

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 $T = 8^\circ\text{C}$

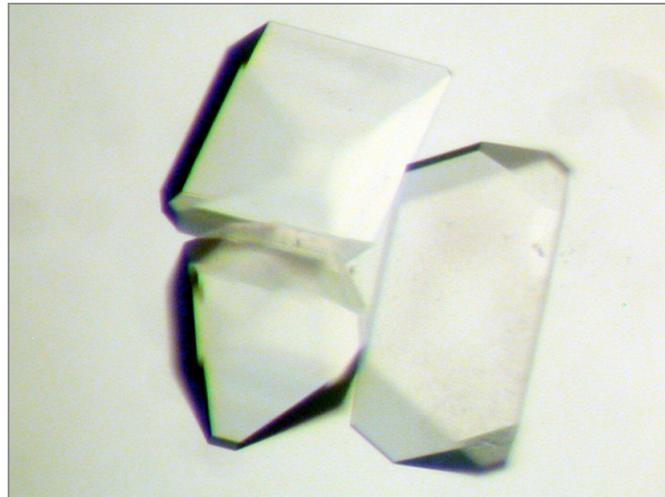


Figure 2.1.1 : Cristaux d'urate oxydase. Taille des cristaux : ~ 200 - $500 \mu\text{m}$.

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 50 mM tampon acétate de sodium pH 4,5 $C_p = 40$ mg/mL Composition des gouttes : $V_p = 3$ μ L, $V_R = 3$ μ L $T = 20^\circ\text{C}$

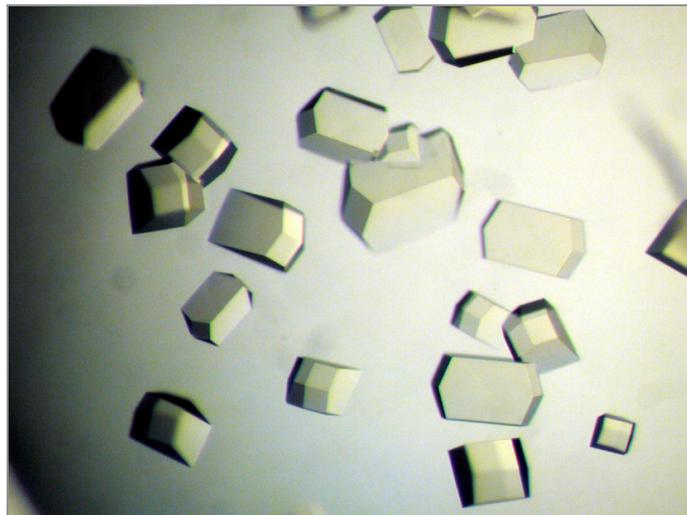


Figure 2.1.2 : Cristaux de lysozyme. Taille des cristaux : ~ 300 - 800 μm .

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 0 - 5% éthylène glycol 100 mM tampon ADA pH 6,6 $C_p = 50$ mg/mL Composition des gouttes : $V_p = 6$ μ L, $V_R = 4$ μ L $T = 20^\circ\text{C}$

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.

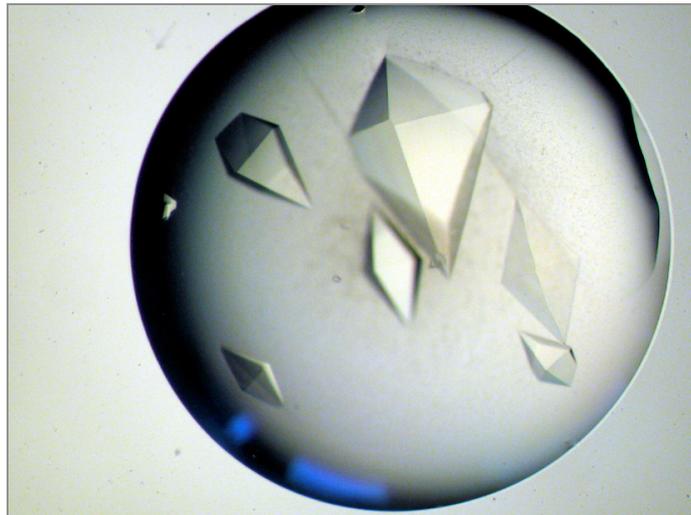


Figure 2.1.3 : Cristaux de thaumatine. Taille des cristaux : $\sim 200-800 \mu\text{m}$.

