



UNIVERSITÉ DE STRASBOURG

Bio - Cristallographie

Déterminations des structures 3D des macromolécules
biologiques par diffraction des rayons X

JEAN CAVARELLI
Université de Strasbourg,
Centre de Biologie Structurale Intégrative,
Institut de Génétique et de Biologie Moléculaire et Cellulaire
1, rue Laurent Fries, - 67404 Illkirch-Grafenstaden Cedex
Tél : 03 69 48 52 74 - Fax : 03 88 65 32 76 - e-mail : jean.cavarelli@unistra.fr

Préambule

Les molécules biologiques responsables de toute vie cellulaire sont des hétéropolymères de très grande taille appartenant à deux familles : les protéines et les acides nucléiques. Les processus biologiques sont les résultats d'interactions complexes et dynamiques (dans l'espace et dans le temps) soit de macromolécules biologiques entre elles, soit de macromolécules avec de petits substrats cellulaires. La connaissance des structures tridimensionnelles de ces macromolécules soient seules, soient engagées dans des complexes spécifiques, est essentielle pour la compréhension (à l'échelle atomique) des fonctions biologiques. Les structures 3D sont des outils précieux dans l'étude des réactions complexes à l'origine des mécanismes du vivant et jouent de plus un rôle intégrateur et fédérateur dans le processus complexe et pluridisciplinaire allant d'une tumeur à son médicament et couvrant des domaines de recherche allant de la génomique intégrative à la modélisation moléculaire. La connaissance de ces structures est l'un des piliers actuels de la biologie moléculaire et représente une source de progrès qui génère des retombées non seulement en recherche fondamentale mais aussi en recherche appliquée (domaine de la santé humaine, biotechnologies). Cela justifie les investissements importants réalisés depuis plusieurs années dans les secteurs publics et privés.

La diffraction des rayons X par des monocristaux est la méthode par excellence pour l'étude des macromolécules biologiques à l'échelle atomique. La cristallographie a permis la détermination des structures tridimensionnelles de plusieurs dizaines milliers de macromolécules biologiques dans des gammes de taille et de complexité très variées : petites protéines, oligonucléotides, acides ribonucléiques de transfert, immunoglobulines, complexes multi-enzymatiques, complexes nucléo-protéiques, virus d'insectes, de plantes ou de mammifères. Les propriétés physico-chimiques intrinsèques des macromolécules biologiques donnent naissance à des cristaux avec de grands paramètres de maille cristalline et un pouvoir de diffraction en général limité en comparaison du standard actuel des petites molécules organiques. Ceci impose des méthodes et des techniques adaptées tout au long du processus cristallographique. Cette méthodologie propre aux macromolécules biologiques va être présentée dans ce chapitre. L'explosion actuelle de cette méthode est due aux progrès réalisés tant au niveau de la technologie (biologie moléculaire, source de rayons X, cryo-cristallographie, détecteurs de rayons X, moyens de calculs) qu'au niveau des logiciels de traitements des données de diffraction (collecte, phasage, affinement). Bien que reposant sur des méthodes et des techniques compliquées, les progrès réalisés permettent une automatisation de plus en plus poussée du processus de détermination de structure. Ceci se traduit par un raccourcissement extraordinaire du délai séparant l'obtention d'un premier cristal de qualité à la détermination de la structure cristalline. Si la limite de diffraction est raisonnable, une structure cristallographique peut être maintenant obtenue en quelques heures dans les cas les plus favorables) après l'obtention des premières données expérimentales et ceci avec un minimum d'intervention humaine. Cette révolution des années 2000 permet à la cristallographie biologique de relever de nouveaux défis à la fois en recherche fondamentale et en recherche appliquée. Chaque nouvelle structure tridimensionnelle bouleverse et balaye la vision (parfois simpliste et très incomplète) du processus par lequel une fonction biologique est assurée. Ces nombreux progrès méthodologiques et techniques permettent d'aborder des problèmes de plus en plus complexes.

La lecture de cet article suppose une connaissance initiale de la cristallographie géométrique et une première initiation à la théorie de la diffraction des rayons X par des monocristaux. Le lecteur pourra consulter quelques références données à la fin de cet article pour compléter sa formation initiale.

Table des matières

1	Présentation générale	6
2	Purification des macromolécules biologiques	8
3	Cristallisation	11
4	Enregistrement des données de diffraction	15
4.1	Source de rayons X	16
4.2	Les détecteurs	17
4.3	Méthodes d'enregistrement	19
4.3.1	Collecte en faisceau monochromatique	20
4.3.2	Collecte en faisceau polychromatique	24
4.4	Cryocristallographie	26
5	Introduction à la détermination des phases	28
6	Méthode de la série isomorphe	33
6.1	Détermination des phases : Cas Idéal	33
6.2	Détermination des phases : Cas réel	35
6.3	Obtention des dérivés lourds	37
6.4	Détermination de la position des atomes lourds.	37
7	Utilisation de la diffusion anormale	38
7.1	Travail à une seule longueur d'onde : Formalisme SAD. Vision graphique	41
7.2	Travail à plusieurs longueurs d'ondes. Vision graphique	43
7.3	Travail à plusieurs longueurs d'ondes. Les mathématiques.	44
7.4	Incorporation des diffuseurs anormaux	45
7.5	Détermination des f' et f''	46
7.6	Développements et Perspectives	46
8	Le remplacement moléculaire	47
8.1	La fonction de rotation	48
8.2	La fonction de translation	49
8.3	Remarques	49
8.4	Détermination des phases	50
9	Cartes de densité électronique	50
9.1	Définitions	50
9.2	Amélioration par des méthodes de modification de densité	52
9.2.1	Nivellement de solvant	52
9.2.2	Atomicité	54
9.2.3	Histogrammes des densités	54
9.2.4	Symétrie non cristallographique	54
9.3	Extension des phases	55
9.4	Construction de la structure	55

10 Affinement d'une structure cristallographique	55
10.1 Introduction	55
10.2 Principes	56
10.3 Les méthodes de moindres carrés	60
10.4 Les méthodes de la dynamique moléculaire	60
10.5 Maximum de vraisemblance	61
10.6 Carte de densité électronique	62
10.7 Particularités des macromolécules biologiques	62
11 Validation et contrôle qualité des structures	63
12 Base de données PDB	65
13 Détermination des structures à haut débit, les Années 2000.	65
14 Perspectives	67
15 Bibliographie	69
Index	70

Avertissement :

L'origine de cette revue est un article qui a été écrit en 1998 pour la revue "Les Techniques de l'Ingénieur" (adresse internet <http://www.techniques-ingenieur.fr>). La première version de cet article, qui a été publiée en Mai 2000, dans le traité "Analyse et Caractérisation", P1090 : 1-20, est parue sous le titre "Cristallographie des macromolécules biologiques". Une mise à jour de l'article initial a été réalisé en 2009 en 2 articles :

1. Détermination des structures 3D des macromolécules biologiques par diffraction des rayons X, Partie I. Techniques de l'Ingénieur, Traité Analyse chimique et Caractérisation, (2009), Vol. TA1, No. P1110, 1-12.
2. Détermination des structures 3D des macromolécules biologiques par diffraction des rayons X, Partie II. Techniques de l'Ingénieur, Traité Analyse chimique et Caractérisation, (2009), Vol. TA1, No. P1111, 1-19.

Toute reproduction de ces articles sans autorisation du Centre français d'exploitation du droit de copie est strictement interdite par la loi.

Mises à jour :

- décembre 2000, plusieurs modifications significatives.
- octobre 2001, plusieurs modifications.
- décembre 2003, plusieurs modifications.
- octobre 2004, modifications mineures.
- décembre 2005, plusieurs modifications.
- décembre 2007, plusieurs modifications.
- Octobre 2008, plusieurs modifications significatives.
- Octobre 2009, plusieurs modifications.
- Octobre 2015, quelques modifications.

Depuis la première version de cet article, de nombreux développements et transformations majeures ont eu lieu, à la fois sur les aspects méthodologiques et sur les développements technologiques. Cela ne modifie pas les principes de bases qui sont présentés ici et qui ne constituent qu'une introduction à la cristallographie des macromolécules.

1 Présentation générale

Les acides nucléiques sont des polynucléotides qui utilisent un alphabet de 4 bases, dont l'unité monomérique contient un sucre furanose, un groupement phosphate et une base hétérocyclique (soit purique (2 possibilités) soit pyrimidiques (2 possibilités)). On distingue deux grandes familles d'acide nucléique, celle de l'ADN (acide désoxyribonucléique) et celle de l'ARN (acide ribonucléique). Dans un ADN, le sucre est un 2'-désoxy- β -D ribose et les deux bases puriques sont l'Adénine (A), la guanine (G), alors que les deux bases pyrimidiques sont la thymine (T) et la cytosine (C). Dans les ARNs, le sucre est un β -D ribose et la thymine est remplacée par l'uracile. Les protéines sont des polymères construits essentiellement avec une bibliothèque principale de 20 acides aminés qui ont chacun des propriétés physico-chimiques originales. La vie d'une cellule est le résultat d'associations dynamiques entre ces macromolécules biologiques, et la fonction biologique d'une macromolécule donnée est gouvernée par sa structure dans l'espace (structure tridimensionnelle). La connaissance des structures tridimensionnelles est donc une condition nécessaire à une compréhension à l'échelle atomique des fonctions biologiques. Connaître une structure, ce n'est pas uniquement la décrire mais essayer de répondre à certaines questions de manière à expliquer et à prévoir :

- Expliquer, c'est pouvoir décrire le fonctionnement d'une molécule, les rapports entre la séquence, la structure et la fonction, les mécanismes par lesquels elle interagit avec d'autres partenaires.
- Prévoir, c'est prédire le comportement dynamique de cette molécule, les interactions avec ses partenaires. On pourra alors essayer de les modifier dans les cas où par exemple ces interactions sont néfastes pour la vie cellulaire (pharmacologie, thérapie génique).

On comprend ici que la structure tridimensionnelle est le noeud central de communication entre la séquence d'une macromolécule et sa (ou ses) fonction(s) biologique(s) et donne des clés pour comprendre l'évolution au niveau moléculaire. La visualisation de cette structure tridimensionnelle ouvre et révèle un champ vaste (et souvent inimaginable) à la connaissance. Elle apporte une vision des mécanismes biologiques que l'on essayait (parfois en vain) de délimiter par des expériences indirectes.

Deux méthodes sont actuellement utilisées pour déterminer les structures tridimensionnelles des macromolécules biologiques à l'échelle atomique : la diffraction des rayons X et la résonance magnétique nucléaire (RMN). Près de 85% des structures connues à ce jour ont été déterminées (on dit résolues) par diffraction des rayons X. Contrairement à la RMN, la cristallographie ne souffre pas de limitations en taille de la macromolécule étudiée et ne présente qu'une seule barrière : être capable d'obtenir des cristaux de la macromolécule étudiée.

En cristallographie, on utilise l'interaction élastique cohérente des rayons X avec les électrons des atomes pour avoir une information structurale à l'échelle atomique. Un échantillon quelconque (de volume V) est caractérisé en chaque point de l'espace (repéré par le vecteur \vec{r}) par une densité électronique $\rho(\vec{r})$. Lorsqu'une onde plane monochromatique, de vecteur d'onde \vec{s}_0 , interagit avec un échantillon, il y a diffusion dans toutes les directions de l'espace. Pour une direction de diffusion caractérisée par un vecteur d'onde \vec{s} (qui fait un angle 2θ avec l'onde incidente), l'amplitude et la phase de l'onde diffusée peuvent être représentés par un nombre complexe, appelé facteur de structure. Pour une direction de diffusion donnée, le facteur de structure s'exprime de manière simple en fonction du vecteur de diffusion $\vec{S} = \vec{s} - \vec{s}_0$:

$$F(\vec{S}) = \int_V \rho(\vec{r}) e^{2i\pi(\vec{S} \cdot \vec{r})} d^3\vec{r} \quad (1)$$

Si l'échantillon considéré est un cristal, c'est à dire un arrangement périodique ordonné d'un motif dans l'espace tridimensionnel¹ (supposé infini à l'échelle atomique), on peut montrer que la diffusion est alors restreinte à certaines directions privilégiées de l'espace (phénomène de diffraction). Les conditions de diffraction sont exprimées par les équations de Laue, qui sont satisfaites si et seulement si le vecteur \vec{S} est un vecteur du réseau réciproque du cristal (construit sur les $\vec{a}^*, \vec{b}^*, \vec{c}^*$) :

$$\vec{S} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^* \quad (2)$$

avec $(h,k,l) \in \mathbb{Z}$. Pour (h,k,l) donné, le facteur de structure de l'onde diffusée dans la direction $\vec{s} = \vec{S}_{hkl} + \vec{s}_0$, sera alors noté F_{hkl} et on parlera alors de la réflexion hkl .

Inversement, la connaissance de l'ensemble des facteurs de structure F_{hkl} permet de calculer la fonction densité électronique correspondante. Cette fonction reflète une distribution continue d'électrons dans la maille cristalline et présente des maxima dans les régions occupées par les atomes. Elle est donc en général interprétée en termes de positions atomiques.

Pour une forme cristalline, la densité électronique $\rho(x, y, z)$ en un point de coordonnées relatives (x, y, z) de la maille cristalline est calculée en théorie par la relation :

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F_{hkl} e^{-2i\pi(hx+ky+lz)} \quad (3)$$

où F_{hkl} est le facteur de structure pour la réflexion hkl et V le volume de la maille cristalline. Le facteur de structure, nombre complexe, s'écrit :

$$F_{hkl} = |F_{hkl}| e^{i\phi_{hkl}} \quad (4)$$

où $|F_{hkl}|$ est le module et ϕ_{hkl} la phase du facteur de structure pour la réflexion hkl . Le facteur de structure est relié à la nature et la position des N atomes présents dans la maille par la relation :

$$F_{hkl} = \sum_{j=1}^N q_j f_j w_j^{hkl} e^{2i\pi(hx_j+ky_j+lz_j)} \quad (5)$$

où f_j^{hkl} est le facteur de diffusion de l'atome j de coordonnées relatives (x_j, y_j, z_j) , q_j le facteur d'occupation, w_j^{hkl} le facteur de Debye-Waller. Le facteur f_j caractérise l'interaction d'un atome j , ou plus exactement de son nuage électronique, avec les rayons X incidents. Les facteurs q_j , w_j^{hkl} traduisent le désordre et le mouvement de chaque atome². Une structure cristallographique correspond à une double moyennation dans le temps et dans l'espace, car elle représente la moyenne des structures présentes dans le cristal et pendant le temps de l'expérience (de collecte des données de diffraction).

1. le cristal est décrit par une maille directe définie par trois vecteurs $(\vec{a}, \vec{b}, \vec{c})$ et une symétrie correspondant à l'un des 230 groupes d'espace.

2. Dans de nombreux cas, on utilise une simplification du facteur de Debye-Waller par un facteur dit "d'agitation thermique" isotrope qui sera alors caractérisé par un paramètre atomique unique noté B_j , (voir paragraphe 10) On utilise maintenant le terme de paramètres de déplacements anisotropes.

La détermination de la phase ϕ_{hkl} de chaque réflexion constitue l'une des difficultés majeures de la cristallographie, car les ondes diffractées par un cristal ne peuvent être refocalisées et les phases sont perdues.

Seules les amplitudes des ondes diffusées sont accessibles à l'expérience, un détecteur ne mesurant pour une réflexion donnée que l'intensité I_{hkl} de l'onde diffractée. Celle-ci est proportionnelle au carré du module du facteur de structure (hypothèse d'un cristal de parfaite mosaïcité) :

$$I_{hkl} \propto (|F_{hkl}|)^2 = F_{hkl}F_{hkl}^* \quad (6)$$

où F_{hkl}^* est le complexe conjugué de F_{hkl} . Pour résoudre ce problème, les méthodes directes utilisées en cristallographie de petites molécules (seulement quelques centaines d'atomes) ne permettent pas de résoudre directement les structures des macromolécules de grande tailles. En effet, ces méthodes directes utilisent des relations de contraintes entre facteurs de structure (équation de Sayre, formule de la tangente) dont l'application nécessite des données de diffraction à très haute résolution (de l'ordre de 1Å, 1Å = 10⁻¹⁰m). La cristallographie biologique utilise trois grandes techniques pour déterminer les phases : le remplacement isomorphe, la diffusion anormale, le remplacement moléculaire. Les phases connues, il est alors possible de calculer une carte de densité électronique (équation 3) qui permet, selon la limite de diffraction des données et l'exactitude des phases, de déterminer une première structure de la molécule qui sera ensuite affinée.

Dans cette revue, nous allons décrire les étapes nécessaires à la détermination de la structure tridimensionnelle d'une macromolécule biologique. On peut schématiquement décrire ce processus en six étapes : obtention de la macromolécule à l'état pur (où des macromolécules dans le cas d'assemblages), cristallisation, collecte de données de diffraction, phasage, construction du modèle par interprétation des cartes de densité électronique, affinement et validation de la structure. La figure 1 présente une vue globale du processus cristallographique.

Pour le problème des phases, nous présenterons trois méthodes. Deux méthodes, le remplacement isomorphe multiple et la diffusion anormale, ne font aucune hypothèse sur la structure tridimensionnelle recherchée. Une autre méthode, le remplacement moléculaire, suppose la connaissance d'une structure parente de la molécule étudiée. De par sa fonction et son but, la cristallographie des macromolécules biologiques se situe à l'interface de plusieurs domaines complémentaires comme la biologie, la chimie, la physique et nécessite de bonnes connaissances de l'outil informatique et des mathématiques. Nous nous contenterons de présenter quelques notions de base qui devraient permettre une compréhension des problèmes posés et la manière de les résoudre actuellement.

2 Purification des macromolécules biologiques

L'obtention de cristaux de qualité nécessite de disposer le plus souvent de plusieurs milligrammes de la macromolécule d'intérêt à l'état soluble et avec une très grande pureté. Les premières structures tridimensionnelles de macromolécules correspondaient à des molécules qui pouvaient facilement être obtenues en grande quantité et qui cristallisaient relativement facilement. Aujourd'hui, on aborde une étude cristallographique d'une macromolécule pour résoudre un problème biologique donné. La macromolécule cible peut être disponible en faible quantité, voir pratiquement insoluble. Néanmoins, la croissance importante du nombre de

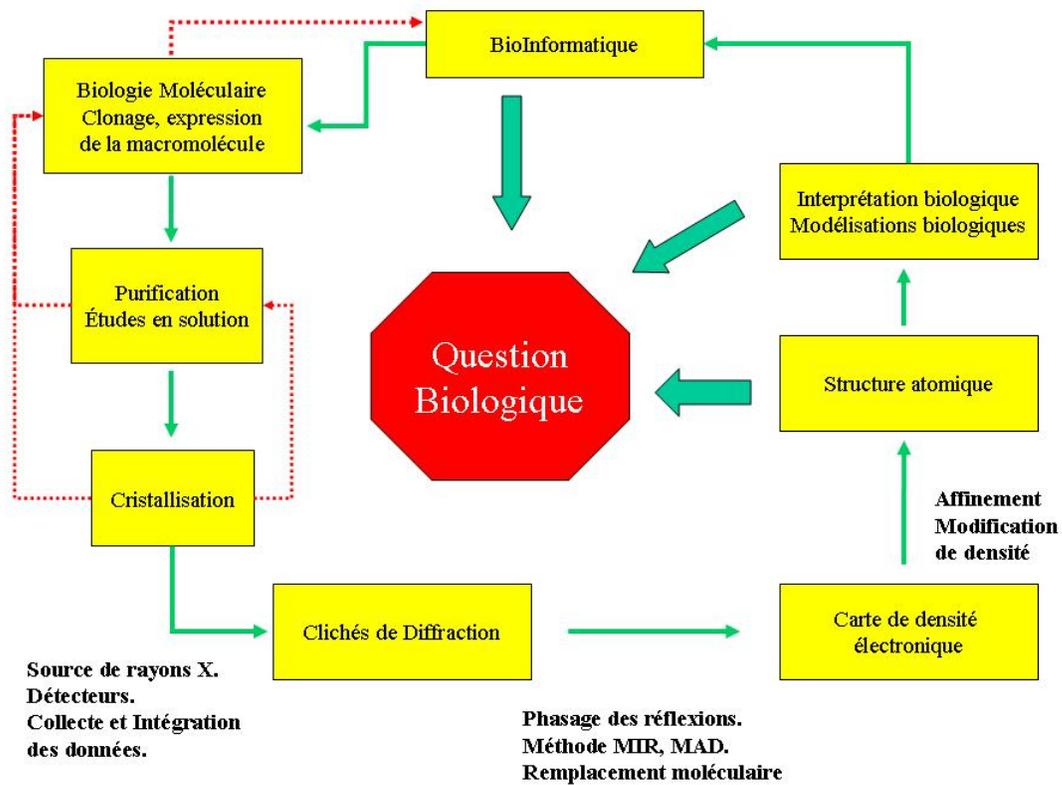


FIGURE 1 – Résolution de la structure d’une macromolécule biologique par diffraction des rayons X. Vue schématique des différentes étapes.

structures tridimensionnelles résolues vient de l'utilisation des techniques de la biologie moléculaire. Dès que le gène correspondant à la protéine étudiée a pu être indentifié, il peut être cloné dans un vecteur d'expression et exprimé dans une cellule hôte procaryotique ou eucaryotique. Les systèmes de sur-expression les plus utilisés pour les études structurales sont la bactérie *Escherichia coli*, les cellules d'insectes associées à un vecteur baculovirus, des cellules de mammifères et la levure. Le système bactérien est le plus facile d'utilisation et permet dans de nombreux cas un niveau élevé de surexpression. Néanmoins, ce système présente deux inconvénients majeurs : (a) de nombreuses protéines eucaryotiques ne se replient pas correctement et précipitent sous forme d'amas insoluble de protéines appelés corps d'inclusion, (b) de nombreuses modifications post-traductionnelles spécifiques aux systèmes d'origine eucaryotique ne sont pas effectuées. Dans de nombreux vecteurs d'expression, la protéine est fusionnée soit avec une autre protéine particulière soit avec une queue peptidique, ce qui permet une première étape de purification par colonne d'affinité. Par exemple, l'addition d'une queue polyhistidine permet une première purification de la protéine sur une colonne d'affinité en présence d'ions nickel que chélate la séquence polyhistidine. Un site de clivage à une protéase donnée permet ensuite l'obtention de la protéine seule.

Les techniques de génie génétique permettent aussi d'apporter des modifications ponctuelles de la séquence de la protéine pouvant modifier à bon escient son comportement physico-chimique. De telles modifications sont parfois cruciales pour le succès du projet. On peut classer ces modifications en plusieurs catégories en fonction du but recherché :

- Amélioration de la solubilité de la protéine. Une modification de certains acides aminés peut complètement modifier la solubilité et permettre ainsi l'étude structurale. Dans certains cas, une seule mutation ponctuelle peut changer drastiquement la solubilité de la macromolécule.
- Renforcement des contacts cristallins. Les contacts entre molécules dans un cristal se font par des interactions entre acides aminés exposés à la surface des protéines. La mutation d'acides aminés susceptibles d'être exposés en surface peut permettre de renforcer les contacts cristallins et d'améliorer notablement la qualité de diffraction.
- Détermination des phases. La mutagénèse dirigée permet aussi d'introduire si nécessaire des acides aminés susceptibles de réagir avec des composés d'atomes lourds nécessaires pour le processus de détermination des phases. L'introduction de sélénium dans les protéines, réalisée en remplaçant une méthionine par une sélénométhionine permet d'utiliser la méthode de la diffusion anormale à plusieurs longueurs d'onde (Méthode MAD, voir paragraphe 7).

De nombreuses protéines de grandes tailles sont composées de plusieurs domaines fonctionnels, chacun des domaines pouvant avoir une structure et une fonction propre et indépendante. Les techniques de la biologie moléculaire permettent alors de cloner un domaine spécifique afin d'étudier une fonction donnée.

Le processus de purification qui suit l'étape de surexpression et qui permet d'isoler la molécule d'intérêt est en général constitué d'une succession d'étapes qui vont exploiter les propriétés physico-chimiques propres à une macromolécule donnée pour la séparer d'un mélange. Elles impliquent le plus souvent :

- une précipitation par des sels, des solvants organiques ou des polymères,
- des méthodes chromatographiques qui utilisent la différence de partition d'une macromolécule entre une phase stationnaire et une phase mobile. Les propriétés physico-chimiques qui sont mises à profit peuvent être :

- la distribution de charges à la surface de la macromolécule (chromatographies par échanges d'ions),
- la présence de groupement particuliers (chromatographies d'affinité)
- l'hydrophobicité (chromatographies d'interactions hydrophobes),
- la taille et la forme de la macromolécule (gel filtration).

Cette étape très importante du processus cristallographique ne sera néanmoins pas détaillée dans cet article et le lecteur pourra consulter les références et liens donnés en bibliographie pour une étude plus complète.

3 Cristallisation

La cristallisation est un changement de phase qui fait passer un composé à l'état solide ordonné. Un cristal est un arrangement périodique ordonné dans l'espace tridimensionnel, la maille cristalline étant l'unité de base qui se repète par translations selon la loi décrite par un réseau ponctuel. En 3 dimensions, il existe 14 réseaux ponctuels possibles (dit de Bravais) qui sont décrits uniquement par 7 mailles : les 7 systèmes cristallins. Chaque système est caractérisé par une symétrie, ce qui implique des contraintes sur les paramètres de maille (voir tableau 1). En combinant les différentes symétries, il existe uniquement 230 groupes d'espaces qui décrivent un nombre limité de manière d'ordonner périodiquement un motif dans l'espace.

L'empilement des molécules dans un cristal possède donc un ordre à longue distance, les liens entre molécules étant identiques d'une maille cristalline à l'autre. Par opposition, on parlera de précipité amorphe pour décrire un empilement non ordonné, sans symétries de translations. Les macromolécules sont cristallisées à partir d'une solution aqueuse, dans des conditions physico-chimiques (température, pH, pression, etc...) où elles gardent leur activité biologique. Bien que plusieurs centaines de macromolécules biologiques aient déjà été cristallisées, permettant ainsi de connaître leurs structures, la plupart des facteurs responsables de la cristallisation sont mal maîtrisés et il n'existe pas de règles directement applicables à une nouvelle macromolécule. Pourtant, la cristallisation des macromolécules biologiques est régie par les mêmes lois thermodynamiques que celles qui gouvernent la cristallisation des petites molécules organiques. L'obtention d'un cristal peut se décomposer en quatre étapes successives ou simultanées : formation d'un état de sursaturation, nucléation, croissance cristalline et cessation de croissance. Néanmoins, les macromolécules biologiques présentent de nombreuses sources potentielles d'hétérogénéités qui peuvent nuire fortement à leur capacité de cristalliser. On peut citer par exemple la présence :

- de molécules contaminantes, macromolécules ou petites molécules étrangères qui n'ont pu être éliminées par les étapes de purification,
- de micro-hétérogénéités de séquences,
- de variations génétiques et de dégradation telles que protéolyse, modifications post-traductionnelles partielles ou hétérogènes (glycosylation, phosphorylation, méthylation),
- de micro-hétérogénéités conformationnelles : régions flexibles, équilibres entre différentes formes oligomériques, aggrégation, dénaturation.

