0,10;<br>0,12           | 38;40;39;<br>41         | 0,23 | 0,24  | 1,40                        |
| Gd-DTPA-BMA | 2                       | 0                       | 0,34; 0,28                        | 62; 62                  | 0,24 | 0,28  | 2,98                        |

Tableau 2.4.8 : Statistiques d'affinement pour les dérivés affinés de la protéine X.

### Classement des sites de fixation

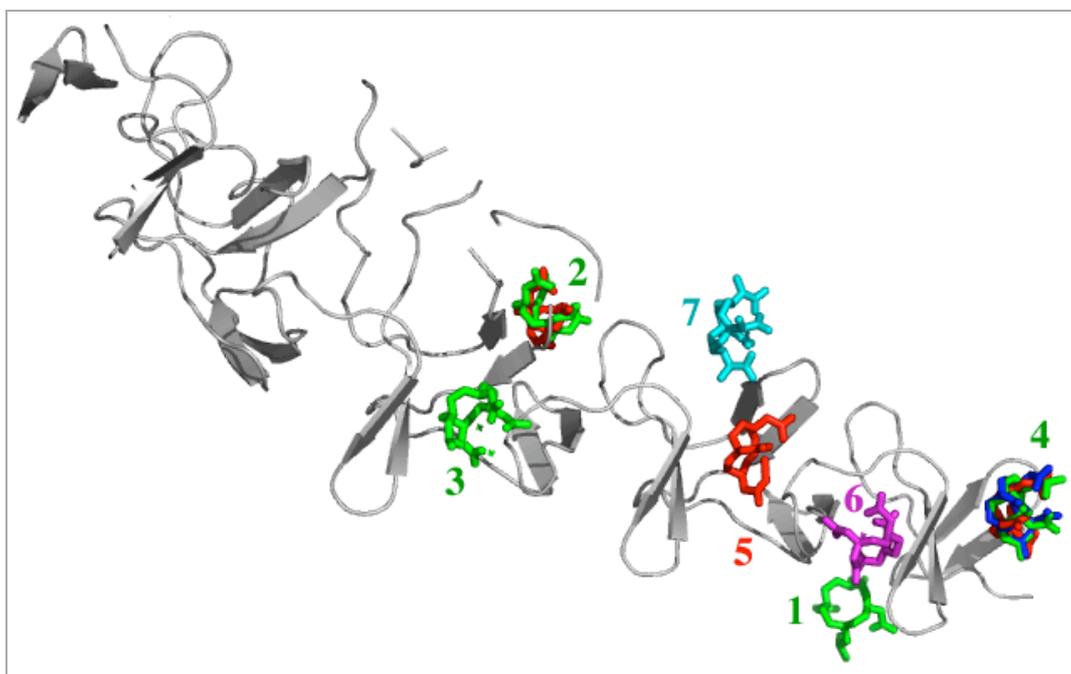


Figure 2.4.28 : Modèle affiné de la protéine X en représentation cartoon. Les sites de fixation des différents complexes sont indiqués et nommés de 1 à 7. En vert : le complexe Gd-HPDO3A, en rouge : Gd-DO3A, en magenta : Gd-DTPA, en cyan : Gd-DOTMA, en bleu : Gd-DOTA-BOM, représenté par le modèle de Gd-DOTMA.

| Site 1   |      | Site 2   |      | Site 3   |      | Site 4   |      | Site 5   |      |
|----------|------|----------|------|----------|------|----------|------|----------|------|
| complexe | Qcns |
| HPDO3A   | 0,87 | HPDO3A   | 0,66 | HPDO3A   | 0,60 | HPDO3A   | 0,72 | HPDO3A   | 0,13 |
| DO3A     | 0,30 | DO3A     | 0,66 |          |      | DO3A     | 0,63 | DO3A     | 0,83 |
|          |      |          |      | DOTA-BOM | 0,17 | DOTA-BOM | 0,95 |          |      |

Tableau 2.4.9 : Taux d'occupation des sites 1 à 5 avec les complexes Gd-HPDO3A, Gd-DO3A et Gd-DOTA-BOM (le nom des complexe est abrégé par le sigle du ligand).

### Les sites de fixation des différents complexes

La protéine X fait partie des protéines qui lient la choline (choline binding proteins - ChBP). Le motif fixant la choline (choline binding motive : cbm) est composé d'une séquence particulière d'approximativement 20 acides aminés (dont au moins deux tryptophanes) (Garcia *et al.*, 1988). Ce motif est intégré de manière répétée dans la structure de la protéine, les sites cbm étant tous localisés dans la partie C-terminale de la protéine (à droite dans la figure 2.4.28). Les sites de fixation 1 à 5 correspondent à des sites cbm. Ils lient de préférence les complexes Gd-HPDO3A et Gd-DO3A et aussi le complexe Gd-DOTA-BOM. Le taux d'occupation des différents sites est différent pour les trois complexes (voir tableau 2.4.9).

Les sites 6 et 7 ne correspondent pas à des sites cbm. C'est sur ces deux sites que se lient les complexes Gd-DTPA et Gd-DOTMA. Chacun des deux complexes n'a qu'un seul site de fixation avec cette protéine. On remarque qu'il n'y a pas de site de fixation des complexes sur la partie N-terminale de la protéine, dont la surface est chargée très négativement.

Dans la discussion qui suit, les numéros de sites correspondent à la dénomination définie plus haut (figure 2.4.28) et non pas à la valeur du taux d'occupation.

### Mode de fixation des différents complexes

#### Gd-DTPA (charge : -2)

##### Site 6 (taux d'occupation : 0,99)

Le site est localisé vers le côté C-terminal de la protéine, correspondant à une partie de surface globalement chargée positivement (figure 2.4.29.a). On remarque la coïncidence d'une zone de surface de charge positive avec la fixation du seul complexe de charge -2. La fixation se fait dans une orientation où tous les groupements carboxyles sont tournés vers la surface. Le complexe se lie avec son côté ionique tourné vers la surface. Son orientation permet l'interaction de l'ion  $Gd^{3+}$  et de la plupart des groupements carboxyliques du ligand avec les résidus de la surface protéique. La chaîne du complexe est tournée vers le solvant. L'ion  $Gd^{3+}$  est coordonné par l'atome d'oxygène OD de l'Asn266 (Gd-O : 2,4 Å) qui agit comme neuvième ligand de l'ion. Les interactions directes entre les atomes d'oxygène du ligand et des résidus protéiques sont indiquées dans la figure 2.4.29.b. Deux atomes d'oxygène d'un même bras du ligand forment des liaisons H avec les atomes d'azote amide de la chaîne principale (O(loin Gd)-N(Asn269) : 2,8 Å, O(proche Gd)-N(Gly268) : 2,9 Å). Les atomes d'oxygène de deux autres bras forment une liaison H avec l'atome d'oxygène OH de la Tyr264 (O(ligand)-OH : 2,7 Å) et un pont salin avec l'atome d'azote NZ de la Lys273 (O(ligand)-NZ : 2,6 Å). De plus, un réseau de molécules d'eau ordonnées entre le ligand et la surface renforce la fixation (figure 2.4.30.a).

Cinq molécules d'eau sont directement liées aux atomes d'oxygène du ligand et en particulier aux atomes d'oxygène des deux bras du ligand qui n'interagissent pas directement avec des résidus protéiques. La densité électronique correspondant au ligand dans ce dérivé est si bien définie qu'il a été possible de construire le modèle du complexe à partir de la densité 1Fo-Fc, calculée sans le modèle du complexe, et ainsi de résoudre la structure du complexe jusque-là inconnue.

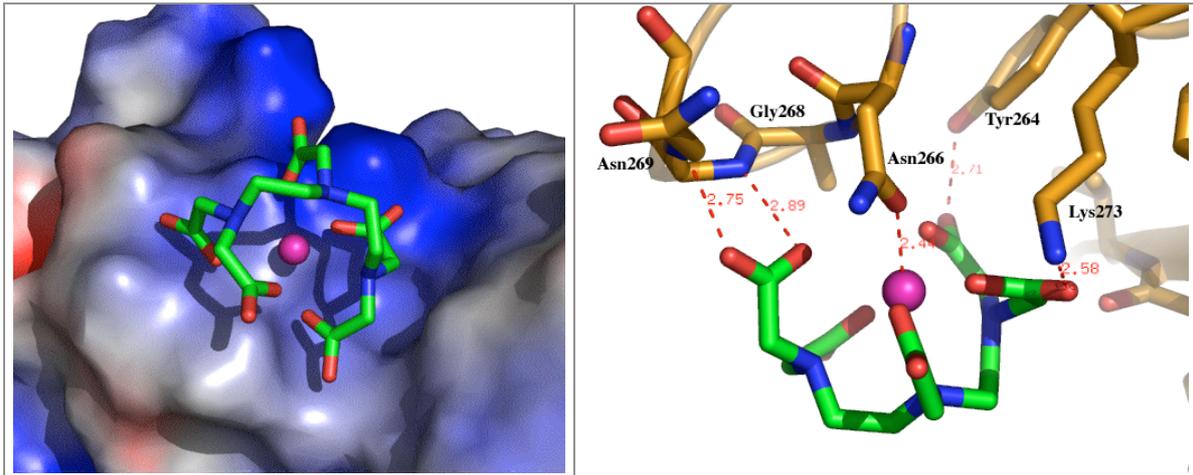


Figure 2.4.29 : Site de fixation 6 du dérivé protéine X + Gd-DTPA. a) Représentation en mode surface électrostatique. b) Interactions directes entre le complexe et la surface protéique.

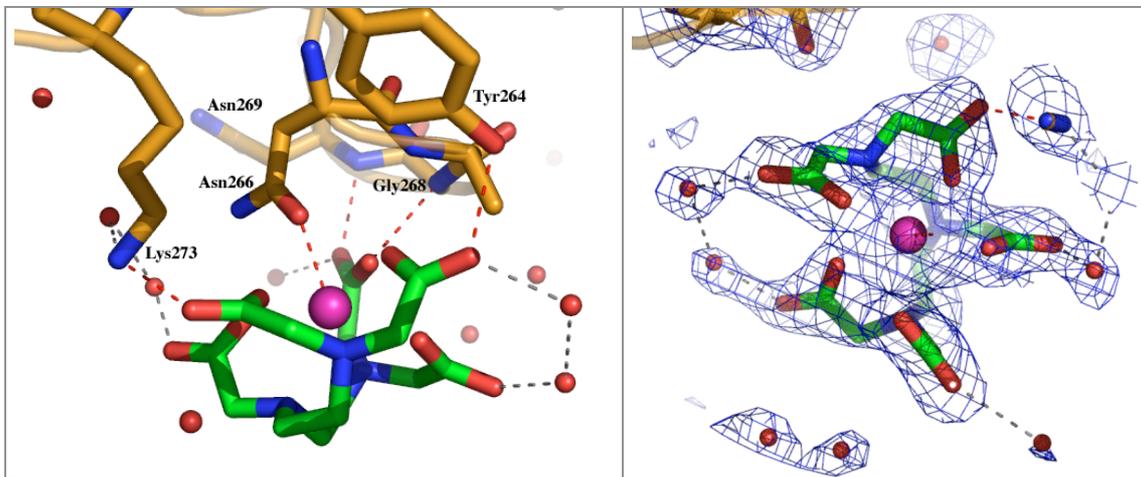


Figure 2.4.30 : Site de fixation 6 du dérivé protéine X + Gd-DTPA. a) Les molécules d'eau coordonnées par le ligand du complexe. b) La densité 2Fo-Fc du modèle affiné du complexe et des molécules d'eau est très bien définie.

### Gd-DOTMA (charge : -1)

#### Site 7 (taux d'occupation : 0,65)

Le complexe se lie par un côté formé par deux bras du ligand à une partie de surface légèrement concave et chargée positivement (figure 2.4.31.a). Sa fixation semble reposer sur le réseau de molécules d'eau ordonnées entre la surface protéique et le ligand, incluant quatre molécules d'eau coordonnées au ligand à des distances comprises entre 2 et 3 Å. Il n'y a pas de densité claire qui correspondrait à un atome agissant de neuvième ligand pour l'ion  $Gd^{3+}$ . Bien que la densité du modèle affiné ne soit pas bien définie pour la partie du ligand tournée vers le solvant, elle permet d'identifier sans ambiguïté l'orientation du complexe.

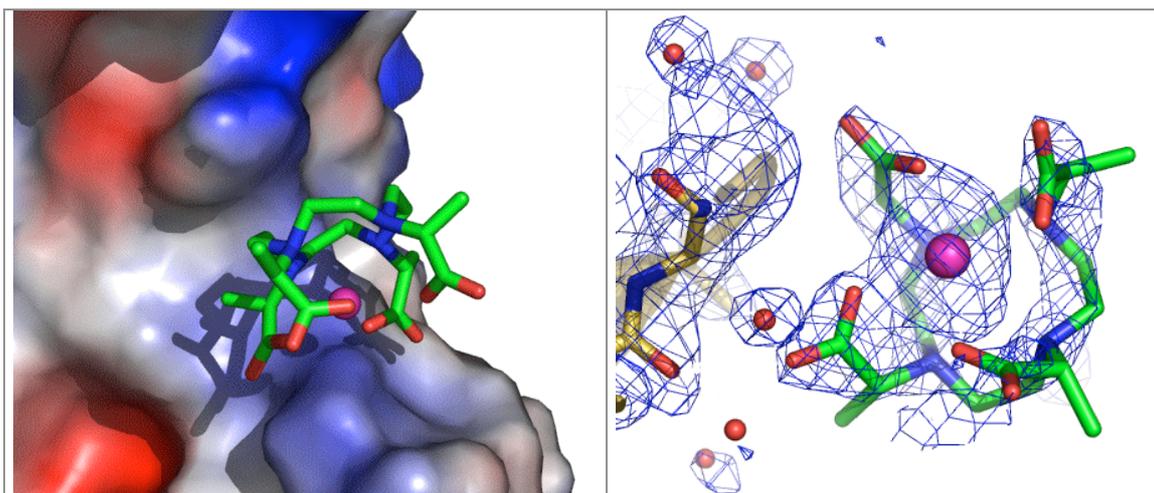


Figure 2.4.31 : Site de fixation 7 du dérivé protéine X + Gd-DOTMA. a) Représentation en mode surface électrostatique. b) La densité 2Fo-Fc du modèle affiné.

### Gd-HPDO3A (charge : 0)

Tous les sites de fixation du complexe correspondent à des cavités cbm. La charge de la surface des cavités est légèrement négative. Les cavités cbm sont formées par au moins deux tryptophanes dont les cycles aromatiques sont approximativement perpendiculaires. La fixation du complexe Gd-HPDO3A se fait par interaction hydrophobe entre le macrocycle du complexe et le cycle aromatique du tryptophane situé au fond de la cavité, les plans des deux cycles étant parallèles (distance entre plans parallèles  $\sim 3,7$  Å). Le cycle du deuxième tryptophane de la cavité est parallèle à un côté du complexe formé par deux bras du ligand. Pour les sites 1, 2 et 3, la densité électronique montre la présence d'une molécule d'eau liée à l'ion  $Gd^{3+}$ , agissant comme neuvième ligand de l'ion. Pour les sites 1 et 2, cette même molécule d'eau pont, par liaison H, l'atome d'azote ND2 d'une asparagine du bord de la cavité de fixation.

Par la suite, je décris d'éventuelles interactions supplémentaires propres aux différents sites.

#### Site 1 (taux d'occupation : 0,87) (figure 2.4.32)

La liaison du complexe implique une molécule de protéine symétrique, un atome d'oxygène du ligand formant un pont salin avec l'atome d'azote NZ de la Lys196, résidu qui appartient à la molécule symétrique (O-NZ : 2,7 Å). Un atome d'oxygène du ligand interagit avec l'atome d'azote de la même asparagine, qui est pontée par la molécule d'eau liée au  $Gd^{3+}$  (O(ligand)-OD : 2,7 Å). Un réseau de molécules d'eau entoure le ligand, dont cinq molécules directement liées au ligand. La densité 2Fo-Fc du modèle affiné est bien définie (figure 2.4.32.c).

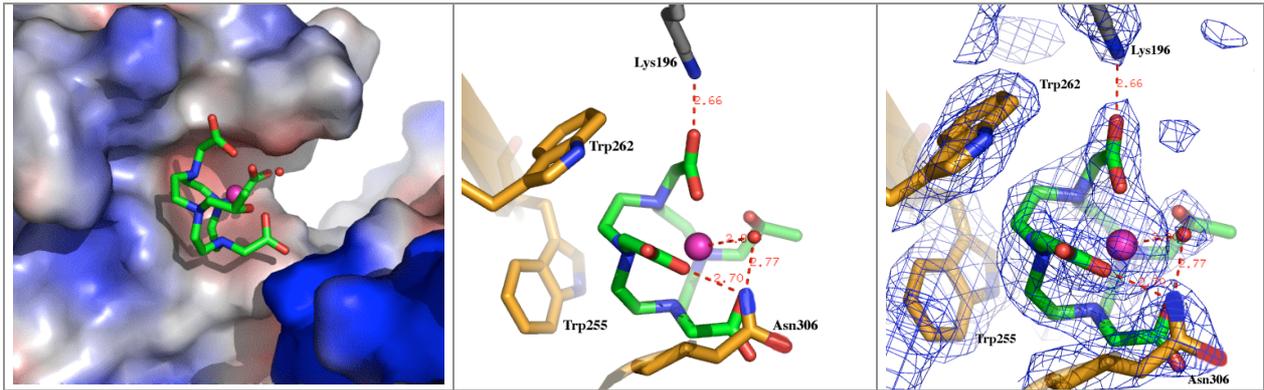


Figure 2.4.32 : Site de fixation 1 du dérivé protéine X + Gd-HPDO3A. a) Représentation en mode surface électrostatique de la cavité de fixation. En gris, la lysine 196 appartenant à une molécule de protéine symétrique. b) Interactions entre le complexe et la protéine et une molécule d'eau qui sert de neuvième ligand à l'ion  $Gd^{3+}$ . c) La densité 2Fo-Fc du modèle est bien définie.

### Site 2 (taux d'occupation : 0,66) (figure 2.4.33.a)

La densité électronique, non montrée ici, est relativement bien définie et permet d'identifier l'orientation du complexe. De manière analogue au site 1, l'asparagine impliquée dans la coordination de la molécule d'eau liée au gadolinium interagit également avec un atome d'oxygène du ligand. Du réseau de molécules d'eau entourant le complexe cinq molécules sont directement liées au ligand.

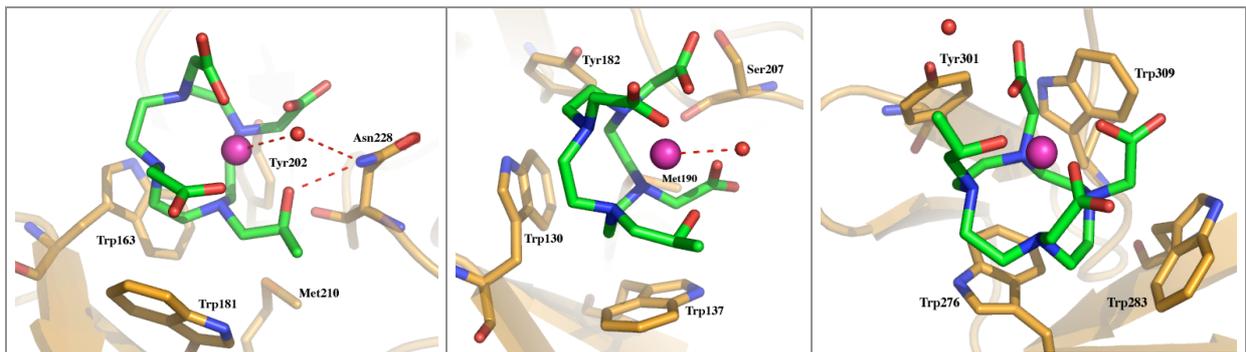


Figure 2.4.33 : Sites de fixation du dérivé protéine X + Gd-HPDO3A : a) site 2, b) site 3, c) site 4.

### Site 3 (taux d'occupation : 0,60) (figure 2.4.33.b)

Un réseau de molécules d'eau entourant le complexe inclut 3 molécules directement liées au ligand.

### Site 4 (taux d'occupation : 0,72) (figure 2.4.33.c)

Le site 4 possède trois tryptophanes perpendiculaires. Le cycle d'un des tryptophanes est parallèle au macrocycle du ligand et les cycles des deux autres sont parallèles à deux côtés du complexe respectivement formés par deux bras du ligand. Un autre côté de la cavité est formé par une tyrosine, perpendiculaire à deux des tryptophanes. L'ion  $Gd^{3+}$  pointe vers le solvant. Une molécule d'eau est directement liée au ligand.

### Gd-DO3A (charge : 0)

Tous les sites de fixation du complexe sont des cavités cbm. Dans tous les sites de fixation, le complexe Gd-DO3A est lié par interaction hydrophobe, son macrocycle étant parallèle au cycle aromatique d'un des tryptophanes formant la cavité, à une distance légèrement supérieure à 3 Å. Pour deux des complexes affinés (site 5 et site 2), le cycle du deuxième tryptophane de la cavité est approximativement parallèle au côté du complexe formé par deux

bras de ligand. Pour ces deux mêmes sites, l'ion  $Gd^{3+}$  est coordonné par l'atome d'oxygène OD de l'asparagine du bord de la cavité, qui fournit ainsi un huitième ligand à l'ion. Cette coordination engendre un mouvement du résidu vers l'ion (voir Figure 2.4.35.b). La densité électronique ne permet pas d'identifier un neuvième ligand éventuel de l'ion. Pour aucun des sites, la fixation n'implique la formation de liaisons hydrogène ou ioniques entre le ligand et des résidus protéiques. Des caractéristiques supplémentaires des différents sites de fixation, sont données ci-dessous

#### Site 5 (taux d'occupation : 0,83)

Le site le plus occupé par le complexe Gd-DO3A correspond à un site faiblement occupé (13%) du complexe Gd-HPDO3A. L'orientation du complexe Gd-HPDO3A n'a pas été affinée, mais la densité 1Fo-Fc indique que le complexe se lie dans une orientation perpendiculaire au complexe Gd-DO3A, alors que pour les autres sites leur fixation se fait dans la même orientation. Le taux de fixation plus élevé du complexe Gd-DO3A pourrait être dû à la taille plus petite de cette cavité. Quatre molécules d'eau sont liées au ligand, dont une qui ponte un atome d'oxygène amide de la chaîne principale.

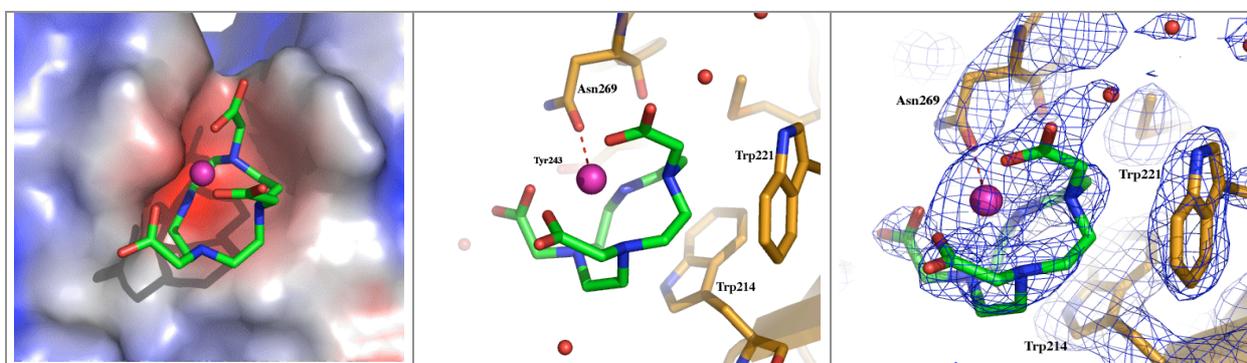


Figure 2.4.34 : Site de fixation 5 du dérivé protéine X + Gd-DO3A. a) Représentation en mode surface électrostatique. b) Coordination de l'ion  $Gd^{3+}$  par l'Asp269. c) Densité 2Fo-Fc du modèle affiné.

#### Site 2 (taux d'occupation : 0,66) (figure 2.4.35.a, b et c)

La densité n'indique aucune interaction par liaison H du ligand ni avec des résidus protéiques ni avec des molécules d'eau.

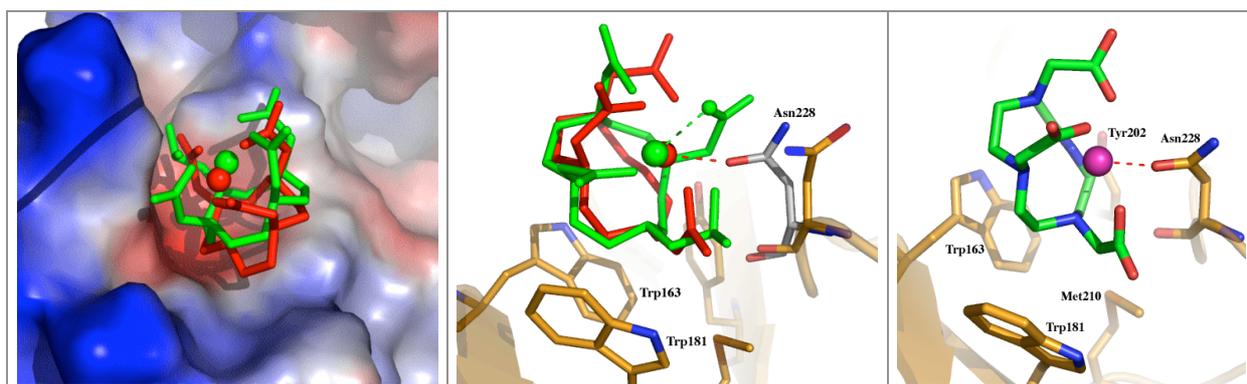


Figure 2.4.35 : Site de fixation 2 du dérivé protéine X + Gd-DO3A. a) et b) En rouge : le modèle de Gd-DO3A, en vert le modèle de Gd-HPDO3A. a) Représentation en mode surface de la cavité de fixation (la surface est calculée à partir de la conformation des résidus dans le dérivé du complexe Gd-HPDO3A). b) En gris : la conformation du résidu Asn228 dans le dérivé de Gd-DO3A, en or : sa conformation dans le dérivé de Gd-HPDO3A. Pour la fixation du complexe Gd-DO3A, le mouvement du résidu Asn228 vers l'ion  $Gd^{3+}$  permet la coordination directe de l'ion  $Gd^{3+}$ . c) Fixation du complexe Gd-DO3A.

**Site 4** (taux d'occupation : 0,63)

Cette cavité est formée par trois tryptophanes et une tyrosine dont les cycles sont perpendiculaires. Les deux tryptophanes qui ne sont pas parallèles au macrocycle du ligand se trouvent chacun face à un bras du ligand. L'ion  $Gd^{3+}$  pointe vers le solvant. La densité de ce site n'est pas très bien définie et ne permet pas de détecter d'éventuels huitième et neuvième ligand de l'ion  $Gd^{3+}$ .

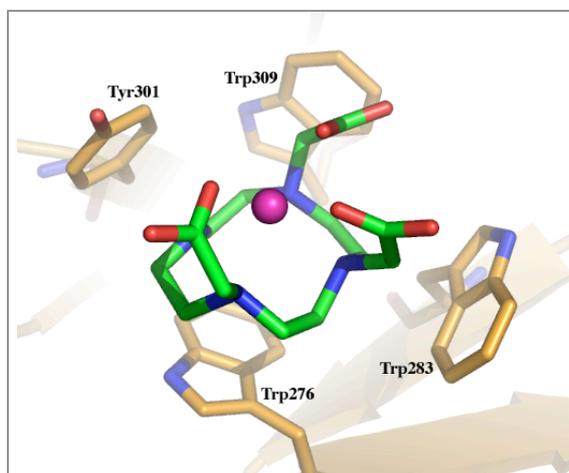


Figure 2.4.36 : Site de fixation 4 du dérivé protéine X + Gd-DO3A.

Les trois sites de fixation mineurs du complexe (taux d'occupation ~30%) se trouvent également dans des cavités cbm.

**Gd-DOTA-BOM** (charge -1)

**Site 4** (taux d'occupation : 0,95)

Le complexe se lie avec un taux d'occupation très élevé dans la cavité cbm du site 4. Cette cavité, relativement large, est formée par trois tryptophanes et une tyrosine perpendiculaires. La densité électronique expérimentale et 1Fo-Fc autour de l'ion  $Gd^{3+}$  est très forte, mais elle ne semble pas correspondre à une orientation ordonnée du complexe (figure 2.4.37.a) et ne permet pas de modéliser le complexe fixé.

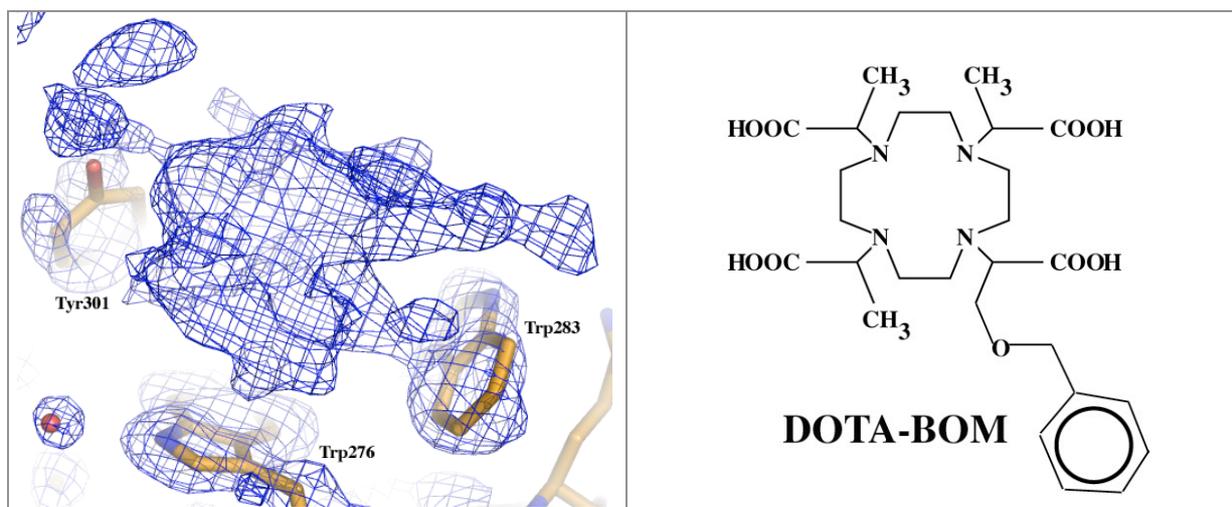


Figure 2.4.37 : a) Densité électronique expérimentale au site 4 du dérivé protéine X + Gd-DOTA-BOM.  
b) Formule du ligand DOTA-BOM.

### **Gd-DTPA-BMA (charge 0) :**

Le complexe se lie avec des taux d'occupation de l'ordre de 30% dans les cavités cbm du site 2 et du site 4.

## **II.4.11. Résultats sur la fixation des complexes**

L'affinement des structures dérivées montre que ces dernières sont assez isomorphes avec la structure native. Généralement, la fixation des complexes engendre peu de changement de la structure protéique, des paramètres de maille et de la forme cristalline. Pour certains dérivés la liaison du complexe engendre des mouvements de chaîne latérale de résidus en interaction directe avec le complexe. La fixation des complexes sur les protéines est non-covalente et de faible affinité. Généralement, elle résulte de multiples interactions faibles, impliquant le plus souvent un réseau de molécules d'eau coordonnées entre le complexe et la surface protéique et elle dépend de la forme de la surface protéique.

La fixation des complexes peut combiner différents types d'interaction dont les suivants :

**1)** L'interaction directe entre l'ion  $Gd^{3+}$  et des atomes chargés négativement des résidus de la surface protéique. Pour des complexes où l'ion  $Gd^{3+}$  est entouré par 8 atomes de coordination du ligand, l'ion interagit avec un seul atome de la surface protéique, par exemple avec l'atome d'oxygène de la chaîne latérale des résidus Asn, Gln, Asp ou Glu. Pour le Gd-DO3A fournissant 7 atomes de coordination, l'ion pourra interagir avec deux atomes chargés négativement (les deux atomes d'oxygène carboxyliques de l'Asp ou du Glu).