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)

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 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 MgCl₂
 50 mM tampon tris pH7
 C_p ~ 30 mg/mL
 Composition des gouttes : V_p = 4 μL, V_R = 2 μL
 T = 20°C

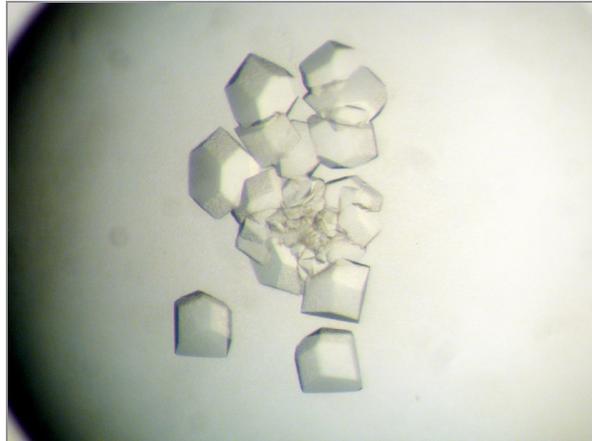


Figure 2.1.4 : Cristaux de GI obtenus en présence de MPD. Taille des cristaux : ~300 μm.

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) :

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 ₄) ₂ SO ₄
40 mM MgCl ₂	50 mM MgCl ₂
50 mM tampon tris pH 7,0	50 mM tampon tris pH 7,0
C _p ~ 30 mg/mL	C _p ~ 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% $(\text{NH}_4)_2\text{SO}_4$ 0,1 M tampon Bicine pH 9,0 5% glycérol T = 20°C

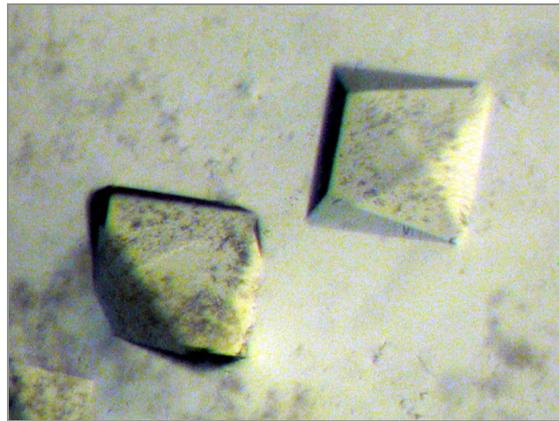


Figure 2.1.5 : Cristaux de YGGV. Taille des cristaux : ~200 μm .

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 0,1 M acétate d'ammonium 0,1 M tampon bis tris pH = 5,5
Composition des gouttes	$V_p = 4 \mu\text{L}$, $V_R = 4 \mu\text{L}$ + 1 μL 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 0,2 M NaCl 10-20% glycérol 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 3%(v/v) dioxane 0,1 M tampon dihydrate de citrate trisodique pH 5,6 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- β -lactamase de classe B de *Pseudomonas aeruginosa* (230 résidus/ua)

Conditions de cristallisation	30% PEG8000 100 mM tampon cacodylate de sodium pH 6,5 200 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 100 mM tampon tris maléate pH 6,5
C _p	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 100 mM tampon imidazole pH 7,8 1 mM DTT
C _p	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.

