

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 ρ_{PH} du

cristal dérivé peut alors être considérée comme la somme de la densité électronique ρ_P de la macromolécule seule, identique à celle du cristal natif, et de la densité électronique ρ_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° 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 ω_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 ω est la pulsation du rayonnement incident et α 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.

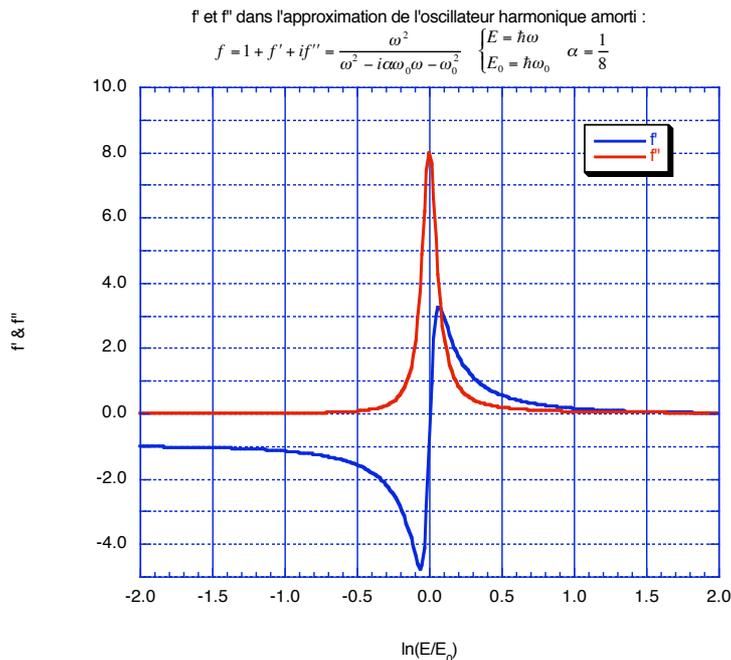


Figure 1.2.1 : Variation avec l'énergie d'excitation des parties réelles et imaginaires f' et f'' du facteur de diffusion avec le modèle de l'oscillateur harmonique amorti.

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 μ 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, λ la longueur d'onde du rayonnement, μ/ρ 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.

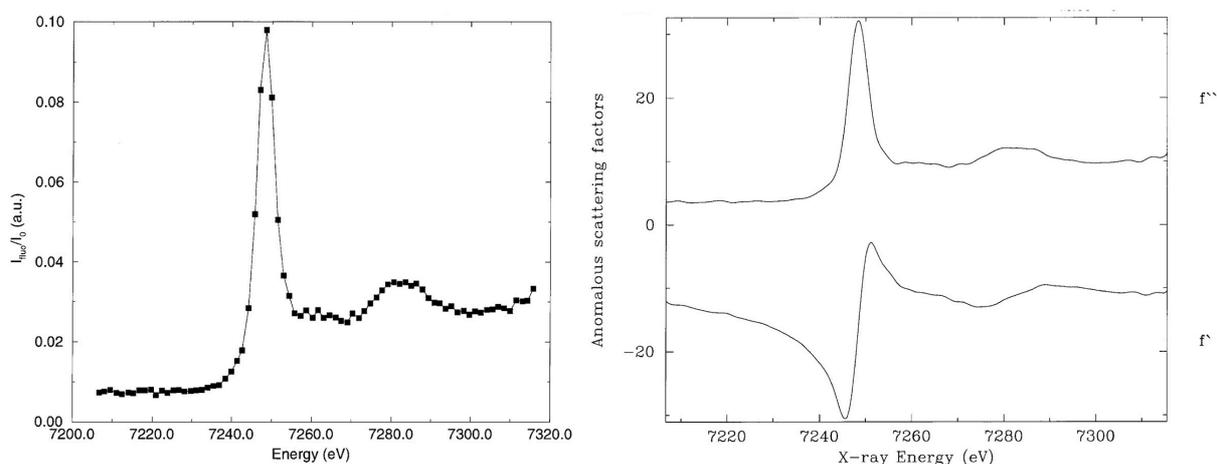


Figure 1.2.2 : Spectre d'absorption, enregistré en mode fluorescence à l'ESRF sur une solution contenant un complexe de gadolinium au seuil L_{III} du Gd (gauche) et les facteurs de diffusion anomaux f' et f'' du gadolinium en fonction de l'énergie du rayonnement incident obtenus à partir de ce spectre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