**1b)** L'ion  $Gd^{3+}$  coordonne une molécule d'eau. Cette même molécule d'eau pontre l'ion  $Gd^{3+}$  à un ou plusieurs résidus protéiques.

**2)** Des ponts salins ou des liaisons H directes entre les atomes d'oxygène des bras des ligands (les atomes d'oxygène et d'azote pour Gd-DTPA-BMA) et des atomes d'oxygène ou d'azote de la surface protéique.

**3)** L'interaction hydrophobe entre le macrocycle (ou "pseudo-macrocycle") et des résidus aromatiques de la surface protéique.

**4)** L'établissement d'un réseau de liaisons H à travers la coordination de molécules d'eau par les atomes d'oxygène (et d'azote) des bras des ligands. Ces molécules d'eau établissent éventuellement des ponts vers la surface protéique.

**5)** L'interaction entre deux complexes menant à la fixation coopérative de deux molécules de complexe voisines (lysozyme).

Le tableau 2.4.9 résume quelle sorte d'interaction décrite ci-dessus intervient lors de la fixation des complexes pour les différents dérivés.

| Gd-HPDO3A (0) |            | Gd-DO3A (0) |            |      |      |       |
|---------------|------------|-------------|------------|------|------|-------|
| lysozyme      | protéine X | lysozyme    | protéine X | YGGV | yeaZ | GI    |
| 3 4 5         | 3 1b 2 4   | 1 3 4 5     | 1 3 4      | 1    | 1 2  | 1 2 4 |
| 4 5           | 3 1b 2 4   | 1 3 4 5     | 1 3        |      |      |       |
|               | 3 4        |             | 3          |      |      |       |
|               | 3 4        |             |            |      |      |       |

| Gd-DOTMA (-1) |       |      |      | Gd-DTPA-BMA (0) |            |
|---------------|-------|------|------|-----------------|------------|
| protéine X    | urate | Yggv | yeaZ | urate           | thaumatine |
| 4             | 3 2 4 | 2 4  | 1 4  | 1 2 4           | 3          |
|               |       |      |      | 1 2 4           |            |
|               |       |      |      | 1b 2 4          |            |

| Gd-DTPA (-2) |        |            | Gd-DOTA (-1) |            | Gd-DOTA-BOM (-1) |            |
|--------------|--------|------------|--------------|------------|------------------|------------|
| protéine X   | GI     | thaumatine | lysozyme     | thaumatine | lysozyme         | thaumatine |
| 1 2 4        | 1b 2 4 | 3 2 4      | 3 1 2 5      | 3 2        | 3 2 4            | 3          |
|              |        |            | 3 5          |            |                  | 2 4        |

Tableau 2.4.9 : Interactions intervenant pour la fixation des complexes avec les différentes protéines (entre parenthèses la charge totale du complexe). Plusieurs lignes correspondent à différents sites de fixation.  
Abréviations des noms de protéine : GI : glucose isomérase, urate : urate oxydase.

Les complexes se lient souvent dans des cavités plus ou moins prononcées, formées par la surface de la protéine (thaumatine) ou par l'empilement cristallin (lysozyme), ce qui indique l'importance de facteurs stériques pour la fixation. La représentation des surfaces électrostatiques permet de constater que les complexes chargés négativement ont plus tendance à se lier sur des parties de surface chargées positivement que les complexes neutres. La représentation de la surface montre que le complexe se lie de préférence selon une orientation qui accorde sa forme à la forme de la surface, typiquement avec les bras carboxyliques ou méthylamides qui se nichent dans des creux de la surface. On remarque que deux complexes de structure très proche peuvent se lier selon des modes de fixation tout à fait différents (exemples : Gd-DOTA-BOM - Gd-DOTMA dans les protéines lysozyme, urate oxydase et protéine X, Gd-HPDO3A - Gd-DOTA dans le lysozyme et la thaumatine). La possibilité de liaison par interaction hydrophobe entre un complexe particulier et le cycle aromatique de tryptophanes ou de phénylalanines semble dépendre, outre de la nature du résidu aromatique, de l'environnement du résidu aromatique ou peut-être des conditions physico-chimiques. Ainsi dans le lysozyme les complexes Gd-HPDO3A et Gd-DO3A interagissent avec le(s) tryptophane(s), mais non pas le Gd-DOTMA, alors que c'est l'inverse pour l'urate oxydase. Les complexes qui interagissent avec la phénylalanine dans les dérivés de thaumatine sont d'autres complexes, dont les deux complexes avec un ligand linéaire qui, eux, ne semblent pas interagir avec les tryptophanes.

#### II.4.11.a. Quelques caractéristiques récurrentes de la fixation des différents complexes

##### Gd-DO3A

Le mode de fixation principal du complexe correspond à la coordination de l'ion  $Gd^{3+}$  par les deux atomes d'oxygène carboxyliques d'un aspartate ou glutamate de la surface protéique. Un mode de fixation alternatif est l'interaction hydrophobe du macrocycle avec le cycle aromatique de tryptophanes.

## **Gd-HPDO3A et Gd-DOTA-BOM**

Les deux complexes ont tendance à établir des liaisons hydrophobes avec les tryptophanes.

Les complexes Gd-DTPA-BMA, Gd-DTPA, Gd-DOTA, Gd-DOTMA et Gd-DOTA-BOM interagissent de différentes manières avec les protéines.

Le mode d'interaction peut, en partie, expliquer l'influence de la présence du complexe sur la solubilité de la protéine. L'interaction hydrophobe pourrait solubiliser la protéine, car, dans ces conditions, les complexes couvrent des parties hydrophobes de la protéine. Inversement, la fixation d'un complexe à des résidus ioniques de la surface par son côté ionique et tournant son côté apolaire vers le solvant pourrait diminuer la solubilité de la protéine.

## **Conclusions**

L'utilisation du complexe Gd-DO3A mène souvent à des dérivés avec des sites de fixation multiples et des taux de fixation moyens, mais suffisants pour un bon phasage. L'utilisation du complexe Gd-HPDO3A est indiquée pour des protéines avec un taux élevé de tryptophanes dans la séquence, pourvu que ces derniers soient supposés situés en surface.

La forte fixation des complexes Gd-HPDO3A et Gd-DO3A sur le motif structural cbm en fait des candidats de premier choix pour préparer des dérivés avec des protéines qui lient la choline (ChBP). Il serait envisageable d'introduire des sites cbm dans d'autres protéines par mutagenèse dirigée. Pourvu qu'une telle introduction n'agisse trop fortement sur le repliement, la solubilité et la cristallisation de la protéine les sites cbm pourraient alors agir comme sites de fixation des complexes.

## **II.5. Conclusions**

L'étude cristallographique de la fixation des complexes dans les cristaux de protéines a été menée avec un grand nombre de dérivés. Pour toutes les protéines testées, il a été possible d'obtenir des cristaux dérivés avec un fort pouvoir de phasage avec au moins un des complexes testés.

Pendant ma thèse, l'utilisation des complexes a par ailleurs permis la résolution de la structure de quatre nouvelles protéines.

Les résultats de l'étude confirment le potentiel des complexes pour se fixer fortement dans les cristaux et permettre ainsi l'obtention de dérivés à fort pouvoir de phasage. Ils confirment la facilité d'utilisation des complexes et leur applicabilité générale dans de nombreuses conditions testées.

Pour une combinaison de protéine et de complexe quelconque, le taux de succès pour obtenir un dérivé performant est de l'ordre de 50%.

Pour préparer un dérivé anomal avec une nouvelle protéine, en choisissant le complexe à utiliser de manière ciblée, tenant compte des caractéristiques de la protéine, et éventuellement en testant deux ou trois complexes différents, il est très probable d'obtenir un bon dérivé.

Les résultats des affinements du mode de fixation des différents complexes permettent de mieux orienter le choix du complexe à utiliser en fonction des propriétés de la protéine. Néanmoins, sauf pour des cas exceptionnels, comme celui des protéines ChBP, il n'est pas possible de prévoir avec certitude la fixation d'un complexe sur une protéine. Cela est dû au fait que la fixation des complexes résulte généralement de nombreuses interactions faibles, impliquant, entre autres, le réseau de molécules d'eau ordonnées à la surface de la protéine en combinaison avec des conditions stériques favorables à la liaison du complexe.

Il serait donc d'un grand intérêt de trouver un méthode relativement simple et facile à mettre en œuvre comme alternative à la cristallographie pour détecter la fixation ou non d'un complexe avec une protéine.

## **Partie III.**

# **Méthodes physico-chimiques pour détecter la fixation d'un complexe sur une protéine**



### **III.0.Méthodes physico-chimiques pour caractériser la fixation des complexes sur des protéines - Introduction**

Les résultats des affinements du mode de fixation des complexes sur les protéines ne permettent pas, dans le cas général, de prévoir la fixation ou non d'un complexe sur une nouvelle protéine. C'est pourquoi nous avons cherché une méthode physico-chimique, alternative à la cristallographie, qui permettrait de caractériser l'interaction des complexes avec des protéines, en solution ou dans le cristal. On pourrait ainsi déterminer parmi les complexes le meilleur candidat pour se fixer sur la protéine avant d'éventuelles expériences cristallographiques. Le but est donc de trouver une méthode relativement simple et facile à mettre en œuvre pour déterminer la constante de dissociation caractérisant l'interaction du complexe avec la protéine. L'ordre de grandeur attendu pour cette constante est déterminé à partir des taux d'occupations observés lors de études cristallographiques (voir chapitre I.5.4). Elle est de l'ordre de 100 mM.

Boggon & Shapiro (2000) et Lott *et al.* (2003) recommandent d'utiliser l'électrophorèse de gels de polyacrylamide natifs pour cribler des potentiels dérivés d'atomes lourds. L'analyse de la migration du mélange protéine avec le composé lourd permettrait de détecter d'éventuelles incompatibilités qui provoqueraient la dénaturation de la protéine et seraient révélées sur le gel. En outre, la fixation d'un composant chargé sur la protéine pourrait modifier sa migration, ce qui permettrait de détecter la fixation du complexe lourd sur la protéine. Ils recommandent également l'utilisation de la spectrométrie de masse pour détecter l'éventuelle fixation du composant sur la protéine. Des expériences de spectrométrie de masse en utilisant des dérivés de protéine obtenus avec nos complexes ont montré qu'il était possible de détecter l'association non covalente des complexes avec différentes protéines. En revanche, il semble difficile d'établir par cette méthode la spécificité de l'interaction et de comparer les différents complexes entre eux (Eric Girard, communication personnelle).

Le choix de la méthode de détection de la fixation du composant va dépendre des propriétés physico-chimiques de l'atome lourd utilisé. Pour la recherche d'une telle méthode, nous nous sommes appuyés sur les propriétés physico-chimiques particulières des lanthanides qui permettraient leur détection.

La forte fluorescence des lanthanides dans leur seuil d'absorption  $L_{III}$  les rend facilement détectables par spectroscopie des rayons X. Cependant, cette méthode ne semble pas bien adaptée pour détecter la fixation d'un de nos complexes sur une protéine. Comme nous travaillons à des concentrations de complexe élevée, il y a une grande quantité de molécules de complexe non-liées dans les canaux de solvant et il serait difficile de distinguer ces dernières des molécules liées. En effet, la fluorescence des lanthanides ne varie que peu en fonction de l'environnement. D'éventuels changements de l'environnement de l'ion lors de la fixation ne modifieraient pas le signal détecté.

### **III.0.1. Difficultés liées aux caractéristiques des complexes utilisés**

Pour détecter la fixation de nos complexes, il faut tenir compte des particularités des complexes et de leur interaction avec les protéines.

Premièrement, la faible affinité des complexes pour les protéines fait qu'ils doivent être utilisés avec une concentration élevée. Il semble difficile de distinguer les molécules non-liées, présentes en grande quantité dans les canaux de solvant du cristal des molécules liées à la protéine dans des sites spécifiques. Un trempage dans une solution ne contenant pas de complexe de gadolinium afin d'enlever le complexe des canaux de solvant (back-soak) risque de provoquer la dissociation des sites de fixation des molécules de complexe liées.

Pour des méthodes de détection qui se servent des ions de lanthanides comme sondes pour détecter la fixation du complexe, un problème est que la fixation des complexes à la protéine n'entraîne pas de forts changements de l'environnement de l'ion, car la sphère de coordination de l'ion est presque saturée par les atomes du ligand.

### **III.0.2. Considérations expérimentales**

La plupart des méthodes ont été testées en utilisant le lysozyme de blanc d'œuf de poule comme protéine modèle, car elle est disponible en grandes quantités et pas chère.

Comme nous ne connaissions pas l'influence des conditions physico-chimiques sur la fixation des complexes sur les protéines et afin de pouvoir comparer les résultats obtenus par différentes méthodes avec les résultats obtenus par cristallographie, nous avons travaillé dans des conditions physico-chimiques proches des conditions de cristallisation, si possible.

## III.1. Calorimétrie de titrage isotherme

### III.1.1. Introduction

Une expérience de calorimétrie de titrage isotherme (ITC : Isothermal Titration Calorimetry) permet l'étude de l'interaction d'un ligand avec une macromolécule biologique. Elle permet de déterminer l'enthalpie de la liaison du ligand avec la macromolécule ( $\Delta H_b$ ), ainsi que la constante d'association ( $K_a$ ) et la stœchiométrie de la réaction ( $n$ ).

Dans une expérience de titrage ITC, le "ligand", contenu dans une seringue, est titré dans une solution de la "macromolécule", contenue dans une cellule (ou *vice versa*). L'expérience est réalisée à température constante. Après chaque ajout d'une petite quantité de ligand, la chaleur libérée ou absorbée est mesurée par rapport à une cellule de référence remplie de solution tampon. La différence de chaleur est exprimée comme la puissance électrique requise pour maintenir une température constante entre la cellule contenant l'échantillon et la cellule de référence, qui sont toutes deux placées dans un récipient adiabatique. Quand la macromolécule est saturée par le ligand, le signal de chaleur diminue et il ne reste plus que la chaleur de dilution.

La figure 3.1.1 montre le schéma du montage d'une expérience ITC.

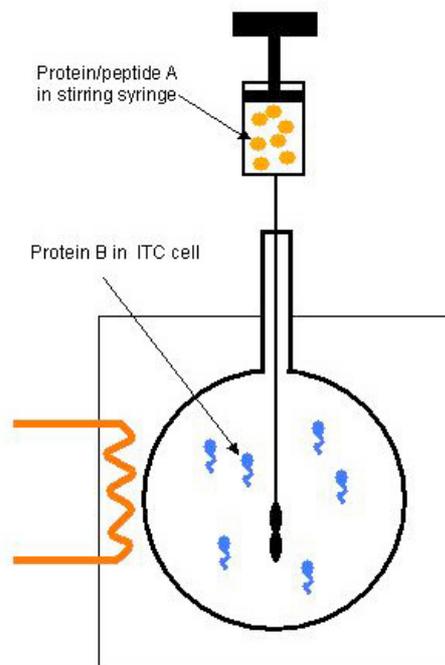


Figure 3.1.1 : Diagramme d'une cellule ITC et seringue. La cellule contenant l'échantillon et la cellule de référence (qui n'est pas montrée ici) sont placées dans un récipient adiabatique, dont la température est contrôlée par le calorimètre (McMahon, 2005).

Le domaine de constantes d'association directement mesurables est de  $10^2$  à  $10^9$   $M^{-1}$  (Pierce *et al.*, 1999). Les constantes d'affinité typiques, que nous avons déterminées pour nos complexes à partir des expériences de cristallographie sont de l'ordre de  $10$   $M^{-1}$ .

La figure 3.1.2 montre des courbes de mesures typiques pour une réaction endothermique. La pente au point d'inflexion de la courbe de chaleur intégrée augmente avec la constante d'association.

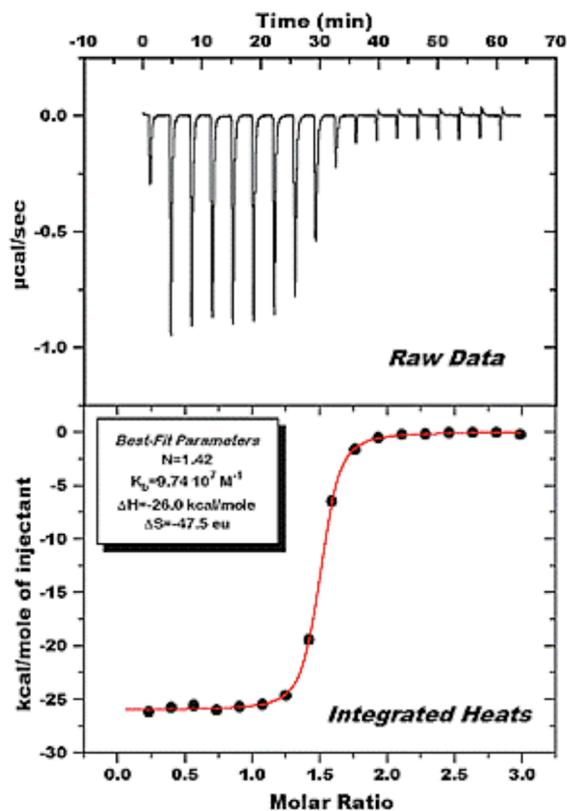


Figure 3.1.2 : Courbes de mesure typiques pour une réaction endothermique. La figure du haut montre la chaleur absorbée à chaque ajout de ligand. La figure du bas montre la chaleur absorbée intégrée par mole de ligand ajouté (<http://www.microcalorimetry.com>).

La figure 3.1.3 montre les isothermes de liaison simulées pour des valeurs différentes du paramètre  $c$ , qui caractérise l'isotherme expérimentale, avec :  $c = K_a[M]n$ , où  $[M]$  est la concentration de la macromolécule dans la cellule et où  $n$  caractérise la stœchiométrie  $n : 1$  de l'association entre le ligand et la macromolécule.

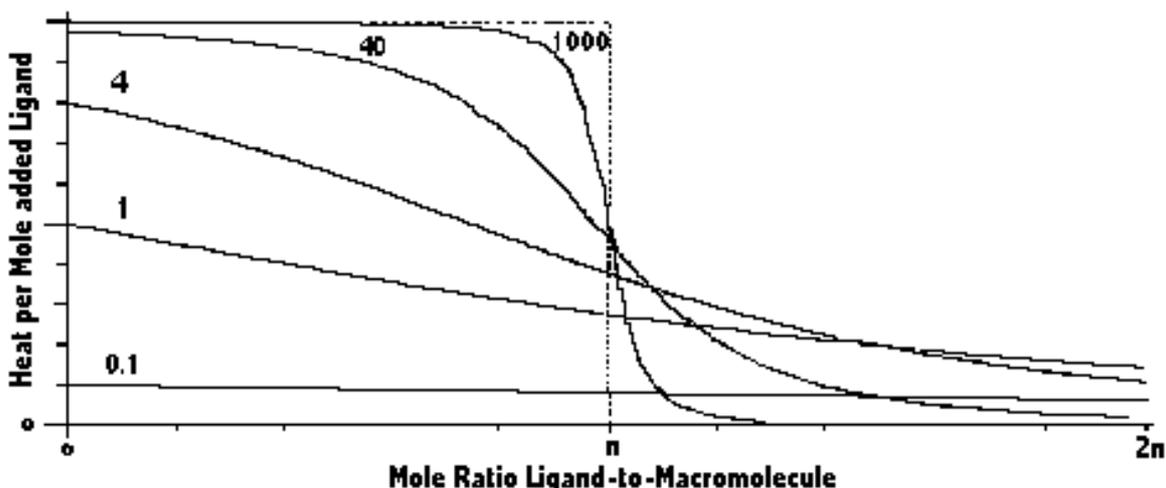


Figure 3.1.3 : Isothermes de liaison simulées pour différentes valeurs du paramètre "c" (Wiseman *et al.* 1989).

Pour que la courbe expérimentale puisse être interprétée, il est nécessaire que la valeur de  $c$  soit comprise entre 1 et 1000. Pour des faibles valeurs de  $c$ , la forme caractéristique sigmoïdale du graphe est perdue : la transition devient très large, s'approchant d'une relation

linéaire. Le point d'équivalence, correspondant au rapport molaire  $n$  entre ligand et protéine, ne peut pas être déterminé.

Dans nos conditions expérimentales, avec une constante d'association de  $10 \text{ M}^{-1}$  et une concentration de protéine de  $1,5 \text{ mM}$ , nous obtenons une valeur approximative de  $c = K_a[M]n = 10 \text{ M}^{-1} * 1,5 * 10^{-3} \text{ M} * 2 = 0,03$ .

### III.1.1.a. Détermination des paramètres thermodynamiques

Pour un système avec une stœchiométrie 1 :1, la chaleur absorbée ou libérée lors de l'ajout du ligand est de

$$Q = \frac{V_0 \Delta H_b [M]_t K_a [L]}{1 + K_a [L]} \quad (\text{I})$$

Avec :  $V_0$  : volume de la cellule,  $[M]_t$  : concentration de la macromolécule totale (incluant molécules libres et liées),  $[L]$  : concentration de ligand libre.

Avant l'intégration des pics des données brutes, les chaleurs de liaison sont normalisées en fonction de la concentration du ligand et corrigées de l'effet de dilution.

En exprimant le relation (I) en fonction de la concentration totale de ligand  $[L]$ , les valeurs de  $K_a$  et  $\Delta H_b$  sont déterminées en ajustant les paramètres à déterminer de façon à faire concorder la courbe calculée à la courbe expérimentale.

Avant l'expérience, titrant et macromolécule devraient être dialysés, de préférence dans un même récipient, afin de minimiser les artefacts dus à des différences de tampon.

### III.1.2. Étude de l'association d'une solution de lysozyme avec le complexe Gd-HPDO3A par calorimétrie de titrage isotherme

Des expériences calorimétriques de titrage isotherme ont été réalisées en collaboration avec Pierre le Parlouër (Thermal Consulting) avec la protéine lysozyme de blanc d'œuf de poule et le complexe Gd-HPDO3A.

#### III.1.2.a. Appareillage

Les expérimentations ont été réalisées avec le calorimètre de titrage isotherme ITC CSC 4200, à l'entreprise Sétaram à Caluire. Le calorimètre assure les fonctions suivantes dans une gamme de température de  $0$  à  $110 \text{ °C}$  : programmation et régulation de la température du calorimètre, acquisition et numérisation des signaux de température et de chaleur.

### III.1.3. Conditions expérimentales

Les conditions expérimentales ont été choisies de manière à être proche des conditions de cristallisation (tampon, pH, concentration du complexe de gadolinium). Il était cependant nécessaire de travailler avec une concentration de NaCl plus basse que pour la cristallisation, afin d'éviter la formation de cristaux de lysozyme lors de l'expérience de titrage. La concentration de la protéine doit être élevée pour obtenir une valeur élevée du coefficient  $c$ .

Nous avons choisi de réaliser les deux expérimentations suivantes :

### Expérience 1

Pour la première expérience, le titrant est une solution très concentrée de la protéine, est c'est la solution de complexe qui est placée dans la cellule (à l'inverse des conditions habituelles, où c'est la protéine qui est placée dans la cellule). Les deux agents sont préparés dans le même tampon.

|  |   |
|--|---|
| Solution de titrage<br>250 $\mu\text{L}$     | 200 g/L lysozyme<br>50 mM acétate de sodium pH = 4,6                |
| Solution de la cellule<br>1300 $\mu\text{L}$ | 110 mM Gd-HPDO3A<br>0,55 M NaCl<br>50 mM acétate de sodium pH = 4,6 |

### Expérience 2

Pour la deuxième expérience, le titrant est une solution concentrée du complexe Gd-HPDO3A. La cellule contient la solution protéique.

|  |  |
|--|--|
| Solution de titrage<br>250 $\mu\text{L}$     | 1,0 M Gd-HPDO3A<br>0,5 M NaCl<br>50 mM acétate de sodium pH = 4,6  |
| Solution de la cellule<br>1300 $\mu\text{L}$ | 22 g/L lysozyme<br>0,55 M NaCl<br>50 mM acétate de sodium pH = 4,6 |

Le calorimètre régule la température à 20°C.

Pour les deux expériences, la cellule de référence, servant à mesurer la chaleur de dilution du titrant lors des injections, est remplie d'une solution contenant 0,55 M de NaCl et tamponnée avec 50 mM d'acétate de sodium à pH 4,6.

Pour chacune des deux expériences, 16 injections de 10  $\mu\text{L}$  sont réalisées. L'intervalle de temps séparant 2 injections successives est de 5 minutes.

### III.1.4. Résultats

Le tableau 3.1.1 résume les valeurs de chaleur mesurées à chaque injection de ligand pour l'expérience 1 et 2. Les courbes correspondantes sont montrées dans les figures 3.1.4 et 3.1.5.

| Expérience 1          |                    |  | Expérience 2          |                    |
|-----------------------|--------------------|--|-----------------------|--------------------|
| Chaleur<br>( $\mu$ J) | Vinj<br>( $\mu$ L) |  | Chaleur<br>( $\mu$ J) | Vinj<br>( $\mu$ L) |
| 44,01693              | 10                 |  | 32863,33              | 10                 |
| -287,579              | 10                 |  | 22479,15              | 10                 |
| -362,934              | 10                 |  | 22008,67              | 10                 |
| -419,445              | 10                 |  | 20797,67              | 10                 |
| -432,448              | 10                 |  | 19612,49              | 10                 |
| -469,076              | 10                 |  | 18678,75              | 10                 |
| -489,097              | 10                 |  | 17771,14              | 10                 |
| -470,921              | 10                 |  | 17165,28              | 10                 |
| -501,129              | 10                 |  | 16536,61              | 10                 |
| -530,293              | 10                 |  | 15940,74              | 10                 |
| -544,488              | 10                 |  | 15348,6               | 10                 |
| -609,12               | 10                 |  | 14749,4               | 10                 |
| -633,114              | 10                 |  | 14164,71              | 10                 |
| -649,918              | 10                 |  | 13509,32              | 10                 |
| -698,847              | 10                 |  | 12991,81              | 10                 |
| -701,745              | 10                 |  | 12132,76              | 10                 |

Tableau 3.1.1 : Valeurs de chaleur mesurées pour l'expérience 1 et 2.

## Expérience 1

La figure 3.1.4 montre la différence de chaleur pour les 16 injections de la solution protéique dans la cellule contenant le complexe. Le premier pic est endothermique et les 15 suivants sont exothermiques. La chaleur de dilution de la protéine dans le tampon a été soustraite.

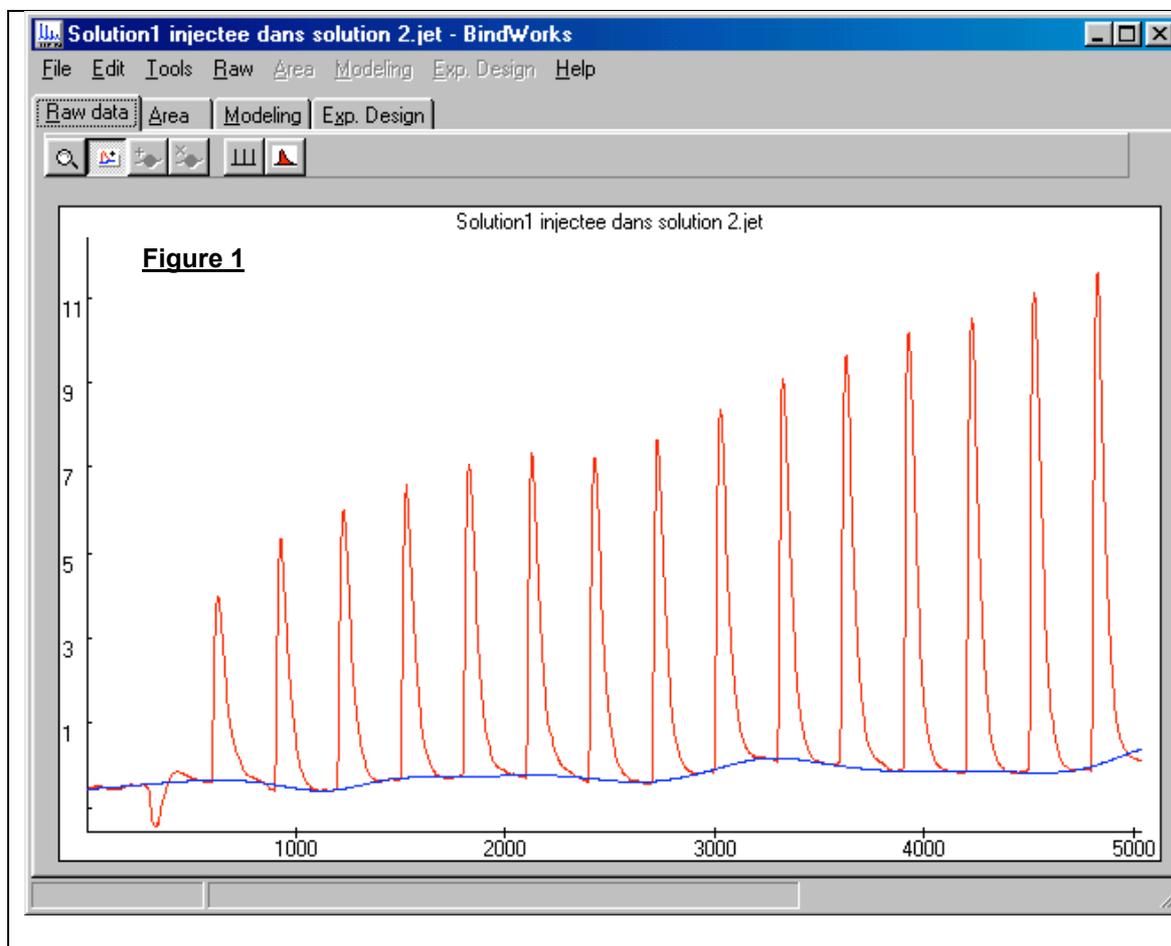


Figure 3.1.4 : Chaleur dissipée (absorbée) en fonction du temps. Chaque pic correspond à l'injection de 10  $\mu\text{L}$  de solution protéique dans la cellule contenant le complexe. L'unité en abscisses est en secondes, l'unité en ordonnées est arbitraire.

## Expérience 2

La figure 3.1.5 montre la différence de chaleur pour les 16 injections de la solution de complexe dans la cellule contenant la protéine. Tous les pics sont endothermiques. La chaleur de dilution du complexe dans le tampon a été soustraite.

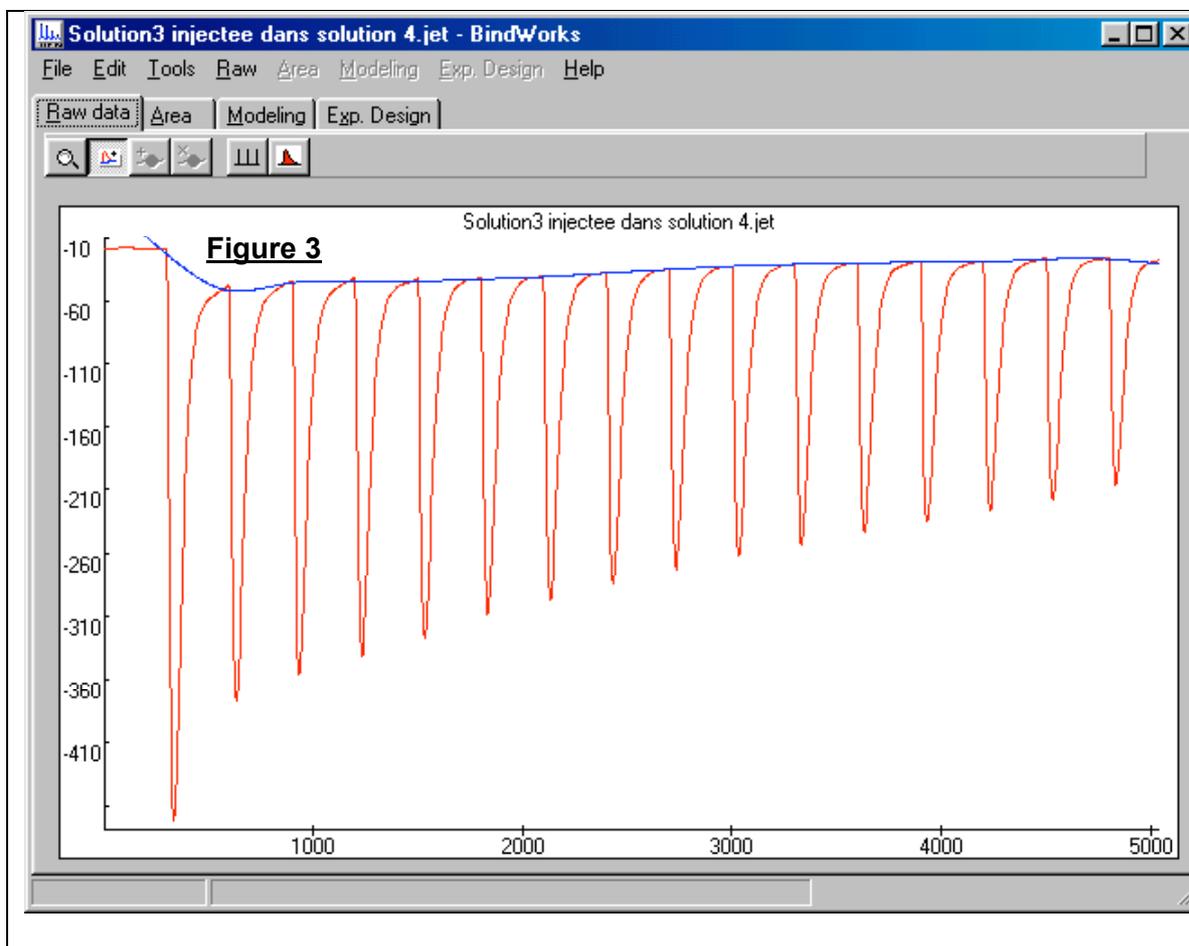


Figure 3.1.5 : Chaleur absorbée en fonction du temps. Chaque pic correspond à l'injection de 10  $\mu\text{L}$  de solution de complexe dans la cellule contenant la protéine. L'unité en abscisses est en secondes, l'unité en ordonnées est arbitraire.