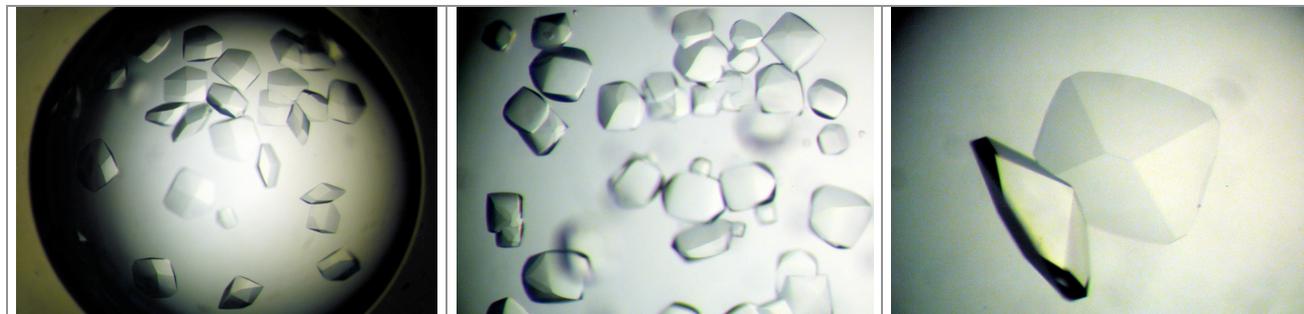


Figure 2.1.6 : Cristaux de lysozyme obtenus en présence de 300 mM de différents complexes. À gauche : Gd-HPDO3A, au milieu : Gd-DOTA, à droite : Gd-DO3A. On remarque l'influence de la présence des différents complexes sur la morphologie des cristaux. L'ajout du complexe mène à des cristaux "arrondis"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

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/σ : Rapport signal sur bruit

Les indicateurs R_{fac} et R_{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/σ 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 300 mM	Gd-DO3A 300 mM	Gd-HPDO3A 300 mM	Gd-DOTA 300 mM	
Groupe d'espace	P ₄ ,2 ₁ ,2				
Paramètres de maille [Å]	a = b = 77,8; c = 37,4	a = b = 77,3 ; c = 37,3	a = b = 77,1 ; c = 38,3	a = b = 77,2 ; 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 (1,53-1,62)	1,54 (1,54-1,62)	1,53 (1,53-1,62)	1,54 (1,54-1,62)	1,64 (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 _{fac}	0,068 (0,387)	0,061 (0,26)	0,057 (0,079)	0,040 (0,15)	0,055 (0,086)
R _{ano}	0,034 (0,138)	0,039 (0,13)	0,050 (0,076)	0,035 (0,098)	0,062 (0,096)
Domaine angulaire	180°	180°	180°	180°	180°

Urate oxydase	Gd-DOTMA 300 mM		Gd-DTPA-BMA 300 mM		Gd-DO3A 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 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 (1,34-1,41)	2,16 (2,16-2,28)	1,45 (1,45-1,53)	2,70 (2,70-2,84)	1,45 (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 _{fac}	0,051 (0,32)	0,090 (0,37)	0,040 (0,13)	0,037 (0,051)	0,039 (0,19)
R _{ano}	0,037 (0,23)	0,098 (0,24)	0,044 (0,097)	0,110 (0,148)	0,023 (0,13)
Domaine angulaire	180°	180°	180°	180°	180° par pas de 1°

Thaumatine	Gd-DOTA 300 mM		Gd-DOTA-BOM 300 mM	
Groupe d'espace	P4 ₁ 2 ₁ 2		P4 ₁ 2 ₁ 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 (1,45-1,53)	2,69 (2,69-2,83)	1,45 (1,45-1,53)	2,69 (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 _{fac}	0,056 (0,24)	0,030 (0,056)	0,053 (0,24)	0,066 (0,096)
R _{ano}	0,019 (0,079)	0,054 (0,066)	0,029 (0,11)	0,096 (0,013)
Domaine angulaire	180°	80°	180°	180°

Thaumatine	Gd-DTPA-BMA 300 mM		Gd-DTPA 300 mM		Gd-HPDO3A 100 mM
Groupe d'espace	P4 ₁ 2 ₁ 2		P4 ₁ 2 ₁ 2		P4 ₁ 2 ₁ 2
Paramètres de maille [Å]	a = b = 57,9 c = 150,4		a = b = 57,9 c = 150,1		a = b = 57,9 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 (1,45-1,53)	2,66 (2,66-2,81)	1,45 (1,45-1,53)	2,67 (2,67-2,82)	1,55 (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 _{fac}	0,040 (0,15)	0,043 (0,063)	0,037 (0,097)	0,032 (0,052)	0,039 (0,092)
R _{ano}	0,021 (0,053)	0,058 (0,072)	0,025 (0,042)	0,049 (0,062)	0,024 (0,037)
Domaine angulaire	180°	180°	180°	180°	217°

