

Étude historique de la visualisation des protéines

De la représentation ancienne de « noyau organique azoté » du XIX^e siècle à la vision statique en « boules-tiges de fer » des années cinquante aux méthodes actuelles de modélisation en dynamique moléculaire

Jean-Frédéric Bruch^{a,*}, Loïc Metairy^b, Myriam El Gani^a, Arthur Pearson^b, Thibault Kervarrec^c, Flore Tabareau-Delalande^d, Junior Samuel López Yépez^e

RÉSUMÉ

L'histoire de la visualisation des protéines est inséparable de celle de la biologie structurale qui étudie la structure et l'organisation spatiale des macromolécules biologiques. Les anciens avaient initialement imaginé les protéines comme des noyaux organiques azotés entourés d'une copule minérale. La détermination à l'échelle atomique de la structure 3D des protéines fait appel à des techniques d'approche biophysiques élaborées dans la première moitié du vingtième siècle. La cristallographie par diffraction des rayons X permet d'obtenir une carte de densité électronique. La spectroscopie par résonance magnétique nucléaire (RMN) définit une carte des distances inter-protons et des angles de torsion du squelette de la protéine. La cryomicroscopie électronique donne une image directe de la molécule saisie dans son milieu aqueux. Ces trois familles de techniques évoluent de manière spectaculaire pour grandir en précision, en définition et en analyse spatiale et temporelle. Aujourd'hui ces méthodes convergent grâce à des bases de données interactives et des outils bioinformatiques de modélisation dynamique vers la définition de modèles de liaisons « protéines-ligands », la découverte d'inhibiteurs pharmacologiques, la prédiction de structure macromoléculaire et une compréhension accrue des maladies du repliement anormal des protéines.

Bases de données - cryomicroscopie électronique - diffraction des rayons X - dynamique moléculaire - magnétique nucléaire - maladies du repliement - modélisation moléculaire - repliement des protéines - spectroscopie par résonance.

a Service de Pathologie

Centre Hospitalier Louis Pasteur
4 rue Claude Bernard 28630 Le Coudray

b Service d'Imagerie,

Centre Hospitalier de Blois, 41016, Blois Cedex

c Service d'Anatomie-cytologie Pathologique

CHRU de Tours 37044 Tours cedex 9

d Service de Pathologie

Centre Hospitalier régional d'Orléans
1 Porte Madeleine, 45032 Orléans, France

e Department of Chemistry

Langelandsgade 140 8000 Aarhus C Denmark

* Correspondance

jfbruch@ch-chartres.fr

article reçu le 15 octobre 2016, accepté le 31 octobre 2016.

© 2016 – Elsevier Masson SAS – Tous droits réservés.

SUMMARY

Historic study of the display of proteins

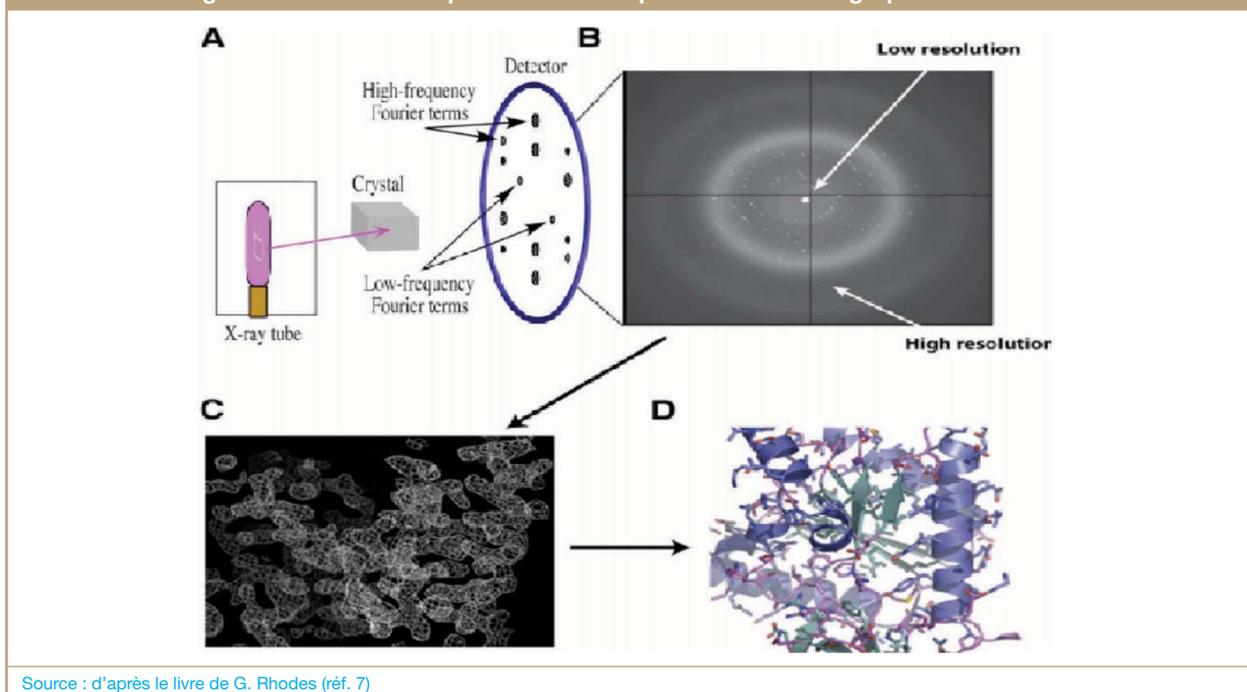
The history of the visualization of proteins is inseparable from that of structural biology that studies the structure and spatial organization of biological macromolecules. The ancients had initially imagined proteins as organic nitrogen nuclei surrounded by a mineral copula. The determination at the atomic scale of the 3D structure of proteins uses biophysical approach techniques developed in the first half of the twentieth century. Crystallography by X-ray diffraction provides an electron density map. Nuclear magnetic resonance spectroscopy defines a map inter-proton distances and angles of torsion of the backbone of the protein. Electron cryomicroscopy gives a direct image of the input molecule in its aqueous medium. These three families of techniques are evolving dramatically to grow in accuracy, resolution and spatial and temporal analysis. Today these methods converge with interactive databases and dynamic modeling bioinformatics tools towards defining models of «protein ligands» bounds, the discovery of pharmacological inhibitors, the prediction of macromolecular structure and an increased understanding of folding diseases abnormal proteins.

X ray diffraction - Cryoelectron microscopy - NMR - Protein folding

1. Introduction

Retracer l'histoire de la visualisation moléculaire des protéines nécessite de nous repositionner dans le temps en fonction des différents événements qui ont jalonné l'histoire des sciences de la chimie, de la physique et de la biologie des protéines. « Voir » n'est pas toujours « comprendre » et un cadre de pensée est nécessaire à la compréhension éclairée des images moléculaires, qu'elles soient destinées à la recherche, à l'enseignement ou à la vulgarisation.

Figure 1 – schéma récapitulatif de la séquence de cristallographie moléculaire.



Source : d'après le livre de G. Rhodes (réf. 7)

Nous étudierons différentes périodes pour en extraire les principaux paradigmes et suivre pas à pas l'évolution des idées, des découvertes et des faits en fonction du contexte de chaque époque.

2. Courant dix-neuvième siècle : les premières notions

La notion de protéine est apparue en 1838 sous l'impulsion du chimiste néerlandais Gerhardus Johannes Mulder pour désigner un ensemble de substances végétales ou animales douées de propriétés nutritives et dont l'analyse élémentaire montrait à peu près la même formule empirique en carbone, hydrogène, azote et oxygène selon une règle de proportionnalité. (400/620/100/120). Ce noyau organique serait entouré d'une « copule » minérale formée d'atomes de soufre et de phosphore individuels [1]. Une copule est un composé imaginaire dont la présence déguise toutes les propriétés chimiques des corps auxquels il est uni. Durant le XIX^e siècle une douzaine d'acides aminés seront isolés et identifiés notamment par Vauquelin et Robiquet, Liebig puis caractérisés structuralement au siècle suivant. En 1894 un chimiste allemand Emil Fischer propose un paradigme d'interaction entre les protéines enzymatiques et les substances dont elles catalysent la formation selon un modèle clé (molécule transformée) et serrure (enzyme). Fischer parvient à établir le type de liaison chimique qui connecte les acides aminés entre eux pour former des chaînes : la liaison peptidique ce qui lui vaudra l'obtention du prix Nobel de chimie en 1902.

