



**HAL**  
open science

## Structure du récepteur pour l'antigène du lymphocyte T. Structures de protéines à très haute résolution.

Dominique Housset

► **To cite this version:**

Dominique Housset. Structure du récepteur pour l'antigène du lymphocyte T. Structures de protéines à très haute résolution.. Biologie structurale [q-bio.BM]. Université Joseph Fourier, 1999. tel-03032705

**HAL Id: tel-03032705**

**<https://hal.univ-grenoble-alpes.fr/tel-03032705>**

Submitted on 1 Dec 2020

**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.

Mémoire de Diplôme d'Habilitation à Diriger les Recherches

Comité de Physique

Université Joseph Fourier – Grenoble

Structure du récepteur pour l'antigène du  
lymphocyte T

Structures de protéines à très haute résolution.

par Dominique HOUSSET

Soutenu le 28 octobre 1999

Composition du jury de soutenance :

Pr. R. FOURME

Pr. C. LECOMTE

Dr. P. ALZARI

Pr. E. PEBAY-PEROULA

Dr. G. MAZZA

## Résumés

### ***A - Structure du récepteur pour l'antigène du lymphocyte T.***

Le processus de présentation de l'antigène (Ag) à la cellule T est un événement clef de la réponse immunitaire cellulaire. Un tel processus implique trois partenaires principaux: pour les cellules présentatrices d'Ag, une protéine codée par le complexe d'histocompatibilité majeur (MHC), un peptide antigénique lié à ce dernier, et pour les cellules T, un récepteur (TCR). La compréhension de la présentation de l'antigène, à l'échelle moléculaire, a largement progressé ces dernières années, notamment grâce aux structures de plusieurs molécules de MHC déterminées par les équipes de P. Bjorkman et D. Wiley. Ces études ont montré que le MHC possède une cavité qui lie le peptide antigénique et le présente au TCR. Le TCR est composé de deux chaînes,  $\alpha$  et  $\beta$  de type Immunoglobuline, contenant chacune un domaine variable, un domaine constant et une région transmembranaire.  $V\alpha$  et  $V\beta$ , les domaines variables du TCR se fixent à la région du MHC qui lie l'antigène. Récemment, la structure du TCR 2C (murin) et la structure du complexe TCR A6/peptide/HLA-A2 humain ont été déterminées par les groupes de I. Wilson et D. Wiley. La connaissance de plusieurs TCR et complexes, reconnaissant le même MHC de type I, mais des peptides différents, devrait permettre d'élucider le mécanisme de reconnaissance spécifique d'un peptide.

En collaboration avec le groupe de Bernard Malissen (CIML, Marseille), nous avons très récemment déterminé la structure d'un domaine variable de TCR (KB5-C20), complexé à un Fab monoclonal anticlonotypique (Désiré-1). Cette structure nous a montré que certaines régions CDR devaient changer de conformation lors la fixation au complexe MHC/peptide. La structure de plusieurs complexes TCR KB5-C20/peptide/MHC H-2K<sup>b</sup>, en cours d'étude, devrait donc nous apporter des informations nouvelles sur le mode de reconnaissance du peptide et du MHC et la réactivité croisée du TCR.

### ***B - Structures de protéines à très haute résolution.***

Dans certains cas favorables, des données de diffraction à résolution atomique ( $d < 1.2$  Å) ou sub-atomiques ( $d < 0.9$  Å) peuvent être obtenues pour des cristaux de protéines. Quelles informations supplémentaires peut-on extraire de ce type de données ? Et quels outils faut-il développer ? La structure de la toxine II du scorpion *Androctonus australis* Hector, affinée à très haute résolution (0,96 Å), a été l'occasion de tester un certain nombre d'approches. Une telle résolution, rarement atteinte pour une protéine, a permis d'obtenir une bonne estimation de l'anisotropie de mouvement des atomes et de localiser un grand nombre d'atomes d'hydrogène. Les données de diffraction ont permis de tester avec succès une méthode de résolution de structure *ab initio*, fondée sur la technique « bake & shake ». L'utilisation d'un modèle de densité électronique atomique non-sphérique, fondé sur un développement multipolaire de la densité électronique (collaboration avec l'équipe de C. Lecomte, Université de Nancy), a permis de faire un point précis des détails observables à cette résolution, et des perspectives pour des résolutions plus élevées.

**PARCOURS DE RECHERCHE.....I**

<b>INTRODUCTION.....</b>	<b>I</b>
<b>PROJETS RÉALISÉS DEPUIS LA THÈSE.....</b>	<b>I</b>
STRUCTURE DE MUTANTS DE L'INHIBITEUR DE LA TRYPSINE BOVINE.....	I
STRUCTURE DE DEUX ASPARAGINASES, UNE ENZYME UTILISÉE COMME AGENT THÉRAPEUTIQUE CONTRE CERTAINES LEUCÉMIES.....	II
STRUCTURE DE LA FERRÉDOXINE I DE DESULFOVIBRIO AFRICANUS.....	II
STRUCTURE DE LA 17 $\beta$ -HYDROXYSTÉROÏDE DEHYDROGÉNASE HUMAINE DE TYPE 1.....	III
STRUCTURE D'UNE TOXINE DE SERPENT, LA FASCICULINE.....	III
STRUCTURE D'UNE AMINOPEPTIDASE DE LACTOCOCCUS LACTIS : PEP <sup>C</sup> .....	IV
<b>LISTE DES PUBLICATIONS .....</b>	<b>V</b>

**A STRUCTURE DU RÉCEPTEUR POUR L'ANTIGÈNE DU LYMPHOCYTE T ..... 1**

<b>A.1 INTRODUCTION GÉNÉRALE .....</b>	<b>1</b>
A.1.1 PATHOLOGIES IMPLIQUANT LE LYMPHOCYTE T.....	2
A.1.2 RECONNAISSANCE SOI/NON-SOI.....	2
A.1.3 PROBLÈME DE DÉFINITION DU SOI.....	3
<b>A.2 CONTEXTE DE L'ACTION DES LYMPHOCYTES T.....</b>	<b>3</b>
A.2.1 APERÇU GÉNÉRAL DES PROTAGONISTES DU SYSTÈME IMMUNITAIRE.....	3
A.2.2 UN SCÉNARIO EXTRÊMEMENT SIMPLIFIÉ D'UNE AGRESSION.....	6
<b>A.3 LES PROTAGONISTES DE LA RÉPONSE CELLULAIRE.....</b>	<b>7</b>
A.3.1 LES MOLÉCULES MHC .....	7
A.3.2 LE TCR.....	8
A.3.3 RÔLE DE L'INTERACTION TCR $\alpha\beta$ /pMHC DANS LES DIFFÉRENTS STADES DE DÉVELOPPEMENT DU LYMPHOCYTE T.....	10
A.3.4 DE L'INTERACTION DES LYMPHOCYTES T AVEC UNE CELLULE À AU SIGNAL D'ACTIVATION. 11	
A.3.5 SIGNAL D'ACTIVATION DU LYMPHOCYTE T .....	11
<b>A.4 STRUCTURE DES DIFFÉRENTS PARTENAIRES DE LA RECONNAISSANCE .....</b>	<b>13</b>
A.4.1 LA STRUCTURE DU TCR .....	15
A.4.2 LES MOLÉCULES MHC .....	16
A.4.3 L'INTERFACE TCR / pMHC .....	17
<b>A.5 LA STRUCTURE DU TCR KB5-C20.....</b>	<b>19</b>
A.5.1 INTERFACE DU COMPLEXE CLONOTYPE (TCR KB5-C20) / ANTI-CLONOTYPE (FAB DÉSIRÉ-1) 20	
A.5.2 STRUCTURE DU SCFV.....	21
A.5.3 LE "SWITCH" DU BRIN $\beta$ C".....	22
A.5.4 UNE BOUCLE CDR3 $\beta$ PARTICULIÈREMENT LONGUE.....	24
A.5.5 INTERACTION TCR/PEPTIDE/MHC: UN MODÈLE DU COMPLEXE KB5-C20 /KB1/ H-2K <sup>B</sup> 24	
<b>A.6 ELUCIDER LE LIEN ENTRE LES INTERACTIONS TCR/pMHC ET L'ACTIVATION DE LA CELLULE T.....</b>	<b>26</b>
A.6.1 BASES STRUCTURALES DE LA RÉACTIVITÉ CROISÉE DU TCR.....	26
<b>A.7 PERSPECTIVES .....</b>	<b>28</b>
<b>A.8 BIBLIOGRAPHIE .....</b>	<b>29</b>

<b>B</b>	<b>LES STRUCTURES DE PROTÉINES À TRÈS HAUTE RÉOLUTION.....</b>	<b>36</b>
<b>B.1</b>	<b>INTRODUCTION.....</b>	<b>36</b>
<b>B.2</b>	<b>APPORT D'UNE RÉOLUTION ATOMIQUE OU SUBATOMIQUE.....</b>	<b>37</b>
B.2.1	LES AVANTAGES MÉTHODOLOGIQUES.....	37
B.2.2	L'INTÉRÊT BIOLOGIQUE.....	38
B.2.3	LES MOYENS NÉCESSAIRES.....	38
B.2.4	LES DIFFICULTÉS MAJEURES.....	39
<b>B.3</b>	<b>ÉTAT DE L'ART.....</b>	<b>40</b>
B.3.1	DES PIONNIERS ...	40
B.3.2	... AUX RÉSULTATS ACTUELS.....	40
<b>B.4</b>	<b>UN NOUVEAU MODÈLE DE DENSITÉ ÉLECTRONIQUE ATOMIQUE.....</b>	<b>42</b>
B.4.1	UN TRANSFERT DU SAVOIR-FAIRE ISSU DES PETITES MOLÉCULES.....	43
B.4.2	QUELQUES EXEMPLES D'APPLICATION.....	43
<b>B.5</b>	<b>LIMITES ACTUELLES.....</b>	<b>53</b>
<b>B.6</b>	<b>PERSPECTIVES.....</b>	<b>54</b>
<b>B.7</b>	<b>DÉFINITIONS ET ABRÉVIATIONS:.....</b>	<b>54</b>
<b>B.8</b>	<b>BIBLIOGRAPHIE.....</b>	<b>55</b>

# Parcours de Recherche

## Introduction

La biologie structurale, et plus particulièrement la cristallographie des protéines sont au centre de mon activité scientifique depuis la fin de ma thèse. Après une formation universitaire en physique, la thèse a été l'occasion d'un premier contact avec la biologie des protéines, par une approche théorique, en étudiant leurs propriétés dynamiques par simulation sur supercalculateurs. Cette étude m'a permis d'apprécier la grande difficulté de modéliser les molécules complexes que sont les protéines, et l'importance de valider au maximum les résultats des simulations par l'expérience.

Par la suite, mon travail a pris une orientation résolument expérimentale, en abordant la détermination de structures de protéines par cristallographie aux rayons X et l'analyse des relations structure tridimensionnelle - fonction biologique. Avec pour point commun la technique expérimentale, la cristallographie macromoléculaire, les différents projets réalisés au cours des sept dernières années m'ont permis d'aborder plusieurs thèmes de biologie structurale (stabilité du repliement, catalyse enzymatique, interactions protéine-protéine et protéine-substrat, reconnaissance antigénique, très haute résolution) ainsi qu'une grande variété de méthodes utilisées pour la résolution et l'affinement des structures de protéines.

A l'heure actuelle, mon travail de recherche se focalise sur deux thématiques principales: la première porte sur l'étude structurale des récepteurs pour l'antigène des lymphocytes T, dans le but de mieux comprendre, à l'échelle atomique, le mécanisme de reconnaissance de l'antigène. Le deuxième projet, d'ordre méthodologique, concerne les structures de protéines à résolution subatomique. Le but est d'estimer dans quelle mesure la très haute résolution et les densités électroniques très précises peuvent apporter des informations sur la charge atomique, le potentiel électrostatique, le mouvement des atomes, etc. Ces deux thèmes de recherche seront présentés en détail dans ce mémoire.

Les autres projets, pour la plupart achevés, feront l'objet d'une description succincte. Celle-ci permettra tout de même de mieux appréhender la continuité de mon parcours de recherche, et illustrera la diversité méthodologique et le grand apport de la cristallographie à la connaissance des macromolécules biologiques.

## Projets réalisés depuis la thèse.

### *Structure de mutants de l'inhibiteur de la trypsine bovine.*

Nous avons résolu et affiné la structure de 5 mutants du BPTI (*Bovine Pancreatic Trypsin Inhibitor*). Les mutations concernent des résidus enfouis, soit dans le cœur rigide de la protéine (F22A, Y23A, N43G, F45A), soit à l'interface entre le cœur et les boucles se fixant à la trypsine (Y35G). Si les 4 premiers mutants ont une structure très similaire à la protéine native, leur stabilité, mesurée par calorimétrie du repliement ou par taux d'échange de protons, est moindre. Le mutant Y35G présente, par contre, une conformation très

différente pour l'une des boucles de fixation à la trypsine. La mutagenèse dirigée, combinée à l'analyse structurale et les mesures thermodynamiques permettent de mieux comprendre le rôle structural de certains acides aminés. Sur le plan méthodologique, ces études ont fait appel à la méthode du remplacement moléculaire, la structure de la protéine native étant déjà connue. L'efficacité de la dynamique moléculaire, utilisée comme méthode d'affinement, a permis la reconstruction de la boucle du mutant Y35G sans intervention « manuelle », malgré un déplacement de plus de 5 Å.