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

Protéine X	Gd-DOTMA 100 mM	Gd-DTPA-BMA 100 mM	Gd-DO3A 100 mM
Groupe d'espace	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Paramètres de maille [Å]	a = 49,2 b = 115,1 c = 75,9	a = 48,7 b = 114,0 c = 76,3	a = 49,3 b = 115,7 c = 76,7
λ [Å]	1,711	1,711	1,711
Source de RX	ID29	ID23-1	ID29
Résolution [Å]	2,17 (2,17-2,29)	2,98 (2,98-3,14)	2,17 (2,17-2,29)
I/ σ	10,9 (4,7)	7,2 (2,6)	7,2 (2,2)
<I/ σ >	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 _{fac}	0,053 (0,15)	0,081 (0,24)	0,068 (0,27)
R _{ano}	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 [σ]	59	22 ; 21 ; 11	46 ; 39 ; 29 ; 17 ; 16 ; 15

Glucose isomérase	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	
λ [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/σ	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_{fac}	0,036 (0,29)	0,034 (0,088)	0,043 (0,29)	0,034 (0,072)
R_{ano}	0,027 (0,13)	0,082 (0,102)	0,024 (0,112)	0,039 (0,060)
Domaine angulaire	180°	180°	180°	180°

Glucose isomérase	Gd-DOTA-BOM 100 mM		Gd-DOTMA 100 mM		Gd-DTPA-BMA 100 mM
Groupe d'espace	I222		I222		I222
Paramètres de maille [Å]	a = 92,7 c = 102,3	b = 98,3	a = 92,7 c = 102,6	b = 98,0	a = 92,7 b = 97,9 c = 102,6
λ [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/σ	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_{fac}	0,031 (0,33)	0,030 (0,078)	0,050 (0,31)	0,047 (0,11)	0,059 (0,16)
R_{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°

YGGV	Gd-DO3A 100 mM		
Groupe d'espace	P4 ₃ 2 ₁ 2		
Paramètres de maille [Å]	a = b = 78,7 c = 77,2		
λ [Å]	pic 1,711	infl1 1,712	infl2 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/σ	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_{fac}	0,066 (0,47)	0,050 (0,31)	0,051 (0,29)
R_{ano}	0,106 (0,24)	0,063 (0,28)	0,065 (0,23)
Domaine angulaire	180°	90°	90°

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

YEAZ	Gd-DOTMA 100 mM	
Groupe d'espace	P2 ₁ 2 ₁ 2 ₁	
Paramètres de maille [Å]	a = 76,3 b = 97,6 c = 141,9	
λ [Å]	1,711	0,98
Source de RX	FIP	FIP
Résolution [Å]	2,70	2,28
I/ σ	13,6 (6,9)	9,3 (1,7)
Redondance	4,7 (1,3)	5,0 (2,6)
R _{fac}	0,034 (0,096)	0,054 (0,23)
R _{ano}	0,043 (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_{ano}, 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 ($\sim 1,5 \text{ \AA}$) avec un signal anormal relativement faible ($f'' \sim 6 e^-$) et un jeu à basse résolution ($\sim 2,7 \text{ \AA}$) 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 $2,7 \text{ \AA}$ 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_{libre}** : 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 300 mM	Gd-DO3A 300 mM	Gd-HPDO3A 300 mM	Gd-DOTA 300 mM
Méthode	SAD	SAD	SAD	MAD
λ [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; 0,16	1,06; 0,85	0,82; 0,52
Bsha [Å ²]	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 _{libre} (nb cycles dm)	0,216 (2)	0,180 (3) 0,185 (2)	0,138(1)	0,172 (3)

Urate oxydase	Gd-DOTMA 300 mM	Gd-DTPA-BMA 300 mM	Gd-DO3A 100 mM
Méthode	MAD	MAD	SAD
λ [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 ²]	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 _{libre} (nb cycles dm)	0,207 (3)	0,174 (3))	0,203 (2)