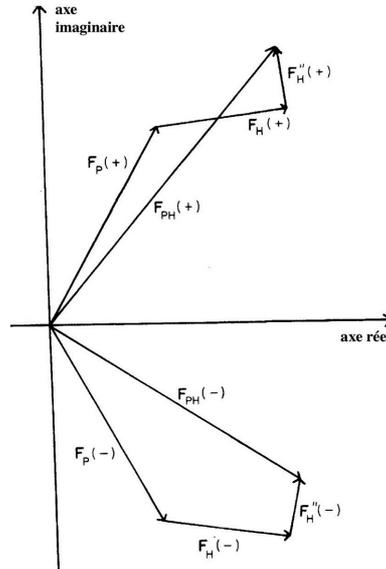


Figure 1.2.3 : Diagramme démontrant la brisure de la loi de Friedel, où "P" est la contribution des atomes qui diffractent normalement et "H" la contribution des atomes qui diffractent anormalement.

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

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 :

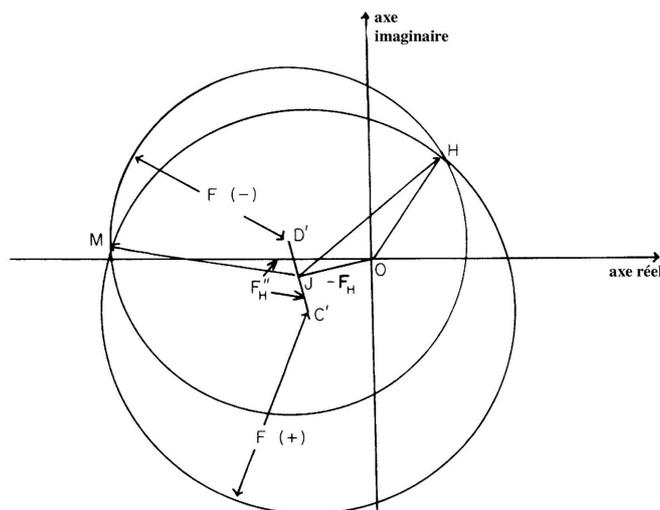


Figure 1.3.1 : Diagramme de Harker, illustrant la détermination des phases dans le cas de la diffusion anormale.

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 α_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 θ ,

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 λ^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 ($\lambda = 1,85 -0,72 \text{ \AA}$).

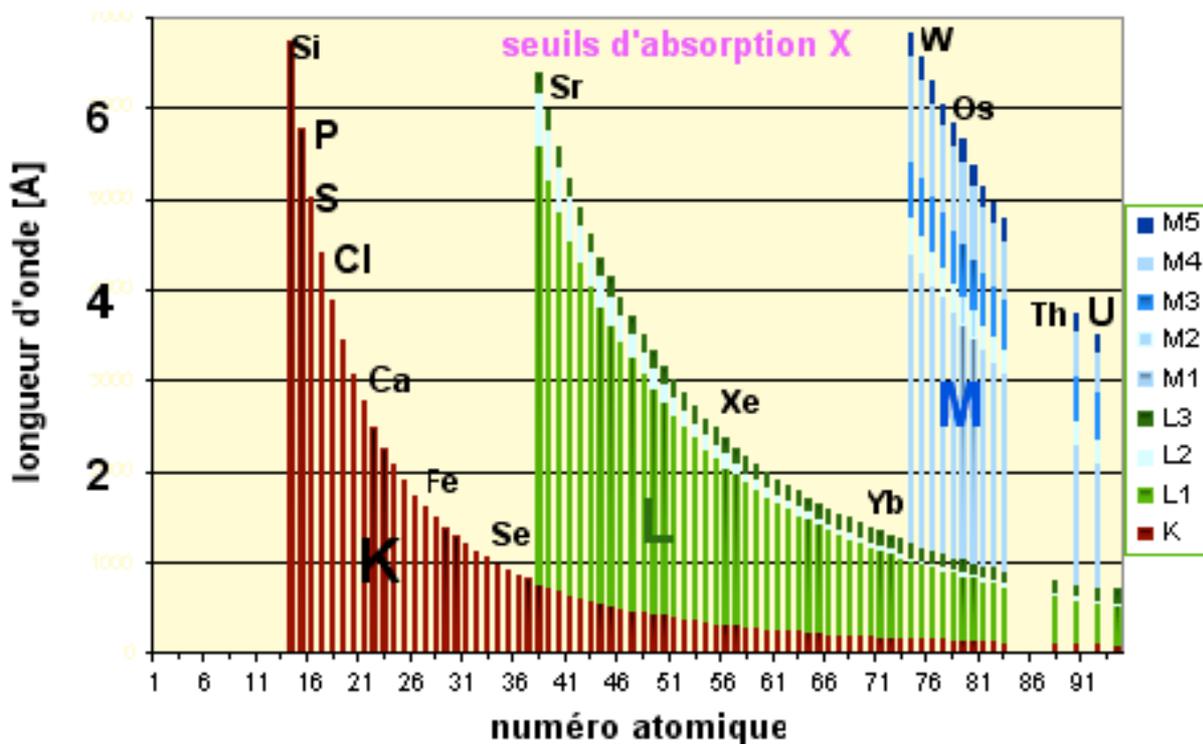


Figure 1.4.1 : Seuils d'absorption K, L, et M pour les éléments stables du tableau périodique.