La vision des protéines dans les années vingt était celle d'une sphère, pour les protéines dites globulaires et des filaments pour les protéines fibreuses[2].

3. Années 1890-1950 : les découvertes essentielles en physique qui vont poser les bases de l'étude des structures macromoléculaires

3.1. Années 1890-1950 : cristallographie rayons X

En 1895 Wilhelm Roentgen découvre un soir d'hiver, en actionnant un tube cathodique enveloppé dans un carton rigide que de mystérieux rayons invisibles traversent le carton et viennent impressionner un écran. [3]. Cette découverte des rayons dénommés « X » vaut à son auteur l'attribution du premier prix Nobel de physique en 1901. Parallèlement deux chimistes réputés Franz Hofmeister et Emil Fischer proposent en 1897 indépendamment et simultanément que les protéines sont des polymères linéaires d'acides aminés liés par des liaisons peptidiques [2]. En 1912 X Max von Laue découvre que les rayons X sont diffractés par des cristaux et suspecte que la longueur d'onde du rayon X pourrait être comparable à la distance inter-atomique. William Bragg et son fils Willie établissent en 1913-1915 les bases physiques du phénomène de diffraction des rayons X, par analogie avec la diffraction de la lumière dans un réseau de diffraction de microscope optique, en notant que les feuilles d'atomes jouent le rôle de plans de réflexion pour les rayons X.

Willie Bragg formule en 1913 une loi mathématique définissant les conditions nécessaires pour produire une diffraction effective de rayons X à travers un cristal moléculaire dont les feuillettes sont séparés d'une distance « d ». La formule $\{2 \sin \theta \times d = l \times n\}$ exprime la condition

pour laquelle un maximum d'interférences survient arrivant par sommations successive à produire des rayons suffisamment intenses pour produire une tâche visible sur un écran. Cette formule indique que deux rayons incidents en phase, touchant chacun un plan du cristal, suivant une orientation appropriée d'angle « alpha » suivent un chemin dont la différence de longueur est égale au produit de la longueur d'onde du rayon X par un nombre entier. Laue et les deux Braggs (père et fils) seront nobélisés respectivement en 1914 et 1915 [3].

Le cristal diffracte la source de rayons X en une multitude de rayons produisant chacun une tâche sur le film radiographique. Les ondes diffractées sont la transformée de Fourier des objets moléculaires agencés dans l'unité cellulaire du cristal. La position et l'intensité des tâches de diffraction contiennent les informations nécessaires pour déterminer la structure moléculaire. L'intensité de la tâche est fonction de l'amplitude du rayon diffracté. La position de la tâche permet de déduire la direction du rayon dans un système de coordonnées 3D des plans de diffraction (h,k,l) de l'espace dit « réciproque ». L'image moléculaire est redessinée dans l'espace « réel » à partir d'une carte de contours de densité électronique (coordonnées x,y,z) obtenue à partir de la transformée de Fourier inverse de chaque rayon diffracté (figure 1).

En 1926 le chimiste James Summer découvre la cristallisation du premier enzyme : l'uréase (prix Nobel de chimie en 1946). Dorothy Crowfoot Hodgkin travaillait avec J. Desmond Bernal (dénommé « le Sage ») sur la cristallisation des protéines et publia en 1934 les premières images de diffraction protéique à partir de cristaux de pepsine.

La mathématicienne britannique Dorothy Wrinch formule en 1936 une théorie imaginaire dite des « cyclols » selon laquelle la chaîne d'acide aminé se replie selon une figure géométrique cyclique reproductible quasiment à l'infini. Cette théorie fut rapidement invalidée par Linus Pauling en tant que contredisant les lois élémentaires de la chimie. Dorothy Crowfoot Hodgkin poursuit ses travaux de diffraction moléculaire (cholestérol, pénicilline) sans pouvoir élucider la structure cristallographique des protéines.

Le chaînon manquant était que les structures ne pouvaient pas être résolues avec précision en raison d'un problème de phase inconnue ne permettant pas de recalculer l'ensemble des ondes diffractées au sein de chaque unité cellulaire [4-5-6-7].

3.2. RMN : années 1890-1950

Le physicien Wolfgang Pauli décrit dès 1924 une propriété magnétique du noyau atomique : en tournant sur lui-même le noyau chargé positivement crée un champ magnétique selon la loi de Faraday appliquée à un atome : ce phénomène caractérise le spin nucléaire magnétique. Plus tard le physicien américain Isidore Isaac Rabi montre en 1938 un premier exemple de résonance magnétique nucléaire (RMN).

La RMN va prendre ses lettres de noblesse à la fin de la deuxième guerre mondiale, les américains F. Bloch et E. Purcell décrivent indépendamment chacun dans son laboratoire, le phénomène de résonance et de relaxation. Ils obtiendront le prix Nobel de physique en 1952. [8]

4. Les années cinquante-soixante : les bâtisseurs & les fondations

4.1. Cristallographie rayons X: Années 50-60

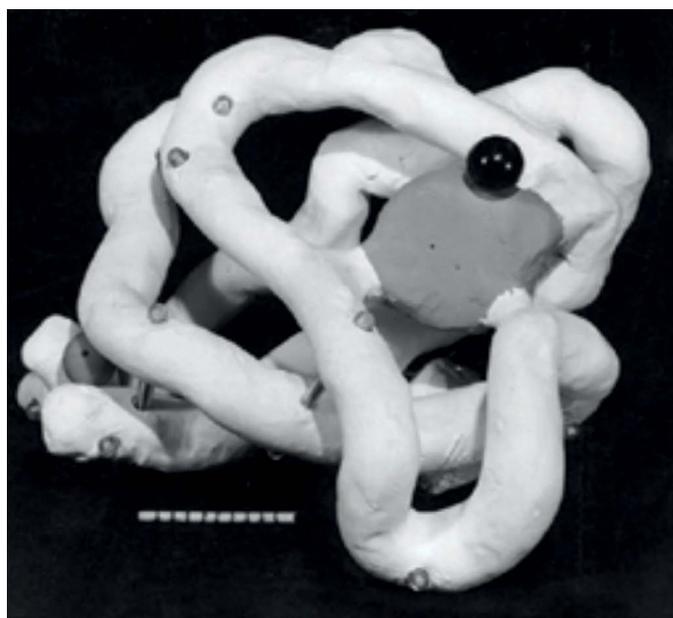
C'est aux USA que les bases de la structure secondaire des protéines seront posées grâce aux travaux de Linus Pauling. Alors que Pauling n'avait jamais résolu de structure, il déduisit de manière théorique, en 1951, la structure de l'hélice alpha à partir de la géométrie des liaisons chimiques reliant entre eux les acides aminés des protéines. En 1953 il décrit la structure des feuillets Béta. Il suivit la prédiction d'un de ses maîtres qui affirmait que toute structure moléculaire ne pouvait s'échafauder qu'en « avançant » « et en « tournant » à la manière d'une hélice [9]. Sa connaissance fine de la chimie lui permit de prédire de façon presque exacte le pas et le rayon de l'hélice.

C'est en Angleterre que seront déterminées, grâce à la solubilisation du problème de phase par la technique du remplacement isomorphe consistant à substituer un ou plusieurs atomes par un atome lourd, les premières structures cristallographiques de protéines par les chercheurs John Kendrew et Max Perutz (originaire d'Autriche et immigré en 1936). Ces derniers publient respectivement la structure de la myoglobine en 1958 et l'hémoglobine en 1960 [5-10-11] (figure 2 a).

