



ETUDES EN CRISTALLOGRAPHIE DES PROTEINES

Wim P Burmeister

► **To cite this version:**

Wim P Burmeister. ETUDES EN CRISTALLOGRAPHIE DES PROTEINES. Biologie structurale [q-bio.BM]. Université Joseph Fourier, 2000. <tel-01301836>

HAL Id: tel-01301836

<http://hal.univ-grenoble-alpes.fr/tel-01301836>

Submitted on 13 Apr 2016

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

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

Université Joseph Fourier – Grenoble I

Sciences & Géographie

Faculté de Pharmacie

Habilitation à diriger les recherches

Wilhelm Pascal Burmeister

ETUDES EN CRISTALLOGRAPHIE DES PROTEINES

Soutenance le 26.6.2000 à 14 :00 devant le jury composé de

Président : Pr. Emmanuel Drouet

Rapporteurs : Pr. Marcel Knossow

Dr. François Penin

Pr. Joseph Zaccai

Examineurs : Dr. Stephen Cusack

Dr. Bernard Henrissat

TABLE DES MATIERES

1.	CURRICULUM VITAE	2
2.	LISTE DES PUBLICATIONS	3
3.	INTRODUCTION	6
4.	VERS LES LIMITES DE LA CRISTALLOGRAPHIE AUX RAYONS X	7
4.1.	Introduction	7
4.2.	La formidable boîte à outils de la cristallographie des protéines	7
4.3.	L'instrumentation synchrotron	11
4.4.	Les effets de l'irradiation	14
5.	VARIATIONS SUR UN THEME : LES GLYCOSIDASES	17
5.1.	Introduction	17
5.2.	Neuraminidase	18
5.3.	Myrosinase	21
6.	COMMENT RELIER STRUCTURE ET FONCTION	24
6.1.	Introduction	24
6.2.	Récepteur FcRn	24
6.3.	Phosphoprotéine du virus de Sendai	26
6.4.	Récepteur de l'adénovirus Ad3	28
7.	CONCLUSION ET PERSPECTIVES	30

1. CURRICULUM VITAE

Wilhelm Pascal Burmeister
5, place St. André,
38000 Grenoble

Travail: 04 76 20 72 82
Privé: 04 76 15 20 95
E-mail: wpb@embl-grenoble.fr

Laboratoire de Virologie, Domaine de la Merci
38706 La Tronche

wim.burmeister@ujf-grenoble.fr

36 ans, marié, nationalité allemande

Formation

- 1974 - 83 Lycée à Norderstedt, Schleswig-Holstein, Allemagne, conclu avec le baccalauréat ("Abitur")
- 1983 - 84 Service militaire
- 1984 - 87 Etudes de physique à l'Ecole Polytechnique d'Aix-la-Chapelle (RWTH Aachen), Allemagne ("Vordiplom", 1986)
- 1987 - 88 Maîtrise de Physique à l'Université Joseph Fourier à Grenoble
- 1988 - 89 Magistère et D.E.A. Science des Matériaux à l'Université Joseph Fourier, Grenoble
- 1989 - 92 Doctorat de Physique à Grenoble (Thèse de l'Université Joseph Fourier) avec mention très honorable et les félicitations du jury

Emplois

- 1985 Stage de deux mois à Philips Medical Systems à Hamburg au laboratoire de recherche.
- 1989 - 1992 Thèse au sein du laboratoire de Dr. S. Cusack au Laboratoire Européen de Biologie Moléculaire (EMBL) à Grenoble. Bourse EMBL.
- 1992 - 1994 Chercheur associé dans le groupe d'immunologie structurale du Pr. P. Bjorkman au California Institute of Technology, Pasadena, USA. Bourse EMBO.
- 1995 -1999 Chercheur (Beamline scientist) à l'Installation Européenne de Rayonnement Synchrotron (ESRF), Grenoble dans le groupe de Dr. S. Wakatsuki et responsable de la ligne de lumière ID14-3.
- 2/2000 –8/2000 Chercheur à l'Installation Européenne de Rayonnement Synchrotron et au Forschungszentrum Jülich, Grenoble.
- Depuis 9/2000 Maître de conférence en Physiques et Mathématiques Pharmaceutiques à la Faculté de Pharmacie de l'Université Joseph Fourier. Laboratoire d'accueil : Laboratoire de Virologie (E. Drouet) et EMBL Grenoble (S. Cusack et R. Ruigrok).

2. LISTE DES PUBLICATIONS

classées par ordre chronologique.

- (1) Burmeister, W. P., Daniels, R. S., Dayan, S., Gagnon, J., Cusack, S., & Ruigrok, R. W. H.
Sequence and Crystallization of Influenza B/Beijing/1/87 Neuraminidase
Virology **180**, 266-272 (1991).
- (2) Burmeister, W. P., Ruigrok, R. W. H., & Cusack, S.
The 2.2 Å resolution crystal structure of influenza B neuraminidase and its complex with sialic acid
EMBO Journal **11**, 49-56 (1992).
- (3) Burmeister, W. P., Henrissat, B., Bosso, C., Cusack, S. & Ruigrok, R. W. H.
Influenza B neuraminidase can synthesize its own inhibitor
Structure **1**, 19-26 (1993).
- (4) Burmeister, W. P., Baudin, F., Cusack, S., & Ruigrok, R. W. H.
Comparison of Structure and Sequence of B/Yamagata and B/Beijing Neuraminidases Shows a Conserved “Head” but Much Greater Variability in the “Stalk” and NB Protein
Virology **192**, 683-686 (1993).
- (5) Burmeister, W. P., Cusack, S., & Ruigrok, R. W. H.
Calcium is Needed for the Thermostability of Influenza B Neuraminidase
Journal of General Virology **75**, 381-388 (1994).
- (6) Burmeister, W. P., Gastinel, L. N., Simister, N. E., Blum, M. L. & Bjorkman, P. J.
Crystal structure at 2.2 Å resolution of the MHC-related neonatal Fc receptor
Nature **372**, 336-343 (1994).
- (7) Burmeister, W. P., Huber, A. H. & Bjorkman, P. J.
Crystal structure of the complex of rat neonatal Fc receptor with Fc
Nature **372**, 379-383 (1994).
- (8) Bjorkman, P. J. & Burmeister, W. P.
Structures of Two Classes of MHC Molecules Elucidated: Crucial Differences and Similarities
Curr. Opinion in Struct. Biology **4**, 852-856 (1994).
- (9) Hammersley, A. P., Brown, K., Burmeister, W. P., Claustre, L., Gonzalez, A., McSweeney, S. M., Mitchell, E., Moy, J. P., Svensson, S. O. & Thompson, A. W.
Calibration and application of an x-ray image intensifier/charge-coupled device detector for monochromatic macromolecular crystallography
J. Synchrotron Rad. **4**, 67-77 (1997).

- (10) Cipriani, F, Castagna, J.C., Claustre, L., Wilkinson, C., Tomizaki, T., Burmeister, W.P. & Wakatsuki, S.
A large x-ray image plate scanner for a Weissenberg camera on ID14
ESRF Newsletter **28**, 30-32 (1997).
- (11) Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S. & Henrissat, B.
The crystal structures of *Sinapis alba* myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase
Structure **5**, 663-675 (1997).
- (12) Wakatsuki, S., Belrhali, H., Mitchell, E.P., Burmeister, W.P., McSweeney, S.M., Kahn, R., Bourgeois, D., Yao, M., Tomizaki, T., & Theveneau, P.
ID14 'Quadriga', a Beamline for Protein Crystallography at the ESRF,
J. Synchrotron Rad., **5**, 215-221 (1998).
- (13) Burmeister, W. P., Bourgeois, D., Kahn, R., Belrhali, H., Mitchell, E. P., McSweeney, S. M. & Wakatsuki, S.
Diamonds, a multilayer and a sagittally focusing crystal as optical elements on the ID14/Quadriga-3 beamline at ESRF
Proc. Soc. Photo-Opt. Instrum. Eng. **3448**, 188-196 (1998).
- (14) Kozielski, F., De Bonis, S. Burmeister, W. P., Cohen-Addad, C., & Wade, R. H.
The crystal structure of the minus-end-directed microtubule motor protein ncd reveals variable dimer conformations
Structure, **7**, 1407-1416 (1999).
- (15) Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R. W. H. & Burmeister, W. P.
On the domain structure and the polymerization state of the Sendai virus P Protein.
Virology, **266**, 99-109 (2000).
- (16) Burmeister, W. P.
Structural changes in cryo-cooled protein crystals owed to radiation damage
Acta Cryst D **56**, 328-341 (2000).
- (17) Tarbouriech, N., Curran, J., Ruigrok, R.W.H. & Burmeister, W. P.
Sendai virus phosphoprotein oligomerization domain: First structure of a natural homotetrameric coiled coil.
Nature Struct. Biol. **9**, 777-781 (2000).
- (18) Burmeister, W.P. Cottaz, S., Rollin, P., Vasella, A. & Henrissat, B.
High resolution X-ray crystallography shows that ascorbate is acting as a cofactor for myrosinase providing the catalytic base
J Biol. Chem. **275**, 39385-39393 (2000).

- (19) Arzt, S., Baudin, F., Barge, A., Timmins, P. Burmeister, W.P. & Ruigrok, R.W.H
Combined Results from solution Studies on Intact Influenza Virus M1 Protein and from a
New Crystal form of its N-Terminal Domain Show That M1 Is an Elongated Monomer
Virology, **279**, 439-446 (2001).
- (20) Czjzek, M., Cicek, M., Zamboni, V., Burmeister, W.P., Bevan, D.R., Henrissat, B. &
Esen, A.
Crystal structure of monocotyledon (maize ZMGlu1) β -glucosidase and a model of its
complex with p-nitrophenyl- β -D-thioglucoside.
Biochem. J. **354**, 37-46 (2001).
- (21) Durmort, C., Mitraki, A., Stehlin, C., Schoehn, G., Drouet, E., Cusack, C. &
Burmeister, W. P.
Structure of the adenovirus fibre receptor domain of Ad3, a non CAR-binding serotype
Virology, in press.
- (22) Marion, D., Tarbouriech, N. Ruigrok, R. W. H., Burmeister, W. P. & Blanchard, L.
Assignment of the ^1H , ^{15}N and ^{13}C resonances of the nucleocapsid-binding domain of the
Sendai virus Phosphoprotein
sousmis.
- (23) Tarbouriech, N. & Burmeister, W. P.
A structure determination using crystals with a rare pseudo-merohedral twinning
en préparation.

3. INTRODUCTION

Dans les chapitres suivants je présenterai mes travaux dans le domaine de la biologie structurale, qui ont porté, dans un ordre chronologique, sur la neuraminidase du virus de la grippe, le récepteur FcRn du rat, la myosinase, la protéine P du virus de Sendai et le domaine récepteur de la fibre d'adénovirus. Deux projets n'ont pas pu être achevés à ce jour, et ne seront pas abordés : la base du penton, une protéine de la capsid de l'adénovirus, et la protéase du virus Epstein-Barr, une protéine impliquée dans la formation de la capsid virale dont la structure est en train d'être résolue.

J'ai regroupé ces sujets, et aussi le travail d'installation d'une ligne de lumière à l'ESRF dédiée à la cristallographie des protéines, en trois grands thèmes : méthodes, glycosidases et interactions protéine-protéine.

4. VERS LES LIMITES DE LA CRISTALLOGRAPHIE AUX RAYONS X

4.1. Introduction

Depuis le début de ma carrière la cristallographie des macromolécules a progressé rapidement et une étape a encore été franchie en l'an 2000 avec la détermination de la structure du ribosome, quand la grande sous-unité, qui comporte environ 100 000 atomes (sans les atomes d'hydrogène), a été résolue à 2.4 Å de résolution dans le groupe de T. Steitz et la petite sous-unité à 3 Å et 3.3 Å de résolution par les groupes de V. Ramakrishnan et d'A. Yonath.

Les développements technologiques, qui ont contribué à cette montée en puissance ont été les suivants:

- L'introduction des stations graphiques dans les années 80 pour la modélisation des structures
- le développement des détecteurs électroniques bi-dimensionnels, d'abord surtout le détecteur à plaque ("image plate"), plus récemment les détecteurs utilisant des CCD*
- l'introduction de l'indicateur R_{libre} par Brünger en 1992, qui a permis de valider un affinement et de vérifier si une structure est correcte. Ceci a accru énormément la confiance dans les structures déterminées par cristallographie aux rayons X
- l'utilisation de la basse température pour collecte de données et en parallèle le développement des sources synchrotron de plus en plus brillantes permettant l'utilisation d'échantillons de plus en plus petits. Ces deux techniques ont créé une synergie positive car les échantillons plus petits sont plus faciles à congeler.
- finalement l'explosion de la puissance de calcul des ordinateurs et la "popularisation" des logiciels grâce à des interfaces utilisateur plus conviviales, des formats de fichiers plus facilement échangeables entre différents logiciels et finalement une réduction du nombre des bogues dans les logiciels.

Quand en 1989, j'ai commencé à travailler dans le domaine de la biologie structurale tous ces développements étaient à leur début ou venait d'être introduits.

4.2. La formidable boîte à outils de la cristallographie des protéines

Lors de mon travail, j'ai appris à utiliser un grand nombre d'outils adaptés à la résolution de structures cristallographiques. Certains de ces outils sont essentiels pour aborder la résolution de structures difficiles.

Aujourd'hui, la résolution d'une structure de protéine peut se résumer à : Produire de la protéine sélénométhionylée, faire pousser les cristaux, les congeler et faire une expérience MAD# au synchrotron, utiliser un logiciel automatique de phasage et calculer la carte de densité électronique après aplatissage du solvant, et si possible la tracer d'une façon automatique (*i.e.* Arp/Warp). Cependant cette recette est loin d'être applicable à tous les problèmes. Souvent, les cristaux ne diffractent pas à assez haute résolution pour mener avec succès une expérience MAD. Comme je détaillerai dans le chapitre 4.4 (effets de l'irradiation), il n'est pas toujours possible d'extraire l'information nécessaire à partir d'un petit cristal dont le pouvoir diffractant est faible. Il se peut que le système d'expression ne se prête pas à la production de protéine sélénométhionylée, que la protéine soit trop grande ou simplement ne comporte pas de méthionine.

Enfin, les cristaux maclés constituent aussi un obstacle très important.

* CCD: Charge Coupled Device

MAD: Dispersion anormale à des multiples longueur d'onde, Multiple wavelength Anomalous Dispersion

Lors du développement de la cristallographie des macromolécules, surtout pendant les vingt dernières années, un grand nombre de techniques permettant de récupérer ou d'améliorer l'information de phase ont été mises au point, dont certaines sont devenues très populaires car faciles d'utilisation (par exemple grâce aux interfaces graphiques et aux formats des fichiers échangeables).