De plus, les macromolécules peuvent subir des changements de conformation, ou de charge en fonction de la température, du pH ou de l'état de solvatation. Elles sont très sensibles à

Système	Symétrie génératrice	Groupes ponctuels de symétrie	Contraintes imposées	Réseaux de Bravais
Triclinique	Identité ou inversion	$1, \bar{1}$	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma$	P
Monoclinique	1 symétrie d'ordre 2	$2, m, 2/m$	$a \neq b \neq c$ $\alpha = \gamma = 90^\circ$ β	P, C
Orthorhombique	3 symétrie d'ordre 2 \perp	$222, mm2, mmm$	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	P, C, I, F
Quadratique	1 axe 4	$4, \bar{4}, 4/m$ $422, 4mm, \bar{4}2m$ $4/mmm$	$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	P, I
Hexagonal	1 axe 6	$6, \bar{6}, 6/m$ $622, 6mm, \bar{6}2m$ $6/mmm$	$a = b \neq c$ $\alpha = \beta = 90^\circ \quad \gamma = 120^\circ$	P
Trigonal Rhomboédrique	1 axe 3	$3, \bar{3}$ $32, 3m, \bar{3}m$	$a = b \neq c$ $\alpha = \beta = 90^\circ \quad \gamma = 120^\circ$ axe 3 $\parallel \vec{c}$ $a = b = c$ $\alpha = \beta = \gamma$ axe 3 $\parallel \vec{a} + \vec{b} + \vec{c}$	P R
Cubique	4 axes 3	$23, m\bar{3}$ $423, \bar{4}3m, m\bar{3}m$	$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$	P, I, F

Tableau 1 – Les 7 systèmes cristallins et leurs symétries.

l'environnement et peuvent facilement se dégrader ou se dénaturer. Il est donc nécessaire de maintenir un état d'hydratation, des conditions de pH et de température physiologiques.

Les techniques de la chimie analytique fine (diffusion de lumière, spectroscopie de masse, dichroïsme circulaire) sont maintenant appliquées aux macromolécules pour contrôler leur état en solution (agrégation, hétérogénéité de population en masse ou en charge, structures non repliées). Une modification des étapes de purification, voire de clonage et de surexpression est souvent nécessaire pour obtenir des cristaux de qualité suffisante pour une étude structurale.

La plupart des méthodes de cristallisation utilisent un agent précipitant pour diminuer la solubilité de la macromolécule. Les paramètres que l'on va faire varier pour obtenir la formation d'un cristal sont nombreux et une telle étude est très gourmande en quantité de macromolécules. On peut citer par exemple le pH, la nature et la concentration du tampon, la force ionique, la température, la concentration en macromolécule, la nature et concentration de l'agent précipitant, la présence d'additifs. Quatre grandes catégories d'agents précipitants sont généralement utilisés :

- les sels. Parmi ceux-ci, le sulfate d'ammonium est incontestablement le plus utilisé, pour sa grande efficacité à précipiter les protéines et sa forte solubilité dans l'eau (3,9M à 0°C, 4,1M à 25°C).
- les polymères à longue chaîne et les polymères de bas poids moléculaires. Le polyéthylène glycol (PEG) est le plus utilisé. Le PEG est un polymère que l'on peut trouver en différents poids moléculaires, habituellement de 400 à 20000.
- les solvants organiques et les composés organiques non volatils. Pour les protéines, à l'exception du méthyl-2,4-pentanediol (MPD), les solvants organiques sont plutôt utilisés comme additifs à un autre agent précipitant. Les plus utilisés sont l'éthanol, acétone, et les butanols. Ils doivent parfois être utilisés avec précaution en raison de leurs capacités dénaturantes. Par contre, les solvants organiques ont été utilisés avec succès pour la cristallisation des acides nucléiques (ARN de transfert, oligonucléotides). Ceci est dû d'une part, à la plus grande tolérance des polynucléotides pour ces composés organiques, et d'autre part, à leur surfaces polyanioniques qui sont plus sensibles que les protéines aux variations de constante diélectrique.

Deux techniques de cristallisation sont principalement utilisées en cristallographie biologique :

- la diffusion de vapeur. Elle peut se faire par évaporation de solvant ou échange de solvant contre un réservoir plus concentré, ou échange de solutés volatils.
- la diffusion de liquide. Celle-ci est réalisée soit directement par une interface liquide-liquide, soit à l'aide d'une membrane à dialyse ou de gels.

Nous ne présenterons que la méthode par diffusion de vapeur. Dans celle-ci, la sursaturation est atteinte par évaporation d'espèces volatiles. Dans une enceinte fermée, on place sans les mettre en contact, deux solutions :

- une solution de faible volume (goutte de $1\mu\text{l}$ à $10\mu\text{l}$ en général pour les expériences manuelles, voir maintenant de 10nl à 100nl avec les robots de cristallisation) contenant la macromolécule, un agent précipitant (ou plusieurs agents précipitants) et un tampon,
- une solution de grand volume (de $100\mu\text{l}$ à 1 ml de volume, voir maintenant de $50\mu\text{l}$ à $100\mu\text{l}$ avec les robots de cristallisation) contenant un agent déshydratant. Ce compartiment (appelé le plus souvent "le puits") se comportera alors comme un réservoir au sens

thermodynamique du terme : on supposera donc que sa composition ne change pas au cours du processus d'échange de vapeur.

Le rôle de l'agent précipitant est de diminuer la solubilité de la macromolécule dans le petit volume (goutte). L'augmentation de la concentration de la macromolécule et/ou de l'agent précipitant pourra conduire à la sursaturation de la macromolécule qui est nécessaire à la nucléation et à la croissance cristalline. Une manière d'augmenter la concentration de la macromolécule et/ou de l'agent précipitant est de déhydrater la goutte. Le rôle de l'agent déhydratant est de réduire la pression de vapeur d'eau au dessus du réservoir. De même, l'agent cristallisant, le tampon (et la macromolécule) réduisent la pression de vapeur d'eau au dessus de la goutte. Les deux solutions ayant des pressions partielles différentes, les espèces volatiles (eau, alcool...) vont diffuser jusqu'à l'équilibre thermodynamique. Si la pression de vapeur d'eau au dessus du réservoir est initialement inférieure à celle au dessus de la goutte, il y aura transfert d'eau de la goutte vers le réservoir jusqu'à ce que les deux pressions de vapeurs deviennent identiques. En conséquence, tous les composants de la goutte vont se concentrer (macromolécule, agent précipitant, impuretés).

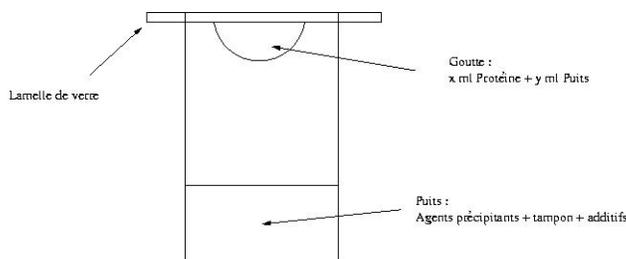


FIGURE 2 – Méthodes de cristallisation par diffusion de vapeur. Technique de la goutte suspendue.

Un montage souvent employé utilise la même espèce chimique pour agent déhydratant et agent cristallisant. Le réservoir est préparé à une concentration donnée en agent déhydratant. La goutte est préparée en ajoutant un volume x d'une solution stock de macromolécule à un volume y de solution de réservoir. Deux dispositifs expérimentaux sont utilisés pour cette technique, l'un appelé goutte suspendue (figure 2), l'autre goutte assise. En plus de l'aspect thermodynamique, la sursaturation et la formation des cristaux dépendent de la cinétique de transfert des échanges de vapeur. Comme tous processus d'évaporation, cette technique dépend de nombreux paramètres : nature et concentration initiales des constituants, température, taille et forme de la goutte, distance goutte - réservoir. Le lecteur se reportera à la bibliographie pour une étude plus détaillée de ces phénomènes.

Les cristaux de macromolécules biologiques présentent des propriétés particulières qui les distinguent fortement des autres cristaux :

- ils contiennent une forte proportion de solvant désordonné (de 20 % à 90 %). Un cristal de macromolécule biologique est un état particulier de la matière (état semi-solide, semi-liquide). Les canaux de solvant permettent d'accommoder des variations de l'ordre cristallin.
- les forces impliquées dans l'empilement cristallin sont faibles, le nombre de contacts (liaisons hydrogène, ponts salins, interactions hydrophobes) entre entités voisines sont limitées (en proportion de leur masse par rapport aux petites molécules), la cohésion du

réseau est donc plus faible. Les cristaux sont alors en général difficiles à obtenir et ils sont fragiles, de taille limitée (quelques centaines de μm) dans les cas les plus favorables.

- les macromolécules sont sensibles, une perte de la structure native étant possible. On doit donc se limiter à certains domaine de pH, de température, de force ionique. Il est aussi nécessaire de maintenir un état d'hydratation autour des cristaux.
- les macromolécules sont flexibles et présentent une dynamique interne qui provoquent des variations locales de la chaîne polypeptidique
- la qualité de diffraction, conséquence directe de tous ces désordres, est assez souvent limitée (2Å-3Å).

Pour que la structure résolue ait un intérêt au niveau biologique, il est primordial que la macromolécule soit cristallisée dans des conditions physico-chimiques (comme par exemple la température, le pH, la pression) où elle garde son activité biologique. De nombreux exemples de cristaux d'enzyme (protéines avec une activité catalytique) ont montré que l'activité étaient conservée dans l'état cristallin.

4 Enregistrement des données de diffraction

Dans le cas de la collecte des données par la méthode d'oscillation, qui est la méthode de référence actuelle pour les cristaux de macromolécules biologiques (voir paragraphe 4.3.1), l'énergie diffractée E_{hkl} pour chaque réflexion hkl est reliée au module du facteur de structure $|F_{hkl}|$ par la relation de Darwin (hypothèse d'un cristal de parfaite mosaïcité³ en rotation avec une vitesse angulaire constante) :

$$E_{hkl} = C \ I_0 \ \frac{V_c}{V^2} \ \lambda^3 \ A_{hkl} \ L_{hkl} \ P_{hkl} \ |F_{hkl}|^2 \quad (7)$$

où C est un facteur constant, A_{hkl} le facteur d'absorption, L_{hkl} le facteur de Lorentz, P_{hkl} le facteur de polarisation, V le volume de la maille cristalline directe, V_c le volume du cristal, λ la longueur d'onde du faisceau incident d'intensité I_0 .

La collecte des données de diffraction des macromolécules biologiques doit résoudre plusieurs problèmes :

- un nombre important de mesures à enregistrer et à traiter : le nombre de réflexions à mesurer est directement proportionnel au volume de la maille directe (et varie avec l'inverse du cube de la limite de diffraction des cristaux (exprimée en Å dans l'espace direct),
- une intensité relativement faible des ondes diffractées : l'intensité mesurée de chaque onde diffractée est proportionnelle au volume du cristal illuminé, à l'intensité du faisceau incident et varie comme l'inverse du carré du volume de la maille cristalline.

On a donc cherché à compenser les défauts mentionnés ci-dessus en jouant sur trois éléments essentiels pour l'enregistrement des données de diffraction : la source de rayons X, le

3. Dans un cristal réel, la périodicité cristalline n'est jamais parfaite. Des défauts ponctuels (dislocations, défauts d'empilement, grains et joints de grains, macles) affectent l'ordre cristallin à courte distance. Un cristal est constitué d'agrégats de petits cristaux ou grains reliés entre eux par des surfaces de raccordement appelées joints de grains. Les dislocations des grains et leurs frontières caractérisent la mosaïcité du cristal. Celle-ci se traduit par un élargissement des pics de diffraction.

détecteur de rayons X qui enregistre les intensités des réflexions diffractées, et la méthode d'enregistrement.

Les développements algorithmiques et méthodologiques de la génomique structurale (voir paragraphe 13) ont permis de fortement automatiser cette étape de collecte de diffraction permettant ainsi de tester rapidement de nombreux cristaux, d'enregistrer un espace complet de diffraction pour les meilleurs d'entre eux et ceci sans intervention humaine. A partir d'un bon cristal, avec un source de rayonnement synchrotron intense et un détecteur approprié, un espace complet de diffraction peut s'effectuer en moins de vingt minutes.

4.1 Source de rayons X

Les sources de rayons X utilisées en cristallographie biologique sont soit une source (dite) classique de laboratoire soit le rayonnement synchrotron. Dans une source classique, une cible de métal pur (anticathode) est soumise à un bombardement d'électrons accélérés à une énergie suffisante, le spectre d'émission associé présente des raies caractéristiques du métal qui la compose. Deux systèmes sont utilisés : le tube scellé et l'anticathode tournante. Ils impliquent un travail en faisceau monochromatique et présentent des limitations en puissance.

L'utilisation du rayonnement synchrotron pour les études structurales en biologie est l'une des révolutions qui s'est produite dans le domaine pendant les années 1980-1990. Un électron (ou un positron) ultrarelativiste émet un rayonnement électromagnétique, appelé rayonnement synchrotron, lorsqu'il est défléchi par un champ magnétique. Au départ, ce rayonnement constituait une simple curiosité et une nuisance pour les physiciens des hautes énergies qui travaillaient sur les anneaux de collision. En effet, dans les synchrotrons à électrons ou dans les anneaux de stockage utilisés pour la physique des particules élémentaires, l'émission synchrotron est un sous-produit inévitable de la circulation des particules, qui provoque une importante perte d'énergie. Initialement phénomène nuisible, le rayonnement synchrotron possède des caractéristiques si exceptionnelles qu'il est devenu un outil majeur dans une multitude de disciplines dont la cristallographie. Les principales caractéristiques de ce rayonnement sont les suivantes :

- un rayonnement blanc très intense. En raison de la focalisation naturelle dans le plan de l'orbite, la brillance spectrale du faisceau est très élevée, de plusieurs ordres de grandeur supérieure à celle d'un tube à rayons X classique.
- une émission très directionnelle et de structure temporelle pulsée. Dans le plan de l'orbite, la divergence du faisceau dépend des fentes ou de l'optique utilisée : elle est de l'ordre de quelques milliradians. Dans le plan vertical, le rayonnement est presque aussi bien focalisé que l'émission laser (de l'ordre de quelques dixièmes de milliradian). En cristallographie classique des protéines, la focalisation et la monochromatisation du faisceau incident sont réalisées par un monochromateur à cristal courbé et/ou des doubles miroirs focalisants. La direction des photons incidents est fortement corrélée avec leur énergie ce qui induit une divergence du faisceau fortement asymétrique. Ce phénomène ne se produit pas pour les sources conventionnelles sauf en présence de doubles miroirs focalisants. La structure temporelle pulsée est transparente pour de nombreuses expériences. Néanmoins, une utilisation est envisageable pour la cristallographie résolue dans le temps.
- un rayonnement polarisé. Le rayonnement synchrotron émis par une particule est polarisé linéairement dans le plan de l'orbite.

La première génération de machines a utilisé le rayonnement émis dans les aimants de courbure des anneaux de collisions conçus pour les physiciens des hautes énergies. On assiste maintenant à l'utilisation de machines qui ont été construites et optimisées pour l'utilisation du rayonnement synchrotron et qui permettent d'obtenir des sources de plusieurs ordres de grandeur plus intenses. Ces gains sont dus à une utilisation plus efficace du rayonnement par l'ajout d'éléments magnétiques d'insertion de grandes dimensions (wigglers multipoles ou onduleurs). Ces aimants permettent une extension de la distribution spectrale vers les basses longueurs d'onde et une augmentation de l'intensité disponible. Une soixantaine de synchrotrons existent actuellement dans le monde et possèdent des lignes entièrement dédiées à la biologie structurale. L'une des sources les plus puissantes (synchrotron dit de 3^e génération) est installée à l'ESRF ("European Synchrotron Radiation Facility") à Grenoble. Une nouvelle source, SOLEIL (Source Optimisée de Lumière d'Énergie Intermédiaire du LURE (Laboratoire pour l'Utilisation du Rayonnement électromagnétique)), a été ouverte fin 2006 sur le Plateau de Saclay à Saint Aubin (Essonne). En cristallographie des protéines, le rayonnement synchrotron, utilisé avec une optique de focalisation et/ou de monochromatisation performante a permis des progrès considérables. On peut citer par exemple :

- une réduction des temps d'exposition de plusieurs ordres de grandeur, qui se traduit par une augmentation du nombre de réflexions que l'on peut enregistrer par cristal avant qu'il ne soit détruit par suite de l'action du rayonnement. L'augmentation de l'intensité du faisceau est particulièrement importante pour les cristaux de petite taille et/ou avec de grandes mailles. On peut utiliser maintenant des cristaux de plus en plus petits (de l'ordre de quelques dizaines de microns) que l'on ne sait pas faire grossir.
- une meilleure résolution spatiale des taches de diffraction. Des structures macromoléculaires de plus en plus complexes, correspondant à des mailles cristallines de plus en plus grandes (plus de 1000 Å dans le cas du virus de la langue bleue) peuvent être résolues.
- un gain en résolution dû à une amélioration du rapport signal sur bruit.
- une utilisation du rayonnement blanc qui permet le choix de la longueur d'onde. Il est possible de travailler à courte longueur d'onde pour minimiser le problème de l'absorption et pour réduire la dégradation due aux rayonnements. Un développement fondamental concerne l'optimisation de l'utilisation de la diffusion anormale par la méthode MAD (voir paragraphe 7) qui permet d'obtenir des phases expérimentales de très bonne qualité. Ce rayonnement blanc intense a aussi permis le renouveau de la méthode de Laue, et le développement des méthodes de cristallographie résolue dans le temps.

4.2 Les détecteurs

Le choix d'un "détecteur" de rayons X dépend de ses caractéristiques selon plusieurs critères : coefficient d'absorption pour les rayons X, dynamique, résolution spatiale, uniformité spatiale et temporelle de la réponse, linéarité de la réponse avec le rayonnement reçu, bruit de fond. De plus, pour les cristaux de macromolécules biologiques, il faut tenir compte de leur fragilité importante lors de l'exposition aux rayons X qui limite la durée de vie d'un cristal sous irradiation, de leur faible pouvoir diffractant qui implique des temps d'exposition aux rayons X importants, et du nombre important de données à enregistrer. On verra aussi ci-dessous que les conditions d'enregistrement optimum pour la méthode d'oscillation consistent à utiliser un angle d'oscillation minimum avec une distance cristal-détecteur maximum. Ceci impose deux limitations qui sont directement liées à la technologie des détecteurs : un grand

nombre de données doivent être lues et stockées rapidement, le détecteur doit avoir une taille la plus grande possible.

Les progrès technologiques réalisés ces dernières années sur la technologie des détecteurs ont été un facteur essentiel de la révolution cristallographique. Pour des cristaux à grandes mailles cristallines ($\geq 200 \text{ \AA}$), le film photographique a été pendant de nombreuses années le détecteur le mieux adapté à ces critères avec cependant trois principaux inconvénients : la nécessité d'une étape de numérisation pour les traitements ultérieurs, une détection des photons spatiale et non temporelle, et un bruit de fond chimique intrinsèque. Depuis le début des années 90, le film a été remplacé d'abord par des détecteurs bidimensionnels puis par des détecteurs connus sous le nom de "plaque image". Ces détecteurs marient une grande résolution spatiale, un très faible bruit de fond, un haut coefficient d'absorption, une grande dynamique et une très bonne linéarité. Ils sont de plus réutilisables. Une "plaque image" comprend une couche sensible ($\simeq 150 \mu\text{m}$ d'épaisseur) constituée de centres colorés (ions $BaFBr : Eu^{2+}$). Lors de l'absorption d'un photon X, certains centres colorés passent à un état d'énergie plus élevé métastable ce qui permet de créer une image latente du spectre de diffraction sur la plaque. On stimule alors la désexcitation de ces états par un faisceau laser hélium-néon, et on mesure la fluorescence engendrée par le laser. Ce type de détecteurs a réellement constitué une révolution dans le domaine. Ils permettent d'enregistrer directement sur le disque d'un ordinateur les données du cliché de diffraction. Avec l'amélioration des performances des logiciels, on peut alors analyser et traiter complètement les images de diffraction au fur et à mesure de leur enregistrement. L'un des seuls défauts de ces plaques images est qu'elles nécessitent un temps de lecture de quelques minutes. Les premières plaques image avaient un diamètre de 180mm et nécessitaient un temps de lecture de l'ordre de trois minutes. Les plaques les plus performantes ont actuellement un diamètre de 350mm et un temps de lecture de l'ordre de la minute. Avec les sources de rayonnement synchrotron actuelles où les temps d'exposition sont de l'ordre de la seconde, le temps de lecture des plaques images devient un facteur limitant. L'utilisation de détecteurs CCD (Charge-Coupled Device) a permis de réduire le temps de lecture (de l'ordre quelques secondes). Dans un détecteur CCD, les photons incidents sont convertis en photons visibles par un scintillateur déposé à la surface d'une fibre optique couplé par d'autres composants optiques et électronique à un tube cible en silicium. Ils sont néanmoins de taille limitée (diamètres variant de 55mm à 165mm) ce qui constitue un handicap pour les cristaux à grande maille cristalline⁴. En effet, plus le détecteur est petit, plus il faut le rapprocher du cristal pour ne pas perdre les ondes diffractées à haute résolution. Plus le détecteur est proche du cristal, plus les taches de diffractions seront voisines sur la surface du détecteur. Pour les grandes mailles cristallines, on aura alors des superpositions de taches de diffraction, qui rendront le cliché inutilisable. Pour les détecteurs CCD, on peut remédier à ce problème en maintenant plusieurs CCD côte à côte. De nos jours, tous les grands laboratoires de cristallographie biologique sont équipés de plaques images, et le plus souvent de détecteurs CCD.

Les recherches se sont poursuivies afin de limiter les défauts respectifs de chaque système (temps de lecture pour les plaques images et taille pour les CCDs) et mettre au point de nouveau type de détecteurs. Actuellement, une nouvelle famille de détecteurs a conduit à une troisième révolution dans le domaine. Ces détecteurs, présents à quelques exemplaires sur quelques sites "synchrotrons" utilise une technologie dite "hybride-pixel" (détecteurs Megapixel

4. Pour une distance cristal-détecteur donnée, si le dispositif expérimental le permet, on peut bouger un détecteur sur le cercle 2θ pour collecter des réflexions à plus haute résolution.

Pilatus) où les photons X sont directement convertis en signal électrique qui est ensuite amplifié et traité par une matrice 2D de diodes au silicium à jonction p-n couplés à une matrice de canaux de lecture utilisant la technologie "CMOS". Ces détecteurs, développés initialement pour la physique des hautes énergies, marient un très grande dynamique, une grande taille ($431 \times 448 \text{mm}^2$ pour le "Pilatus 6M") et un très faible bruit de fond à un temps de lecture de l'ordre de quelques millisecondes. Des progrès majeures ont été réalisés avec ce nouveau type de détecteurs en particulier pour des problèmes complexes (cristaux de très petite taille et grande maille cristalline).

4.3 Méthodes d'enregistrement

Lorsqu'un cristal est soumis à un faisceau de rayons X, les conditions de diffraction (diffusion cohérente dans certaines directions privilégiées de l'espace) sont exprimées par les équations de Laue :

$$\vec{a} \cdot \vec{S} = h, \quad \vec{b} \cdot \vec{S} = k, \quad \vec{c} \cdot \vec{S} = l \quad (8)$$

où $\vec{S} = \vec{s} - \vec{s}_0$, \vec{s}_0 et \vec{s} correspondant respectivement au vecteur d'onde du faisceau incident et diffusé, $(\vec{a}, \vec{b}, \vec{c})$ les vecteurs de base du réseau direct du cristal et avec $(h, k, l) \in \mathbb{Z}$. Ce qui signifie que le vecteur \vec{S} est un vecteur du réseau réciproque du cristal (construit sur les $\vec{a}^*, \vec{b}^*, \vec{c}^*$) :

$$\vec{S} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^* \quad (9)$$

Ceci est souvent traduit simplement par la loi de Bragg qui montre que la famille de plans réticulaires d'indice de Miller [hkl] donne lieu à un faisceau diffracté si et seulement si :

$$2 d_{hkl} \sin\theta = \lambda \quad (10)$$

où d_{hkl} est la distance interréticulaire de la famille de plans [hkl], θ l'angle entre la direction du faisceau incident de longueur d'onde λ et la famille de plans. La loi de Bragg n'exprime donc que le lieu géométrique des ondes diffractées. Un enregistrement d'un espace complet de diffraction nécessite de mesurer l'intensité de toutes les ondes diffractées⁵, il faut donc amener toute les familles de plans en position de diffraction. D'après la loi de Bragg, il y a donc deux possibilités :

- un travail en faisceau monochromatique, où λ est constant ce qui nécessite de faire varier θ . On impose alors au cristal un mouvement de rotation.
- un travail en faisceau polychromatique (le faisceau contient toutes les longueurs d'onde) où θ est constant (le cristal est immobile).

5. Dans la pratique, un cristal a une diffraction limitée. On nomme limite de diffraction d'un cristal correspond, (dans les conditions expérimentales optimales), à la norme du plus grand vecteur de diffusion \vec{S} , noté S_{max} , au-delà duquel le rapport signal sur bruit des mesures devient insuffisant. Dans la pratique courante, cette limite est exprimée dans l'espace direct par $d_{min} = \frac{1}{|S_{max}|}$. Les méthodes de collecte de données de diffraction introduisent aussi une troncature à basse résolution. Concrètement, les réflexions hkl pour lesquelles $|\vec{S}| < |S_{min}|$ et $|\vec{S}| > |S_{max}|$ ne peuvent pas être enregistrées expérimentalement. Les réflexions enregistrées correspondent donc au domaine $[|S_{min}|; |S_{max}|]$ dans l'espace réciproque ou au domaine $[d_{min}; d_{max}]$ dans l'espace direct (avec $d_{max} = \frac{1}{|S_{min}|}$).

4.3.1 Collecte en faisceau monochromatique

Les méthodes d'enregistrement sont le reflet des progrès et des limites des techniques et des logiciels capables d'indexer et de mesurer les réflexions diffractées par un cristal. L'étude des cristaux à grande maille cristalline, qui correspondent à des réseaux réciproques fortement peuplés, ferme la porte à l'utilisation du diffractomètre automatique qui mesure l'intensité d'une seule réflexion à la fois. La nécessité d'enregistrer un maximum de données avec un minimum de cristaux condamne les méthodes d'enregistrement dites "avec écran" telles que la méthode de précession et la chambre de Weissenberg, et a abouti à l'apogée de la méthode d'oscillation. Cette méthode est la plus utilisée aujourd'hui. Son principe consiste à donner au cristal un mouvement de rotation de faible amplitude (de l'ordre de $0,2^\circ$ à 1°) autour d'un axe perpendiculaire aux rayons X incidents. On recueille les intensités des réflexions diffractées sur un détecteur placé perpendiculairement à l'axe des rayons X. Un espace complet de diffraction est alors obtenu en enregistrant une série d'oscillations permettant de balayer l'espace réciproque indépendant (tableau 2). Un cliché de diffraction est constitué d'une série caractéristique de lunes dont les limites représentent les projections, sur le détecteur, des plans du réseau réciproque lors de leur traversée de la sphère d'Ewald entre le début et à la fin de l'oscillation. La figure 3 montre un exemple de cliché d'oscillation d'un cristal de virus *Nudaurelia Capensis* ω .

L'angle d'oscillation maximum possible qui évite la superposition, sur le détecteur, de deux points du réseau réciproque, dépend des paramètres d'orientation du cristal, de la résolution obtenue et du domaine de réflexion du cristal (mosaïcité et divergence de l'optique). La portion de l'espace réciproque à enregistrer dépend du groupe de symétrie de Laue du cristal. Le tableau 2 donne pour chaque groupe de Laue et en fonction de l'orientation du cristal, le domaine angulaire à enregistrer pour obtenir un espace complet de diffraction. Le choix de l'axe de rotation et l'angle d'oscillation est un compromis qui est fait de façon à :

- enregistrer un nombre de clichés minimum, ce qui pour un domaine de rotation donné, implique un angle d'oscillation maximum par cliché.
- minimiser les superpositions de réflexions sur le cliché de diffraction, ce qui restreint l'angle d'oscillation.
- maximiser le rapport signal sur bruit sur le cliché de diffraction, ce qui correspond à une oscillation la plus petite possible.