Anfinsen propose en 1957 un premier modèle de pliage moléculaire expliquant la structure tridimensionnelle des protéines [12].

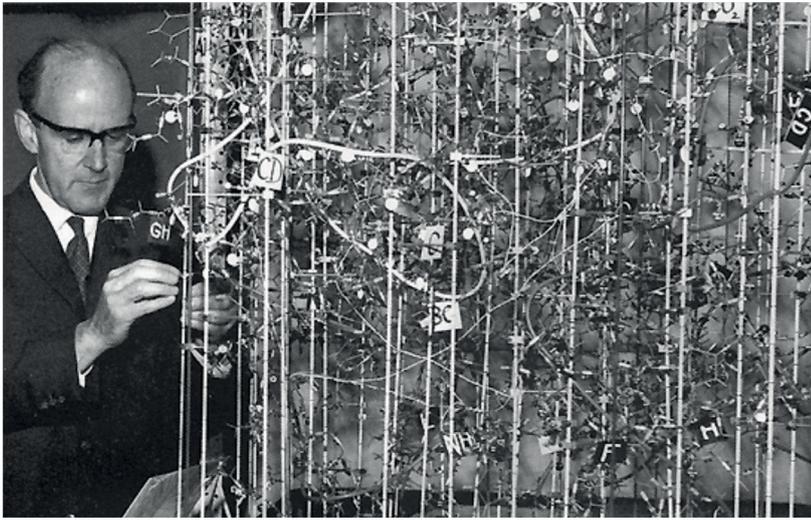
En 1962, dans un laboratoire à Cambridge (USA), Michael Rossmann décrit une variante du remplacement isomorphe appelée « remplacement moléculaire », permettant aux chercheurs d'utiliser une structure existante pour calculer l'orientation et la position d'une structure apparentée, mais inconnue.

Figure 2 a – Image en « pâte à modeler » de la myoglobine.



Source: Kendrew, 1958

Figure 2 b – Image en « boules-fils de fer » de l'hémoglobine.



Source: Perutz, 1960

En 1964 Aaron Klug, travaillant à Cambridge (Angleterre), applique les principes de diffraction des rayons X à la microscopie électronique qui devient « cristallographique ». Cette découverte lui vaudra le prix Nobel de chimie en 1982. En 1965 David C Phillips (Londres) résout la structure du lysozyme et en 1967 les premiers cristallographes américains résolvent plusieurs enzymes ainsi qu'en 1968 la structure des immunoglobulines. [13]

4.2. RMN : années 1950-60

Sur le plan de la RMN trois découvertes essentielles voient le jour : le déplacement chimique, l'effet Overhauser et le couplage J. Le déplacement chimique a été découvert en 1950 par WG Proctor et FC Yu à Stanford University (USA) et WC Dickinson au Massachusetts Institute of Technology. Le phénomène de déplacement chimique est une propriété de modification de la fréquence de résonance du noyau en fonction de son environnement électronique et de sa position dans la molécule. Le physicien américain Albert Overhauser prédit en 1953 que la polarisation de spin nucléaire peut être transférée d'une population de spin à une autre en fonction de la distance interatomique. Le couplage J est une interaction entre plusieurs spins à travers les liaisons chimiques [14]. Les premiers spectromètres RMN à électro-aimant sont commercialisés en 1955. Jardetski publie le premier spectre RMN d'une protéine (la ribonucléase) avec un spectromètre de 40 MHz [14].

Les années 60 furent marquées par l'introduction en 1962 du premier spectromètre doté d'aimant supraconducteur à 200 MHz. En 1964-65 Ernst et Anderson vont mettre au point la séquence de transformée de Fourier pulsée. Ces recherches vont conduire à la conception d'une nouvelle machine intégrant directement la transformée de Fourier par l'entreprise Bruker.

4.3. La cryo-EM : années 50-60

En 1968 David de Rosier et Aaron Klug [15] présentent une technique de reconstruction 3D de microscopie électronique d'une queue hélicoïdale de virus bactériophage T4. Ils uti-

lisent les principes de Fourier et les fonctions de Bessel pour combiner de multiples vues 2D des molécules naturellement présentes dans la structure symétrique du virus. Les trois principales difficultés sont : la nécessité de mettre l'échantillon sous vide et de le dessécher (ce qui ne convient pas à l'examen des macromolécules dont le milieu naturel est hydraté), les dommages causés par le rayonnement et le très faible contraste spontané. Rosier et Klug introduisent la coloration par les sels de métaux lourds (acétate d'uranyle) qui génèrent une image plus contrastée et réduisent les dégâts d'irradiation.

4.4. Modélisation : années 50-60

Dans les années 60, les cristallographes, les biologistes informaticiens et les chimistes s'intéressent aux méthodes les plus appropriées de visualisation, mémorisation et communications des données moléculaires. Cependant la méthodologie pour contenir les données n'était pas bien définie. Dans les années avant l'internet, il

était nécessaire d'envoyer par voie postale les informations sous forme de boîtes de cartes perforées ou des bandes magnétiques. [16]

Le premier modèle moléculaire fut bâti par Kendrew en 1957 pour la myoglobine en pâte à modeler à une échelle de 5 cm/Å et un plus tard il construisit un modèle géant désigné « forest of rods ». Il fut construit avec 2 500 barres métalliques disposées dans un cube de 2 mètres de côté. Des petites boules colorées étaient attachées aux tiges pour représenter les atomes. Cette forêt de tiges métalliques obscurcissait la vision globale de la molécule. [16-17] (figure 2 b).

En 1968 Fred Richards introduit un comparateur optique permettant de fabriquer plus rapidement les modèles en laiton de style Kendrew. Les densités électroniques résultant d'études cristallographiques étaient imprimées par ordinateur sur papier. Un système de miroir permettait de superposer la carte de densité électronique sur le modèle laiton. Ce procédé de visualisation fut dénommé la Richard's Box. [16-17] Parallèlement, apparaissent les premiers programmes informatiques pour la visualisation des protéines assistées par ordinateur. Levinthal utilise un oscilloscope commandé par ordinateur pour afficher l'image en rotation d'une structure protéique.

5. Les années 1970-80 : résolution par diffraction RX et RMN des petites protéines solubles et les débuts de la cryomicroscopie électronique

5.1. Années 70-80 : cristallographie rayons X

Pendant les années 70 la procédure de résolution moléculaire par cristallographie rayons X nécessite globalement 5 à 7 ans, commençant par une année consacrée à la purifi-

cation et la cristallisation, puis recueillir des centaines de photographies de diffractions de dizaines de cristaux monts dans les capillaires à température ambiante, numériser les photographies dans un scanner, résoudre les Patterson à la main et calculer les phases avec de programmes faits maison puis construire le modèle style Richard's box. Dans les années 80, La technique de cristallisation progresse avec une plus grande quantité de protéines produites à partir de gènes clonés et une collecte plus rapide avec les détecteurs de zones multifils et l'utilisation du rayonnement synchrotron. Le temps total de résolution est d'environ 2 à 5 ans [4]. En 1985 Deisenhofer J. et son groupe mettent au point la résolution d'une grosse protéine membranaire: un chromophore du centre de réaction photosynthétique d'une bactérie pourpre [13].