Thaumatine	Gd-DOTA 300 mM	Gd-DOTA-BOM 300 mM	Gd-DTPA-BMA 300 mM	Gd-DTPA 300 mM	Gd-HPDO3A 100 mM
Méthode	MAD	MAD	MAD	MAD	SAD
λ [Å]	0,92 1,71	0,92 1,71	0,92 1,71	0,92 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 ; 0,14
Bsha [Å ²]	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 _{libre} (n b cycles dm)	0,108 (16) 0,135 (3)	0,090 (5) 0,094 (3))	0,128 (3)	0,101(3) 0,096 (5)	0,113 (3)

Protéine X	Gd-HPDO3A 100 mM	Gd-DOTA-BOM 100 mM	Gd-DTPA 100 mM
Méthode	SAD	MAD	SAD
λ [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; 0,07	1,06	0,89
Bsha [Å ²]	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 _{libre} (nb cycles dm)	0,103 (3) 0,099 (4)	0,195 (3) 0,191 (5)	0,113 (2)

Glucose isomérase	Gd-DO3A 100 mM	Gd-DTPA 100 mM	Gd-DOTA-BOM 100 mM	Gd-DOTMA 100 mM	Gd-DTPA-BMA 100 mM
Méthode	MAD	MAD	MAD	SAD	SAD
λ [Å]	0,92 1,71	0,92; 1,71	0,92 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 0,75 0,46	0,47 0,12	0,21 0,22	0,40	0,13 0,08 0,25
Bsha [Å ²]	aniso aniso 20	22,05 1,00	16 13	28	43 4 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 _{libre} (nb cycles dm)	0,174 (3)	0,123 (3)	0,199 (3)	0,147 (2)	0,191 (3)

YGGV	Gd-DO3A 100 mM	Gd-DOTA-BOM 100 mM	Gd-DOTMA 100 mM
Méthode	MAD	MAD	MAD
λ [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 ²]	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 _{libre} (nb cycles dm)	0,111 (3)	0,161 (1)	0,100 (4) 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^- . 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.

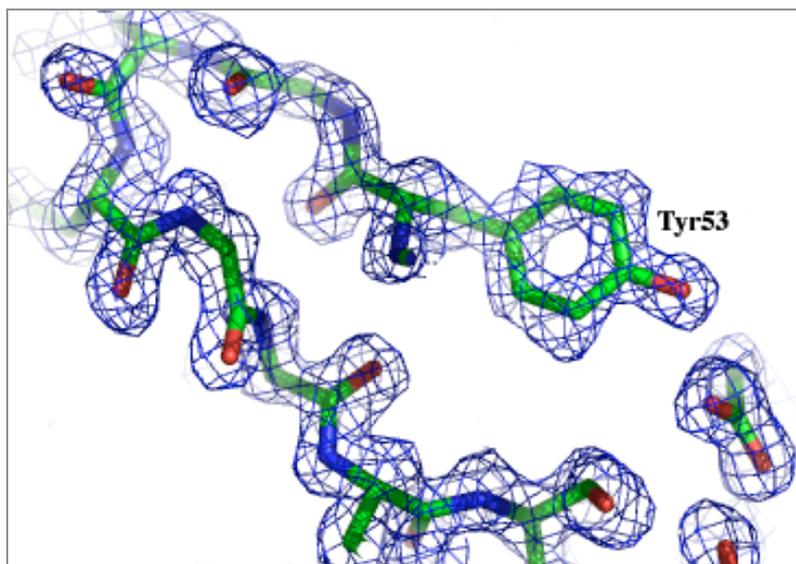


Figure 2.3.1 : Carte de densité électronique (contour 1σ) après SOLOMON pour le dérivé du lysozyme avec le Gd-HPDO3A calculée avec des données à une résolution de 1,53 Å.

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.