Housset, D., Kim, K.-S., Fuchs, J., Woodward, C. and Wlodawer, A. (1991). Crystal Structure of a Y35G mutant of Bovine Pancreatic Trypsin Inhibitor. *J. Mol. Biol.*, **220**, 757-770

Danisheski, A.T., Housset, D., Kim, K.S., Tao, F., Fuchs, J., Woodward, C., and Wlodawer, A. (1993). Crevice-forming mutants in the rigid core of bovine pancreatic trypsin inhibitor: Crystal Structures of F22A, Y23A, N43G and F45A. *Protein Science*, **2**, 577-587.

Kim, K.S., Tao, F., Fuchs, J., Danisheski, A.T., Housset, D., Wlodawer, A., and Woodward, C. (1993) Crevice-forming mutants of bovine pancreatic trypsin inhibitor: Stability changes and new hydrophobic surface. *Protein Science*, **2**, 588-596.

### **Structure de deux asparaginases, une enzyme utilisée comme agent thérapeutique contre certaines leucémies.**

Les asparaginases sont des enzymes qui catalysent la transformation de la L-asparagine en L-aspartate et ammoniac. Elles sont utilisées depuis de nombreuses années pour le traitement de certaines leucémies. Par contre, les différentes études biochimiques n'avaient pas permis d'élucider le mécanisme réactionnel. La détermination des structures d'asparaginases a fait l'objet de nombreuses tentatives pendant près de 20 ans dans différents laboratoires, avant d'être finalement déterminée en 1993, dans le groupe d'A. Wlodawer. La structure de l'enzyme, active sous la forme d'un homo-tetramère, a permis de bien décrire le site actif, à l'interface de deux sous-unités et de poser les bases du mécanisme enzymatique. La structure a été résolue par la méthode du remplacement isomorphe, sur la base d'un seul dérivé d'atome lourd (9 sites de fixation). L'utilisation d'un modèle préliminaire (fournissant l'enveloppe de la structure) ainsi que la moyennation de la densité électronique des quatre sous-unités du tétramère ont permis d'obtenir une carte de densité électronique interprétable.

Swain, A.L., Jaskólski, M., Housset, D., Mohana Rao, J.K. and Wlodawer, A. (1993) Crystal Structure of Escherichia coli L-Asparaginase, an Enzyme used in Cancer Therapy. *Proc. Natl. Acad. Sci. (USA)*, **90**, 1474-1478.

Lubkowski, J., Wlodawer, A., Housset, D., Weber, I., Ammon, H.L., Murphy, K.C., Swain, A.L. (1994). Refined crystal structure of *Acinetobacter glutaminasificans* Glutaminase-Asparaginase. *Acta Cryst.*, **D50**, 826-832.

### **Structure de la ferrédoxine I de *Desulfovibrio africanus*.**

Les bactéries sulfato-réductrices ont la faculté unique de coupler la production d'ATP à la réduction des sulfates. Pour cette réaction, ces bactéries possèdent un certain nombre de protéines transporteuses d'électrons, dont les ferrédoxines, qui sont notamment impliquées dans la réduction du sulfite en sulfure.

La résolution de la structure de la ferrédoxine de *Desulfovibrio africanus* a été rendue possible grâce à la combinaison du signal anomal (présence d'un cluster 4Fe-4S) et du remplacement moléculaire. En effet, une première tentative, n'utilisant que le signal anomal, s'est révélée infructueuse (données trop incomplètes) et le remplacement moléculaire seul (avec un modèle possédant 39% d'homologie stricte) ne permettait pas de reconstruire

certaines régions. La combinaison des phases et la moyennation de densité électronique ont permis de reconstruire complètement le modèle.

Séry, A., Housset, D., Serre, L., Bonicel, J., Hatchikian, C., Frey, M. et Roth, M. (1994) Crystal structure of the ferredoxin I from *Desulfovibrio africanus* at 2.3 Å résolution. *Biochemistry*, **33**, 15408-15417.

### **Structure de la 17 $\beta$ -hydroxystéroïde déhydrogénase humaine de type I.**

La 17 $\beta$ -hydroxystéroïde-déshydrogénase (17 $\beta$ -HSD) catalyse la réaction de réduction de l'œstrone pour former le 17 $\beta$ -œstradiol, ainsi que la réaction d'oxydation inverse. Cette enzyme régule le taux d'œstradiol intracellulaire et celui-ci semble jouer un rôle important pour les cancers hormono-dépendants du sein. Ainsi, la 17 $\beta$ -HSD apparaît comme une cible appropriée pour la conception d'inhibiteurs bloquant la production de l'œstrogène.

La structure de la 17 $\beta$ -HSD de type I recombinante (17 $\beta$ -HSD1) complexée avec l'œstradiol et le NADP<sup>+</sup> a été résolue au laboratoire, en conjuguant la technique du remplacement moléculaire et celle des dérivés d'atomes lourds. Le complexe binaire 17 $\beta$ -HSD1-œstradiol a été affiné à 1,7 Å de résolution, avec des données enregistrées à la température ambiante. Le complexe ternaire 17 $\beta$ -HSD1-œstradiol-NADP<sup>+</sup> a été affiné à 2,2 Å de résolution, avec des données enregistrées à basse température (-150°C). Les deux complexes cristallisent dans le groupe d'espace C2 (a=122 Å b=45 Å c=61 Å  $\beta$ =99°), avec une molécule d'enzyme par unité asymétrique. L'analyse de ces deux modèles a permis de déterminer précisément la façon dont le stéroïde et le NADP<sup>+</sup> interagissaient avec l'enzyme: l'œstradiol est stabilisé à ses deux extrémités par des liaisons hydrogènes impliquant les chaînes latérales des résidus Ser142, Tyr155 et His221, le corps du stéroïde étant maintenu par un ensemble de résidus hydrophobes. Le NADP<sup>+</sup> adopte une conformation étendue, le groupement nicotinamide étant à proximité du cycle D de l'œstradiol. Un mécanisme catalytique peut être proposé sur la base de cette structure. Toutefois, une boucle proche du site actif (résidus 192 à 198) et la partie C-terminale de la protéine (40 derniers résidus) sont désordonnées et ne peuvent être construites dans la carte de densité électronique. La résolution des structures de 3 mutants de cette enzyme (H221L, H221Q, Mutant Cter délété) a montré l'importance du positionnement précis de l'œstradiol pour la réaction enzymatique. D'autre part, la boucle 192-198 ne semble stabilisée que lorsque le site NADP<sup>+</sup> est occupé à 100%.

Breton, R., Housset, D., Mazza C. and Fontecilla-Camps, J.C. (1996). The Structure of a complex of Human 17- $\beta$  hydrosteroid-dehydrogenase with estradiol and NADP<sup>+</sup> identifies two principal targets for the design of inhibitors. *Structure*, **4**, 905-915.

Mazza, C., Breton, R., Housset, D. and Fontecilla-Camps, J.C. (1998) Unusual charge stabilization of NADP<sup>+</sup> in 17 $\beta$ -hydroxysteroid dehydrogenase. *J. Biol. Chem.*, **273**, 8145-8152

### **Structure d'une toxine de serpent, la fasciculine.**

La fasciculine est une toxine de serpent de la famille de l'érabutoxine (à trois doigts). La connaissance de la structure était une étape essentielle pour comprendre comment elle interagissait avec son récepteur, l'acétylcholinestérase et ainsi connaître la base structurale de sa toxicité.

Les structures de deux fasciculines ont été résolues au laboratoire et ont montré la grande flexibilité d'une boucle impliquée dans l'interaction avec l'acétylcholinestérase.



Le Du, M.H., Housset, D., Marchot, P., Bougis P.E., Navaza, J., and Fontecilla-Camps, J.C. (1996). Crystal Structure of Fasciculin 2 from Green Mamba Snake Venom : Evidence for Unusual Loop Flexibility. *Acta Cryst.*, **D52**, 87-92.

Housset, D. & Fontecilla-Camps, J.C. (1996). The structure and evolution of snake toxins of the three-finger folding type. in *Protein Toxin Structure*, M. W. Parker ed., Springer-Verlag, Heidelberg, Germany.

### **Structure d'une aminopeptidase de *Lactococcus lactis* : PepC.**

Les bactéries lactiques possèdent un équipement enzymatique spécifique pour la nutrition azotée. Pour disposer d'acides aminés, ces bactéries procèdent schématiquement en trois étapes: dans un premier temps, des protéines présentes dans le milieu extérieur (la caséine dans le lait) sont dégradées par des protéases de paroi. Ensuite, les peptides ainsi obtenus sont transportés dans le cytoplasme. Là, les peptides sont hydrolysés par des peptidases, et la libération séquentielle d'acides aminés en N-terminal permet de pourvoir la bactérie en acides aminés simples. L'étude structurale de ces peptidases permet de mieux comprendre leur spécificité vis-à-vis des acides aminés libérés. L'enjeu est la maîtrise de ce système protéolytique, essentiel pour la croissance des bactéries lactiques et pour la qualité nutritive des aliments. Ainsi, les techniques de génie génétiques devraient permettre de disposer de souches mieux adaptées aux exigences de l'industrie laitière.

La protéine PepC que nous étudions, en collaboration avec l'équipe de J.C. Grippon (INRA, Jouy-en Josas), fait partie de ces peptidases. PepC se présente sous la forme d'un homo-hexamère de 300 kDa, chaque sous-unité comprenant 436 résidus. Son gène a été cloné à partir d'une souche de *Lactococcus lactis*. D'après ses propriétés biochimiques et sa structure primaire, PepC appartient à la famille des peptidases à cystéines. Une forte identité de séquence (35%) existe avec la Bléomycine hydrolase, protéine de résistance à un antibiotique peptidique: la Bléomycine, et dont la structure de cette protéine chez la levure (Gal6) a été très récemment publiée (Joshua-Tor *et al.*, 1995).

La structure de PepC a été résolue par la méthode du remplacement moléculaire avec le modèle de Gal6, et affiné à 2,1 Å de résolution. L'étude des structures de complexes entre un mutant inactif (C67A) et des tripeptides est en cours. Elle devrait permettre de dégager une image précise du site de liaison du substrat. La synthèse d'inhibiteurs spécifiques pourra alors être envisagée. De tels inhibiteurs auront un intérêt pharmacologique certain, compte tenu des fortes analogies structurales et fonctionnelles entre PepC et la bléomycine hydrolase humaine, enzyme impliquée dans la résistance de certains tissus à la bléomycine, un médicament anticancéreux.

## Liste des publications

Housset, D., Kim, K.-S., Fuchs, J., Woodward, C. and Wlodawer, A. (1991). Crystal Structure of a Y35G mutant of Bovine Pancreatic Trypsin Inhibitor. *J. Mol. Biol.* , **220**, 757-770

Danishefski, A.T., Housset, D., Kim, K.S., Tao, F., Fuchs, J., Woodward, C., and Wlodawer, A. (1993). Crevice-forming mutants in the rigid core of bovine pancreatic trypsin inhibitor: Crystal Structures of F22A, Y23A, N43G and F45A. *Protein Science*, **2**, 577-587.

Kim, K.S., Tao, F., Fuchs, J., Danishefski, A.T., Housset, D., Wlodawer, A., and Woodward, C. (1993) Crevice-forming mutants of bovine pancreatic trypsin inhibitor: Stability changes and new hydrophobic surface. *Protein Science*, **2**, 588-596.

Swain, A.L., Jaskólski, M., Housset, D., Mohana Rao, J.K. and Wlodawer, A. (1993) Crystal Structure of Escherichia coli L-Asparaginase, an Enzyme used in Cancer Therapy. *Proc. Natl. Acad. Sci. (USA)*, **90**, 1474-1478.

Housset, D., Habersetzer-Rochat, C. Astier, J.P., and Fontecilla-Camps, J.C. (1994). The crystal structure of toxin II from the scorpion *Androctonus australis* Hector refined at 1.3 Å resolution. *J. Mol. Biol.*, **238**, 88-103.

Lubkowski, J., Wlodawer, A., Housset, D., Weber, I., Ammon, H.L., Murphy, K.C., Swain, A.L. (1994). Refined crystal structure of *Acinetobacter glutaminasificans* Glutaminase-Asparaginase. *Acta Cryst.*, **D50**, 826-832.

Séry, A., Housset, D., Serre, L., Bonicel, J., Hatchikian, C., Frey, M. et Roth, M. (1994) Crystal structure of the ferredoxin I from *Desulfovibrio africanus* at 2.3 Å resolution. *Biochemistry*, **33**, 15408-15417.

Le Du, M.H., Housset, D., Marchot, P., Bougis P.E., Navaza, J., and Fontecilla-Camps, J.C. (1996). Crystal Structure of Fasciculin 2 from Green Mamba Snake Venom : Evidence for Unusual Loop Flexibility. *Acta Cryst.*, **D52**, 87-92.

Breton, R., Housset, D., Mazza C. and Fontecilla-Camps, J.C. (1996). The Structure of a complex of Human 17-β hydrosteroid-dehydrogenase with estradiol and NADP<sup>+</sup> identifies two principal targets for the design of inhibitors. *Structure* , **4**, 905-915.

Housset, D. & Fontecilla-Camps, J.C. (1996). The structure and evolution of snake toxins of the three-finger folding type. in *Protein Toxin Structure*, M. W. Parker ed., Springer-Verlag, Heidelberg, Germany.