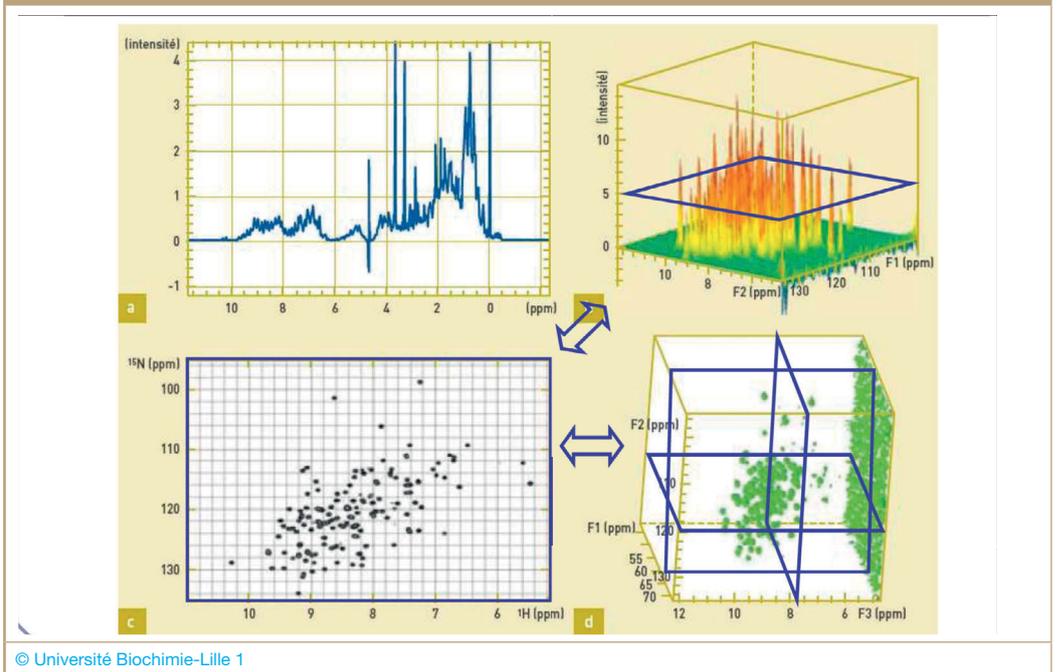
5.2. Années 70-80: la RMN

Les années 70 voient progresser la puissance des machines qui passe de 270 MHz en 1971 à 500 en 1979 par la compagnie Bruker. En 1975 l'équipe suisse menée par Richard Ernst qui sera nobélisé en 1991 concrétise l'idée du chercheur belge Jean Jeener d'une double transformée de Fourier avec spectroscopie bidimensionnelle.[18] Le principe général est que les pics de résonance des protons individualisés par le déplacement chimique peuvent secondairement être mis en relation par des phénomènes de couplage direct ou indirect avec transfert d'aimantation d'un proton à un autre. Deux grands types de séquence vont être définis: la séquence «COSY» (Correlated Spectroscopy) qui spécifie les couplages indirects et «NOESY» (Nuclear Overhauser Effect Spectroscopy) qui spécifie les couplages directs. Le NOESY est un effet «distance» variant dans l'espace (contrainte de proximité maximale de 5 angströms) alors que COSY est un effet «liaison covalente» entre atomes.

Les années 80 de la RMN: Kurt Wuthrich et ses collaborateurs mettent au point à partir des expériences 2D une technique d'attribution spectrale séquentielle des acides aminés permettant de résoudre par RMN la structure de la plupart des petites protéines de poids inférieur à 30 kd. Le principe se résume en deux étapes:

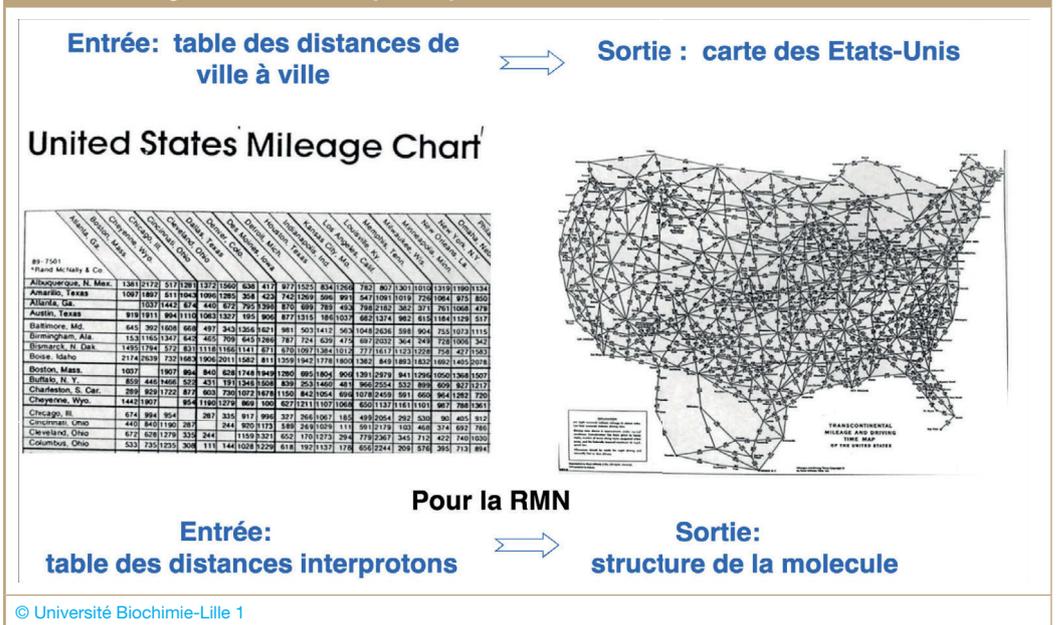
- Identification du type d'acide aminé par transfert à l'intérieur d'un même résidu (interaction scalaire, couplage J: séquence COSY)
- Position dans la séquence par transfert au résidu suivant ou précédent (interaction dipolaire, distance inter proton: séquence NOESY) (figure 3).

Figure 3 a – RMN 1D, 2D, 3D, 4D.



© Université Biochimie-Lille 1

Figure 3b – Des couples dipolaires à la structure tridimensionnelle.



Des calculs de transfert à plus longue distance par interaction dipolaire permettent, par des méthodes itératives de reconstituer la structure 3D [19]. Une métaphore permet de mieux comprendre comment la table des distances interprotoniques en entrée peut générer en sortie la structure de la molécule : c'est comme si en entrée nous avions la table des distances de ville à ville et en sortie la carte des états unis (*figure 3b*) [20].

Plusieurs publications voient le jour et définissent les structures de petites protéines comme l'inhibiteur de trypsine pancréatique bovine en 1982.

Parallèlement aux USA sont mises au point des techniques 2D hétéronucléaires, par les chercheurs Bodenhausen et Ruben du MIT qui corrèlent le déplacement chimique du noyau azote ou carbone au déplacement du proton auquel ils sont liés.

5.3. Cryo-microscopie électronique : années 70-80

En 1974 Ken Taylor et Robert Glaeser montrent la faisabilité d'utiliser une technique de spécimen congelé-hydraté en microscopie électronique permettant de se rapprocher des conditions biologiques réelles.

Vers 1980, au Laboratoire Européen de Biologie Moléculaire (EMBL) à Heidelberg, Jacques Dubochet trouve le moyen de garder l'eau dans le microscope électronique en refroidissant l'objet tellement rapidement que l'eau s'immobilise sans avoir le temps de cristalliser en glace. Ce procédé s'appelle la vitrification de l'eau. À -170 degrés, l'eau garde la structure du liquide et le spécimen reste intact [21].

Les chercheurs de Cambridge (Klugg, Crowther, Henderson) et d'autres ont mis au point différentes techniques de reconstruction 3D de différentes projections 2D de l'objet correctement remise en ordre. Chaque spécimen observé au microscope ne contient une seule particule mais des milliers voir des millions immobilisés au hasard dans la mince couche de suspension vitreuse.

5.4. Modélisation / PDB : années 1970-80

Les années 70 virent naître, sous l'égide de Walter Hamilton, Edgard Meyer et Olga Kennard la mise en place de la première base de données protéique (Protein Data Bank PDB) au Brookhaven national laboratory suite à un Symposium sur la structure et la fonction des protéines au niveau tridimensionnel en 1971[22].

Dans les années 80, la PDB monte en puissance, pilotée par un comité d'expert issu de l'IUCr (L'Union Internationale de Cristallographie). Les efforts conjoints de Fred Richards et des membres du comité aboutirent à la publication d'un guide pratique en 1989. Le dépôt des données structurales dans la PDB est un pré-requis pour toute publication.