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 \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 I₂/KI peuvent produire des dérivés incorporant les ions I, I₃⁻ ou I₅⁻ (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⁻ pour l'iode avec le rayonnement K α 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_{III}, 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^e 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 β_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 :

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

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	λ [Å]	f'' [e ⁻]	$\frac{\Delta F^{\pm}}{F}$ [%]	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 ₂ HgI ₄	1	1	seuil : 1,0	10,1	7,7	-
K ₂ HgI ₄	1	1	1,54	7,7	5,9	-
K ₂ PtCl ₄	1	0,5	seuil : 1,1	10,2	3,9	-
UO ₂ (C ₂ H ₃ O ₂) ₂	2	0,25	1,54	16	4,3	-
Xe	3	0,3	1,54	7,4	2,9	-
Gd ³⁺	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	$\lambda_{\text{seuil}}[\text{Å}]$	$f''_{\text{seuil}}[e^-]$	$f''_{\text{CuK}\alpha} [e^-]$ ($\lambda=1,54\text{Å}$)	$f''_{\text{CrK}\alpha} [e^-]$ ($\lambda=2,29\text{Å}$)
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.

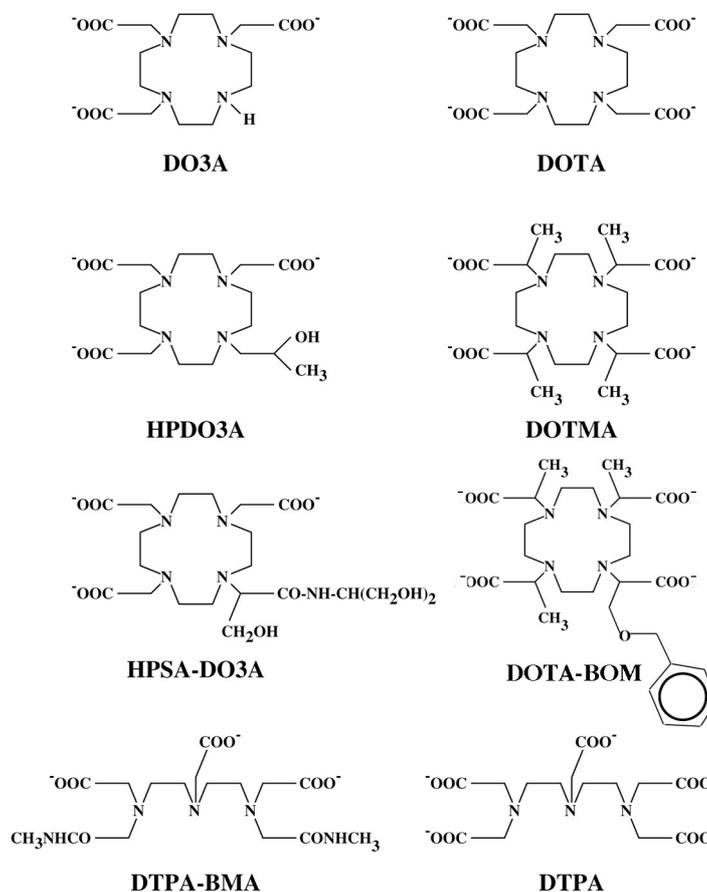


Figure 1.4.1. : Formules développées des ligands des complexes utilisés.

La figure 1.4.1. montre les ligands des complexes utilisés. Tous ces ligands chélatent un ion de Gd^{3+} mais peuvent tout aussi bien chélater les autres ions, Ln^{3+} , 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 pKa 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 $\log K \sim 22$ (Gd-DO3A) et $\log K \sim 28$ [Gd-DOTA] (à titre de comparaison, pour le complexe [Gd-EDTA], $\log K = 17,35$) (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 ($pH \leq 2$) 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 Ln^{3+} . Pour l'ion 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 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 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 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 (MAGNEVISTTM, 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'' .

Élément	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

Élément	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.