L'allure des courbes obtenues (figure 3.1.4 et 3.1.5) ne correspond pas à l'allure des courbes qui est habituellement observée, et qui tend vers une valeur constante quand le point d'équivalence (rapport molaire [titrant / agent dans la cellule] =  $n$ ) est atteint.

Un calcul des rapports molaires atteints dans nos conditions en montre la raison.

Expérience 1 : le rapport molaire protéine / ligand atteint lors de la dernière injection est de  $2,24 \mu\text{M}/143 \mu\text{M} = 0,02$ .

Expérience 2 : le rapport molaire ligand / protéine atteint lors de la première injection est de  $160 \mu\text{M}/2 \mu\text{M} = 5$ .

Pour l'expérience 1, le point d'équivalence ( $n$  : rapport molaire protéine / ligand = 0,5) n'est jamais atteint, alors que pour l'expérience 2, il est dépassé dès la première injection ( $n$  : rapport molaire ligand / protéine = 2).

Pour obtenir de meilleures conditions expérimentales, nous sommes limités d'un côté par la solubilité de la protéine (expérience 1) et de l'autre côté, en travaillant avec une concentration de complexe plus faible, ou en injectant des volumes plus faibles (expérience 2), il est

probable que les différences de chaleur dues à l'association spécifique des deux molécules deviennent négligeables devant les effets parasites de l'expérience.

Le résultat obtenu n'est pas surprenant, puisque, pour l'expérience 2, la valeur  $c = K_a [M]n = 10M^{-1} * 1,5 * 10^{-3} M * 2 = 0,03$  est significativement plus faible que les valeurs recommandées.

En outre, nous avons négligé de dialyser les différentes solutions dans un même tampon, ce qui peut engendrer des chaleurs parasites dues à d'éventuelles différences de pH et de concentration.

### **III.1.5. Conclusion**

L'affinité du complexe Gd-HPDO3A, qui se fixe avec des populations élevées sur la protéine lysozyme lors de la cristallisation, est trop faible pour pouvoir caractériser la fixation par des expériences de calorimétrie. Cette méthode ne semble donc pas permettre de détecter la fixation d'un de nos complexes sur une protéine en solution.

## **III.2. Expériences de résonance plasmonique de surface pour caractériser l'interaction de la protéine lysozyme avec le complexe Gd-HPDO3A**

La résonance plasmonique de surface (RPS) est une méthode de caractérisation de la liaison d'un "ligand" sur un "récepteur" adsorbé à la surface d'une couche métallique. Le système détecte la faible variation de l'indice de réfraction de l'interface quand le ligand se fixe aux récepteurs. Le signal mesuré est le changement de l'angle de résonance.

Nous avons réalisé des expériences de RPS (Biacore) avec la protéine lysozyme et le complexe Gd-HPDO3A. Les courbes obtenues n'ont pas indiqué d'interaction spécifique entre la protéine et le complexe. Il est probable que la constante de dissociation de l'ensemble protéine et complexe de gadolinium soit trop faible pour que l'interaction puisse être caractérisée par RPS. En effet, le domaine de constantes de dissociation typique pour des expériences de RPS est compris entre  $K_d = 10 \text{ mM}$  et  $K_d = 1 \text{ pM}$  (McDonnell, 2001).

### III.3.Électrophorèse de gels d'acrylamide natifs

L'électrophorèse de gels natifs permet la séparation des protéines en fonction de leur charge et de leur taille hydrodynamique.

Nous aimerions voir si la fixation des complexes sur une protéine modifie la vitesse de migration de la protéine. Pour des complexes chargés, leur charge ajoutée à la charge de la protéine pourrait modifier la vitesse de migration de l'ensemble constitué de la protéine et du complexe. Pour une migration en sens direct, de l'anode vers la cathode, la fixation d'un complexe chargé négativement devrait augmenter la vitesse de migration ; la migration en sens inverse devrait être freinée par la fixation. Que les complexes soient chargés ou non, la fixation du complexe pourrait freiner la migration en augmentant la taille de la molécule. La masse approximative des complexes est de 650 g/mol ; la protéine la plus petite que nous avons étudiée, le lysozyme, a une masse de 14 kDa. Pour deux sites de fixation par molécule de protéine, la différence relative de masse entre une molécule avec et sans complexe fixé serait de  $\Delta m/m \sim 10\%$ . Si le complexe se fixe dans une cavité de la protéine, l'influence sur la taille hydrodynamique pourra être moindre.

Nous nous sommes posé la question suivante : pour être proche des conditions de cristallisation, quelle concentration de complexe faut-il ajouter à la protéine ? Dans la solution déposée dans les puits du gel, la concentration de la protéine est d'environ 1 mg/mL, alors que dans la solution de cristallisation, la concentration protéique est de 40 mg/mL (lysozyme). Faut-il choisir une concentration de complexe proche de la concentration utilisée pour la cocristallisation ( $C_{\text{complexe}} = 100 \text{ mM}$ ) ou faut-il garder un rapport molaire protéine / complexe similaire à celui des conditions de cristallisation ( $C_{\text{complexe}} = 2,5 \text{ mM}$ ) ? Dans un premier temps, nous avons choisi de travailler avec une concentration de complexe élevée, pour voir si l'ajout du complexe pouvait modifier la vitesse de migration.

#### III.3.1. Préparation des gels

Habituellement on prépare le gel natif pour la migration de la protéine avec un tampon et un pH choisi en fonction du point isoélectrique (pI) de la protéine. La concentration d'acrylamide est choisie en fonction de la taille de la protéine. Cette concentration détermine la taille du maillage du gel polymérisé. On choisit cette concentration plus faible dans la partie supérieure du gel ("stacking"), qui sert à concentrer la protéine, que dans la partie "running" qui sert à la séparation des bandes. Habituellement, le gel est préparé à un pH supérieur au pI de la protéine. Dans le gel, la protéine sera alors chargée négativement. Le sens habituel de la migration se fait alors de haut en bas, de l'anode, chargée négativement, vers la cathode, chargée positivement. La tension appliquée entre les deux électrodes, et le courant résultant, sont choisis de sorte que la chaleur dissipée dans le gel n'échauffe pas trop l'échantillon (valeurs typiques : 100 V et 30 mA). Pour la composition des gels, je me suis appuyé sur l'article de McLellan (1982).

### **III.3.2. Préparation du gel pour la migration du lysozyme de blanc d'œuf de poule**

#### **Tampon "running"**

25 mM histidine

30 mM tampon MOPS [acide 3-(N-Morpholino)propanesulfonique] équilibré à pH 6,6 avec NaOH

#### **Composition du gel**

Le gel a été préparé avec le tampon "running" à pH 6,6 et 7,5% d'acrylamide. Je n'ai pas ajouté de partie "stacking".

Pour un volume de 30 mL, le gel se prépare en mélangeant :

21,6 mL tampon "running"

7,5 mL acrylamide (30%)

900 µL APS (1%) (persulfate d'ammonium,  $(\text{NH}_4)_2\text{S}_2\text{O}_8$ )

35 mL TEMED (N,N,N',N'-Tetramethylethylenediamine, catalyseur pour la polymérisation des gels de polyacrylamide)

#### **Préparation des échantillons de protéine à déposer et dépôt des échantillons**

Dans chaque puits, j'ai déposé 7 µL de solution contenant 5 ou 10 µg de protéine, 100 mM de complexe de gadolinium et 2 µL de la solution décrite ci-dessous :

#### **Solution de dépôt**

20% tampon "running"

80% glycérol

une petite quantité de bleu de bromophénol

#### **Tampon de migration**

25 mM de tampon tris pH 8,3

92 mM glycine

Le gel, placé dans le tampon de migration, est relié au générateur de courant. La migration a été réalisée à température ambiante pendant ~2 h à une tension de 100 V, dans des appareils d'électrophorèse de Amersham Biosciences "mighty small II" reliés à un générateur de courant EPS1.

Le pI de la protéine (pI du lysozyme = 11) est supérieur au pH du gel : la protéine sera chargée positivement dans le gel. Pour faire migrer la protéine dans le gel, il faut donc inverser le sens des électrodes par rapport au sens habituel.

Le gel est coloré avec le bleu de Coomassie.

### **III.3.3. Résultats obtenus avec le lysozyme de blanc d'œuf de poule**

La figure 3.3.1 montre le gel d'électrophorèse natif obtenu avec le lysozyme de blanc d'œuf de poule et différents complexes de gadolinium.

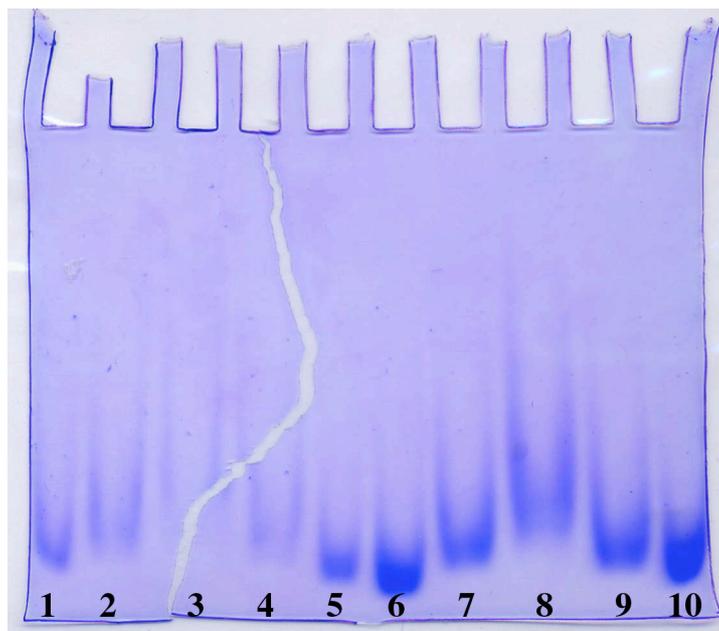


Figure 3.3.1 : Gel d'électrophorèse natif où 7  $\mu$ L de solution contenant 5 ou 10  $\mu$ g de lysozyme et un des complexes de gadolinium à une concentration de 100 mM ont été déposés dans les puits.

Le tableau 3.3.1 indique les complexes correspondant au numéro de puits. La charge des complexes est indiquée entre parenthèses.

|                     | Sans additif | Gd-DOTA-BOM (-1) | Gd-DOTA (-1) | Gd-DOTMA (-1) | Gd-HPDO3A (0) |
|---------------------|--------------|------------------|--------------|---------------|---------------|
| Lysozyme 5 $\mu$ g  | 1            | 2                | 3            | 4             | 5             |
| Lysozyme 10 $\mu$ g | 6            | 7                | 8            | 9             | 10            |

Tableau 3.3.1 : Quantité de protéine déposée dans les puits du gel de la figure 3.3.1 et complexe ajouté.

La figure 3.3.1 montre que la migration de la protéine a été ralentie pour tous les puits contenant des complexes chargés. Parmi les complexes chargés, ce sont les complexes Gd-DOTA-BOM et Gd-DOTA qui se fixent dans les cristaux de lysozyme, le premier avec un site de fixation unique, le deuxième avec deux sites de fixation (chapitre II.4.8). Le complexe Gd-DOTMA ne se fixe pas de manière significative dans les cristaux de lysozyme. Alors que le fort ralentissement de la migration de la protéine dans les puits 3 et 8 est en accord avec la fixation du complexe chargé Gd-DOTA sur deux sites de fixation, les deux complexes Gd-DOTA-BOM et Gd-DOTMA ralentissent la migration de manière comparable, alors que ce n'est que le premier qui se fixe significativement dans les cristaux de la protéine. Pour le complexe non chargé, Gd-HPDO3A, il ne semble pas y avoir de ralentissement significatif de la migration (puits 5 et 10), alors que, dans le cristal, le complexe se fixe fortement.

### III.3.4. Essais d'électrophorèse de gels natifs avec d'autres protéines

Des essais d'électrophorèse de gel natif ont été réalisés avec les protéines urate oxydase d'*Aspergillus flavus* et glucose isomérase de *Streptomyces rubiginosus*. Les conditions de préparation des gels sont résumées dans le tableau 3.3.2. Cependant, aucun des gels n'a jamais permis de discerner de différence de vitesse de migration dû à l'ajout des complexes.

|                   | Urate oxydase |        | Glucose isomérase |        |
|-------------------|---------------|--------|-------------------|--------|
| pI                | 7,3           |        | 5                 |        |
| pH du gel         | 6,8           | 8,8    | 6,8               | 8,8    |
| sens de migration | inverse       | direct | direct            | direct |

Tableau 3.3.2 : Conditions de préparation de gels d'acrylamide natifs pour la migration des protéines urate oxydase et glucose isomérase.

La figure 3.3.2 montre l'exemple d'un gel d'acrylamide avec la protéine glucose isomérase. Le gel a été préparé avec un tampon TRIS pH 6,8, avec 3,75 % d'acrylamide dans la partie "stacking" et 10% d'acrylamide dans la partie "running". Dans chaque puits, j'ai déposé 30 µg de protéine glucose isomérase avec 100 mM de complexe de  $Gd^{3+}$ , sauf dans le puits 9, dans lequel j'ai déposé la protéine mélangée à 10 mM d' $EuCl_3$ . Les puits 2 à 8 et 10 correspondent aux huit différents complexes. La migration a été faite en sens direct.

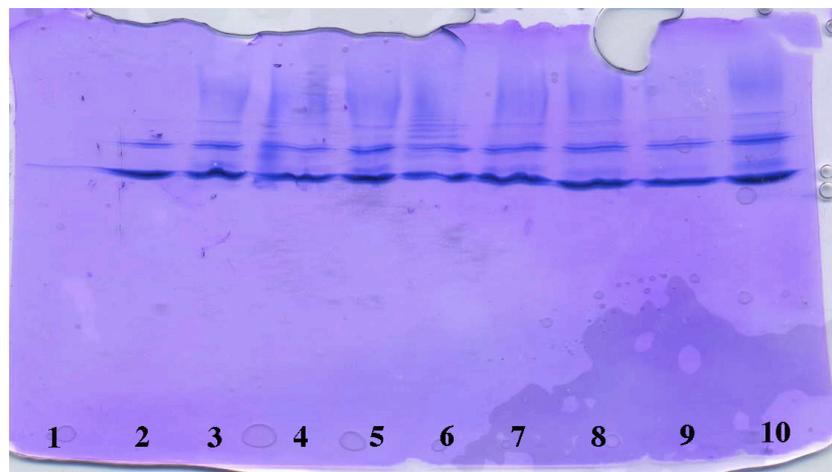


Figure 3.3.2 : Gel d'acrylamide de la protéine glucose isomérase et différents complexes de gadolinium.

Le gel correspond au cas le plus souvent observé avec les gels que j'ai préparés : il n'y a pas de différence de la vitesse de migration significative entre les différents puits. Même l'ajout de l'ion  $Eu^{3+}$ , qui se fixe fortement à la glucose isomérase dans deux sites de fixation et qui devrait provoquer une différence de charge de +6, ne ralentit pas de manière significative la migration.

### III.3.5. Conclusions

Les résultats du gel de lysozyme montrent que, pour les conditions utilisées ici, la migration est ralentie par l'ajout des complexes chargés, et même par l'ajout de ceux qui ne se fixent pas dans le cristal. Il reste à déterminer, par des expériences supplémentaires, si un rapport peut être trouvé entre le taux de fixation du complexe dans le cristal et la migration, peut-être en travaillant avec des concentrations de complexe plus basses. Pour le lysozyme, la fixation du complexe non chargé ne modifie pas de manière suffisante la taille hydrodynamique de la protéine pour provoquer un ralentissement de la fixation. Ceci veut dire que la méthode n'est probablement pas bien adaptée à la détection de la fixation des complexes non chargés (qui représentent la moitié des complexes étudiés).

Il reste cependant à optimiser les conditions de préparation de gel et de migration pour les gels avec d'autres protéines, afin de pouvoir observer une différence de migration quand on ajoute un complexe chargé. Les conditions à modifier seraient la concentration d'acrylamide

qui détermine la vitesse de migration, le temps de migration, le pH et la concentration de complexe utilisée.

## **III.4. Expériences de résonance magnétique nucléaire**

### **III.4.1. Introduction**

Le noyau de certains atomes possède un moment magnétique (spin) non nul. Dans un champ magnétique externe ( $B_0$ ) l'alignement du spin des noyaux avec le champ correspond à des niveaux d'énergie discrets. Pour faire passer le spin d'un noyau d'un niveau d'énergie à un autre niveau, d'énergie supérieure, il faut appliquer un rayonnement électromagnétique dont la fréquence correspond exactement à la différence d'énergie entre les deux niveaux. Les fréquences typiques sont de l'ordre du MHz (fréquence radio). Cette fréquence de résonance dépend de l'intensité du champ magnétique  $B_0$  et de l'environnement chimique du noyau. Un spectre de résonance magnétique nucléaire (RMN) unidimensionnel est obtenu en balayant un domaine de fréquences radio et en mesurant l'absorption pour toutes les fréquences. Une fréquence d'absorption particulière correspond à un type de noyau magnétique avec un environnement chimique particulier. Dans les protéines sans marquage isotopique, seuls les protons ont un spin non nul. La fréquence de résonance des différents types de protons est mesurée par rapport à la fréquence de résonance d'un composé de référence, la différence relative par rapport à cette référence est appelée le déplacement chimique du proton et ne dépend pas du champ magnétique appliqué. Le temps de relaxation d'un proton est la durée pendant laquelle le spin reste dans l'état excité. La largeur de la bande d'absorption correspondante est inversement proportionnelle à la durée de vie de l'état excité.

### **III.4.2. Détection de la fixation d'un complexe sur une protéine**

En théorie, il y a deux possibilités de détecter la fixation du complexe sur la protéine :

a) Une première possibilité consiste à utiliser comme sonde les molécules d'eau liées à l'ion  $Gd^{3+}$  dans le complexe en solution, molécules qui peuvent être remplacées par des résidus protéiques lors de la fixation à la protéine. En effet, le voisinage du centre paramagnétique  $Gd^{3+}$  provoque une forte baisse du taux de relaxation des protons des molécules d'eau (Aime *et al.*, 2000; Huskens *et al.*, 1995). C'est cet effet qui est utilisé pour l'imagerie médicale. Le problème, si nous utilisons cette méthode, serait que, dû à la faible affinité des complexes pour les protéines, nous devrions travailler avec un large excès de complexe non lié, dont le signal couvrirait alors le signal des molécules de complexe liées à la protéine.

b) La deuxième possibilité consiste à utiliser les protons de la protéine proches du site de fixation du complexe comme sonde. En effet, le centre paramagnétique  $Gd^{3+}$  provoque une forte baisse du temps de relaxation des protons situés à proximité.

Avec un complexe dans lequel l'ion  $Gd^{3+}$  est remplacé par un autre ion de lanthanide non paramagnétique (par exemple le lutétium), la fixation du complexe entraînerait un changement du déplacement chimique des protons de la protéine à proximité du site de fixation. Dans les deux cas, le problème est que, pour avoir une occupation suffisamment

forte des sites de fixation sur la protéine, il faut travailler avec un large excès de complexe. Comme l'effet relaxant du  $Gd^{3+}$  est très fort, en travaillant avec une concentration élevée de complexe de  $Gd^{3+}$ , on risque d'avoir un élargissement des pics de l'ensemble des protons de la protéine, dû à la présence du complexe dans le solvant.

### **III.4.3. Expériences de RMN 1D avec le lysozyme de blanc d'œuf de poule et les complexes Gd-HPDO3A et Lu-HPDO3A**

Nous avons choisi d'étudier si la deuxième possibilité (méthode b) permettrait de détecter la fixation d'un complexe sur une protéine. Comme, en outre, nous disposions du complexe Lu-HPDO3A, nous avons décidé d'utiliser les complexes Gd-HPDO3A et Lu-HPDO3A, en combinaison avec la protéine lysozyme de blanc d'œuf de poule. En effet, le choix du lysozyme comme protéine test nous a paru préférable à plusieurs égards :

- nous pouvions facilement disposer d'une grande quantité de la protéine pure, qui n'est pas chère,
- nous avons étudié la fixation des complexes avec la protéine : les complexes de lanthanides Ln-HPDO3A se fixent fortement dans les cristaux de la protéine,
- la protéine a également été étudiée en RMN et son spectre est attribué (Redfield & Dobson, 1988).

Le complexe Ln-HPDO3A se lie au lysozyme à proximité de trois tryptophanes (voir chapitre II.4.8.). Dans le spectre RMN 1D  $^1H$  du lysozyme, les protons liés aux atomes d'azote des noyaux indole de cinq des six tryptophanes donnent lieu à des raies uniques (Laurents & Baldwin, 1997). Nous voulions donc tester si l'ajout du complexe Ln-HPDO3A (avec Ln = Gd ou Lu) aurait une influence sur la position, l'intensité et la largeur des raies correspondant aux trois tryptophanes à proximité des sites de fixation. Des expériences 1D  $^1H$  présentent l'avantage de pouvoir se faire dans un temps expérimental très court (acquisition du spectre en quelques secondes) et d'être facile à interpréter. Elles ne nécessitent pas de marquage isotopique de la protéine.

### **III.4.4. Préparation des échantillons**

Pour être proche des conditions de cristallisation, nous avons travaillé avec le même tampon et au même pH que pour la cristallisation. Nous avons travaillé avec la protéine à la concentration  $C_p = 2$  mM et avec les deux différents lanthanides Gd et Lu en faisant varier la concentration de complexe de 2 mM à 100 mM. En supposant une constante de dissociation de 100 mM pour la fixation du complexe sur la protéine, ceci donnerait des taux d'occupation du site de fixation de la protéine de 1,9%, de 8,9% et de 49,8% pour des concentrations respectives de complexe de 2, 10 et 100 mM. Sauf pour la concentration de complexe la plus élevée, le signal de la protéine qu'on aimerait détecter, celui qui résulte de la forme liée au complexe, correspond à la forme largement minoritaire.

Toutes les solutions contiennent 2 mM de lysozyme, 50 mM de tampon acétate de sodium pH 4,6 et 10% de  $D_2O$ . Le volume d'échantillon nécessaire est 500  $\mu L$ .

Les solutions ont été préparées en ajoutant à la solution décrite ci-dessus les agents suivants :

- a) blanc : solution de protéine seule
- b) Gd-HPDO3A 2 mM, c) 10 mM, d) 100 mM
- e) Lu-HPDO3A 2 mM, f) 10 mM, g) 100 mM

### III.4.4.a. Spécifications techniques

Les expériences ont été réalisées au Laboratoire de RMN (LRMN) de l'IBS, en collaboration avec Rémy Sounier qui a enregistré pour nous les spectres 1D et 2D.

Tous les spectres ont été enregistrés avec le spectromètre 600 MHz du laboratoire, avec un temps d'acquisition d'une minute et à la température de 25°C.

## III.4.5. Spectres 1D $^1\text{H}$

### III.4.5.a. Protéine native

Le spectre RMN 1D  $^1\text{H}$  d'une solution native de lysozyme a tout d'abord été enregistré (figures 3.4.1 et 3.4.2).

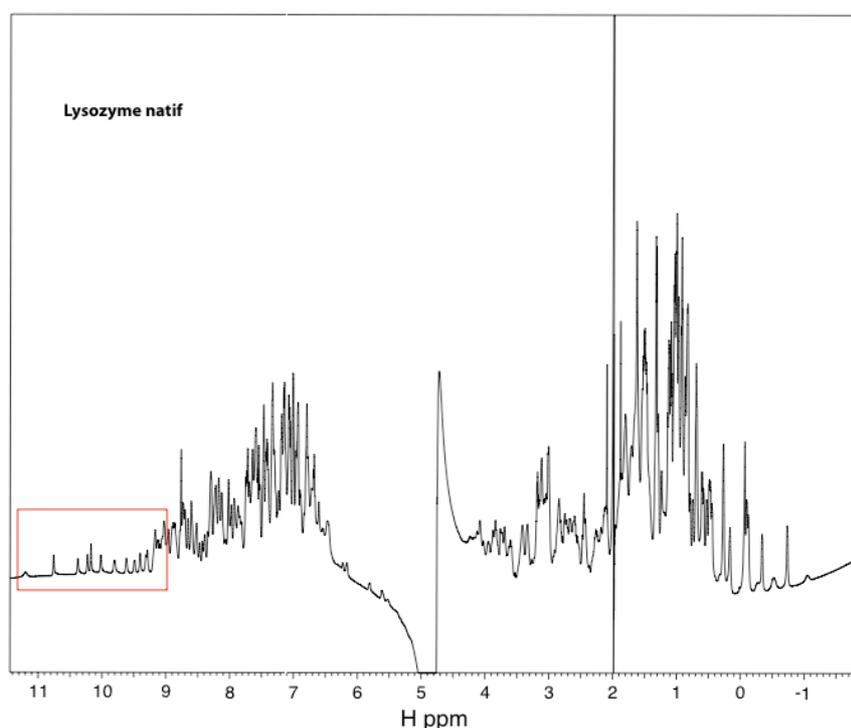


Figure 3.4.1 : Spectre RMN 1D  $^1\text{H}$  d'une solution de lysozyme sans ajout de complexe. En rouge : le domaine où apparaissent les pics des tryptophanes. Le pic du  $\text{H}_2\text{O}$  apparaît à 4,7 ppm. Le large pic à 2 ppm correspond au tampon.

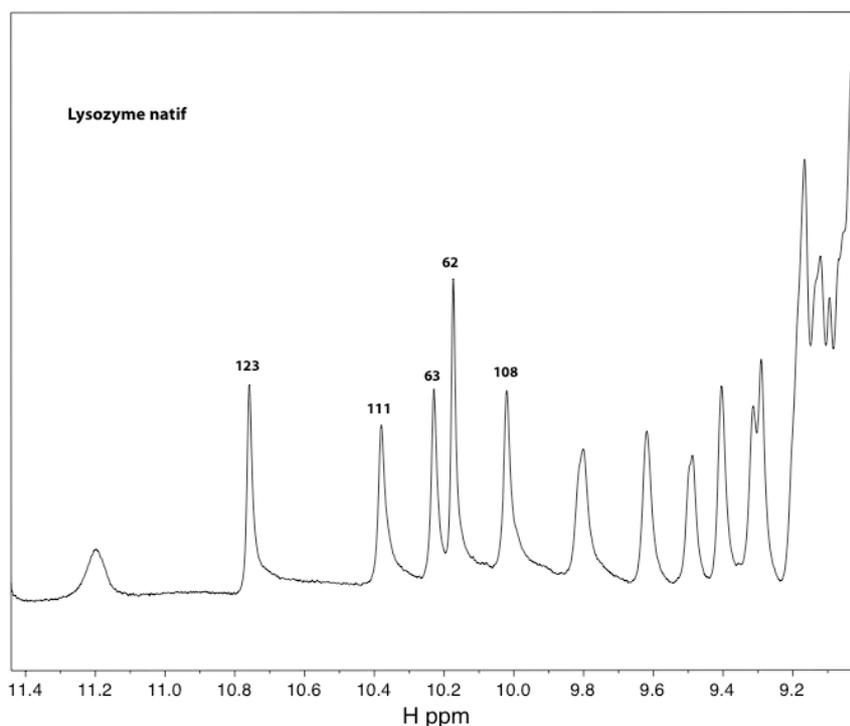


Figure 3.4.2 : Spectre RMN 1D  $^1\text{H}$  d'une solution de lysozyme sans ajout de complexe. Aggrandissement du domaine des tryptophanes. Les numéros correspondent aux numéros des tryptophanes dans la séquence du lysozyme.

La figure 3.4.1 montre les pics de résonance des protons de la solution de lysozyme. La figure 3.4.2 correspond à l'agrandissement du domaine où apparaissent les pics des protons liés aux atomes d'azote des noyaux indole de cinq tryptophanes du lysozyme. L'attribution des pics est celle de Laurents & Baldwin (1997). Les positions des pics sont les suivantes : Trp123 : 10,755 ppm ; Trp111 : 10,38 ppm ; Trp63 : 10,225 ppm ; Trp62 : 10,16 ppm ; Trp108 : 10,015 ppm.

### III.4.6. Protéine avec le complexe Gd-HPDO3A

Les figures 3.4.3 et 3.4.4 montrent les pics des protons des tryptophanes pour des solutions supplémentées avec, respectivement, 2 mM et 10 mM de Gd-HPDO3A.

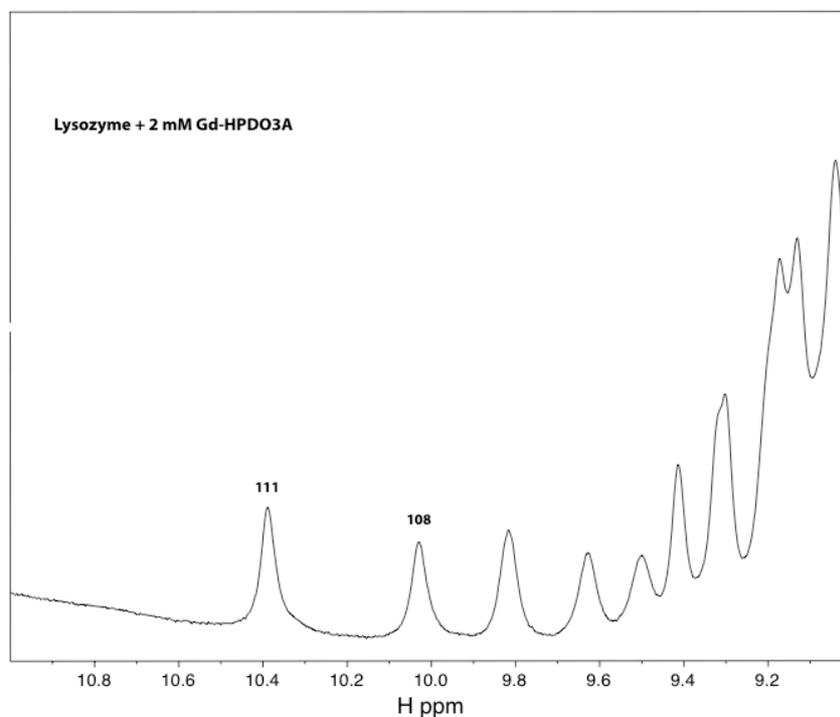


Figure 3.4.3 : Spectre RMN 1D  $^1\text{H}$  d'une solution de lysozyme avec 2 mM de Gd-HPDO3A. Agrandissement du domaine des tryptophanes.