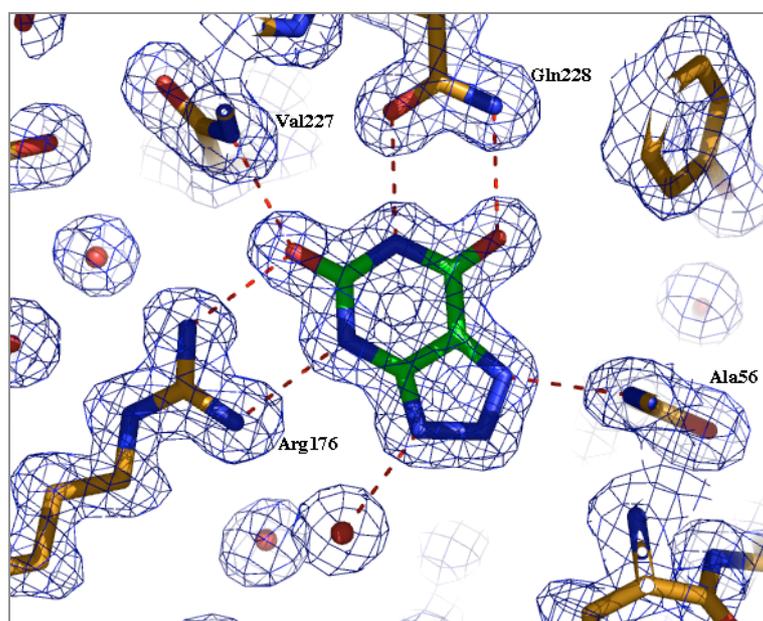


Figure 2.3.2 : Carte de densité électronique (contour 1σ) après SOLOMON pour le dérivé de l'urate oxydase avec le Gd-DTPA-BMA calculée avec des données à une résolution de 1,45 Å. La carte montre l'inhibiteur 8-azaxanthine et ses interactions avec les résidus voisins.

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.

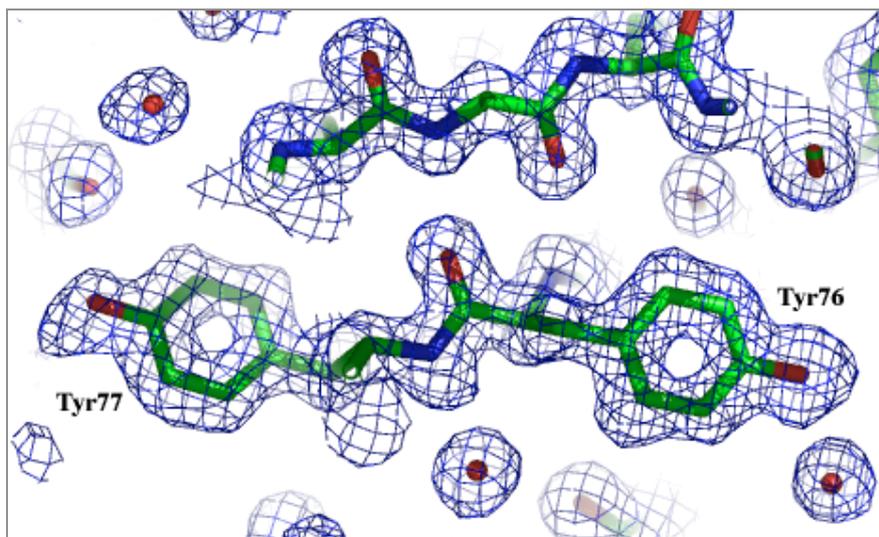


Figure 2.3.3 : Carte de densité électronique (contour 1 σ) après SOLOMON pour le dérivé protéine X + Gd-DOTA-BOM calculée avec des données à une résolution de 1,40 Å.

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.

YeaZ	Gd-DOTMA 100 mM
Méthode	MAD
λ [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 ²]	20; 62; 103; 1; 49; 49; 49; 49
FOM	0,436
FOMshasol	0,821
FOMdm	0,791
R _{libre} (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.