Smith, G.D., Blessing, R.H., Ealick, S.E., Fontecilla-Camps, J.C., Hauptman, H.A., Housset, D., Langs, and Miller, R. (1997). Ab initio structure determination and refinement of a scorpion protein toxin. *Acta Cryst.*, **D53**, 551-557.

Housset, D., Mazza, G., Grégoire, C., Piras, C., Malissen, B., Fontecilla-Camps, J.C. (1997). The three-dimensional structure of a T-cell antigen receptor V $\alpha$ V $\beta$  heterodimer reveals a novel arrangement of the V $\beta$  domain. *Embo J.*, **16**, 4205-4216.

Mazza, C., Breton, R., Housset, D. and Fontecilla-Camps, J.C. (1997) Unusual charge stabilization of NADP<sup>+</sup> in 17 $\beta$ -hydroxysteroid dehydrogenase. Accepted in *J. Biol. Chem.*

Mazza, G., Housset, D., Piras, C., Grégoire, C., Lin, S.Y., Fontecilla-Camps, J.C. and Malissen, B. (1998). Glimpses at the recognition of peptide/MHC complexes by T-cell antigen receptor. *Immunological Rev.*, **163**, 187-196

Mazza, G., Housset, D., Piras, C., Grégoire, C., Fontecilla-Camps, J.C. and Malissen, B. (1999) Structural features of the interaction between an anti-clonotypic antibody and its cognate T-cell antigen receptor. *J. Mol. Biol.*, **287**,773-780.

Housset, D., Benabicha, F., Pichon-Pesme, V., Jelsch, C., Maierhofer, A., David, S., Fontecilla-Camps, J.C. and Lecomte, C. (2000) Towards the charge density study of proteins : a room temperature scorpion toxin structure at 0.96 Å resolution as a first test case. *Acta Cryst.*, **D56**, 151-160.

# A Structure du récepteur pour l'antigène du lymphocyte T

## A.1 Introduction générale

Les lymphocytes T font partie des principaux acteurs du système immunitaire, en permettant à l'organisme de se débarrasser de différents types de pathogènes (virus, bactéries, protozoaires). Ils sont les médiateurs de la réponse cellulaire, par opposition à la réponse humorale dont les lymphocytes B et le système complémentaire sont les acteurs. La spécificité de la réponse cellulaire est de pouvoir détecter la présence de molécules ou micro-organismes étrangers à l'intérieur des cellules de l'hôte, et d'activer une chaîne d'évènements qui permettra l'élimination de l'intrus. La clef moléculaire qui permet au lymphocyte T d'établir cette distinction entre cellules saines et cellules infectées est un récepteur membranaire, le TCR (*T cell receptor*).

L'homologie de séquence et de structure entre ce récepteur et le fragment Fab des anticorps cache un mode de reconnaissance de l'antigène très différent. Si l'anticorps interagit directement avec l'antigène entier, avec une forte affinité, le TCR, pour sa part, reconnaît l'antigène sous la forme de fragments peptidiques, issus de la dégradation des protéines de l'agent pathogène par la machinerie enzymatique de la cellule hôte. A la surface des cellules infectées, ces peptides antigènes sont présentés aux lymphocytes T, sous la forme d'un complexe avec une des molécules codées par le complexe majeur d'histocompatibilité (MHC pour *major histocompatibility complex*). Ainsi, à la différence des anticorps, le TCR reconnaît une surface composite formée par une molécule du soi (la molécule MHC) et un peptide antigénique, avec une affinité modérée ( $K_D$  de  $10^{-5}$  à  $10^{-6}$  M). A cause du caractère polymorphique des molécules MHC, le répertoire des TCR, aussi vaste que celui des anticorps, ne pourra reconnaître les antigènes que dans le cadre restrictif des molécules MHC du soi (restriction du MHC). Mais le rôle du TCR ne se limite pas à l'interaction avec le complexe peptide/MHC (pMHC) : formant un complexe non-covalent avec les molécules CD3, il transmet le signal de reconnaissance du peptide antigène au lymphocyte T. Ce signal aura des conséquences diverses pour le lymphocyte selon le stade de développement de celui-ci et sa localisation (activation et prolifération d'un lymphocyte mature, déclenchement d'un lymphocyte activé, état d'anergie pour rendre tolérant au soi, survie ou mort cellulaire lors de la sélection thymique).

La compréhension de ce mécanisme de reconnaissance antigénique, très sélectif bien que basé sur des interactions faibles avec des complexes pMHC peu différents, a beaucoup progressé ces dernières années, notamment grâce aux études structurales.

C'est dans le dessein de connaître les détails de l'interaction entre le TCR et le peptide antigénique, pour comprendre quels sont les déterminants de la reconnaissance antigénique, que nous avons entrepris le projet de déterminer la structure atomique d'un TCR. Même si l'action des lymphocytes T est d'une grande complexité, impliquant un grand nombre de cellules, de molécules et de cascades de réactions, la spécificité du lymphocyte T pour sa cible s'exprime exclusivement dans ce TCR. La structure de ce dernier, et la façon dont il interagit avec ses différents partenaires devraient nous faire avancer grandement sur la nature des interactions à l'origine de la spécificité et des mécanismes de restriction vis-à-vis du MHC.

Dans un contexte plus général, en comprenant les mécanismes de reconnaissance, on peut espérer pouvoir agir sur les mécanismes d'activation des cellules T pour immuniser ou

activer la réponse immunitaire (infections rétro-virales, cancers, ...) ou au contraire accroître la tolérance à certains antigènes (greffe, maladie auto-immune).

### **A.1.1 Pathologies impliquant le lymphocyte T.**

Les lymphocytes T permettent à l'organisme de lutter contre une grande variété de pathologies. Le point commun entre ces différentes maladies est que le micro-organisme (ou son produit) impliqué se loge à l'intérieur des cellules de l'hôte. Deux types de lymphocyte T vont permettre de répondre à une grande variété de situations. Les lymphocytes T cytotoxiques s'attaquent aux pathogènes se développant dans le cytosol. Ainsi, ils agissent dans la quasi-totalité des affections virales (vaccin, grippe, rage, sida, ...) ainsi que contre certaines affections microbiennes (*listeria*), en tuant les cellules infectées. Les lymphocytes T auxiliaires (*helper*) vont lutter contre les micro-organismes se développant dans les vésicules des macrophages (tuberculose, lèpre, leishmaniose) ou indirectement, contre des toxines bactériennes extracellulaires ou certains virus, en activant la production d'anticorps des lymphocytes B spécifiques (tétanos, pneumonie, staphylocoque doré, virus de la polio, ...). Ces lymphocytes T agissent en activant d'autres acteurs du système immunitaire (macrophages, lymphocyte B) qui élimineront les micro-organismes ou neutraliseront leurs produits. Si cette action n'est pas parfaitement contrôlée, les lymphocytes T peuvent soit ne pas répondre à un agent pathogène, soit attaquer les tissus sains et être la cause de maladies auto-immunes. Ces dernières ont pour cause l'activation de lymphocytes T auto-spécifiques, qui peut intervenir soit par l'intermédiaire de superantigènes (arthrite, Cole & Griffiths, 1993, diabète insulino-dépendant Conrad *et al.*, 1994, encéphalomyélite, Brocke *et al.*, 1993, ...), soit par l'intermédiaire de peptides viraux ou bactériens homologues à un peptide immunodominant du soi (sclérose multiple, Wucherpfennig & Strominger, 1995 ...). Enfin, les lymphocytes T ont un rôle permanent de surveillance, permettant d'éliminer les cellules devenues cancéreuses. Même si le système immunitaire n'agit pas de façon efficace contre le développement de nombreux cancers, les lymphocytes T sont parfois capables de reconnaître spécifiquement ces cellules (Boon *et al.*, 1994). L'activation de ces lymphocytes par différentes voies permet de lutter efficacement contre certains cancers.

### **A.1.2 Reconnaissance soi/non-soi.**

La façon dont le lymphocyte T est capable de faire la distinction entre soi et non-soi reste la question centrale de la réponse immunitaire cellulaire. Une très grande sélectivité est impérative pour lutter efficacement contre les pathogènes et ne pas détruire les tissus sains. Le processus de sélection thymique (cf. §A.3.2.1) permet de construire un répertoire de lymphocytes T capables de reconnaître une très grande variété d'antigènes étrangers, et de rester tolérants aux peptides du soi. Pour autant, tous les lymphocytes T auto-réactifs ne sont pas éliminés : un grand nombre de tels lymphocytes, pouvant réagir aux peptides spécifiques de certains tissus, circulent en permanence dans notre corps. Par contre, ils ne seront, en principe, jamais activés, et seront éventuellement anergisés, c'est-à-dire rendus inactivables. Différentes stratégies sont mises en place pour maintenir la tolérance vis-à-vis du soi. La compréhension de ce contexte est fondamentale pour décrypter l'impact d'une interaction TCR/pMHC. Le lymphocyte T, par l'intermédiaire du TCR, n'est pas capable à lui seul, de discriminer le soi du non-soi ; Il possède une spécificité pour un certain nombre de peptides, présentés par une ou plusieurs molécules MHC. Le contexte de la présentation de l'antigène et de l'activation du lymphocyte T, dans les tissus lymphoïdes périphériques permettra d'assurer la tolérance vis-à-vis du soi.

### A.1.3 Problème de définition du soi.

Définir le soi n'est pas aussi simple qu'il n'y paraît. Dans son acception la plus générale, elle inclut toutes les molécules de l'hôte, le non-soi représentant donc le reste du monde. Simple et séduisante, cette définition ne semble pas convenir pour qualifier l'action du système immunitaire. Pour preuve l'existence d'anticorps dirigés contre l'ADN ou la kératine, ou l'absence de réponse immunitaire engendrée par différents haptènes. Dans une acception des plus restreintes, le soi peut être défini comme l'ensemble des molécules du soi accessibles aux acteurs de la réponse immunitaire adaptative, lors de leur phase de maturation. Là encore, cette définition ne convient que partiellement, car les lymphocytes sont, en pratique, tolérants pour un grand nombre de peptides spécifiques de certaines lignées cellulaires de l'hôte, jamais rencontrés lors du processus de sélection. Une définition du non-soi, peut-être plus proche de la distinction que semble opérer le système immunitaire et proposée par P. Matzinger (1994), est tout ce qui représente un danger pour l'hôte. Dans ce mémoire, nous garderons la première définition, plus universelle, tout en indiquant schématiquement les mécanismes par lesquels le système immunitaire est capable de focaliser son action contre les pathogènes.

## A.2 Contexte de l'action des lymphocytes T.

Il est difficile d'aborder le mode d'action des lymphocytes T et le rôle de son récepteur pour l'antigène sans décrire un minimum le contexte de cette action. Effet, le lymphocyte T n'agit pas seul, mais fait partie intégrante d'un ensemble complexe de défenses que forme le système immunitaire. Comme nous allons le voir succinctement, les différentes cellules du système immunitaire coopèrent étroitement pour parvenir à l'élimination des micro-organismes invasifs. Dans le cas des lymphocytes T, le processus de présentation de l'antigène par une cellule présentatrice d'antigène (APC) fixe le cadre de reconnaissance. L'objectif de cette partie est donc de présenter quelques généralités sur le système immunitaire, permettant de situer l'action des lymphocytes T, pour mieux comprendre le rôle de la reconnaissance TCR $\alpha\beta$  / antigène dans les différentes phases d'action du lymphocyte T.

### A.2.1 Aperçu général des protagonistes du système immunitaire

Pour se protéger contre l'invasion de différents types d'organismes (bactéries, virus, champignons, parasites), ou éviter la prolifération de cellules du soi devenues cancéreuses, les vertébrés disposent de tout un arsenal de cellules et molécules qui sont collectivement capables de reconnaître et éliminer la plupart de ces intrus. Cet ensemble d'outils de défense qui forme le système immunitaire est extrêmement complexe chez les mammifères. On peut distinguer deux formes de réponses immunitaires :

- Une réponse innée, non-adaptative, qui se fait par l'intermédiaire de cellules phagocytaires non-spécifiques, capables d'internaliser une grande variété de bactéries, avec l'aide éventuelle du système complémentaire (réponse humorale non-adaptative).
- Un autre type de réponse, dit adaptatif, permet de sélectionner et adapter certaines cellules (les lymphocytes B ou T) pour lutter spécifiquement contre un micro-organisme et éventuellement procurer une protection permanente. Cette réponse spécifique se fera grâce aux anticorps (historiquement appelée réponse humorale) ou aux cellules T (réponse cellulaire).



En réalité, les phagocytes comme les deux types de lymphocytes coopèrent étroitement pour apporter une protection aussi efficace que possible.

### **A.2.1.1 Les leucocytes**

Ces cellules, comme toutes les cellules sanguines, proviennent des cellules hématopoïétiques dites pluripotentes, dans la moelle osseuse. Ces cellules se différencient ensuite pour former les globules rouges, les plaquettes et les globules blancs. Ces derniers, appelés aussi leucocytes, sont les médiateurs des différents types de réponses immunitaires. Les leucocytes se subdivisent en deux lignées (tableau A.1) :

#### A.2.1.1.1 la lignée myéloïde

Elle comprend différents types de phagocytes :

Les macrophages, forme mature des monocytes circulant dans le sang, sont présents dans la plupart des tissus.