Les conditions optimales d'enregistrement sont d'utiliser une oscillation la plus faible possible et à une distance cristal-détecteur la plus grande possible. Cela induit un temps d'enregistrement plus long, une augmentation considérable des données à stocker et à manipuler, ce qui n'est pas réellement un problème avec les moyens informatiques actuels. De plus, une redondance des mesures est nécessaire pour améliorer la qualité des données enregistrées (en particulier le rapport signal sur bruit) et l'un des problèmes actuels est la dégradation du cristal (même en cryo-cristallographie) lors de l'enregistrement, en particulier si on utilise un rayonnement synchrotron. Des choix des paramètres optimaux de l'enregistrement doivent donc être effectués, ces choix nécessitant encore une expérience de la cristallographie pour réaliser les bons compromis. Si l'on dispose de plusieurs bons cristaux, on peut optimiser ces paramètres après un enregistrement initial. Au cours des dernières années, on a automatisé de plus en plus ce processus, ce qui a supposé une mise en place de logiciels d'aide à la décision ou de décision automatique.

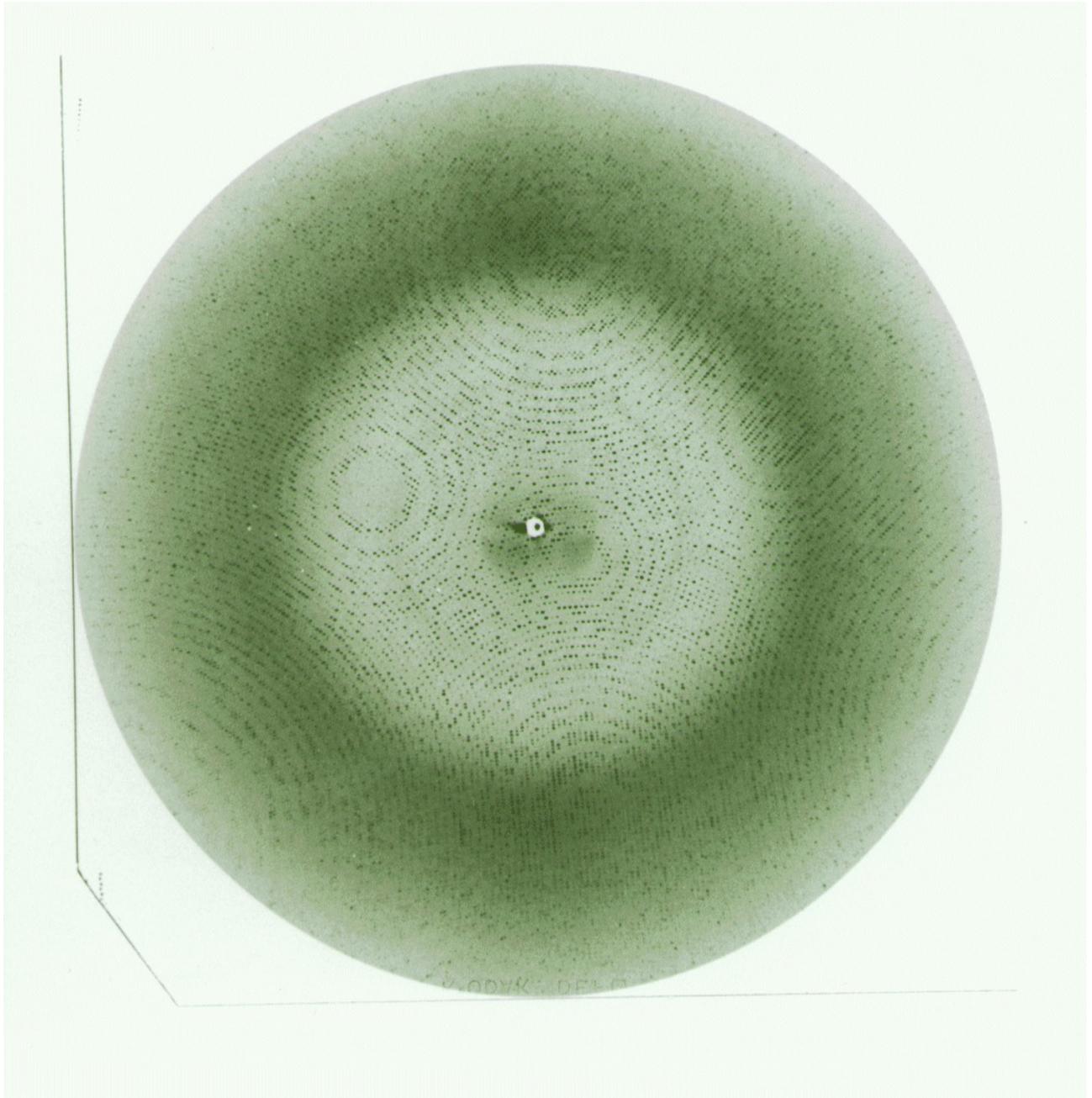


FIGURE 3 – Cliché d'oscillation. Exemple d'un cristal de virus *Nudaurelia Capensis* ω . Cliché enregistré avec la source de synchrotron de Brookhaven (U.S.A.) dans les conditions expérimentales suivantes : $\lambda = 1,22\text{\AA}$, distance cristal-film = 120mm, angle oscillation $0,3^\circ$. Paramètres cristallins : $a = 414,0\text{\AA}$, $b = 410,7\text{\AA}$, $c = 420,1\text{\AA}$, $\alpha = 59,1^\circ$, $\beta = 58,9^\circ$, $\gamma = 64^\circ$. L'angle d'oscillation est de $0,3^\circ$. Un total de 15245 réflexions ont été mesurées sur ce cliché avec une limite de diffraction de $2,7\text{\AA}$. Dans ce groupe d'espace triclinique, un domaine angulaire de 180° est nécessaire pour obtenir un espace de diffraction complet, ce qui correspond à 600 clichés de ce type.

Système	Groupe du Point	Standard	Diffusion Anomale
Triclinique	1	180°	180° + 2 θ_{max}
Monoclinique	2	180°(b) ; 90°(ac)	180°(b); 180° + 2 θ_{max} (ac)
Orthorhombique	222	90° (ab ou ac ou bc)	90° (ab ou ac ou bc)
Quadratique	4 422	90° (c ou ab) 45°(c); 90°(ab)	90°(c); 90° + θ_{max} (ab) 45°(c); 90°(ab)
Cubique	23 432	environ 60° environ 35°	environ 70° environ 45°
Trigonal/Rhomboédrique	3 32	60°(c); 90°(ab) 30°(c); 90°(ab)	60° + 2 θ_{max} (c); 90° + θ_{max} (ab) 30° + θ_{max} (c); 90°(ab)
Hexagonal	6 622	60°(c); 90°(ab) 30°(c); 90°(ab)	60°(c); 90° + θ_{max} (ab) 30°(c); 90°(ab)

Tableau 2 – Méthode d'oscillation : Domaine angulaire minimum à enregistrer pour chaque système cristallin pour obtenir un espace complet de diffraction. (a), (b), (c) représente l'axe du cristal qui est parallèle à l'axe de rotation, ac signifie un vecteur quelconque dans le plan ac. θ_{max} est l'angle de diffusion maximum. En cas de diffusion anormale (voir le paragraphe 7), il est nécessaire de mesurer les réflexions hkl et les réflexions $\bar{h}\bar{k}\bar{l}$. Tableau adapté de Zbigniew Dauter, Acta Cryst., (1999), D55, 1703-1717.

On peut schématiquement diviser une collecte de données en plusieurs étapes :

1. Inspection visuelle d'une première image de diffraction pour estimer la qualité de diffraction du cristal.
2. Indexation de une ou plusieurs images, ce qui permet de déterminer une première estimation des paramètres de la maille cristalline et de l'orientation du cristal. On pourra alors choisir des paramètres optimaux pour l'enregistrement : distance cristal-film, angle oscillation, domaine angulaire minimal à mesurer, temps d'exposition.
3. Collecte d'un espace complet de diffraction.
4. Affinement des paramètres géométriques de l'enregistrement.
5. Mesures de l'intensité des réflexions.
6. Mise à l'échelle de tous les clichés de diffraction.
7. Réduction des données. Détermination du groupe d'espace, moyennation des réflexions équivalentes. Analyse statistique et estimations des erreurs sur les mesures.

On rappelle que les éléments de symétrie présents dans un cristal et qui caractérisent le groupe d'espace du cristal introduisent des contraintes entre certains facteurs de structures. Chaque élément de symétrie d'un cristal est caractérisé par un opérateur de symétrie qui permet de relier deux positions équivalentes d'un cristal. Soit M et M' deux points cristallographiquement équivalents par l'élément de symétrie noté s . M est repéré par le vecteur \vec{r}_j dans la maille cristalline, M' est repéré par le vecteur $\vec{r}_{j's}$. On a alors : $\vec{r}_{j's} = R_s \vec{r}_j + T_s$ où R_s est une matrice (3 lignes, 3 colonnes) et T_s est une matrice (3 lignes, 1 colonne). En notant $H^t = [hkl]$, on peut montrer que :

$$F_{R_s^t H} = F_H \exp(-2i\pi H^t T_s) \quad (11)$$

où X^t est la transposée de la matrice X .

Les réflexions $H' = R_s^t H$ et H sont dites équivalentes, ce qui implique une contrainte sur les modules et sur les phases de ces deux réflexions :

$$\begin{aligned} |F_{R_s^t H}| &= |F_H| \\ \alpha_{R_s^t H} - \alpha_H &= -2\pi H^t T_s \end{aligned}$$

De plus, en absence de diffuseur anomal, on montre facilement que (en notant F^* le complexe conjugué de F) :

$$F_{\bar{h}\bar{k}\bar{l}} = F_{hkl}^* \quad (12)$$

Cette équation est connue sous le nom de loi de Friedel. L'analyse des données de diffraction permet de mettre en évidence les contraintes imposées par la relation 11 ce qui donne accès au groupe d'espace d'un cristal donné. La qualité des données est caractérisée par un facteur de désaccord entre réflexions équivalentes, nommé le plus souvent R_{sym} (ou parfois R_{merge}), exprimé le plus souvent par la relation :

$$R_{sym}(I) = \frac{\sum_{hkl} \sum_{j=1}^{j=n_{hkl}} | \langle I_{hkl} \rangle - I_{hkl,j} |}{\sum_{hkl} \sum_{j=1}^{j=n_{hkl}} I_{hkl,j}} \quad (13)$$

$$\langle I_{hkl} \rangle = \frac{1}{n_{hkl}} \sum_{j=1}^{j=n_{hkl}} I_{hkl,j} \quad (14)$$

où $\langle I_{hkl} \rangle$ représente l'intensité moyenne de la réflexion d'indice hkl calculée à partir des n_{hkl} mesures.

Classiquement R_{sym} est de l'ordre de 0,02 – 0,10. Bien que très utilisé, le coefficient R_{sym} défini par la relation 13 n'est pas un excellent indicateur de la qualité des données mesurées car l'expression utilisée dépend de la redondance des mesures effectuées. Pour supprimer de ce biais, plusieurs auteurs ont proposé d'introduire dans les programmes de traitement d'autres expressions, dont les plus utilisées sont :

1. un facteur de désaccord nommé $R_{r.i.m.}$ ("Redundancy-independent merging R-factor") donné par la relation :

$$R_{r.i.m.}(I) = \frac{\sum_{hkl} \sqrt{\frac{n_{hkl}}{n_{hkl}-1}} \sum_{j=1}^{j=n_{hkl}} |\langle I_{hkl} \rangle - I_{hkl,j}|}{\sum_{hkl} \sum_{j=1}^{j=n_{hkl}} I_{hkl,j}} \quad (15)$$

2. un facteur de désaccord nommé $R_{p.i.m.}$ ("Precision-indicating merging R-factor") donné par la relation :

$$R_{p.i.m.}(I) = \frac{\sum_{hkl} \sqrt{\frac{1}{n_{hkl}-1}} \sum_{j=1}^{j=n_{hkl}} |\langle I_{hkl} \rangle - I_{hkl,j}|}{\sum_{hkl} \sum_{j=1}^{j=n_{hkl}} I_{hkl,j}} \quad (16)$$

R_{rim} (noté aussi R_{meas}) reflète la qualité des données non-moyennées (les observations), tandis que R_{pim} reflète la qualité des données moyennées (réflexions uniques). Les travaux de Karplus, Diederichs en 2012⁶ ont montré que ces facteurs ne sont, ni de bons marqueurs de la qualité et de l'utilité des données collectées, ni de bons indicateurs de la qualité de la structure qui sera construite à partir de ces données. Inspirés de critères utilisés en cryo-microscopie électronique, ils ont montré la nécessité d'utiliser 2 nouveaux indicateurs $CC_{\frac{1}{2}}$, CC^* ("Coefficients de corrélation"), plus judicieux pour évaluer la qualité intrinsèque des données.

4.3.2 Collecte en faisceau polychromatique

La méthode de Laue est une technique très ancienne. Elle consiste à irradier le cristal étudié par un rayonnement polychromatique. Le premier cliché de diffraction de rayons X, obtenu en 1912 par Friedrich, Knipping et Von Laue était de ce type. Cette méthode était tombée en désuétude, du moins pour les études structurales, jusqu'à ces dernières années, en raison de la faible intensité des sources de rayons X conventionnelles, de leur largeur spectrale peu étendue et des difficultés à déchiffrer les clichés de diffraction obtenus. Un exemple de cliché de Laue d'une macromolécule biologique est donnée à la figure 4. Il s'agit d'un cliché de Laue de l'aspartyl-ARNt synthétase de *Thermus thermophilus*.

L'utilisation du rayonnement synchrotron en cristallographie biologique est à l'origine du renouveau de cette méthode. Elle permet :

6. Karplus PA, Diederichs K. Linking crystallographic model and data quality. 2012, Science, May 25;336(6084) :1030-33

- une utilisation optimale du rayonnement,
- une réduction des temps d'exposition,
- l'enregistrement simultané de plusieurs milliers de réflexions. Selon la symétrie du groupe d'espace du cristal, rarement plus de trois clichés de diffraction sont nécessaires en théorie pour obtenir un espace complet de diffraction.
- une obtention directe des intensités intégrées, l'intégration étant faite sur le domaine de longueur d'onde, par opposition aux domaines angulaires pour les expériences monochromatiques. Ceci est très important pour l'étude des réactions très rapides.

De nombreuses études ont montré que certaines molécules conservent leur activité dans le cristal et que la fixation d'un ligand peut provoquer un changement de conformation des molécules dans le cristal. Les seuls changements structuraux autorisés sont bien évidemment ceux qui sont compatibles avec la cohésion du réseau cristallin. Les autres changements provoquent la rupture de la symétrie ou de la périodicité du cristal et ne peuvent pas donner lieu à une étude par diffraction des rayons X. L'objectif premier de la résolution de la structure tridimensionnelle d'une macromolécule est de connaître la structure pour comprendre la fonction. Dans le cas d'un enzyme par exemple, la connaissance complète et fine du mécanisme moléculaire nécessiterait de connaître la structure de la protéine seule et de tous les intermédiaires de la réaction. Ceci est un obstacle majeur pour la cristallographie conventionnelle dans le cas d'une réaction à cinétique rapide car elle ne permet pas de visualiser les intermédiaires de réaction à courte durée de vie qui sont importants pour la compréhension du mécanisme. Il est possible d'étudier une réaction dans un cristal sous réserve de disposer d'une méthode d'enregistrement des données de diffraction qui soit synchronisée avec la durée du phénomène à observer. Les méthodes de diffraction des rayons X monochromatiques permettent difficilement d'étudier des phénomènes dont la durée de vie est inférieure à 1 heure. C'est le défi qui a été relevé par la méthode de Laue.

Malgré plusieurs exemples de belles réussites correspondant à de réelles prouesses technologiques et méthodologiques (voir bibliographie), de nombreux problèmes intrinsèques incomplètement résolus limitent fortement son utilisation. Le problème majeur est celui de la mosaïcité. Le fait que le cristal reste immobile durant toute l'expérience rend la méthode de Laue très sensible à tout défaut qui pourrait être présent dans le cristal (macles, forte mosaïcité). Un cristal qui donne des clichés de diffraction de très bonne qualité avec un rayonnement monochromatique peut très bien avoir un comportement médiocre dans le faisceau polychromatique. De plus ce comportement peut varier d'un cristal à l'autre pour le même type de cristaux. Cela a été observé dans plusieurs exemples et c'est finalement aujourd'hui le problème majeur de cette technique.

4.4 Cryocristallographie

Un autre développement majeur pour l'enregistrement des données de diffraction a consisté à collecter les données de diffraction en refroidissant le cristal à très basse température (classiquement au environ de 100K). On parle alors de cryocristallographie. L'intérêt majeur de l'enregistrement de données à très basse température est l'augmentation considérable de la durée de vie du cristal, car à ces températures, la progression des radicaux libres provoqués par les photons X dans le cristal est fortement ralentie, ce qui limite de façon importante les dommages dus à l'irradiation en fonction du temps. Cela provoque de plus une réduction des désordres d'ordre dynamique présents dans le cristal. Dans ces conditions, un espace de

diffraction complet peut être enregistré, dans la plupart des cas, à partir d'un seul cristal. De plus, la qualité initiale de diffraction est conservée dans de nombreux cas pendant tout l'enregistrement. On obtient alors des données plus complètes et de meilleure qualité.

La collecte est réalisée en plaçant le cristal dans un flux gazeux froid issu d'un cryostat. Le gaz le plus couramment utilisé est l'azote. Deux méthodes sont utilisées le plus souvent :

- la première consiste à transférer directement le cristal (monté sur une boucle appropriée) dans le flux froid.
- la seconde utilise un liquide cryogénique intermédiaire (propane, éthane). Dans ce cas, le cristal (monté sur une boucle) est d'abord transféré dans un petit récipient rempli du liquide cryogénique ; ce récipient est baigné dans un bain d'azote liquide, ce qui provoquera assez rapidement la solidification du contenu. Le petit récipient contenant le cristal et le liquide cryogénique solide est alors placé dans un cryostat contenant de l'azote liquide. Les cristaux congelés peuvent être conservés et stockés presque indéfiniment à très basse température dans l'azote liquide, dans l'attente d'une utilisation ultérieure. On peut même assez souvent les réutiliser après une première exposition aux rayons X, ce qui permet par exemple de sélectionner préalablement les meilleurs cristaux par des tests de diffraction au laboratoire avant d'effectuer les mesures sur un site synchrotron.

Le refroidissement du cristal sans l'endommager reste le problème majeur de cette méthode. Les cristaux de macromolécules biologiques contenant une grande proportion de solvant, il faut éviter la formation de glace cristalline qui détruirait le cristal. Pour cela, on refroidit le cristal en présence d'une solution cryoprotectrice appropriée dont le choix dépend de la macromolécule et des conditions de cristallisation. Un bon cryoprotectant doit d'une part, favoriser la formation de glace amorphe lorsqu'il est refroidi instantanément et, d'autre part, préserver le pouvoir diffractant du cristal. Les cryoprotectants les plus utilisés sont le glycérol, l'éthylène glycol, le 2-méthyl-pentanediol (MPD). On peut soit les ajouter directement dans le milieu de cristallisation, soit effectuer des transferts successifs d'un cristal dans une solution appropriée de plus en plus concentrée en cryoprotectant. La recherche du mélange cryoprotectant idéal et des conditions physico-chimiques adéquates constituent l'étape la plus fastidieuse de la méthode car on ne peut agir que par essais et erreurs.

La cryocristallographie est indispensable lors de mesures effectuées avec des sources synchrotrons dont le faisceau très brillant détruit rapidement les cristaux à température ambiante. Pratiquement tous les enregistrements s'effectuent maintenant dans ces conditions de très basses températures (autour de 100K). On peut aussi utiliser cette technique pour piéger des états de transitions de certaines réactions pas trop rapides et ainsi réaliser ces enregistrements en faisceau monochromatique en évitant les problèmes inhérents à la méthode de Laue. Avec les sources de rayonnement synchrotron de troisième génération, la dégradation (du cristal et de la structure) devient le facteur limitant lors de la collecte des données de diffraction. Les développements actuels concernent l'utilisation d'hélium à la place de l'azote, ce qui permet de travailler à plus basse température (autour de 15K).

D'autre part, les techniques expérimentales utilisées pour la cryocristallographie sont facilement exécutables par des automates, ce qui a permis de réduire fortement l'intervention humaine lors de la collecte des données de diffraction. Le processus de collecte s'est donc fortement automatisé. Il est actuellement possible d'envoyer des échantillons congelés sur un site de rayonnement synchrotron et de piloter à distance la collecte des données de diffraction. L'automatisation permet aussi de tester rapidement de nombreux cristaux, d'évaluer la qua-

lité de diffraction et de ne collecter les données finales que pour les meilleurs cristaux. Pour plus de détails pratiques, le lecteur pourra se référer aux ouvrages cités en bibliographie.

5 Introduction à la détermination des phases

La diffraction élastique des rayons X par un cristal donne accès à une fonction appelée densité électronique qui reflète une distribution continue d'électrons dans la maille cristalline. En théorie, cette fonction est calculée pour chaque point de coordonnées relatives (x, y, z) d'une maille cristalline par la relation

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F_{hkl} e^{-2i\pi(hx+ky+lz)} \quad (17)$$

où F_{hkl} est le facteur de structure pour la réflexion hkl et V le volume de la maille cristalline. Le facteur de structure, nombre complexe, s'écrit :

$$F_{hkl} = |F_{hkl}| e^{i\phi_{hkl}} \quad (18)$$

où $|F_{hkl}|$ est le module et ϕ_{hkl} la phase du facteur de structure pour la réflexion hkl .

La détermination de la phase ϕ_{hkl} de chaque réflexion constitue l'une des difficultés majeures de la cristallographie, car les ondes diffractées par un cristal ne peuvent être refocalisées et les phases sont perdues. Seules les amplitudes des ondes diffusées sont accessibles à l'expérience, un détecteur ne mesurant pour une réflexion donnée que l'intensité I_{hkl} de l'onde diffractée. Celle-ci est proportionnelle au carré du module du facteur de structure (hypothèse d'un cristal de parfaite mosaïcité) :

$$I_{hkl} \propto (|F_{hkl}|)^2 = F_{hkl} F_{hkl}^* \quad (19)$$

où F_{hkl}^* est le complexe conjugué de F_{hkl} .

Il est essentiel de remarquer ici que dans le calcul de la fonction densité électronique (comme dans toute série de Fourier), l'information la plus importante pour l'exactitude de la fonction calculée est l'information de phase. En utilisant le théorème de Parseval, on peut montrer que si, pour chaque réflexion hkl , F_{hkl}^v est le facteur de structure réel et si F_{hkl}^u est le facteur de structure utilisé dans le calcul, l'erreur quadratique moyenne commise sur la densité est alors donnée par la relation

$$\langle (\Delta\rho)^2 \rangle = \left(\frac{1}{V^2}\right) \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |F_{hkl}^v - F_{hkl}^u|^2 \quad (20)$$

La figure 5 montre que les erreurs sur les phases est l'effet dominant de $|F_{hkl}^v - F_{hkl}^u|$. En conséquence, si l'erreur commise sur les phases est trop importante, la fonction densité électronique ne sera pas interprétable. Il est donc essentiel de disposer de phases initiales qui soient le plus proche possibles des phases exactes.

Pour déterminer la phase de chaque réflexion (on dit résoudre le problème des phases), les méthodes directes utilisées en cristallographie de petites molécules (comportant seulement quelques centaines d'atomes) ne peuvent en général pas s'appliquer. En effet, ces méthodes utilisent des relations de contraintes entre facteurs de structure (équation de Sayre, formule

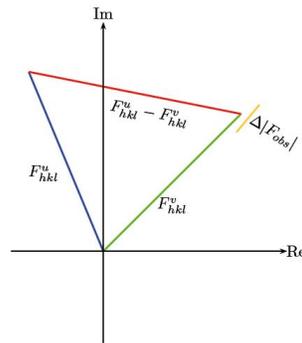


FIGURE 5 – Effets des erreurs sur le calcul de la densité électronique. Pour une réflexion donnée, l'erreur dans la fonction de densité électronique est proportionnelle à $|F_{hkl}^v - F_{hkl}^u|^2$. Expérimentalement, l'erreur sur le module des facteurs de structure est faible (de l'ordre de quelques %, notée ici $\Delta|F_{obs}|$) alors que l'erreur sur les phases peut être importante (plusieurs dizaines de degrés). Sur cette figure, $|F_{hkl}^v| = |F_{hkl}^u|$

de la tangente) dont l'application nécessite des données de diffraction à très haute résolution (de l'ordre de 1\AA , $1\text{\AA} = 10^{-10}\text{m}$). Plusieurs exemples d'utilisation de ces méthodes dans des cas pour la détermination de structure de macromolécules "ab-initio" ont été publiés dans des situations favorables (voir ouvrages cités en bibliographie).

Trois familles de méthodes sont utilisées pour déterminer les phases :

- Les méthodes directes qui utilisent des contraintes entre facteurs de structure, ce qui nécessite une résolution atomique. C'est les méthodes par excellence pour la cristallographie des petites molécules organiques. En cristallographie biologique, ces méthodes sont surtout utilisées pour localiser la structure d'un petit nombre d'atomes.
- Les méthodes qui utilisent des différences d'intensités entre facteurs de structure pour donner accès à l'information de phases. Deux méthodes sont utilisées couramment à ce jour : le remplacement isomorphe ("Multiple Isomorphous Replacement" (MIR) et la diffusion anormale ("Single wavelength Anomalous Scattering" (SAD) et Multi-wavelength Anomalous Scattering" (MAD)). Des développements récents concernent l'utilisation avec succès des variations d'intensités provoquées par la dégradation (modifications locales de la structure) due au rayonnement lors des expériences de diffraction comme source d'informations de phases⁷.
- les méthodes qui nécessitent la connaissance de la structure d'une molécule parente : le remplacement moléculaire.

Les phases connues, il est alors possible de calculer la fonction densité électronique. L'obtention d'un jeu de phases initiales qui seront le plus souvent améliorées par des techniques de modification de densité dont l'objectif sera de diminuer pour chaque réflexion de l'erreur par rapport à la phase exacte. Bien que traité historiquement de manière séparée, la détermination des phases par la méthode du remplacement isomorphe multiple ou avec l'utilisation de la diffusion anormale repose sur un principe commun. Dans ces méthodes, la source de phasage vient du fait que les variations des facteurs de structures et donc de leurs modules sont reliés à des phénomènes d'origine physique ou chimique dues à un sous-groupes d'atomes.

7. "Radiation-Induced Phasing (RIP)" voir la bibliographie pour un développement de cet aspect.

Pour chaque réflexion "indépendante" notée H, on dispose de N_H mesures du module du facteur de structure $|F_{H,j}|$, $j = 1 \rightarrow N_H$. Pour chaque réflexion, le facteur de structure peut se décomposer en 2 termes :

$$F_{H,j} = P_H + V_{H,j} \quad (21)$$

avec

- P_H une partie constante, inconnue, la même pour toutes les sources de données j
- $V_{H,j}$ une partie variable, due à un sous-groupe d'atomes donnés

$V_{H,j}$ reflète un modèle atomique (positions, paramètres de déplacements atomiques), qui peut être représenté par une variable vectorielle θ . Les relations entre les mesures expérimentales, les modules des facteurs de structure sont :

$$|F_{H,j}(\theta)| = |P_H + V_{H,j}(\theta)|, \quad j = 1 \rightarrow N_H \quad (22)$$

En théorie, chaque équation peut être résolue très simplement dans le plan complexe. Graphiquement on utilise la construction de Harker où la solution correspond à l'intersection des N_H cercles centrés en $-V_{H,j}$ et de rayon $|F_{H,j}|$ (Figure 6). Si les $V_{H,j}$ ne sont pas colinéaires, les systèmes d'équations sont mathématiquement résolus si $N_H > 2$ pour réflexions non-centriques ($N_H > 1$ pour réflexions centriques). Dans le cas réel, plusieurs catégories d'erreurs vont polluer la simplicité des équations et nécessiter un traitement statistique complexe.