Je vais présenter deux des techniques que j'ai utilisées lors de mon travail et qui ne sont pas tout à fait standard :

- L'utilisation de la symétrie non-cristallographique pour l'amélioration des phases,
- la recherche 6-dimensionnelle de fragments dans l'espace réel pour un phasage initial quand un remplacement moléculaire classique n'était pas possible.

Finalement, j'écrirai quelques lignes sur les problèmes liés aux cristaux maclés que j'ai rencontrés.

On parle de symétrie non-cristallographique quand plusieurs monomères d'une protéine sont présents à l'intérieur de l'unité asymétrique. Cette unité asymétrique forme la base à partir de laquelle tout le cristal peut être généré en appliquant les opérations de symétrie et des translations d'un nombre entier de mailles dans les différentes directions. La présence d'une symétrie non-cristallographique impose des contraintes sur les phases d'une réflexion par rapport aux réflexions liées par la matrice de rotation de l'opérateur de symétrie non-cristallographique. Ainsi, il est possible d'améliorer considérablement la qualité des phases. La connaissance de la seule enveloppe permet déjà d'améliorer les phases car elle lie les phases entre réflexions voisines : c'est la technique de l'aplatissement de solvant. Cette technique est utilisée d'une façon systématique dans le phasage expérimental.

Pour l'utilisation de la symétrie non-cristallographique, il faut d'abord définir les opérateurs de symétrie et l'enveloppe du domaine de la protéine à laquelle ces symétries s'appliquent.

Dans le cas de la neuraminidase du virus de la grippe (réf. 2), qui sera discutée dans le chapitre 5.2., il était facile de trouver l'opérateur non-cristallographique à partir de la structure tétramérique connue de la protéine et des symétries cristallographiques. Ici, l'axe d'ordre 4 du tétramère et un axe d'ordre 2 cristallographique coïncident. L'enveloppe a pu être trouvée en calculant d'abord une carte de corrélation entre la densité électronique et la densité électronique après une rotation de 90° autour de l'axe d'ordre 4. En utilisant la méthode de Wang, l'enveloppe a été définie, analogue à la définition de l'enveloppe pour l'aplatissement du solvant, car la corrélation n'existe que dans les régions qui obéissent à la symétrie non-cristallographique alors qu'elle disparaît dans les autres régions. On obtient donc l'enveloppe du tétramère. Avec l'opérateur et l'enveloppe ainsi obtenus, après quelques cycles d'amélioration des phases, la carte de densité électronique était suffisamment bonne pour tracer la structure.

La structure de la neuraminidase a été déterminée avant l'introduction de l'indicateur R_{libre} . Cet indicateur consiste à exclure une partie (~ 5 %) des réflexions de l'affinement et de calculer un facteur R avec ces réflexions. Ce facteur R n'est pas biaisé par l'affinement et permet donc de vérifier la qualité d'une structure, si on a tenu compte de la symétrie non-cristallographique lors de la sélection des réflexions. Il est nécessaire d'éviter toute relation entre réflexions utilisées lors de l'affinement et réflexions du jeu pour le calcul de R_{libre} . Dans l'affinement de la neuraminidase je me suis aidé d'un autre indicateur à savoir la différence entre les phases calculées et les phases du phasage expérimental, qui devrait se réduire tout au long de l'affinement de la structure car l'erreur des phases calculées par rapport aux "vraies" phases diminue au cours de l'affinement, alors que l'erreur des phases expérimentales reste constant. Néanmoins, cet indicateur présente le désavantage que sa valeur théorique n'est pas connue et qu'il faut assumer que $\langle \varphi_{\text{exp}} \rangle = \langle \varphi_{\text{vrai}} \rangle$.

Pour mon deuxième projet, la détermination de la structure du récepteur FcRn (réf. 6, Fig. 17) pendant mon post-doctorat, celle-ci était essentiellement connue d'avance. Mais

malgré des homologies de séquence fortes avec la molécule CMH* classe I, le remplacement moléculaire s'est avéré impossible. Ceci était certainement dû à la présence de trois molécules dans l'unité asymétrique et des similitudes du repliement des différents domaines. En plus de l'homologie de séquence, la chaîne lourde de la molécule CMH classe I et la chaîne lourde du FcRn sont toutes les deux liées à une molécule de β -microglobuline, une petite protéine d'une centaine d'acides aminés avec un repliement d'immunoglobuline.

Après l'échec du remplacement moléculaire, la résolution de la structure du récepteur FcRn s'est révélée particulièrement délicate car seules des phases MIRAS* d'une qualité très médiocre étaient disponibles du à la sensibilité des cristaux aux rayons X. En effet les données n'ont pu être collectées qu'à température ambiante avec un générateur rayons X du fait de l'impossibilité d'une congélation.

Nous avons donc adopté une stratégie consistant à placer la protéine domaine par domaine dans la carte MIRAS après aplatissage du solvant, par recherche "6-dimensionnelle" dans l'espace direct et en s'aidant de l'information sur l'empilement cristallin, les positions des atomes lourds des dérivés et la position relative de la chaîne lourde par rapport à la chaîne légère. La première étape a consisté à localiser une molécule de β -microglobuline dans la carte de densité électronique avec une recherche en 6-dimensions en utilisant sa trace carbone- α . Domaine par domaine, le dimère de FcRn qui était présent dans l'unité asymétrique a pu être assemblé. La position de la molécule ainsi obtenue a été utilisée pour les calculs des enveloppes et la détermination des opérations de symétrie non-cristallographique. Les phases expérimentales ont ainsi été améliorées en appliquant ces relations. Elles ont l'avantage de ne pas être biaisées par le modèle de départ du remplacement moléculaire, un net avantage lors du traçage de la chaîne. Une fois la structure déterminée dans cette première forme, des cristaux d'une deuxième forme diffractant à haute résolution (2.2 Å) ont pu être utilisés avec une procédure de remplacement moléculaire traditionnelle pour améliorer le modèle final.

Pendant l'affinement nous nous sommes heurtés à des difficultés dues à la présence inattendue d'un monomère libre dans l'unité asymétrique, à côté d'un dimère. C'est un cas très rare et inattendu, surtout parce-que dans 3 autres formes de cristaux seuls des dimères sont présents à l'intérieur du cristal.

Finalement, une combinaison de MIRAS avec la technique de la recherche "6-dimensionnelle" a permis de résoudre la structure d'un complexe FcRn-Fc à 4.5 Å de résolution (réf. 7).

Il est surprenant que les méthodes de recherche 6-dimensionnelle ne se soient guère développées, car même à l'époque, avec une puissance de calcul limitée, le temps nécessaire pour cette recherche était tout à fait abordable (quelques jours maximum).

Comparée à ces problèmes, la résolution de la structure de la myosinase par remplacement moléculaire avec un affinement à 1.2 Å de résolution s'est révélée tout à fait classique, même en l'absence de la séquence primaire (réf. 11, réf. 18).

La résolution de la structure du domaine récepteur de l'adénovirus Ad3 (réf. 21) a été extrêmement facile et a montré la puissance des outils disponibles aujourd'hui pour des cristaux possédant un bon pouvoir diffractant (1.6 Å de résolution). Après le remplacement moléculaire, 95 % des résidus ont été tracés automatiquement dans la journée en utilisant le logiciel Arp/Warp et l'affinement a été complété en quelques jours.

Finalement je vais présenter le travail avec des cristaux maclés dans ce chapitre plutôt technique.

Une première forme de cristaux du domaine de multimérisation de la protéine P du virus de Sendai (réf. 15) a montré un type de macule rare, correspondant à une pseudo-mérohédrie dans un réseau centré C222₁. Ce macule correspond à une rotation de 90° autour de l'axe *b*. Ce

* CMH: Complexe Majeur d'Histocompatibilité

* MIRAS: Multiple Isomorphous Replacement using Anomalous Scattering

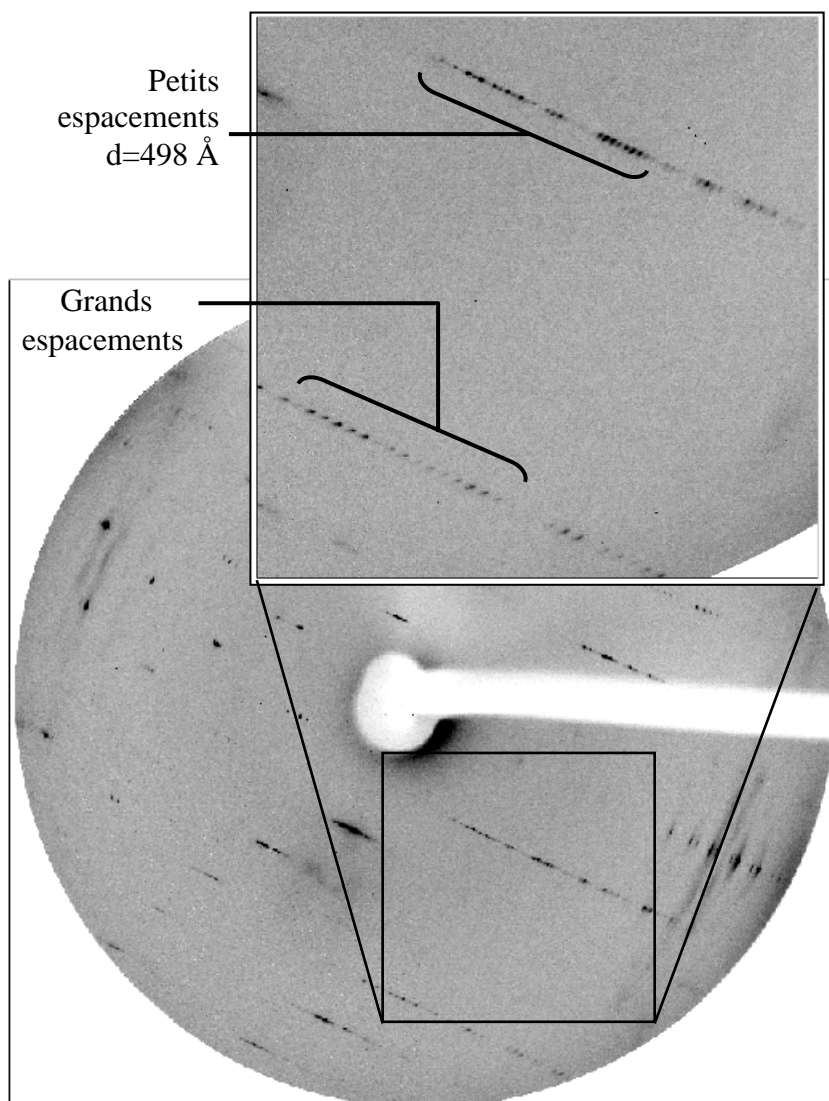


Fig. 1: Cliché de diffraction d'un cristal du domaine de multimérisation de la protéine P maclé avec une pseudo-mérohédrie. Insert : agrandissement

Distance du détecteur
MarCCD 133mm: 300 mm
 λ : 0.9475 Å
Résolution: 4 Å au bord

type de macle se manifeste par une loi d'extinction "impossible" (Fig. 1). Il faut aussi noter les paramètres de maille inhabituels de $a=c=49$ Å, $b=498$ Å, possédant un rapport 1/10 entre les deux axes a et b . Dans ce cas spécial de pseudo-mérohédrie, il est possible de trier les réflexions en réflexions maclées et non-maclées.

Mais d'abord la structure d'une protéine légèrement tronquée coté C-terminal a été résolue (chap. 6.3, réf. 17) en utilisant une forme tétragonale de cristaux. Ces cristaux partagent les deux paramètres de maille de 49 Å, indiquant une orientation des molécules identique à la forme orthorhombique. La structure des cristaux maclés a pu être résolue en recherchant la position du modèle dans l'espace direct qui minimise le facteur R cristallographique. Ici le problème se limite à des recherches 3-dimensionnelles car l'orientation du modèle est connue. Le remplacement moléculaire classique n'était pas possible, probablement à cause de la forme extrêmement allongée de la molécule (Fig. 20). L'information biologique de la structure ainsi déterminée reste limitée, du fait de la résolution d'environ 4 Å. Néanmoins, elle semble permettre de localiser une partie de l'extension C-terminale de la protéine dans les cristaux maclés, par rapport à la protéine tronquée. Cette structure reste un joli exemple des macles des cristaux et de leur explication par l'empilement des molécules (réf. 23).

J'ai encore rencontré des problèmes de macle lors du travail sur les cristaux de la protéase Epstein-Barr. Ici la situation est compliquée par la présence simultanée d'une pseudo-symétrie en plus des macles ce qui fait apparaître les limites des outils disponibles pour l'analyse des macles. Le travail se concentre maintenant d'un coté sur la production d'une protéine mutée qui, nous espérons, ne présentera pas ce macle, et aussi sur une détermination

de la structure à partir des données maclées. Finalement, ce travail semble aboutir et l'affinement de la structure est en cours.

Dans le domaine du traitement des données des cristaux maclés, à l'avenir, des progrès seront certainement encore possibles.

4.3. L'instrumentation synchrotron

Quand le projet de la ligne de lumière ID14 à été lancé, il avait pour but de combler rapidement un retard de l'ESRF dans le domaine de la biologie structurale. Une conception qui prévoyait l'installation de multiples stations expérimentales sur la même élément d'insertion, la source de rayons X, a été choisie. Je faisais partie de l'équipe de scientifiques qui ont été chargés de mettre en œuvre ce projet.

La solution technique consistait à construire trois stations expérimentales latérales grâce à des monochromateurs largement transparents aux rayons X. Ces monochromateurs utilisent des lames fines coupées dans des monocristaux de diamant. Une quatrième station, permettant de changer l'énergie des monochromateurs rapidement pour faire du MAD était prévue sur la ligne directe (réf.12, Fig. 2).

Mon travail a largement porté sur la réalisation d'un système optique basé sur des cristaux de diamant. Il a nécessité une importante phase de caractérisation et de développement. Les lames d'une taille d'environ 7*7 mm ont été obtenues à partir de cristaux de diamant synthétiques, après une caractérisation approfondie par l'étude de courbes de diffraction et d'images de topographie, réalisées essentiellement sur la ligne optique BM5 de l'ESRF. Il a aussi fallu mettre en œuvre toute l'instrumentation du système optique, formé d'un deuxième cristal de germanium courbé pour la focalisation verticale et d'un miroir multicouche pour la focalisation horizontale. Ce miroir multicouche est formé d'une alternance de couches très fines (périodicité 65 Å) d'un matériau lourd, le ruthénium et d'un matériau léger, le carbure de bore B₄C, déposées sur un substrat de silicium collé sur un support en céramique piézoélectrique. Les couches reflètent les rayons X suivant la loi de Bragg. Le support piézoélectrique (bimorphe) peut être courbé par l'application d'une tension pour obtenir le rayon nécessaire pour la focalisation horizontale en utilisant le même principe qu'une lame bimétallique.