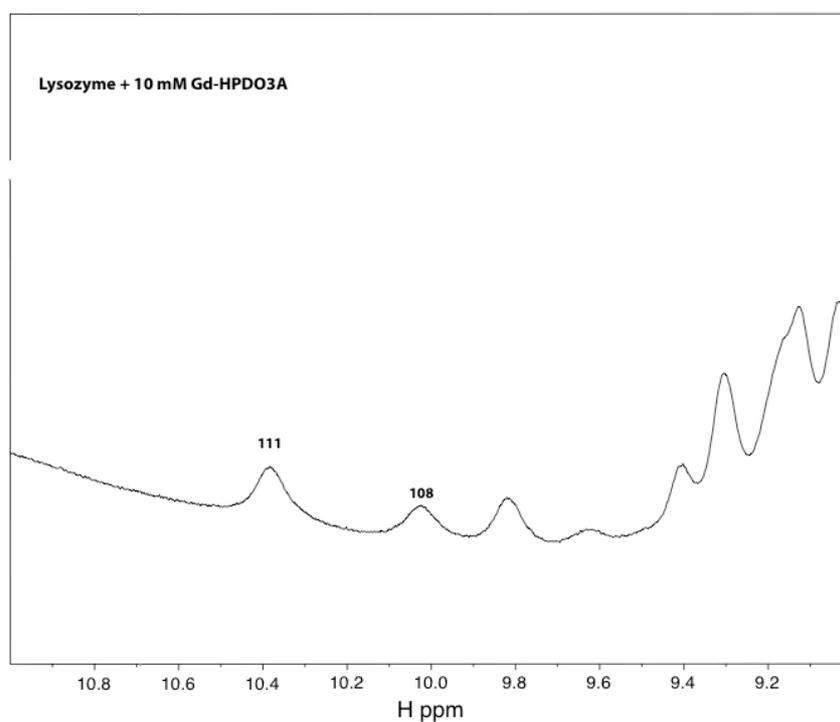


Figure 3.4.4 : Spectre RMN 1D  $^1\text{H}$  d'une solution de lysozyme avec 10 mM de Gd-HPDO3A. Agrandissement du domaine des tryptophanes.

Déjà dans la figure 3.4.3, il ne reste plus que les pics des tryptophanes 111 et 108, et les pics encore présents se sont beaucoup affaiblis et se sont élargis. Cet effet est dû à la chute du temps de relaxation provoqué par le  $Gd^{3+}$ . Dans la figure 3.4.4, correspondant à une concentration de complexe de 10 mM, l'affaiblissement des pics est encore plus fort. C'est pourquoi nous n'avons pas mesuré de spectre correspondant à 100 mM de complexe, car, déjà à des concentrations plus faibles, la présence du  $Gd^{3+}$  entraîne l'affaiblissement de tous les pics de la protéine. C'est ce que montre la comparaison du spectre de la protéine native (figure 3.4.1) au spectre complet de la protéine avec 10 mM de Gd-HPDO3A (figure 3.4.5).

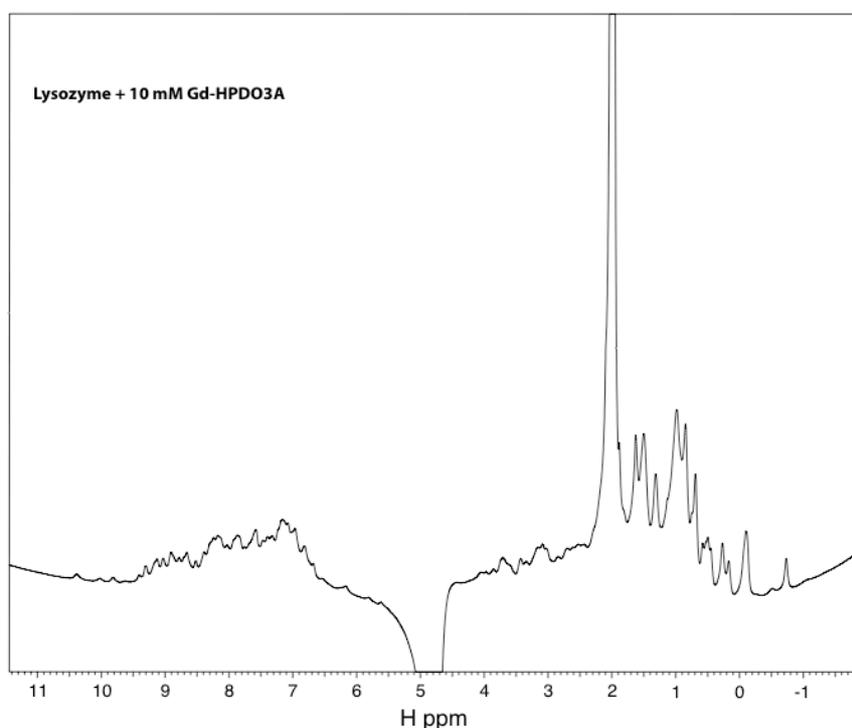


Figure 3.4.5 : Spectre RMN 1D  $^1H$  d'une solution de lysozyme avec 10 mM de Gd-HPDO3A.

### III.4.7. Protéine avec le complexe Lu-HPDO3A

Les figures 3.4.6 à 3.4.8 correspondent au spectres enregistrés en ajoutant le complexe Lu-HPDO3A à la solution de lysozyme.

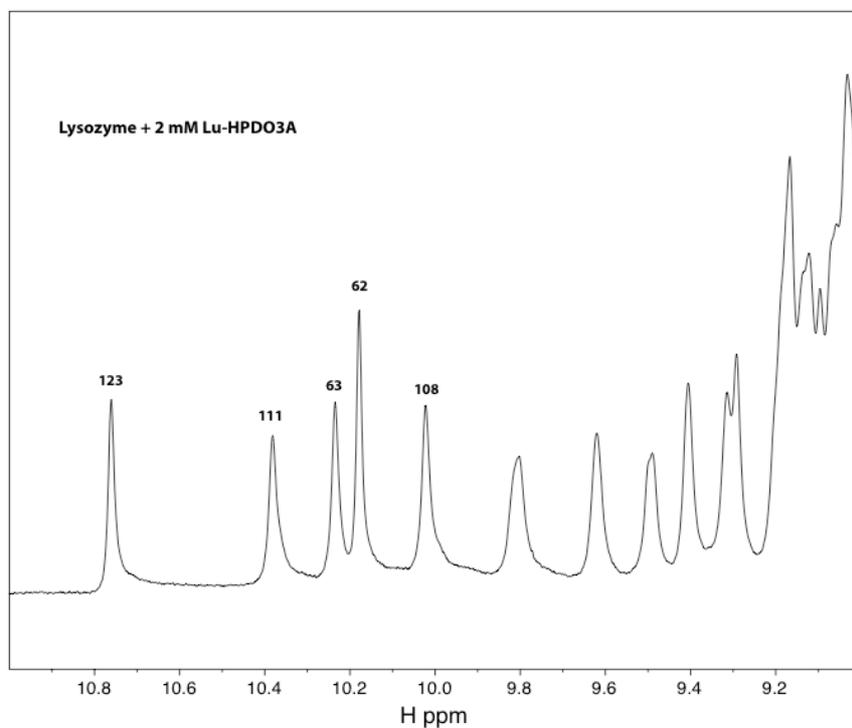


Figure 3.4.6 : Spectre RMN 1D <sup>1</sup>H d'une solution de lysozyme avec 2 mM de Lu-HPDO3A. Agrandissement du domaine des tryptophanes.

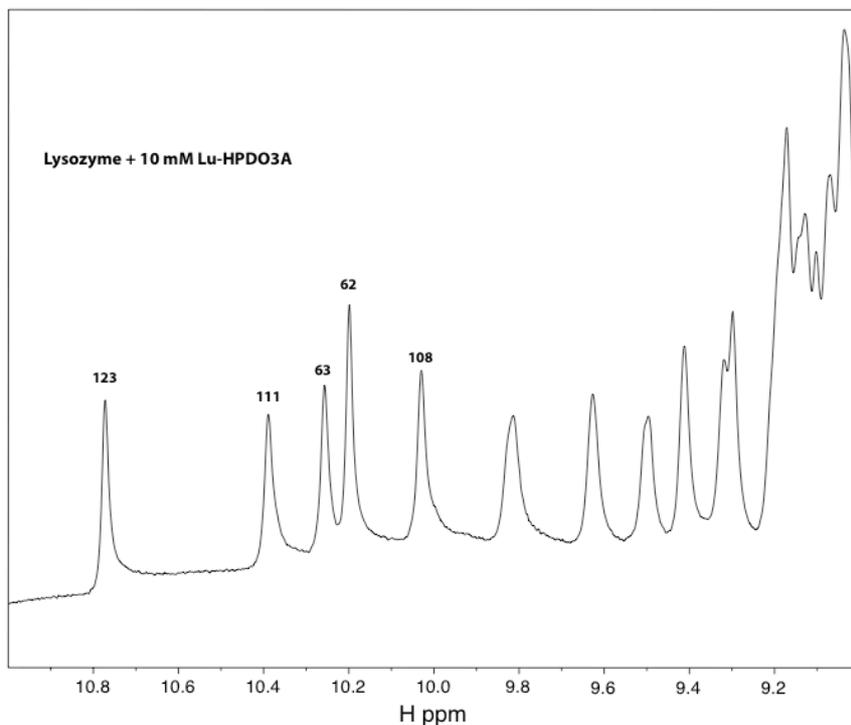


Figure 3.4.7 : Spectre RMN 1D  $^1\text{H}$  d'une solution de lysozyme avec 10 mM de Lu-HPDO3A. Agrandissement du domaine des tryptophanes.

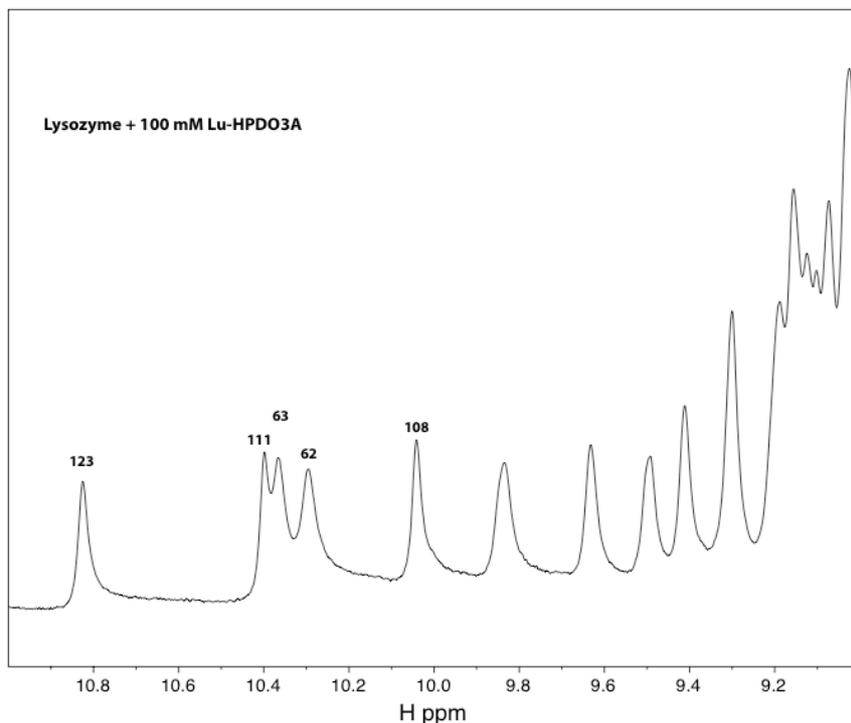


Figure 3.4.8 : Spectre RMN 1D  $^1\text{H}$  d'une solution de lysozyme avec 100 mM de Lu-HPDO3A. Agrandissement du domaine des tryptophanes.

Les figures 3.4.6 à 3.4.8 montrent l'évolution des pics des tryptophanes en ajoutant respectivement 2, 10 et 100 mM de Lu-HPDO3A à la solution de protéine. On remarque le déplacement et la baisse d'intensité relative graduelles de certains pics. Dans la figure 3.4.8, les positions des pics sont les suivantes : Trp123 : 10,82 ppm ; Trp111 : 10,39 ppm ; Trp63 :

10,36 ppm ; Trp62 : 10,29 ppm ; Trp108 : 10,03 ppm. Par rapport à la protéine native, on remarque un décalage notable pour les pics des Trp62 et 63 et un peu moins fort pour le pic du Trp123. Ce sont également ces mêmes pics qui subissent une baisse d'intensité par rapport aux pics des Trp111 et 108, qui changent relativement peu. Comme le montre la figure 3.4.9, l'ajout d'une forte concentration du complexe de Lu<sup>3+</sup> modifie l'intensité de certains pics des résidus de la protéine, mais ne les affecte pas tous.

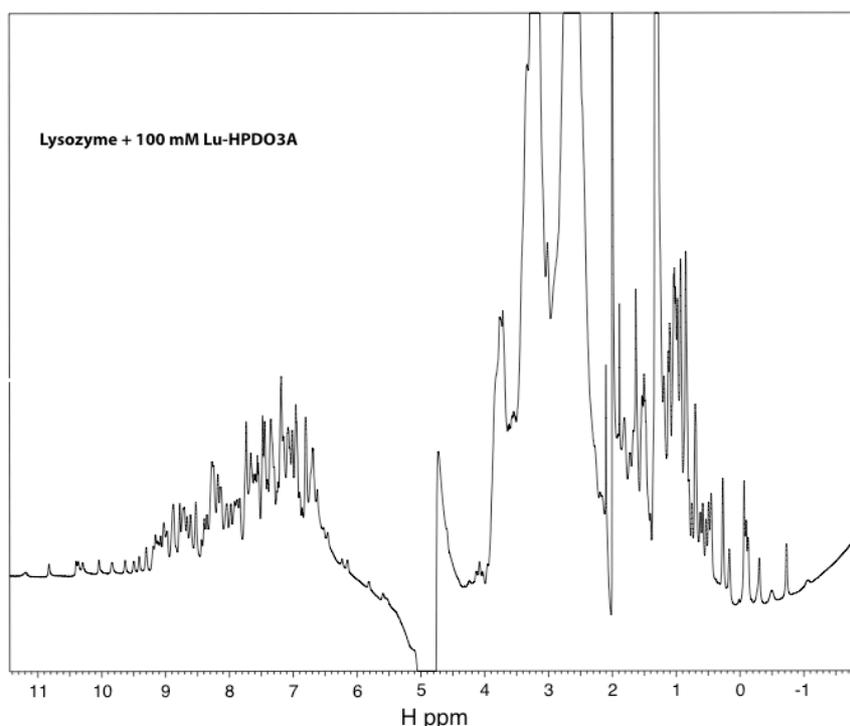


Figure 3.4.9 : Spectre entier de la solution de protéine avec 100 mM de Lu-HPDO3A. Les pics très intenses entre 1 et 5 ppm correspondent aux protons du ligand HPDO3A.

### III.4.8. Interprétation des spectres

Dans le cristal, ce sont les tryptophanes Trp62, Trp63 et Trp123 qui sont proches des molécules de complexe fixées, avec des distances entre l'atome d'azote indole et l'ion Ln<sup>3+</sup> de 6,8 Å, 8,6 Å, 7,5 Å et 15,0 Å pour le Trp62, 63, 123 et 108 respectivement. Le tableau 3.4.1. montre la vitesse d'échange du proton lié à l'azote du noyau indole des tryptophanes de lysozyme dans D<sub>2</sub>O d'après Laurents & Baldwin (1997). La vitesse d'échange correspond directement à l'exposition des résidus au solvant. Les tryptophanes avec des petites vitesses d'échange sont donc plutôt enfouis dans la protéine.

| Résidu                               | Trp62       | Trp63                        | Trp123                      | Trp111     | Trp108                      |
|--------------------------------------|-------------|------------------------------|-----------------------------|------------|-----------------------------|
| Vitesse d'échange [s <sup>-1</sup> ] | très rapide | rapide<br>6x10 <sup>-3</sup> | lente<br>2x10 <sup>-5</sup> | très lente | lente<br>4x10 <sup>-5</sup> |

Tableau 3.4.1 : Vitesse d'échange des protons d'azote indole des tryptophanes de lysozyme dans D<sub>2</sub>O d'après Laurents & Baldwin (1997). Pour les vitesses très lente et très rapide les valeurs numériques ne pouvaient pas être mesurées.

Si les changements d'intensité et de position des pics des protons de tryptophanes sont dus à la présence du complexe de lanthanide dans le solvant, alors les changements devraient le plus

fortement affecter les tryptophanes les plus exposés au solvant, donc d'abord le Trp62, ensuite le Trp63 et finalement le Trp123 et Trp108. Si, par contre, le changement est dû à la fixation spécifique du complexe sur les sites de fixation, ce devrait être en premier lieu les Trp62, 63 et 123 qui sont affectés.

Les pics les plus affectés par la présence des complexes correspondent aux tryptophanes 62 et 63 et, semble-t-il un peu moins fort, le Trp123. Les tryptophanes 62 et 63, situés au voisinage des sites de fixation, sont également les plus exposés au solvant, le changement d'intensité et de position des pics peut donc être dû à la présence des ions dans le solvant. Le pic du Trp111 est le moins affecté. Ce dernier résidu est éloigné des sites de fixation et est peu exposé au solvant.

Le pic du Trp108 semble moins affecté par la présence des ions  $\text{Lu}^{3+}$  et  $\text{Gd}^{3+}$  que le pic du Trp123, alors que leur exposition au solvant est analogue. Ceci indique un effet de fixation spécifique en proximité du Trp123. Mais cet effet peut aussi n'être dû qu'au hasard, car le déplacement et le changement d'intensité du pic peuvent être différents selon les conditions particulières qui affectent le résidu.

Comme les tryptophanes les plus exposés au solvant sont, à l'exception du Trp108, également les plus proches des sites de fixation, nous avons décidé d'enregistrer des spectres 2D  $^1\text{H}$ - $^1\text{H}$  NOESY pour obtenir de l'information supplémentaire sur d'éventuelles interactions du complexe avec la protéine.

### **III.4.9. Spectres 2D $^1\text{H}$ - $^1\text{H}$ NOESY**

Des spectres de RMN bidimensionnels sont obtenus en appliquant plusieurs impulsions de radiofréquence à l'échantillon avant de mesurer la décroissance du signal. Dans une expérience NOESY (Macura *et al.*, 1981) on mesure les effets dus à l'interaction de protons voisins dans l'espace. Ils interagissent par transfert d'énergie de résonance d'un proton excité à un proton non excité pendant le temps de mélange entre deux impulsions. Une expérience NOESY permet d'obtenir de l'information sur l'interaction de protons distants de moins de 6 Å, grâce aux pics de corrélation dus à l'effet Overhauser nucléaire (NOE).

### III.4.9.a. Enregistrement des spectres 2D

Les échantillons utilisés correspondent aux échantillons a) et g) décrits ci-dessus. En outre, le spectre 1D  $^1\text{H}$  du complexe Lu-HPDO3A seul a été enregistré pour identifier les positions des pics correspondant aux protons du ligand (figure 3.4.10).

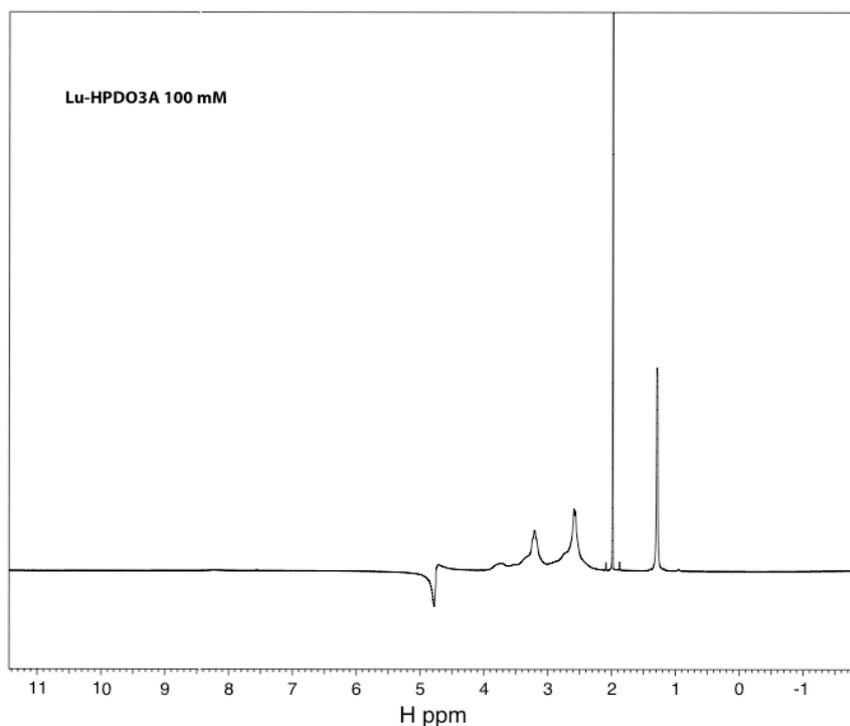


Figure 3.4.10 : Spectre 1D des protons du complexe Lu-HPDO3A.

### III.4.9.b. Spécifications techniques

Les spectres 2D sont enregistrés avec le spectromètre 600MHz du LRMN de l'IBS. Le temps de mélange est de 300 ms et le temps de mesure pour un spectre est d'une nuit.

### III.4.10. Spectres 2D $^1\text{H}$ - $^1\text{H}$ NOESY de lysozyme et de lysozyme avec 100 mM de Lu-HPDO3A

La figure 3.4.11 correspond au spectre NOESY de la protéine seul, la figure 3.4.12 au spectre de la protéine avec, en plus, 100 mM du complexe Lu-HPDO3A.

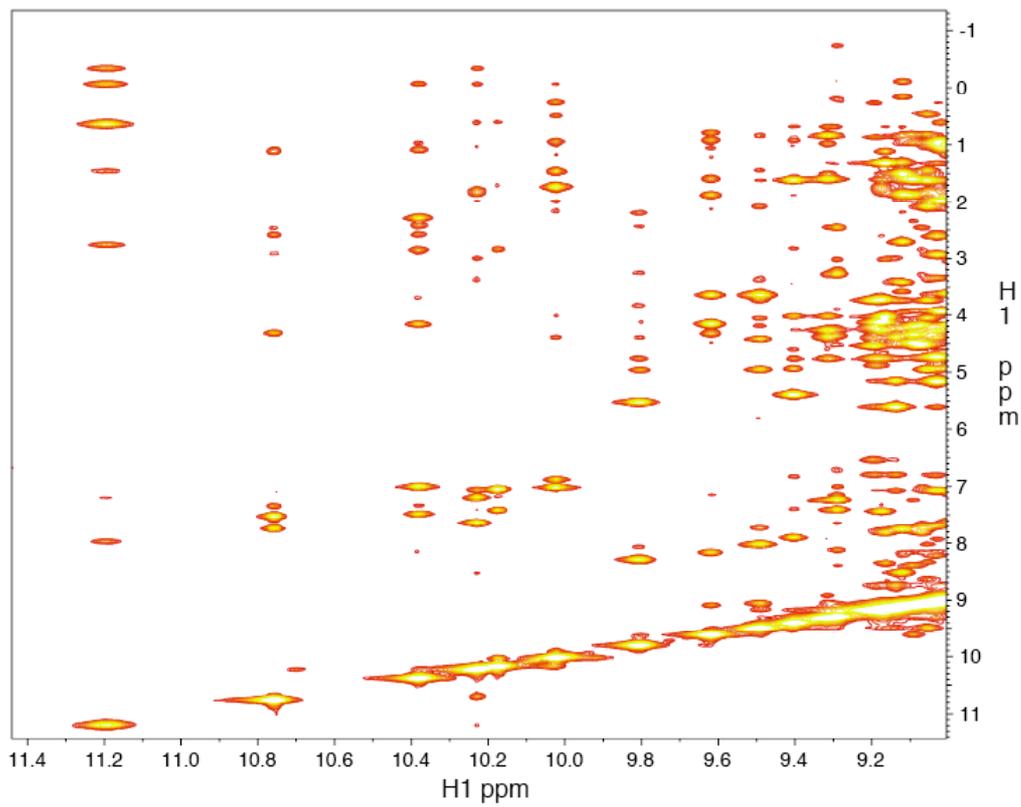


Figure 3.4.11 : Spectre  $^1\text{H}$ - $^1\text{H}$  NOESY enregistré sur une solution de 2 mM de lysozyme. Agrandissement du domaine des tryptophanes.

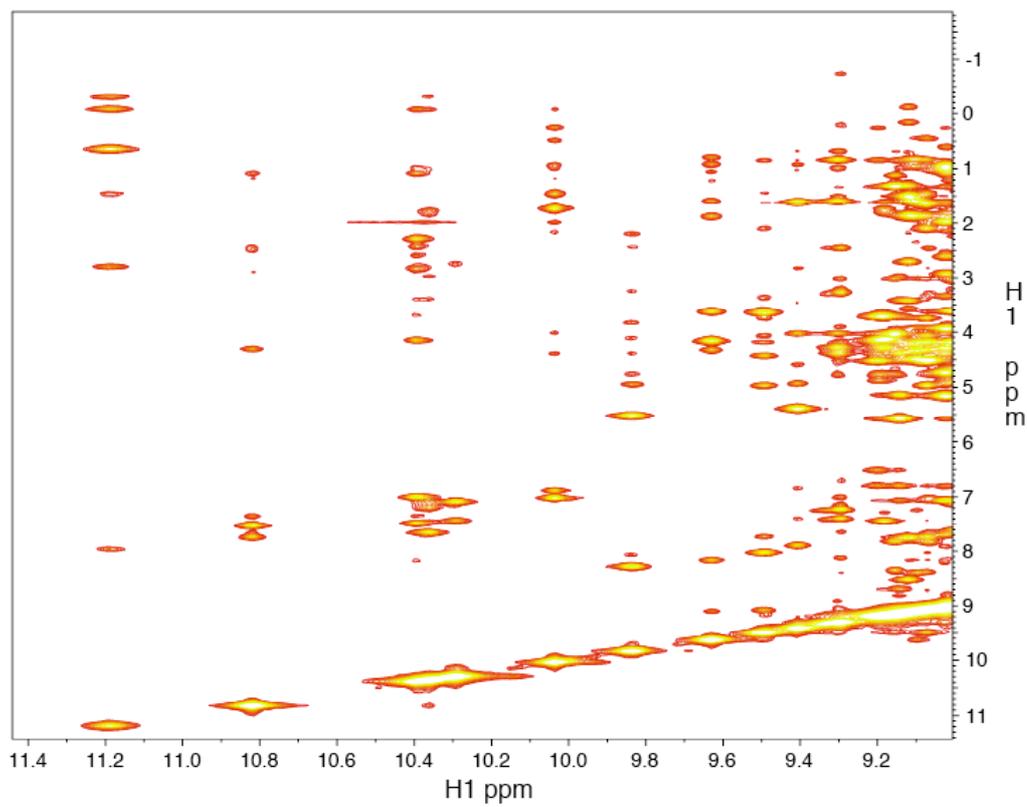


Figure 3.4.12 : Spectre  $^1\text{H}$ - $^1\text{H}$  NOESY enregistré sur une solution de 2 mM de lysozyme avec 100 mM de Lu-HPDO3A. Agrandissement du domaine des tryptophanes.

Dans les spectres des figures 3.4.11 et 3.4.12, on observe, en bas du spectre, entre 10,0 et 10,8 ppm, les pics de résonance diagonaux des protons des tryptophanes. Ce sont les pics qui correspondent aux bandes observées dans les spectres 1D. On observe le même déplacement entre les pics diagonaux des tryptophanes des deux spectres que celui qui est observé dans les spectres 1D quand on ajoute le complexe de  $\text{Lu}^{3+}$  (figure 3.4.12). Le déplacement relatif entre les pics des deux spectres apparaît plus clairement dans la figure 3.4.13, qui superpose les spectres des figures 3.4.11 et 3.4.12.

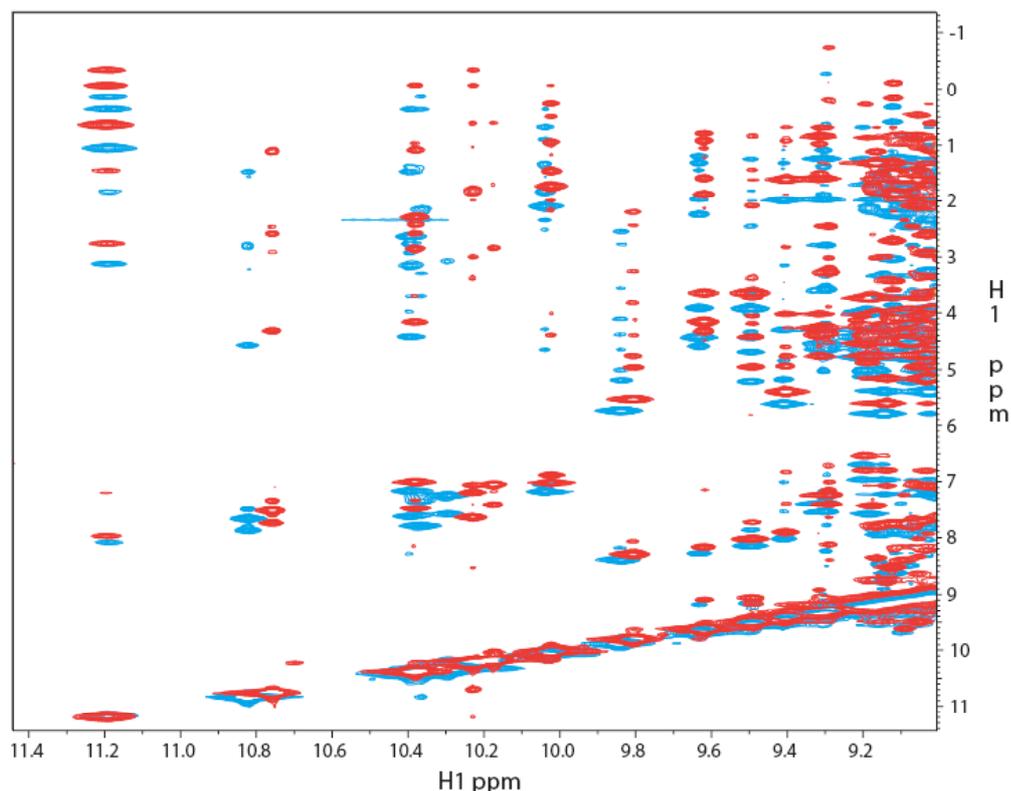


Figure 3.4.13 : Superposition des spectres enregistrés avec la solution protéique sans (rouge) et avec (bleu) 100 mM de Lu-HPDO3A. Agrandissement du domaine des tryptophanes.

Les pics diagonaux correspondant aux protons du ligand HPDO3A n'apparaissent pas dans les figures 3.4.12 et 3.4.13, car ils sont situés entre 1 et 5 ppm (voir spectre 1D du ligand seul : figure 3.4.10).

La figure 3.2.14 montre le spectre complet de la protéine avec le complexe Lu-HPDO3A. En raison de la concentration élevée du complexe, les pics correspondant aux protons du ligand forment des tracés entre 1 et 5 ppm. La forte intensité de ces pics par rapport au reste du spectre, empêche la détection d'éventuels pics de corrélation entre les protons du ligand et des protons de la protéine, qui résonnent dans le même domaine. Il est plus aisé de chercher de tels pics dans le domaine des fréquences de résonance des tryptophanes.