Dans l'équation 22, la signification de j dépend de la méthode utilisée :

- "Single Isomorphous Replacement (SIR), Multiple Isomorphous Replacement (MIR)" : j fait référence aux différents dérivés lourds
- "Single-wavelength anomalous diffraction (SAD), Multi-wavelength anomalous diffraction (MAD)" : j fait référence aux différences de "Bijvoet"
- "(site specific) Radiation damage-Induced phasing (RIP)" : j fait référence aux degrés d'irradiations

Dans les trois prochains paragraphes (6, 7, 8), nous présenterons les fondements des trois méthodes suivantes : le remplacement isomorphe multiple, la diffusion anormale, le remplacement moléculaire. Grâce aux développements méthodologiques et technologiques réalisés au cours des dix dernières années, nous verrons que l'utilisation efficace de la diffusion anormale (suivie de méthodes de modification de densité) a permis des progrès spectaculaires dans ce domaine. Si la molécule cristallisée contient des diffuseurs anormaux dont le signal anormal est significatif dans la gamme de longueur d'onde utilement accessible, et si les cristaux sont de qualité suffisante, la détermination des phases ne pose pas réellement de problèmes, ce travail devenant de plus en plus un exercice de "routine". On verra en effet que la justesse des phases "MAD" ne dépendra que de la qualité des données de diffraction. L'utilisation de la diffusion anormale s'est maintenant imposée comme la méthode de phasage de référence en cristallographie des macromolécules biologiques. Dans les cas favorables, une première carte de densité électronique de qualité peut être obtenue quelques dizaines de minutes après de la collecte des données de diffraction, et un premier modèle quelques heures plus tard. Néanmoins, si la molécule inconnue à une structure très proche de celle d'une molécule parente connue, le remplacement moléculaire sera une méthode puissante qui donnera accès à l'information de phases sans aucune autres données supplémentaires. Pour une vision plus complète de la résolution du problème des phases en bio-cristallographie et des références complètes aux travaux originaux, le lecteur pourra se référer aux ouvrages et liens cités en bibliographie.

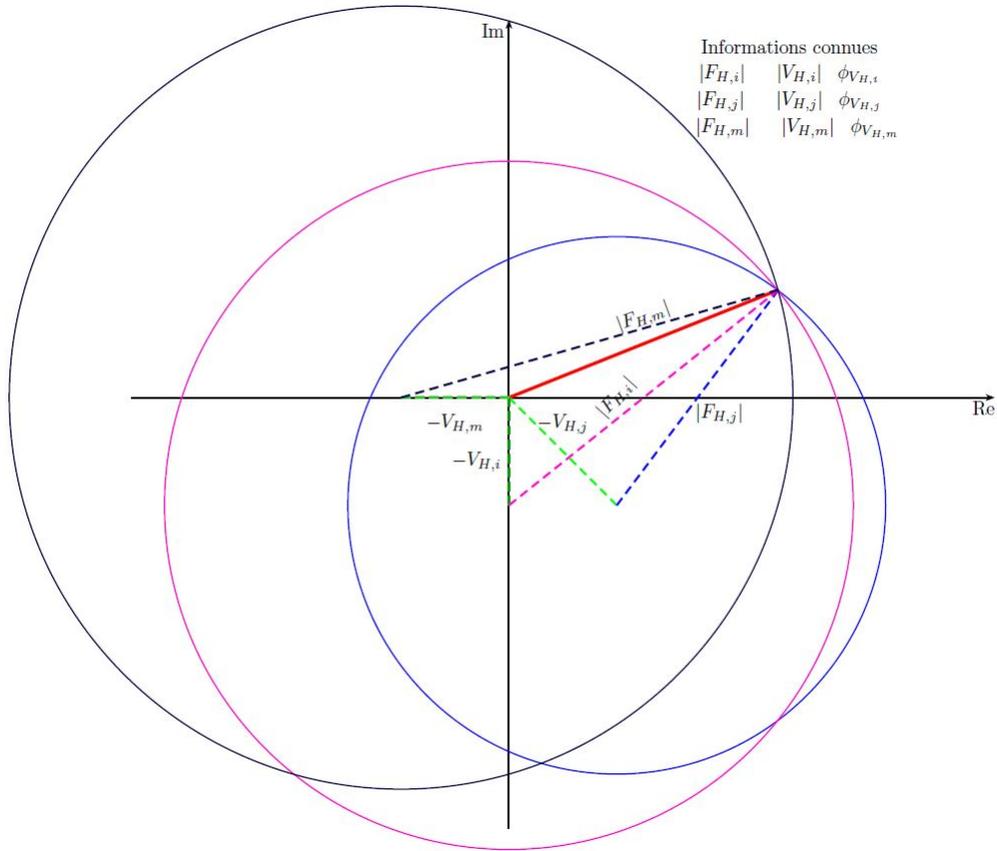


FIGURE 6 – Phasage expérimental, construction de Harker. Détermination du facteur de structure P_H (module et phase) à partir de 3 mesures (cas idéal sans erreurs). Le problème à résoudre est de trouver P_H qui vérifie les 3 relations : $(F_{H,j} = P_H + V_{H,j})$, $(F_{H,i} = P_H + V_{H,i})$ et $(F_{H,m} = P_H + V_{H,m})$. La construction de Harker montre que la connaissance de $|F_{H,i}|$, $|F_{H,j}|$, $|F_{H,m}|$, et des contributions des 3 sous-structures ($|V_{H,i}|$, $\phi_{V_{H,i}}$, $|V_{H,j}|$, $\phi_{V_{H,j}}$, $|V_{H,m}|$, $\phi_{V_{H,m}}$) permet de trouver P_H (en rouge sur la figure).

6 Méthode de la série isomorphe

Introduite en 1954 grâce au travail de pionnier de Max Perutz, cette méthode a permis l'essor de la cristallographie biologique. Elle est néanmoins de plus en plus remplacée par la méthode "SAD/MAD" que nous verrons au paragraphe 7. Pour la résolution d'une structure, on essaie de se placer le plus possible dans une optique SAD/MAD, que les diffuseurs anomaux soient déjà présents dans la structure ou qu'ils aient été introduits par une logique de type "atomes lourds".

6.1 Détermination des phases : Cas Idéal

Notons P la structure tridimensionnelle d'une macromolécule et supposons que l'on ajoute à la structure P un nombre d'atomes supplémentaires correspondant à une structure notée D. On appellera P la structure native et $(P + D)$ la structure dérivée. On supposera aussi que le cristal natif et le cristal dérivé sont isomorphes, ce qui signifie que :

- la symétrie cristallographique et les paramètres cristallins sont conservés.
- la molécule n'a subi ni rotation ni translation dans la maille.

Les changements dans la densité électronique de la molécule native consistent alors uniquement en une addition de densité électronique pour les positions de l'espace occupées par les atomes supplémentaires introduits. Pour chaque réflexion hkl , on peut alors écrire

$$F_{P+D} = F_P + F_D \quad (23)$$

où F_P est le facteur de structure du cristal natif (module $|F_P|$ et phase ϕ_P), F_{P+D} est le facteur de structure du cristal dérivé (module $|F_{P+D}|$ et phase ϕ_{P+D}) et F_D est le facteur de structure des atomes D seuls (module $|F_D|$ et phase ϕ_D).

A partir de la relation 23, on obtient :

$$|F_{P+D}|^2 = |F_P|^2 + |F_D|^2 + 2|F_P||F_D|\cos(\phi_P - \phi_D) \quad (24)$$

ce qui permet de trouver 2 valeurs possibles pour la phase ϕ_P de la réflexion hkl :

$$\begin{aligned} \phi_P^{sol1} &= \phi_D + \cos^{-1}\left(\frac{|F_{P+D}|^2 - |F_P|^2 - |F_D|^2}{2|F_P||F_D|}\right) \\ \phi_P^{sol2} &= \phi_D - \cos^{-1}\left(\frac{|F_{P+D}|^2 - |F_P|^2 - |F_D|^2}{2|F_P||F_D|}\right) \end{aligned}$$

On voit donc que si l'on connaît la nature et la position des atomes supplémentaires D dans la maille (on peut alors calculer $|F_D|, \phi_D$), et si l'on dispose des intensités diffractées ($|F_{P+D}|^2, |F_P|^2$), l'indétermination sur les phases est ramenée à deux valeurs possibles pour chaque réflexion. On remarquera que ces deux valeurs sont symétriques par rapport à la phase des atomes D seuls ϕ_D (pour une réflexion donnée hkl). Ceci est illustré graphiquement par la construction de la figure 7 (triangle OUW). Dans le plan complexe, la relation $F_P = F_{P+D} - F_D$ est vérifiée pour les points W et M qui sont à l'intersection du cercle natif (centré en O et de rayon $|F_P|$) et du cercle D (centré en U et de rayon $|F_{P+D}|$).

L'utilisation d'un deuxième dérivé $(P+B)$ (atomes supplémentaires introduits B) permettra donc de lever complètement l'indétermination et d'obtenir la phase de chaque réflexion.

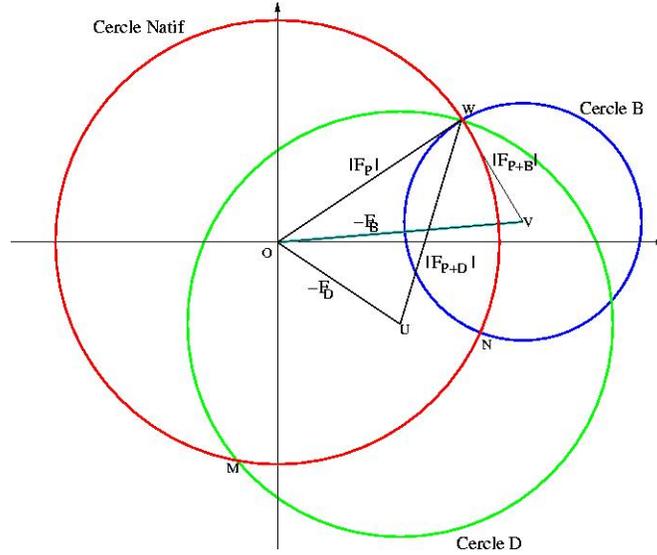


FIGURE 7 – Méthode MIR. Détermination de la phase du facteur de structure F_P pour une réflexion d'indice hkl donné (cas idéal sans défauts d'isomorphisme et sans erreurs). Les points W et M vérifient la relation $F_P = F_{P+D} - F_D$, les points W et N vérifient la relation $F_P = F_{P+B} - F_B$. Le point W détermine la vraie phase de cette réflexion.

En effet, ce deuxième dérivé permettra de trouver deux valeurs possibles pour la phase ϕ_P de la réflexion hkl :

$$\begin{aligned}\phi_P^{sol3} &= \phi_B + \cos^{-1}\left(\frac{|F_{P+B}|^2 - |F_P|^2 - |F_B|^2}{2|F_P||F_B|}\right) \\ \phi_P^{sol4} &= \phi_B - \cos^{-1}\left(\frac{|F_{P+B}|^2 - |F_P|^2 - |F_B|^2}{2|F_P||F_B|}\right)\end{aligned}$$

Parmi les quatre possibilités de phase (pour une réflexion d'indices hkl) notées ϕ_P^{sol1} , ϕ_P^{sol2} , ϕ_P^{sol3} , ϕ_P^{sol4} , deux d'entre elles sont identiques et correspondent à la phase vraie de la réflexion d'indice hkl . Ceci est illustré à la figure 7. Dans le cas de B seul, la relation $F_P = F_{P+B} - F_B$ est vérifiée pour les points W et N qui sont à l'intersection du cercle natif et du cercle B (centré en V et de rayon $|F_{P+B}|$). Le point W qui est à l'intersection des trois cercles vérifie les deux relations ($F_P = F_{P+B} - F_B$) et ($F_P = F_{P+D} - F_D$) et permet de calculer la phase de la réflexion.

La résolution de la structure native P est donc ramenée à la détermination de la position des atomes des structures D et B . Celles-ci étant composées en général de quelques atomes, on sait les résoudre par des techniques appropriées (fonction de Patterson, méthodes directes) qui n'utilisent que des modules de facteurs de structures. On verra au paragraphe 7 que si le dérivé possède une contribution anormale significative à la longueur d'onde de travail, un seul dérivé sera alors nécessaire pour obtenir en théorie l'information de phase pour chaque réflexion.

Le pouvoir de phasage d'un dérivé est lié au rapport $\frac{\langle \Delta F_{iso} \rangle}{\langle |F_P| \rangle}$ où $\Delta F_{iso} = ||F_{P+D}| - |F_P||$ et où $\langle x \rangle$ représente la variation quadratique moyenne de la variable x . Ce rapport correspond à la variation statistique moyenne du module du facteur de structure provoquée

par l'addition de N_D atomes de facteur de diffusion f_D à N_P atomes de facteur de diffusion f_P et est donnée par la relation suivante :

$$\frac{\langle \Delta F_{iso} \rangle}{\langle |F_P| \rangle} \simeq \frac{\gamma}{2} \sqrt{\frac{N_D}{N_P}} \frac{f_D}{f_P} \quad (25)$$

avec $\gamma = 2$ pour les réflexions centriques et $\gamma = \sqrt{2}$ pour les réflexions non centriques.

Dans la méthode de la série isomorphe, on introduit des atomes supplémentaires qui ont un nombre d'électrons élevés (on parle alors d'atomes lourds). La relation 25 montre que l'on peut obtenir des différences $|\Delta F_{iso}|$ très significatives en jouant sur la taille et le nombre d'atomes introduits quel que soit la taille de la macromolécule étudiée. L'exemple extrême à ce jour concerne l'étude cristallographique de sous-unités de ribosomes de différentes espèces. Un ribosome est un gros complexe nucléoprotéique (plusieurs millions de Dalton) qui est le siège de la synthèse protéique. Le phasage par la méthode "MIR" a nécessité la mise au point de "cluster" d'atomes lourds à base de tungstène $((AsW_9O_{33})_2(PhSn)_4)$ ou de tallium $Ta_6Br_{12}^{2+}$. Ces travaux ont permis d'aboutir dans le cas de la sous-unité 50S du ribosome d'*Haloarcula marismortui* à une structure à 2,4Å de résolution (publiée en septembre 2000). Les structures de nombreux virus ayant un poids moléculaire encore plus élevé ont été déterminés à haute résolution par cristallographie. Néanmoins, ces virus sont des molécules symétriques (symétrie hélicoïdale ou symétrie icosahédrique) et leurs cristaux appartiennent soit à un groupe d'espace à symétrie élevée et/ou possèdent un degré de symétrie non cristallographique élevé, ce qui facilite grandement le processus de phasage (voir le paragraphe 9).

6.2 Détermination des phases : Cas réel

En fait, le cas idéal n'existe pas en raison de la présence de nombreuses erreurs. Ces erreurs ont trois origines principales :

- erreurs sur la mesure des intensités,
- erreurs sur le modèle des atomes lourds additionnels (positions, occupations, facteurs d'agitation thermique),
- les modifications structurales (défauts d'isomorphisme)

Il est important d'essayer de séparer dans les différences d'intensité entre une native et un dérivé, la contribution due à la présence des atomes lourds et la contribution due au non isomorphisme. Une des manifestations les plus simples d'un non isomorphisme concerne la variation des paramètres de maille qu'il convient donc de mesurer de la manière la plus précise. Toutefois, des défauts d'isomorphisme peuvent ne pas se manifester par des différences sur les paramètres de maille.

Dans ces conditions, les trois cercles (natif, D et B) de la figure 7 se coupent pas en un point ou ne se coupent pas du tout. Dans ce cas, pour chaque dérivé j, pour chaque réflexion et pour toutes les valeurs possibles de la phase, on définit une densité de probabilité $P_j(\phi_P)$. Cette densité de probabilité est calculée à partir des informations suivantes :

- $|F_P|, \sigma(F_P)$ où $\sigma(F_P)$ est l'incertitude sur la mesure de $|F_P|$,
- $|F_{P+D_j}|, \sigma(F_{P+D_j})$ où $\sigma(F_{P+D_j})$ est l'incertitude sur la mesure de $|F_{P+D_j}|$,
- un modèle des atomes lourds $q_i^j, x_i^j, y_i^j, z_i^j, B_i^j$, ($i = 1, m^j$) où m^j est le nombre d'atomes lourds pour le dérivé j. Cela permet de calculer F_{D_j} et ϕ_{D_j} .

— un modèle des erreurs et du non-isomorphisme.

La connaissance de cette distribution de probabilité de phase permet de trouver F_P optimum (module et phase) qui soit en accord avec toutes les amplitudes mesurées et optimise les paramètres des atomes lourds et les phases.

L'utilisation de plusieurs dérivés lourds permet de calculer une distribution totale de probabilité totale (pour chaque réflexion) :

$$P(\phi_P) = \prod_j P_j(\phi_P) \quad (26)$$

Cette distribution n'est généralement pas unimodale. Le calcul de l'expression rigoureuse de cette distribution de probabilité et de l'affinement des paramètres impliqués dépasse le cadre de cet article. Plusieurs approximations ont été utilisées qui ne sont toujours justifiées et qui dans les cas difficiles peuvent ne pas permettre d'obtenir des cartes interprétables. Les fondements mathématiques corrects ce travail ont été initiés il y a de nombreuses années par Gérard Bricogne et ont conduit à l'utilisation de logiciels appropriés par la communauté internationale⁸. Le lecteur est invité à consulter la bibliographie pour une présentation détaillée de ces méthodes.

Il a été montré que l'utilisation du centroïde de la distribution, appelé facteur de structure le meilleur noté F_{hkl}^{best} , dans calcul des cartes de densités électronique permet d'obtenir une carte où l'erreur quadratique moyenne est la plus faible.

$$F_{hkl}^{best} = \frac{\int_0^{2\pi} P(\phi) F_{hkl}(\phi) d\phi}{\int_0^{2\pi} P(\phi) d\phi} \quad (27)$$

Pour chaque réflexion, La qualité de l'information sur la phase est mesurée par un nombre m_{hkl} , appelé figure de mérite qui est le reflet de l'étroitesse de la distribution de probabilité, m_{hkl} tendant vers la valeur 1 pour une distribution très étroite et vers la valeur 0 pour une distribution très aplatie. La figure de mérite est calculée par l'expression :

$$m_{hkl} = \frac{|F_{hkl}^{best}|}{|F_{hkl}|} \quad (28)$$

Malheureusement, elle n'est pas un bon indicateur de la justesse des phases mais simplement de la précision du calcul. En première approximation, pour une réflexion donnée, la figure de mérite représente la valeur moyenne du cosinus de l'erreur sur la phase calculée.

La carte de densité électronique qui sera calculée et que l'on va chercher à interpréter pour construire la structure de la molécule sera alors⁹ :

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{hkl \in D^*} m_{hkl} |F_{hkl}| e^{i\phi_{hkl}^{best}} e^{-2i\pi(hx+ky+Lz)} \quad (29)$$

8. Traitée ici dans le cas du phasage "MIR", cette distribution de probabilité sera généralisée pour inclure l'ensemble des informations de phases disponibles incluant ainsi d'autres sources potentielles (par exemple des phases issues des méthodes "MAD/SAD", "RIP", remplacement moléculaire).

9. voir le paragraphe 9 pour la définition de D^*

6.3 Obtention des dérivés lourds

C'est l'étape limitante de ce processus car c'est parfois un processus lent et fastidieux, utilisant le plus souvent une recherche par essais et erreurs qui peut s'avérer infructueuse dans certains cas. Deux méthodes principales sont utilisées pour introduire l'atome lourd et permettre sa fixation sur la macromolécule soit de façon covalente, soit par des liaisons électrostatiques en un ou plusieurs sites. Pour les petites molécules, un atome "léger" est substitué par un autre élément plus lourd. Pour les macromolécules, cette substitution chimique n'est généralement pas réalisable, sauf dans le cas des métalloprotéines, où on peut essayer d'enlever le métal de façon non dénaturante et de le remplacer par un métal plus lourd. Pour les acides nucléiques, on peut substituer une base par une autre base modifiée : par exemple, remplacer une thymine par un 5-iodouridine. Pour les cristaux de macromolécules biologiques, on tire profit de leur forte proportion de solvant (de 20% à 90%) et de la présence de canaux de solvant pour permettre une diffusion de sel d'atome lourd (méthode de trempage). L'atome lourd prend alors la place de molécules désordonnées de solvant, qui ne contribuent que faiblement au pouvoir de diffraction de la macromolécule sauf à basse résolution. Le trempage se fait en diffusant lentement une solution du composé souhaité dans la goutte de cristallisation. En pratique, après formation du cristal, le milieu de cristallisation (on dit la liqueur mère) est échangée petit à petit par la solution contenant l'élément lourd. Le composé que l'on veut diffuser dans le cristal doit être soluble dans la liqueur mère, ne pas donner de réactions secondaires nuisibles avec les ions présents dans le milieu et il doit avoir une certaine affinité pour la macromolécule étudiée. A titre d'exemple, les sels de dérivés mercuriels interagissent en outre avec des acides aminés tels que cystéine et histidine. A l'heure actuelle, il n'existent aucune méthode permettant d'obtenir avec succès un dérivé lourd isomorphe. De nombreux essais faisant varier la nature et la concentration de l'atome lourd (1mM-50mM), les conditions de trempage (de quelques minutes à plusieurs jours) peuvent être nécessaires pour obtenir un cristal dérivé isomorphe acceptable. Dans certains cas, il n'a pas été possible de trouver de tels dérivés lourds. Les défauts d'isomorphisme, toujours présent sauf exceptions, les dégâts provoqués par la diffusion des atomes lourds, réduisent souvent la qualité de diffraction des cristaux et ne permettent d'obtenir des phases qu'à des résolutions limités (3,5 Å- 4,5 Å).

De très bons dérivés lourds de protéines peuvent aussi être obtenus par l'utilisation de gaz rares (Xénon ou Krypton) sous pression (8-20 bars) qui se fixent par des interactions faibles de van der Waals dans des cavités hydrophobes accessibles. Cette méthode a permis dans plusieurs cas de résoudre la structure de protéines réfractaires aux sels d'atomes lourds classiques. Une base de données d'atomes lourds les plus couramment utilisés en cristallographie biologique et répertoriant aussi les conditions expérimentales utilisées dans les structures publiées a été mise en place (voir le site internet "<http://www.bmm.icnet.uk/had/>"). Malheureusement les informations détaillées concernant les conditions expérimentales du processus de phasage sont souvent absentes des publications, ce qui ne facilite pas la construction d'une base de données de connaissance.

6.4 Détermination de la position des atomes lourds.

La connaissance de la position des atomes lourds dans la maille cristalline est nécessaire pour le calcul de la phase de chaque réflexion. Comme les atomes lourds sont en nombre limité dans la maille cristalline, leurs positions seront calculées par des méthodes classiques (fonction de Patterson, méthodes directes) n'utilisant que les modules de facteurs de structures.

Pour utiliser ces méthodes, il est nécessaire de déterminer, pour chaque réflexion, $|F_D^2|$ qui représente le module du facteur de structure dû à la contribution des atomes lourds seuls. Une fonction de Patterson¹⁰ calculée avec ces modules $|F_D^2|$ permettra alors de trouver les positions de atomes lourds dans la maille pour le dérivé D . On ne peut estimer exactement $|F_D^2|$ que dans certains cas particuliers mais plusieurs approximations sont en général utilisées de manière satisfaisante : $(|F_{P+D}| - |F_P|)^2$ ou $(|F_{P+D}|^2 - |F_P|^2)$. Si le dérivé lourd possède une contribution anormale significative (voir paragraphe 7), la meilleure approximation de $|F_D^2|$ est $(|F_{P+D}^+| - |F_{P+D}^-|)^2$ où $|F_{P+D}^+|$ représente le module du facteur de structure du dérivé $(P + D)$ pour la réflexion hkl et $|F_{P+D}^-|$ pour la réflexion $\bar{h}\bar{k}\bar{l}$.

Deux problèmes mineurs ne seront pas abordés ici : le choix d'une origine commune à tous les dérivés lourds et le choix de l'énantiomère correcte. Le lecteur pourra consulter les ouvrages donner en bibliographie pour plus de détails.

7 Utilisation de la diffusion anormale

L'interaction des rayons X avec un atome donné j est modélisée par un facteur de diffusion f_j qui caractérise la diffusion de l'atome en prenant comme unité celle d'un électron libre. Dans les conditions dites à tort 'normales', f_j est un nombre réel c'est à dire qu'il n'y a pas de différence de phase entre l'onde diffusée par l'atome et celle diffusée par un électron libre (diffusion Thompson). Dans la réalité, les électrons d'un atome ne sont pas libres, ils sont répartis en orbitales d'énergies données et l'atome présente des discontinuités d'absorption correspondant à l'excitation de couches électroniques profondes. Les approximations de la théorie classique ne sont plus justifiées si l'énergie des rayons X incident est proche d'un seuil d'absorption de l'atome. Il se produit un phénomène de résonance qui se traduit par un changement phase de l'onde diffusée¹¹ (par rapport à un électron libre). Ce changement de phase dépend de la nature de l'atome et de la longueur d'onde incidente λ . Le facteur de diffusion atomique s'exprime alors par un nombre complexe et s'écrit sous la forme :

$$\begin{aligned} f &= {}^0f + \lambda\delta \\ &= {}^0f + \lambda f' + i \lambda f'' \end{aligned} \quad (32)$$

où 0f représente la diffusion dite 'normale', $\lambda f'$ et $\lambda f''$ deux termes correctifs¹² fonctions de la longueur d'onde λ . La figure 8 montre deux exemples de variation du facteur de diffusion

10. La fonction de Patterson est le produit de convolution de la densité électronique par son image obtenue par un centre de symétrie par rapport à l'origine :

$$P(\vec{u}) = V \rho(\vec{r}) * \rho(-\vec{r}) = V \int_v \rho(\vec{r}) \rho(\vec{r} + \vec{u}) d\vec{r} \quad (30)$$

On peut montrer que les maxima de la fonction de Patterson correspondent aux vecteurs interatomiques de la structure. Dans la relation 30, on peut remplacer la densité électronique par sa valeur en fonction de F_{hkl} (voir la relation 3) et on obtient alors

$$P(x, y, z) = \left(\frac{1}{V}\right) \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |F_{hkl}|^2 \exp(-2i\pi(hx + ky + lz)) \quad (31)$$

où x, y, z sont des coordonnées relatives dans la maille cristalline.

11. un traitement détaillé de l'origine de la diffusion anormale dépasse le cadre de cet article. Pour de plus amples développements, voir la partie bibliographie.