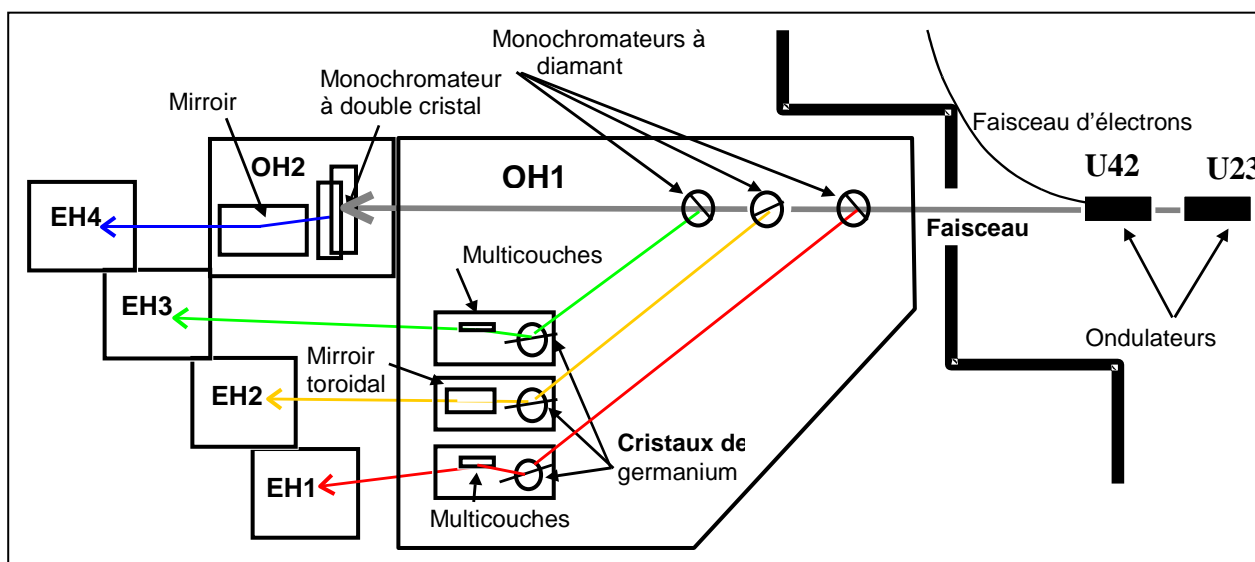


Fig. 2: Schéma de la ligne de lumière ID14 à l'ESRF.

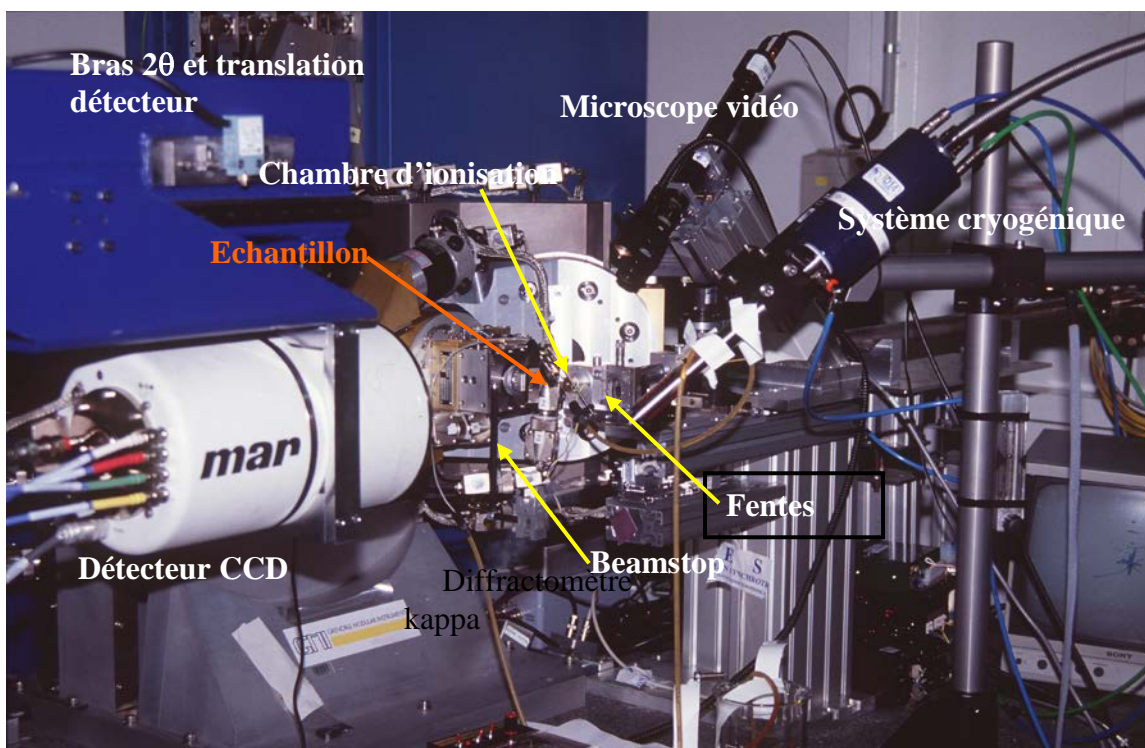


Fig. 3: Environnement échantillon de la station expérimentale ID14-3 à l'ESRF lors d'une expérience utilisant le détecteur CCD.

Ces trois éléments produisent un faisceau de rayons X dans la cabine expérimentale de dimensions 1.0 mm (largeur à mi-hauteur horizontale) x 0.08 mm (largeur à mi-hauteur verticale) (réf.13). La focalisation horizontale relativement médiocre résulte d'effets de dispersion chromatique qui sont inévitables vu l'arrangement des deux cristaux, en mode Laue pour le diamant et en mode Bragg pour le cristal de germanium. Le mode Laue signifie que le faisceau diffracté traverse le cristal, alors qu'il est reflété en mode Bragg. L'utilisation des cristaux de diamant en mode Laue a été imposée par des contraintes mécaniques de polissage. Les phénomènes de dispersion et aussi de focalisation qui se produisent lors de la diffraction en mode Laue des rayons X avec une très faible divergence sont gouvernés par la théorie dynamique de la diffraction et n'avaient par réellement été décrits au moment de la construction de la ligne.

Au niveau de la station expérimentale, la taille du faisceau délimitée par des fentes varie entre $50 * 50 \mu\text{m}$ et $200 * 200 \mu\text{m}$ (Fig. 3). L'intensité du faisceau est mesurée avec des chambres d'ionisation. Le centrage de l'échantillon sur l'axe de rotation du diffractomètre se fait à l'aide d'un microscope. Vu la sphère de confusion du diffractomètre, la taille du faisceau, et les performances du microscope utilisé pour centrer l'échantillon, la taille minimale d'un cristal exploitable est estimée à $10*10*30 \mu\text{m}$, ce qui ne veut pas dire qu'il est possible de collecter un jeu de données complet à partir d'un échantillon de cette taille (voir le chapitre 4.4).

Lors de la conception de la ligne, l'utilisation des collectes de données à basse température venait d'être introduite. Je me suis familiarisé avec cette technique, lors de ma thèse à l'EMBL, où quelques expériences préliminaires ont été faites et pendant mon post-doctorat. A l'EMBL, j'ai utilisé la congélation pour essayer de piéger des complexes enzyme-substrat. Ceci a donné des jeux de données exploitables, même si le complexe enzyme-substrat n'a pas pu être observé.

La construction de la ligne ID14 se déroulait en même temps que les collectes de données "phares" sur la capsid du virus de la langue bleu par D. Stuart et ses collaborateurs. Une congélation des cristaux était impossible à cause des paramètres de maille d'environ 800 \AA et la mosaïcité très faible requise.

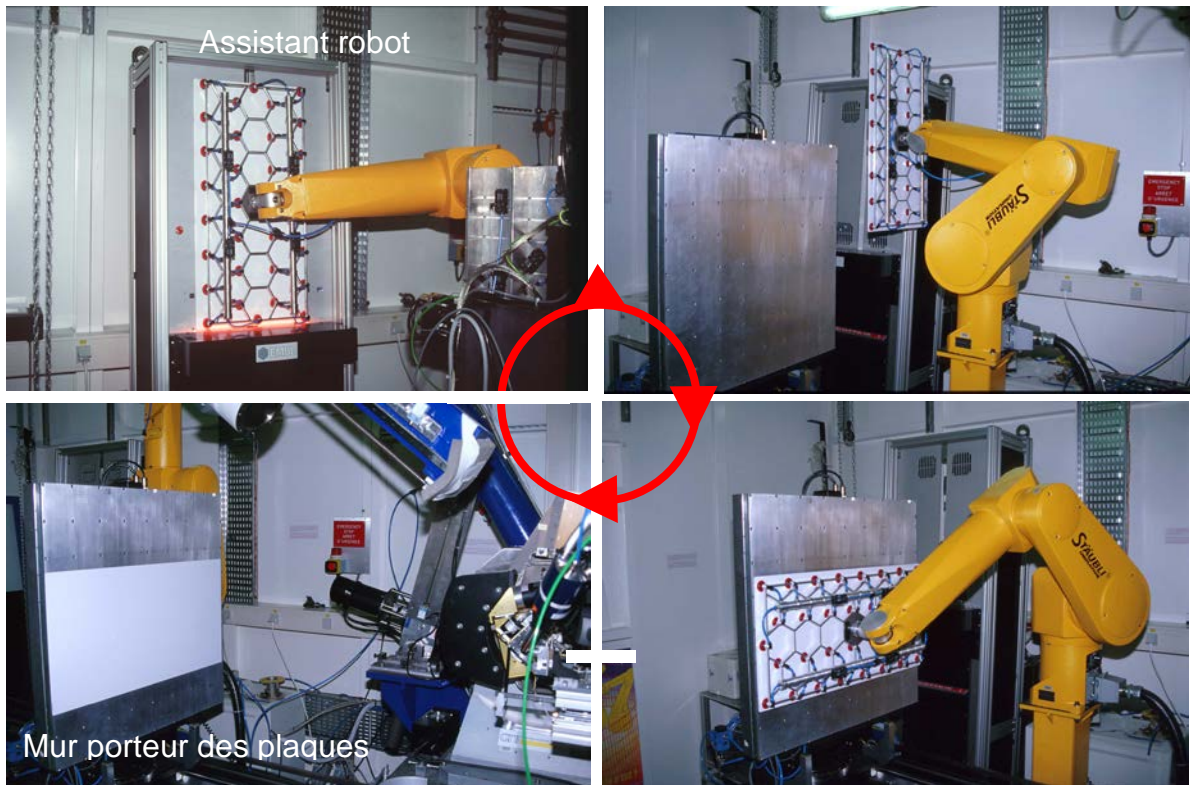


Fig. 4: Cycle du chargement de la caméra Weissenberg robotisée. En haut, à gauche: Présentation et effaçement de l'image plate. A droite: Transfert de l'image plate vers le mur. En bas, à gauche: image plate placé pour l'exposition

Il était donc attendu qu'un grand nombre de projets prestigieux, comme la détermination de la structure du ribosome, impliqueraient aussi des mailles élémentaires très grandes ne pourraient pas tolérer la congélation des cristaux. Une solution au problème consistait à utiliser une caméra Weissenberg avec une très grande surface sensible pour extraire un maximum de données avant la destruction de l'échantillon par le rayonnement et pour résoudre en même temps les taches de diffraction très proches.

Lors de l'installation de la station expérimentale, la mise en oeuvre d'une telle caméra Weissenberg robotisée basée sur des plaques photosensibles ("image plate", réf. 10) a demandé de gros efforts. Cette caméra a une surface sensible de 800*800 mm avec des pixels de 0,1 mm. Un scanner spécialement adapté à ces plaques a été construit à l'EMBL Grenoble, de même qu'un dispositif de présentation de ces plaques appelé "assistant robot". Un robot industriel permettait l'échange automatique des plaques (Fig. 4). La coordination entre les différentes composantes: assistant robot, robot et mur porteur des plaques s'est révélée très complexe. En plus, ce système a du être intégré dans l'environnement échantillon de la station (Fig. 3) qui comporte déjà toutes les composantes d'une station expérimentale pour la diffraction des rayons X, y compris un détecteur CCD pour le test des cristaux.

Finalement, des problèmes scientifiques nécessitant cette caméra manquaient. En fait, le mode Weissenberg n'a jamais été utilisé. Nous avons uniquement profité de la grande taille de la surface sensible pour prendre des clichés avec un angle de rotation relativement élevé sur des cristaux congelés de très grande maille et avec un bon pouvoir diffractant. A cause de la grande surface sensible, il n'y a que peu de recouvrement spatial des taches. Un cliché de diffraction d'un tel projet montre Fig. 5.

Une opération de routine n'a jamais réellement été possible par manque de moyens alloués à la finition du projet. La caméra a fini par être dépassée par des détecteurs CCD de plus en plus grands et bientôt sans doute aussi par des détecteurs "pixel" fondés sur des éléments

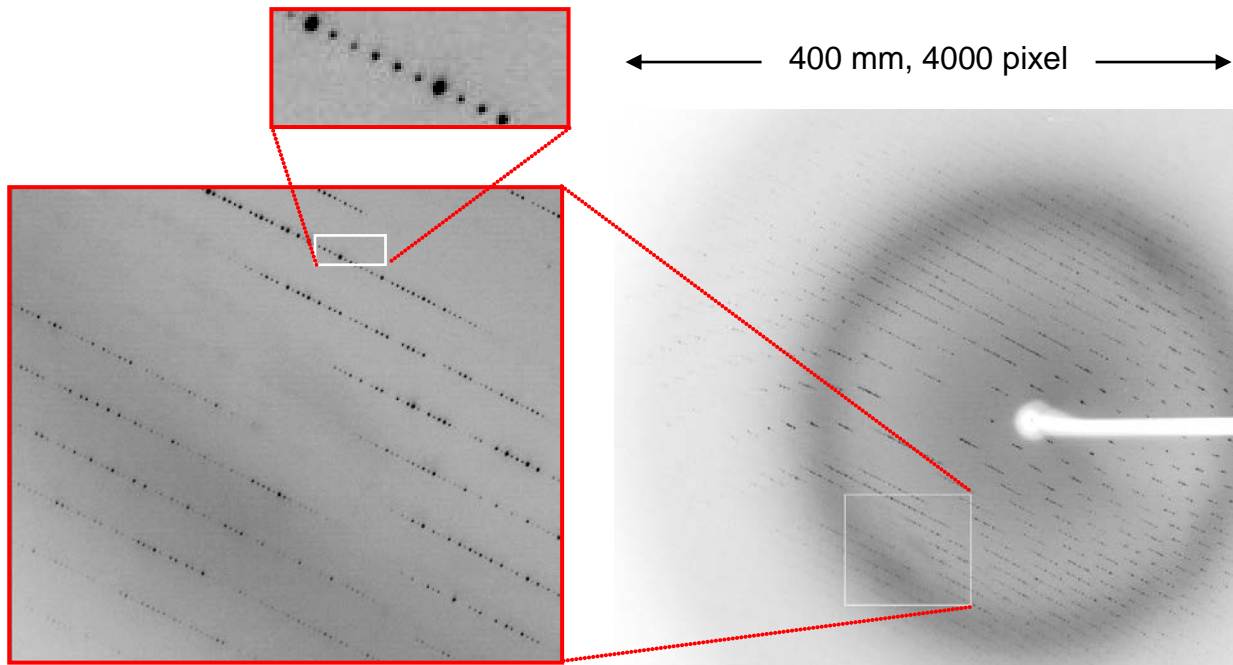


Fig. 5: Clichée de diffraction de la acetylcholinesterase (Y. Bourne et P. Marchot, AFMB Marseille) pris avec la camera LIP (Large Image Plate).
Groupe d'espace: $P6_522$, avec $a=73 \text{ \AA}$ $c=552 \text{ \AA}$. A droite, la moitié de l'image plate est montrée, qui fait 8000×4000 pixels au total. La caméra peut porter deux de ces plaques.

sensibles à base de silicium, qui peuvent atteindre des surfaces de 300×300 mm avec une très bonne résolution spatiale.