Les granulocytes, pour leur part, se divisent en trois catégories différentes : les neutrophiles, les éosinophiles et les basophiles.

Les phagocytes (macrophages, neutrophiles, ...) présents dans les tissus et le sang permettent de lutter immédiatement contre un grand nombre de bactéries par phagocytose directe. Ils interviennent également après l'action des lymphocytes B, en phagocytant les pathogènes opsonnés par les anticorps spécifiques. Les macrophages et les cellules dendritiques ont un rôle essentiel dans la présentation des antigènes aux lymphocytes B et T, en permettant l'activation des lymphocytes naïfs (lymphocytes matures n'ayant jamais été activés).

#### A.2.1.1.2 La lignée lymphoïde

Elle donne, naissance aux différents lymphocytes, dont les lymphocytes B et T sont les représentants majeurs. Après maturation, chaque lymphocyte se caractérise par le récepteur pour l'antigène qu'il exprime à sa surface, qui lui confère une spécificité pour un ou plusieurs antigènes. Ce récepteur appartient à la famille des immunoglobulines. Dans le cas des lymphocytes B, il est la forme membranaire de l'anticorps qui sera secrété suite à l'activation du lymphocyte.

Après la phase de maturation, ces lymphocytes circulent perpétuellement à travers les vaisseaux sanguins ou lymphatiques. La présentation des antigènes aux lymphocytes et l'activation de ceux-ci a lieu dans les organes lymphatiques périphériques (nœuds lymphatiques, plaques de Peyers, rate).

Les lymphocytes B jouent un rôle primordial dans la réponse adaptative aux pathogènes extracellulaires. Les lymphocytes B interagissant spécifiquement avec le pathogène mûrent en plasmocytes, capables d'exprimer des anticorps de haute affinité, pouvant inactiver ou opsonner le pathogène. Pour être activé, les lymphocytes B ont souvent besoin des lymphocytes T (Th2), qui reconnaissent un fragment de l'antigène présenté par les molécules MHC du lymphocyte B. Les lymphocytes B peuvent aussi, en présence d'agents infectieux, activer les cellules T naïves.

Les lymphocytes T ont pour rôle essentiel de lutter contre les pathogènes intracellulaires, qui échappent aux lymphocytes B. Ils interviennent aussi pour certains pathogènes extracellulaires, en activant les lymphocytes B. La particularité du lymphocyte T est qu'il ne reconnaît pas le pathogène entier, mais seulement des fragments peptidiques de celui-ci, présentés à la surface des cellules infectées par le pathogène. Il existe deux sortes de lymphocytes T : les lymphocytes T cytotoxiques et les lymphocytes T auxiliaires (*helper*), qui agissent en fonction de la localisation du pathogène dans la cellule (cytosol ou vésicules).

Pour les pathogènes se développant dans le cytosol (virus), les lymphocytes T cytotoxiques seront activés. Après prolifération et activation, ces lymphocytes sont capables de tuer directement les cellules infectées, en induisant l'apoptose par activation de nucléases.

Dans le cas des bactéries se développant dans les vésicules des macrophages ou pour les protéines internalisées par les phagocytes ou lymphocytes B, des lymphocytes T auxiliaires, Th1 ou Th2, seront activés. Après prolifération et différenciation, les lymphocytes Th1 activeront les macrophages, qui élimineront l'agent pathogène par fusion des vésicules avec les lysosomes. Les lymphocytes Th2 activeront les lymphocytes B, qui se mettront à produire des anticorps spécifiques de la protéine internalisée.

**Tableau A.1** : les différents leucocytes et leurs fonctions

Moëlle osseuse	Sang	Tissus	Fonction et cibles		
cellule hémato-poïétique pluripotente	progéniteur myéloïde	monocyte	macrophage phagocytose directe : bactéries ; Après opsonnisation : bactéries, toxines, ... ; Après activation par Th1 : bactéries, parasites, ...		
		neutrophile	cellule dendritique phagocytose et présentation d'antigènes activation des lymphocytes T		
		eosinophile	-	phagocytose (dans le sang uniquement)	
		basophile	-	phagocytose parasites opsonnés	
		lymphocyte B	-	mal connues. protection des muqueuses, perméabilité vasculaire.	
	progéniteur lymphoïde	lymphocyte cytotoxique (CD8 <sup>+</sup> )		production d'anticorps bactéries, parasites, virus extracellulaires	
		lymphocyte T	Th1	cellule tueuse virus et bactéries se développant dans de cytosol	
			auxiliaire (CD4 <sup>+</sup> )	Th2	active les macrophages pathogènes intravésiculaires
					active les lymphocytes B pathogènes extracellulaires (toxines bactériennes, ...)

### A.2.1.2 Les cellules présentatrices d'antigène (APC)

Si les lymphocytes B sont capables d'interagir directement avec l'agent pathogène ou ses produits directs, les lymphocytes T ont pour fonction de détecter la présence de micro-organisme étranger à l'intérieur des cellules. Pour permettre cette reconnaissance, la plupart des cellules possédant un noyau sont capables de présenter à leur surface des fragments peptidiques issus des protéines dégradées à l'intérieur de la cellule. Cette présentation se fait par le moyen de molécules codées par le complexe majeur d'histocompatibilité (MHC). Lorsqu'un microorganisme est internalisé par la cellule dans laquelle il se duplique éventuellement, les protéines spécifiques à cet intrus, comme celles de la cellule hôte, sont amenées à être dégradées par la machinerie enzymatique (protéasome dans le cytosol, protéases dans les vésicules). Ces protéines sont découpées en fragments peptidiques qui peuvent être fixés, dans le réticulum endoplasmique ou dans les vésicules, aux molécules codées par le MHC. Ces complexes molécule MHC/peptide (pMHC) migrent ensuite vers la



surface cellulaire. La cellule T pourra alors interagir avec cette surface et reconnaître les peptides provenant des protéines du soi et les peptides provenant du pathogène internalisé.

Si théoriquement la plupart des cellules possèdent cette fonction de présentation, elles diffèrent par leur pouvoir d'activation du lymphocyte T. Certaines cellules phagocytaires, issues des cellules hématopoïétiques de la moelle osseuse, ont un rôle primordial dans l'activation de la réponse immunitaire cellulaire et sont appelées cellules présentatrices d'antigène « professionnelle » (APC). Il s'agit notamment des cellules dendritiques, spécialisées pour internaliser les pathogènes dans les tissus périphériques et présenter les peptides antigènes dans les tissus lymphoïdes. Elles sont capables d'activer des cellules T naïves. D'autres types de cellules, comme les macrophages, les lymphocytes B, peuvent également jouer ce rôle d'APC pendant le temps de l'infection, autrement leur action se limite à l'activation des cellules T effectrices. Enfin, certaines cellules, comme les cellules nerveuses, expriment très peu de molécules MHC et sont donc peu sensibles aux lymphocytes T.

## **A.2.2 Un scénario extrêmement simplifié d'une agression**

Lorsqu'un micro-organisme réussit à franchir la barrière épidermique et crée un foyer infectieux, il provoque immédiatement l'activation du système immunitaire.

### **A.2.2.1 Réponse innée**

Dans un premier temps, il s'agit d'une réponse innée, par l'intermédiaire des macrophages. Si cette première réponse est efficace, l'intrus est éliminé et l'infection passe inaperçue. Si au bout de quelques heures, ceux-ci ne réussissent pas à éliminer l'intrus, la réponse immunitaire adaptative est déclenchée.

### **A.2.2.2 Réponse adaptative**

#### A.2.2.2.1 Migration vers les nœuds lymphatiques

Les macrophages libèrent alors des cytokines qui vont permettre aux lymphocytes de migrer en grand nombre des vaisseaux sanguins vers les vaisseaux et les nœuds lymphatiques. Les phagocytes (APC) eux aussi migreront vers ces nœuds lymphatiques où lymphocytes et antigènes pourront interagir. Les lymphocytes naïfs (lymphocytes jamais encore activés) ou les lymphocytes mémoires (si le pathogène a déjà infecté l'hôte par le passé), spécifiques de l'antigène présenté, seront alors activés par les APC.

#### A.2.2.2.2 Activation et prolifération

La phase d'activation se décompose en plusieurs étapes. Dans un premier temps, le lymphocyte se dilate, pour devenir en quelques heures un lymphoblaste. Vient ensuite la phase de prolifération où le lymphoblaste se divise, plusieurs fois par 24 heures, pour se différencier au bout de 4 ou 5 jours en lymphocyte effecteur, capable lors de la reconnaissance du même antigène, de sécréter des anticorps (pour les cellules B) ou de tuer une cellule infectée (pour les cellules T). La reconnaissance de l'antigène provoque le déclenchement (*triggering*) de ces lymphocytes effecteurs. Ces derniers ont une durée de vie courte (quelques jours). Dans cette phase, le nombre de lymphocytes effecteurs (activés) peut atteindre  $10^{12}$  en une semaine.

#### A.2.2.2.3 Arrêt de la réponse et mémoire

Lorsque l'agresseur a disparu, la majorité de ces lymphocytes effecteurs meurent de façon programmée (apoptose) et seul un petit nombre se transforme en cellules mémoires

avec une durée de vie bien plus longue. Ces cellules permettront l'immunisation temporaire ou permanente face à une agression identique ultérieure.

### A.3 Les protagonistes de la réponse cellulaire

Nous allons maintenant nous intéresser plus en détail à la réponse cellulaire, dans laquelle s'inscrit notre thème de recherche. Le rôle du récepteur de la cellule T s'inscrit au cœur de la réponse cellulaire, lors de la reconnaissance, par le lymphocyte T, d'une cellule infectée ou cancéreuse. Cet événement central implique deux partenaires, la cellule présentatrice (APC pour l'activation d'une cellule T naïve), qui présente l'antigène, à l'aide de la molécule MHC et le lymphocyte T, capable de le reconnaître, par l'intermédiaire de son récepteur (TCR). La structure génétique du MHC et du TCR permet de mieux comprendre la variabilité de ces molécules et l'origine de la spécificité de notre répertoire de lymphocytes T pour nos molécules MHC.

#### A.3.1 Les molécules MHC

Les molécules MHC sont réparties en deux classes appelées I et II. Les molécules MHC de classe I présentent les peptides issus de la dégradation des protéines cytosoliques (protéines virales, par exemple) alors que les molécules MHC de classe II vont présenter les peptides présents dans les vésicules liées à la membrane. Elles vont permettre d'indiquer la présence de pathogènes vivant dans les vésicules des macrophages ou internalisées par les phagocytes. Les molécules MHC de classe I permettent d'activer les lymphocytes T cytotoxiques (CD8<sup>+</sup>) alors que les molécules MHC de classe II activent les lymphocytes T auxiliaires (CD4<sup>+</sup>). Ces molécules MHC sont appelées HLA (*Human Leucocyte Antigen*) chez l'homme, H-2 chez la souris.

##### A.3.1.1.1 Le gène du MHC

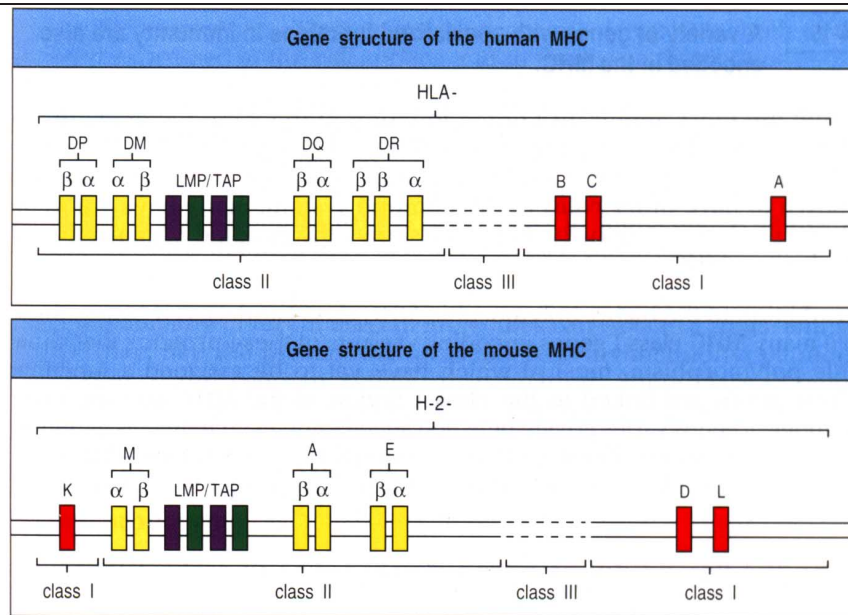
Les molécules MHC sont codées par les gènes du complexe majeur d'histocompatibilité. L'une des particularités du MHC vient du fait qu'il est polygénique et que les produits du MHC sont hautement polymorphiques.

La polygénie traduit le fait que le MHC code pour plusieurs molécules MHC classe I et classe II (chez l'homme, chaque complexe de gènes code pour 3 molécules MHC classe I et quatre molécules MHC classe II, figure A.1).

Le polymorphisme consiste en l'existence de multiples allèles de chaque gène (plus d'une centaine pour certains *loci*).

Étant donnée cette grande variabilité, la plupart des individus sont hétérozygotes pour les *loci* des MHC, et comme les deux allèles présents chez un individu sont exprimés (expression co-dominante), chaque humain exprime six molécules MHC classe I et huit molécules MHC classe II différents à la surface de chacune des cellules présentatrices d'antigène. Par conséquent également, la chance pour que deux individus expriment les mêmes MHC est très faible : elle n'est déjà que de 0.25 pour les frères et sœurs.