12. 0f varie beaucoup avec l'angle θ , par contre $\lambda f'$ et $\lambda f''$ dépendent peu de θ . On rappelle que pour un atome donné, à $\frac{\sin\theta}{\lambda} = 0$, 0f représente le nombre d'électrons de l'atome. On pourrait aussi montrer que

de 2 atomes (Se et Hg)

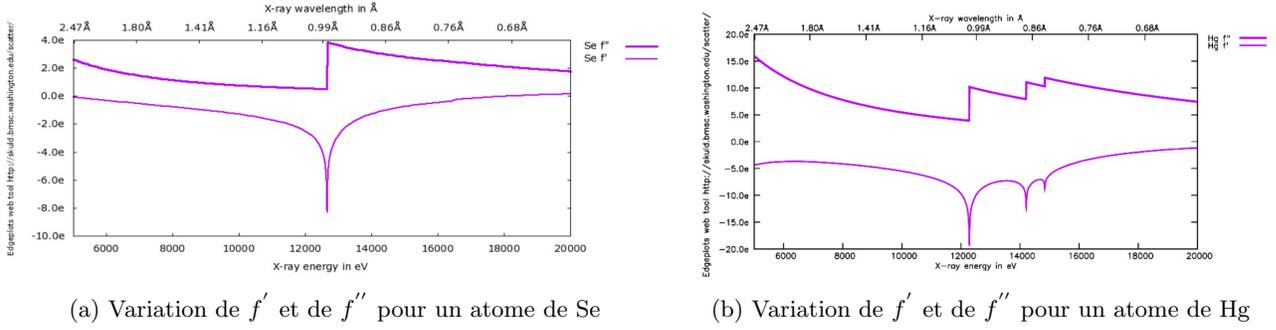


FIGURE 8 – Diffusion anormale. Exemples de variation de f' et de f'' © http://skuld.bmsc.washington.edu/scatter/AS_index.html

Dans la gamme d'énergie utilisée en cristallographie, (6keV ($\lambda = 2\text{Å}$) - 40 keV ($\lambda = 0,3\text{Å}$)), pour la plupart des atomes présents dans les macromolécules, la correction est négligeable et on n'en tient pas compte (diffusion dite 'normale'). Néanmoins, en raison de la potentialité énorme de cette méthode pour la détermination des phases (voir ci-dessous), on va en fait introduire des diffuseurs anomaux dans les macromolécules biologiques ou se placer à une longueur d'onde où la diffusion anormale d'atomes présents dans la structure est significative. Une conséquence importante de la présence d'un diffuseur anomal est que la loi de Friedel n'est plus vérifiée, c'est à dire que F_{hkl} et $F_{\bar{h}\bar{k}\bar{l}}$ ne sont plus des nombres complexes conjugués, ce qui se traduit par $I_{hkl} \neq I_{\bar{h}\bar{k}\bar{l}}$ ¹³. Ceci se déduit facilement de la relation permettant de calculer le facteur de structure d'une réflexion quelconque hkl en fonction des paramètres atomiques. En présence de diffuseurs anomaux, le facteur de structure d'une réflexion quelconque hkl s'écrit :

$$F_{hkl} = \sum_{j=1}^N \begin{pmatrix} f_j + \lambda f'_j + i \lambda f''_j \end{pmatrix} q_j w_j e^{2i\pi(hx_j + ky_j + lz_j)} \quad (33)$$

où (x_j, y_j, z_j) sont les coordonnées relatives dans la maille d'un atome j donné dont les paramètres de déplacements atomiques sont (q_j, w_j) . Ce qui peut s'écrire aussi sous la forme :

$$F_{hkl} = {}^0F_{hkl} + F'_{hkl} + F''_{hkl} \quad (34)$$

¹³ $\lambda f' \leq 0$ ou $\lambda f' \geq 0$, par contre $\lambda f'' \geq 0$.

13. on rappelle que dans le cas d'un cristal de parfaite mosaïcité, $I_{hkl} \propto |F_{hkl}|^2 = F_{hkl} F_{hkl}^*$

avec

$$\begin{aligned}
 {}^0F_{hkl} &= \sum_{j=1}^N {}^0f_j q_j w_j e^{2i\pi(hx_j+ky_j+lz_j)} \\
 F'_{hkl} &= \sum_{j=1}^N \lambda_j f'_j q_j w_j e^{2i\pi(hx_j+ky_j+lz_j)} \\
 F''_{hkl} &= \sum_{j=1}^N i \lambda_j f''_j q_j w_j e^{2i\pi(hx_j+ky_j+lz_j)}
 \end{aligned}$$

Il est alors facile de vérifier que, en notant F^* le complexe conjugué de F :

$$\begin{aligned}
 {}^0F_{\bar{h}\bar{k}\bar{l}} &= ({}^0F_{hkl})^* \\
 F'_{\bar{h}\bar{k}\bar{l}} &= (F'_{hkl})^* \\
 F''_{\bar{h}\bar{k}\bar{l}} &= -(F''_{hkl})^*
 \end{aligned}$$

ce qui permet de démontrer que

$$\begin{aligned}
 (F_{\bar{h}\bar{k}\bar{l}})^* &= {}^0F_{hkl} + F'_{hkl} - F''_{hkl} \\
 F_{hkl} - (F_{\bar{h}\bar{k}\bar{l}})^* &= 2F''_{hkl}
 \end{aligned}$$

Le non-respect de la Loi de Friedel (illustré graphiquement à la figure 9) sera mise à profit pour obtenir une information expérimentale des phases. On peut schématiquement considérer deux cas :

- Travail à une seule longueur d'onde. La contribution anormale permettra alors de lever partiellement l'ambiguïté de phase.
- Travail à plusieurs longueurs d'onde. On utilisera la variation du facteur de structure d'une réflexion à plusieurs longueurs d'onde pour lever complètement l'indétermination de phase et obtenir des phases expérimentales d'excellente qualité.

La méthode est connue sous le sigle anglo-saxon de "MAD" pour "Multiwavelength Anomalous Diffraction". Pour cette méthode, les différences de module de facteurs de structure $|\Delta F_{\pm hkl}| = ||F_{hkl}| - |F_{\bar{h}\bar{k}\bar{l}}||$ et $|\Delta F_{\Delta\lambda}| = ||\lambda_i F| - |\lambda_j F||$ sont en moyenne faibles (de l'ordre de quelques ‰), ce qui nécessite des mesures de très bonne qualité où les sources d'erreurs expérimentales sont minimisées. Cette méthode repose donc sur un rôle crucial de la qualité et de la complétude des données et de la redondance des mesures. De telles mesures sont maintenant classiquement obtenues en cristallographie des macromolécules biologiques grâce aux progrès technologiques (synchrotrons, détecteurs, cryocristallographie) et l'amélioration des logiciels de traitement de données de diffraction.

Deux approches historiques ont été utilisées pour décrire un même phénomène, la diffusion anormale ayant été utilisée en appui de la méthode MIR et dans ce cas un seul dérivé lourd est nécessaire. On trouve donc encore les sigles anglo-saxons pour décrire les conditions de phasage : "SIRAS" ("Single Isomorphous Replacement with Anomalous Scattering"), "MIRAS" ("Multiple Isomorphous Replacement with Anomalous Scattering"). Pour sortir de cette vision historique réductrice, on parle maintenant de phasage "SAD" ("Single-wavelength Anomalous Diffraction") ou "MAD" ("Multi-wavelength Anomalous Diffraction").

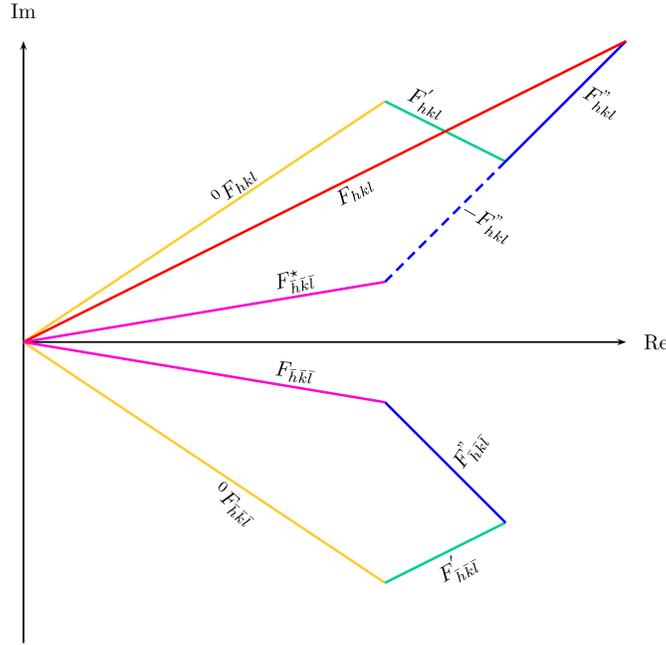


FIGURE 9 – Non respect de la Loi de Friedel en cas de diffusion anormale : Représentation Graphique. $F_{hkl} = {}^0F_{hkl} + F'_{hkl} + F''_{hkl}$, $F_{\bar{h}\bar{k}\bar{l}} = {}^0F_{\bar{h}\bar{k}\bar{l}} + F'_{\bar{h}\bar{k}\bar{l}} + F''_{\bar{h}\bar{k}\bar{l}}$ et $(F_{\bar{h}\bar{k}\bar{l}})^* = {}^0F_{hkl} + F'_{hkl} - F''_{hkl}$. En conséquence, $F_{\bar{h}\bar{k}\bar{l}} \neq F_{hkl}^*$ donc $|F_{hkl}| \neq |F_{\bar{h}\bar{k}\bar{l}}|$.

7.1 Travail à une seule longueur d'onde : Formalisme SAD. Vision graphique

Pour une longueur d'onde donnée λ , les relations 33 montrent que l'on peut écrire le facteur de structure d'une réflexion donnée en le décomposant en deux termes : l'un indépendant de la longueur d'onde et l'autre fonction de la longueur d'onde λ . Dans le cas particulier où parmi les N atomes d'une structure, il n'y a qu'un seul type de diffuseurs anomaux (cas le plus courant dans la pratique), on peut montrer que $\phi_{F''} = \phi_{F'} + \frac{\pi}{2}$, (ϕ correspondant à la phase d'un facteur de structure). Pour une paire de Bijvoet donnée, l'équation 33 peut être représentée graphiquement (Figure 10).

Les figures (11,12) montrent que si on ne connaît que $|F_{hkl}|$, $|F_{\bar{h}\bar{k}\bar{l}}|$ et le modèle complet des diffuseurs anomaux (F'_{hkl} ; F''_{hkl}), alors 2 solutions sont possibles pour 0F (notées 0F et 0F_2) sur la figure (11) (et bien sur 2 solutions possibles pour F^+ et F^-). Ces 2 solutions possibles correspondent à des $|{}^0F|$ différents.

Dans le cas historique du formalisme type "SIRAS", avec une structure P (aucun diffuseur anormal), une structure $(P + A)$ contenant A atomes diffuseurs anomaux supplémentaires et si les deux structures P et $(P + A)$ sont isomorphes, alors la connaissance de $|F_P^{hkl}|$, $|F_{P+A}^{hkl}|$, $|F_{P+A}^{\bar{h}\bar{k}\bar{l}}|$ permet de déterminer ϕ_P de manière unique. Seuls 2 espaces de diffraction sont alors nécessaires. Les relations ci dessus montrent aussi que les deux solutions SIR sont symétriques par rapport F' et les deux solutions SAD sont symétriques par rapport F'' .

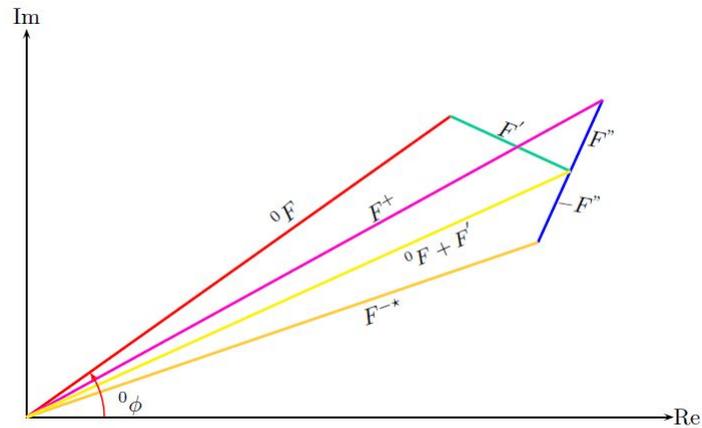


FIGURE 10 – Diffusion anormale. Conséquence à 1 longueur d’onde. Notation utilisée pour une réflexion hkl . ${}^0F = {}^0F_{hkl}$ $F^+ = F_{hkl}$ $F^- = F_{\bar{h}\bar{k}\bar{l}}$ $F^{-*} = (F_{\bar{h}\bar{k}\bar{l}})^*$ $F' = F'_{hkl}$ $F'' = F''_{hkl}$

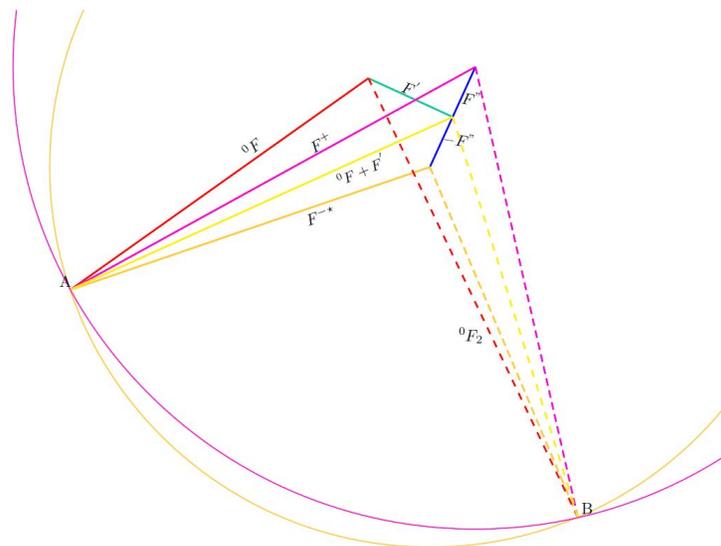


FIGURE 11 – Les 2 solutions de phases SAD. Le cercle centré en $(F' + F'')$ et de rayon $|F^+|$ coupe le cercle centré en $(F' - F'')$ et de rayon $|F^-|$ en A et B. Les points A et B sont symétriques par rapport à la direction portée par F'' . Ils correspondent aux 2 solutions possibles pour 0F (notées 0F et 0F_2)

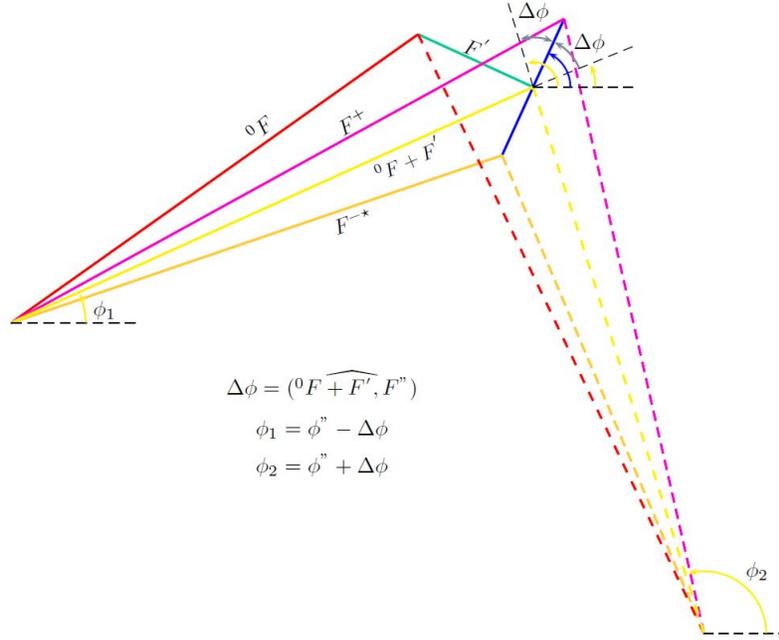


FIGURE 12 – Phasage SAD : les 2 solutions sont symétriques par rapport à F''

Dans le cas SAD, avec la notation ${}^\lambda F(H) = {}^0 F_T(H) + {}^\lambda F'_A(H) + i {}^\lambda F''_A(H)$, on peut montrer (Figure 12) que :

- les 2 solutions pour ${}^0 \phi_T$ sont symétriques par rapport à ${}^0 \phi_A + 90^\circ$
- les 2 phases possibles sont ${}^0 \phi_T^{Vrai}$ et $2{}^0 \phi_A + \pi - {}^0 \phi_T^{Vrai}$, avec ${}^0 \phi_T^{Vrai}$ étant la solution correcte.

En notant \overline{F} le complexe conjugué du facteur de structure F , la densité électronique calculée avec la moyenne des phases SAD peut s'écrire sous la forme :

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{h,k,l} ({}^0 F_T^{hkl} - \overline{{}^0 F_T^{hkl}} e^{2i\phi_A^{hkl}}) e^{-2i\pi(hx+ky+lz)}$$

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{h,k,l} {}^0 F_T^{hkl} e^{-2i\pi(hx+ky+lz)} - \left(\frac{1}{V}\right) \sum_{h,k,l=-\infty}^{+\infty} \overline{{}^0 F_T^{hkl}} e^{2i\phi_A^{hkl}} e^{-2i\pi(hx+ky+lz)} \quad (35)$$

Le deuxième terme de l'équation 35 correspond à un terme de bruit (voir tableau 3 dans le paragraphe 9 pour interprétation de ce terme). Ce bruit peut être fortement filtré par les méthodes de modifications de densité pour aboutir à une carte de densité électronique interprétable d'où les succès nombreux à partir d'un jeu SAD uniquement.

7.2 Travail à plusieurs longueurs d'ondes. Vision graphique

La Figure 11 correspond à une collecte à 1 longueur d'onde notée λ_1 . On peut faire la même figure avec les données collectées à 2 longueur d'onde notées λ_1 et λ_2 (figure 13). On se place à nouveau dans le cas particulier où parmi les N atomes d'une structure, il n'y a qu'un seul type de diffuseurs anomaux (cas le plus courant dans la pratique).

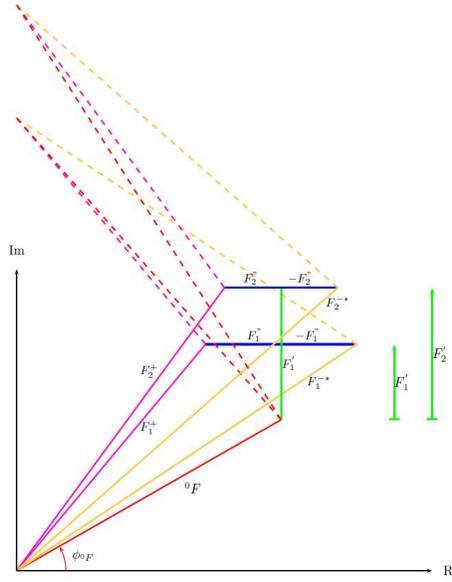


FIGURE 13 – Phasage MAD. Collecte à 2 longueurs d’onde. Les notations utilisées sont celles définies à la figure 10. Pour la longueur d’onde λ_1 , on dispose de $|F_1^+|$ $|F_1^-|$ F_1' F_1'' . Pour une longueur d’onde λ_2 , on connaît $|F_2^+|$ $|F_2^-|$ F_2' F_2'' .

A partir de cette figure 13, on comprend que, pour une réflexion donnée, la connaissance de ($|F_1^+|$, $|F_1^-|$, F_1' , F_1'' , $|F_2^+|$, $|F_2^-|$, F_2' , F_2'') permet de déterminer de manière unique 0F .

7.3 Travail à plusieurs longueurs d’ondes. Les mathématiques.

Considérons un cristal composé d’un nombre A de diffuseurs anomaux parmi un nombre N de diffuseurs dit ‘normaux’. Pour une longueur d’onde donnée λ , on peut écrire le facteur de structure en le décomposant en deux termes : l’un indépendant de la longueur d’onde et l’autre fonction de la longueur d’onde λ :

$${}^\lambda F(H) = {}^0F_N(H) + {}^\lambda F_A(H) \tag{36}$$

où H représente la réflexion d’indice hkl . On peut écrire cette relation sous la forme :

$$\begin{aligned} {}^\lambda F(H) &= {}^0F_N(H) + {}^0F_A(H) + {}^\lambda F_A'(H) + i {}^\lambda F_A''(H) \\ &= {}^0F_T(H) + {}^\lambda F_A'(H) + i {}^\lambda F_A''(H) \end{aligned} \tag{37}$$

où ${}^0F_T(H)$ inclut la contribution dite ‘normale’ de tous les $N + A$ atomes. Dans le cas particulier, qui est le cas rencontré dans la pratique, où les A diffuseurs anomaux sont de même nature (1 seul type de diffuseur anormal), on peut alors écrire ${}^\lambda F_A' = (\frac{\lambda f'}{0f}) {}^0F_A$ et ${}^\lambda F_A'' = (\frac{\lambda f''}{0f}) {}^0F_A$. L’équation (37) s’écrit alors :

$${}^\lambda F(H) = {}^0F_T(H) + (\frac{\lambda f'}{0f} + i \frac{\lambda f''}{0f}) {}^0F_A(H) \tag{38}$$

En utilisant cette relation, on peut calculer le module du facteur de structure $|{}^\lambda F(H)|$ pour la réflexion H et le module du facteur de structure $|{}^\lambda F(-H)|$ pour la réflexion $(-H = \bar{h}\bar{k}\bar{l})$. En appelant

${}^0\phi_T(H)$ et ${}^0\phi_A(H)$ les phases respectives de ${}^0F_T(H)$ et ${}^0F_A(H)$, en notant $|{}^0F_T| = |{}^0F_T(H)|$ $|{}^0F_A| = |{}^0F_A(H)|$ et en posant $\Delta\phi = {}^0\phi_T(H) - {}^0\phi_A(H)$ on obtient alors :

$$\begin{aligned} |{}^\lambda F(H)|^2 &= |{}^0F_T|^2 + a(\lambda) |{}^0F_A|^2 + b(\lambda) |{}^0F_T| |{}^0F_A| \cos\Delta\phi \\ &\quad + c(\lambda) |{}^0F_T| |{}^0F_A| \sin\Delta\phi \end{aligned} \quad (39)$$

et

$$\begin{aligned} |{}^\lambda F(-H)|^2 &= |{}^0F_T|^2 + a(\lambda) |{}^0F_A|^2 + b(\lambda) |{}^0F_T| |{}^0F_A| \cos\Delta\phi \\ &\quad - c(\lambda) |{}^0F_T| |{}^0F_A| \sin\Delta\phi \end{aligned} \quad (40)$$

avec $a(\lambda) = \frac{\lambda f'^2 + \lambda f''^2}{0 f^2}$, $b(\lambda) = 2 \frac{\lambda f'}{0 f}$, et $c(\lambda) = 2 \frac{\lambda f''}{0 f}$.

Les relations 39 et 40 constituent un système linéaire avec une contrainte non-linéaire ($\cos^2\Delta\phi + \sin^2\Delta\phi = 1$). ${}^0f, {}^\lambda f', {}^\lambda f''$ étant connus, les coefficients $a(\lambda), b(\lambda), c(\lambda)$ sont donc des constantes à une longueur d'onde donnée. Le système ci-dessus possède alors quatre inconnues $|{}^0F_T|^2, |{}^0F_A|^2, (|{}^0F_T| |{}^0F_A| \cos\Delta\phi), (|{}^0F_T| |{}^0F_A| \sin\Delta\phi)$, qui ne dépendent pas de la longueur d'onde de travail. Les observations expérimentales sont $|{}^\lambda F(H)|^2$ et $|{}^\lambda F(-H)|^2$. Un travail à une longueur d'onde aboutit donc à un système d'équations avec quatre inconnues et deux équations. En enregistrant un espace complet de diffraction à deux longueurs d'onde différentes λ_1 et λ_2 , on aura alors quatre équations donnant $(|{}^{\lambda_1} F(H)|^2, |{}^{\lambda_1} F(-H)|^2, |{}^{\lambda_2} F(H)|^2, |{}^{\lambda_2} F(-H)|^2)$ et toujours 4 inconnues $(|{}^0F_T|^2, |{}^0F_A|^2, |{}^0F_T| |{}^0F_A| \cos\Delta\phi, |{}^0F_T| |{}^0F_A| \sin\Delta\phi)$. Le système d'équations linéaires pourra donc être résolu mathématiquement. En général, on effectuera néanmoins les mesures avec au moins trois longueurs d'onde différentes afin d'optimiser le signal et d'augmenter la surdétermination du rapport nombre d'observations sur nombre de paramètres. Les courbes de variation de f' et de f'' en fonction de la longueur d'onde sont déterminés juste avant la collecte des données MAD par l'enregistrement du spectre de fluorescence X de l'échantillon. Les longueurs sont choisies pour optimiser le signal : l'une correspond au maximum de f'' , l'autre au minimum de f' et une ou plusieurs autres longueurs d'onde.

Le pouvoir de phasage de cette méthode est lié à la variation du facteur de structure de chaque réflexion entre deux longueurs d'onde λ_1 et λ_2 et peut s'exprimer approximativement dans le cas d'un seul type de diffuseur anormal par la relation :

$$\frac{\langle |{}^{\lambda_1} F| - |{}^{\lambda_2} F| \rangle}{\langle |F| \rangle} \simeq \sqrt{\frac{N_A}{2N_T}} \frac{|{}^{\lambda_1} f' - |{}^{\lambda_2} f' |}{f_N} \quad (41)$$

où $\langle |F| \rangle = \frac{\langle |{}^{\lambda_1} F| + |{}^{\lambda_2} F| \rangle}{2}$. La détermination des phases ${}^0\phi_T$ nécessitent de déterminer les positions des diffuseurs anormaux (on peut alors calculer ${}^0\phi_A$ et déterminer ${}^0\phi_T = \Delta\phi - {}^0\phi_A$). Pour ceci on utilise des fonctions de Patterson ou des méthodes directes à partir des $|{}^0F_A|$.

L'équation(38) et les équations (39) (40) ont été obtenues dans le cas particulier (le plus courant) où il y a un seul type de diffuseur anormal. Des équations plus complexes sont obtenues dans le cas où plusieurs types de diffuseur anormaux sont présents. En présence de q types de diffuseur anormaux, les équations donnant $|{}^\lambda F(H)|^2$ et $|{}^\lambda F(-H)|^2$ comporteront $(q+1)^2$ termes. La résolution du système nécessitera d'enregistrer des espaces complets de diffraction supplémentaires à différentes longueur d'onde.

7.4 Incorporation des diffuseurs anormaux

Pour introduire des diffuseurs anormaux, on peut aussi utiliser la méthode de trempage ; les sels de mercure (Hg, seuil L_{III} , $\lambda = 1.009 \text{ \AA}$), et de platine (Pt, seuil L_{III} , $\lambda = 1.072 \text{ \AA}$) sont les plus utilisés. Dans le cas des métalloprotéines, on peut mettre à profit la présence de métaux déjà présents, comme par exemple le zinc (Zn, seuil K, $\lambda = 1.284 \text{ \AA}$), le cuivre (Cu, seuil K, $\lambda = 1.381 \text{ \AA}$), le fer (Fe, seuil K, $\lambda = 1.743 \text{ \AA}$). Néanmoins, le succès de la méthode MAD vient de l'utilisation de sélénométhionine. Les protéines contenant en moyenne 2% de méthionine, une sélénométhionine est obtenue par la

substitution de l'atome de soufre par un atome de sélénium (le seuil d'absorption K correspond à $\lambda = 0,9793\text{\AA}$). Cette substitution est réalisée par des techniques de génie génétique qui permettent de bloquer la biosynthèse des méthionines dans les cellules et de fournir de la sélénométhionine en remplacement dans le milieu de culture. Les protéines mutantes ainsi obtenues sont fonctionnelles et ne présentent sauf exception pas de changement de conformation par rapport aux protéines naturelles. Récemment, une méthode ingénieuse d'incorporation d'acides aminés non naturels dans les protéines, et en particulier de Iodo-phenylalanine en lieu et place de l'acide aminé Phe classique, a permis d'utiliser la diffusion anormale de l'iode et de résoudre une structure 3D sur un premier cas test. La généralisation de cette méthode ouvre de nouvelles perspectives pour le phasage des macromolécules biologiques. Pour les acides nucléiques, un marquage analogue peut être réalisé en utilisant des bases bromées (le seuil d'absorption K du brome correspond à $\lambda = 0,92\text{\AA}$) comme par exemple la 5-bromouridine qui est un analogue structural de la thymidine.