Néanmoins, le "petit" détecteur CCD (MarCCD) avec un diamètre de 133 mm et 2048×2048 pixels était extrêmement performant, surtout par rapport à la résolution spatiale qui a permis de résoudre jusqu'à 300 ordres de diffraction. Un grand nombre de structures ont pu être résolues. Un exemple qui était particulièrement difficile à cause de la petite taille des cristaux et de leur faible pouvoir diffractant est présenté dans la réf. 14.

4.4. Les effets de l'irradiation

Lors des collectes de données sur la myosinase il s'est avéré que dans les structures à haute résolution les ponts disulfures présentaient très souvent une conformation alternée, qui correspondait à un pont disulfure réduit. J'avais déjà fait cette observation lors de la détermination de la structure du récepteur FcRn (réf. 6). J'ai donc mené une étude plus poussée sur les effets de l'irradiation sur la structure des protéines, car les échantillons peuvent recevoir des doses énormes lors des collectes de données effectuées avec une source synchrotron de 3^{ème} génération. En utilisant le refroidissement des échantillons à 100 K par un flux d'azote, des données peuvent être collectées sur de très petits cristaux. Mais même à basse température, les dégâts causés par l'irradiation deviennent un facteur limitant. Jusqu'à ce jour, les dommages par radiation n'avait été décrits qu'en terme de perte globale de pouvoir diffractant.

J'ai analysé la structure de la myosinase (réf. 11) après exposition à une dose de rayons X croissante. Les expériences ont été faites sur la ligne ID14-3 à l'ESRF. Un effort spécial a été fait pour obtenir une bonne mesure des doses de l'irradiation. Les changements de la structure ont été analysés en utilisant des cartes de Fourier différence et en utilisant des affinements des taux d'occupation des groupes chimiques affectés. La disparition des ponts disulfure, due d'abord à une réduction suivie éventuellement de la perte des atomes de soufre, (Fig. 6) la

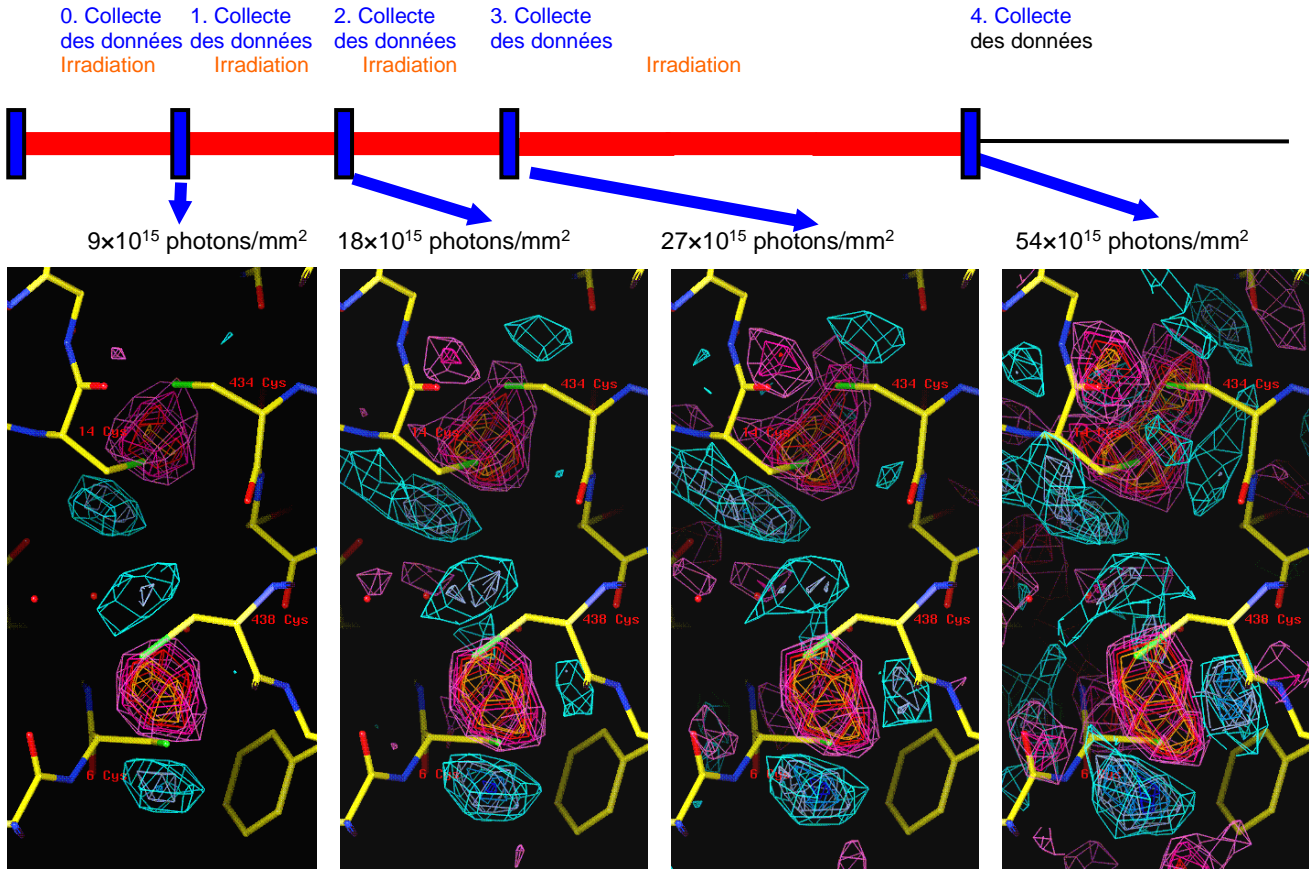


Fig. 6: Effets de l'irradiation sur une protéine : En haut : Schéma de l'expérience, des collectes de données brèves alternent avec des périodes d'irradiation du cristal avec une dose donnée. En bas : Cartes de densité électronique différence par rapport au jeu de données 0 qui sert comme référence. Niveaux des contours à 0.013, 0.021, 0.028 et 0.056 e⁻³. La structure native (avant irradiation) est représenté dans la région de deux ponts disulfures. Des effets similaires sont observées pour les résidus Asp, Glu, Tyr et Met.

décarboxylation des acides aspartiques et glutamiques et la perte des groupes hydroxyle des tyrosines et des groupes méthanthio des méthionines ont été observées. Cette perte de densité électronique due à la radiolyse a été quantifiée (Fig. 7). Une possible relation entre la sensibilité aux rayons X et l'environnement des résidus a été étudiée. Les résultats obtenus sont en accord avec des résultats publiés sur les effets d'une irradiation par des électrons ou des rayons γ sur les protéines ou acides aminés qui ont été obtenus par résonance paramagnétique électronique ou d'autres techniques spectroscopiques.

Ces résultats suggèrent qu'il y existe des limites intrinsèques dans la collecte des données sur des petits échantillons, même à basse température, similaire à ceux décrits dans le cas de la microscopie électronique. La qualité des modèles cristallographiques peut être compromise à cause de la présence de radiolyse dans le cristal suite à une dose d'irradiation excessive. Ces effets peuvent être une source de non-isomorphisme dans des expériences de type MAD. Mais, il est peut-être possible d'exploiter l'effet pour créer des différences isomorphes par radiolyse des résidus de sélénométhionine.

Ces études de grande importance pour l'opération d'une station pour la bio-cristallographie se sont achevées avec une publication (réf. 16). Il faudrait encore étudier la dépendance des dégâts non seulement en fonction de la dose mais aussi en fonction du flux, ainsi que l'effet des agents protecteurs potentiels. Ces études nécessitent un flux plus élevé que celui obtenu sur ID14-3 et je les ai laissées à mes anciens compétiteurs.

Malheureusement, les résultats des études sur le "radiation damage" sont restées sans suite dans l'opération des lignes de lumière. Les tentatives pour quantifier le flux de photons sur les stations expérimentales ont été abandonnées après mon départ. De nombreuses expériences sont compromises par une absence complète de dosimétrie. Souvent les utilisateurs ne notent qu'au moment du traitement des données que la dose maximale tolérable a été dépassée en observant que la résolution du cliché de diffraction baisse fortement au cours d'une collecte de données. Ou, alors, qu'une expérience MAD s'avère inexploitable.

Je suppose, que dans le "petit" monde des synchrotrons il est difficile d'accepter qu'il y ait des limites sur la quantité d'information que l'on peut extraire d'un échantillon d'une taille donnée et qu'il est souhaitable de prendre toutes les mesures possibles pour préserver l'information du signal diffracté. Trop souvent, un bruit de fond parfaitement évitable s'ajoute

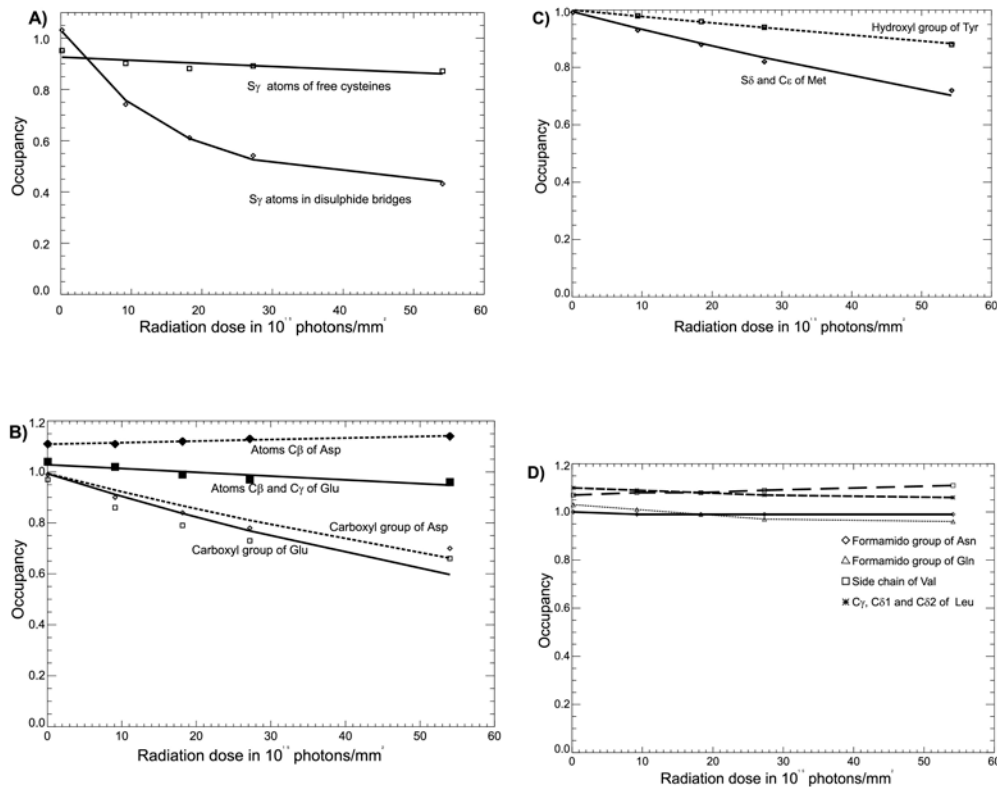


Fig. 7. Diminution du taux d'occupation de différents types de groupe chimique après une exposition à une dose d'irradiation croissante. Les atomes du même type de groupe chimique ont été groupés pour l'affinement du taux d'occupation. a) montre l'effet sur les cystéines b) montre que le groupe carboxylate des résidus acide aspartique et glutamique est affecté d'une façon spécifique. c) montre l'effet sur les tyrosines et méthionines d) montre des groupes stables, qui ont été utilisés pour contrôler la méthode utilisée.

sur les clichés de diffraction. En plus, souvent les données sont dégradées à cause de fluctuations d'intensité des faisceaux rayons X dues à diverses instabilités mécaniques et électroniques qui ne sont pas surveillées.

5. VARIATIONS SUR UN THEME : LES GLYCOSIDASES

5.1. Introduction

Deux de mes sujets de recherche portaient sur des glycosidases, qui sont des protéines qui hydrolysent des liaisons entre un sucre et son aglycone. Très souvent, comme dans le cas de la neuraminidase, celui-ci est un autre sucre. L'aglycone peut aussi être un autre groupe chimique, comme dans le cas de la myrosinase. La famille des glycosidases est très large, comportant des enzymes comme les cellulases et les amylases, qui dégradent les polymères naturels les plus répandus dans le monde: la cellulose et l'amidon. Une autre enzyme très connue en biochimie, le lysozyme, fait aussi partie de cette famille, et malgré son statut de sujet d'étude de prédilection, son mécanisme réactionnel réel n'a été élucidé que tardivement.

Au premier abord, le mécanisme réactionnel d'une glycosidase avec rétention de configuration a l'air simple (Fig. 8) : Un résidu carboxylique agissant comme un acide donne un proton sur l'oxygène de la liaison glycosidique en même temps qu'un autre acide carboxylique, dans sa forme déprotonée mène une attaque nucléophile sur l'atome de carbone anomérique (I). On forme un intermédiaire covalent lié au nucléophile et l'aglycone libre (II). Pour la deuxième étape de la réaction, une molécule d'eau est activée par l'abstraction d'un proton par le premier acide carboxylique, l'acide-base général, qui était déprotoné (III). Cette molécule mène une attaque nucléophile produisant ainsi le sucre libre et la situation initiale est rétablie (IV).

Mais une étude plus approfondie montre que la nature est nettement plus riche comme je vais montrer à l'exemple de la neuraminidase et de la myrosinase, deux enzymes agissant avec rétention de configuration.