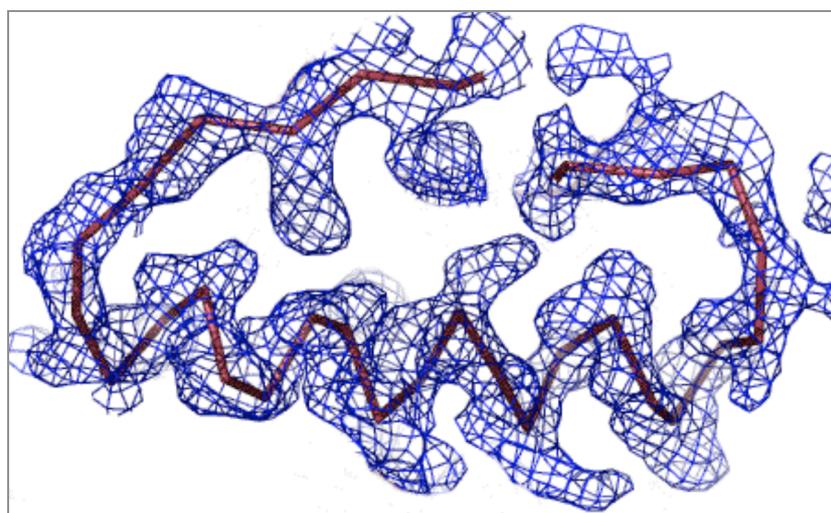


Figure 2.3.3 : Extrait de carte de densité expérimentale après nivellement de solvant pour le dérivé yeaz + Gd-DOTMA (contour 1σ). La carte est calculée avec des données à une résolution de 2,28 Å.

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 ₄) ₂ SO ₄		
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 ₄) ₂ SO ₄	(NH ₄) ₂ SO ₄		
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** : β -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 : $\frac{N_{succès}}{N_{test}} = \frac{23}{43} = 53\%$.

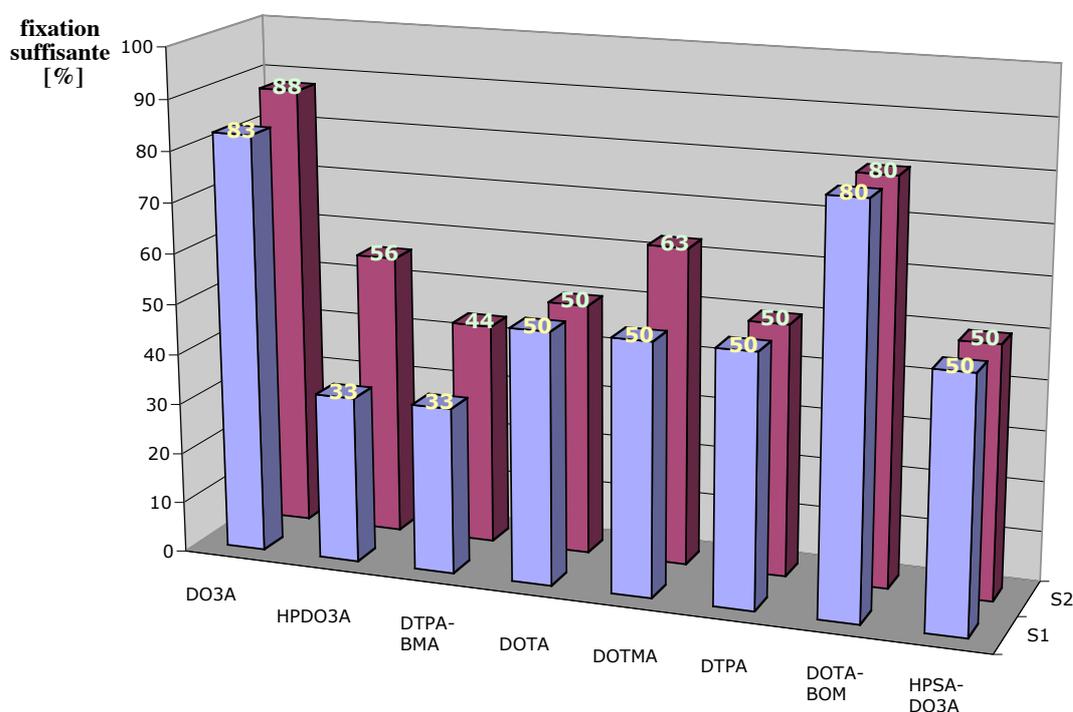


Figure 2.3.4 : Histogramme représentant le taux de succès pour l'obtention de dérivés avec une fixation suffisante du complexe comme fonction du complexe utilisé. Colonnes bleues : S1 : Résultats obtenus avec les protéines testées systématiquement. Colonnes violettes : S2 : Résultats comprenant tous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