Zbigniew Dauter a proposé une méthode alternative pour l'incorporation de diffuseurs anormaux. La méthode consiste en un trempage ultra court du cristal dans une solution cryoprotectrice contenant des ions halogènes (Br^- , I^-), qui peuvent occuper des sites dans les couches de solvant ordonnées autour des macromolécules biologiques. Plusieurs structures ont maintenant été résolues avec cette technique, le temps de trempage étant en général compris dans l'intervalle [15s-45s] avec des concentrations en ions halogénures dans la gamme [0,3M-1M]. Une extension de cette méthode a été récemment proposée par l'incorporation de cations monovalents (comme Cs^+ , Rb^+) ou polyvalents (particulièrement des lanthanides comme par exemple Gd^{3+} , Eu^{3+} , Sm^{3+}).

7.5 Détermination des f' et f''

Bien que des valeurs approchées des coefficients f' et f'' soient tabulées pour un atome donné, ces coefficients f' et f'' doivent être déterminés de manière plus précise pour chaque échantillon pour deux raisons :

- ces coefficients sont sensibles à l'environnement géométrique de l'atome diffuseur anormal,
- il faut tenir compte de la calibration du faisceau de rayons X incidents qui dépend de la longueur d'onde de travail et du type d'installation utilisée.

Le coefficient f'' est proportionnel au coefficient linéaire d'absorption μ , sa variation avec l'énergie E étant donné par la relation :

$$f''(E) = \frac{mc\epsilon_0}{\hbar e^2} E \mu(E) \quad (42)$$

où \hbar , ϵ_0 , m , c , e sont les constantes physiques classiques.

La relation de Kramers-Kronig permet de trouver f' étant donné $f''(E)$:

$$f'(E) = \frac{2}{\pi} \int_0^{+\infty} f''(E_0) \frac{E_0}{E^2 - E_0^2} dE_0 \quad (43)$$

Expérimentalement, pour déterminer f'' , on a besoin de connaître le spectre d'absorption de l'élément considéré (dans l'environnement donné) en fonction de l'énergie des rayons X incidents. Ce spectre d'absorption est déduit du spectre de fluorescence, la fluorescence étant proportionnelle à l'absorption. En pratique, on enregistre le spectre de fluorescence du cristal juste avant la collecte de données cristallographiques. A partir des courbes obtenues, on peut choisir les longueurs d'ondes de travail. Les valeurs de f' et f'' pourront être affinées au cours du processus de phasage.

7.6 Développements et Perspectives

Les équations (39, 40) n'impliquent aucune approximation et la justesse des phases "MAD" ne dépend que de la qualité des données de diffraction. Il est donc possible d'obtenir des phases expérimentales à haute résolution (si la qualité de diffraction du cristal le permet), ce qui n'est généralement

pas le cas de la méthode "MIR" où les problèmes de défauts d'isomorphisme deviennent prépondérants à haute résolution. De plus, avec les techniques de cryocristallographie, il est souvent possible d'effectuer l'ensemble des mesures sur un seul cristal. Les cartes de densité électronique que la méthode MAD permet d'obtenir sont en général d'une très grande qualité. Les limitations de cette méthode sont de deux ordres :

- l'accès à une source de rayonnement synchrotron est nécessaire,
- la détermination des positions des diffuseurs anomaux pose problème si le nombre de diffuseur anomaux est trop grand dans l'unité asymétrique (≥ 80 avec les logiciels actuels).

De nombreuses structures ont maintenant été résolues par la méthode MAD, et c'est la méthode de phasage de référence en cristallographie des macromolécules biologiques.

Dans certains cas très favorable (forte proportion de solvant, symétrie non cristallographique, données d'excellente qualité à une résolution suffisante), des phases SAD ("Single wavelength Anomalous Diffraction"), calculées en utilisant les données provenant d'une collecte à une seule longueur d'onde (maximum de f'' par exemple), améliorées ultérieurement par un processus de modification de densité, sont suffisantes pour obtenir une carte de densité électronique interprétable. De plus en plus de structures sont maintenant résolues avec ce phasage SAD. Cette méthode devient une alternative de plus en plus crédible au phasage MAD, en particulier dans le cas de cristaux trop fragiles et/ou se modifiant au cours des enregistrements.

Des développements récents concernent l'utilisation de diffuseurs anomaux intrinsèques aux macromolécules biologiques : atomes de soufre (cystéine, méthionine des protéines) ou de phosphore (acides nucléiques). Bien que le signal anomal soit très faible dans la gamme de longueur d'onde classiquement accessible, des premiers succès peuvent laisser imaginer que cette méthode deviendra peut être la méthode de choix pour la détermination des phases d'ici quelques années.

8 Le remplacement moléculaire

Le terme de remplacement moléculaire représente une famille de techniques dont l'objet est d'utiliser la présence d'une macromolécule, d'une sous-unité de macromolécule, ou d'un fragment de macromolécule, dans différents environnements cristallographiques, pour déterminer ou améliorer des phases initiales. C'est la méthode de choix lorsque la macromolécule à étudier (molécule cible) est parente à une autre, dont la structure tridimensionnelle est connue (molécule sonde). Ce qui est important ici est la parenté structurale, c'est-à-dire la similitude du repliement tridimensionnel. Cette similarité structurale est en général mise en évidence par des méthodes d'analyse des séquences. La méthode du remplacement moléculaire n'utilise que les modules des facteurs de structures de la protéine native sans nécessité l'ajout d'atomes supplémentaires. Pour utiliser cette méthode, on doit disposer de :

- Un espace complet de diffraction pour la molécule X dont on cherche à déterminer la structure tridimensionnelle,
- La structure d'une molécule M que l'on suppose être parente à X, c'est-à-dire que les deux molécules sont supposées avoir un repliement tridimensionnel commun (partiellement ou complètement).

Le terme remplacement moléculaire est à vrai dire impropre et on devrait plutôt utiliser le terme de placement moléculaire car il s'agit en fait de placer une molécule connue dans la maille cristalline de la molécule à résoudre. Chaque molécule dans l'unité asymétrique est caractérisée par son orientation et sa position, soit par 6 paramètres. Si l'unité asymétrique contient N molécules, il y a donc 6N paramètres à déterminer. Ce problème est actuellement résolu par 2 types de méthodes :

- des algorithmes fondés sur la recherche de similitudes entre la fonction de Patterson de la molécule cible et la fonction de Patterson de la molécule sonde (appelés ici pour simplifier Méthode de placement type A, MPA).
- des méthodes cherchant à déterminer simultanément les 6 paramètres de la molécule sonde (MP type B, MPB). Deux types d'algorithmes sont utilisés :

- une recherche systématique dans l'espace à dimension 6,
- des algorithmes stochastiques de 2 familles :
 - les algorithmes génétiques,
 - les méthodes fondés sur l'algorithme de Monte Carlo.

Les techniques appelées ici par convenance MPB, très gourmande au temps de calcul, se sont fortement développées ces dernières années et ont conduit dans plusieurs cas difficiles à trouver la réponse exacte restée inaccessible à la technique du placement moléculaire de type A. Néanmoins, dans le cadre de cet article, on ne traitera que des algorithmes fondés sur les corrélations entre fonctions de Patterson (MPA) qui présentent l'avantage de s'illustrer simplement. En effet, ils permettent de décomposer le problème de dimension $6N$ en 2 problèmes de dimension $3N$. Dans le cas d'une seule molécule sonde par unité asymétrique ($N=1$), il s'agira d'abord de trouver l'orientation de cette molécule (fonction de rotation) puis de déterminer sa position dans l'unité asymétrique (fonction de translation). C'est la méthode initiée en 1962 par Michael Rossmann et qui a subi plusieurs améliorations par de nombreux auteurs. Une amélioration importante a consistée à introduire une étape supplémentaire entre la fonction de rotation et la fonction de translation. Dans cette étape, connue sous le terme anglo-saxon de "Patterson Correlation refinement", on introduit un certain nombre de degrés de libertés internes pour la molécule sonde (division de la molécule sonde en j modules autonomes).

Le lecteur pourra consulter les références données à la fin de cet article pour approfondir les deux méthodes (MPA, MPB).

La méthode du (rem)placement moléculaire de type A utilise les propriétés de la fonction de Patterson qui est le produit de convolution de la densité électronique par son image obtenue par un centre de symétrie par rapport à l'origine. On peut montrer que les maxima de la fonction de Patterson correspondent aux vecteurs interatomiques d'une structure. Ces vecteurs peuvent être schématiquement divisés en deux classes :

- les vecteurs intramoléculaires, reliant deux atomes de la même molécule (ou d'une même sous-unité). Ces vecteurs ne dépendent que de l'orientation de la molécule par rapport aux axes de symétrie de la maille. Une rotation de la molécule dans la maille impliquera la même rotation des vecteurs intramoléculaires. Une translation de la molécule dans la maille n'affectera pas la distribution de ces vecteurs intramoléculaires.
- les vecteurs intermoléculaires reliant des atomes de deux molécules différentes. Ces vecteurs dépendent de l'orientation et de la position des molécules dans la maille cristalline.

Il n'est pas possible de séparer ces deux types de vecteurs interatomiques sans connaître la structure de la molécule, et de plus ces deux classes ne sont pas disjointes. Toutefois, les vecteurs intramoléculaires sont en moyenne plus courts que les vecteurs intermoléculaires, et se trouveront plus proches de l'origine de la fonction de Patterson. On séparera donc ces deux classes par un critère de longueur.

La fonction de Patterson d'une macromolécule, ou d'une sous-unité de macromolécule, a donc une distribution caractéristique de densité qui reflète son ensemble de vecteurs interatomiques. La "comparaison" de la fonction de Patterson observée (pour une molécule dont la structure est inconnue et que l'on cherche à déterminer) avec la fonction de Patterson calculée à partir de la structure d'une molécule parente connue, permettra de déterminer la position et l'orientation de la molécule parente dans la maille cristalline, et donc de déterminer des phases initiales pour les facteurs de structure de la molécule inconnue.

Les remarques ci-dessus concernant les vecteurs intramoléculaires et les vecteurs intermoléculaires sont à l'origine de la décomposition d'un problème de placement moléculaire de type A en deux étapes successives : rotation, translation. L'étape de rotation (fonction de rotation) déterminera l'orientation des molécules dans l'unité asymétrique, et l'étape de translation (fonction de translation) déterminera les positions des molécules correctement orientées.

8.1 La fonction de rotation

L'idée de base ici est que si l'on donne à la molécule M une orientation similaire à la molécule X dans le cristal, alors la fonction de Patterson calculée P_M sera analogue à la fonction de Patterson

observée pour la molécule X. Pour déterminer l'orientation inconnue de X, il suffit donc, pour toutes les orientations possibles de M de retenir celle pour laquelle P_M et P_X seront similaires. Pratiquement, il s'agit de trouver la ou les rotations (définie par une matrice C) permettant de "superposer" la fonction de Patterson P_X de la molécule inconnue et la fonction de Patterson calculée à partir de la molécule modèle P_M . Le degré de similitude entre les deux fonctions de Patterson est exprimé par une fonction de corrélation calculée sur un volume U, généralement une sphère centrée à l'origine. Ce volume est choisi de façon à éliminer les vecteurs intermoléculaires entre molécules appartenant à des mailles voisines. Mathématiquement on exprimera ceci par un coefficient de corrélation entre les deux fonctions de Patterson :

$$Corr(\theta_1, \theta_2, \theta_3) = \int_U P_X(r) P_M(Cr) dV \quad (44)$$

où $P_X(r)$ représente la valeur de la Patterson du cristal X en r , $P_M(Cr)$ représente la valeur de la Patterson du cristal M en (Cr) , C étant la matrice de rotation construite sur trois angles $(\theta_1, \theta_2, \theta_3)$ qui définissent une rotation quelconque dans l'espace. Les maxima de la fonction de corrélation $Corr$ correspondent aux meilleures superpositions de P_X avec P_M .

Cette fonction de rotation peut être calculée soit dans l'espace direct (opérations sur les fonctions de Patterson), soit dans l'espace réciproque en exprimant les fonctions de Patterson en terme de facteurs de structure. Dans tous les cas, en raison de la taille des macromolécules, les calculs sont longs et coûteux en temps, même pour les supercalculateurs les plus puissants. De nombreux algorithmes ont été développés pour remédier à ceci ; ils diffèrent par les approximations numériques qu'ils utilisent effectuer les calculs et par certains critères de pondération et de sélection.

8.2 La fonction de translation

Le terme fonction de translation se rapporte ici à toute technique permettant de trouver la position d'une molécule correctement orientée dans une maille cristalline. Ceci comprend des méthodes travaillant dans l'espace réciproque et des méthodes de calcul dans l'espace direct. Elles nécessitent toutes d'avoir trouvé la ou les solutions exactes de la fonction de rotation. Plusieurs méthodes sont utilisées. Il s'agit de trouver les maxima de la fonction $T(t)$ définie par :

$$T(t) = \int_V P_x(u) P_c(u, t) du \quad (45)$$

où $P_x(u)$ est la fonction de Patterson observée en un point repéré par le vecteur u pour la molécule étudiée X, $P_c(u, t)$ est la fonction de Patterson calculée pour la molécule modèle M correctement orientée dans la maille et positionnée en un point repéré par le vecteur t et V le volume de la maille. La fonction $T(t)$ présente un maximum pour le vecteur t_{sol} correspondant à un vecteur intermoléculaire reliant deux molécules modèles (donc reliées entre elles par un élément de symétrie cristallographique) correctement positionnées dans la maille cristalline.

8.3 Remarques

Le point central de cette méthode est le choix de la molécule modèle. Si dans certains cas ce choix est relativement simple, il peut être très délicat dans d'autres. Le choix de la molécule parente peut être fait selon deux critères :

- à partir d'analyses de séquence. Si deux macromolécules ont une forte homologie de séquence, elles adopteront certainement le même repliement tridimensionnel.
- à partir d'analyses fonctionnelles. Si deux protéines ont la même fonction biologique, elles peuvent avoir la même structure dans l'espace.

Si la molécule modèle est trop différente (localement ou globalement) de la molécule inconnue, alors la méthode du remplacement moléculaire risque d'échouer ou même de donner une solution fautive. Il faudra donc toujours suivre avec précaution une solution issue du remplacement moléculaire et s'assurer de son bon comportement au cours de l'affinement cristallographique (diminution du facteur de désaccord cristallographique R_{libre} , voir paragraphe 10). Le grand avantage de cette méthode est qu'elle ne nécessite qu'un seul espace de diffraction. En cas de succès, elle peut donc permettre d'aboutir rapidement à la structure tridimensionnelle. Dans certains cas, elle peut mener à une impasse totale. Il faudra alors revenir aux méthodes de phasages MIR (ou MAD) présentées aux paragraphes précédents.

8.4 Détermination des phases

Lorsqu'on dispose d'une molécule homologue modèle et que les problèmes de la rotation et de la translation ont été résolus, le calcul d'un premier jeu de phases est alors possible et permet, avec les modules des facteurs de structure observés, le calcul d'une carte de densité électronique, qui permettra une première construction de la molécule inconnue.

9 Cartes de densité électronique

9.1 Définitions

La diffraction des rayons X par un cristal permet d'accéder à une fonction appelée densité électronique qui reflète une distribution continue d'électrons dans la maille cristalline. Cette fonction présente des maxima dans les régions occupées par les atomes et est donc en général interprétée en termes de positions atomiques. En théorie, la densité électronique $\rho(x, y, z)$ en un point de coordonnées relatives (x, y, z) dans la maille cristalline est calculée par la relation :

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |F_{hkl}| e^{i\phi_{hkl}} e^{-2i\pi(hx+ky+lz)} \quad (46)$$

où $|F_{hkl}|$ et ϕ_{hkl} sont respectivement le module du facteur de structure et la phase de la réflexion hkl , V le volume de la maille cristalline. En absence de diffuseurs anomaux, cette fonction est à valeurs réelles. En présence de diffuseurs anomaux, cette fonction densité électronique est une fonction à valeurs complexes dont la partie imaginaire ne dépendra que de la structure des diffuseurs anomaux.

Il est important de bien comprendre le lien bijectif entre les données de l'espace direct (la structure atomique décrite par des positions et des paramètres de déplacements atomiques comme le paragraphe 10) et l'espace réciproque (les facteurs de structure qui caractérisent les ondes diffusées). Quelques relations sont représentées dans le tableau 3.

Dans la pratique, la diffraction des rayons X permet d'accéder à un nombre limité de réflexions (ensemble noté D^* ci-dessous), chaque réflexion hkl étant caractérisé par un module du facteur de structure $|F_{hkl}^{obs}|$ et une phase calculée ϕ_{hkl}^{calc} . La densité électronique calculée est alors donnée par la relation :

$$\rho(x, y, z)^{obs} = \left(\frac{1}{V}\right) \sum_{hkl \in D^*} |F_{hkl}^{obs}| e^{i\phi_{hkl}^{calc}} e^{-2i\pi(hx+ky+lz)} \quad (47)$$

Il y a donc 3 sources d'erreurs principales qui font que $\rho(x, y, z)^{obs}$ donnée par la relation (47) peut être fortement différente de la densité théorique donnée par la relation (46) et peuvent rendre l'interprétation de cette fonction (en position atomiques) difficile voir impossible. On peut décomposer ces erreurs en deux classes :

Structure	positions atomiques	facteur de diffusion	Facteur de Structure	module FS	Phase FS	Densité électronique
A	\vec{r}_j	f_j	F_H	$ F_H $	α_H	$\rho(r)$
B	$-\vec{r}_j$	f_j	$\overline{F_H}$	$ F_H $	$-\alpha_H$	$\rho_B(r) = \rho(-r)$
C	\vec{r}_j	$-f_j$	$-F_H$	$ F_H $	$\pi + \alpha_H$	$\rho_C(r) = -\rho(r)$
D	$-\vec{r}_j$	$-f_j$	$-\overline{F_H}$	$ F_H $	$\pi - \alpha_H$	$\rho_D(r) = -\rho(-r)$

Tableau 3 – Pour une structure donnée, relations entre un modèle structural et les facteurs de structure. Pour une réflexion donnée notée H , $\overline{F_H}$ représente le complexe conjugué du facteur de structure F_H . Les structures B,C,D sont des modifications de la structure A, la nature des modifications étant décrites dans les colonnes correspondantes

— les erreurs sur la valeur des coefficients utilisés.

Pour chaque réflexion hkl , il s'agit de l'erreur sur la mesure du module du facteur de structure $|F_{hkl}|$ ($|F_{hkl}^{obs}| \neq |F_{hkl}^{vrai}|$, erreur en général de l'ordre de quelques %) et de l'erreur sur la phase ϕ_{hkl} ($\phi_{hkl}^{calc} \neq \phi_{hkl}^{vrai}$, erreur qui peut être de plusieurs dizaines de degrés).

— les erreurs de troncatures.

La densité électronique expérimentale est calculé sur un sous-espace de l'ensemble des hkl (noté D^* dans l'expression (47), et D^* est différent de \mathbb{Z}^3).

En effet expérimentalement, il n'a pas été possible d'enregistrer les réflexions pour lesquelles $|S_{hkl}^{\vec{r}}| < |S_{min}^{\vec{r}}|$ et $|S_{hkl}^{\vec{r}}| > |S_{max}^{\vec{r}}|$ (rappel : $S_{hkl}^{\vec{r}} = ha^* + kb^* + lc^*$). Le domaine D^* correspond dans l'espace réciproque aux réflexions hkl dont le module de $S_{hkl}^{\vec{r}}$ appartient au domaine $[|S_{min}^{\vec{r}}|; |S_{max}^{\vec{r}}|]$. Dans l'espace direct, D^* correspond à l'intervalle $[d_{min}; d_{max}]$ ($d_{min} = \frac{1}{|S_{max}^{\vec{r}}|}$, $d_{max} = \frac{1}{|S_{min}^{\vec{r}}|}$). Dans la pratique, on rappelle que d_{min} est appelé limite de diffraction du cristal. La non-complétude des données dans l'intervalle $[|S_{min}^{\vec{r}}|; |S_{max}^{\vec{r}}|]$ est aussi bien sur une autre source d'erreurs dans la fonction de densité électronique.

Nous avons déjà vu au paragraphe 5 que l'information de phase est cruciale pour l'exactitude du calcul de la fonction densité électronique. Nous ne reviendrons pas ici sur cet aspect essentiel pour la justesse et l'interprétation d'une fonction densité électronique. Nous verrons aussi dans le paragraphe 9.2 que ces erreurs de phases pourront être éventuellement réduites par les méthodes de modification de densité (voir paragraphe 9.2).

Dans la fonction densité électronique expérimentale, l'existence de ces erreurs se traduira par :

- un élargissement des "maxima",
- un déplacement des maxima par rapport aux vraies positions,
- l'apparition de faux maxima,
- l'apparition de minima négatifs,
- des superpositions d'extrema (conséquences des trois effets ci dessus).

En fonction de l'amplitude des erreurs, on ne pourra pas alors interpréter la densité électronique expérimentale.

Dans la pratique, en absence de troncature sévère à basse résolution, la limite de diffraction d'un cristal (d_{min} notée ci-dessus), ou plus exactement la résolution à laquelle un espace complet de diffraction a pu être mesuré avec précision, fixera les barrières de ce que l'on pourra "voir" ou interpréter dans la fonction de densité électronique. Conséquence directe des effets de troncatures, on pourrait démontrer que, dans la fonction de densité électronique expérimentale, la distance minimale pour que 2 détails soient vus de manière isolée est théoriquement de $\simeq 0,7d_{min}$ (en pratique $\simeq d_{min}$). Pour fixer les idées, sous réserve bien sur d'avoir les phases exactes des facteurs de structures :

- si $d_{min} \simeq 5\text{Å}- 6\text{Å}$, on ne pourra définir que la forme grossière de la macromolécule,
- si $d_{min} \simeq 3,5\text{Å}- 4\text{Å}$, on pourra construire les structures secondaires et le repliement de la chaîne principale,
- si $d_{min} \simeq 2\text{Å}- 3\text{Å}$, on construira la structure tridimensionnelle complète, on positionnera les molécules de solvant ordonné, les ions et les petits substrats éventuels,
- si $d_{min} \leq 1,2\text{Å}$, on pourra individualiser les atomes.

Dans la pratique, on pondère dans le calcul de la fonction densité électronique chaque information de phase par la figure de mérite m_{hkl} (définie par l'équation 28) et on calcule une expression du type

$$\rho(x, y, z)^{obs} = \left(\frac{1}{V}\right) \sum_{hkl \in D^*} m_{hkl} |F_{hkl}^{obs}| e^{i\phi_{hkl}^{calc}} e^{-2i\pi(hx+ky+lz)} \quad (48)$$

On appelle carte de densité électronique l'expression sur une grille donnée (donc calculée de façon discontinue) de la fonction densité électronique (qui est une fonction continue). Le pas de la grille est en général choisi à $\frac{d_{min}}{3}$ (voir les références bibliographiques pour plus de détails sur cet aspect).

9.2 Amélioration par des méthodes de modification de densité

Pour améliorer l'interprétation des cartes de densité électronique, il faut donc diminuer l'erreur commise sur chaque phase. Cela est réalisé par des méthodes de modification de densité qui utilisent des contraintes physico-chimiques que la densité électronique doit respecter. Les connaissances utilisées concerne la présence de solvant, l'atomicité et la positivité de la densité électronique, les histogrammes de densité, la symétrie non cristallographique. Ces propriétés introduisent des relations entre certains facteurs de structures (modules et phases) qui se traduisent par des contraintes sur la densité électronique; leur incorporations permettra donc d'améliorer les cartes de densité électronique. L'utilisation de ces contraintes sur la densité est réalisée par un processus cyclique qui alterne des calculs dans l'espace direct (opération sur les cartes de densités) et dans l'espace réciproque (opérations sur les facteurs de structure) et comprend typiquement les étapes suivantes (figure 14) :

1. Calcul d'une carte de densité électronique avec des phases initiales et les facteurs de structures observés $|F_{obs}|, \phi_{ini}$, chaque terme étant pondéré par une figure de mérite m_{ini} ,
2. Modification de la densité électronique, compte-tenu de une ou plusieurs contraintes que l'on détaillera ci-dessous,
3. Calcul de facteur de structure ($|F_{calc}|, \phi_{calc}$) pour chaque réflexion par transformation de Fourier inverse, à partir de la densité électronique modifiée.
4. Combinaisons des facteurs de structure observés avec les facteurs de structure calculés pour aboutir à un nouveau jeu de coefficients $|F_{nouv}|, \phi_{nouv}, m_{nouv}$
5. Retour à l'étape 1, avec les coefficients $|F_{nouv}|, \phi_{nouv}, m_{nouv}$. En général on prendra $|F_{nouv}| = |F_{obs}|$ sauf pour quelques réflexions qui pourraient ne pas avoir pu être enregistrées.

Ce processus cyclique est répété jusqu'à ce que la convergence soit atteinte selon des critères prédéfinis par l'utilisateur. Selon la validité et le type de contraintes appliquées, ces méthodes apportent des améliorations souvent spectaculaires des cartes de densité électronique.

9.2.1 Nivellement de solvant

Les cristaux de macromolécules contiennent souvent une forte proportion de solvant (de 20% à 90%). On s'attend donc à priori à visualiser dans les cartes de densité électronique une zone contenant la macromolécule et une zone de solvant. Cette distinction entre la volume contenant la macromolécule (on dira un masque) et le volume de solvant est souvent suffisamment définie même pour des cartes calculées avec des phases initiales relativement fausses. Cette méthode exploite le fait que dans des cartes à moyenne résolution, les zones de solvant présentent une densité faible et relativement plate

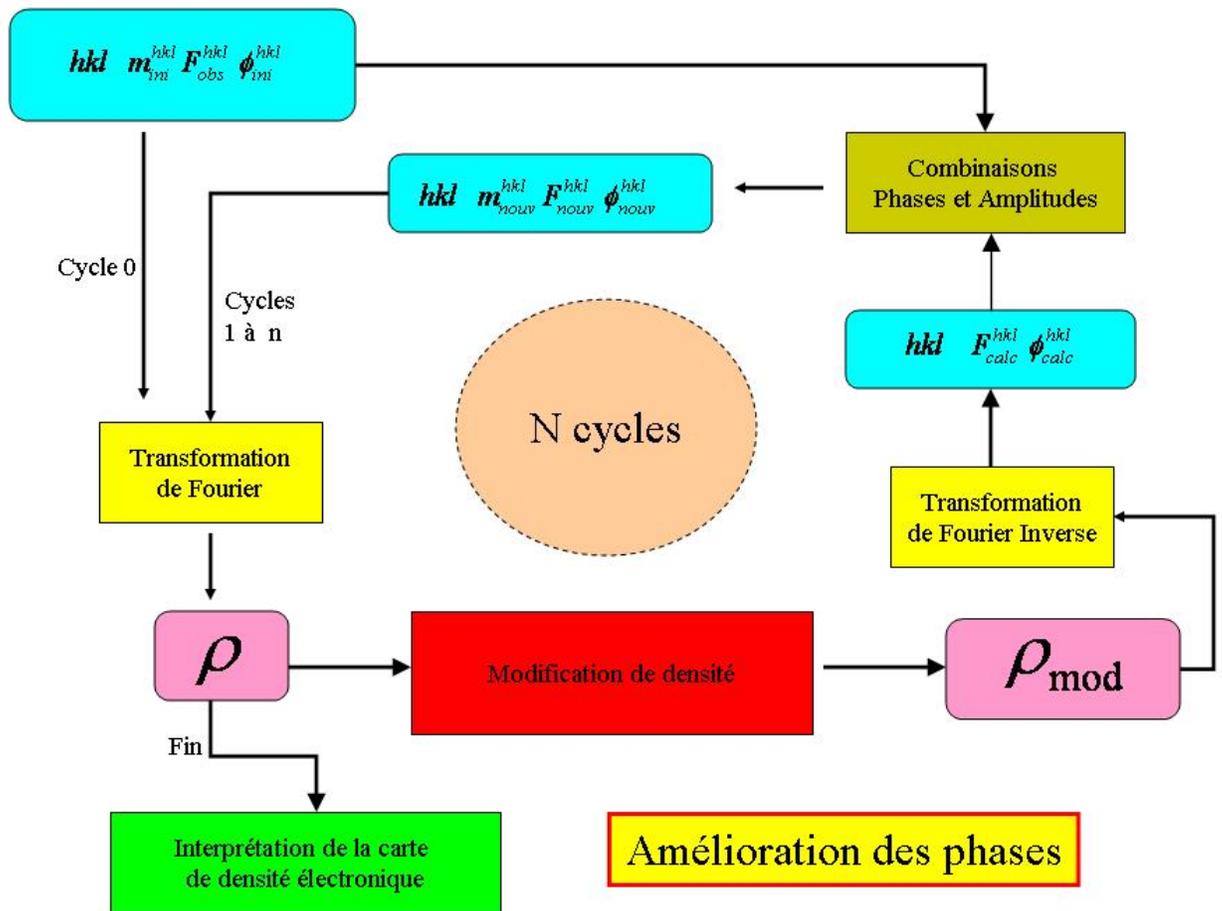


FIGURE 14 – Amélioration des phases par la technique de modification de densité électronique.

avec très peu de fluctuations. Ceci est du au désordre des molécules de solvant non-liées. La contrainte de nivellement de solvant s'applique en définissant au préalable les limites entre la macromolécule et le solvant (définition d'un masque moléculaire) et en remplaçant la densité dans la zone de solvant par une densité uniforme. Ceci est maintenant réalisé par des méthodes automatiques. Plus la proportion de solvant est importante, plus les contraintes imposées sont fortes et productives.