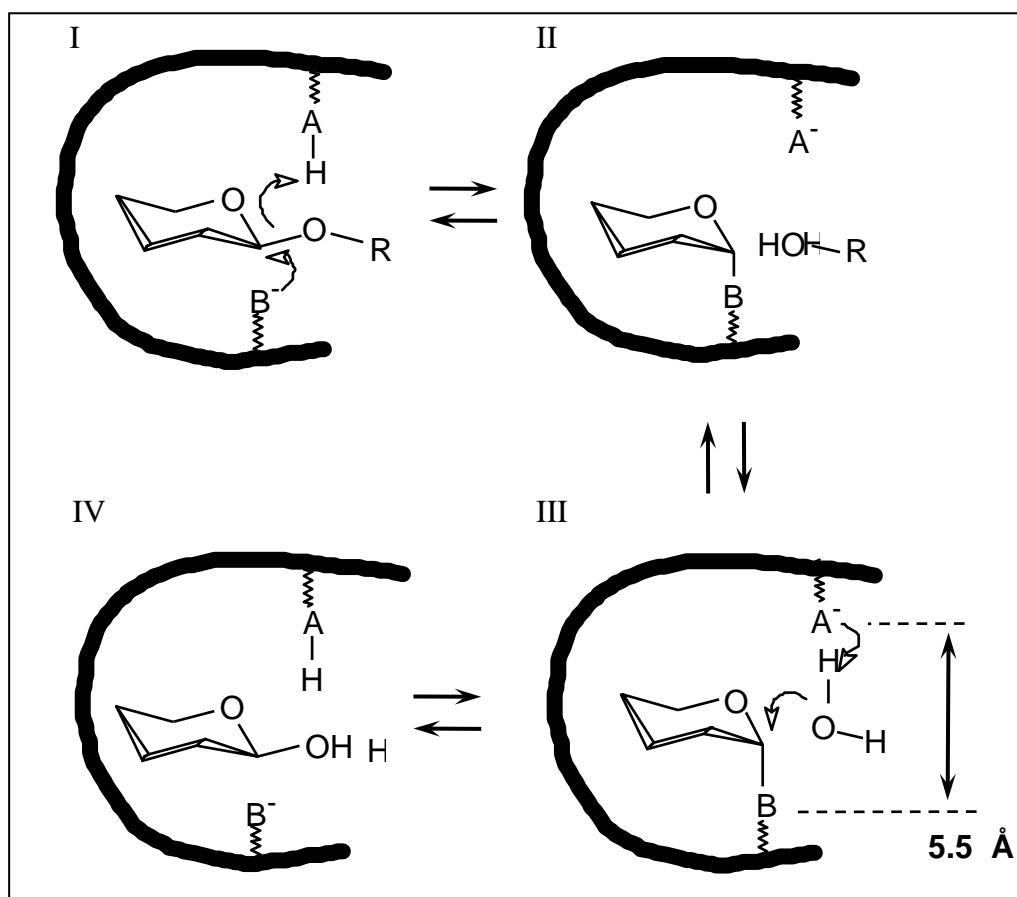


Fig. 8 Mécanisme canonique des glycosidases avec rétention de configuration

5.2. Neuraminidase

La neuraminidase (ou sialidase) du virus de la grippe souche B/Beijing/87, que j'ai étudiée pendant ma thèse a pour rôle de désactiver les récepteurs viraux, en enlevant les résidus d'acide sialique terminaux des glycoprotéines et glycolipides, qui servent comme récepteur viral.

L'étude cristallographique des complexes neuraminidase-acide sialique, le produit de la réaction catalysée (Fig. 9), a révélé les résidus du site actif et leurs interactions avec l'acide sialique (Fig.10, réf. 2). On voit que les deux résidus (Asp 148, Fig. 10 et Glu 275, Fig. 11), qui peuvent être assimilés à des résidus canoniques, sont bien présents. L'acide sialique se trouve dans une conformation bateau contrairement à sa conformation chaise habituelle, ce qui donne déjà une première indication par rapport au mécanisme réactionnel. Après trempage des cristaux de neuraminidase dans une solution de DANA* (Fig. 9, réf. 3) celui-ci a pu être observé dans le centre actif. Le DANA est un inhibiteur de la neuraminidase avec un K_i micromolaire et possède une conformation plane similaire à l'ion oxocarbonium. Cet ion oxocarbonium est un des intermédiaires de la réaction comme on verra ensuite et le DANA représente donc un analogue de l'état de transition.

Nous avons été intrigués par une absence de la densité électronique du groupement hydroxyle à la position 2 qui a été observé sous certaines conditions de trempage des cristaux avec l'acide sialique. En fait le composé lié ressemblait plutôt au DANA ce qui a pu être confirmé dans les mois qui suivaient la soutenance de ma thèse. La production du DANA par la neuraminidase en présence d'un excès d'acide sialique a pu être mis en évidence par la spectrométrie de masse. En plus, nous avons montré que l'enzyme "tourne à vide" sur son produit, l'acide sialique, ce qui veut dire qu'il y a un échange du groupement hydroxyle en position 2 avec le solvant. Ceci a pu être prouvé par une incubation de la neuraminidase avec l'acide sialique en présence d'eau marquée à l'isotope ^{18}O en utilisant la spectrométrie de masse (Fig. 12). Avec une probabilité faible, cette action de l'enzyme sur le substrat mène au produit d'élimination DANA. La capacité de l'enzyme d'agir sur son produit mène à la synthèse d'un inhibiteur qui est fatal pour le fonctionnement de l'enzyme (réf. 3). Il faudrait encore élucider comment l'enzyme évite cette formation de DANA qui serait fatal pour son activité.

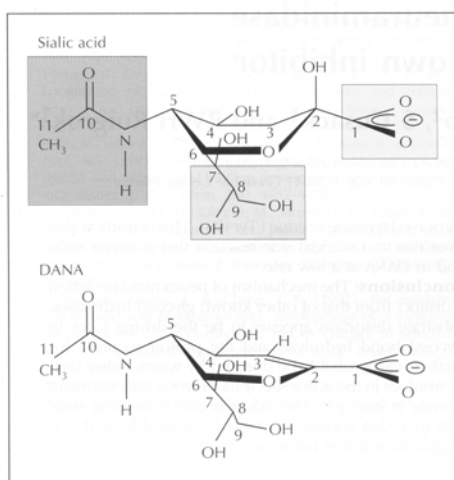


Fig. 9: Structures de l'acide sialique et de l'inhibiteur acide 2,3-deoxy-3-dehydro-neuraminique (DANA). Les trois groupes fonctionnelles: acétamine, glycérol et carboxyle sont encadrés.

* DANA: Acide 2,3-déshydro-2-déoxy-N-acetyl neuraminique

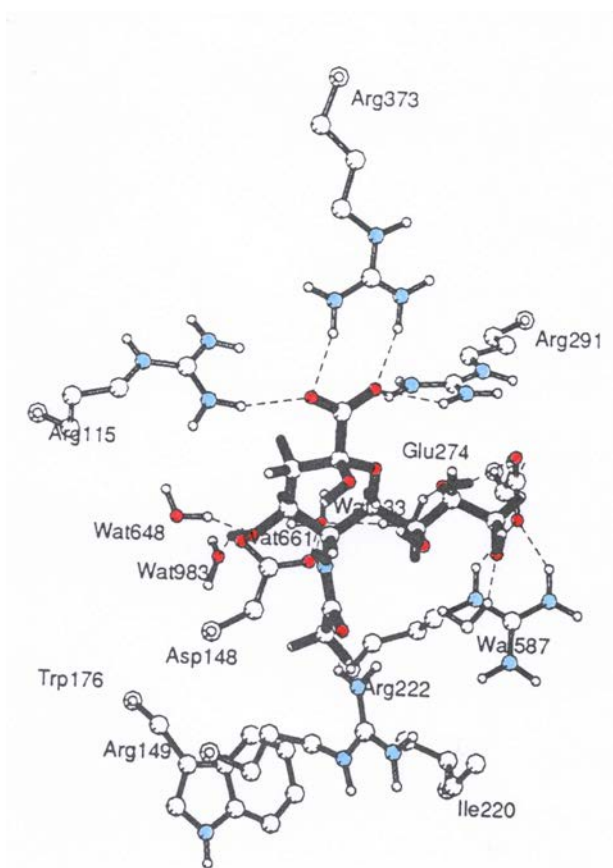


Fig. 10: Reconnaissance de l'acide sialique dans le site actif de la neuraminidase. Les ponts hydrogène sont montrés en pointillé. Atomes d'oxygène en rouge, azote en bleu.

Donc ici, contrairement à la plupart des glycosidases, la présence de l'aglycone n'est pas nécessaire pour un cycle de catalyse. Une activité enzymatique constante pour une variation de pH large souligne que l'état de protonation du résidu "acide-base" n'est pas essentiel et que ce résidu ne semble pas jouer un rôle très important. L'observation du produit d'élimination montre l'importance de l'état de transition de l'ion oxocarbonium, car il est prédisposé à l'élimination d'une molécule d'eau, ce qui forme un composé stable, le DANA. Ici, l'intermédiaire covalent canonique ne semble pas exister. Une reconnaissance très forte entre l'enzyme et l'acide sialique déforme l'acide sialique dans une conformation bateau et favorise l'état de transition plan. Cette distorsion se fait entièrement au niveau du sucre, n'impliquant pas l'aglycone comme pour d'autres enzymes comme le lysozyme. Ceci est possible grâce à la spécificité de l'acide neuraminique: son groupement carboxyle, qui est pris dans une sorte de pince formé par trois résidus d'arginine (Fig. 10).

En conclusion, le mécanisme réactionnel (Fig. 11) est le suivant : Le substrat se lie au site actif dans une conformation bateau, ce qui déstabilise déjà la liaison glycosidique (1). Un donneur de proton, qui peut être une molécule d'eau ou l'acide aspartique 148 à pH bas, va transférer un proton sur la liaison glycosidique. En même temps, une charge négative va se développer sur la tyrosine 408 stabilisant l'état de transition de l'ion oxocarbonium (2). Le résultat est la coupure de la liaison glycosidique et la présence de l'ion oxocarbonium dans le site actif. Une seconde attaque nucléophile par une molécule d'eau va ensuite mener à la formation de l'acide sialique libre (3) qui va changer sa configuration de α à β par mutarotation (avec un équilibre entre les anomères en solution d'environ 95 % β , 5 % α).

Donc, malgré la présence des deux résidus canoniques, le mécanisme réactionnel n'a pas réellement besoin de l'acide-base et ne passe pas par un état de transition covalent.

Avant le début de ma thèse, la structure de la neuraminidase d'une souche A de la grippe avait été résolue. Mais le groupe à l'origine de ce travail était engagé dans une collaboration avec une entreprise pharmaceutique, et, basé sur la structure cristallographique ils étaient en train de développer des inhibiteurs de la neuraminidase, en partant du DANA comme "lead

compound". Des tentatives de collaboration ayant échouées nous avons donc entrepris la détermination de la structure *de novo*. Une bonne partie du travail sur le site actif de la neuraminidase et le mécanisme s'est déroulé en parallèle. Le développement pharmaceutique a mené à un médicament (nom commercial Relenza ou Zanamavir) de Glaxo-Wellcome, basé sur le 4-guanidinyl DANA. Le composé actif est appliqué par l'inhalation d'une poudre en utilisant un inhalateur spécial.

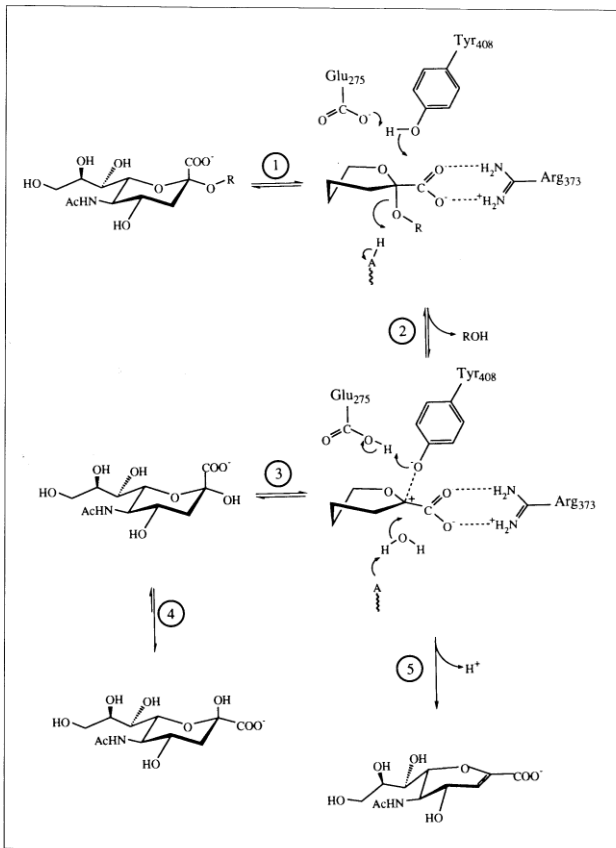


Fig. 11: Mécanisme réactionnel de la neuraminidase.

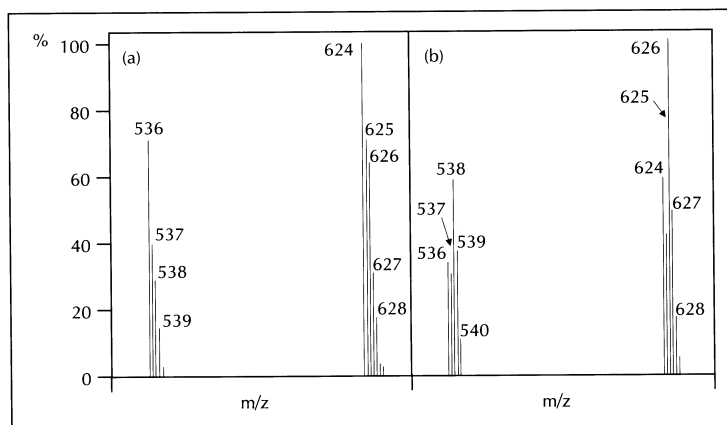


Fig. 12: Spectre de masse de l'acide sialique derivatisé avec TMS dans la région de m/z de 520 à 640 montrant deux pics correspondants à deux fragments contenant l'oxygène O². L'acide sialique a été incubé dans H₂¹⁸O pour 14,5 h à 37 °C soit dans le tampon PBS (gauche) ou en présence de neuraminidase (droite). Les pics les plus hauts sont décalés de 536 et 624 vers 538 et 626 indiquant l'incorporation de l'oxygène marqué.