La variété des molécules MHC au sein d'un individu permet aux cellules présentatrices de présenter une large gamme de peptides, limitant les possibilités pour un pathogène d'échapper à la vigilance du système immunitaire. La variété entre individus diminue la chance que toute une espèce soit sensible à un pathogène donné. La conséquence de cette variété est la spécificité du répertoire de lymphocytes T d'un individu pour son haplotype (combinaison de ses molécules MHC). Cette spécificité est la cause du rejet des greffes provenant d'individus ne possédant pas le même haplotype.



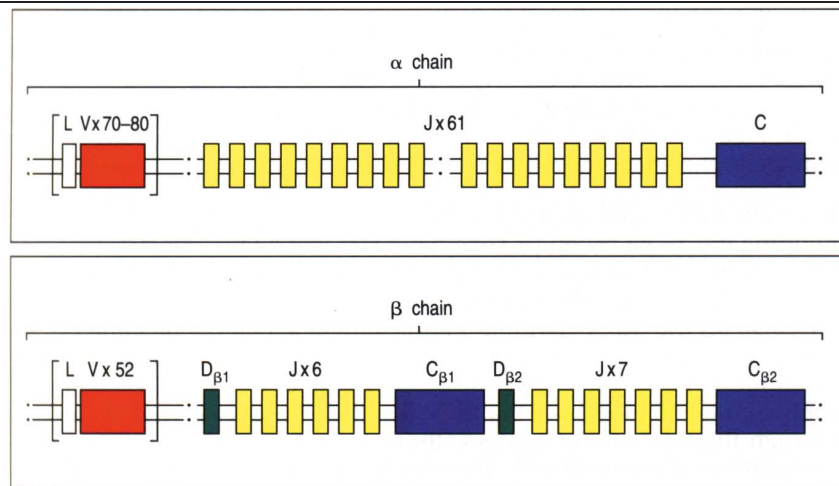
**Figure A.1** : Structure schématique des gènes du MHC pour l'homme et la souris. Chez l'homme, les molécules MHC de classe I sont appelées HLA-A, -B et -C ; les molécules MHC de classe II sont appelées HLA-DR, -DP et -DQ. Chez la souris, les molécules MHC de classe I se nomment H-2K, H-2D et H-2L. D'après *Immuno-Biology*, Janeway & Travers, pp 4:21.

### A.3.2 Le TCR

Les lymphocytes T peuvent exprimer deux types de récepteurs : l'un est formé des chaînes  $\alpha$  et  $\beta$ , l'autre des chaînes  $\gamma$  et  $\delta$ . Ces TCR $\alpha\beta$  et TCR $\gamma\delta$  correspondent à deux lignées différentes de lymphocyte T issues du même progéniteur, la différenciation ayant lieu au début de la phase de maturation. Le rôle des lymphocytes T exprimant le récepteur  $\gamma\delta$ , présent notamment dans les tissus épithéliaux (épiderme, intestin grêle) est mal connu (action antimicrobienne); le TCR $\gamma\delta$  semble plus proche des anticorps que du TCR $\alpha\beta$  et interagit sans doute directement avec l'antigène entier ou des protéines homologues au MHC (MIC-A). La lignée de cellules T  $\alpha\beta$ , de loin la plus abondante, est présente dans les vaisseaux sanguins et les tissus et vaisseaux lymphoïdes, et reconnaît les peptides présentés par les molécules MHC de classe I ou de classe II. C'est à ce type de récepteur que nous nous intéressons. L'analyse de la structure du gène des chaînes  $\alpha$  et  $\beta$  permet de comprendre l'origine de la variabilité et de la spécificité du TCR.

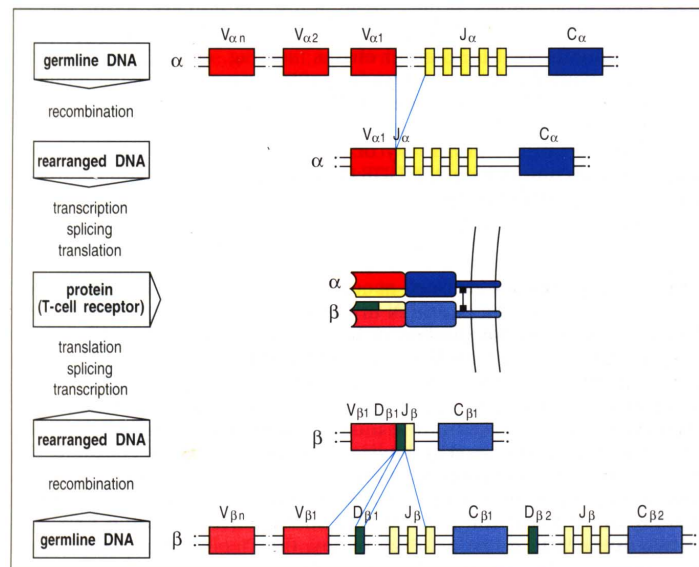
De façon similaire aux anticorps, la recombinaison génétique est à l'origine de la grande diversité du TCR $\alpha\beta$ . L'organisation des gènes des chaînes  $\alpha$  et  $\beta$  est similaire à celle des chaînes légères et des chaînes lourdes des anticorps (figure A.2).

La chaîne  $\alpha$  est le fruit de la recombinaison des gènes  $V\alpha$ ,  $J\alpha$  et  $C\alpha$ . La chaîne  $V\beta$  résulte de la recombinaison des gènes  $V\beta$ , D,  $J\beta$  et  $C\beta$  (figure A.3). Parce que nous possédons plusieurs gènes pour les segments  $V\alpha$ ,  $J\alpha$ ,  $V\beta$ , et  $J\beta$  le mécanisme de recombinaison permet de construire de l'ordre  $10^{16}$  TCR $\alpha\beta$  différents.



**Figure A.2** : Structure schématique des gènes des chaînes  $\alpha$  et  $\beta$  du TCR. Le gène de la chaîne  $\alpha$  contient 70 à 80 segments  $V\alpha$ , environ 61 segments  $J\alpha$  et un segment codant pour le domaine constant. Le gène de la chaîne  $\beta$  contient environ une cinquantaine de segments  $V\beta$ , et deux *clusters* de gènes contenant chacun un segment D, six ou sept segments  $J\beta$  et un segment  $C\beta$ . D'après *Immuno-Biology*, Janeway & Travers, pp 4:35.

La seule différence notable avec les anticorps est l'absence d'hyper-mutations somatiques (mutation ponctuelle après la phase de maturation, sur les résidus de site antigénique), qui permet notamment aux anticorps d'optimiser leur affinité pour l'antigène.



**Figure A.3** : Recombinaison des gènes pour les chaînes  $\alpha$  et  $\beta$  du TCR. D'après *Immuno-Biology*, Janeway & Travers, pp 4:35.

### A.3.2.1 Le processus de sélection thymique

Avant d'être mis en circulation dans les vaisseaux sanguins et lymphatiques, les lymphocytes T sont sujets à une sélection sévère. Il est en effet fondamental que le lymphocyte T soit capable de distinguer le soi du non-soi. Il doit pouvoir ne pas réagir lorsqu'un peptide du soi lui est présenté (faute de quoi il s'attaquerait à nos cellules saines) et

s'activer lors qu'il rencontre un peptide étranger pour lequel il a une certaine spécificité. Cette phase d'apprentissage de nos lymphocytes T garde encore quelques zones d'ombre, mais en voici les grandes lignes.

Les lymphocytes T migrent très tôt vers le thymus où a lieu la phase de maturation. Durant cette phase, le lymphocyte T va exprimer un certain nombre de molécule de surface, dont les chaînes  $\beta$  et  $\alpha$  du récepteur pour l'antigène, fruit du réarrangement des gènes codant pour les segments V, (D), J et C. Suit la phase de sélection qui se décompose en deux étapes : la sélection positive et la sélection négative. La sélection positive, mise en évidence à la fin des années 70 (Bevan, 1977, Zinkernagel *et al.*, 1978) permet de ne garder que les clones spécifiques d'au moins une des molécules MHC présentes ; La sélection négative (Lo *et al.*, 1986, Kappler *et al.*, 1987) permet d'éliminer les clones interagissant trop fortement avec les peptides du soi présentés dans le thymus, dans le contexte des molécules MHC exprimées.

Au départ, le lymphocyte qui pénètre dans le thymus ne possède pratiquement aucune des protéines de surface caractéristiques du lymphocyte T mature ; dans une première phase, certaines molécules de surface sont exprimées (CD2, CD44, CD25) et le réarrangement de la chaîne  $\beta$  a lieu. L'expression d'une pré-chaîne  $\alpha$  et de CD3 met fin au réarrangement de la chaîne  $\beta$  et déclenche l'expression des molécules CD4 et CD8. Suit alors une phase de prolifération des lymphocytes à ce stade de maturation. Lorsque la prolifération cesse, le réarrangement de la chaîne  $\alpha$  démarre pour produire un TCR $\alpha\beta$  complet. A ce stade intervient la sélection positive : pour passer cette sélection et échapper à l'apoptose, les lymphocytes T doivent exprimer un TCR $\alpha\beta$  ayant une certaine affinité pour au moins une des molécules MHC, en complexe avec un quelconque peptide du soi. Durant cette phase, au cours de laquelle 95% des lymphocytes T sont éliminés, le réarrangement de la chaîne  $\alpha$  peut continuer, permettant à un même lymphocyte de tester plusieurs combinaisons  $\alpha\beta$ . La sélection négative, par l'intermédiaire des APC, élimine les lymphocytes T réagissant trop fortement aux peptides du soi présentés, et permet d'éliminer l'essentiel des lymphocytes auto-réactifs. Les lymphocytes sélectionnés évoluent ensuite pour exprimer le TCR $\alpha\beta$  en grand nombre et cesser d'exprimer l'un des deux co-récepteurs CD4 ou CD8.

Après cette sélection, les lymphocytes T naïfs quittent le thymus pour circuler à travers les vaisseaux sanguins et lymphatiques. A la fin de la phase de sélection, on peut distinguer deux types de lymphocytes T : les lymphocytes T cytotoxiques, spécifiques des molécules MHC de classe I, expriment le co-récepteur CD8 (CD8<sup>+</sup>). Les lymphocytes T auxiliaires sont, eux, spécifiques des molécules MHC de classe II et expriment le co-récepteur CD4 (CD4<sup>+</sup>).

### **A.3.3 Rôle de l'interaction TCR $\alpha\beta$ /pMHC dans les différents stades de développement du lymphocyte T.**

L'interaction entre TCR $\alpha\beta$  et la molécule MHC intervient à plusieurs étapes du développement du lymphocyte T :

- Dans la phase de maturation, pour la sélection du répertoire.
- Pour l'activation de la cellule T naïve ou de la cellule T mémoire, qui enclenchera la multiplication des lymphocytes.
- Pour déclencher l'action des cellules T effectrices.
- Rôle dans le maintien de la tolérance et la survie des cellules T mémoires.

Pendant toutes ces phases, le TCR va former ponctuellement des complexes avec des molécules MHC chargées avec des peptides du soi ou du non-soi (figure A.4). En fonction du contexte, la formation d'un tel complexe aura des conséquences complètement différentes, allant de l'absence de réaction, la mise en état d'anergie, l'activation, le déclenchement ou la mort programmée. Ainsi, cette fonction de reconnaissance est fondamentale tout au long de la vie du lymphocyte. La compréhension, à l'échelle atomique, de cet évènement clef doit



permettre de mieux comprendre la façon dont cette interaction est traduite en signaux d'action aussi variés.

### A.3.4 De l'interaction des lymphocytes T avec une cellule à au signal d'activation.

L'interaction entre un lymphocyte T et sa cible s'établit de prime abord par l'intermédiaire de molécules d'adhésion (CD2, LFA-1 et ICAM-3 pour le TCR ; LFA-3, ICAM-1, ICAM-2 et LFA-1 pour la cellule présentatrice). Dans ce contexte d'interaction cellulaire relativement stable, les TCR vont pouvoir « identifier » les différents peptides présentés par les molécules MHC. L'analyse des paramètres cinétiques et thermodynamiques de l'interaction TCR/pMHC permet de nous éclairer sur les bases de la sélectivité du TCR.

Pour qu'il y ait activation du lymphocyte T, l'engagement de soixante à cent TCR avec les pMHC suffit. Ce chiffre est à mettre en rapport avec la centaine de milliers de pMHC présents à la surface de la cellule présentatrice. L'affinité du TCR pour le pMHC est relativement faible ( $1 < K_D < 90 \mu\text{M}$ , Davis *et al.*, 1998) et ne participe pas significativement à l'adhésion cellulaire. Les paramètres cinétiques mesurés pour différents complexes TCR/pMHC montrent une association lente ( $10^2 < K_{on} < 10^4 \text{ M}^{-1}\text{s}^{-1}$ ) et une dissociation rapide ( $10^{-2} < K_{off} < 10^{-1} \text{ s}^{-1}$ ). Un grand nombre d'études ont tenté d'élucider le lien entre les caractéristiques thermodynamiques de l'interaction TCR/pMHC et la sélection ou l'activation du lymphocyte. Une plage de  $K_D$  comprise entre 10 et 60  $\mu\text{M}$  semble être optimale, dans certains cas, pour une sélection positive (Alam *et al.*, 1996). Un autre groupe (Kersh *et al.*, 1998) a par contre montré que le  $K_D$  ne permettait pas de faire de lien direct avec l'activation. La comparaison de différents peptides (agonistes, faibles agonistes, antagonistes et nuls) montre que la cinétique de dissociation ( $K_{off}$ ) peut être discriminante (Alam *et al.*, 1999). D'une façon plus générale, il semble désormais admis que la demi-vie du complexe TCR/pMHC montre une bonne corrélation avec l'effet biologique (Matsui *et al.*, 1994) ; elle semble comprise entre 2 et 10 s (Savage *et al.*, 1999) lorsqu'il y a reconnaissance. Ces données cinétiques ont permis d'élaborer des modèles cinétiques de réponses (McKeithan, 1995), permettant de comprendre comment le lymphocyte T est capable d'une très grande sélectivité, avec un récepteur possédant une faible affinité pour son ligand.