9.2.2 Atomicité

L'atomicité utilise l'équation de Sayre bien connue en cristallographie des petites molécules qui contraint la forme locale de la densité électronique dans le cas d'atomes identiques et bien séparés les uns des autres (résolution atomique). Pour les petites molécules (moins de 200 atomes), ces contraintes sont suffisantes pour résoudre la structure *ab-initio*, c'est-à-dire avec les modules des facteurs de structure de la protéine native seule. Ceci est possible car ces cristaux diffractent jusqu'à très haute résolution ($\leq 1\text{\AA}$). C'est rarement le cas des macromolécules biologiques, qui de plus comportent un nombre trop important d'atomes. Pour les macromolécules biologiques à moyenne résolution (2,5 \AA), les contraintes imposées par l'atomicité sont approximativement valables et elle est surtout employée conjointement à d'autres techniques de modification de densité pour améliorer les cartes de densités électronique.

9.2.3 Histogrammes des densités

Cette technique repose sur l'observation que l'histogramme de répartition de la densité électronique (fraction de la maille cristalline qui a une densité située entre deux valeurs t et $t + \delta t$) calculés pour de nombreuses protéines avec les phases exactes (hors solvant) est très similaire pour des cartes calculées à une résolution donnée. L'histogramme de répartition représente alors une probabilité de distribution de la densité électronique, indépendante de la macromolécule étudiée et ne dépendant que de la résolution à laquelle la densité électronique est calculée. La contrainte s'applique en calculant une fonction correctrice qui cherche à calquer l'histogramme de la densité observé à celui d'une densité idéale. Cette fonction est alors appliquée à la densité électronique.

9.2.4 Symétrie non cristallographique

Dans de nombreux cas de cristaux de macromolécules biologiques, l'unité asymétrique contient plusieurs macromolécules identiques (un dimère, trimère ou un tétramère, voir plus). Il existe alors une symétrie locale (rotation, translation) reliant les macromolécules équivalentes. Cette symétrie supplémentaire est appelée symétrie non-cristallographique car elle est respectée uniquement à l'intérieur d'un volume (on dit une enveloppe) qui est inclus dans l'unité asymétrique. Cette enveloppe ne peut pas remplir complètement l'unité asymétrique. De plus, le réseau cristallin n'est pas conservé par cette symétrie.

Cette symétrie non cristallographique est mise en évidence dans la fonction de Patterson car elle induit des caractéristiques particulières de celle-ci. On peut alors déterminer la nature, l'orientation et la position des éléments de symétrie dans l'unité asymétrique à partir de la fonction de Patterson de la macromolécule native. La symétrie non cristallographique introduit une redondance dans le jeu d'intensités diffractées qui induit des contraintes entre les facteurs de structures (donc entre les phases). En fait c'est l'expression de ces contraintes dans l'espace direct, qui est à la base du formidable succès de cette méthode en tant que processus d'amélioration et extension des phases. La technique de modification de densité consiste alors à imposer cette symétrie en remplaçant la densité de chaque point à l'intérieur de l'enveloppe par la moyenne entre sa densité et celle de ces n points équivalents (n étant l'ordre de la symétrie non cristallographique). La présence de cette symétrie non cristallographique est un facteur puissant pour l'amélioration des phases.

9.3 Extension des phases

Dans la pratique, les méthodes de phasage ne permettent d'obtenir des phases que jusqu'à une résolution limitée, qui est souvent inférieure à la résolution pour laquelle on a pu enregistrer les modules des facteurs de structure. On peut par exemple disposer d'un espace complet de diffraction jusqu'à 2Å de résolution et ne disposer de phases expérimentales que jusqu'à 3,5Å de résolution. Cela peut être typiquement le cas qui se produit dans la méthode du remplacement isomorphe multiple où les défauts d'isomorphisme peuvent sévèrement affecter le pouvoir de phasage. Les méthodes de modifications de densité permettent aussi de calculer à partir de phases connues à basse résolution les phases des réflexions de plus haute résolution. On parlera alors d'extension de phases. Ceci se réalise simplement dans le processus d'écrit par la figure 14 où dans l'étape de transformation de Fourier inverse peut se faire à une résolution légèrement supérieure à celle utilisée pour calculer la carte dans l'étape 1. Cette extension des phases doit être réalisée de manière progressive et son apport est d'autant plus important que les contraintes imposées sont fortes. Les cas extrêmes et typiques concernent les cristaux de virus à symétrie icosaédrique qui possèdent souvent une symétrie non cristallographique très élevée (ordre 15, 30 voir 60) ce qui permet de déterminer les structures cristallines à résolution atomique à partir de phases initiales à basse résolution.

9.4 Construction de la structure

Construire la structure consiste à placer les atomes de la macromolécule dans la densité électronique. Ce travail est d'autant plus facile que la résolution est élevée et que les phases utilisées pour calculer la carte sont proches des valeurs exactes. De nombreux logiciels d'aide à la construction sont disponibles et facilitent grandement le travail. Cette construction utilise des stations graphiques permettant de manipuler en temps réel les cartes de densités électroniques et les macromolécules biologiques. C'est sûrement l'étape la plus vivifiante du processus cristallographique car elle permet de dévoiler la structure atomique de la macromolécule étudiée. Les cartes de densité électronique sont représentées par des surfaces d'isodensité. La première étape d'analyse d'une carte de densité électronique consiste d'abord à repérer le squelette de la macromolécule biologique avant de pouvoir positionner les détails plus fins. Dans de nombreux cas où la qualité des phases est limitée, la première interprétation est délicate et suppose une très bonne expertise. L'obtention d'un premier modèle permet de calculer des phases qui pourront être combinées aux phases expérimentales et d'initier un nouveau processus de modification de densité. Des phases plus exactes pourront alors être obtenues qui donneront des cartes de densité électronique plus interprétables. Le modèle sera alors amélioré.

La construction du modèle dépend fortement de la limite de résolution des données de diffraction. Des données jusqu'à 3 Å sont nécessaires pour construire une structure détaillée de la macromolécule biologique. Pour visualiser les molécules de solvants ou d'ions, qui sont très importants pour les réactions biologiques, il est nécessaire d'avoir une limite de résolution des données de diffraction d'au moins 2,8 Å. La validité de l'information biologique qui pourra être déduit de la structure est fortement contrainte par la qualité des données de diffraction et par la limite de diffraction effective des cristaux.

Grâce aux outils informatiques puissants (machines et logiciels) sans cesse améliorés, le nombre d'heures nécessaires à la construction du premier modèle est de plus en plus court. Cette étape qui nécessitait plusieurs mois de travail il y a encore quelques années peut généralement se faire en quelques jours, voir en quelques heures. Si la limite de diffraction est suffisante et si les phases sont proches des phases exactes, les logiciels de construction (voir bibliographie) permettent maintenant d'obtenir de manière automatique, en général en moins d'une heure, une grande partie de la structure.

10 Affinement d'une structure cristallographique

10.1 Introduction

Le but de l'affinement d'une structure cristallographique est non seulement d'améliorer le modèle construit en terme de positions atomiques lors des étapes précédentes mais aussi de déterminer les

paramètres caractérisants les mouvements des atomes, afin de représenter au mieux la (ou les) structure(s) présentes dans le cristal. Cette étape du processus cristallographique consiste à optimiser le modèle issu de l'interprétation des cartes de densité électronique, par un procédé mathématique qui minimise le plus souvent l'écart entre les modules de facteurs de structure observés et les modules des facteurs de structure calculés. D'une façon générale, l'affinement d'une macromolécule biologique est un processus d'optimisation non linéaire d'une certaine fonction (que l'on définira) dépendant de paramètres et reposant sur des observations.

Schématiquement 4 parties sont à considérer :

- le choix de la fonction à optimiser,
- le choix du processus mathématique d'optimisation,
- les observations disponibles,
- les paramètres (variables) de la fonction.

Dans l'affinement cristallographique, les observations correspondent à tout ce qui est connu sur le cristal avant l'étape d'affinement :

- les paramètres cristallins,
- les modules des facteurs de structures (dont le nombre est fixé par la limite de diffraction du cristal),
- la stéréochimie des macromolécules étudiées,
- éventuellement les phases expérimentales initiales.

Les paramètres à affiner sont le plus souvent les positions atomiques et les paramètres de déplacements atomiques. Ces derniers modélisent le désordre statique et dynamique des atomes et sont généralement connus sous le terme de facteurs de Debye-Waller et facteurs d'occupation.

La fonction à optimiser s'exprime le plus souvent de deux manières : soit sous forme de résiduels de moindres carrés ou de pseudo-fonction d'énergie empirique associée soit sous forme d'un maximum de vraisemblance. Plusieurs méthodes d'optimisation ont été utilisées. La première famille de technique d'optimisation repose sur la méthode du gradient (en matrice complète, matrice creuse, gradients conjugués ou gradients conjugués préconditionnés), qui utilise un développement de la fonction à optimiser en série de Taylor. La seconde famille de technique consiste en l'utilisation d'algorithmes stochastiques, le plus utilisé étant la méthode du recuit simulé. L'étude de ces techniques d'optimisation complexes dépasse le cadre de cet article et ne seront pas traitées ici. Le lecteur est invité à consulter la bibliographie pour une présentation détaillée de ces méthodes. Dans tous les cas, l'affinement cristallographique des macromolécules biologiques est un problème délicat en raison de la non-linéarité et de la complexité de la fonction à optimiser, de l'interdépendance et du nombre très important de paramètres à affiner et finalement du nombre limité d'observations.

10.2 Principes

L'affinement cristallographique consiste le plus souvent à minimiser l'écart entre les modules de facteurs de structure observés et les modules des facteurs de structure calculés¹⁴. Les facteurs de structure sont calculés à partir de la nature et de la position des atomes dans la maille par la relation :

$$F_{hkl}^{calc} = \sum_{j=1}^N q_j f_j w_j^{hkl} e^{2i\pi(hx_j + ky_j + lz_j)} \quad (49)$$

où f_j est le facteur de diffusion de l'atome j de coordonnées relatives (x_j, y_j, z_j) , (q_j, w_j^{hkl}) les paramètres de déplacements atomiques, N est le nombre d'atomes de la maille. Pour un atome j donné, q_j est appelé facteur d'occupation, w_j^{hkl} le facteur de Debye-Waller (on utilise actuellement

14. c'est un biais majeur car on a vu au paragraphe 5 que l'information de phase était l'information cruciale pour la justesse d'une structure (voir plus loin l'introduction du *Ribre* pour tenter de remédier à ce problème)

le terme de facteur ou paramètres de déplacements anisotropes). Les facteurs q_j et w_j sont obtenus au cours du processus d'affinement cristallographique décrit dans ce paragraphe et ils cherchent à traduire la position et les mouvements de chaque atome dans le cristal (notion de désordre). En effet, même si un cristal est en théorie une répétition périodiquement parfaitement ordonnée d'un motif dans l'espace, en pratique (et en particulier pour les cristaux de macromolécules biologiques qui contiennent une proportion importante de solvant), de nombreux écarts à cette symétrie cristallographique (ordre cristallin) existent. Les molécules peuvent alors adopter différentes microconformations dans le cristal (désordre conformationnel). La structure cristallographique correspondant à une double moyennation dans le temps et dans l'espace de l'ensemble des structures présentes dans le cristal pendant le temps de l'expérience, le modèle construit en termes de paramètres atomiques se doit de refléter au mieux les positions et les mouvements des atomes dans le cristal. On distingue deux types de désordre, le désordre statique et le désordre dynamique et pour simplifier :

- Le désordre statique traduit des conformations différentes adoptées dans les molécules du cristal par un atome ou un groupe d'atomes.
- Le désordre dynamique traduit les mouvements de vibration d'un atome autour d'une position d'équilibre, l'amplitude de ces mouvements dépendant entre autre de la température. Les périodes de vibrations atomiques sont grandes vis-à-vis du temps que mettent les rayons X à traverser le cristal. A un instant donné, le spectre de diffraction du cristal correspond à celui d'un cristal "figé" où les atomes sont disposés aléatoirement hors de leurs positions d'équilibre.

Dans de nombreux cas, on utilise une simplification du facteur de Debye-Waller par un facteur appelé improprement "d'agitation thermique"¹⁵ isotrope qui sera alors caractérisé par un paramètre atomique unique noté B_j . Le facteur d'agitation thermique B_j est liée au déplacement quadratique moyen $\overline{U_j^2}$ de l'atome j autour de sa position d'équilibre par la relation : $B = \frac{8\pi^2}{3} \overline{U_j^2}$. Le facteur de Debye-Waller s'écrit alors : $w_j^{hkl} = e^{-B_j \frac{\sin^2 \theta}{\lambda^2}}$

Un modèle plus réaliste consiste à décrire chaque atome comme un oscillateur harmonique anisotrope. Le facteur de Debye-Waller est alors exprimé sous forme d'une matrice tenseur de déplacement atomique et comporte neuf paramètres ($\beta_j^{iv}, i = 1, 3; v = 1, 3$) par atome (seuls 6 paramètres sont indépendants). On a alors :

$$w_j^{hkl} = e^{-(H^T \beta_j H)} \quad (50)$$

où H représente la matrice 3 lignes et 1 colonne de coordonnées (h,k,l) et $H^T = [h, k, l]$ est la matrice 1 ligne 3 colonnes, transposée de H .

Néanmoins, pour un atome donné, l'utilisation de ce tenseur de déplacement atomique n'a de sens que si le nombre de données expérimentales disponibles pour déterminer et affiner ces paramètres sont en nombre suffisant, ce qui nécessite, dans le cas des cristaux de macromolécules biologiques, des limites de diffraction de l'ordre de 1Å-1,3Å. L'anisotropie des déplacements atomiques (et leurs corrélations) peut être néanmoins prise en compte dans le modèle simpliste d'un facteur d'agitation isotrope par l'utilisation de paramètres supplémentaires. Pour les structures à résolution moyenne (1,3Å- 3,0Å) l'approche la plus utilisée actuellement en bio-cristallographie pour décrire l'anisotropie des déplacements atomiques consiste en regrouper les atomes en domaines rigides, chaque domaine possédant des libertés de mouvements (affinement dit TLS pour "Translation, Libration, Screw-rotation"). Enfin, dans de nombreux cas, l'affinement des q_j (apparaissant sous forme d'un facteur multiplicatif dans l'expression 49 et des w_j (intervenant implicitement sous forme d'un facteur exponentiel) est fortement corrélé, ce qui explique que dans de nombreuses structures tridimensionnelles les q_j sont fixés à ($q_j = 1$) pour la grande majorité des atomes. Néanmoins, lorsque la limite de diffraction le permet et/ou que les cartes de densités électroniques montrent des conformations alternatives bien identifiées de groupes d'atomes, plusieurs conformations sont alors construites pour ces groupes d'atomes avec chacune un taux d'occupation approprié.

15. cette utilisation est impropre car elle laisse supposer que ce facteur ne prend en compte que le désordre dynamique.

Comme dit précédemment, l'affinement cristallographique est un processus d'optimisation non-linéaire pour lequel plusieurs méthodes ont été développées. Trois types d'algorithmes sont classiquement utilisés :

- les méthodes de moindres carrés et les méthodes de minimisation d'énergie associées.
- les algorithmes de dynamique moléculaire utilisant une minimisation par la méthode de recuit simulé.
- les méthodes de maximum de vraisemblance.

L'affinement cristallographique des macromolécules biologiques est une étape difficile principalement pour quatre raisons :

- le nombre de données expérimentales mesurées (modules des facteurs de structure) est souvent relativement faible par rapport au nombre de paramètres à optimiser. A des résolutions inférieures à 2,5 Å, le nombre d'observations dépasse rarement le nombre de paramètres.
- le modèle initial de la structure peut être de qualité médiocre en raison des erreurs des phases expérimentales ce qui se traduit par un désaccord entre la structure construite et la structure présente dans le cristal. Un modèle peut être imparfait (régions mal construites) et/ou incomplet (régions non encore construites car non interprétables).
- la fonction à minimiser possède plusieurs minima secondaires en plus du minimum absolu. L'obtention du modèle le mieux adapté aux données de diffraction va donc dépendre du rayon de convergence de la méthode de minimisation utilisée. Plus le modèle initial sera éloigné de la solution correcte, plus une méthode de minimisation à faible rayon de convergence aura tendance à aboutir au minima local le plus proche.
- les cartes de densités électroniques calculées avec des phases issues du modèle présentent des biais qui ont tendance à mettre en valeur la densité électronique du modèle introduit, même s'il est localement inexact, et à détruire la densité électronique de régions du modèle qui n'auraient pas encore été construites car non interprétables. Différents schémas de combinaisons de phases et d'amplitudes ont été introduits pour réduire ces biais.

Dans la pratique l'affinement cristallographique est un processus cyclique à deux étapes (voir figure 15) :

- une étape d'optimisation du modèle par calculs automatiques. On obtient alors de nouveaux paramètres pour décrire la structure cristalline.
- une étape de reconstruction manuelle où l'on interprète les cartes de densités électronique calculées à partir des nouveaux paramètres. Les nouvelles phases permettent de construire un nouveau modèle qui sera soumis au processus d'optimisation automatique.

Ces deux étapes sont très gourmandes en temps de travail humain et en temps de calculs. De nombreux efforts ont été réalisés ces dernières années par l'écriture de nouveaux algorithmes et par l'utilisation de calculateurs de plus en plus puissants. Le déroulement de l'affinement cristallographique est contrôlé par l'évolution d'un facteur de désaccord cristallographique R_{crist}

$$R_{crist} = \frac{\sum_{hkl} \left| \frac{|F_{obs}^{hkl}| - K' |F_{calc}^{hkl}|}{|F_{obs}^{hkl}|} \right|}{\sum_{hkl} |F_{obs}^{hkl}|} \quad (51)$$

Classiquement, un premier modèle peut avoir un facteur de désaccord cristallographique $R_{crist} \simeq 0.4-0.5$ en début du processus d'affinement et $R_{crist} \simeq 0.15-0.2$ en fin d'affinement. Malheureusement ce facteur R_{crist} n'est pas le reflet de l'exactitude du modèle car il est le fruit d'une procédure d'optimisation qui ne tend qu'à réduire le désaccord entre les modules des facteurs de structure observés et calculés. Un modèle avec de nombreuses erreurs, c'est-à-dire qui ne traduit pas la structure présente dans le cristal, peut avoir un très bon R_{crist} . Pour remédier à ceci, il faudrait une méthode d'optimisation qui prendrait explicitement en compte les erreurs sur les phases. L'approche actuelle consiste à évaluer la justesse d'un modèle par la valeur d'un facteur de désaccord cristallographique particulier

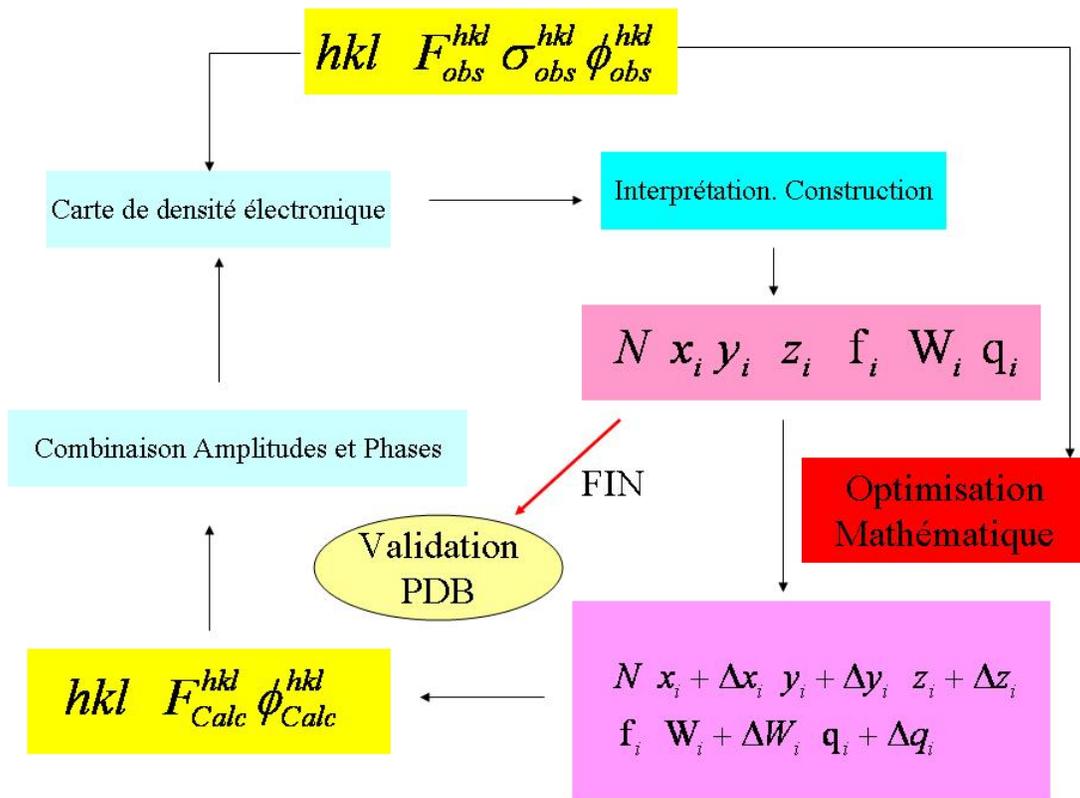


FIGURE 15 – Les différentes étapes de l’affinement d’une macromolécule biologique

appelé R_{libre} . Celui-ci est calculé à partir de réflexions qui, dès le début du processus d'affinement, sont mises de côté et ne sont pas utilisées dans l'optimisation du modèle. Typiquement, de 5 % à 10 %, des réflexions sont utilisées pour le calcul de R_{libre} . Il a été montré, par l'étude de structures tests, que la valeur de R_{libre} est fortement corrélée avec l'erreur commise sur les phases d'un modèle. C'est donc un bon critère pour suivre l'évolution d'un affinement. Une différence trop importante (≥ 0.10) entre R_{libre} et R_{crist} , traduit un processus dit de suraffinement où la baisse artificielle de R_{crist} ne correspond pas à une amélioration du modèle. L'introduction d'algorithmes d'optimisation fondés sur des méthodes de maximum de vraisemblance et sur une vision Bayésienne des distributions statistiques des facteurs de structure permettent d'éviter de tomber dans un certain nombre de problèmes présentés ci-dessous.

10.3 Les méthodes de moindres carrés

Ces techniques cherchent le minimum de la fonction

$$\psi = \sum_{h,k,l} \omega_{hkl} (|F_{obs}^{hkl}| - K' |F_{calc}^{hkl}|)^2 \quad (52)$$

où $|F_{obs}^{hkl}|$ et $|F_{calc}^{hkl}|$ sont respectivement le module du facteur de structure observé et calculé pour la réflexion d'indice hkl , K' un facteur d'échelle et ω_{hkl} est un poids affecté à chaque réflexion. Pour l'affinement des structures biologiques, on doit malheureusement faire face à un nombre d'observations insuffisants en raison de la limite de diffraction restreinte des cristaux (rarement mieux que 1,5 Å). Dans ces conditions, les méthodes de minimisation par moindres carrés ne sont pas stables. Pour augmenter le rapport nombre d'observations sur nombre de paramètres, deux méthodologies sont utilisées :

- augmenter le nombre d'observations. On incorpore dans la fonction à minimiser des informations stéréochimiques connues a priori : géométrie des acides aminés (longueurs et angles de valence), symétrie non cristallographique, corrélations entre atomes liés. On parlera alors d'affinement restreint. La fonction à minimiser comporte alors des termes interdisant à la structure de ne pas trop s'écarter d'une stéréochimie idéale. Les caractéristiques de cette stéréochimie sont définies à partir d'un ensemble de structures affinées à très haute résolution. Ces restrictions seront introduites soit sous formes de restrictions stéréochimiques simples, soit en les incorporant dans une énergie potentielle totale de la macromolécule que l'on cherchera à minimiser (voir paragraphe 10.4).
- diminuer le nombre de paramètres. On réduit le nombre de paramètres à affiner en contraignant certains paramètres à des valeurs fixes. On parle alors d'affinement avec contraintes. On peut ainsi fixer les longueurs de liaisons et les angles de valence et de n'affiner que les angles de torsion.

L'introduction de contraintes et(ou) de restrictions permet d'augmenter le rayon de convergence de la méthode d'optimisation et de diminuer le risque de suraffinement.

10.4 Les méthodes de la dynamique moléculaire

La dynamique moléculaire étudie le comportement dynamique d'un système d'atomes dans un champ de forces moléculaires. Dans un calcul de dynamique moléculaire, chaque atome de masse m_i et de coordonnées r_i est soumis à une force F_i donnée par la loi de Newton :

$$m_i \frac{\delta^2 r_i}{\delta t^2} = - \frac{\delta E_{TOT}}{\delta r_i} = \frac{F_i(r_i)}{m_i} \quad (53)$$

où E_{TOT} est l'énergie du système. Le système d'équations défini par la relation (53) peut être résolu numériquement à partir d'une géométrie de départ et de vitesses atomiques initiales.

La physique statistique relie le paramètre macroscopique qui définit l'agitation thermique moyenne à la température. On peut alors attribuer une vitesse initiale aléatoire à chaque atome à partir d'une

distribution de Maxwell à la température désirée. Connaissant les vitesses et les positions initiales de chaque atome, le système d'équations (53) peut être intégré sur un temps très court ($\simeq 1\text{-}2\text{fs}$, $1\text{fs}=10^{-15}\text{s}$) (algorithme de Verlet) et permet d'obtenir les déplacements à appliquer à chaque atome. On peut alors recommencer le calcul pour un deuxième pas d'intégration. Si le nombre de cycles est suffisant ($\simeq 10^3 - 10^4$), on peut alors atteindre le minimum absolu de l'énergie totale E_{TOT} .