5.3. Myrosinase

Un autre exemple de glycosidase sur laquelle j'ai travaillé est la myrosinase, qui cette fois-ci est une thioglucosidase, c'est à dire qu'elle agit sur des glucosides dont la liaison glycosidique est constituée d'un atome de soufre. La myrosinase des graines de moutarde a été étudiée pour comprendre quelles sont les spécificités d'une action sur une liaison thioglycosidique. La myrosinase est la seule thioglycosidase connue et elle agit sur des glucosinolates (Fig.13a). L'hydrolyse de la liaison thioglycosidique forme du glucose libre et libère l'aglycone, qui est instable et se réarrange spontanément dans la plupart des cas en sulfate et isothiocyanate (Fig.13b). Cette enzyme ne possède qu'un des deux résidus canoniques, le nucléophile, alors l'acide-base est remplacé par un résidu de glutamine.

La détermination de la structure (réf. 11) a montré que le site actif est quasiment identique au site actif de la β -glucosidase cyanogène, une enzyme du trèfle avec une homologie de séquence assez élevée.

Contrairement à la neuraminidase, nous n'avons pas pu observer une affinité de l'enzyme pour le glucose, un des produits de sa réaction. Par contre, en utilisant un substrat modifié sur la position 2, le 2-désoxy-2-fluoro-glucotropaeolin (Fig. 13c), nous avons pu observer l'intermédiaire covalent, lié par une liaison α -glucosidique à l'acide glutamique Glu 409, le nucléophile. L'électronégativité de l'atome de fluor mène à une stabilisation de l'état de transition et un ralentissement généralisé de la réaction. Cet intermédiaire covalent a montré que tous les groupes hydroxyle sont impliqués dans des interactions avec l'enzyme.

La structure de la myrosinase a révélé aussi une poche hydrophobe, qui est adaptée à l'interaction avec la chaîne latérale hydrophobe de l'aglycone des glucosinolates, et aussi deux résidus d'arginine, qui sont parfaitement placés pour une interaction avec le groupe sulfate du substrat. A l'exception du remplacement du glutamate de l'acide/base général par une glutamine, la machinerie catalytique est identique à celle de la β -glucosidase cyanogène. Le mécanisme catalytique de la myrosinase est expliqué par le caractère de l'aglycone, qui est un très bon groupe partant et qui n'a pas besoin d'une assistance d'un acide aminé pour

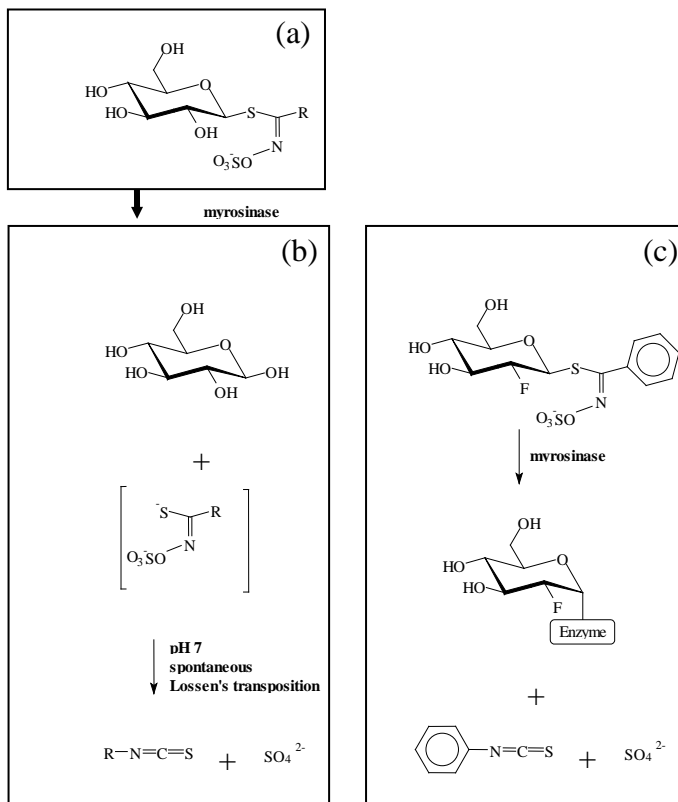


Fig. 13 Glucosinolates et myrosinase. a) Structure des glucosinolates, R peut être par exemple un groupe allyl, hydroxybenzyl ou phényl.

b) Les produits de la réaction

c) Structure de l'inhibiteur 2-f-glucotropaeoline et schéma de la formation de l'intermédiaire covalent.

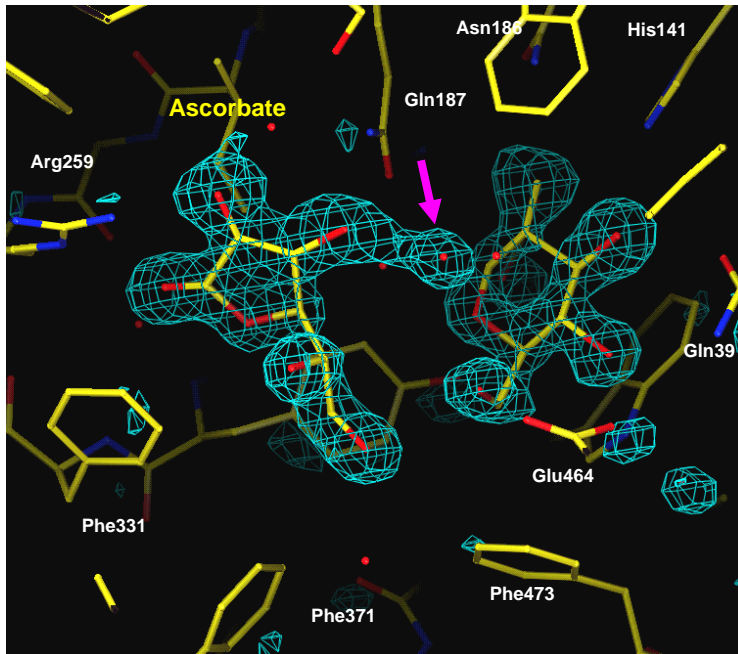


Fig. 14: La molécule d'ascorbate liée dans le site actif de la myrosinase 2-F-glucosylée. Les liaisons hydrogènes reconnaissant l'ascorbate sont montrées. La flèche marque une molécule d'eau placée pour l'attaque nucléophile sur la glucosyl-enzyme et activé par l'ascorbate. La carte de densité électronique F_o-F_c est contournée à 3σ .

l'hydrolyse de la liaison glucosidique. Le remplacement de l'acide/base général des O-glycosidases par un résidu de glutamine suggère que, pour l'hydrolyse de l'enzyme glucosylée lors de la 2^{ème} étape de la catalyse, le rôle de ce résidu soit le positionnement précis d'une molécule d'eau à la place d'une activation par une base. Ces résultats ont pu être confirmés par des complexes avec des analogues de l'état de transition, des composés avec une conformation trigonale à la position de l'atome C2 anomérique qui ont été étudiés à haute résolution (1.2 – 1.6 Å, réf. 18). Un des inhibiteurs, le *D*-gluconhydroximo-1,5-lactame se lie au site actif en même temps qu'un ion sulfate et nous faisons l'hypothèse que cette structure ressemble au complexe entre enzyme et substrat (qui comporte un groupe sulfate).

La myrosinase montre une activation intéressante par l'acide ascorbique qui a été décrite dans la littérature et qui peut atteindre un facteur de 400. Nous avons donc étudié à haute résolution la myrosinase en présence d'acide ascorbique, ainsi que des complexes ternaires entre myrosinase, acide ascorbique et des inhibiteurs (réf. 18). L'acide ascorbique se lie à un site différent du site de liaison du glucose (Fig. 14) mais qui recouvre en partie le site de liaison de l'aglycone des glucosinolates ce qui indique que l'activation a lieu à la deuxième étape de la catalyse, l'hydrolyse de l'enzyme glucosylée. Une présence simultanée des analogues d'état de transition ou du groupe 2-désoxy-2-fluoro-glucosyl et de l'acide ascorbique est possible. Cette activation au niveau de l'hydrolyse de l'intermédiaire covalent a été confirmée par l'étude biochimique de la réactivation de l'enzyme 2-fluoroglucosylée où la vitesse de réactivation augmente d'un facteur 14 en présence d'ascorbate. La structure du complexe ternaire, avec le groupe 2-F-glucosyl et l'acide ascorbique montre qu'une molécule d'eau, qui est très bien placée pour une attaque nucléophile sur l'enzyme 2-fluoroglucosylée, peut être activé par l'acide ascorbique (Fig. 14, 15). L'acide ascorbique est donc recruté au milieu de la réaction et est utilisé pour remplacer la base catalytique absente de la myrosinase.

Donc il y a encore une nouvelle variation sur le thème du mécanisme réactionnel des glycosidases (Fig. 16): Avec un bon groupe partant, l'acide-base généralisé peut simplement être absent pour la première étape de la catalyse, mais un groupe fonctionnel extérieur peut aussi être recruté au milieu de la réaction pour remplacer cette fonction !

Ces exemple montrent aussi la valeur d'une approche expérimentale : Une approche basée uniquement sur la séquence de la protéine n'aurait jamais pu prédire toutes ces variantes du mécanisme réactionnel !

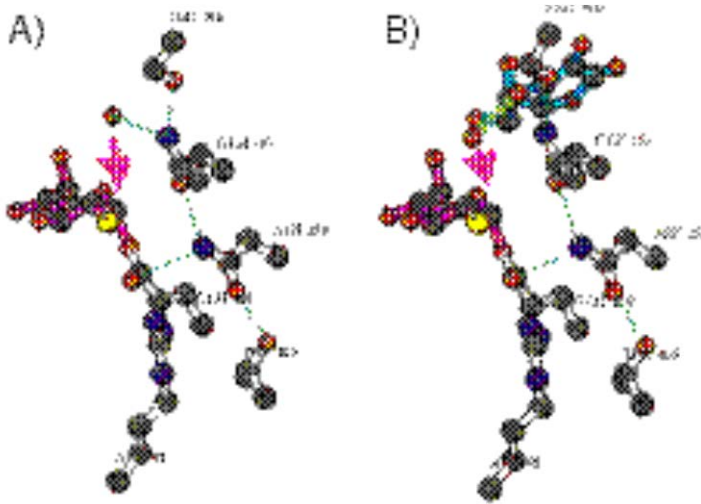


Fig. 15: Le glucosyl-enzyme en absence a) et en présence b) d'ascorbate. Dans a), une molécule d'eau est positionnée par le résidu Gln187 pour l'attaque nucléophile, pendant que la molécule d'eau est activée par l'ascorbate dans b).

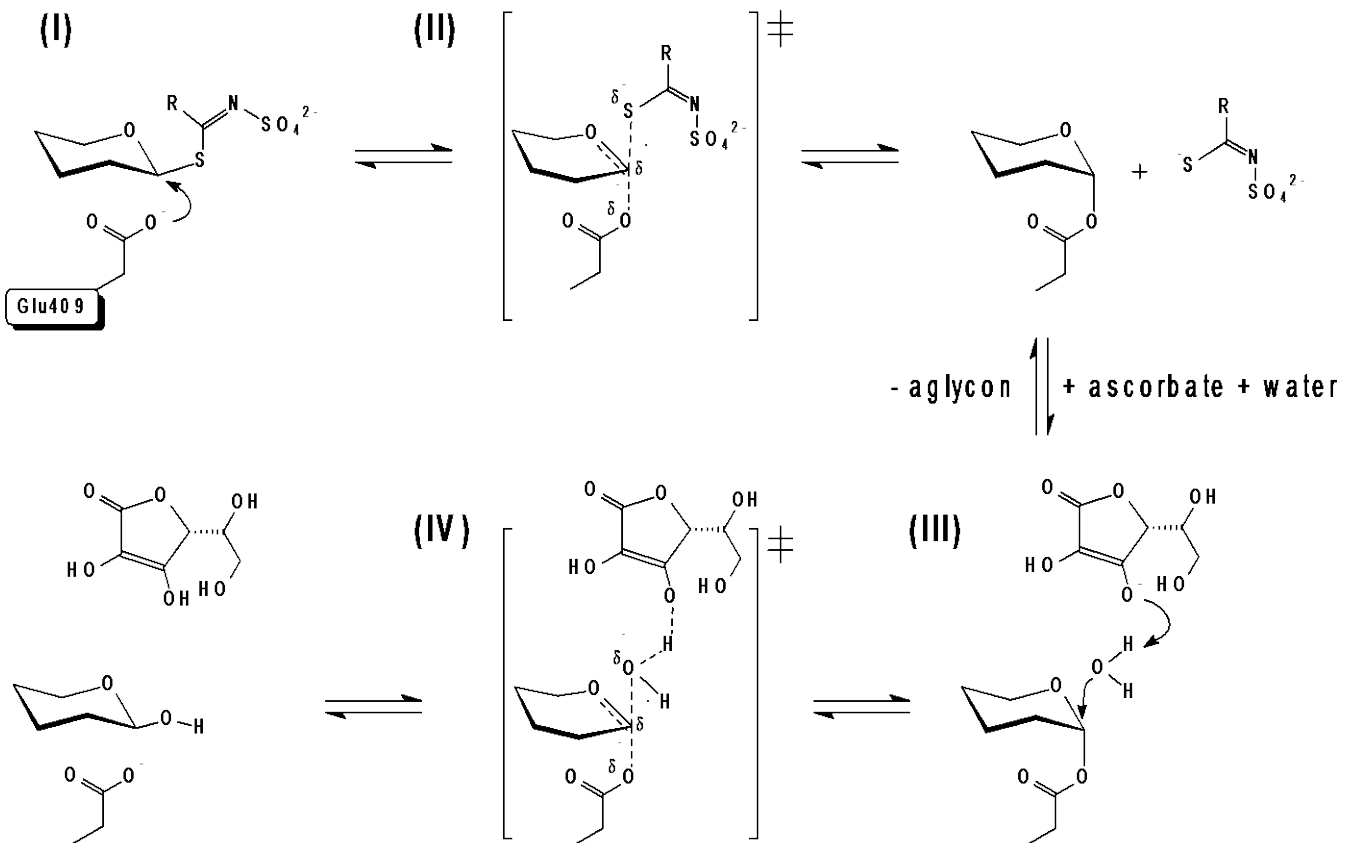


Fig. 16: Mécanisme réactionnel de la myrosinase en présence d'ascorbate

6. COMMENT RELIER STRUCTURE ET FONCTION

6.1. Introduction

Un des aspects frustrants du travail sur les récepteurs ou d'autres protéines impliquées dans les interactions protéine-protéine est l'absence d'information que l'on peut tirer d'une structure tridimensionnelle à elle seule. En effet, tout l'intérêt du travail réside dans le recoupement de la connaissance de la structure avec des informations biochimiques ou d'une étude cristallographique des complexes apportant une réponse directe aux questions par rapport aux interactions des partenaires. Trois sujets de mon travail de recherche rentrent dans cette classe de protéines: le récepteur FcRn, la phosphoprotéine du virus de Sendai et le domaine récepteur de l'adénovirus du sérotype Ad3.