### A.3.5 Signal d'activation du lymphocyte T

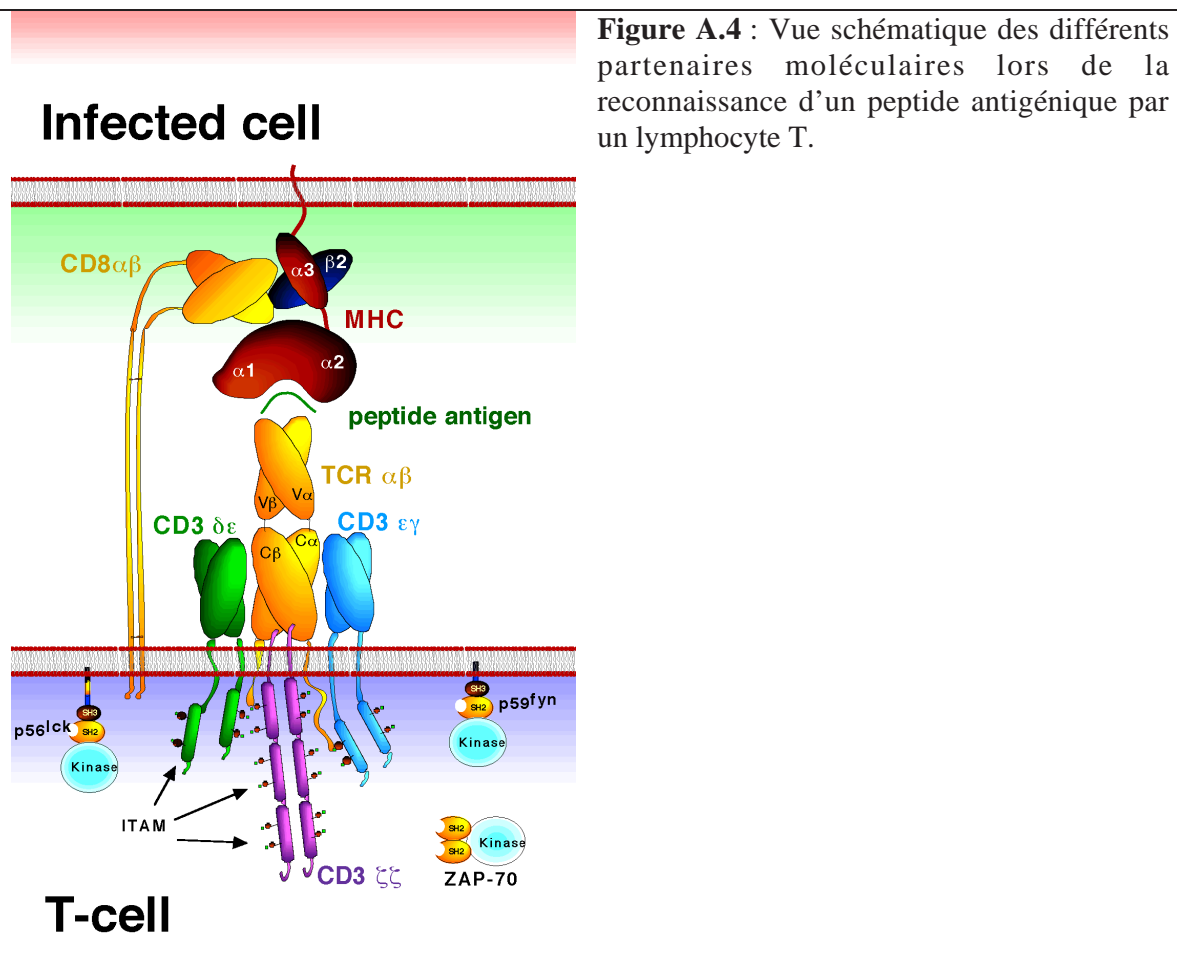
Une question importante pour laquelle il n'y a pas encore de réponse définitive est la façon dont la cellule T est informée que son ou ses récepteurs sont engagés dans une interaction avec un antigène. Deux hypothèses différentes sont généralement avancées.

Une première suppose un changement structural du TCR lors de la reconnaissance d'un complexe pMHC, qui permettrait de traduire à l'intérieur de la cellule l'évènement de reconnaissance. Cette hypothèse s'appuie sur le fait que le TCR présente une certaine rigidité par rapport aux anticorps, notamment grâce à l'interface  $V\beta/C\beta$  importante (800  $\text{\AA}^2$  enfouie, Bentley *et al.*, 1995). Cependant, les différentes structures de TCR, complexées ou non au MHC, n'ont montré aucun changement de conformation significatif, notamment à l'interface  $V\beta/C\beta$ . D'autre part, l'interface montre tout de même une certaine flexibilité. Cette hypothèse est donc de moins en moins étayée par les récents résultats structuraux (Garcia *et al.*, 1999).

La deuxième hypothèse postule que le signal d'activation est transmis grâce à un phénomène d'oligomérisation des TCR et des co-récepteurs, permettant d'activer une cascade de réactions par l'intermédiaire de tyrosine-kinases (Kolanus *et al.*, 1993, Weis *et al.*, 1994). En effet, chacun des domaines CD3 (associés au  $\text{TCR}\alpha\beta$ ) possède un ou deux motifs ITAM (*Immunoreceptor Tyrosine-based Activation Motif*) sur leur segment intracellulaire (1 ITAM pour les chaînes  $\gamma$ ,  $\delta$ ,  $\epsilon$  ; 2 ITAM pour la chaîne  $\zeta$ ). L'oligomérisation des TCR aurait pour

conséquence le recrutement de la tyrosine-kinase  $p59^{fyn}$ , sans doute associée à la membrane, alors que la formation du complexe TCR/pMHC/CD8 ou TCR/pMHC/CD4 permettrait, en plus, le recrutement de la tyrosine-kinase  $p56^{lck}$ , associée de façon non-covalente à l'extrémité intracellulaire du co-récepteur (figure A.4). Ces kinases pourront alors phosphoryler les motifs ITAM. Une fois phosphorylés, les ITAM vont pouvoir recruter une autre tyrosine-kinase, ZAP70 ( $\zeta$  Associated Protein), qui va se lier aux ITAM doublement phosphorylés de la chaîne  $\zeta$  et s'activer en étant elle-même phosphorylée. Ainsi commence la chaîne de réaction intracellulaire qui aboutira à l'activation de la cellule T.

Ce schéma permet d'expliquer plusieurs modes d'activation, comme celui par les anticorps bivalents (Désiré-1, H57, etc., Meuer *et al.*, 1983), où la seule dimérisation des TCR par un anticorps spécifique suffit à enclencher la prolifération (la kinase  $p59^{fyn}$  peut donc suffire à déclencher la cascade d'activation). Par contre, le Fab monovalent de ces mêmes anticorps ne provoque aucune activation. De façon similaire, des multimères de pMHC solubles sont capables d'activer les lymphocytes T (Boniface *et al.*, 1998). Le rôle du co-récepteur est de favoriser la stabilité du complexe TCR/pMHC (Luescher *et al.*, 1990, Garcia *et al.*, 1996a), en interagissant avec le MHC avec une faible affinité et indépendamment du TCR (Wyer *et al.*, 1999). Il facilite l'activation des ITAMs, mais n'est pas forcément indispensable, notamment si le TCR possède une grande affinité pour le pMHC.



Par contre, la façon dont les complexes TCR/pMHC peuvent dimériser reste un mystère. Même si la molécule MHC classe II cristallise sous la forme de dimère, montrant la

plausibilité de tel dimère en surface des APC, aucune des structures cristallographiques de complexe TCR/pMHC ne montre une telle organisation (Garcia *et al.*, 1999).

Ce mécanisme nous éclaire aussi un peu plus sur la façon dont le TCR peut faire la différence entre deux peptides dont l'affinité diffère très peu. Il cadre mieux avec l'idée que la réponse du lymphocyte n'est pas du tout ou rien, mais quelque chose de graduel, en fonction du nombre de complexes TCR/pMHC qui peuvent être formés à l'interface cellule T : cellule présentatrice.

#### **A.3.5.1 L'activation des cellules T naïves nécessite deux signaux**

Pour être activée et proliférer, la cellule T naïve doit recevoir simultanément deux signaux : un signal de reconnaissance de l'antigène, par l'intermédiaire du TCR $\alpha\beta$ , et un signal co-stimulant, sans doute par l'intermédiaire des molécules B7 (CD80 et CD86) à la surface des APC. Cette exigence est essentielle pour le maintien de la tolérance vis-à-vis des peptides du soi.

Seuls trois types de cellules APC sont capables de fournir ce deuxième signal : les cellules dendritiques dans les tissus lymphoïdes, les macrophages, après ingestion d'un micro-organisme, et les lymphocytes B, en présence de certains constituants bactériens. Les cellules dendritiques sont en permanence capables de co-stimuler les lymphocytes T et sont plus spécifiquement chargées de présenter des fragments de protéines virales. La phagocytose déclenche l'expression des molécules MHC classe II et des molécules B7 du macrophage. Ce dernier acquiert alors la fonction d'APC. Son rôle est principalement d'activer les lymphocytes T contre les microbes. Les lymphocytes B, en présence de constituants bactériens comme les lipopolysaccharides, sont capables d'internaliser les récepteurs membranaires liés à un antigène et de présenter des fragments peptidiques de l'antigène au moyen des molécules MHC de classe II, tout en exprimant les molécules B7 permettant la co-stimulation.

L'activation des cellules T naïves est donc possible dans un contexte très restreint. De plus, lorsqu'une cellule T naïve reçoit un signal de reconnaissance sans signal co-stimulateur, elle entre dans un état d'anergie. C'est à dire qu'elle devient inactivable, même par une APC. Ce mécanisme permet donc de rendre tolérant les lymphocytes reconnaissant des peptides du soi en dehors du système lymphatique ou du cadre d'une infection.

Ce processus est sans doute une des pierres angulaires de la sélectivité soi/non-soi de la réponse cellulaire. En imposant aux lymphocytes T d'être activés dans un contexte du soi semblable à celui de la sélection intrathymique, on évite toute activation par des antigènes du soi absents du système lymphatique. En conséquence, le répertoire de lymphocytes T issu de la sélection thymique n'est que partiellement tolérant au soi. C'est le contexte de l'activation qui permettra de compléter cette sélectivité. De fait, chaque lymphocyte T peut reconnaître une certaine gamme de peptides dont certains sont endogènes (Evavold *et al.*, 1995, Kersh *et al.*, 1996).

### **A.4 Structure des différents partenaires de la reconnaissance**

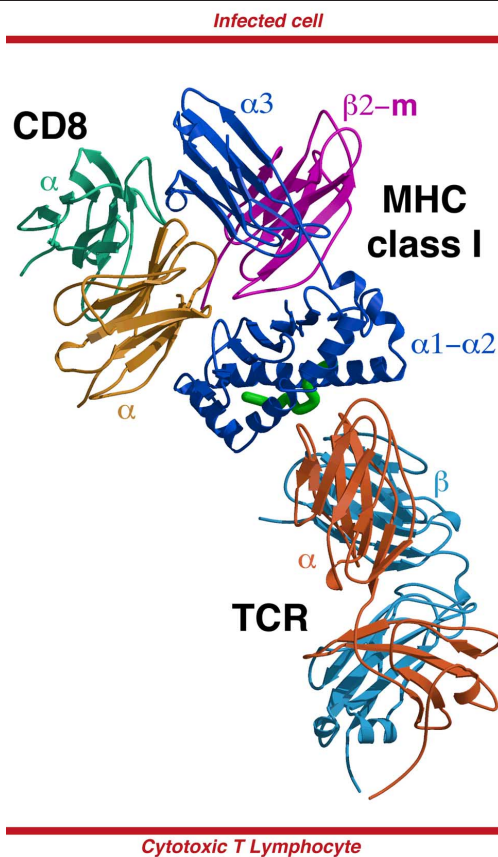
La connaissance de la structure à l'échelle atomique des différents partenaires impliqués dans la reconnaissance antigénique a été une étape cruciale pour la compréhension de la double spécificité du TCR pour le peptide et la molécule MHC. Comme parallèle, l'ensemble des études structurales sur les anticorps, débuté il y a près d'un quart de siècle, a permis des avancées considérables dans la compréhension de la base structurale de la spécificité anticorps-antigène (Chothia *et al.*, 1989, Padlan, 1994). De nombreux groupes de par le monde se sont lancés dans cette tâche de longue haleine que représentait la résolution de la



structure des différents partenaires impliqués dans la reconnaissance de l'antigène par le lymphocyte T (TCR $\alpha\beta$ , molécule MHC classe I et classe II, co-récepteurs CD4 et CD8, CD3). La détermination en 1984, de la première séquence d'un TCR par le groupe M. Davis (Hedrick *et al.*, 1984) laissait espérer une issue rapide de ces travaux, mais un certain nombre de difficultés devaient être au rendez-vous. Les fragments solubles de TCR se sont avérés difficiles à sur-exprimer en grande quantité, malgré la très grande variété de systèmes d'expression et constructions testés. L'obtention de cristaux de bonne qualité allait être également ardue.