Jane Richardson, du laboratoire de cristallographie du Duke University présente en 1977 un modèle général de classification de protéines à partir des motifs structuraux récurrents (hélice alpha, feuillet B, boucle ou coude) [23] (*figure 4*).

Elle publie en 1981 les dessins (*figure 4*) de son interprétation de la structure des protéines. Ces dessins avaient pour but de rendre les structures plus compréhensibles [17]. Effectivement, leur éloquence fera qu'ils seront rapidement adoptés par la communauté scientifique, le «*ribbon diagram*» était né. En 1982, Lesk et Hardman conçoivent un logiciel de génération automatique de structures.

6. Les années 90-2005 : résolution par diffraction RX et RMN des protéines de membrane et des grosses protéines ; les gains en résolution de la cryo-microscopie électronique

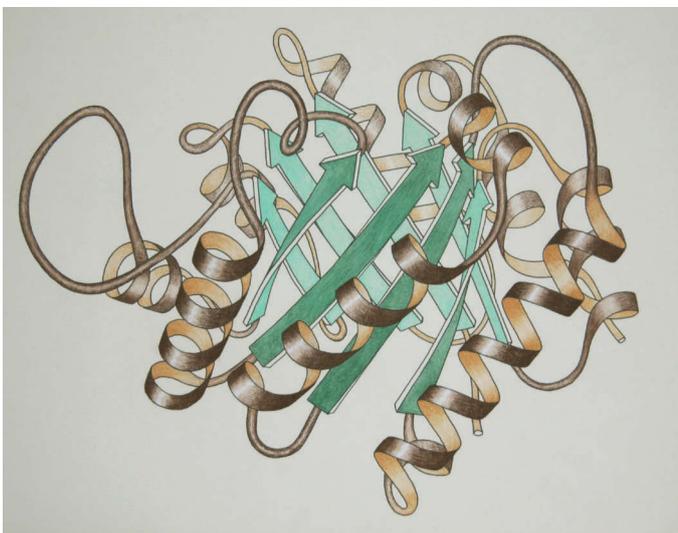
6.1. Cristallographie Rayons X : années 90-2005

Wayne Hendrickson et ses collègues mettent au point la méthode de diffraction «*anomale*» multi longueur d'onde (multiwavelength diffraction anormale (MAD) par incorporation d'atomes de sélénométhionine remplaçant la méthionine naturelle. Cette méthode permet de réduire le temps de résolution à 1 à 2 ans.[4]

Durant les années 90, les protéines kinases AMP cyclique dépendent font l'objet d'études structurales approfondies. Suzan Taylor et ses collaborateurs proposent, à partir des études structurales et par comparaison d'alignement de séquences un nouveau concept de «*colonne vertébrale moléculaire*» reliant les 2 sous unités de la protéine kinase permettant une flexibilité conformationnelle nécessaire à la fixation de l'ATP préalable à la phosphorylation sur la boucle d'activation [24]. Ses recherches ouvrent la voie à l'idée d'une régulation allostérique induite par des changements de conformation moléculaire. Ce champ de connaissance est le préalable à la découverte de nouveau type d'inhibiteur dans le cadre des thérapies ciblées.

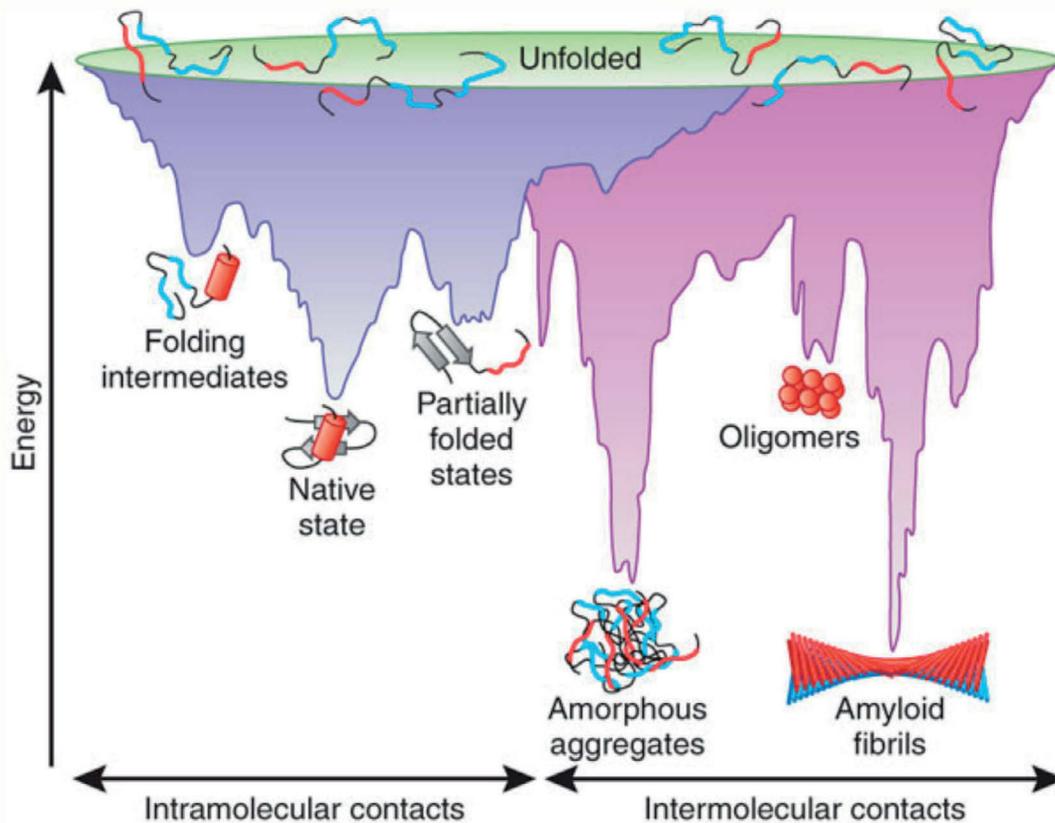
Parallèlement, Peter Wolynes [25] énonce la théorie du paysage énergétique moléculaire comme un nouveau modèle de repliement des protéines faisant intervenir des

Figure 4 – Dessin de Jane Richardson de la triose phosphate isomérase.



© Jane Richardson, 1981

Figure 5 – diagramme énergétique du repliement des protéines normales ou pathologiques.



Source : Ulrich H, Hayer-Hartl M. Converging concepts of protein folding *in vitro* and *in vivo* F. Nature Struct & Molec Biol 2009; 16 (6): 574-581

notions thermodynamiques. La représentation en forme d'entonnoir du paysage énergétique décrit la transition de conformations de protéines du désordre à l'ordre (figure 5).

Rod Mac Kinnon publie en 1998 la première structure à haute résolution d'un canal ionique.

Pour ce travail Mc Kinnon a partagé le prix Nobel 2003 de chimie [26].

6.2. La RMN : Années 90-2005

La méthode d'attribution séquentielle de Wuthrich est limitée aux petites protéines (moins de 80AA ou moins de 30kd). Les protéines plus grosses sont à l'origine d'artefacts de superposition et d'élargissement des pics.

L'équipe de Bax aux USA ouvre la voie des analyses multidimensionnelles des protéines marquées. Bax lui-même déclarait ne pas avoir assez de patience pour décrypter des documents de plus de un mètre de côté constellés d'innombrables minuscules tâches de résonance devant être attribuées « manuellement » de manière plus ou moins fiable aux différentes structures atomiques. L'enrichissement isotopique, l'amélioration des sondes radiofréquence, des champs magnétiques plus puissants et la résolution du problème de bruit de fond protonique lié à l'eau ont grandement contribué à améliorer la qualité de la résolution spectrale, la fiabilité et la vitesse d'exécution des expériences. [27]

Les années 90 ont vu aboutir la résolution RMN des protéines membranaires qui par définition échappent aux

techniques « en solution » et grâce au concept d'angle magique de rotation (« *magic angle spinning* » MAS) [28]. Un chercheur avait noté en tournant sa cuillère dans sa tasse de thé qu'il serait utile de faire tourner la molécule selon un angle dit magique pour obtenir le tenseur de déplacement chimique anisotropique et recalculer les pics.