6.2. Récepteur FcRn

Des récepteurs pour la partie Fc des anticorps jouent de rôles nombreux et importants dans le système immunitaire. En particulier, une classe des récepteurs Fc (FcRn) a été décrite : ils transfèrent des anticorps du type IgG (immunoglobuline γ) maternels vers le nouveau-né, et donnent donc une immunité humorale au nouveau-né avant que son système immunitaire soit pleinement développé. Il a été montré ultérieurement, que ce récepteur FcRn est très largement répandu chez les mammifères et sert pour le recyclage des IgG après une endocytose 'accidentelle' de ces molécules par des cellules de l'organisme. Les souris et rats allaités expriment ce récepteur Fc dans leur intestin qui sert pour le transport des anticorps du lait ingérés vers la circulation sanguine. Il lie l'IgG maternel au pH acide du lait (pH 6.5). Les complexes FcRn/IgG sont transportés par transcytose vers le côté basal des cellules où l'IgG se dissocie du récepteur au pH plus basique du sang (pH 7.5). L'affinité des récepteurs FcRn pour l'IgG varie donc de plusieurs ordres de grandeur quand le pH change d'une unité. La séquence primaire de ces récepteurs montre une grande similarité aux séquences des chaînes lourdes des molécules d'histocompatibilité majeure (MHC) de classe I qui servent pour la présentation d'antigènes aux cellules tueuses. Dans les deux cas, la chaîne lourde est associée à une molécule de β -microglobuline pour former le récepteur FcRn ou la

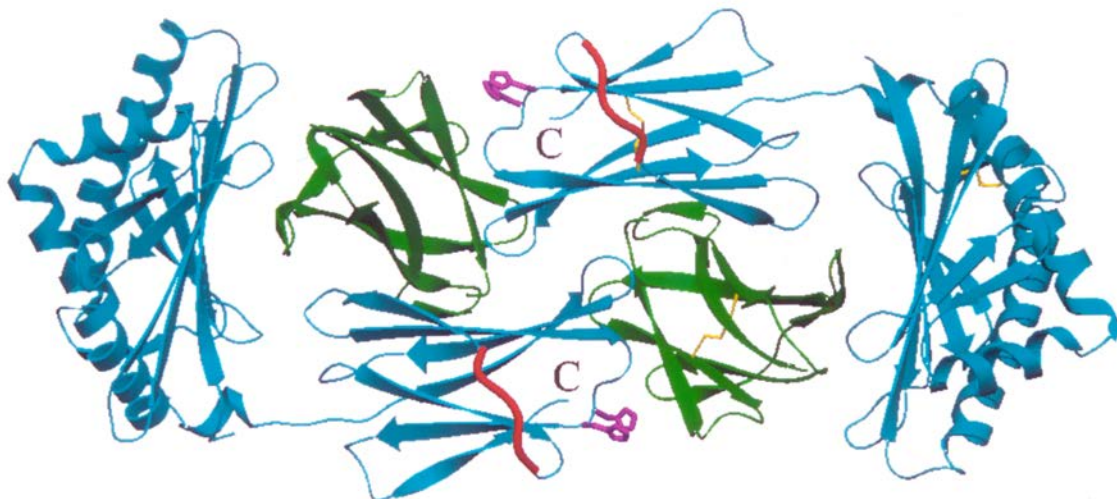


Fig. 17: Vue du dimère du récepteur FcRn. La β -microglobuline est montrée en vert, des résidus d'histidine impliqués dans la dimérisation en magenta, et l'équivalent d'une boucle interagissant avec le co-récepteur CD8 dans les molécules CMH class I en rouge.

molécule CMH classe I. Aussi, la chaîne lourde comporte une hélice transmembranaire et quelques résidus d'acides aminés cytosoliques.

La partie extracellulaire de la protéine a été exprimée au laboratoire dans des bioréacteurs utilisant des cellules CHO* transformées qui secrètent cette molécule dans le milieu de culture. Une purification de celle-ci est facile en se servant d'une colonne portant de l'IgG immobilisée et en utilisant la variation de l'affinité du récepteur fonctionnel en fonction du pH. L'élucidation de ce mécanisme d'affinité dépendant du pH est la motivation principale du travail de résolution de structure. J'ai présenté quelques aspects cristallographiques de ce travail dans le chapitre 4.2.

Un dimère de la molécule a été observé dans chacune des quatre formes cristallines que nous avons obtenues (réf. 6). Dans une des formes, un monomère coexiste avec le dimère. Il était très tentant d'attribuer un rôle à cette dimérisation (Fig. 17), surtout car elle implique deux résidus d'histidine. Comme le changement d'affinité du récepteur pour son ligand est proche du pKa de l'histidine, il est probable que des résidus d'histidine seraient impliqués dans le mécanisme de la protéine. Un équilibre monomère-dimère peut aussi expliquer une partie de la coopérativité nécessaire responsable pour ce changement d'affinité très brutal autour de pH 7.0.

Malgré cette première idée sur le mode d'action, il n'était pas encore possible de faire des conclusions par rapport aux surfaces d'interaction. Nous avons dû attendre la structure d'un complexe entre le récepteur FcRn et le ligand Fc pour résoudre cette question (réf. 7). Le domaine Fc a été obtenu par digestion protéolytique des anticorps IgG, qui produit le fragment variable Fab (un hétérodimère) et le fragment constant Fc (un homodimère). Mais les cristaux du complexe, malgré une cristallisation facile, ne diffractaient qu'à 4.5 Å de résolution. Une image détaillée de l'interaction n'était donc pas disponible. Cette résolution a quand même permis de placer la structure d'un fragment Fc, connu, et celle du récepteur FcRn dans l'unité asymétrique du cristal. Avec les axes de symétrie cristallographique d'ordre 2 et le caractère dimérique des deux partenaires, nous avons obtenu un réseau compliqué d'interactions où il était difficile d'identifier les unités biologiques interagissantes. Nous avons retenu une interaction entre un dimère de FcRn et un dimère de Fc (Fig. 18). Ce complexe devrait être asymétrique car un rapport 1:1 des monomères de FcRn:Fc a été observé en solution, alors que le dimère de FcRn présente deux sites de liaison. Une analyse de la surface d'interaction montrait qu'un grand nombre de résidus d'histidine du fragment Fc, assez largement conservés pour les différents sous-types, a été utilisé. Il semble donc que les résidus contribuant à la dépendance du pH de l'affinité du récepteur se trouvent surtout coté ligand, excepté des résidus d'histidine impliqués dans l'équilibre monomère-dimère de



Fig. 18: Vue du complexe de FcRn (vert et bleu) – Fc (rouge). Le dimère de FcRn interagit avec un dimère de Fc d'une façon asymétrique. L'unité asymétrique cristallographique des cristaux du complexe ne contient qu'un monomère de FcRn et un monomère de Fc.

* CHO : Cellules Ovaires du Hamster Chinois

FcRn.

Le mode d'interaction de FcRn avec Fc est complètement différent de l'interaction de la molécule MHC classe I avec le peptide, le récepteur des cellules tueuses (TcR) et le co-récepteur CD8. Par ailleurs, sur le fragment Fc, le site d'interaction situé entre les deux domaines d'immunoglobuline de chaque chaîne, est similaire au site utilisé par d'autres protéines liant la partie Fc de l'anticorps.

Par la suite, le modèle du fonctionnement du récepteur que nous avons proposé a été la base de nombreuses expériences en biochimie et biologie moléculaire. Ces expériences ont confirmé le modèle basé sur la structure que nous avons proposé à l'époque.

6.3. Phosphoprotéine du virus de Sendai

Le virus de Sendai est un virus de la famille de paramyxovirinae qui infecte des rongeurs, surtout des souris. Il est devenu le virus modèle de cette famille de virus, qui contient des membres importants comme le virus de la rougeole, de la rinderpest, le virus des oreillons et des virus parainfluenza. Ce sont des virus enveloppés par une membrane. Leur information génétique est présente sous forme d'une molécule d'ARN à brin simple protégé par la nucléoprotéine N (Fig. 19a), avec 6 bases liées par une molécule de nucléoprotéine. Ce complexe est la matrice pour l'ARN polymérase ARN-dépendante du virus. La phosphoprotéine P est présente à l'intérieur des cellules infectées par le virus de Sendai, et à ce jour, deux rôles lui sont connus : Elle agit comme chaperonine pour stabiliser la nucléoprotéine du virus nouvellement produite et surtout pour l'empêcher de lier l'ARN cellulaire d'une façon non-spécifique. Enfin, c'est un cofacteur essentiel de la polymérase lors de la transcription du génome viral. Elle interagit en même temps avec la nucléoprotéine du complexe ribonucléique du virus et avec la polymérase, et elle permet la progression de celle-ci lors de la réplication du génome.

La protéine P comporte plusieurs domaines (Fig. 19b) et il a été montré que la partie C-terminale (Cterm) est suffisante pour le rôle dans la transcription du génome, alors qu'une séquence coté N-terminus est essentielle pour l'activité de chaperonine.

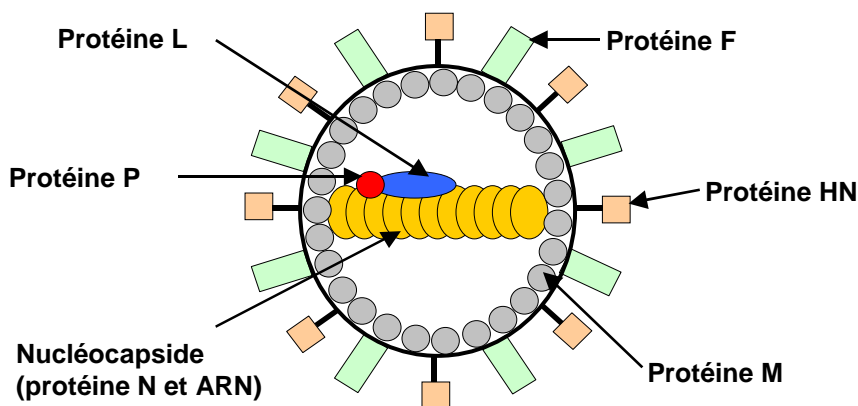
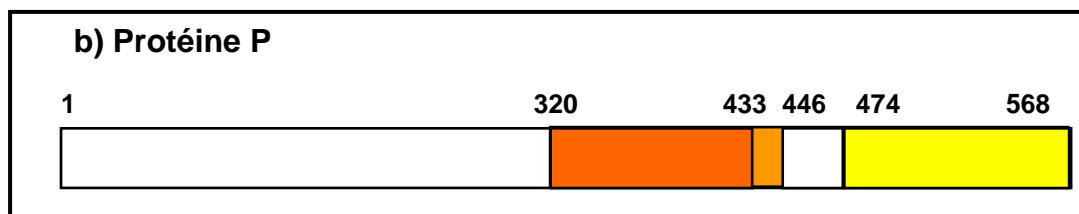


Fig. 19: a) Schéma du virus de Sendai b) Schéma de la protéine P, les différentes constructions (fragments) utilisées : Cterm : 320-568, X : 475-568, PM : 320-446, PM' : 320-433



Le travail sur cette protéine était le sujet de thèse de N. Tarbouriech, mon étudiant en thèse à l'ESRF. Du fait de l'impossibilité d'obtenir des cristaux de la partie C-terminale intégrale, nous avons dû diviser ce domaine en deux, le domaine de multimérisation (PM) et le domaine de la protéine X lors de l'étude structurale. Cette protéine X est produite dans les cellules infectées par un site d'entrée ribosomale dans l'ARN messager de P. La détermination de sa structure par RMN est en cours en collaboration avec l'Institut de Biologie Structurale (IBS) à Grenoble (réf. 22). La partie de la protéine P responsable pour la multimérisation a pu être cristallisée et la structure a été résolue par une combinaison des méthodes MIRAS et MAD après l'élimination de quelques acides aminés coté C-terminus (PM'). Des études biophysiques (réf. 15) comme la diffusion des rayons X et des neutrons des petits angles et les mesures de dichroïsme circulaire, ont montré que les deux domaines, le domaine de la protéine X et le domaine de multimérisation PM, se replient d'une façon indépendante et que les propriétés du fragment C-terminal (Cterm) total sont la somme des propriétés des deux domaines, en négligeant la partie considéré comme "linker" laisse en blanc dans Fig. 19b. Dans ce genre d'approche réductionniste il se pose toujours la question d'un repliement indépendant des fragments, que nous avons pu confirmer dans ce cas.

La partie de la protéine dont la structure a finalement été résolue (Fig. 20) montre une très longue hélice α quadruple responsable de la tétramérisation et un petit domaine globulaire. Cette tétramérisation avait déjà été prédite par des études biochimiques (réf. 15). A ce stade, les implications fonctionnelles restent hypothétiques. En analysant des mutants décrits dans la littérature, nous avons vu qu'un certain nombre d'entre eux affecte l'extrémité N-terminale de la quadruple hélice, un domaine qui a été décrit comme le site de liaison de la polymérase L. Ces résidus affectent la synthèse d'ARN, soit en réduisant la processivité (la polymérase arrivant moins bien à traverser les régions intergéniques) soit en réduisant simplement la quantité d'ARN produite. Au niveau de la structure, ces mutations sont soit placées dans le noyau hydrophobe de la quadruple hélice, qui pourrait s'ouvrir plus facilement à son extrémité C-terminale, soit elles affectent des résidus de surface chargés. La forme du domaine de multimérisation suggère qu'elle fonctionne peut-être comme un axe moléculaire,

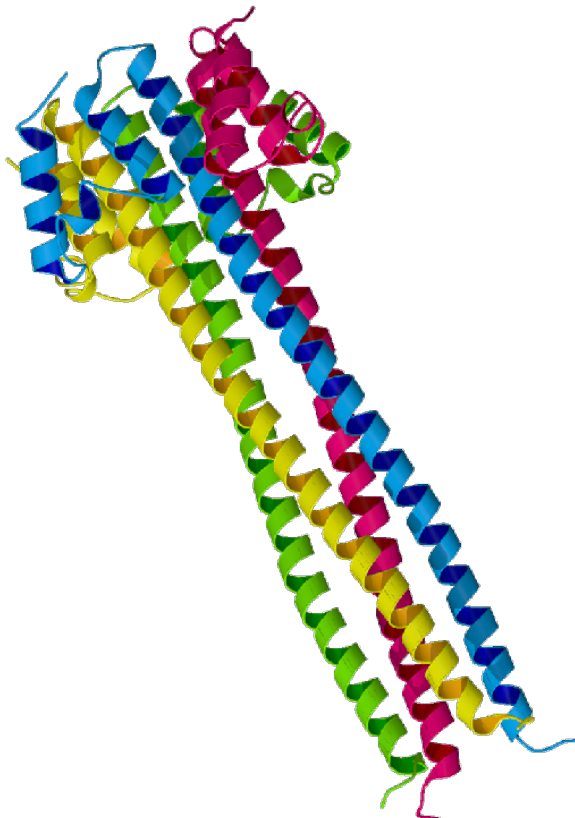


Fig. 20: Représentation en rubans de la structure du tétramère du domaine de multimérisation de la protéine P du virus de Sendai.

liant la progression de la polymérase d'une nucléoprotéine à une autre après la transcription de 6 bases avec une rotation continue lors de la progression de la transcription le long de ces 6 nucléotides qui sont liés à chaque molécule de nucléoprotéine. Les domaines X attachés, semblaient-ils par des attaches "linker" flexibles, peuvent peut-être se lier sur des nucléoprotéines successives en faisant la roue (terme anglais "to cart-wheel") le long les protéines N du complexe nucléoprotéine - ARN.