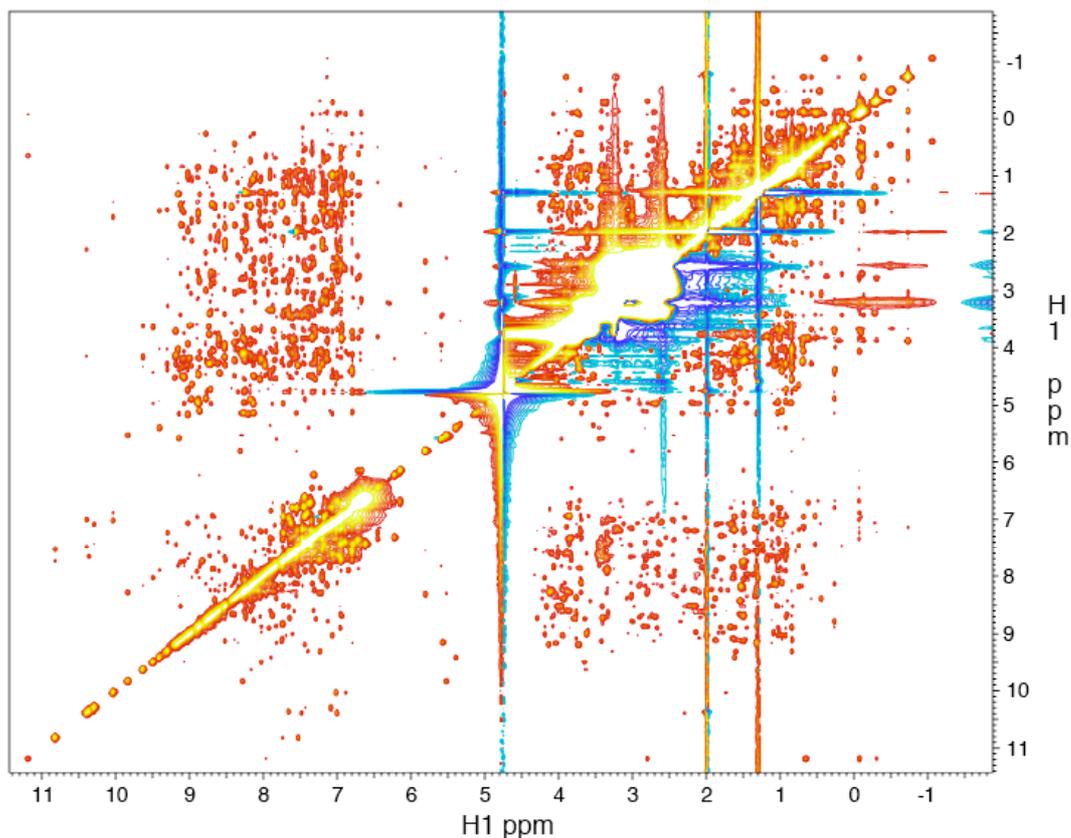


Figure 3.2.14 : Spectre  $^1\text{H}$ - $^1\text{H}$  NOESY enregistré sur une solution de 2 mM de lysozyme avec 100 mM de Lu-HPDO3A. En raison de la concentration élevée du complexe, les pics correspondant aux protons du ligand forment des tracés entre 1 et 5 ppm.

### III.4.11. Interprétation des spectres

S'il y avait fixation spécifique du complexe Lu-HPDO3A à proximité des tryptophanes 62, 63 et 123, des pics de corrélation entre les protons du ligand et le tryptophane considéré pourraient apparaître dans le spectre de la figure 3.4.12, alors que ces pics ne seraient pas présents dans la figure 3.4.11. Ces pics devraient apparaître aux fréquences de résonance des tryptophanes concernés. La comparaison du spectre de la figure 3.4.12 au spectre de la figure 3.4.11 ne montre aucune évidence d'apparition de tels pics.

### III.4.12. Conclusions sur les expériences RMN 1D et 2D

Les expériences 1D et 2D n'ont pas permis de détecter la fixation du complexe sur la protéine, pourtant observée dans le cristal. Ceci peut être dû à différentes raisons : la fixation peut être trop peu spécifique, avec une constante de dissociation trop élevée pour permettre de détecter l'interaction.

Il se peut également que le complexe ne se fixe pas sur la protéine en solution. Il semble probable que la fixation du complexe dans le cristal de lysozyme soit favorisée par l'empilement cristallin qui forme une cavité autour des sites de fixation, impliquant trois molécules de protéine symétrique (voir chapitre II.4.8.).

Il serait intéressant de savoir si le résultat d'expériences analogues avec une protéine de la famille des ChBP serait le même, car pour ces protéines la cavité de fixation est également présente en solution.

Même si les résultats étaient plus probants, ce type d'expérience ne permettrait pas de détecter la fixation d'un complexe sur une protéine, de manière rapide et simple. En effet, l'expérience demande un équipement particulier (spectromètres RMN). De plus, l'interprétation des spectres est très compliquée, car les pics de la protéine ne sont généralement pas attribués et les différents complexes ne se fixent pas forcément sur les tryptophanes, et, enfin, l'expérience nécessite une grande quantité de protéine, de l'ordre de la dizaine de milligrammes.

## **III.5. La luminescence des lanthanides**

### **III.5.1. Propriétés photophysiques des ions de lanthanides trivalents**

Les cations trivalents des lanthanides ont la configuration électronique (Xe)  $4fn$ , avec  $n$  variant de 1 ( $Ce^{3+}$ ) à 14 ( $Lu^{3+}$ ). Les transitions des électrons  $f$  sont responsables des propriétés optiques des ions de lanthanides. La luminescence des lanthanides est caractérisée par une longue durée de vie des états excités, de l'ordre de la milliseconde (en comparaison, la durée typique de fluorescence de fond dans du matériel biologique (résidus aromatiques) est de l'ordre de la nano- à la microseconde). Les spectres d'absorption et d'émission sont composés de raies fines très marquées, et sont remarquablement insensibles à des changements environnementaux, dues au blindage des électrons  $f$  contre des perturbations extérieures par les orbitales  $5s$  et  $5p$  remplies.

C'est l'interdiction des transitions  $f-f$  qui est responsable de la longue durée de vie des états excités, mais cette première résulte aussi en de faibles coefficients d'extinction des ions (de l'ordre de  $0,3$  ( $Tb^{3+}$ ) à  $3$  ( $Gd^{3+}$ )  $L.mol^{-1}.cm^{-1}$  aux raies d'absorption (Binnemans & Görller-Walrand, 1995)), qui rendent difficile la photoexcitation directe des ions. Pour des ions de lanthanides en solution aqueuse, le rendement de luminescence est très faible, car les molécules d'eau de la première sphère de coordination donnent lieu à la désexcitation par voie non-radiative des états excités par couplage vibronique à des états vibrationnels des liaisons O-H (et N-H). La liaison des molécules d'eau entraîne également un fort raccourcissement de la durée de vie de la luminescence.

L'intensité de luminescence est proportionnelle à la concentration du lanthanide (Parker & Williams, 1996), dans la limite de concentrations "raisonnables", ne donnant pas lieu au phénomène d'extinction par effet de concentration (« concentration quenching »).

#### **III.5.1.a. Choix des ions**

De tous les ions  $Ln^{3+}$ , seuls les ions  $Eu^{3+}$ ,  $Tb^{3+}$ ,  $Gd^{3+}$ ,  $Ce^{3+}$ ,  $Dy^{3+}$  et  $Sm^{3+}$  luminescent de manière significative en solution aqueuse, à température ambiante (Horrocks, 1993). Au vu des conditions expérimentales accessibles, nous avons choisi de travailler avec l'euprimum et le terbium, leurs raies d'absorption et d'émission étant situées dans le domaine de l'ultraviolet et du visible. Nous avons également essayé d'utiliser la luminescence du  $Gd^{3+}$ , qui absorbe et qui émet dans l'ultraviolet. L'utilisation de la luminescence du  $Gd$  nous permettrait d'utiliser nos complexes sans essayer de remplacer l'ion  $Gd^{3+}$  des complexes utilisés par un autre ion de lanthanide. Cependant, comme le comportement de fixation des complexes sur les protéines

semble indépendant de l'ion  $\text{Ln}^{3+}$  chélaté (Girard *et al.*, 2003c) l'utilisation des complexes avec le gadolinium n'est pas impératif pour notre étude.

### III.5.1.b. Émission stimulée par effet d'antenne (sensitized emission)

Pour les applications habituelles de lanthanides comme marqueurs luminescents, la faiblesse des coefficients d'absorption est généralement contournée en coordonnant l'ion avec un ligand qui contient un chromophore. Le chromophore, servant d'antenne, absorbe de la lumière et transfère de l'énergie sur l'ion qui luminesce. Le ligand sert également à exclure des molécules de solvant de la sphère de coordination de l'ion, ce qui empêche la désexcitation non-radiative de l'ion. En augmentant ainsi l'absorption de lumière et le rendement de luminescence, on obtient un signal fortement amplifié qui permet de détecter de faibles concentrations de l'agent luminescent, même en utilisant des sources de lumière peu intenses. Même pour des ions de lanthanides sans complexe "antenne", l'amplification de l'émission peut être importante. Ainsi, pour des ions de  $\text{Tb}^{3+}$  liés à des acides nucléiques, ces derniers servent d'antenne quand on irradie l'ensemble avec du rayonnement ultraviolet : les acides nucléiques absorbent la lumière et transmettent une partie de l'énergie aux ions  $\text{Tb}^{3+}$  qui transforment partiellement cette énergie en luminescence (Feig *et al.*, 1999). De manière analogue, les tryptophanes des protéines absorbent la lumière UV et transmettent de l'énergie aux ions  $\text{Tb}^{3+}$  liés à la protéine. Cet effet dépend de la distance entre l'antenne et l'ion, mais il peut être encore observé pour des distances sensiblement supérieures à 10 Å (Horrocks & Collier, 1981). Ce même effet n'est pas observé pour les ions  $\text{Eu}^{3+}$ .

### III.5.1.c. Influence des molécules d'eau coordonnées à l'ion sur sa luminescence

Les molécules d'eau de la première sphère de coordination de l'ion provoquent le raccourcissement de la durée de vie et la baisse d'intensité de la luminescence. Le rendement de la désexcitation non-radiative, via les molécules d'eau liées, dépend de la différence d'énergie entre l'état excité, émissif, de l'ion, et le niveau vibrationnel de l'état fondamental le plus élevé. Plus la différence est faible, plus l'effet de désexcitation par voie non-radiative est fort. Pour les ions d'Eu, de Tb et de Gd, l'effet est le plus fort pour l'Eu, avec un rapport  $I_D/I_H^*$  de 40. Ce rapport est de 10 pour le Tb. Il est de 1 pour le Gd, pour lequel la différence d'énergie est la plus élevée, ce qui empêche le transfert d'énergie aux états vibrationnels des molécules d'eau.

\*  $I_D/I_H$  : rapport des intensités de luminescence émises pour des ions dans une solution de  $\text{D}_2\text{O}$  et dans une solution de  $\text{H}_2\text{O}$ , les liaisons D-H ne permettant pas la désexcitation de l'ion.

### III.5.1.d. Durée de vie de la luminescence

Des expériences en temps résolu, qui consistent à mesurer l'intensité émise après le délai  $t$  qui suit l'arrêt de l'excitation, permettent de déterminer la durée de vie de l'état excité de l'ion luminescent. Pour un seul environnement des ions luminescents, le déclin de la luminescence émise obéit à une loi exponentielle :  $I(t) = I_0 \exp(-k_{obs} t)$ , avec la constante de décroissance de luminescence observée dans l'eau :  $k_{obs}^{\text{H}_2\text{O}} = k^0 + \sum k_i^{nr} + k_{OH}$ , avec  $k^0$  : constante de décroissance radiative,  $\sum k_i^{nr}$  : la somme des constantes de processus de désexcitation non-radiative autres que par les liaisons O-H de molécules d'eau liées,  $k_{OH}$  : constante de désexcitation non-radiative par transfert d'énergie aux oscillateurs O-H de molécules d'eau liées (Parker & Williams, 1996).

Le tableau 3.5.1 montre des durées de vie mesurées pour trois complexes d'euporium, pour des complexes en solution de  $\text{H}_2\text{O}$  et de  $\text{D}_2\text{O}$  (Anelli *et al.*, 1991). Dans les deux premiers

complexes, la sphère de coordination de l'ion comporte une molécule d'eau ; dans le complexe Eu-DO3A elle comporte deux molécules d'eau (voir chapitre I.5.2). Si, lors de la fixation du complexe à la protéine, ces molécules d'eau sont remplacées par des résidus protéiques, alors l'effet sur la durée de vie de la luminescence du complexe lié devrait être le même que lorsqu'on remplace les molécules d'eau entourant le complexe par des molécules de D<sub>2</sub>O.

| Complexe     | $\tau_{D_2O}$ , 77K [ms] | $\tau_{H_2O}$ , 77K [ms] | $\tau_{D_2O}$ , 300K [ms] | $\tau_{H_2O}$ , 300K [ms] |
|--------------|--------------------------|--------------------------|---------------------------|---------------------------|
| Eu-DOTA      | 2,3                      | 0,89                     | 2,6                       | 0,65                      |
| Eu-HPSA-DO3A | 2,4                      | 0,93                     | 2,3                       | 0,62                      |
| Eu-DO3A      | 0,74                     | 0,38                     | 0,70                      | 0,31                      |

Tableau 3.5.1 : Durées de vie mesurée pour les différents complexes d'euporium dans une solution de D<sub>2</sub>O et dans une solution de H<sub>2</sub>O, à température ambiante et à température cryogénique. La durée de vie est mesurée pour l'émission correspondant à la transition <sup>5</sup>D<sub>0</sub> -> <sup>7</sup>F<sub>1</sub>.

### III.5.1.e. Influence de l'environnement sur la forme du spectre d'émission

La forme du spectre d'émission est relativement peu sensible à l'environnement de l'ion, car les électrons de la couche f sont blindés contre des perturbations extérieures par les électrons 5s et 5p. Pour les complexes des lanthanides, le changement du profil du spectre est d'autant moins important que l'environnement de l'ion est principalement imposé par le ligand qui l'entoure. L'environnement de l'ion peut avoir une influence sur l'intensité relative de certaines raies d'émission (transitions hypersensibles) (Eu : 615 nm, Tb : 490 nm) (Horrocks & Albin, 1991).

### III.5.1.f. Applications en biologie

La luminescence des lanthanides, en particulier à travers l'utilisation de complexes de lanthanides blindant l'ion contre le solvant et comportant un chromophore, joue un rôle important pour les marqueurs employés dans la microscopie de fluorescence, ou, pour des dosages immunologiques par fluorescence (fluoroimmunoassays), en utilisant un complexe contenant un ion Ln<sup>3+</sup> lié à l'anticorps ou à l'antigène de la substance à détecter. La longue durée de vie de la luminescence permet, par des expériences en temps résolu, de distinguer la luminescence du marqueur de la fluorescence de fond du matériel biologique.

Une autre application est l'étude des sites de fixation des ions Ca<sup>2+</sup> ou Mg<sup>2+</sup> dans les protéines. Avec un rayon ionique proche de celui de ces derniers ions, les ions de lanthanides ont une forte affinité pour la protéine à ces sites de fixation. L'inspection de la luminescence des ions Ln<sup>3+</sup> liés à la protéine dans ces sites permet de caractériser leur environnement, en particulier le nombre de molécules d'eau remplacées lors de la liaison de l'ion, et, par là même, de déterminer le nombre de sites de fixation différents (Bünzli & Pfefferle, 1994; Elbanowski & Makowska, 1996). Le titrage de la protéine avec des ions de lanthanides permet, par une expérience de compétition, la comparaison d'affinités de différents ions pour le site de fixation (Feig *et al.*, 1999).

## III.5.2. But : détecter la fixation de nos complexes sur les protéines

Le but de l'expérience est de détecter la fixation d'un complexe à une protéine, soit en solution, soit dans le cristal. Le problème est que l'émission de luminescence des molécules de

complexe liées à la protéine sera superposée à l'émission des molécules présentes à forte concentration dans le solvant. En outre, comme les ions sont engagés dans les complexes, leur environnement direct ne change que très peu lors de l'interaction avec la protéine et il y a donc peu de chance de pouvoir observer de fortes différences dans les caractéristiques de la lumière émise.

Il n'est pas possible de se débarrasser des ions non-liés présents dans les canaux de solvant, car il est probable que le lavage du cristal (back-soak) dans une solution ne contenant pas de complexe entraînerait non seulement une diminution de la concentration de complexe dans le solvant mais aussi une diminution du taux d'occupation des molécules de complexe dans les sites de fixation. En effet, des études préliminaires sur des cristaux dérivés contenant un complexe d'euprium de taille comparable à la taille de nos complexes et utilisant la luminescence des lanthanides, indiquent que, dans un cristal dérivé trempant dans une solution qui ne contient pas de complexe, les molécules de complexe se dissocient des sites de fixation avec un temps caractéristique de l'ordre de quelques minutes. (G. Pompidor, IBS-LCM, communication privée)

Il serait intéressant de voir, si, avec des mesures précises et en choisissant de manière astucieuse les conditions expérimentales, il serait possible de détecter certains effets dus à la fixation d'un complexe sur une protéine. De tels effets pourraient être :

- a) Le changement de durée de vie par le remplacement de molécules d'eau, avec un effet attendu le plus fort pour un complexe d'Eu. Ceci nécessite que la fixation du complexe se fasse via la coordination de l'ion par des résidus protéiques, remplaçant ainsi la (ou les) molécule(s) d'eau liée(s) à l'ion (voir, au chapitre II.4, les modes de fixation affinés). Une expérience en temps résolu permettrait de distinguer la luminescence des molécules de complexe non-liées, hydratées, de la luminescence des ions liés à la molécule qui décroîtrait moins rapidement.
- b) Le changement d'intensité résultant des trois effets suivants :
  - dans le cas d'un cristal, de l'augmentation de concentration du complexe due à la fixation du complexe à la protéine,
  - de l'éventuel remplacement de molécules d'eau liées à l'ion de lanthanide par des résidus protéiques,
  - éventuellement, avec les ions  $Tb^{3+}$  et en utilisant une source de lumière UV, par une augmentation de la luminescence par transfert d'énergie absorbée par les tryptophanes sur les ions.
- c) Distinguer les molécules de complexe de  $Tb^{3+}$  en solution de celles qui sont liées à des tryptophanes (lysozyme, ChBP), par excitation avec un rayonnement absorbé par les tryptophanes mais non pas par les ions. Cependant, il n'est pas sûr que le complexe et la protéine soient liés pendant suffisamment longtemps pour permettre le transfert d'énergie. Et dans ce cas, le transfert se fera-t-il de manière privilégiée sur les ions des complexes liés et non pas sur les ions dans le solvant ?

L'étude du changement d'intensité de luminescence dû à la fixation du complexe à la protéine dans le cristal n'est possible qu'en comparant différents complexes dans des cristaux d'une même protéine, car la concentration apparente dépend du contenu de solvant du cristal. Pour une concentration de 100 mM des molécules de complexe non-liées dans les canaux de solvant d'un cristal de lysozyme et de glucose isomérase, avec des pourcentages de solvant respectifs de 29% et 48%, la concentration apparente est de 29 mM et de 48 mM. Si on considère que les molécules de complexe liées s'ajoutent aux molécules de complexe libres et en supposant qu'une molécule de complexe se lie à une molécule de protéine, avec la concentration de protéine dans le cristal de 14 mM pour la glucose isomérase (voir chapitre

I.5.4.) et de  $C_p = \frac{16}{V_{\text{maille}} \times N_A} = 120mM$  pour le lysozyme, la fixation du complexe entraînerait

donc une augmentation notable de la concentration, de 30 % et de 400% respectivement.

Les effets attendus dépendront des conditions expérimentales, de l'ion de lanthanide utilisé, mais également du mode de fixation du complexe sur la protéine (fixation par le côté ionique, par le côté du macrocycle, via des tryptophanes...).

La mesure des effets mentionnés ci-dessus nécessite un montage expérimental qui permette de choisir une longueur d'onde d'excitation dans le domaine d'absorption des ions ou d'éventuelles antennes et détecter le rayonnement à la longueur d'onde d'émission. Il est nécessaire de pouvoir quantifier la luminescence mesurée et de pouvoir réaliser des expériences en temps résolu avec des temps caractéristiques de l'ordre de la milliseconde.

### III.5.3. Expériences de luminescence avec des complexes et des ions de lanthanides avec des cristaux et des solutions de protéine

#### III.5.3.a. Choix de la longueur d'onde de la lumière incidente

Pour l'excitation directe des ions, les spectres d'absorption correspondent aux spectres d'excitation, ce qui correspond à un rendement relativement constant (Anelli *et al.*, 1991). Pour une excitation directe efficace il faut donc viser des raies d'absorption. Le tableau 3.5.2 montre les domaines d'absorption et d'émission des ions  $\text{Eu}^{3+}$  et  $\text{Tb}^{3+}$  décrits dans la littérature.

|                  | Raies d'absorption [nm]                     | Raies d'émission [nm]                       |
|------------------|---|---|
| $\text{Eu}^{3+}$ | 250, 280-330, 355-360, 370-400              | 585-600, 610-630, 577-581, 700              |
| $\text{Tb}^{3+}$ | 245-300 <sup>a</sup> , 300-380 <sup>b</sup> | 280-330, 355-360, 370-400, 485-500, 540-555 |

Tableau 3.5.2 : Longueurs d'onde des raies d'absorption et émission pour les ions  $\text{Eu}^{3+}$  et  $\text{Tb}^{3+}$ . <sup>a</sup> : forte absorption, <sup>b</sup> : faible absorption. (Elbanowski & Makowska, 1996 ; Binnemans & Görller-Walrand, 1995 ; Binnemans & Görller-Walrand, 1995 ; Anelli *et al.*, 1991). Dans Horrocks & Albin (1991), les spectres sont enregistrés avec des solutions de  $[\text{LnEDTA}]$ , dans Binnemans & Wallrand (1995) avec des solutions de perchlorate de Ln(III). Toutes les gammes de longueur d'onde sont approximatives.

Pour une absorption de la lumière par les résidus de protéine aromatiques il faut travailler avec une source de lumière UV. Le maximum d'absorption des tryptophanes est situé à 280 nm (Brooks, 2005).

En travaillant avec une source de lumière de longueur d'onde supérieure à 300 nm on exclue le transfert d'énergie via les résidus aromatiques.

#### III.5.3.b. La luminescence de l'ion $\text{Gd}^{3+}$

L'ion  $\text{Gd}^{3+}$  présente plusieurs raies d'absorption vers 275 nm (Binnemans & Görller-Walrand, 1995). La seule raie d'émission relativement intense est située vers 310 nm (Anelli *et al.*, 1991), elle coïncide avec le domaine d'émission des résidus aromatiques de la protéine et ne pourra probablement en être distinguée que par des expériences en temps résolu.

#### III.5.3.c. Synthèse des complexes de lanthanides

Notre premier problème pour procéder aux expériences de luminescence a été que nous ne disposions que des complexes de  $\text{Gd}^{3+}$  et non pas des complexes analogues d' $\text{Eu}^{3+}$  ou de  $\text{Tb}^{3+}$ . Par la suite, nous avons pu obtenir les complexes  $\text{Eu-DOTA}$  et  $\text{Eu-HPSA-DO3A}$  de la part de

Pier Lucio Anelli de la société Bracco Imaging, qui nous avait déjà fourni la plupart des complexes de gadolinium. Nous avons également contacté l'équipe de David Parker (Parker & Williams, 1996) qui devait nous fournir du complexe Eu-DO3A qui ne nous est pas encore parvenu.

En fait, lorsqu'on dispose du ligand et du sel de lanthanide, la synthèse des complexes est relativement facile et nous aurions pu la faire nous-mêmes. Grâce à la forte affinité des ligands pour les ions de lanthanides, il suffit d'ajouter la quantité stœchiométrique de sel de lanthanide à la solution du ligand. Cependant, parmi les huit ligands, trois seulement sont disponibles commercialement : les ligands DTPA, DOTMA et DOTA. Remplacer un ion de lanthanide chélaté par le ligand contre un autre ion est beaucoup plus délicat : pour dissocier le ligand de l'ion d'origine, il faut travailler dans des conditions très acides tout en chauffant pendant plusieurs jours ce qui nécessite des équipements dont nous ne disposons pas à l'IBS.

### III.5.4. Conditions expérimentales

Les expériences de luminescence ont été réalisées avec le microspectrophotomètre du " cryobench ", équipement installé à l'ESRF en collaboration avec l'IBS, et qui est habituellement utilisé pour des études d'absorption et fluorescence de cristaux de protéines dans l'UV/visible (Bourgeois *et al.*, 2002). Les expériences ont été réalisées avec l'aide d'Antoine Royant (LCCP, IBS) qui m'a montré comment utiliser cet équipement.

#### III.5.4.a. Spécifications techniques du microspectrophotomètre utilisé

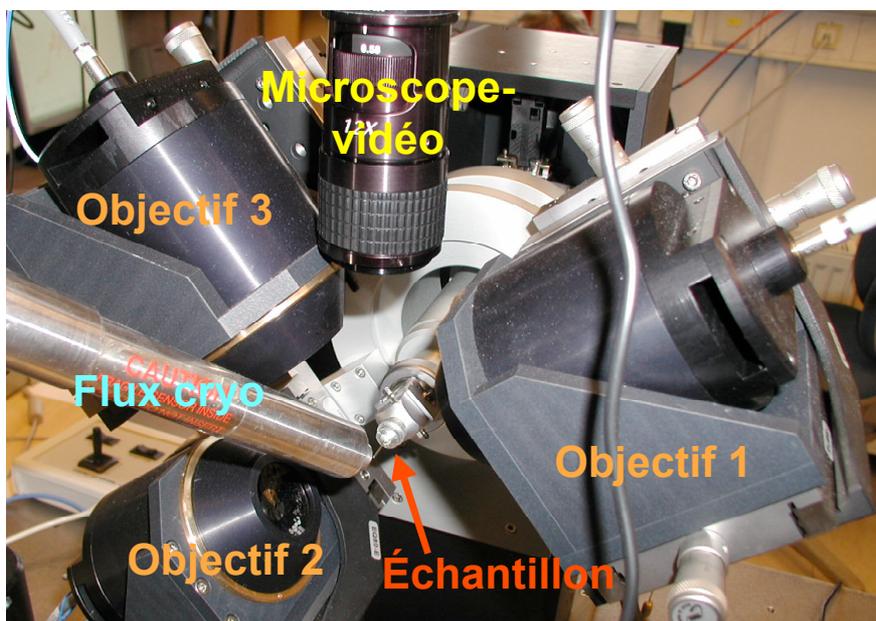


Figure 3.5.1 : Montage expérimental sur le " cryobench ".

La figure 3.5.1 montre le montage expérimental utilisé. L'échantillon est monté dans une boucle standard sur une tête goniométrique. L'expérience peut se faire soit à température ambiante, soit avec un échantillon congelé sous flux d'azote gazeux. La lumière est amenée sur l'échantillon à travers un premier objectif qui est relié à la source de lumière par une fibre optique. Un deuxième objectif, face au premier, recueille la lumière envoyée au détecteur qui permet de mesurer des spectres d'absorption. Un troisième objectif, à 90° de la lumière

incidente, permet l'acquisition de spectres de fluorescence. Un microscope permet d'observer l'échantillon et de régler les objectifs et les fentes pour pouvoir concentrer la lumière incidente sur l'échantillon et recueillir la lumière transmise ou émise à travers les deux autres objectifs. L'échantillon peut être un cristal de protéine ou une goutte de solution. Dans le deux cas, la quantité de protéine utilisée est assez faible, avec un volume de l'échantillon de l'ordre de  $10^{-12}$  -  $10^{-9}$  L. Nous avons effectué toutes les expériences à température cryogénique, ce qui empêche l'échantillon de sécher et peut retarder d'éventuels dégâts d'irradiation. Les sources de lumière disponibles sur le microspectrophotomètre sont des lasers avec des longueurs d'onde d'émission de 266 nm, 355 nm et 532 nm et une lampe de lumière blanche émettant un spectre avec une intensité relativement constante sur l'ensemble du domaine UV/visible. Le détecteur CCD « one-shot » utilisé, permet l'acquisition rapide de spectres d'émission UV/visible avec une résolution de 2,3 nm.

### **III.5.4.b. Problèmes expérimentaux**

Le montage, simple, fait que le signal à détecter doit être relativement intense (pertes dans les objectifs, passage à travers l'air, lumière diffusée...). Ceci, ajouté à la faible taille de l'échantillon, fait que nous n'avons pas pu détecter de signaux de luminescence, ni acquérir des spectres d'absorption en utilisant la source de lumière blanche. Nous n'avons pu détecter des signaux de luminescence provenant des ions de lanthanide qu'en utilisant les lasers. Notre choix de longueur d'onde d'excitation a donc été restreint aux longueurs d'onde 266 nm et 355 nm. Il n'est pas possible, sur ce montage, de faire des expériences en temps résolu et l'acquisition des spectres se fait en même temps que l'irradiation. C'est pourquoi nous ne pouvons ni déterminer la durée de vie des états excités des ions luminescents, ni distinguer la luminescence d'ions dans des environnements différents (liés, non-liés), ni enfin nous débarrasser de la fluorescence de fond du matériau biologique utilisé.

De nombreux phénomènes peuvent venir perturber la mesure : l'intensité du signal de luminescence dépend fortement de l'orientation de l'échantillon dans le trajet de la lumière, il peut y avoir des réflexions, des perturbations par le porte-échantillon, etc... Il n'est donc pas possible de réaliser des mesures d'intensité quantitatives absolues qui permettraient de comparer les signaux de différents échantillons, mais il est possible de comparer l'intensité relative de différentes raies d'émission. La nécessité de travailler avec une lumière excitatrice intense peut entraîner une dégradation rapide de l'échantillon lorsqu'on travaille avec le laser ultraviolet (voir figure 3.5.15).

### **III.5.4.c. Porte-échantillons utilisés**

Pour nos premières expériences, nous avons utilisé des boucles en nylon comme porte-échantillon. Pour les expériences suivantes, nous avons travaillé avec des boucles lithographiées en mylar, car le nylon est lui-même fluorescent. On s'est alors aperçu que, sur les boucles de mylar utilisées pour la première fois, des impuretés fluorescaient fortement lorsqu'on les irradiait avec de la lumière ultraviolette, et leur fluorescence pouvait, selon la géométrie utilisée, perturber fortement nos spectres d'émission (figure 3.5.12).

### **III.5.5. Expériences de luminescence**

Dans un premier temps, nous avons voulu déterminer des conditions expérimentales, comme la longueur d'onde et l'intensité de la lumière excitatrice, adaptées à l'acquisition des spectres de luminescence avec les différents échantillons.

La longueur d'onde d'excitation ( $\lambda_{exc}$ ) et le temps d'acquisition ( $t$ ) sont indiqués sur tous les spectres présentés. Dans ces spectres, la raie observée à la longueur d'onde de la lumière excitatrice correspond à un reflet de la lumière incidente, la raie à 730 nm à un artefact du détecteur.

Les protéines utilisées sont le lysozyme de blanc d'œuf de poule et la glucose isomérase (GI) de *Streptomyces rubiginosus*.

#### **III.5.5.a. Préparation des échantillons**

Les cristaux utilisés sont obtenus avec les conditions de cristallisation décrites en II.1.3. Sauf mention contraire, les cristaux sont cryoprotégés par trempage dans une solution cryoprotectrice et les solutions de protéine et de complexes de lanthanides par ajout de la quantité nécessaire d'agent cryoprotecteur afin d'éviter la formation de cristaux de glace, qui provoqueraient une forte diffusion de la lumière émise.

### **III.5.6. Spectre de fluorescence d'un cristal et d'une solution de protéine native**

Les figures 3.5.2 et 3.5.3 montrent la fluorescence des échantillons de protéine, dont il faut tenir compte pour l'interprétation des spectres qui suivent, puisque, avec ce montage expérimental, on ne peut pas séparer la fluorescence de fond, de courte durée de vie, de la luminescence émise par les lanthanides. Les échantillons biologiques et la boucle de nylon fluorescent, avec un large domaine d'émission entre 400 et 500 nm (figure 3.5.2), fluorescence à laquelle s'ajoute, lorsque l'échantillon est irradié avec de la lumière ultraviolette, la forte fluorescence des résidus aromatiques, avec un maximum d'émission autour de 330 nm (figure 3.5.3).

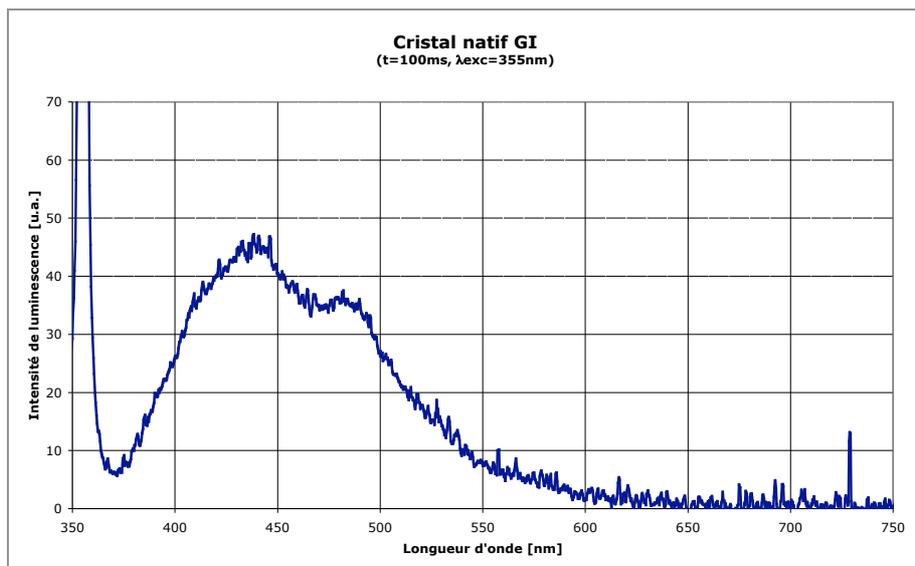


Figure 3.5.2. : Spectre de luminescence d'un cristal de glucose isomérase native, irradié avec un laser de longueur d'onde 355 nm. La raie à 355 nm correspond à un reflet de la lumière excitatrice, la raie à 730 nm à un artefact du détecteur. L'émission entre 400 et 500 nm provient de la fluorescence de la protéine et de la boucle de nylon.