En 1997 l'équipe de Wuthrich et al. met au point une nouvelle séquence dénommée TROSY (Transverse Optimized Relaxation Spectroscopy). Le principe est d'optimiser le phénomène de relaxation transverse (par allongement du temps de relaxation T2) est élaborée par K. Pervushin et K. Wuthrich [29] afin d'obtenir un spectre RMN affiné pour les grosses protéines dont l'étude était classiquement limitée par la largeur des pics qui augmentent avec la masse moléculaire pour résoudre les grosses protéines.

En 2000 Tycko et ses collaborateurs finalisent la technique du couplage dipolaire résiduel (residual dipolar coupling RDC) qui donne des informations structurales de distances et d'angles analogues aux techniques précédentes mais applicables aux molécules de grandes tailles [30]. L'effet RDC est observé dans une solution dite « légèrement orientante » obtenue grâce à l'introduction de cristaux liquides ou diverses substances.

L'étude des pseudos contact de déplacement chimique (Pseudocontact chemical shift PCS) grâce aux protéines para magnétiques permet de caractériser les tenseurs d'anisotropie.

Les 2 techniques conjointes RDC et PCS permettent grâce à des logiciels appropriés (Rosetta) de résoudre des structures de grosses protéines.

L'analyse de la relaxation du à la modulation des interactions dipôle – dipôle entre moments magnétiques de spin permet de remonter aux mouvements moléculaires sur une échelle de temps de l'ordre de quelques nanosecondes à la picoseconde. Les mouvements plus lents à l'échange de la microseconde, voire de la seconde peuvent aussi être abordés par RMN [20].

6.3. Cryo-EM: années 90-2005

Les chercheurs continuent de progresser dans la résolution atomique par cryo-EM et le développement d'outils informatiques de reconstruction 3D à partir des différentes images de projection 2D appelés reconstruction de particules uniques. Des avancées sont obtenues sur des molécules symétriques ou bien hélicoïdales bien ordonnées (pouvant être considérées comme des cristaux 1D). Ainsi Unwin de Cambridge en 1995 publie son étude du récepteur à l'acétylcholine en utilisant un processus de congélation rapide pour piéger l'état ouvert du canal avant inactivation [31].

En 1997, Crowther (Cambridge) résout la structure du virus de l'hépatite B (symétrie icosaédrale) à 7 Angström. En 1998 une avancée majeure est réalisée avec la mise en œuvre de technique statistique (maximum de vraisemblance) pour la reconstruction tridimensionnelle permettant de réduire les défauts de résolution liés aux bruits de fond.

6.4. PDB: années 90-2005

Dans les années 90 le NIGMS (*National Institute of General Medical Science*) devient la 1^{re} agence à imposer des pré-requis pour le dépôt des structures.

L'ancien format de fichier PDB crée en 1974 évolue pour intégrer des structures moléculaires plus complexes. Le fichier mmCIF est introduit en 1996 après une série de réunion et de congrès. Joel Sussman prend la direction de la PDB en 1994. Cette formation évolue vers un consortium en 1998 sous l'égide de la *Research Collaboratory for Structural Bioinformatics* (RCSB). La RSBP PDB collabore avec des centres européens et japonais. Ceci débouche en 2003 sur la création de la *WorldWide Protein Data Bank* (wwPDB) [32].

Les années 90 voient fleurir la multiplication des logiciels de visualisation moléculaire. Parmi ceux-ci, *RasMol* en 1993 est un logiciel gratuit pouvant fonctionner sur un ordinateur personnel et largement accessible [16]. Plusieurs autres logiciels sont sortis depuis: *Visual Molecular Dynamics* (VMD) en 1995, Chime en 1996 et PyMOL en 1999.

Le scientifique et dessinateur David Goodsell réalise des dessins au niveau moléculaire pour expliquer les concepts de la biologie cellulaire. Ses illustrations constituent une aide pédagogique pour l'enseignement de la biochimie [33]. Une vidéo est produite en 2006: «*the inner life of the cell*» [16-17].

En 1998 la technique de visualisation moléculaire par «*space filling model*» est implémentée sur logiciel pour l'étude des protéines. Cette méthodologie est particulièrement utile pour l'étude des surfaces de Van-der-Waals et l'examen de contact atomique entre protéines ou protéine-ligand.

7. Les années 2005-2016: la convergence bio-informatique des techniques (diffraction RX, RMN, cryo-EM), la standardisation et l'interopérabilité des bases de données et la modélisation dynamique

7.1. Cristallographie Rayons X: années 2005-2016

En 2007, Brian Kobilka et ses collègues résolvent la première structure d'une protéine G couplée à un récepteur. Cette prouesse technologique fut rendue possible par une innovation dans la technique de cristallisation couplée à un anticorps permettant de stabiliser les structures macromoléculaires de membranes [34].

L'utilisation du laser à électrons libres à rayons X (XFEL) avec impulsions femtosecondes sur des flux de nano cristaux permet à partir des années 2000 de franchir un cap décisif en vitesse de résolution et en protection de structure moléculaires contre le rayonnement ionisant. La Cristallographie femtosecondes rend désormais possible l'imagerie des cellules vivantes à des résolutions subnanométriques [35]. À terme cette méthode pourrait même ne plus nécessiter d'étape de cristallisation.

7.2. RMN: années 2005-2016

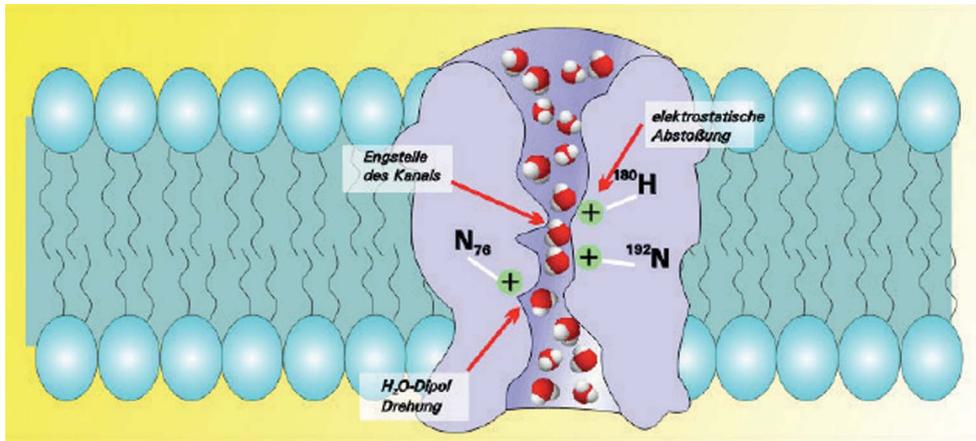
Les techniques récentes d'études RMN donnant des informations d'orientation à longue portée comme le couplage dipolaire résiduel (RDC) et les déplacements chimiques de pseudo contact (PCS) sont mises en œuvre pour analyser les aspects dynamiques des macromolécules. Ces études permettent d'étudier par exemple la dynamique du protéasome. [36]

La RMN intra-cellulaire introduite en 2001 par Serber pour les procaryotes devient éligible pour les eukaryotes grâce aux travaux des chercheurs japonais Inomata et al. Le principe est d'internaliser des protéines marquées grâce à des méthodes de pénétration transmembranaire. Il s'agit d'étudier les protéines dans leur environnement cellulaire *in vitro* ce qui représente une avancée majeure pour la compréhension des mécanismes cellulaires physiologiques ou pathologiques [37].

La RMN se positionne aussi comme une méthode de choix d'étude des protéines intrinsèquement désordonnée (protéine tau, protéine p53, synucléine, calcineurine) [38].