Il aura fallu attendre 1995 pour que la structure d'une première chaîne de TCR soit résolue (Bentley *et al.*, 1995). L'année suivante, la structure de deux TCR complets en complexe avec une molécule MHC classe I était résolue et donnait une première image de l'interface TCR/pMHC (Garboczi *et al.*, 1996, Garcia *et al.*, 1996b).

Le bilan des dix dernières années est riche de résultats qui permettent à l'heure actuelle d'avoir une bonne image de la plupart des partenaires impliqués (MHC classe I (Bjorkman *et al.*, 1987) et classe II (Brown *et al.*, 1993), CD8 (Leahy *et al.*, 1992), CD4 (Ryu *et al.*, 1990, Brady *et al.*, 1993), TCR $\alpha\beta$ , ainsi que les complexes TCR $\alpha\beta$ /pMHC et CD8/pMHC, tableau A.2, figure A.5). Manque encore CD3, le complexe CD4/pMHC et des complexes TCR/pMHC classe II. Mais l'ensemble des structures dont on dispose permet déjà de comprendre les bases structurales de la reconnaissance TCR / antigène.



**Figure A.5 :** Modèle d'un complexe CD8/pMHC/TCR $\alpha\beta$  construit sur la base des structures cristallographiques des complexes CD8 $\alpha$ /pHLA-A2 (Gao *et al.*, 1997), TCR A6/pHLA-A2 (Garboczi *et al.*, 1996) et TCR 2C (Garcia *et al.*, 1996b). Ce modèle confirme bien la façon dont le TCR et CD8 peuvent interagir simultanément avec le pMHC.

**Table A.2** : Structures cristallographiques de TCR seul ou en complexe, connues à ce jour, et pour la plupart, déposée dans la *Protein Data Bank* (Bernstein *et al.*, 1977).

Type de structure	Domaines	Organisme	Résolution (Å)	Code PDB	Référence
TCR 14.3.d (V $\beta$ 8.2)	$\beta$	<i>Mus musculus</i>	1.7	1bec	Bentley <i>et al.</i> , 1995
TCR 1934.4 (V $\alpha$ 4.2J $\alpha$ 4.0)	V $\alpha$ :V $\alpha$	<i>Mus musculus</i>	2.3	nd	Fields <i>et al.</i> , 1995a
TCR 1934.4 (V $\alpha$ 4.2J $\alpha$ 4.0 muté)	V $\alpha$ :V $\alpha$	<i>Mus musculus</i>	2.3	1ac6	Li <i>et al.</i> , 1997
TCR 2C (V $\alpha$ 3J $\alpha$ 58C $\alpha$ :V $\beta$ 8.2D $\beta$ 2J $\beta$ 2.4C $\beta$ 2)	$\alpha$ : $\beta$	<i>Mus musculus</i>	2.5	1tcr	Garcia <i>et al.</i> , 1996b
TCR A6/tax/HLA-A2 (hV $\alpha$ 2.3J $\alpha$ 24C $\alpha$ :hV $\beta$ 12.3D $\beta$ 2J $\beta$ 2.7C $\beta$ 2)	$\alpha$ : $\beta$ / p / $\alpha$ : $\beta$ 2-m	<i>Homo sapiens</i>	2.6	1ao7	Garboczi <i>et al.</i> , 1996
TCR A6/pept/HLA-A2 (hV $\alpha$ 2.3J $\alpha$ 24C $\alpha$ :hV $\beta$ 12.3D $\beta$ 2J $\beta$ 2.7C $\beta$ 2)	$\alpha$ : $\beta$ / p / $\alpha$ : $\beta$ 2-m	<i>Homo sapiens</i>	2.8		Ding <i>et al.</i> , 1999
HLA-A2/CD8 $\alpha\alpha$	$\alpha$ : $\beta$ / $\alpha$ : $\alpha$	<i>Homo sapiens</i>	2.7	1akj	Gao <i>et al.</i> , 1997
TCR KB5-C20/Désiré-1 Fab (V $\alpha$ 2.3J $\alpha$ A10:V $\beta$ 2D $\beta$ 2J $\beta$ 2.3)	scFv V $\alpha$ :V $\beta$ / L:H	<i>Mus musculus</i>	2.6	1kb5	Housset <i>et al.</i> , 1997
TCR ES204 (V $\delta$ 3D $\delta$ 2J $\delta$ 1)	V $\delta$	<i>Homo sapiens</i>	1.9	1tvd	Li <i>et al.</i> , 1998
TCR N15/H57 Fab (V $\alpha$ 8J $\alpha$ 5C $\alpha$ :V $\beta$ 5.2D $\beta$ 2J $\beta$ 2.6C $\beta$ 2)	$\alpha$ : $\beta$ / L:H	<i>Mus musculus</i>	2.8	1nfd	Wang <i>et al.</i> , 1998
TCR 2C/dEV8/MHC-H-2K <sup>b</sup> (V $\alpha$ 3J $\alpha$ 58C $\alpha$ :V $\beta$ 8.2D $\beta$ 2J $\beta$ 2.4C $\beta$ 2)	$\alpha$ : $\beta$ / p / $\alpha$ : $\beta$ 2-m	<i>Mus musculus</i>	3.0-4.2	2ckb	Garcia <i>et al.</i> , 1998
TCR B7/tax/HLA-A2 (hV $\alpha$ 17.2J $\alpha$ 54C $\alpha$ :hV $\beta$ 12.3D $\beta$ 2J $\beta$ 2.7C $\beta$ 2)	$\alpha$ : $\beta$ / p / $\alpha$ : $\beta$ 2-m	<i>Homo sapiens</i>	2.5	1bd2	Ding <i>et al.</i> , 1998
TCR V $\alpha$ 2.6J $\alpha$ 38 (V $\alpha$ 2.6J $\alpha$ 38)	V $\alpha$ :V $\alpha$	<i>Mus musculus</i>	2.5	1b88	Plaksin <i>et al.</i> , 1999
TCR 14.3.d/Sag EB (V $\beta$ 8.2)	$\beta$	<i>Mus musculus</i>	2.4	1sbb	Li <i>et al.</i> , 1998
TCR 14.3.d/Sag EC3 (V $\beta$ 8.2)	$\beta$	<i>Mus musculus</i>	3.5	1jck	Fields <i>et al.</i> , 1996

#### A.4.1 La structure du TCR

Le TCR est un complexe multicaténaire dans lequel les fonctions de reconnaissance et de couplage sont l'attribut de chaînes polypeptidiques distinctes. Ainsi, parmi les 8 chaînes polypeptidiques trans-membranaires participant à la constitution de ce complexe (la stœchiométrie exacte de ce complexe est toujours incertaine), seules les chaînes TCR $\alpha$  et TCR $\beta$  sont directement impliquées dans la fonction de reconnaissance antigénique. Les chaînes CD3  $\gamma\epsilon$ ,  $\epsilon\delta$  et  $\zeta\zeta$  sont impliquées dans la transduction du signal d'activation (cf. §A.2.5, figure A.4).

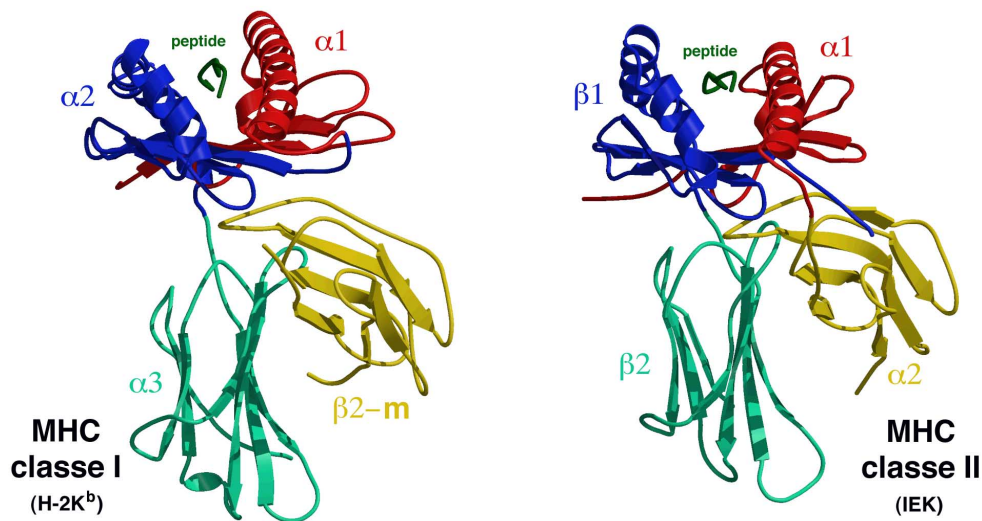
Comme le révélaient déjà la structure des gènes et la séquence, la structure des TCR $\alpha\beta$  présente de nombreuses similitudes avec les fragments Fab des anticorps. Ils sont formés de deux chaînes comportant chacune une région variable (V) amino-terminale et une région constante (C) carboxy-terminale. Les régions V $\alpha$  et V $\beta$ , localisées de manière distale par rapport au plan de la membrane, interagissent entre elles afin de former un site de fixation antigénique partagé. Les domaines C $\alpha$  et C $\beta$  interagissent également entre eux, avec un pont disulfure interchaîne, et se terminent par une extrémité C-terminale trans-membranaire. La

structure des domaines  $V\alpha$  et  $V\beta$  rappelle celle des domaines  $V_L$  et  $V_H$ , faite d'un sandwich de deux feuillets  $\beta$ . Le site de reconnaissance de l'antigène est formé de trois boucles hypervariables par domaine, appelées CDR pour *Complementarity Determining Region*. Une quatrième région hypervariable, HV4, spécifique des TCR se situe aux extrémités du site antigénique. Du fait du mécanisme de réarrangement intrathymique, la variabilité des régions CDR3, à cheval sur les segments V (D) J, est bien plus importante que celle des autres régions hypervariables, limitée par le nombre de segments codant pour les  $V\alpha$  et les  $V\beta$  (Davis & Bjorkman, 1988, Jores *et al.*, 1990).

#### A.4.2 Les molécules MHC

La structure tridimensionnelle des molécules MHC de classe I et II est similaire (rms différence de l'ordre de 1.8 Å, Madden, 1995). Par contre la connectivité des quatre domaines diffère (figure A.6).

La molécule MHC de classe I est formée de deux chaînes polypeptidiques. La première, appelée aussi chaîne lourde, codée dans le gène du complexe majeur d'histocompatibilité, composée de trois domaines, comporte le site de fixation du peptide ( $\alpha1$ - $\alpha2$ ) et un domaine  $\alpha3$  qui s'ancre dans la membrane. La deuxième chaîne est la  $\beta2$ -microglobuline, de repliement de type Ig également ; elle interagit de façon non-covalente avec les domaines  $\alpha3$  et  $\alpha1$ - $\alpha2$ . Le site de fixation du peptide est une gorge profonde formée par deux longues hélices  $\alpha$  posées sur un large feuillet  $\beta$  antiparallèle. Des peptides de huit à dix acides aminés peuvent être fixés dans cette gorge. Les résidus variables liés à la nature polygénique et polymorphique de la molécule MHC sont pour la plupart situés dans la gorge de fixation du peptide ou à la surface des hélices  $\alpha1$  et  $\alpha2$ .



**Figure A.6** : Comparaison des structures des molécules MHC de classe I (pept. KB5-C20/H-2K<sup>b</sup>, résultat non publiés) et II (I-E<sup>k</sup>, Fremont *et al.*, 1996), présentées dans la même orientation.

Les molécules MHC de classe II sont également formées de deux chaînes. La chaîne  $\alpha$  est constituée de deux domaines  $\alpha1$  et  $\alpha2$  (équivalents aux domaines  $\alpha1$  et  $\beta2$ -m des MHC classe I) et d'une extrémité C-terminale ancrée dans la membrane cellulaire. La chaîne  $\beta$  est elle aussi constituée de deux domaines  $\beta1$  et  $\beta2$  équivalents aux domaines  $\alpha2$  et  $\alpha3$  des MHC classe I et est aussi ancrée à la membrane par son extrémité C-terminale. La gorge de fixation

du peptide est également formée par les hélices  $\alpha$  des domaines  $\alpha 1$  et  $\beta 1$ , mais plus ouverte aux extrémités. De ce fait, elle peut fixer des peptides plus longs (au moins 13 résidus).

#### A.4.3 L'interface TCR / pMHC

A l'heure actuelle, trois structures de complexe ternaire TCR/pMHC ont été déterminées par cristallographie. 2C/dEV8/H-2K<sup>b</sup> (Garcia *et al.*, 1998) est un complexe entre un TCR murin, un peptide mitochondrial du soi et la molécule MHC H-2K<sup>b</sup> spécifique de ce TCR. Les deux autres complexes concernent deux TCR humains A6 et B7, reconnaissant tous deux le peptide Tax, issu du virus HTLV-1, présenté par la molécule HLA-A2 (Garboczi *et al.*, 1996, Ding *et al.*, 1998, 1999).