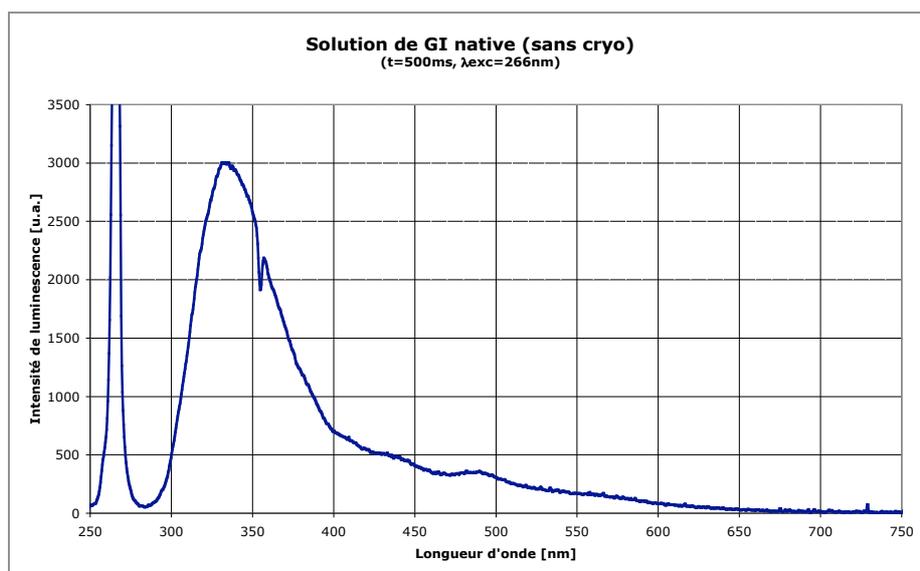


Figure 3.5.3. : Spectre de luminescence d'un cristal de glucose isomérase native, irradié avec un laser de longueur d'onde 266 nm. La forte luminescence avec un maximum autour de 330 nm correspond à la fluorescence des résidus aromatiques. L'incision vers 355 nm est un artefact du détecteur.

Le tableau 3.5.3 résume les caractéristiques de la luminescence des résidus aromatiques. On remarque les valeurs assez élevées du coefficient d'absorption et du rendement de fluorescence, en particulier pour le tryptophane.

|               | Absorption           |                                  | Fluorescence         |                   |                           |
|---------------|----------------------|----------------------------------|----------------------|-------------------|---------------------------|
|               | Longueur d'onde [nm] | Coefficient d'absorption molaire | Longueur d'onde [nm] | Durée de vie [ns] | Rendement de fluorescence |
| Tryptophane   | 280                  | 5600                             | 348                  | 2,6               | 0,20                      |
| Tyrosine      | 274                  | 1400                             | 303                  | 3,6               | 0,14                      |
| Phénylalanine | 257                  | 200                              | 282                  | 6,4               | 0,04                      |

Tableau 3.5.3. : Caractéristiques de la fluorescence des résidus aromatiques (Brooks, 2005).

### III.5.7. Résultats d'expériences préliminaires avec des ions de lanthanides

Comme nous ne disposons pas de complexes d' $\text{Eu}^{3+}$  ou de  $\text{Tb}^{3+}$  nous avons d'abord procédé à des expériences préliminaires avec des ions de lanthanides.

#### III.5.7.a. Interaction des ions $\text{Ln}^{3+}$ avec les protéines

Il est connu que la glucose isomérase lie fortement les ions de lanthanides sur ses sites de fixation des métaux divalents Ca, Mn ou Mg (Rey *et al.*, 1988). Le lysozyme, quant à lui, ne lie pas les ions de lanthanides.

#### III.5.7.b. Préparation des échantillons

Les cristaux dérivés de protéine avec les ions de lanthanides ont été préparé en ajoutant de la solution de chlorure d'euprium,  $\text{EuCl}_3$ , ou de nitrate de terbium,  $\text{Tb}(\text{NO}_3)_3$ , à la goutte contenant les cristaux natifs de protéine. Les volumes ajoutés ont été calculés pour obtenir une concentration de 1 à 10 mM de sel de lanthanide dans la goutte. Les gouttes ont été mises en équilibre avec le puits pendant plusieurs jours. Lors de la manipulation des échantillons contenant des sels de lanthanides, j'ai bien fait attention d'éviter toute contamination de personne ou de matériel, car, contrairement aux complexes qu'on utilise habituellement, les sels des lanthanides sont toxiques.

#### III.5.7.c. Expériences de luminescence avec l'ion $\text{Tb}^{3+}$

La figure 3.5.4 représente le spectre enregistré avec une solution de  $\text{Tb}^{3+}$  à une concentration de 50 mM. Ce spectre présente les raies d'émission caractéristiques du  $\text{Tb}^{3+}$ . La raie d'émission à 580 nm est dédoublée. La figure 3.5.5 représente le spectre enregistré avec une solution contenant la protéine glucose isomérase et des ions de terbium. On y distingue toujours clairement les raies d'émission du terbium.

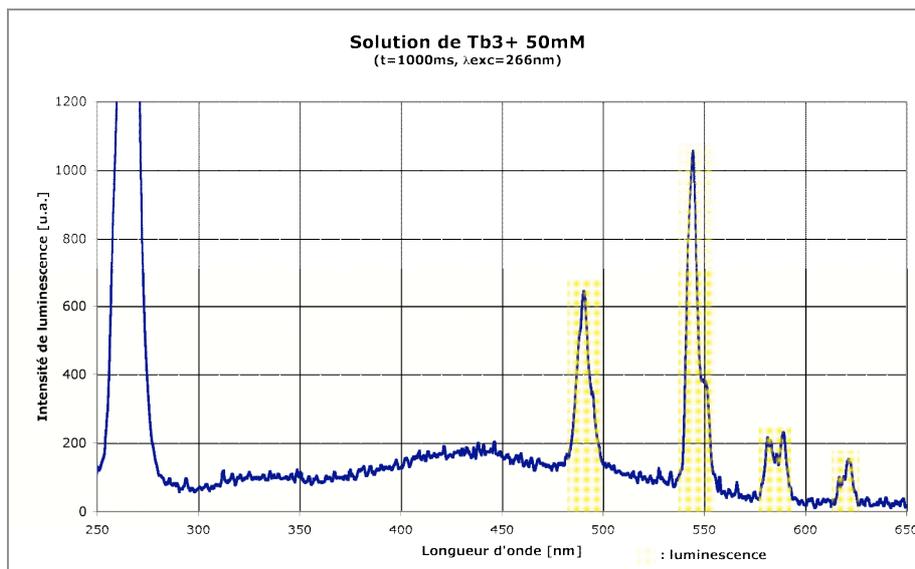


Figure 3.5.4. : Spectre d'émission d'une solution de  $Tb^{3+}$ . Les raies d'émission sont situées aux longueurs d'onde 490, 544, 583 et 621 nm.

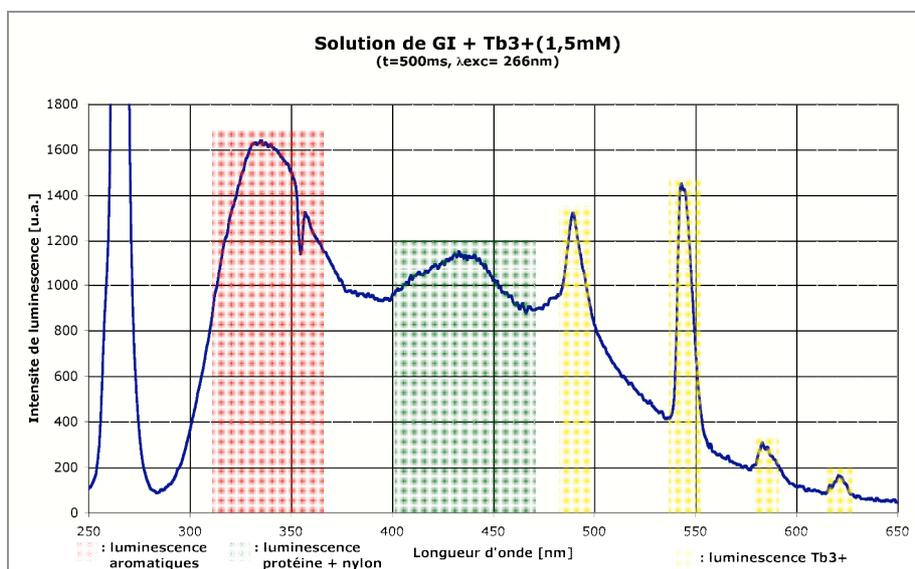


Figure 3.5.5. : Spectre d'émission d'une solution de glucose isomérase avec 1,5 mM de  $Tb^{3+}$ . Les raies d'émission du  $Tb^{3+}$  sont superposées au spectre d'émission de la protéine, mais se distinguent clairement.

### III.5.7.d. Résultats des expériences avec les ions $Tb^{3+}$

Avec le laser à 266 nm, on excite dans une zone de forte absorption (voir figure 3.5.6). Ainsi, nous avons une forte excitation directe (figure 3.5.4) des ions  $Tb^{3+}$  en solution qui conduit à des raies d'émission très prononcées, malgré le faible rendement de luminescence des ions en solution aqueuse. Les raies d'émission sont situées aux longueurs d'onde de 490, 544, 583 et 621 nm, avec des raies dédoublées à 544 et à 621 nm.

Les raies d'émission des ions en présence de la glucose isomérase, qui lie les ions, (figure 3.5.5) sont situées aux mêmes longueurs d'onde et sont à peu près aussi intense que les raies du spectre de la figure 3.5.4. Cependant, le temps d'acquisition du spectre enregistré avec la protéine est deux fois plus court et la concentration des ions  $Tb^{3+}$  est 33 fois moins élevée que pour le spectre de l'ion en solution. Que l'intensité de luminescence soit aussi forte pour le spectre enregistré avec la protéine peut avoir deux causes : a) la liaison des ions  $Tb^{3+}$  par les résidus protéiques qui remplacent des molécules d'eau de la sphère de coordination, ce

qui augmente le rendement de luminescence des ions et, éventuellement, b) l'émission amplifiée par transfert d'énergie des tryptophanes de la protéine (la glucose isomérase a huit tryptophanes dans sa séquence, dont deux à proximité des sites de fixation des ions, à 5,5 et 6,8 Å respectivement). Nous ne savons pas quelle est la raison du dédoublement des raies d'émission du spectre de l'ion en solution.

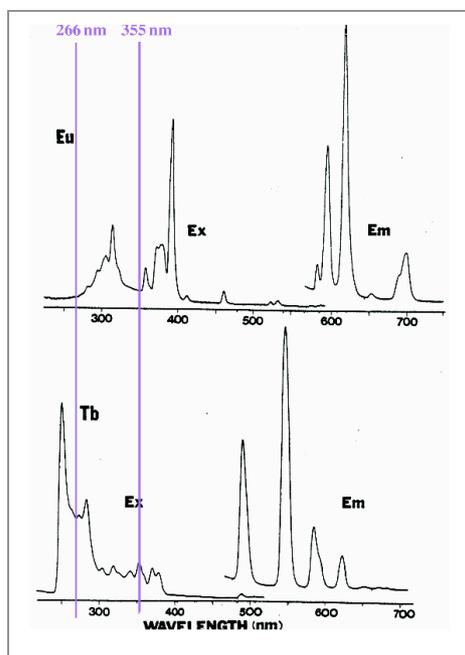


Figure 3.5.6. : Spectres d'excitation et d'émission pour l'euprimum et le terbium enregistrés sur des solutions de [LnEDTA]<sup>-</sup> (Horrocks & Albin, 1991).

### III.5.7.e. Expériences de luminescence avec l'ion $\text{Eu}^{3+}$

La comparaison des spectres obtenus avec deux cristaux de protéine trempés dans une solution de 10 mM d' $\text{Eu}^{3+}$  et irradiés avec de la lumière à 355 nm, montre que le signal obtenu avec le cristal dans lequel les ions de lanthanides sont fortement liés (figure 3.5.8) montre des raies de luminescence de l'ion beaucoup plus fortes que le signal obtenu pour le cristal dans lequel les ions ne sont pas liés (figure 3.5.7).

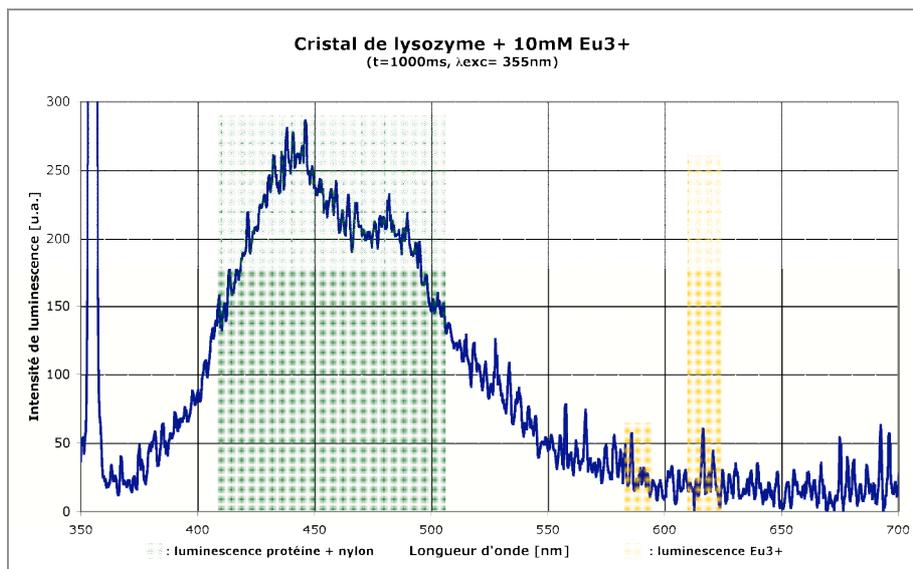


Figure 3.5.7. : Spectre de fluorescence d'un cristal de lysozyme trempé dans une solution contenant 10 mM d'Eu<sup>3+</sup>, irradié avec un laser de longueur d'onde 355 nm.

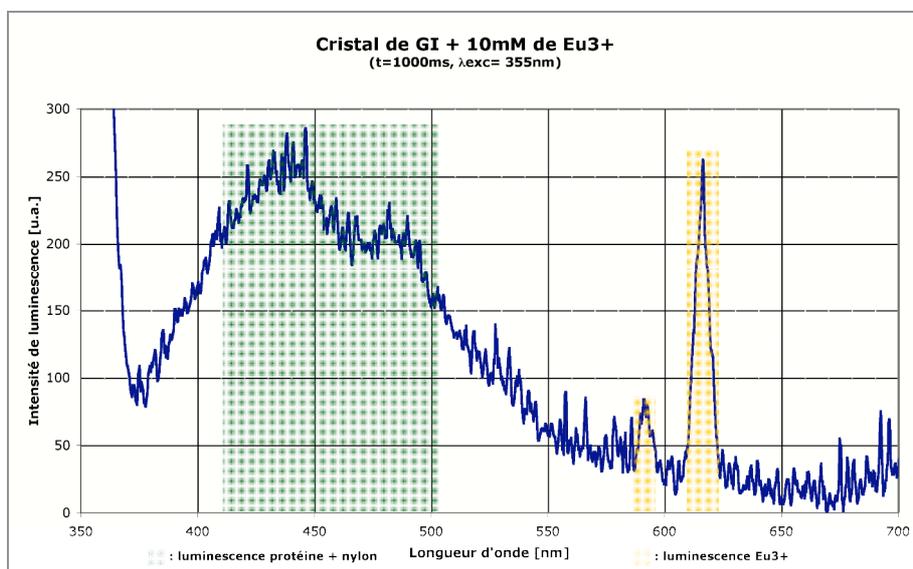


Figure 3.5.8. : Spectre de fluorescence d'un cristal de glucose isomérase trempé dans une solution contenant 10 mM d'Eu<sup>3+</sup>, irradié avec un laser de longueur d'onde 355 nm.

L'amplification du signal de luminescence due à la fixation des ions dans le cristal de protéine peut avoir deux causes : a) l'augmentation de la concentration des ions dans le cristal où les ions sont liés et b) le remplacement des molécules d'eau de la sphère de coordination des ions liés par des atomes des résidus protéiques.

### III.5.8. Expériences avec les complexes de lanthanides

Alors qu'avec les ions de lanthanides, il est facile de détecter leur fixation sur des protéines, ceci s'avère moins facile avec les complexes. Le rendement de luminescence des ions complexés est plus fort que celui des ions en solution aqueuse et l'effet d'amplification du signal lors de la fixation du complexe, dû au remplacement des molécules d'eau, est donc beaucoup plus faible, sinon inexistant. L'affinité beaucoup plus faible des complexes pour les protéines, comparée à celle des ions pour la glucose isomérase par exemple, fait que les

différences de concentration entre les molécules de complexe liées et non-liées seront également plus faibles.

Les résultats obtenus avec les deux seuls complexes d'Eu<sup>3+</sup> dont nous disposons et avec plusieurs complexes de Gd<sup>3+</sup> sont présentés ci-dessous.

### III.5.8.a. Expériences avec les complexes d'euporium

La figure 3.5.9 montre les raies d'émission d'une solution de Eu-DOTA irradiée avec de la lumière ultraviolette. Le spectre présente des raies étroites et intenses qui correspondent exactement aux spectres donnés dans la littérature pour des complexes d'euporium de même type (Anelli *et al.*, 1991). On remarque la différence du spectre du complexe par rapport au spectre de l'ion (irradié à 355 nm), ce dernier ne présente que la raie très intense vers 615 nm et une raie beaucoup moins forte vers 590 nm.

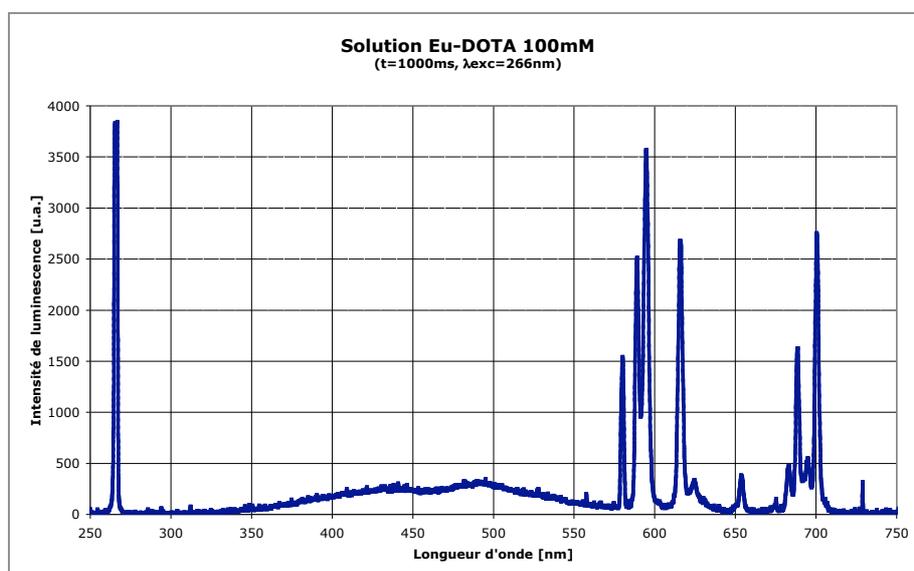


Figure 3.5.9. : Spectre de fluorescence d'une solution du complexe Eu-DOTA à 100 mM.

### III.5.8.b. Préparation des cristaux dérivés

Les cristaux dérivés de glucose isomérase ont été préparés par trempage des cristaux natifs dans des solutions contenant 100 mM du complexe de lanthanide. Les cristaux dérivés de lysozyme ont été préparés par cocristallisation avec 100 mM du complexe correspondant.

### III.5.8.c. Spectres obtenus

Les figures 3.5.10 et 3.5.11 montrent les spectres de fluorescence des cristaux dérivés du lysozyme et de la glucose isomérase avec le complexe Eu-DOTA. Le lysozyme lie le complexe Eu-DOTA avec des taux d'occupation relativement élevés alors que la glucose isomérase ne lie pas le complexe. Dans le lysozyme, le complexe se fixe dans des sites très proches de tryptophanes (voir chapitre II.4.8). Dans les deux spectres, les raies d'émission du complexe se détachent bien de la fluorescence de fond et correspondent au spectre observé pour le complexe seul. Malheureusement, une comparaison quantitative des deux spectres n'est pas possible, car l'intensité du signal dépend de l'orientation du cristal. On ne peut pas non plus se référer à la hauteur du pic de fluorescence des résidus aromatiques, car ceux-ci se dégradent très vite sous irradiation avec de la lumière ultraviolette (voir figures 3.5.15).

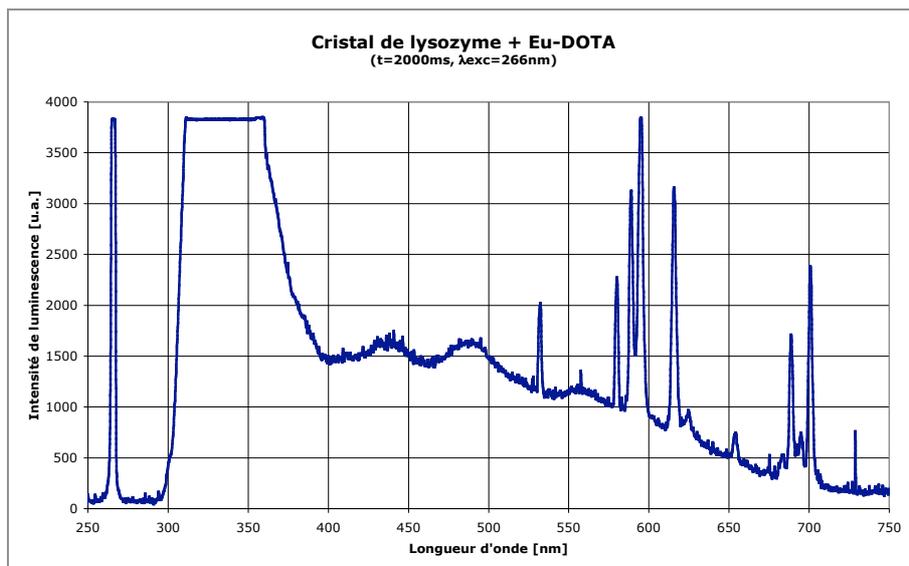


Figure 3.5.10. : Spectre de fluorescence d'un cristal de lysozyme cocristallisé avec 100 mM de Eu-DOTA. Les émissions de fluorescence dans le domaine ultraviolet saturent le détecteur. La raie à 532 nm est un artefact expérimental et correspond à deux fois la longueur d'onde excitatrice.

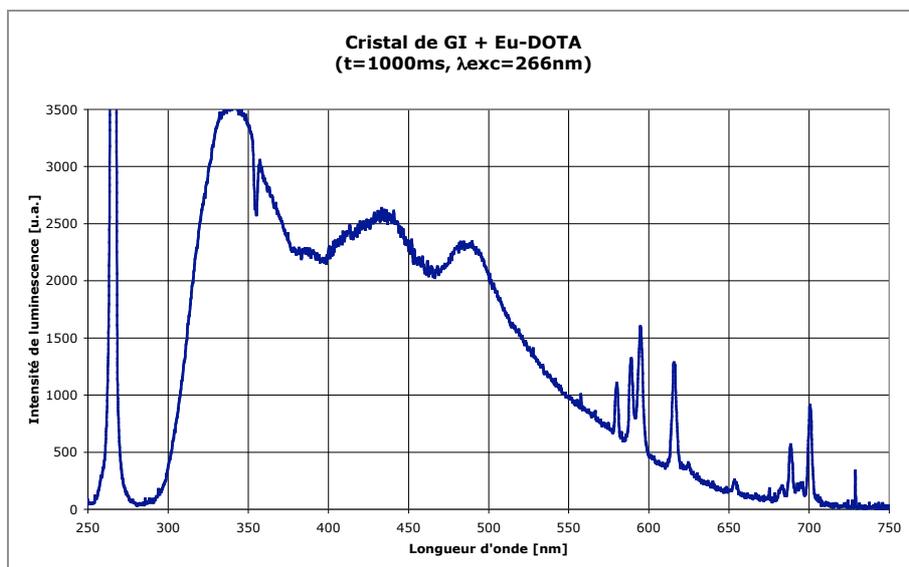


Figure 3.5.11. : Spectre de fluorescence d'un cristal de glucose isomérase trempé dans 100 mM de Eu-DOTA.

La figure 3.5.12 montre le signal de fluorescence d'un cristal dérivé de lysozyme obtenu avec 100 mM du complexe Eu-HPSA-DO3A. Le spectre du complexe est identique au spectre du complexe Eu-DOTA excité à la même longueur d'onde et se détache également bien de la fluorescence de fond. Le complexe Eu-HPSA-DO3A ne se lie pas au lysozyme.

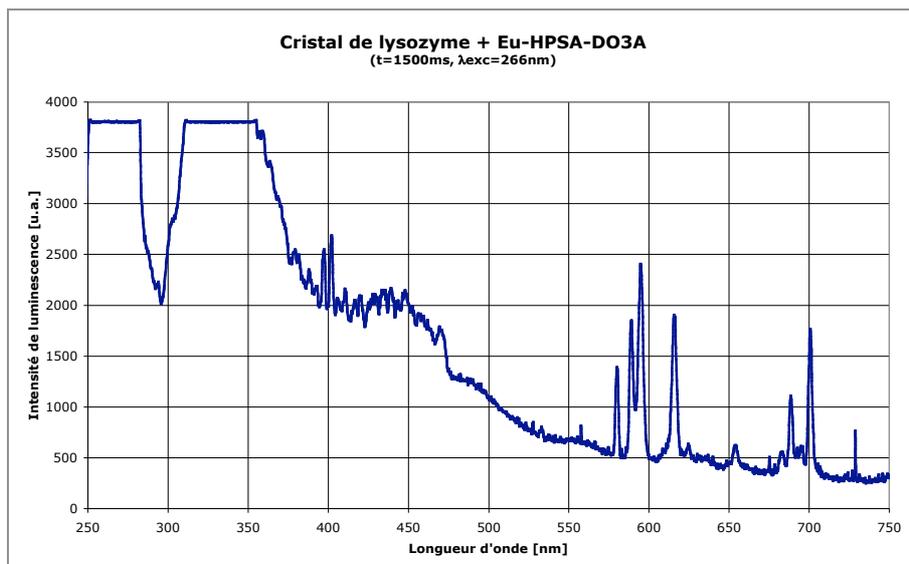


Figure 3.5.12. : Spectre de fluorescence d'un cristal de lysozyme cocristallisé avec 100 mM du complexe Eu-HPSA-DO3A. La fluorescence entre 350 et 450 nm correspond à des impuretés sur la boucle de mylar utilisée.

Tous les spectres présentés ci-dessus ont été obtenus en irradiant avec le laser ultraviolet à 266 nm. Les expériences avec le laser à 355 nm ou avec la source de lumière blanche n'ont pas permis d'obtenir des spectres avec des raies d'émission de l'euprium.

### III.5.8.d. Expériences avec les complexes de gadolinium

Les figures 3.5.13 et 3.5.14 montrent les spectres de luminescence de solutions à haute concentration des complexes Gd-DTPA et Gd-DOTA. Les deux spectres présentent un seul pic d'émission étroit à 312 nm. On remarque l'intensité relativement faible du pic d'émission pour les deux complexes.

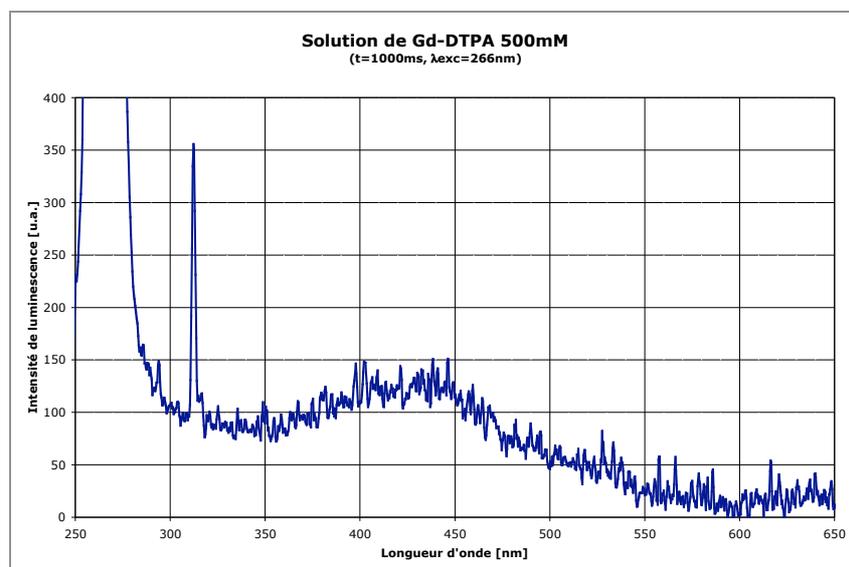


Figure 3.5.13. : Spectre de fluorescence enregistré avec une solution de Gd-DTPA à une concentration de 500mM.

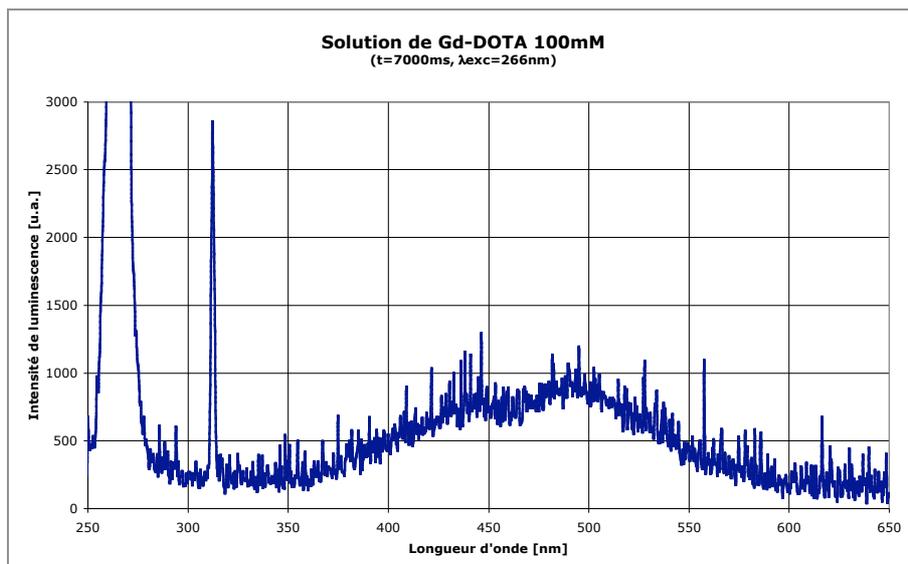


Figure 3.5.14. : Spectre de fluorescence enregistré avec une solution de Gd-DOTA à une concentration de 100 mM.

La figure 3.5.15 montre les spectres enregistrés avec un cristal dérivé de lysozyme cocrystallisé avec 100 mM de Gd-DOTA. Dans aucun des trois spectres, on ne distingue la raie d'émission du  $Gd^{3+}$ . En effet, la raie dans le domaine ultraviolet est entièrement dominée par l'émission des résidus aromatiques, cette dernière étant beaucoup plus intense que la première.

### III.5.8.e. Dégâts d'irradiation III - inconvénient de l'irradiation dans l'UV

La figure 3.5.15 montre la baisse de l'émission de fluorescence des résidus aromatiques au cours de l'irradiation avec une lumière ultraviolette. Cette baisse est due à la dégradation très rapide des résidus aromatiques. En effet, l'échantillon se dégrade en quelques secondes d'irradiation avec le laser à 266 nm.