Mais finalement, nous sommes encore très loin d'une compréhension réelle du fonctionnement de la protéine P, et d'autres études fonctionnelles et structurales sur les complexes entre la nucléoprotéine N, l'ARN et la protéine P seront nécessaires.

6.4. Récepteur de l'adénovirus Ad3

Les adénovirus sont des virus à ADN non-enveloppés. Ils sont à la cause d'infections respiratoires, gastro-intestinales et oculaires aiguës, particulièrement chez l'Homme mais aussi chez certains animaux. Les virus humains comportent plus de 50 types sérologiques différents. Les particules sont composées d'une capsidie icosaédrique formée de la protéine hexon, présente en 240 copies, et de 12 copies de la base du penton pentamérique dans laquelle la fibre trimérique est insérée (Fig. 21). La fibre comporte une tige formée d'un motif en feuillet β caractéristique et du domaine récepteur globulaire à son extrémité C-terminale. Ce domaine est responsable pour l'attachement des particules virales à des cellules cibles, alors que la base du penton porte une séquence RGD liant des intégrines nécessaires pour l'internalisation du virus. Pour de nombreuses souches, la molécule CAR* a été identifiée comme ligand de la fibre. Nous travaillons sur le sérotype Ad3, qui lie un récepteur encore

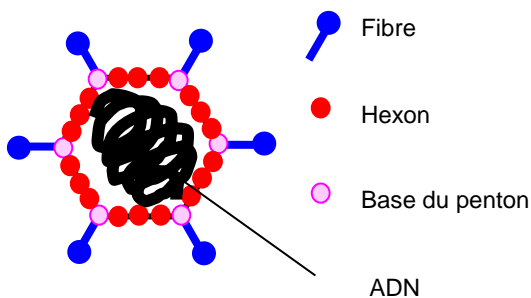


Fig. 21: Schéma de l'adénovirus

inconnu.

Un travail de détermination de la structure à 1.6 Å de résolution très rapide (réf. 21) a été suivi d'un travail d'interprétation plus long qui n'est toujours pas achevé. La structure globale du domaine récepteur (Fig. 22) est très semblable aux structures connues antérieurement, dont une, celle d'Ad12 est aussi connue en complexe avec le ligand CAR. Toutes les différences entre ces structures concernent la conformation des différentes boucles de surface. Nous nous sommes intéressés à la localisation du site de liaison du ligand.

Des analyses de l'homologie des séquences ne sont pas très fructueuses, car à part quelques sérotypes liant le CAR, les sérotypes qui partagent le même ligand ne sont pas connus. Il n'est donc pas possible d'associer la conservation de certains résidus à un ligand donné. D'autant plus qu'il existe une grande variation entre les résidus du site d'interaction des sérotypes liant la molécule CAR. Cette variabilité est certainement due à la pression antigénique sur le site de reconnaissance. Celui-ci n'est pas à l'abri de la réponse immunitaire, contrairement à d'autres

* Coxsackie and Adenovirus Receptor



Fig. 22: Structure du domaine récepteur d'Ad3.
Les feuillets β sont colorés en bleu, les hélices α en rouge.

virus qui enfouissent leur site de reconnaissance du ligand pour le mettre à l'abri des anticorps.

La seule information qu'on a pu extraire de la structure était basée sur la conformation des boucles formant le site d'interaction avec CAR. Si on regarde la conformation des trois sérotypes liant CAR (Fig. 23) on voit que la conformation de la partie N-terminale de la boucle AB est essentiellement conservée, formant ensemble avec le résidu lysine 429 le centre de l'interface. D'une façon surprenante, cette conformation est aussi conservée pour Ad3, qui lie, pour mémoire, un ligand inconnu. Par contre, le degré de variation pour les autres boucles est similaire, avec l'exception de la boucle EG, où Ad3 possède une conformation plus ouverte. Une conclusion peut être que le ligand d'Ad3 se lie à la même position mais qu'il est légèrement plus grand que le domaine de la molécule CAR.

Une autre piste pour l'interprétation de la structure est l'hydrophobicité de la surface. La surface du récepteur Ad3 est relativement hydrophobe (Fig. 24). On observe en particulier une crevasse avec des résidus hydrophobes exposés propices à une interaction avec un ligand, qui

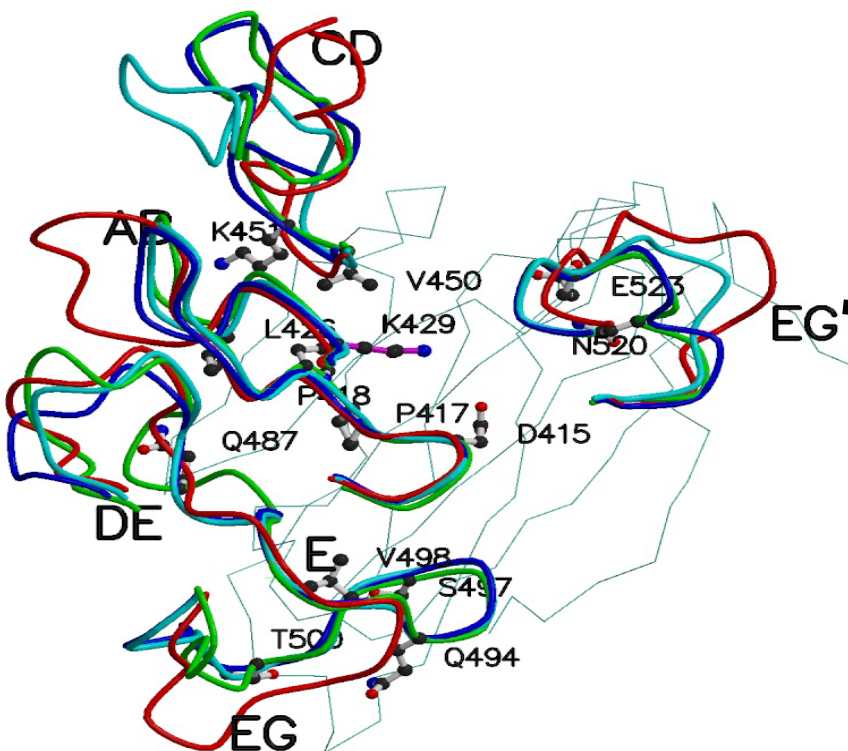


Fig. 23: Boucles impliquées dans l'interface entre le domaine récepteur d'Ad12 et CAR et leurs équivalents dans les autres sérotypes.

Rouge: Ad3; Bleu: Ad5; Cyan: Ad2; Vert: Ad12. La trace $C\alpha$ de CAR est montrée en trait. La numérotation se réfère à Ad12, dont les résidus sont montrés. EG' fait partie de la boucle EG d'un autre monomère du trimère.

n'est pas nécessairement de nature protéique. Sans l'identification du ligand ou des expériences de mutagenèse dirigée modifiant les résidus clés, on ne pourra pas trancher entre ces deux hypothèses. Les deux propositions sont même compatibles, car il n'est pas exclu que le domaine récepteur d'Ad3 lie différents types de ligand. Par conséquent, ce projet se poursuit avec des travaux de la recherche du ligand de cette protéine.

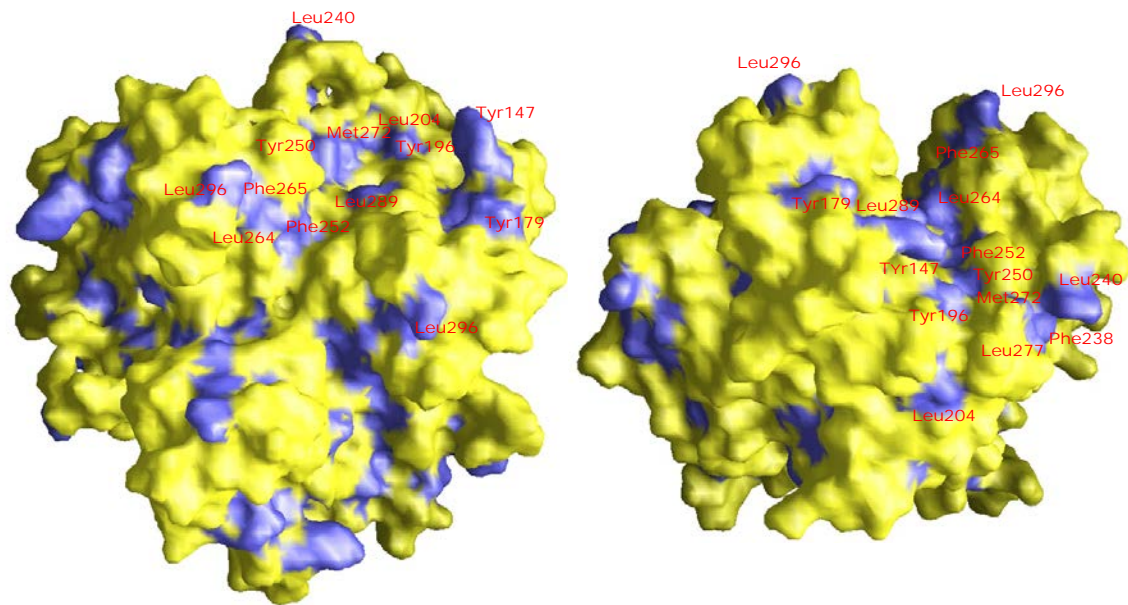


Fig. 24: Surface du domaine récepteur d'Ad3. Vue du dessus et de côté. Les chaînes latérales des résidus hydrophobes (Trp, Tyr, Phe, Leu, Ile, Val, Met, Cys) sont colorées en bleu.

7. CONCLUSION ET PERSPECTIVES

Le développement technique est en train de diviser le domaine de la cristallographie biologique et bientôt il y aura deux approches en cristallographie des protéines :

Résoudre des structures ou résoudre des problèmes.

La première approche est très en vogue et se présente sous la forme de la "génomique structural". Une bonne partie des opérations nécessaires pour la détermination d'une structure se prêtent à l'automatisation :

Clonage des gènes et expression des protéines, purification des protéines et leur cristallisation, manipulation de l'échantillon, collecte de données et détermination de la structure. Les premiers problèmes peuvent être résolus par des systèmes robotiques ; une fois les données collectées, essentiellement avec la méthode MAD, il est possible de résoudre une structure quasi-automatiquement, en associant les logiciels de solution de structure (*i.e.* SOLVE) avec des logiciels de traçage automatique (*i.e.* Arp/Warp). Cependant, il y existe de nombreuses difficultés qui peuvent être rencontrées lors de ce processus et qui mènent à l'échec de l'automatisation: Manque d'expression de la protéine ou manque de solubilité, difficultés de purification, impossibilité d'obtenir des cristaux, problèmes de macles, problèmes cristallographiques (*i.e.* pseudosymétries) dans le cristal ou une qualité des cristaux insuffisante. La résolution d'une structure est donc largement aléatoire et il est peu probable d'obtenir de cette façon une structure qui répond à un problème particulier. Dans les projets pilotes sur des protéines bactériennes solubles, on obtient la structure d'environ une protéine

sur 5 qui rentre dans le processus*. Evidemment, en multipliant les systèmes d'expression utilisés, en incluant des procédures de renaturation ou en variant les méthodes de purification on peut augmenter le taux de réussite. Mais en même temps, l'approche devient de plus en plus lourde.

Par opposition à cette démarche il y a l'approche plus classique, pour ne pas dire artisanale, de résoudre une structure pour résoudre un problème. Il est possible d'être aidé par des techniques décrites précédemment, et dans l'avenir il serait essentiel d'intégrer ces approches, mais la différence reste de ne pas abandonner le projet à la première difficulté. C'est la question à résoudre qui justifie un investissement plus grand dans la structure. Pour surmonter ces problèmes évoqués, il est possible de faire appel à de nombreuses techniques physiques et biochimiques pour caractériser la protéine (diffusion de lumière, sédimentation, dichroïsme circulaire, protéolyse limitée, pontage chimique, spectrométrie de masse, diffusions aux petits angles, mutagénèse dirigée, etc.), dont un bon nombre a été présenté dans les chapitres précédents ou dans les publications qui en font la base. Tout ce savoir supplémentaire augmente considérablement les chances d'aboutir dans la résolution de notre structure. Ces informations peuvent aider à modifier la protéine par une mutagénèse pour surmonter des problèmes de cristallisation.

Une fois les cristaux obtenus (*cf.* chapitre 4.2), un grand nombre d'outils est disponible pour résoudre les problèmes cristallographiques, qui se présentent pour chaque cristal sous une forme différente.

Quand finalement la structure a été résolue, c'est le moment de l'interpréter : Il faut encore passer par un autre cycle d'utilisation des techniques complémentaires (mutagenèse, résonance de plasmons de surface, spectrométrie de masse etc.) pour apporter une réponse à la question posée. Ensuite, il faut adresser la question de l'interaction avec des ligands et nous sommes amenés vers la résolution d'une autre structure, celle du complexe de protéines ou d'une protéine avec l'ARN ou l'ADN. Celle-ci, en général est encore plus difficile.

Dans un tel travail les approches de la génomique structurale ont aussi leur place : Beaucoup de manipulations de routine en amont du travail cristallographique, surtout en biologie moléculaire, dans la purification des protéines et dans la cristallisation se prêtent à l'automatisation. Pour rester compétitif, dans l'avenir, tout laboratoire de recherche devrait centraliser et automatiser ces opérations, pour que les chercheurs puissent se concentrer sur l'essentiel du travail scientifique.

Dans cette vision du travail en biologie structurale, dans une pluridisciplinarité, je souhaite développer des projets de recherche dans le domaine des protéines virales. J'aimerais que ce travail s'intègre dans le cadre du futur Institut de Virologie Structurale et Fonctionnelle où je souhaite former des étudiants en thèse car je pense que la biologie structurale a encore des jours radieux devant elle.

* Projet pilote de génomique structural S2F CARB/TIGR sur *Haemophilus influenzae* (en mai 2001, 13 structures pour 76 protéines exprimées)