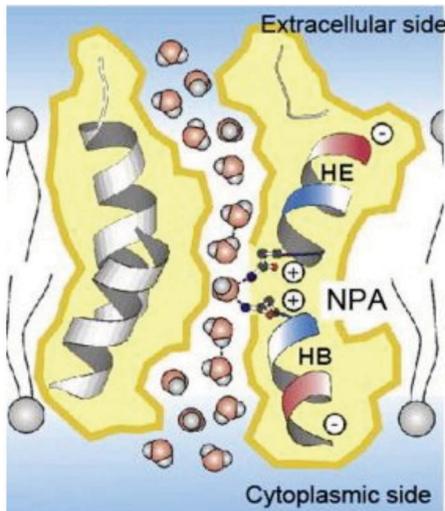
En effet ces protéines mal définies, hautement flexibles, nécessitent une approche plurielle combinant RMN, Cristallographie et modélisation dynamique. Les protéines amyloïdes comme les protéines prions ont la particularité de s'oligomériser suite à un changement de conformation puis de se propager de part en part et précipiter en fibres amyloïdes créant le substratum moléculaire de plusieurs grandes maladies neuro-dégénératives. Une connaissance accrue de ces mécanismes est le pré-requis pour la découverte de nouvelles thérapies moléculaires appropriées.

Figure 6a – Schéma générique de l'aquaporine.



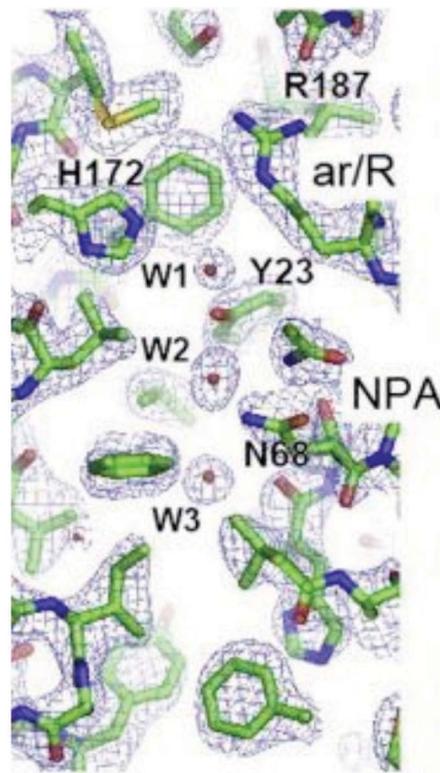
Source : Peter Wolber

Fig 6b – Aquaporine 1 : modélisation moléculaire d'après cristallographie électronique.



Source : d'après *Biochimica et Biophysica Acta* - Yoshinori Fujiyoshi 2000 et 2014

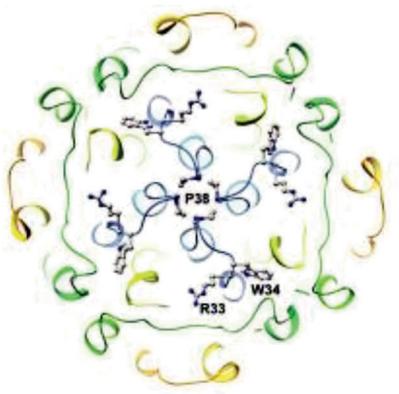
Fig 6c – Aquaporine 0 avec les molécules d'eau (W1, W2 et W3) visibles par la technique de cryomicroscopie.



Les molécules d'eau ne seraient pas visibles par cristallographie rayons X; cette technique permet d'accréditer l'hypothèse que les molécules d'eau sont bien distantes au-delà de la distance autorisée pour les liaisons hydrogène; la rupture des liaisons hydrogène dans cette zone de rétrécissement moléculaire expliquerait le mécanisme d'imperméabilité sélective de l'aquaporine aux protons.

Source : d'après *Biochimica et Biophysica Acta* - Yoshinori Fujiyoshi 2000 et 2014

Figure 6d – Vue extracellulaire de l'aquaporine en cryomicroscopie électronique. la proline 38 forme une structure en rosette au centre du tétramère.



Source : Tom Walz, 2005

En 2015 des chercheurs de l'école polytechnique fédérale de Lausanne, de l'institut de biologie structurale de Grenoble et de l'école normale supérieure de Lyon viennent de mettre au point une technique d'observation directe de la dynamique des protéines par RMN multi nucléaires à angle magique en utilisant une méthodologie de décongélation progressive en portant les protéines à -168 C° , les chercheurs ont pu stopper les mouvements moléculaires puis en augmentant la température, observer les mouvements de la molécule de façon hiérarchique : molécule du solvant puis chaîne latérale et enfin colonne vertébrale [39].

7.3. Cryo-EM: années 2005-2016

Une performance en très haute résolution est obtenue en 2005 par Cryo electro microscopie sur cristal 2D de l'aquaporine à $1,9\text{ \AA}$ par Tom Walz et Steve Harrison [40] (figure 6). Les molécules d'eau ne seraient pas visibles par cristallographie rayons X; cette technique permet d'accréditer l'hypothèse que les molécules d'eau sont bien distantes au-delà de la distance autorisée pour les liaisons hydrogène; la rupture des liaisons hydrogène dans cette zone de rétrécissement moléculaire expliquerait le mécanisme d'imperméabilité sélective de l'aquaporine aux protons d'après Yoshinori Fujiyoshi [45] (figure 6).

Les différentes techniques de reconstruction 3D citées précédemment en cryo-EM sont progressivement incorporées dans des logiciels de visualisation moléculaires (SPARX, Xmipp, RELION) [41]

En 2012, une petite révolution scientifique et technique est l'introduction d'une nouvelle gamme de caméra de détection directe des électrons (*Direct Detecteur Device DDD*) en lieu et place des traditionnelles caméras CCD (*Charge Coupled Device*). Les anciennes caméras CCD nécessitaient une conversion des électrons du microscope en photons par interaction avec un scintillateur. Cette découverte dont Grigorieff N. (*Massachusetts Institute of Technology*) était le pionnier, a permis de faire rentrer la cryo-EM sous la barre des 2-3 Angström de résolution [41].

L'enthousiasme lié aux progrès enregistrés dans la résolution de plus en plus précise des images de cryo-EM a été tempéré par une controverse sur une structure de protéine virale HIV publiée dans la revue PNAS en 2013 dont les experts ont prouvé ultérieurement qu'il s'agissait d'une image totalement erronée.[42] Cette controverse reprise par Henderson Richard (Cambridge, Royaume-Unis) la même année décrivant l'effet « *Einstein from noise* », c'est-à-dire la production d'une image purement construite à partir de bruit de fond.

Cette controverse a renforcé la communauté scientifique internationale dans l'absolu nécessité de définir des standards contrôlés et validés.

7.4. PDB: années 2005-2016

La standardisation et l'harmonisation des procédures d'intégration de données permettent la mise en place d'un langage commun entre différentes techniques et différentes communautés. Ainsi le format de fichier mmCIF permet de faire évoluer le dictionnaire vers une intégration des données RMN et 3D cryo aboutissant

à renommer les données au format « mmCIF » au sein d'un dictionnaire d'échange appelé « PDBx » [17, 22]. Le format mmCIF de PDBx sera généralisé à partir de 2011 au décours d'un congrès entre les représentants de la wwPDB et les principaux développeurs de logiciels de cristallographie [32].

Les données déposées dans la PDB sont rigoureusement évaluées et contrôlées. En 2008, 12 structures publiées dans des revues et insérées dans la base sont considérées comme erronées aboutissant à la mise en place d'une commission appelée Validation Task Force désignée pour mettre en place des recommandations de bonnes pratiques. La commission utilisera différentes méthodes pour évaluer l'intégralité de la base et proposer aux reviewers des critères précis d'éligibilité des structures.

En 2009 Ulrich et al présentent une vue conceptuelle des concepts de repliement appliqués aux protéines normales et pathologiques. Le paysage énergétique montre que le puits maximal de stabilité est obtenu pour la forme de polymère amyloïde qui représente aussi un état irréversible de la protéinopathie. L'accent est mis sur les formes intermédiaires dites « oligomériques » qui seraient le substratum toxique générateur d'agrégation [43].