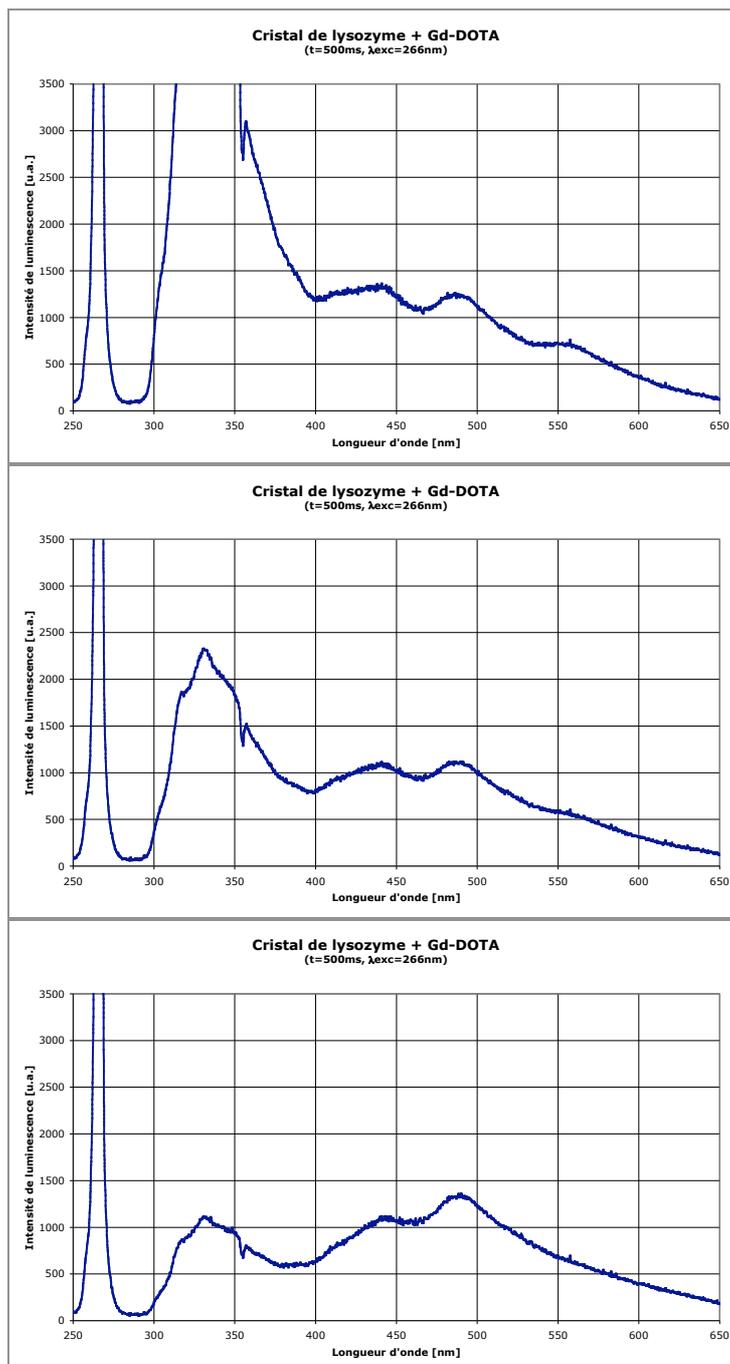


Figure 3.5.15. : Les spectres sont enregistrés sur un même cristal de lysozyme cocrystallisé avec 100 mM de Gd-DOTA. On voit la dégradation des résidus aromatiques due à l'irradiation du cristal avec de la lumière ultraviolette. Entre les différents spectres le cristal n'est irradié que pendant quelques secondes.

Les figures 3.5.16 et 3.5.17 montrent les spectres enregistrés avec des échantillons contenant le complexe Gd-DOTA-BOM. Le spectre de fluorescence de ce complexe est très différent des autres complexes grâce au groupement (phénylméthoxy)méthyle du ligand. Dans le spectre enregistré avec un cristal dérivé de lysozyme avec le complexe Gd-DOTA-BOM on distingue aisément la fluorescence intense du complexe de la fluorescence de la protéine. Le complexe Gd-DOTA-BOM se lie assez fortement au lysozyme dans les cristaux.

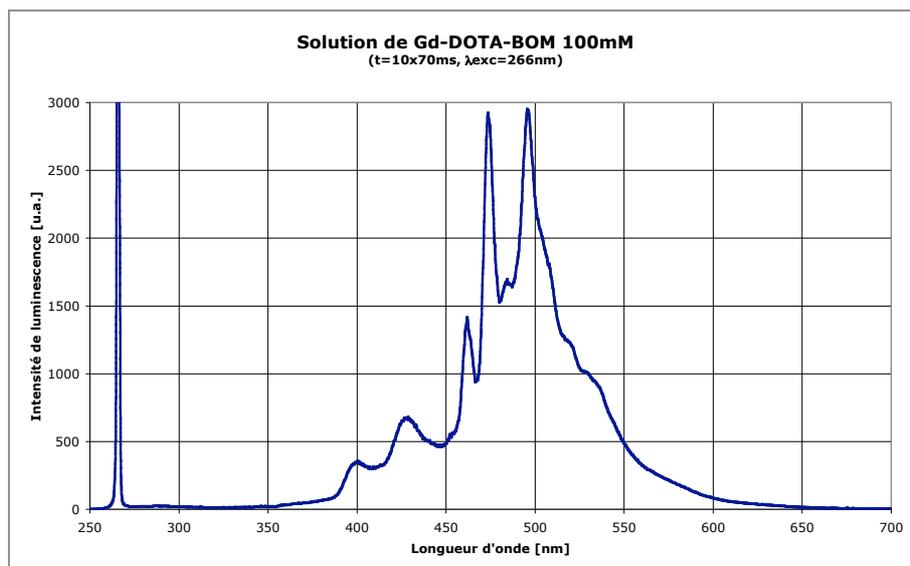


Figure 3.5.16. : Spectre de fluorescence enregistré sur une solution du complexe Gd-DOTA-BOM à une concentration de 100 mM.

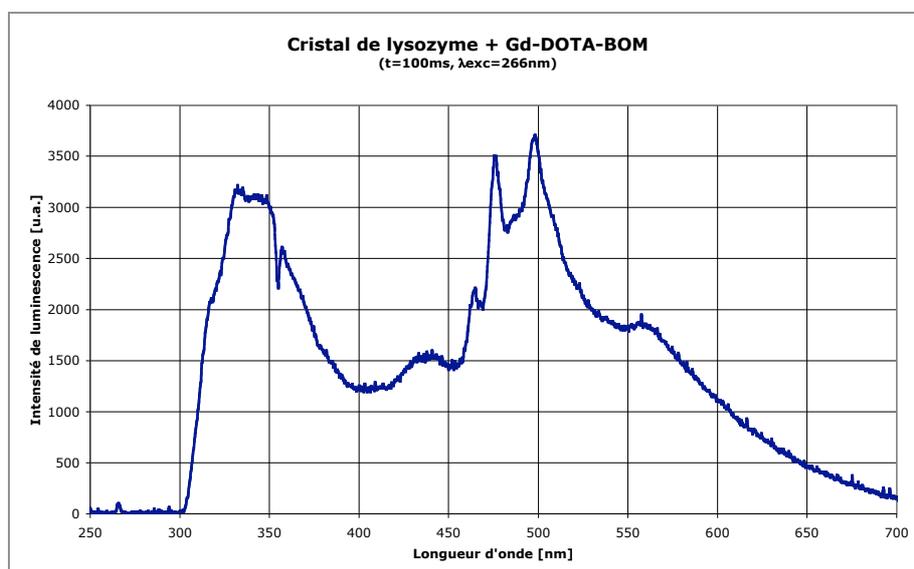


Figure 3.5.17. : Spectre de fluorescence enregistré avec un cristal de lysozyme cocrystallisé avec 100 mM de Gd-DOTA-BOM.

### III.5.9. Conclusions et perspectives pour les expériences de luminescence des lanthanides

À Grenoble, il existe plusieurs groupes de recherche spectroscopique qui utilisent des montages permettant de faire des expériences de luminescence des lanthanides sous des conditions différentes de celles du microspectrophotomètre du "cryobench". Un des montages expérimentaux permet, entre autre, de choisir la longueur d'onde de lumière incidente. Ce montage ne permet que des expériences sur des échantillons en solution, et avec un volume beaucoup plus élevé. La géométrie du montage permet de faire des mesures quantitatives de l'intensité de luminescence et on peut faire des expériences en temps résolu (Spectrophotomètres UV-visible du Laboratoire d'Electrochimie Organique et Photochimie Redox (LEOPR), Université Joseph Fourier Grenoble).

Dans notre équipe et en collaboration avec l'équipe de Patrice Baldeck du Laboratoire de Spectrométrie Physique de l'Université Joseph Fourier de Grenoble des expériences sur la luminescence d'un complexe d'euporium dans des cristaux de protéine sont en cours. Le complexe utilisé présente des caractéristiques proches des complexes décrits dans ce travail, sauf qu'il n'y a aucune molécule d'eau dans la sphère de coordination de l'ion  $\text{Eu}^{3+}$ . Ces expériences de luminescence permettent de déterminer la concentration du complexe dans l'échantillon. Une expérience au cours de laquelle le cristal dérivé est trempé dans une solution de complexe dans lequel l'ion  $\text{Eu}^{3+}$  a été échangé contre un ion non-luminescent, et l'analyse de la dynamique de la décroissance du signal de luminescence du complexe dans le cristal, pourrait permettre de distinguer entre les molécules de complexe des canaux de solvant du cristal et des molécules de complexes liées à la protéine. Les études de la luminescence de cristaux trempés dans différentes solutions de complexe permettent en outre d'étudier la dynamique de diffusion et de fixation des complexes dans les cristaux.

### III.6. Conclusions sur les différentes méthodes physico-chimiques testées

Bien que nous n'ayons pas trouvé de méthode qui permette de détecter la fixation d'un complexe sur une protéine, je vais comparer les caractéristiques des différentes méthodes testées.

Afin qu'une méthode puisse être appliquée dans un contexte plus général, il faut qu'elle soit facilement applicable, qu'elle consomme peu de protéine et que les appareils nécessaires soient facilement accessibles.

Le tableau 3.6.1 résume les avantages et les inconvénients des différentes méthodes testées.

| Méthode      | $m_{\text{protéine}}$ [mg] | État               | Matériel nécessaire | Succès |
|--------------|----------------------------|--------------------|---------------------|--------|
| Calorimétrie | ~30                        | solution           | -                   | non    |
| Gel          | ~ $10^{-2}$                | solution           | +                   | ?      |
| RMN          | ~10                        | solution           | -                   | ?      |
| Luminescence | ~ $10^{-6}$ à 1            | cristal & solution | -                   | ?      |

Tableau 3.6.1 : Résumé des avantages et des inconvénients des différentes méthodes testées.  $m_{\text{protéine}}$  : masse de protéine nécessaire pour tester un complexe. Pour la luminescence, la quantité peut varier de plusieurs ordres de grandeur en fonction du montage expérimental utilisé. Pour le matériel nécessaire un "-" signifie que les expériences nécessitent un équipement particulier et, le plus souvent, difficilement accessible, le signe "+" signifie que le matériel nécessaire est facilement accessible dans un laboratoire de biochimie classique.

Aucune des méthodes testées n'a permis, dans les conditions utilisées, de distinguer un dérivé avec un taux de fixation du complexe élevé d'un dérivé avec un taux de fixation bas. La calorimétrie ne permet clairement pas d'étudier les interactions caractérisées par d'aussi faibles affinités. La RMN est une méthode qui est difficile à mettre en œuvre et dont l'analyse des résultats est également difficile. Il reste à déterminer si, en optimisant les conditions d'utilisation, l'utilisation de gels de polyacrylamide natifs pourrait permettre la détection de la fixation des complexes chargés. Cependant, ces conditions devraient probablement être réadaptées pour chaque nouvelle protéine. La luminescence est la seule méthode qui se réalise aussi bien avec les cristaux qu'avec la solution de protéine. Elle permet l'étude de phénomènes associés à la diffusion de complexes de lanthanides dans les cristaux de protéine, mais, même en optimisant les conditions expérimentales, il semble peu probable qu'elle puisse permettre la détection de la fixation d'un complexe dans une protéine.

Un problème associé aux études de la protéine en solution est que l'interaction du complexe avec la protéine n'est pas nécessairement la même en solution que dans le cristal : certaines cavités qui permettent de piéger une molécule de complexe ne résultent que de l'empilement cristallin, alors que d'autres sites de fixation, accessibles en solution, se trouvent dans des zones de surface impliquées dans des contacts intermoléculaires dans le cristal, ce qui empêche la fixation du complexe dans ces sites. Le mode de fixation affiné des complexes dans les dérivés de lysozyme (chapitre II.4.8) laisse supposer que l'empilement cristallin privilégie la fixation des complexes. Nous avons couramment utilisé le lysozyme comme protéine modèle, mais il semble que, pour les méthodes de caractérisation en solution, il serait plus judicieux de travailler avec d'autres protéines.



**Partie IV.**  
**Conclusions**



## IV. Conclusions et perspectives

Le fort potentiel des complexes pour la préparation de dérivés présentant un fort pouvoir de phasage, tant avec un générateur de laboratoire qu'avec le rayonnement synchrotron, a été confirmé.

Dans de nombreux cas, des cristaux dérivés obtenus grâce à l'utilisation des complexes ont mené à des cartes de densité électroniques expérimentales d'excellente qualité. Ces cartes permettent non seulement de construire aisément le modèle de la protéine, mais également de placer d'éventuels ligands (voir par exemple le cas de l'inhibiteur  $\delta$ -azaxanthine dans la protéine urate oxydase, figure 2.3.2 du § II.3.6).

Outre leurs propriétés de fixation dans les cristaux et leur forte diffusion anormale, l'avantage des complexes est leur facilité d'utilisation et leur bonne tolérance par les protéines. La complémentarité du mode fixation des différents complexes sur les protéines fait que, en testant plusieurs d'entre eux, on augmente les chances d'obtenir un dérivé avec un taux de fixation élevé.

Les résultats des affinements du mode de fixation observé pour différentes combinaisons de complexe et de protéine permettent, dans certains cas, d'orienter le choix du complexe à utiliser avec une nouvelle protéine. Toutefois, ces résultats ne permettent pas, dans le cas général, de prédire de manière certaine la fixation d'un complexe sur une nouvelle protéine.

Dans les cas étudiés, les méthodes alternatives à la cristallographie qui permettent de caractériser l'interaction d'une protéine avec un ligand n'ont pas permis d'estimer le taux de fixation d'un complexe avec la protéine utilisée.

Néanmoins, l'utilisation des complexes par d'autres groupes de cristallographie biologique, comme alternative à des méthodes établies de préparation de dérivés lourds, serait tout à fait souhaitable. Dans ce dessein, il faudrait rendre facilement accessible l'ensemble des complexes par des équipes intéressées.

Ceci n'est pas le cas actuellement (voir § I.5.2)\*\*. Il serait souhaitable, que l'ensemble des complexes soit commercialisé sous forme de crible qui permette de tester les différents complexes pour l'obtention d'un dérivé lourd. Dans ce dessein, le fournisseur de matériel de cristallisation Hampton Research a contacté notre équipe dans le passé. Les négociations qui ont suivi entre les entreprises Hampton Research et Bracco Imaging n'ont pas abouti à la commercialisation d'un tel crible. La levée, récente, d'un certain nombre de brevets protégeant la synthèse des complexes pourrait permettre la relance d'une opération similaire.

Un tel crible, en combinaison avec des procédures de traitements de données de diffraction rapides, automatisés qui permettent l'estimation du pouvoir de phasage d'un dérivé pendant l'acquisition des données de diffraction (Panjikar *et al.*, 2005), permettrait l'obtention efficace de dérivés avec un bon pouvoir de phasage, leur vérification et, le cas échéant, la mise à l'écart de dérivés qui ne permettent pas d'obtenir les phases.

Les résultats des études sur l'obtention des dérivés et sur les modes de fixation des complexes analysés à partir des structures affinées, pourraient aider à concevoir de nouveaux ligands d'atomes lourds, optimisés pour la fixation dans les cristaux de protéines.

## Références

- J. Abrahams (1997). "Bias reduction in phase refinement by modified interference functions: introducing the gamma correction." *Acta Cryst. D***52**, 43-48.
- S. Aime, E. Gianolio, E. Terreno, G. Giovenzana, R. Pagliarin, M. Sisti, G. Palmisano, M. Botta, M. Lowe & D. Parker (2000). "Ternary Gd(III)L-HSA adducts: evidence for the replacement of inner-sphere water molecules by coordinating groups of the protein. Implications for the design of contrast agents for MRI." *JBIC*, **5**, 488-497.
- P. Anelli, V. Balzani, L. Prodi & F. Uggeri (1991). "Luminescence properties of Eu<sup>3+</sup> complexes of highly polydentate ligands." *Gazetta Chimica Italiana*, **121**, 359-364.
- A. Bianchi, L. Calabi, F. Corana, S. Fontana, P. Losi, A. Maiocchi, L. Paleari & B. Valtancoli (2000). "Thermodynamic and structural properties of Gd(III) complexes with polyamino-polycarboxylic ligands: basic compounds for the development of MRI contrast agents." *Coord. Chem. Rev.* **204**, 309-393.
- K. Binnemans & C. Görller-Walrand (1995). "On the color of the trivalent lanthanide ions." *Chemical Physics Letters*, **235**, 163-174.
- D. Blow & F. Crick (1959). *Acta Cryst.* **12**, 794-802.
- T. Boggon & L. Shapiro (2000). "Screening for phasing atoms in protein crystallography." *Structure*, **8**, R143-R149.
- F. Bonneté, D. Vivarès, C. Robert & N. Colloc'h (2001). "Interactions in solution and crystallization of *Aspergillus flavus* urate oxidase" *J. Crystal Growth* **232**, 330-339.
- D. Bourgeois, X. Vernede, V. Adams, E. Fioravanti & T. Ursby (2002). "A microspectrophotometer for UV-visible absorption and fluorescence studies of protein crystals." *J. Appl. Cryst.* **35**, 319-326.
- G. Bricogne (1988). "A Bayesian statistical theory of the phase problem. I. A multichannel maximum-entropy formalism for constructing generalized joint probability distributions of structure factors." *Acta Cryst. A***44**, 517-545.
- G. Bricogne, C. Vonrhein, C. Flensburg, M. Schiltz & W. Paciorek (2003). "Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0." *Acta Cryst. D***59**, 2023-2030.
- D. Brooks (2005). <http://dwb.unl.edu>.
- M. Brown, R. Peierls & E. Stern (1997). "White lines in x-ray absorption." *Physical Review B*. **15**, **2**, 738-744.
- A. Brünger (1992). "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." *Nature* **355**, 472-475

A. Brünger, F. Temple-Burling, P. Grost, R. Grosse-Kunstleve, J. Jiang, J. Kuszewski, M. Nilges, N. Pannu, R. Read, L. Rice, T. Simonson & G. Warren (1998). "Crystallography & NMR System: A new software suite for macromolecular structure determination." *Acta Cryst.* **D54**, 905-921.

J. Bünzli & J. Pfefferle (1994). "Bovine  $\alpha$ -Lactalbumin: Identification of two metal-ion-binding sites using the europium (III) luminescent probe." *Helv. Chim. Acta*, **77**, 323.

W. Burmeister (2000). "Structural changes in a cryo-cooled protein crystal owing to radiation damage." *Acta Cryst.* **D56**, 328-341.

H. Carrell, H. Hoier & J. Glusker (1994). "Modes of binding substrates and their analogs to the enzyme D-xylose isomerase." *Acta Cryst.* **D50**, 113-123.

C. Carubelli, A. Massabni & S. Leite (1997). "Study of the binding of  $\text{Eu}^{3+}$  and  $\text{Tb}^{3+}$  to L-phenylalanine and L-tryptophan." *J. Braz. Chem. Soc.*, **8**, 597-602.

C. Chang, L. Francesconi, M. Malley, K. Kumar, J. Gougoutas, M. Tweedle, D. Lee & L. Wilson (1993). "Synthesis, Characterization, and Crystal Structures of  $\text{M}(\text{DO3A})$  ( $\text{M} = \text{Fe}, \text{Gd}$ ) and  $\text{Na}[\text{M}(\text{DOTA})]$  ( $\text{M} = \text{Fe}, \text{Y}, \text{Gd}$ )." *Inorg. Chem.* **32**, 3501-3508.

C. Charron, R. Giegé & B. Lorber (2004). "Structure of thaumatin in a hexagonal space group: comparison of packing contacts in four crystal lattices." *Acta Cryst.* **D60**, 83-89.

M.-L. Chesne (2002). "Utilisation de la diffusion anormale de l'uranium pour déterminer la structure de macromolécules biologiques - des rayons X durs aux rayons X mous." *Thèse de doctorat, Université Joseph Fourier, Grenoble*.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760-763.

K. Cowtan & P. Main (1996). "Phase combination and cross validation in iterated density-modification calculations" *Acta Cryst.* **D52**, 43-48.

Z. Dauter (2002). "New approaches to high-throughput phasing." *Current opinion in Structural Biology*, **12**:674-678.

Z. Dauter, M. Dauter & E. Dodson (2002). "Jolly SAD." *Acta Cryst* **D58**, 494-506.

Z. Dauter & R. Nagem (2002). "Direct way to anomalous scatterers." *Z. Kristallogr.* **217**, 694-702

W. DeLano (2002). "The PyMOL Molecular Graphics System on World Wide Web." <http://www.pymol.org>

J.-P. Dubost, J.-M. Leger, M.-H. Langlois, D. Meyer & M. Schaefer (1991). "Structure d'un agent de contraste utilisé en imagerie de résonance magnétique le complexe  $\text{DOTA Gd C}_{16}\text{H}_{24}\text{N}_4\text{O}_8\text{NaGd}, 5\text{H}_2\text{O}$ ." *C. R. Acad. Sci. Paris*, **t. 312, Série II**, 349-354.

- L. Ehnebom & B. Pedersen (1992). "Molecular and Crystal Structure of a Lanthanide Complex, Dy-DTPA-BMA Hydrate." *Acta Chemica Scandinavica* **46** 126-130.
- M. Elbanowski & B. Makowska (1996). "The lanthanides as luminescent probes in investigations of biochemical systems." *Journal of Photochemistry and Photobiology A: Chemistry*, **99**, 85-92.
- E. Ennifar, P. Carpentier, J.-L. Ferrer, P. Walter, P. Dumas (2002). "X-ray-induced debromination of nucleic acids at the Br K absorption edge and implications for MAD phasing." *Acta Cryst.* **D58**, 1262-1268.
- G. Evans & G. Bricogne (2002). "Triiodide derivatization and combinatorial counter-ion replacement: two methods for enhancing phasing signal using laboratory Cu K $\alpha$  X-ray equipment" *Acta Cryst.* **D58**, 976-991.
- G. Evans & R. Pettifer (2001). "CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra." *J. Appl. Cryst.* **34**, 82-86.
- A. Feig, M. Panek, W. Horrocks & O. Uhlenbeck (1999). "Probing the binding of Tb(III) and Eu(III) to the hammerhead ribozyme using luminescence spectroscopy." *Chemistry & Biology*, **6**, 801-810.
- E. Garcia, J. Garcia, P. Garcia, A. Arraras, J. Sanchez-Puelles & R. Lopez (1988). "Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages." *Proc Natl Acad Sci U S A*, **85**(3), 914-918.
- E. Garman & J. Murray (2003). "Heavy-atom derivatization." *Acta Cryst.* **D59**, 1903-1913.
- E. Garman & C. Nave (2002). "Radiation damage to crystalline biological molecules: current view." *J. Synchrotron Rad.* **9**, 327-328.
- É. Girard (2001). "Possibilités nouvelles d'études structurales de macromolécules biologiques : utilisation de la diffusion anormale avec des complexes de lanthanides." *Thèse de doctorat, Université Joseph Fourier, Grenoble*.
- É. Girard, P. Anelli, J. Vicat & R. Kahn (2003c). "High-phasing-power lanthanide derivatives: taking advantage of ytterbium and lutetium for optimized anomalous diffraction experiments using synchrotron radiation." *Acta Cryst.* **D59**, 1877-1880.
- É. Girard, L. Chantalat, J. Vicat & R. Kahn (2002). "Gd-HPDO3A, a complex to obtain high-phasing-power heavy-atom derivatives for SAD and MAD experiments: results with tetragonal hen egg-white lysozyme." *Acta Cryst.* **D58**, 1-9.
- É. Girard, M. Stelter, P. L. Anelli, J. Vicat & R. Kahn (2003a). "A new class of gadolinium complexes employed to obtain high-phasing-power heavy-atom derivatives: results from SAD experiments with hen egg-white lysozyme and urate oxidase *Aspergillus flavus*." *Acta Cryst.* **D59**, 118-126.

É. Girard, M. Stelter, J. Vicat & R. Kahn (2003b). "A new class of lanthanide complexes to obtain high-phasing-power heavy-atom derivatives for macromolecular crystallography" *Acta Cryst.* **D59**, 1914–1922.

S. Gras, B. Fernandez, V. Chaumont, P. Carpentier, J. Armengaud & D. Housset (2005). "Expression, purification, crystallization and preliminary crystallographic analysis of the PAB0955 gene product." *Acta Cryst.* **F61**, 208-211.

D. Green, V. Ingram & M. Perutz (1954). "The structure of haemoglobin IV. Sign of determination by isomorphous replacement method." *Proc. R. Soc. London SER. A* **225**, 287-307.

[www.hamptonresearch.com/support/guides/7100G.pdf](http://www.hamptonresearch.com/support/guides/7100G.pdf)

W. Hendrickson (1999). "Maturation of MAD phasing for the determination of macromolecular structures." *J. Synchrotron Rad.* **6**, 845-851.

W. Hendrickson, J. Horton & D. LeMaster (1990). "Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure." *EMBO J.*, **9**, 1665-1672.

W. Hendrickson & E. Lattman (1970). "Representation of phase probability distributions for simplified combination of independent phase information." *Acta Cryst.* **B26**, 136-143.

W. Hendrickson & C. Ogata (1997). "Phase determination from multiwavelength anomalous diffraction measurements." *Methods Enzymol.* **276**, 494-523.

W. Hendrickson & M. Teeter (1981). "Structure of the Hydrophobic Protein Crambin Determined Directly from Anomalous Scattering of Sulfur." *Nature*, **290**, 107-113.

J. Hermoso, L. Lagartera, A. González, M. Stelter, P. García, M. Martínez-Ripoll, J. García & M. Menéndez (2005). "Insights into pneumococcal pathogenesis from the crystal structure of the modular teichoic acid phosphorylcholine esterase Pce." *Nature Structural & Molecular Biology* **12**, 533-538.

W. Horrocks (1993). "Luminescence spectroscopy." *Methods Enzymol.*, **226**, 495-538.

W. Horrocks & M. Albin (1991). "Lanthanide Ion Luminescence in Coordination Chemistry and Biochemistry." *Prog. Inorg. Chem.*, **31**, 1-104.

W. Horrocks & W. Collier (1981). "Lanthanide ion luminescence probes. Measurement of distance between intrinsic protein fluorophores and bound metal ions: Quantitation of energy transfer between tryptophan and terbium (III) or europium (III) in the calcium-binding protein parvalbumin." *J. Am. Chem. Soc.*, **103**, 2856-2862.

J. Ibers, D. Templeton, B. Vainshtein, G. Bacon & K. Lonsdale (1985). "Atomic Scattering Factors." in *International Tables for X-ray Crystallography, III, Physical and Chemical tables*. Dordrecht (Holland): D. Reidel Publishing Company for International Union of Crystallography, 201-246.

S. Jeudy, M. Stelter, B. Coutard, R. Kahn & C. Abergel (2005). "Preliminary crystallographic analysis of the Escherichia coli YeaZ protein using the anomalous signal of a gadolinium derivative." *Acta Cryst.* **F61**, 848-851.

T. Jones, J. Zou, S. Cowan & M. Kjeldgaard (1991). "Improved methods for binding protein models in electron density maps and the location of errors in these models." *Acta Cryst.* **A47**, 110-119.

W. Kabsch (1993). "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants." *J. Appl. Cryst.* **26**, 795-800.

R. Kahn, P. Carpentier, C. Berthet-Colominas, M. Capitan, M.-L. Chesne, E. Fanchon, S. Lequien, D. Thiaudière, J. Vicat, P. Zielinski & H. Stuhmann (2000). "Feasibility and review of anomalous X-ray diffraction at long wavelengths in materials research and protein crystallography." *J. Synchrotron Rad.* **7**, 131-138.

R. Kahn, R. Fourme, R. Bosshard, M. Chiadmi, J.L. Risler, O. Dideberg, & J.P. Wery (1985). "Crystal structure study of Opsanus tau parvalbumin by multiwavelength anomalous diffraction." *FEBS lett.* **179(1)**, 133-137.

J. Karle & H. Hauptman (1956). "A theory of phase determination for the four types of non-centrosymmetric space groups 1P222, 2P22, 3P<sub>1</sub>2, 3P<sub>2</sub>2." *Acta Cryst.* **9**: 635-651.

G. Kleywegt & T. Jones (1998). "Databases in protein crystallography." *Acta Cryst* **D54**, 1119-1131.

T.-P. Ko, J. Day, A. Greenwood & A. McPherson (1994). "Structures of three crystal forms of the sweet protein thaumatin." *Acta Cryst.* **D50**, 813-825.

H. Kramers (1927). "La diffusion de la lumière par les atomes." *Atti. Congr. Int. Phis. Como* **2**: 545-557.

R. de Kronig (1926). "On the theory of dispersion of X-rays." *J. Opt. Soc. Am.* **12**: 547-557.

K. Kumar, C. Chang, L. Francesconi, D. Dischino, M. Malley, J. Gougoutas & M. Tweedle (1994). "Synthesis, Stability, and Structure of Gadolinium(III) and Yttrium(III) Macrocyclic Poly(amino carboxylates)." *Inorg. Chem.*, **33**, 3567-3575.

E. de La Fortelle & G. Bricogne (1997). "Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods." *Methods Enzymol.* **276**, 472-494.

L. Lagartera, A. González, M. Stelter, P. García, R. Kahn, M. Menéndez & J. Hermoso (2005). "Crystallization and preliminary X-ray diffraction studies of the pneumococcal teichoic acid phosphorylcholine esterase Pce." *Acta Cryst.* **F61**, 221-224.

V. Lamzin, A. Perrakis & K. Wilson (2001). "The ARP/WARP suite for automated construction and refinement of protein models." In *Int. Tables for Crystallography. Vol. F: Crystallography of biological macromolecules* (Rossmann, M.G. & Arnold, E. eds.), Dordrecht, Kluwer Academic Publishers, The Netherlands, 720-722.

A. Leslie, E. Duke, D. Brown, J. Naismith, C. Nave, S. Phillips, D. Stuart, A. Thompson, M. Walsh & G. Evans (2004). "Diamond Beamline Proposal 044 - A Tuneable Beamline for Macromolecular Crystallography Optimised for Long Wavelengths (1.5-2.5Å)." Doc. No: BLS-044-PRO-0003.

P. Machin (1985). Editor. *Molecular Replacement*. Proceedings of the Daresbury Study Weekend, 15-16 Feb. 1985. SERC Daresbury Laboratory, Warrington, England.

A. McPherson (2001). "A comparison of salts for the crystallization of macromolecules." *Protein Science*, **10**:418-422.

G. Mie (1908). "Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen." *Ann. Phys. (Leipzig)* **25**: 377–452 (Eq. 98).

J. Müller & U. Heinemann (2005). "Use of anomalous scattering for solving the phase problem in protein crystallography." dans : *Encyclopedic Reference of Genomics and Proteomics* (Oschkinat, H. & Schmieder, P. eds) Springer, Heidelberg, sous presse.

J. Murray, E. Garman & R. Ravelli (2004). "X-ray absorption by macromolecular crystals: the effects of wavelength and crystal composition on absorbed dose." *J. Appl. Cryst.* **37**, 513-522.

J. Navaza (1994). "AMoRe: an Automated Package for Molecular replacement." *Acta Cryst. A***50**: 157-163.

C. Nave & E. Garman (2005). "Towards an understanding of radiation damage in cryocooled macromolecular crystals." *J. Synchrotron Rad.*, **12**, 257-260.