L'idée de base de la dynamique moléculaire est de porter le système à une température initiale suffisamment haute pour franchir des barrières énergétiques et explorer une grande gamme de conformations. Le système est alors ramené lentement à température ambiante en diminuant progressivement l'énergie cinétique des atomes (technique dit de recuit simulé). Cela permet à l'affinement d'éviter de rester piégé dans de faux minima et éventuellement d'atteindre le minimum absolu de la fonction énergétique. Pour l'affinement des structures cristallographiques, on définit une fonction énergie totale de la macromolécule E_{TOT} , qui se décompose en un deux termes :

$$E_{TOT} = E_{POT} + wE_{crist} \quad (54)$$

où E_{POT} représente une fonction énergie potentielle de la macromolécule, E_{crist} une pseudo énergie cristallographique et w un facteur d'échelle. E_{POT} dépend de la géométrie induite par les interactions covalentes et des interactions non-covalentes (Van der Waals, Liaisons hydrogène, interactions électrostatiques). Elle permet d'introduire les contraintes et les restrictions introduites au paragraphe précédent. E_{crist} est défini par :

$$E_{crist} = \sum_{h,k,l} \omega_{hkl} (|F_{obs}^{hkl}| - K' |F_{calc}^{hkl}|)^2 \quad (55)$$

où ω_{hkl} est un poids attribué à chaque réflexion.

La méthode de recuit simulé permet une augmentation du rayon de convergence de l'affinement ce qui permet de réduire les étapes de reconstruction manuelle. En contrepartie, elle nécessite des moyens informatiques relativement puissants pour rester dans des temps de calculs raisonnables. Malheureusement, cette méthode augmente fortement le danger de suraffinement, car elle est capable de bouleverser fortement certaines régions de la structure sans que cela corresponde à un modèle plus exacte. C'est pour remédier à ce problème que la notion de R_{libre} a été introduite en cristallographie ; l'amélioration d'un modèle doit se traduire par une diminution conjointe de R_{crist} et de R_{libre} .

Il faut ici noter que pour des structures déterminées à partir de données de diffraction à très haute résolution, le terme E_{crist} seul suffit à gouverner l'affinement. C'est d'ailleurs en grande partie à partir des structures à très haute résolution que les paramètres stéréochimiques utilisés pour la définition de E_{POT} sont calculés.

10.5 Maximum de vraisemblance

Pour remédier aux difficultés de l'affinement d'une macromolécule biologique, il est nécessaire d'introduire dans la fonction cristallographique que l'on cherche à optimiser un terme qui permette de prendre en compte l'erreur sur les phases. La minimisation d'un terme cristallographique donné par la relation (52) conduit à des suraffinements car elle cherche à obtenir un accord entre modules de facteurs de structure et considère les phases comme constantes (c'est à dire sans erreurs). Ce type de minimisation est parfaitement justifié dans le cas de petites molécules où le rapport nombre d'observations sur nombre de paramètres est énorme, ce qui permet d'obtenir un traitement mathématique stable et une estimation correcte des erreurs.

Pour les macromolécules biologiques, il est important de prendre en compte les erreurs sur les phases dues soit à des modèles incomplets (certains atomes manquants) et/ou des modèles imparfaits (certains atomes mal positionnés). Les fondements mathématiques de ce travail ont été initiés il y a de nombreuses années par Gérard Bricogne et ont conduit à l'utilisation de logiciels appropriés par la communauté internationale. Ces méthodes font appels aux méthodes de maximum de vraisemblance et à l'application appropriée des statistiques Bayésiennes à la cristallographie des macromolécules biologiques. Le but de ces méthodes est de définir la probabilité $P(|F_{obs}| | F_{calc})$ de mesurer les modules

des facteurs de structure observés, étant donné le modèle et les estimations des différentes erreurs. Le meilleur modèle correspond au maximum de cette probabilité. Le traitement mathématique de cet aspect dépasse le cadre de cet article, le lecteur désirant approfondir la question pourra se référer à la bibliographie donnée en annexe. Dans la pratique, on minimise l'inverse de $\text{Log}(P(|F_{obs}| | F_{calc}))$, ce qui permet d'utiliser ce terme comme une énergie cristallographique :

$$E_{crist} = - \sum_{h,k,l} \text{Log}(P(|F_{obs}| | F_{calc})) \quad (56)$$

On peut ici remarquer que la détermination d'une structure par diffraction des rayons X sur des monocristaux comporte de nombreuses étapes où des optimisations de fonctions sont utilisées. Dans chaque cas, on peut à priori utiliser une fonction de score construite sur les méthodes de maximum de vraisemblance s'affranchissant ainsi de certaines simplifications qui provoquent des biais sur l'optimisation de la fonction étudiée. Lorsqu'elles ont été proposées, ces fonctions ont conduit à des améliorations majeures du processus d'optimisation concerné. C'est le cas des processus de phasage des réflexions (MIR, MAD) et de l'affinement cristallographique.

10.6 Carte de densité électronique

Après l'étape d'optimisation mathématique du modèle, on calcule de nouvelles cartes de densité électronique. Tout modèle est en général incomplet (certains atomes sont manquants) et/ou imparfait (certains atomes mal positionnés). La reconstruction du modèle s'effectue dans des cartes de densité électronique calculées avec des coefficients qui essaient de mettre en valeur ces défauts et de proposer des solutions nouvelles. Il est déconseillé d'utiliser comme coefficient les amplitudes $|F_{obs}|$ et les phases ϕ_{calc} (calculées à partir du modèle). En effet, la prédominance des phases dans les séries de Fourier donnerait une carte résultante identique au modèle introduit (on parle alors de biais). De nombreux travaux ont été réalisés dans ce domaine pour réduire ce biais. Deux types de cartes sont utilisées dans la pratique :

- des cartes dites de "Fourier différences" qui sont des cartes calculées avec des amplitudes du type $(u \ m \ | \ F_{obs} \ | \ -v \ D \ | \ F_{calc} \ |)$ et des phases ϕ_{calc} . Dans cette expression, u et v sont des entiers. m et D sont deux coefficients permettant d'estimer la fraction exacte de chaque facteur de structure F_{calc} . Ils sont calculés à partir d'une distribution conditionnelle de probabilité $P(|F_{obs}|, F_{calc})$ qui doit prendre en compte toutes les erreurs. Les coefficients les plus couramment utilisés pour le couple (u,v) sont (3,2) ou (2,1) ou (1,1). On parle alors de carte $(3F_{obs} - 2F_{calc}), (2F_{obs} - F_{calc}), (F_{obs} - F_{calc})$.
- des cartes de type "OMIT MAP". L'idée de base est d'enlever certaines zones du modèle pour le calcul des phases (et de l'affinement). Bien que ces atomes n'interviennent pas dans le calcul des cartes de densité électronique, on peut montrer qu'on devrait quand même les voir, à leur position correcte, grâce à l'information apportée par les modules des facteurs de structure $|F_{obs}|$.

La reconstruction s'effectue en général en travaillant sur plusieurs type de cartes simultanément. En effet, comme il est impossible de supprimer complètement le biais introduit par le modèle, toutes les cartes ne sont pas de même qualité et elles apportent des informations souvent complémentaires. On considère que l'affinement cristallographique est terminé lorsqu'il n'y a plus d'évolution de R_{libre} et que les cartes de densité électronique ne permettent plus de modifications du modèle.

10.7 Particularités des macromolécules biologiques

En fin d'affinement, on obtient classiquement un facteur de désaccord cristallographique R_{crist} de l'ordre de 0,15 – 0,20 avec un facteur de désaccord cristallographique libre R_{libre} de l'ordre de 0,15 – 0,30. C'est en général à ce niveau qu'on estime que le travail est acceptable pour publication et que les coordonnées sont déposées dans une base de données (voir paragraphe 12). Les facteurs R_{crist} et R_{libre} obtenus pour les macromolécules biologiques sont relativement élevés par rapport au standard

des petites molécules. Pour les cristaux de petites molécules (moins de 200 atomes), le facteur R_{crist} est de l'ordre de quelques % en fin d'affinement. Les valeurs élevées de R_{libre} sont dues à plusieurs facteurs spécifiques des cristaux de macromolécules biologiques qui sont peu ou mal pris en compte dans les paramètres décrivant la structure :

- le comportement dynamique des macromolécules biologiques. Elles présentent une flexibilité intrinsèque dans les cristaux, certaines zones pouvant par exemple simultanément adopter plusieurs conformations qu'il serait alors nécessaire d'introduire.
- la proportion importante de solvant (de 20% à 90%) présent dans les cristaux,
- la limite de diffraction en général modeste des cristaux (2Å-3Å).

Un cristal de macromolécule biologique est un état particulier de la matière qui contient une phase solide ordonnée et une phase liquide (état semi-solide, semi-liquide). En effet, dans les cristaux, les macromolécules sont entourées d'une couche de molécules d'eau ordonnées plongées dans un volume occupé par des molécules d'eau désordonnées. Ceci est très avantageux car l'environnement de la macromolécule dans le cristal est proche de l'environnement *in vivo*, et il est de plus possible de faire diffuser des substrats dans les canaux de solvant. Dans le cas d'enzymes, on a pu voir, sur de nombreux exemples, que la macromolécule conservait son activité biologique dans les cristaux. La présence de solvant est aussi très fortement mise à profit dans le processus d'amélioration des phases. Ses avantages se paient au niveau de la limite de diffraction souvent modeste des cristaux et au niveau de l'affinement cristallographique. En effet, on ne sait pas correctement modéliser ces molécules de solvant désordonnées. Elles sont le plus souvent approximées par un modèle de solvant avec une densité constante qui n'est satisfaisant qu'à basse résolution.

La prise en compte du désordre statique et dynamique des macromolécules biologiques introduit des paramètres supplémentaires dans l'affinement cristallographique. En général, ces paramètres ne peuvent pas être pris en compte en raison de la limite de diffraction des données. Il faut noter que cette flexibilité intrinsèque n'est pas éliminer par les techniques de cryocristallographie. Néanmoins, des données à très haute résolution (mieux que 1,2Å) ont pu être collectées pour plusieurs structures de macromolécules biologiques et ont permis d'obtenir des structures finales avec des valeurs de R_{crist} et de R_{libre} de l'ordre de 10%, ce qui se rapproche du standard des petites molécules.

De plus, il faut aussi noter que la fonction énergétique définie par la relation 54 peut posséder, pour une macromolécule biologique, non pas un minimum absolu, mais plusieurs minima d'énergies voisines.

11 Validation et contrôle qualité des structures

Un des dernières étapes dans le processus de détermination d'une structure 3D est une étape de validation et de contrôle qualité de la structure qui sera déposée à la base de données. En effet cette structure est le résultat de plusieurs étapes comportant des interprétations (qui peuvent être subjectives) des données et d'un protocole de détermination de structure (choix de programmes et paramètres utilisés par ces programmes). Une structure cristallographique peut donc comporter des erreurs (locales ou globales, correspondant à un modèle partiellement faux ou incomplet) qui peuvent résulter d'une qualité médiocre des données, d'une inadaptation des programmes utilisés ou peut être d'une relative inexpérience des scientifiques qui ont construit et affinés cette structure. Cette étape de validation et de contrôle qualité a pour but de minimiser ces erreurs et de fournir la structure tridimensionnelle qui correspond, dans l'état actuel des méthodes et des techniques utilisées, au meilleur accord avec les données expérimentales. La validation et le contrôle qualité d'une structure est réalisée non seulement avant le dépôt de la structure à la base de données PDB (voir paragraphe 12), mais aussi au cours du processus de construction et d'affinement pour améliorer la structure construite.

L'étape de validation et de contrôle qualité comporte 3 axes :

1. Validation et contrôle qualité des données cristallographiques utilisées pour déterminer la structure,
2. Validation et contrôle qualité de la structure, du point de vue physico-chimique,
3. Validation et contrôle qualité de la structure versus les données. La structure doit expliquer et refléter la(les) conformations de toutes les molécules présentes de cristal.

Les critères permettant un contrôle qualité d'une structure construite (ou en phase finale de construction) peuvent être schématiquement divisés en deux classes :

- ceux qui peuvent être évalués avec la connaissance de la structure 3D seule (critères stéréochimiques par exemple),
- ceux dont l'évaluation nécessite l'accès à la structure 3D et aux données cristallographiques (modules des facteurs de structures observés, phases utilisées, clichés de diffraction).

Il faut tout d'abord rappeler que la précision d'une structure 3D cristallographique est gouvernée par la limite de diffraction des données. Une structure 3D doit donc toujours être exploitée en gardant en mémoire la résolution à laquelle cette structure a été déterminée. Le fichier "PDB" de cette structure (voir paragraphe 12) contiendra toutes les informations techniques et méthodologiques permettant à un utilisateur averti de saisir la justesse et la précision de la structure utilisée.

A partir de la structure atomique, la conformité de la structure avec un certain nombre de critères de contrôle-qualité peut être évaluée. Il s'agit par exemple (liste non exhaustive) :

- de la conformité de la structure avec la chimie connue des acides aminés et des nucléotides (longueurs de liaison, angles de valence),
- la non-existence de mauvais contacts entre atomes,
- la présence non justifiée d'angles de torsion à priori défavorables (où peu fréquents) d'un point de vue énergétique. A ce titre le diagramme de Ramachandran d'une protéine est un outil précieux de validation.
- de paramètres de déplacements anisotropiques très différents pour des atomes liés,
- de la levée d'ambiguïté (dans les cartes de densité électronique) de l'orientation de certains groupements.

Pour cette étape de validation, de nombreux programmes et serveurs sont disponibles à ce jour. Des liens vers les plus connus d'entre eux sont donnés en bibliographie (paragraphe 15).

D'autres critères de contrôle-qualité nécessitent de disposer de la structure 3D et des données cristallographiques (pour chaque réflexion hkl , $|F_{hkl}^{obs}|$, ϕ_{hkl}^{calc}). Parmi ces critères, on peut citer par exemple :

- la limite de diffraction des données et leurs qualités : rapport signal sur bruit, complétude, redondance des mesures, qualité des données non-moyennées (les observations) (R_{sym} , R_{rim} ou R_{meas}), qualité des données moyennées (réflexions uniques) (R_{pim} , $CC_{\frac{1}{2}}$, CC^*)
- la conformité du modèle avec les données : R_{crist} , R_{libre} , comparaison de la densité expérimentale avec la densité calculée à partir du modèle,
- conformité globale des facteurs dit d'agitations thermiques,

La plupart de critères de contrôle-qualité figureront ou seront liés au fichier PDB de la structure et un utilisateur averti pourra y accéder. Pour une étude détaillée, le lecteur se rapportera à la bibliographie.

12 Base de données PDB

Les structures tridimensionnelles sont déposées dans une base de données spécialisée, appelée "Protein Data Bank" (PDB,RSCB), située actuellement aux Etats Unis (Rutgers, New Jersey). Cette base de données est d'accès libre et gratuit et permet à la communauté internationale d'avoir accès aux coordonnées des structures résolues. Le dépôt des coordonnées d'une macromolécule est devenue une condition préalable à la publication des travaux de recherche dans une revue internationale de haut niveau. Il faut se rappeler que les publications sont le critère premier de jugement du travail scientifique d'un chercheur du secteur public. Le dépôt de coordonnées est une pratique moins courante dans le monde industriel où au contraire on ne souhaite pas communiquer aux compétiteurs les données si précieuses que constituent les coordonnées atomiques d'une structure. A la date de remise à jour de cette revue, la "PDB" contenait plus de 109000 structures dont 90% de structures cristallographiques et un rythme de croissance d'environ 28 nouvelles structures par jour.

Chaque macromolécule est répertoriée dans la "PDB" par un nom de code arbitraire et peut être accédée via une interface graphique simple et performante. Chaque entrée¹⁶ de la "PDB" contient :

- les coordonnées atomiques des atomes et leurs paramètres de déplacements associés,
- des informations techniques simples et concises sur la méthode de détermination de la structure,
- un certain nombre de critères de validation,
- de nombreux liens vers d'autres bases de données spécialisées permettent de rassembler et d'organiser un maximum de connaissances biologiques, structurales et fonctionnelles sur la macromolécule d'intérêt.

A partir de la structure 3D connue, analysée et commentée, la base de données PDB cherche à être un lien privilégié pour accéder à de nombreuses autres bases de données biologiques pour permettre une corrélation entre la séquence, l'évolution de la séquence, la structure et la (ou les) fonctions d'une macromolécule biologique. Cette organisation de la connaissance qui nécessite une interconnexion dynamique entre plusieurs bases de données de natures et d'organisations différentes est un des défis actuels de la bio-informatique structurale. De nombreux outils de communication ont été développés ces dernières années et d'autres restent encore à inventer.

13 Détermination des structures à haut débit, les Années 2000.

L'une des révolutions des années 2000 a été le séquençage complet de plusieurs centaines de génomes ce qui a produit une quantité importante de séquences de gènes dont la fonction est inconnue, c'est à dire non prédictible à partir de la connaissance de la séquence seule. Bien que cet aspect n'ait pas été développé dans cette revue, la connaissance de la structure tridimensionnelle d'une macromolécule peut apporter des informations cruciales sur la (ou les) fonction(s) possible(s). Pour un gène donné, si on connaît la structure de la macromolécule correspondante, on pourra alors, en comparaison à la base de données actuelle, assigner une ou plusieurs fonctions à ce gène. Depuis plusieurs années, des efforts importants ont donc été réalisés à l'échelle mondiale pour déterminer des structures 3D à grande échelle. On peut pour simplifier expliquer ce recours massif à la cristallographie des rayons X par la volonté de relever plusieurs défis :

1. Comprendre et résoudre "le problème du repliement". On peut considérer deux objectifs différents :
 - trouvez le dictionnaire complet des repliements utilisés par la nature, car on sait que ce répertoire est un ensemble fini.
 - construire une base de données de connaissances pour essayer de prévoir le ou les repliements susceptibles d'être adoptés par une macromolécule donnée.

16. Pour le format utilisé dans ce fichier, voir la documentation à <http://www ww pdb.org/docs.html>

2. Comprendre l'évolution en terme structural. Bien que cette partie n'ait pas été abordé dans cet article, les structures 3D sont des outils précieux pour comprendre l'évolution des fonctions. En effet, dans les cas difficiles (et donc les plus intéressants pour une activité de recherche fondamentale), des similarités de fonctions ne sont souvent découvertes qu'à partir de similarités locales ou globales de structures 3D.
3. Assigner une fonction à chaque protéine d'un protéome donné.
4. Apporter des réponses à des problèmes de santé humaine. Il s'agit en particulier d'identifier :
 - de nouvelles cibles pharmacologiques et de déterminer leurs structures 3D,
 - de nouveaux composés à actions pharmacologiques. Dans le processus de conception rationnelle de médicaments, la possibilité de résoudre très rapidement une structure 3D permet à la cristallographie des rayons X d'être utilisée non seulement pour améliorer un premier composé prometteur mais surtout pour découvrir de nouveaux composés à action pharmacologique.

Ce recours massif à la cristallographie, commencé au début des années 2000, a nécessité et nécessite la mise en place d'outils de plus en plus automatisés pour toutes les étapes d'un projet structural. Plusieurs centres de génomique structurale en France et dans le monde ont été créés. Le choix des cibles étudiées et l'état actuel des travaux sont accessibles en partie par exemple via la base de données "The Structural Biology Knowledgebase¹⁷". Ces programmes associent le plus souvent plusieurs laboratoires¹⁸) dans le but d'améliorer et de développer les méthodes et les techniques utilisées pour les études structurales à haut débit. Comme on l'a déjà noté dans les paragraphes précédents, les progrès technologiques et méthodologiques réalisés dans ce domaine ont fortement modifié la manière de travailler en proposant des protocoles de plus en plus automatisés et en conséquence facilitent l'étude des systèmes biologiques complexes. Du point de vue appliqué, de nombreuses petites entreprises ont vu le jour. On peut citer par exemple les développements faits par la société "astex therapeutics" (Cambridge, Angleterre, <http://www.astex-therapeutics.com/>) qui a développé le concept de "fragment-based drug discovery" illustrant ainsi le rôle crucial de la cristallographie dans le processus pharmacologique.

De nombreux programmes, intègrent l'ensemble du processus cristallographie à l'aide d'interfaces graphiques, permettant la résolution d'une structure cristallographique avec une intervention humaine le plus souvent réduite à quelques points cruciaux de décision. Une liste de ces programmes est donné en bibliographie 15.

17. <http://sbkb.org/>

18. voir le programme européen INSTRUCT, <https://www.structuralbiology.eu/>

14 Perspectives

La cristallographie des macromolécules biologiques occupe une place centrale dans la biologie moderne. La structure tridimensionnelle d'une macromolécule permet d'apporter un éclairage neuf et original qui est essentiel pour comprendre la fonction biologique. Au cours de la décennie précédente, des avancées technologiques majeures ont été réalisées qui permettent, dans les cas simples, de résoudre rapidement et avec un minimum d'intervention humaine une structure 3D. On entend ici par cas simple, des macromolécules qui, dans l'état actuel des connaissances méthodologiques et technologiques, peuvent être facilement purifiées et cristallisées, présentant un pouvoir de diffraction suffisant et pour lesquelles le problème des phases peut être résolu de manière classique (en particulier par la méthode MAD/SAD). Il est bon de rappeler que les cas simples d'aujourd'hui étaient assez souvent les défis et prouesses technologiques d'hier. On détermine aujourd'hui en routine des structures 3D à partir d'un nombre très limité de cristaux et souvent de très petite tailles. Toutes ces avancées permettent aux structuralistes d'attaquer des problèmes de plus en plus complexes. La purification de la (ou des) macromolécule(s) et l'obtention de cristaux de qualité (limite de diffraction meilleure que 3\AA , c'est à dire $d_{min} \leq 3\text{\AA}$) restent toujours deux étapes limitantes d'un projet structural. De plus, la résolution du problème des phases est encore dans certains cas un processus long et fastidieux, en particulier pour les gros édifices macromoléculaires. Pour l'interprétation des cartes de densité électronique à résolution moyenne (3\AA - $3,5\text{\AA}$), l'expertise et l'intelligence humaine peut faire encore la différence avec les algorithmes de construction automatique. Les challenges d'aujourd'hui se déclinent en deux axes principaux :

- la détermination de structures à haut débit.

Il s'agit ici de poursuivre les efforts méthodologiques et technologiques présentés au paragraphe 13 afin rendre le processus cristallographique encore plus rapide et plus automatique. On aimerait aussi créer une base de connaissance permettant de sur-exprimer, purifier en grande quantité et cristalliser n'importe quelle macromolécule ou système macromoléculaire.

- l'étude des "macromolécules complexes".

Cette complexité peut d'abord résulter d'assemblages de protéines et/ou d'acides nucléiques en édifices ou entités fonctionnelles de haut poids moléculaires (généralement sans symétrie interne) et qui sont le centre de fonctions biologiques primordiales. Ces complexes sont souvent composés de nombreuses molécules, certaines de ces molécules comportant des domaines structurés et des régions ou des domaines non structurés ne s'organisant qu'en présence du ou des bons partenaires. Certains de ces complexes fondamentaux pour la vie cellulaire ont une courte de vie courte ("transient molecular complexes") et sont donc difficiles à isoler. La sur-expression et la purification de ces complexes (stables ou "transients") dans des quantités compatibles avec une étude structurale est l'un des défis majeurs des 10 prochaines années. De plus, l'étude de la dynamique de ces assemblages (évolution dans le temps et dans l'espace) constituera un autre degré de complexité qui sera éventuellement abordable par cristallographie. On peut aussi appeler molécules complexes, les macromolécules non-solubles (cas des protéines membranaires par exemple), les macromolécules (protéines, acides nucléiques) comportant une grande dynamique interne et les protéines intrinsèquement non structurées. Ces macromolécules sont essentielles pour la vie cellulaire.

L'étude des molécules complexes nécessitent un investissement lourd et coûteux (en hommes et en matériels) sur une échelle de temps relativement longue. De nombreuses avancées méthodologiques et techniques sont encore à développer pour relever ce défi.

Parallèlement à la cristallographie biologique, d'autres techniques d'imagerie ont connu et connaissent un essor important et permettent d'avoir une vision moins précise mais néanmoins importante des systèmes biologiques.

Plusieurs révolutions des années 2010 bousculent et questionnent la cristallographie biologique dans 2 axes :

- le développement des lasers à électrons libres dans la gamme des rayons X (XFEL, X Ray Free Electron Laser) qui ouvrent des champs extraordinaires pour la cristallographie biologique et l'étude de molécules isolées.
- les progrès méthodologiques et technologiques de la cryo-microscopie qui dans certains domaines atteint aujourd'hui des gammes de résolution que l'on pensait réserver à la cristallographie biologique.

Le lecteur trouvera plusieurs articles en bibliographie [15](#) pour des compléments sur ces aspects. L'utilisation conjointe de toutes ces méthodes et techniques, associées aux développements de la biologie moléculaire et cellulaire permettra peut être aux biologistes de passer le cap d'une vision réductionniste des phénomènes à une vision cellulaire intégrée à l'échelle atomique (de la connaissance des mots à la compréhension de la littérature biologique) .

15 Bibliographie

Les articles ci-dessous développent, illustrent et complètent, certains aspects présentés dans cette revue :

- Laskowski R A, Thornton J M, "Understanding the molecular machinery of genetics through 3D structures", Nat Rev Genet. 2008 Feb ;9(2) :141-51
- Alber F, Orster F, Korkin D, Topf M, Sali A. "Integrating Diverse Data for Structure Determination of Macromolecular Assemblies" Annu. Rev. Biochem. 2008, 77 :443-77.
- Chruszcz M, Wlodawer A, Minor W "Determination of Protein Structures. A Series of Fortunate Events", Biophysical Journal, Volume 95 July 2008, 1-9.
- Wlodawer A , Minor W, Dauter Z and Jaskolski M "Protein crystallography for non-crystallographers", FEBS Journal 2008, Jan ;275(1) :1-21.
- Minor DL Jr, "The Neurobiologist's Guide to Structural Biology : A Primer on Why Macromolecular Structure Matters and How to Evaluate Structural Data", Neuron 54, May 24, 2007, 511-533.
- Schmidt A, Lamzin S, "From atoms to proteins", Cell. Mol. Life Sci. 64 (2007) 1959-1969.
- Sua XD, Zhang H, Terwilliger TC, Liljas A, Xiao J and Dong Y. "Protein crystallography from the perspective of technology developments", Crystallography Reviews, 2015, Vol.21, Nos 1-2, 122-153.

Pour une étude approfondie, on pourra consulter les revues de synthèse ci-dessous :

- Acta Cryst D73, February 2017." Protein-ligand complexes : understanding biological chemistry".
- Acta Cryst D72, March 2016." Advances in experimental phasing".
- Acta Cryst D71, January 2015."Crystallography and complementary methods".
- Acta Cryst D69, July 2013. "Data collection and processing".
- Acta Cryst D68, April 2012."Model building, refinement and validation",
- Acta Cryst D67, Part 4, April 2011. "From crystal to structure with CCP4"
- Acta Cryst D66, Part 4, April 2010. "Experimental phasing and radiation damage"
- Acta Cryst D65, Part 2, February 2009."Low-resolution structure determination and validation"
- Acta Cryst D64, Part 1, January 2008. "Molecular replacement".
- Acta Cryst D63, Part 1, January 2007. "Crystallography of complexes".
- Acta Cryst D62, Part 10, October 2006. "Structural Proteomics IN Europe".

Quelques liens "Internet" :

- Base de données PDB (RCSB). Toutes les structures publiées : <http://www.rcsb.org/pdb/>
- [Une liste d'articles classés par rubriques](#)
- [Une sélection de programmes](#)