En 2011 est mis en place par la wwPDB [32] un ambitieux système d'annotations et de dépôts sous forme modulaire. Les données sont soumises à travers un unique portail aboutissant à une procédure automatisée de haute qualité. Ces progrès permettent une interopérabilité des ressources et des systèmes ainsi qu'une lecture plus fluide et sécurisée des données macromoléculaire. Ces méthodes intégratives et multi-échelles qui nécessitent la coordination entre plusieurs disciplines représentent une force motrice qui combine les données expérimentales ainsi que les modélisations moléculaires indispensables pour la découverte de nouvelles molécules thérapeutiques et une meilleure compréhension des mécanismes du vivant. La PDB atteignait les 100 000 entrées en 2014.

Les efforts conjoints de la communauté internationale scientifique pour élaborer ces critères qualité ouvrent la voie d'une nouvelle branche de la biologie structurale dite « intégrative » tirant parti des effets de complémentarité et de validation mutuelle des différentes méthodologies, permettant en les combinant le cas échéant d'obtenir une vue élargie des complexes macromoléculaires [44].

8. Conclusion

Le rêve des biologistes et des médecins de voir, avec un détail atomique, comment différentes biomolécules interagissent entre elles et avec des substrats ou des molécules thérapeutiques suivant des schémas dynamiques bien définis se réalise progressivement. Ce rêve est aussi le fruit de découvertes de grande ampleur cédées par les générations antérieures : nous sommes, ainsi que le disait Bernard de Chartres, Maître à penser du XII^e siècle, comme des « *nains juchés sur des épaules de géants* ».

Liens d'intérêts: les auteurs déclarent ne pas avoir de liens d'intérêts avec le contenu de cet article.

Références

- [1] Briand Y, Brion P, Lafont R et al, « Protéines », Encyclopædia Universalis <http://www.universalis.fr/encyclopedie/proteine>
- [2] Tanford C, Reynolds J. Nature's robots : a history of proteins. Oxford university press, 2001. 304p.
- [3] Authier A. Early day of X-ray cristallography. Oxford university press, 2013. 441p.
- [4] Jaskolski M, Dauter Z, Wlodawer A. A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. FEBS J 2014 ; 281 : 3985-4009
- [5] Strandberg B, Dickerson RE, Rossmann MG. 50 years of Protein Structure Analysis. J. Mol. Biol. 2009; 392: 2-32.
- [6] Isaacs N. A history of experimental phasing in macromolecular cristallography. Acta cryst 2016; D72 : 293-295.
- [7] Rhodes G. Crystallography Made Crystal Clear. Third edition. Academic press, 2006. 305p.
- [8] Goldman M. Histoire de la RMN : Autour de ses premiers acteurs... et après. Actual Chim 2004 ; 273 : 57-60 .
- [9] Pauling L. How my interest in proteins developed. Protein Sci 1993; 2(6): 1060-1063.
- [10] Kendrew JC, Bodo G, Dintzis HM et al. A three dimensional model of the myoglobin molecule obtained by X-ray analysis. Nature 1958; 181: 662-666
- [11] Perutz MF, Rossmann MG, Cullis AF, et al. Structure of haemoglobin: a three dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. Nature 1960; 185: 416-422.
- [12] Histoire de la science des protéines. Yon-Kahn J. Grenoble, EDP sciences 2006: 312p.
- [13] Shi Y. A glimpse of structural biology through X-Ray crystallography. Cell 2014; 159: 995-1014.
- [14] Emsley JW, Feeney J. Milestones in the first fifty years of NMR. Prog Nucl Mag Res Sp 1995; 28: 1-9.
- [15] De Rosier DJ, Klug A. Reconstruction of Three Dimensional Structures from Electron Micrographs. Nature 1968; 217: 130-134.
- [16] Dorey S. Les logiciels de visualisation moléculaire dans l'enseignement des sciences de la vie: conceptions et usages. Thèse de doctorat - ENS Cachan, 2012 : 279p
- [17] Martz E, Francoeur E. History of visualisation of biological macromolecules. www.umass.edu/microbio/rasmol/history.htm
- [18] Aue W. P., Bartholdi E., and Ernst R.R. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. J Chem Phys 1976 ; 64 (5): 2229-2246.
- [19] Mispelter J. Principes de la détermination de la structure tridimensionnelle de protéines en solution par Résonance Magnétique Nucléaire (RMN). 2007 Institut Curie, INSERM.
- [20] www.biochimie-lille1.overblog.com
- [21] Dubochet, J., Lepault, J., Freeman et al. Electron microscopy of frozen water and aqueous solutions. J. Microsc 1982; 128: 219-237.
- [22] Berman HM, Kleywegt GJ, Nakamura H et al. The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future. Structure 2012; 20 (3): 391-396.
- [23] Richardson JS. B-sheet topology and the relatedness of proteins. Nature 1977; 268: 495-500.
- [24] Kornev AP, Taylor SS. Dynamics-Driven Allostery in Protein Kinases. Trends Biochem Sci 2015; 40 (11) : 628-647.
- [25] Frauenfelder H, Sligar SG, Wolynes PG. The energy landscape and motions of proteins. Scienceimage001.gif1991; 254: 1598.
- [26] Doyle DA, Morais Cabral J, Pfuetzner RA et al. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. Science 1998; 280(5360): 69-77.
- [27] Bax A. Triple resonance three-dimensional protein NMR: Before it became a black box. J Magn Reson 2011; 213(2): 442-445.
- [28] Shoolery J N. The development of experimental and analytical high resolution NMR. Prog Nucl Mag Res Sp 1995 ; 28 : 37-52.
- [29] Pervushin K, Riek R, Wider G et al. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. PNAS 1997; 94 (23): 12366-12371.
- [30] Tycko, R.; Blanco, F. J.; Ishii, Y. Alignment of Biopolymers in Strained Gels: A New Way to Create Detectable Dipole-Dipole Couplings in High-Resolution Biomolecular NMR. J Amer Chem Soc 2000. 122 (38): 9340-9341.
- [31] Unwin N. Acetylcholine receptor channel imaged in the open state. Nature 1995; 373: 37-43.
- [32] www.wwPDB.org
- [33] Goodsell DS. The machinery of life. Second edition. Springer Science, 2010. 167p.
- [34] Rasmussen SG, Choi HJ, Rosenbaum DM et al. Crystal structure of human beta2 adrenergic G-protein coupled receptor. Nature. 2007; 450(7168): 383-7.
- [35] Bergh M. Feasibility of imaging living cells at subnanometer resolutions by ultrafast X-ray diffraction. Q Rev Biophys 2008; 41: 181-204.
- [36] Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature 2007. 445(7128):618-22.
- [37] Inomata K, Ohnon A, Tochio H et al. High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. Nature 2009 ; 458: 106-110.
- [38] Ozenne V. Etude de la dynamique conformationnelle des protéines intrinsèquement désordonnées par résonance magnétique nucléaire. Thèse. Université de Grenoble, 2012.
- [39] Lewandowski JR, Halse ME, Blackledge M et al. Direct observation of hierarchical protein dynamics. Science 2015; 348 (6234): 578-581
- [40] Gonen, T. Cheng YF, Sliz P et al. T. Lipid-protein interactions in double-layered two-dimensional AQPO crystals. Nature. 2005;438:633-638.
- [41] Nogales E. The development of cryo-EM into a mainstream structural biology technique. Nature Methods 2016; 13 (1): 24-27.
- [42] Henderson R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. PNAS 2013; 110(45): 18037-18041.
- [43] Ulrich H, Hayer-Hartl M. Converging concepts of protein folding *in vitro* and *in vivo* F . Nature Struct & Molec Biol 2009; 16 (6): 574-581.
- [44] Van den Bedem H, Fraser J. Integrative, dynamic structural biology at atomic resolution-it's about time. Nature Methods 2015; 12(4) : 307-319.
- [45] Tani K, Fujiyoshi Y. Water channel structures analysed by electron crystallography. Biochimica et Biophysica Acta 2014; 1840:1605-1613.