J. Ohana, L. Jacquamet, J. Joly, A. Bertoni, P. Taunier, L. Michel, P. Charrault, M. Pirocchi, P. Carpentier, F. Borel, R. Kahn & J.-L. Ferrer (2004). "CATS: a Cryogenic Automated Transfer System installed on the beamline FIP at ESRF" *J. Appl. Cryst.* **37**, 72-77.

Z. Otwinowski & W. Minor (1997). "Processing of X-ray diffraction data collected in oscillation mode." *Methods Enzymol.* **276**, 307-326.

A. Patterson (1935). *Zeitschrift für Kristallograpie* **90**, 517.

M. Purdy, P. Ge, J. Chen, P. Selvin & M. Wiener (2002). "Thiol-reactive lanthanide chelates for phasing protein X-ray diffraction data." *Acta Cryst. D***58**, 1111-1117.

U. Ramagopal, M. Dauter & Z. Dauter (2003a). "SAD manganese in two crystal forms of glucose isomerase." *Acta Cryst. D***59**, 868-875.

U. Ramagopal, M. Dauter & Z. Dauter (2003b). "Phasing on anomalous signal of sulfurs: what is the limit ?" *Acta Cryst. D***59**, 1020-1027.

R. Ravelli, H. Schröder Leiros, B. Pan, M. Caffre & S. McSweeney (2003). "Specific Radiation Damage Can Be Used to Solve Macromolecular Crystal Structures." *Structure*, **11**, 217-224.

- R. Ravelli & S. McSweeney (2000). "The fingerprint that X-rays can leave on structures." *Structure*, 8, 315-328.
- P. Retailleau, N. Colloc'h, D. Vivares, F. Bonnete, B. Castro, M. El Hajji & T. Prange (2005). "Urate Oxidase from *Aspergillus Flavus*: New Crystal-Packing Contacts in Relation to the Content of the Active Site." *Acta Cryst.* D**61**, 218-229.
- G. Rieck (1985). "Tables Relating to the production, wavelengths and intensities of x-rays." in *International Tables for X-ray Crystallography, III, X-rays and their interaction with crystals*. Dordrecht (Holland): D. Reidel Publishing Company for International Union of Crystallography, 59-72.
- M. Rossmann (1972). Editor. *The Molecular Replacement Method*. New York: Gordon and Breach.
- R. Sanishvili, A. Joachimiak, A. Edwards, A. Savchenko & T. Skarina "Putative Ribosomal Protein" en prep.
- D. Sayre (1952) "The squaring method: a new method for phase determination." *Acta Cryst.* **5**: 60-65.
- M. Schiltz, P. Dumas, E. Ennifar, C. Flensburg, W. Paciorek, C. Vonrhein & G. Bricogne (2004). "Phasing in the presence of severe site-specific radiation damage through dose-dependent modelling of heavy atoms." *Acta Cryst.* D**60**, 1024-1031.
- M. Schiltz, R. Fourme & T. Prangé (2003). "Use of noble gases xenon and krypton as heavy atoms in protein structure determination." *Methods Enzymol.* **374**, 83-119.
- M. Schiltz, Å. Kvik, O. Svensson, W. Shepard, E. de LaFortelle, T. Prangé, R. Kahn & R. Fourme (1997). "Protein Crystallography at Ultra-short Wavelengths : Feasibility Study of Anomalous Dispersion Experiments at the Xenon K-Edge." *J. Synchrotron Rad.* **4**, 287-297.
- T. Schneider & G. Sheldrick (2002). "Substructure solution with SHELXD." *Acta Cryst.* D**58**, 1772-1779.
- A. Schuettkopf & D. van Aalten (2004). "PRODRG - a tool for high-throughput crystallography of protein-ligand complexes." *Acta Cryst.* D**60**, 1355-1363.
- G. Sheldrick (1998). In *Direct Methods for Solving Macromolecular Structures*. Edited by Fortier S. Dordrecht, The Netherlands: Kluwer Academic Publishers; 401-411.
- T. Terwilliger (2002). "Automated main-chain model-building by template-matching and iterative fragment extension." *Acta Cryst.* D**59**, 34-44.
- T. Terwilliger & J. Berendzen. (1999). "Automated MAD and MIR structure solution." *Acta Cryst.* D**55**, 849-861.
- I. Usón & G. Sheldrick (1999). "Advances in direct methods for protein crystallography." *Current Opinion in Structural Biology*, **9**, 643-648.

M. Vaney, S. Maignan, M. RiesKautt & A. Ducruix, A. (1996). "High-resolution structure (1.33 angstrom) of a HEW lysozyme tetragonal crystal grown in the APCF apparatus. Data and structural comparison with a crystal grown under microgravity from SpaceHab-01 mission." *Acta Cryst.* **D52**, 505-517.

J. Victoreen (1949). "The Calculation of x-ray Mass Absorption Coefficients." *Journal of Applied Physics* **20**, 1141-1147.

D. Vivarès & F. Bonneté (2002). "X-ray scattering studies of *Aspergillus flavus* urate oxidase: towards a better understanding of PEG effects on the crystallization of large Proteins." *Acta Cryst.* **D58** 472-479.

C. Weeks & M. Miller (1999). "The design and implementation of SnB v2.0." *J. Appl Crystallogr* **32**, 120-124.

M. Weik, R. Ravelli, G. Kryger, S. McSweeney, M. Raves, M. Harel, P. Gros, I. Silman, J. Kroon & J. Sussman (2000). "Specific chemical and structural damage to proteins produced by synchrotron radiation." *Proc. Natl. Acad. Sci.* **97**, 623-628.

M. Weiss, G. Mander, R. Hedderich, K. Diederichs, U. Ermler & E. Warkentin (2004). "Determination of a novel structure by a combination of long-wavelength sulfur phasing and radiation-damage-induced phasing." *Acta Cryst.* **D60**, 686-695.

## Annexe 1

### Calcul de la carte de Patterson des atomes lourds (Patterson, 1935)

#### A1.1. La fonction de Patterson

La fonction de Patterson est définie comme le développement en série de Fourier des intensités  $I = |F(\mathbf{h})|^2$ , qui sont connues expérimentalement.

$$P(\mathbf{u}) = \frac{1}{V_c} \sum_{\mathbf{h}} |F(\mathbf{h})|^2 \exp(-2\pi i \mathbf{h} \cdot \mathbf{u})$$

avec  $\mathbf{u} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$

Dans le cas où la diffusion anormale est faible, et, par conséquent, la loi de Friedel est respectée, la fonction de Patterson devient la fonction d'autocorrélation de la densité électronique  $\rho(\mathbf{r})$  :

$$P(\mathbf{u}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) = \int_V \rho(\mathbf{r}) \rho(\mathbf{u} + \mathbf{r}) d^3 \mathbf{r}$$

Pour des densités électroniques décrites avec des atomes ponctuels, la densité s'écrit :

$$\rho(\mathbf{r}) = \sum_j Z_j \delta(\mathbf{r} - \mathbf{r}_j)$$

où  $Z_j$  est le nombre d'électrons de l'atome  $j$  situé en  $\mathbf{r}_j$ . Dans ce modèle, la fonction de Patterson s'écrit alors

$$\begin{aligned} P(\mathbf{u}) &= \int_V \sum_j Z_j \delta(\mathbf{r} - \mathbf{r}_j) \sum_{j'} Z_{j'} \delta(\mathbf{u} + \mathbf{r} - \mathbf{r}_{j'}) d^3 \mathbf{r} = \sum_j \sum_{j'} Z_j Z_{j'} \int_V \delta(\mathbf{r} - \mathbf{r}_j) \delta(\mathbf{u} + \mathbf{r} - \mathbf{r}_{j'}) d^3 \mathbf{r} \\ &= \sum_j \sum_{j'} Z_j Z_{j'} \delta(\mathbf{u} + \mathbf{r}_j - \mathbf{r}_{j'}) \end{aligned}$$

et la fonction de Patterson devient :

$$P(\mathbf{u}) = \sum_j \sum_{j'} Z_j Z_{j'} \delta(\mathbf{u} - \mathbf{u}_{jj'})$$

avec les vecteurs interatomiques de la maille  $\mathbf{u}_{jj'} = \mathbf{r}_{j'} - \mathbf{r}_j$

La fonction de Patterson est réduite à des pics de hauteur  $Z_j Z_{j'}$  situés aux extrémités des vecteurs interatomiques.

#### A1.2. La fonction de Patterson "anormale"

Dans le cas où la diffusion anormale n'est pas négligeable, on peut calculer la fonction de Patterson anormale avec les coefficients  $(D'_{ANO})^2 = \left\{ \frac{(|F(+)| - |F(-)|) f^o}{2 f''} \right\}^2$  à la place du carré des modules des facteurs de structure :

$$P_{ANO}(\mathbf{u}) = \frac{1}{V_c} \sum_{\mathbf{h}} |D'_{ANO}(\mathbf{h})|^2 \exp(-2\pi i \mathbf{h} \cdot \mathbf{u})$$

Si on écrit le facteur de structure de la molécule

$$F = F_p + F_H \text{ où :}$$

$F_p$  est le facteur de structure de la protéine native,

$F_H$  le facteur de structure des atomes lourds.

On peut montrer (\*) que la fonction  $P_{ANO}(\mathbf{u})$  correspond, à un bruit près, à une fonction de Patterson dont les coefficients sont les carrés des modules des facteurs de structure des atomes lourds :

$$P_H = \frac{1}{V_c} \sum_{\mathbf{h}} |F_H(\mathbf{h})|^2 \exp(-i2\pi\mathbf{h} \cdot \mathbf{u}) \approx P_{ANO}$$

La fonction de Patterson anormale représente alors la carte des vecteurs interatomiques des atomes lourds. De la même manière la carte de Patterson peut être calculé avec comme coefficients les carrés des différences isomorphes  $D_{iso}^2 \propto \|F_{PH} - F_P\|^2$  ou des différences dispersives  $D_{dis}^2 \propto \| |F|^{\lambda_i} - |F|^{\lambda_j} \|^2$ .

### A1.3. Interprétation de la fonction de Patterson

Compte tenu du grand nombre de pics interatomiques, dans le cas d'un cristal de protéine, la fonction de Patterson n'est, en général, pas interprétable. La carte de Patterson anormale présente des pics beaucoup moins nombreux et qui correspondent aux vecteurs interatomiques des atomes diffusant anormalement. À partir de cette carte, on cherche à trouver les positions des atomes lourds dans la maille. La présence de pics marqués dans la carte de Patterson anormale indique la fixation de diffuseurs anormaux sur la protéine.

Les éléments de symétrie, imposés par le groupe d'espace, mènent à des concentrations de pics dans certaines régions de la carte de Patterson, appelées " sections de Harker ". Les pics trouvés dans ces sections indiquent les extrémités de vecteurs interatomiques d'atomes anormaux occupant des positions équivalentes. Les pics apparaissant en dehors des sections de Harker sont relatifs à des diffuseurs anormaux ne se trouvant pas dans des positions équivalentes, ce sont des pics croisés de différents sites. Connaissant les relations qui relient les positions équivalentes, les coordonnées x,y,z des diffuseurs anormaux peuvent être déduites de coordonnées u,v,w des pics des sections de Harker. La cohérence du résultat est vérifiée par inspection des pics croisés.

Alors que pour un faible nombre de sites anormaux, l'interprétation de la fonction de Patterson peut se faire "à la main", on utilise généralement des programmes de détection des sites des atomes lourds.

Le programme SHELXD (Sheldrick, 1998; Schneider & Sheldrick, 2002) alterne la recherche des sites lourds dans l'espace direct et le calcul et l'affinement des phases dans l'espace réciproque et utilisent les méthodes de Patterson et les méthodes directes.

\* Le facteur de structure de la molécule s'écrit  $F = F_p + F_H$ , où  $F = F(\mathbf{h}) = F(+)$ .

Avec  $F_H = \sum_j f_H \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j)$ ,  $f_H = f^o + f' + if''$  et  $F_H^o = \sum_j f^o \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j)$ ,

on obtient, dans le cas d'un seul type de diffuseur anormal

$$F = F_p + F_H^o \frac{f_H}{f^o} = F_p + F_H^o + F_H^o \left( \frac{f'}{f^o} + \frac{if''}{f^o} \right) = F^o + F_H^o \left( \frac{f'}{f^o} + \frac{if''}{f^o} \right)$$

avec  $F^o = F_p + F_H^o = |F^o| \exp(i\phi^o)$ ,  $F_H^o = |F_H^o| \exp(i\phi_H)$  et

$$F(-) = F(-\mathbf{h}) = |F^o| \exp(-i\phi^o) + |F_H^o| \exp(-i\phi_H) \left( \frac{f'}{f^o} + \frac{if''}{f^o} \right) \text{ et avec } I(+) = F(\mathbf{h})F^*(\mathbf{h}) \text{ on}$$

obtient :

$$I(\pm) = |F^o|^2 + |F_H^o|^2 \left( \left( \frac{f'}{f^o} \right)^2 + \left( \frac{f''}{f^o} \right)^2 \right) + 2 \frac{f'}{f^o} |F^o| |F_H^o| \cos(\phi^o - \phi_H) \pm 2 \frac{f''}{f^o} |F^o| |F_H^o| \sin(\phi^o - \phi_H)$$

On en déduit  $I(+)-I(-) = 4 \frac{f''}{f^o} |F^o| |F_H^o| \sin(\phi^o - \phi_H)$

avec la différence anormale

$$D_{ANO} = |F(+)| - |F(-)| \text{ on a}$$

$$I(+)-I(-) = D_{ANO} (|F(+)| + |F(-)|)$$

$$\Rightarrow D_{ANO} = \frac{I(+)-I(-)}{|F(+)|+|F(-)|} \approx \frac{I(+)-I(-)}{2|F^o|} \approx 2 \frac{f''}{f^o} |F_H^o| \sin(\phi^o - \phi_H)$$

$$\Rightarrow (D_{ANO})^2 = 4 \left( \frac{f''}{f^o} \right)^2 |F_H^o|^2 \sin^2(\phi^o - \phi_H) = 2 \left( \frac{f''}{f^o} \right)^2 |F_H^o|^2 (1 - \cos(2(\phi^o - \phi_H)))$$

Dans la synthèse de Fourier le terme en cosinus, du côté droit de l'expression correspond à du bruit. La fonction de Patterson calculée avec les coefficients

$$\frac{1}{2} (D_{ANO})^2 \left( \frac{f^o}{f''} \right)^2 = 2 (D'_{ANO})^2 \text{ correspond alors à la fonction de Patterson calculée avec les}$$

coefficients  $|F_H^o|^2$ , la fonction de Patterson des atomes lourds.

## Annexe 2

### Comparaison de la fixation des complexes pour des cristaux dérivés de glucose isomérase (GI) obtenus en présence de différents agents précipitants (MPD, PEG, sulfate d'ammonium (AS))

Nous avons procédé à des tests d'obtention de dérivés avec des cristaux de GI obtenus dans différentes conditions de cristallisation. La comparaison de la fixation est cependant rendue difficile par des difficultés de congélation des cristaux. La qualité de diffraction des cristaux obtenus dans les conditions alternatives est largement moins bonne que celle des cristaux obtenus en MPD. De plus, les conditions d'enregistrement des données sont différentes (longueurs d'onde). L'affinement de l'occupation des sites de fixation avec les différents jeux permettrait une comparaison quantitative, mais, faute de temps, nous n'avons pas effectué l'affinement des structures. Voici une comparaison qualitative effectuée à l'aide des cartes de Patterson anormales des différents jeux de données. Toutes les cartes sont calculées avec des données coupées à une résolution de 2,7 Å.

Tous les dérivés appartiennent au groupe d'espace I222 avec des paramètres de maille très proches de  $a = 93$  Å,  $b = 99$  Å et  $c = 103$  Å.

|    | Précipitant | Complexe    | $\lambda$ [Å] | Source RX | $R_{\text{fac}}$ | $R_{\text{ano}}$ |
|----|-------------|-------------|---------------|-----------|------------------|------------------|
| 1) | PEG         | Gd-DTPA-BMA | 1,28          | FIP       | 0,081            | 0,036            |
| 2) | PEG         | Gd-DOTMA    | 0,98          | FIP       | 0,057            | 0,033            |
| 3) | MPD         | Gd-DO3A     | 1,711         | FIP       | 0,034            | 0,082            |
| 4) | PEG         | Gd-DO3A     | 1,54          | labo      | 0,062            | 0,038            |
| 5) | AS          | Gd-DO3A     | 0,98          | FIP       | 0,045            | 0,022            |
| 6) | MPD         | Gd-DTPA     | 1,711         | FIP       | 0,034            | 0,039            |
| 7) | PEG         | Gd-DTPA     | 1,28          | FIP       | 0,117            | 0,093            |

Tableau A2.1. : Indicateurs statistiques, longueur d'onde d'enregistrement, complexe et précipitant utilisé pour sept dérivés de la glucose isomérase.

#### Dérivés avec les complexes Gd-DTPA-BMA et Gd-DOTMA

Pour les dérivés 1 et 2, aucune fixation notable des complexes n'a pu être détectée (figure A2.1.).

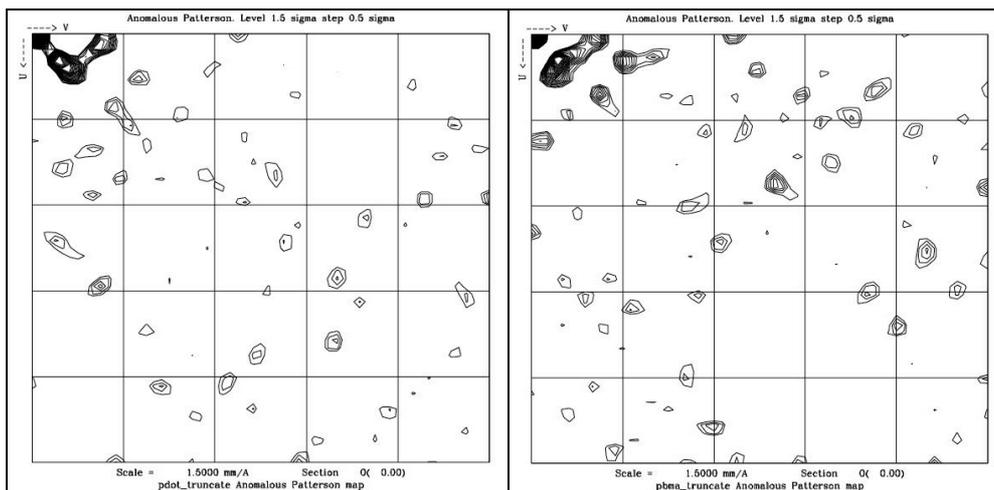


Figure A2.1.: Les sections de Harker  $w=0$  pour les cartes de Patterson anormale des dérivés 1 (gauche) et 2 (droite).

### Dérivés avec le complexe Gd-DO3A

Les jeux de données 3, 4 et 5 correspondent à des dérivés obtenus avec le complexe Gd-DO3A. Comme le montre la valeur du facteur  $R_{\text{fac}}$ , la qualité de diffraction est relativement bonne pour les trois cristaux dérivés (tableau A2.1.). La comparaison des cartes de Patterson pour les trois dérivés est difficile, car les longueurs d'onde d'enregistrement sont très différentes (3:  $\lambda=1,711 \text{ \AA}$ ,  $f''(\text{Gd}) = 28 e^-$ , 4:  $\lambda = 1,54 \text{ \AA}$ ,  $f''(\text{Gd}) = 12 e^-$ , 5:  $\lambda = 0,98 \text{ \AA}$ ,  $f''(\text{Gd}) = 6,7 e^-$ ). Une comparaison qualitative des sections de Harker ( $w=0$  et  $u=0$  ou  $v=0$ ) des cartes de Patterson anormales semble indiquer une fixation semblable entre le dérivé obtenu en MDP (figure A2.2.a) et le dérivé obtenu en PEG (figure A2.2.b). Les pics correspondant aux trois sites de fixation (A,B,C) apparaissent dans les deux cartes avec une fixation moins forte, semble-t-il, pour le site B pour le dérivé en PEG. La carte de Patterson du dérivé obtenu en AS (figure A2.2.c) ne présente aucun pic correspondant aux sites A, B ou C. Elle semble cependant indiquer un site de fixation différent (voir section  $w=0,11$ ,  $v=0,5$ ), ce qui pourrait suggérer une influence de l'agent précipitant sur la fixation du complexe.

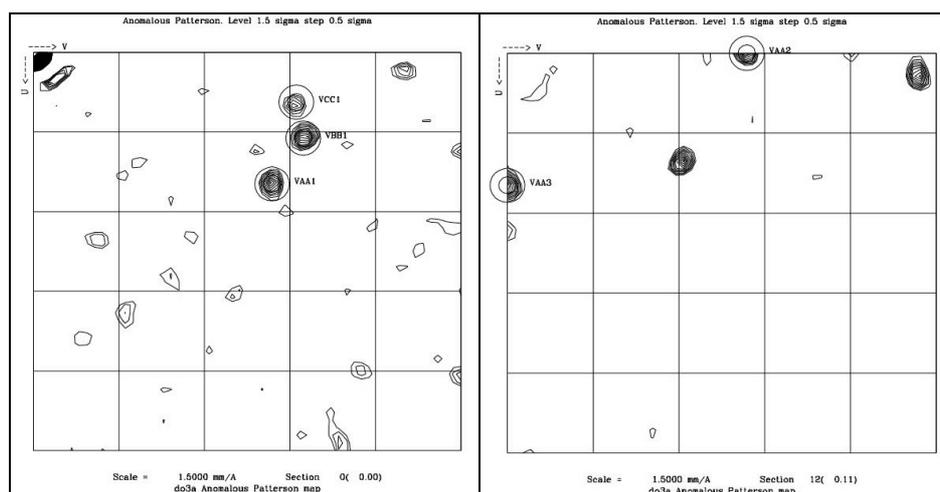


Figure A2.2.a. : Sections  $w=0$  et  $w=0,11$  de la carte de Patterson anormale du dérivé GI-Gd-DO3A obtenu en MPD.

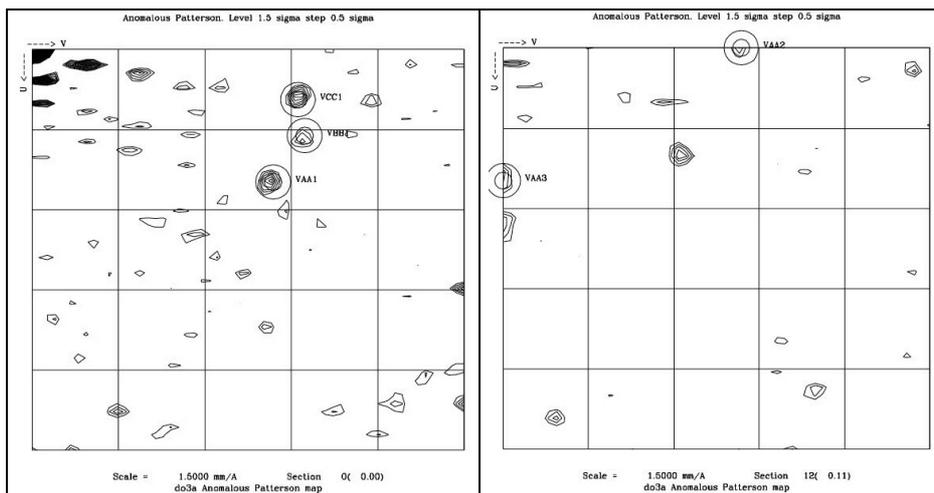


Figure A2.2.b. : Sections  $w=0$  et  $w=0,11$  de la carte de Patterson anormale du dérivé GI-Gd-DO3A obtenu en PEG.

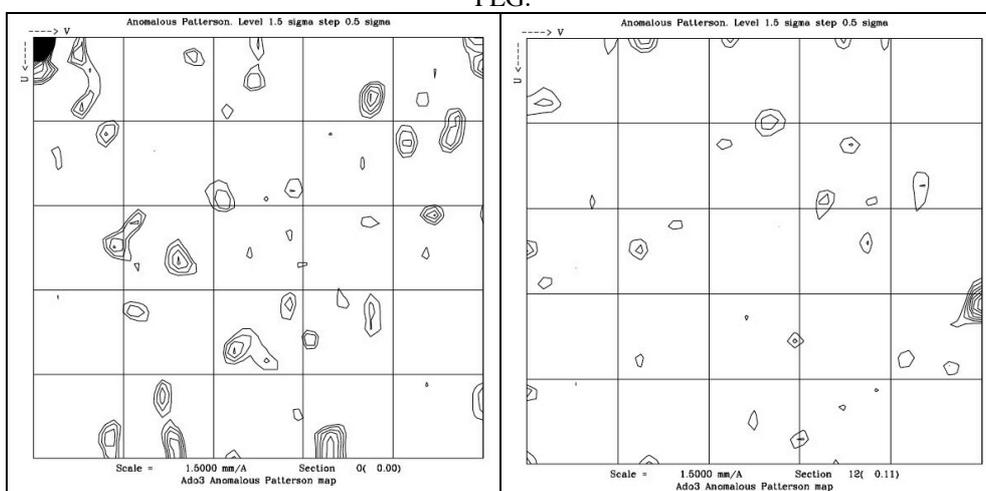


Figure A2.2.c. : Sections  $w=0$  et  $w=0,11$  de la carte de Patterson anormale du dérivé GI-Gd-DO3A obtenu en AS.

### Dérivés avec le complexe Gd-DTPA

Pour le complexe Gd-DTPA aucun bon dérivé en AS n'a pu être obtenu. Comme l'indique la valeur relativement élevée du facteur  $R_{fac}$  (tableau A2.1.) le dérivé obtenu en PEG diffracte relativement mal, ce qui mène à une carte de Patterson anormale assez bruitée (figure A2.3.b). On y retrouve les pics correspondant au pic principal du dérivé en MPD (figure A2.3.a) cependant très faibles pour le jeu 7. Les cartes de jeux de qualité très différente se prêtent difficilement à interprétation.

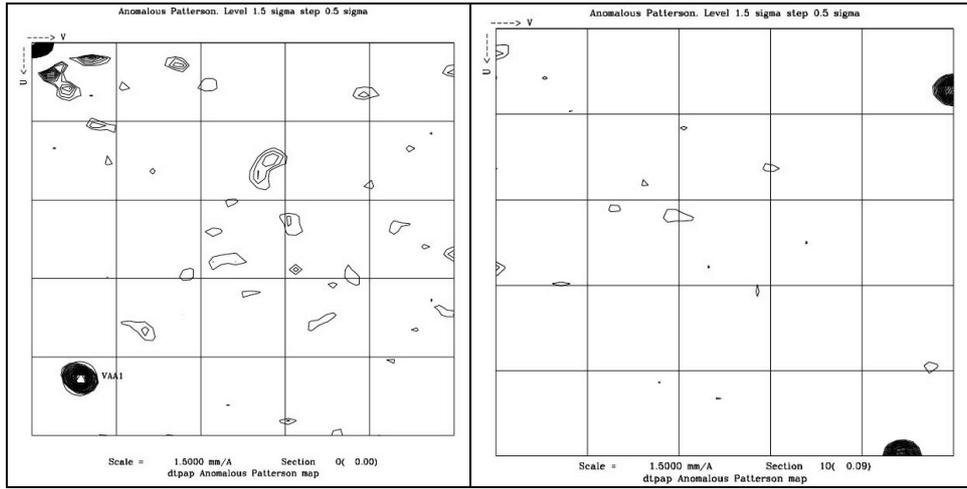


Figure A2.3.a. : Sections  $w=0$  et  $w=0,09$  de la carte de Patterson anormale du dérivé GI-Gd-DTPA obtenu en MPD.

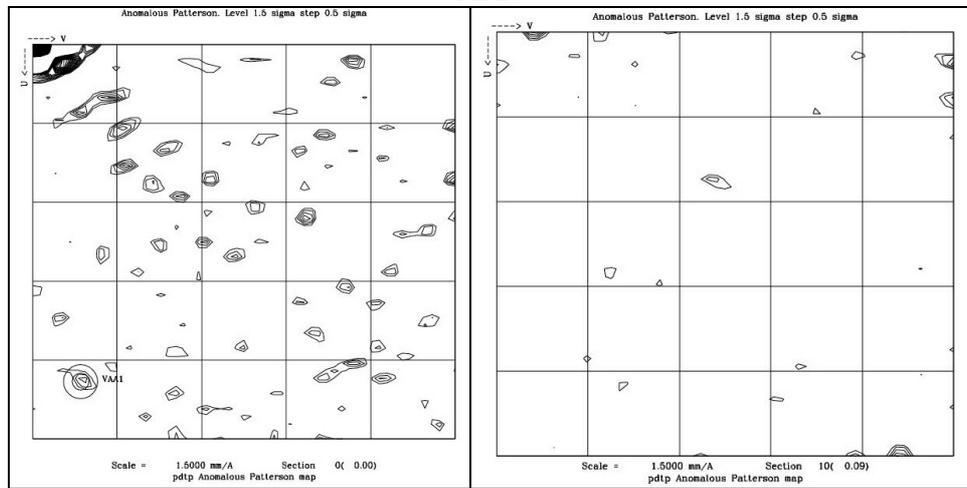


Figure 3b : Sections  $w=0$  et  $w=0,09$  de la carte de Patterson anormale du dérivé GI-Gd-DTPA obtenu en PEG.



### **Étude de complexes à forte diffusion anormale pour la détermination rapide de la structure de protéines par la méthode MAD**

Pour résoudre *de novo* la structure de protéines par cristallographie, il est nécessaire de calculer les phases des facteurs de structure à partir des intensités des rayons X diffractés. Pour ce but, les méthodes SAD et MAD se servent de la diffusion anormale d'atomes présents dans le cristal de la protéine. L'introduction de ces diffuseurs anormaux dans les cristaux est une des étapes clés de la résolution de structure des protéines.

Nous avons étudié une classe de huit complexes de gadolinium servant à insérer les diffuseurs anormaux dans les cristaux de protéine. Le gadolinium présente une forte diffusion anormale et avec le rayonnement X d'un générateur de laboratoire et dans son seuil d'absorption  $L_{III}$ .

Une étude cristallographique menée avec cette classe de complexes et avec huit protéines différentes a permis de démontrer le potentiel élevé des complexes pour la préparation de dérivés à fort pouvoir de phasage. En effet, pour un grand nombre des dérivés testés, les phases expérimentales calculées ont mené à des cartes de densité expérimentale d'excellente qualité, permettant la construction aisée du modèle de la protéine.

L'affinement de la structure des complexes liés à la protéine a permis de comprendre l'interaction des différents complexes avec les protéines.

L'utilisation des complexes a permis de résoudre la structure de quatre nouvelles protéines.

Nous avons également étudié des méthodes physico-chimiques alternatives à la cristallographie dans le dessein de détecter la fixation d'un complexe sur une protéine en tenant compte de la particularité de l'interaction qui est caractérisée par une constante de d'association faible.

**Mots clés :** cristallographie des protéines, détermination des phases expérimentales, diffusion anormale, méthodes MAD et SAD, complexes de lanthanides, dérivés lourds.

### **Investigation of complexes exhibiting strong anomalous scattering for the fast structure determination of proteins by the MAD method**

For the *de novo* structure resolution of proteins by crystallography, it is necessary to calculate the phases of the structure factors from the diffracted X-ray intensities. For this purpose, the SAD and MAD methods take advantage of the anomalous scattering properties of atoms inside the protein crystals. The introduction of the anomalous scatterers into the crystals is one of the key steps of protein structure resolution.

We have investigated a class of eight gadolinium complexes used to introduce the anomalous scatterers into the protein crystals. Gadolinium exhibits high anomalous scattering with a laboratory X-ray generator as well as in its  $L_{III}$  absorption edge.

A crystallographic study carried out with the class of complexes and eight different proteins allowed to demonstrate the great potential of the complexes for the preparation of derivatives with high phasing power. Indeed, for a great number of derivatives, the calculated experimental phases have led to experimental electron-density maps of excellent quality, allowing the easy construction of the protein model.

The refinement of the structure of the complexes bound to the proteins allowed to gain understanding of the interaction between the different complexes and the proteins.

The utilization of the complexes allowed the structure resolution of four new proteins.

Besides the crystallographic studies we attempted to detect the binding of a complex to a protein with different physico-chemical methods taking into consideration the weak binding constant that characterizes the interaction.

**Keywords:** protein crystallography, experimental phase determination, anomalous scattering, MAD and SAD methods, lanthanide complexes, heavy-atom derivatives.