Ces trois structures montrent un mode d'interaction commun entre le TCR et la molécule MHC. Le TCR présente son site de reconnaissance de l'antigène (formé par les boucles CDR et HV4) diagonalement par rapport aux axes des hélices  $\alpha 1$  et  $\alpha 2$  du MHC (figure A.7, A.10). D'un point de vue stérique, l'axe d'élongation du site antigénique du TCR se positionne entre les deux points hauts de la surface formée par les deux hélices  $\alpha$  du MHC et le peptide (figure A.7). Le domaine V $\alpha$  est situé au-dessus de l'extrémité N-terminale du peptide et le domaine V $\beta$  au-dessus de l'extrémité C-terminale du peptide. Ce mode diagonal d'interaction n'avait pas été prédit par les premières tentatives de modélisation du complexe ternaire, qui se fondaient sur l'hypothèse que les CDR1 et 2, moins variables, interagissaient exclusivement avec le MHC, alors que le CDR3, très variable et central, reconnaissait le peptide (Novotny *et al.*, 1986, Chotia *et al.*, 1988, Claverie *et al.*, 1989, Davis & Bjorkmann, 1988, Hong *et al.*, 1992, Jorgensen *et al.*, 1992). Seules deux études de mutagenèse dirigée sur les MHC de classe II avaient proposé une orientation diagonale du TCR par rapport aux hélices  $\alpha$  du MHC (Sun *et al.*, 1995, Sant'Angelo *et al.*, 1996). Un autre mode d'interaction, où le TCR est tourné de 180° selon le pseudo-axe 2 de son domaine variable, a été prédit par des études de mutagenèse dirigée (Sim *et al.*, 1997) ; il n'a pour l'instant jamais été confirmé par une structure.

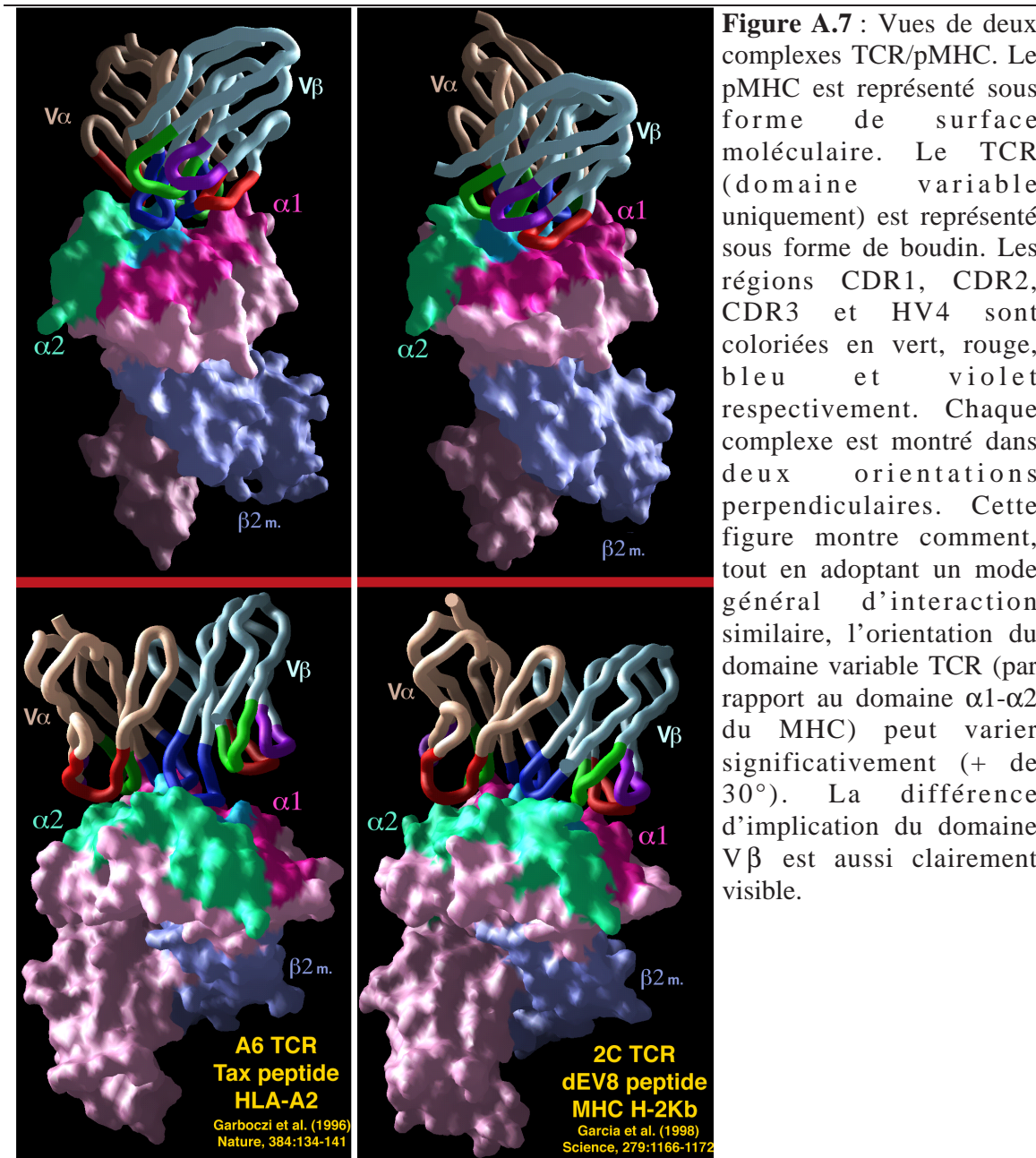
L'interaction du TCR n'est pas symétrique par rapport aux chaînes  $\alpha$  et  $\beta$ . Dans tous les complexes dont la structure a été déterminée, une remarquable conservation de l'interface V $\alpha$ /pMHC est observée. Il semble que l'interaction V $\alpha$ /pMHC soit prépondérante et guide l'orientation du TCR par rapport au pMHC. L'implication du V $\beta$  est beaucoup plus variée. Dans le cas de 2C/dEV8/H-2K<sup>b</sup>, tous les CDR du domaine V $\beta$  sont impliqués, formant 45% de l'interface TCR/pMHC. Ce TCR se caractérise par des boucles CDR3 particulièrement courtes (7 et 6 résidus) qui interviennent peu dans l'énergie d'interaction 2C/pH-2K<sup>b</sup> (Manning *et al.*, 1998). Dans le complexe B7/tax/HLA-A2, seul le CDR2 $\beta$  et le CDR3 $\beta$  interagissent avec le pMHC. Dans le dernier complexe (A6/tax/HLA-A2), le domaine V $\beta$  n'intervient que par l'intermédiaire du CDR3 (figure A.7).

Tout en conservant le mode diagonal d'interaction, l'orientation relative du TCR par rapport au MHC peut ainsi varier de plus de 30°, et dépend sans doute en partie de la longueur des boucles CDR3 (figure A.7). Une longue boucle CDR3 $\beta$  va sans doute imposer un mode d'interaction proche de celui observé pour le complexe TCR A6 / tax / HLA-A2, où le CDR3 $\beta$  est largement prépondérant dans l'interface V $\beta$ /pMHC.

La surface enfouie dans l'interface est d'environ 1000 Å<sup>2</sup> pour le TCR ou le pMHC. Le peptide, avec entre 220 (octapeptide dEV8) et 320 (nonapeptide tax) Å<sup>2</sup> impliqués dans cette interface, ne représente que 20 à 30 % de la surface enfouie.

Les boucles CDR1 sont plus vers les extrémités du peptide et peuvent interagir avec le peptide et le MHC. Les boucles CDR2 interagissent essentiellement avec le MHC: le CDR2 $\alpha$  avec l'hélice  $\alpha 2$  et le CDR2 $\beta$  avec l'hélice  $\alpha 1$ . Les boucles CDR3 sont au centre de l'interface TCR/MHC et interagissent avec le peptide et le MHC.





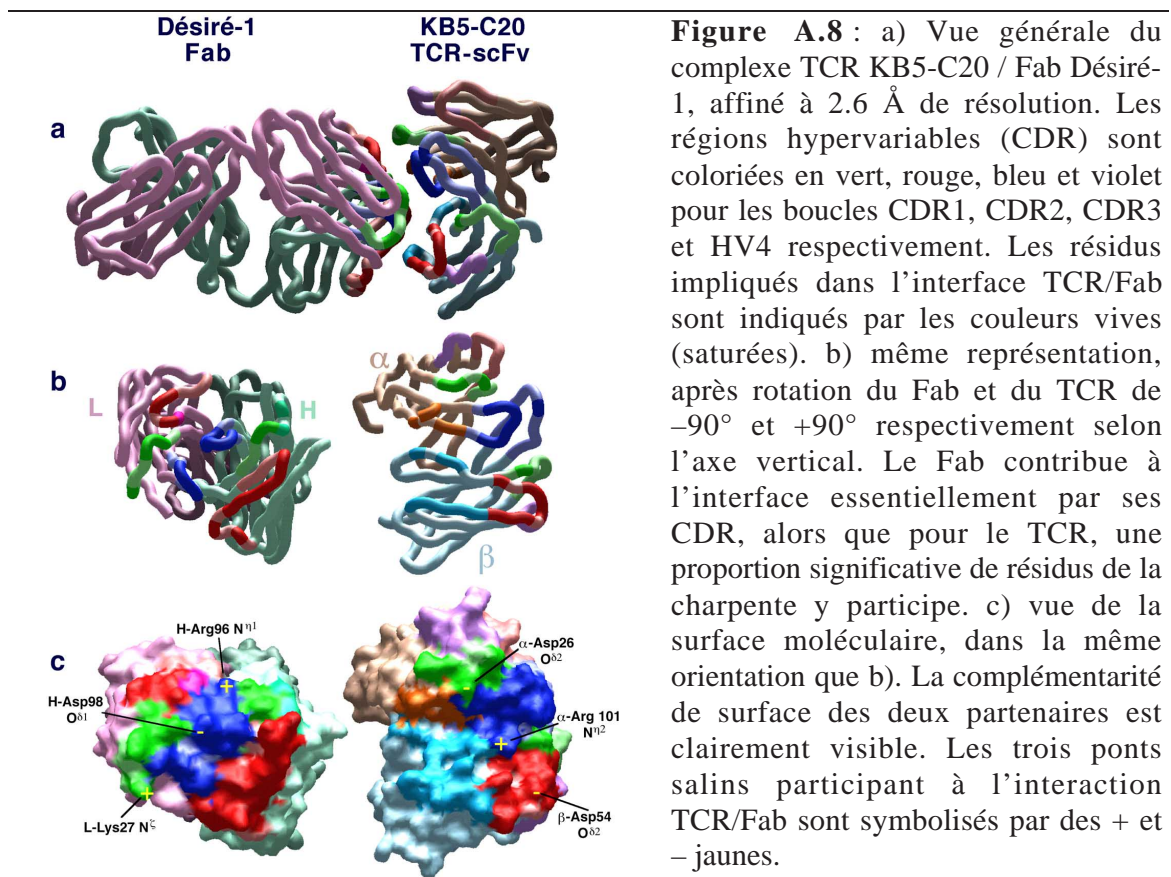
Peu de changements conformationnels sont observés entre le pMHC seul et en complexe avec le TCR. Par contre, les CDR du TCR peuvent se déplacer ou changer de conformation de façon significative lors de l'interaction avec le pMHC. Pour le TCR 2C, des changements de 3 à 6 Å sont mis en évidence pour les CDR1 et 3 (García *et al.*, 1998). Ces changements confirment les mesures thermodynamiques indiquant qu'une barrière d'énergie doit être franchie lors de l'association et la dissociation du complexe TCR/pMHC. La contribution entropique indique une diminution significative du désordre lors de la formation du complexe (Willcox *et al.*, 1999). La flexibilité des boucles CDR3 est également observée pour un TCR scFv, dont la structure a été résolue par RMN (Hare *et al.*, 1999). La conformation des boucles CDR2 est très conservée, conséquence probable de leur interaction avec les hélices  $\alpha$  du MHC.

D'un point de vue plus général, l'interface TCR/pMHC se caractérise par une faible complémentarité de surface, lorsqu'on la compare aux interfaces protéine-inhibiteur ou anticorps-antigène (Ysern *et al.*, 1998)

La comparaison des complexes A6/tax/HLA-A2 et B7/tax/HLA-A2 permet d'observer comment deux TCR différents reconnaissent une même cible. Seul 1 des 17 résidus de B7 en contact avec le pMHC est conservé dans le TCR A6. Par contre, les deux TCR ont en commun 13 positions impliquées dans l'interaction avec la molécule MHC. De plus, si les résidus ne sont pas conservés, la nature des interactions est souvent conservée (Ding *et al.*, 1998). Cet exemple donne une idée de la variété des modes de reconnaissance que peuvent avoir les TCR, malgré le contexte restrictif de présentation de l'antigène.

## A.5 La structure du TCR KB5-C20

Notre projet, fruit d'une collaboration avec le CIML de Marseille-luminy qui a débuté en 1990, avait pour objectif de déterminer la structure du récepteur pour l'antigène d'un lymphocyte T cytotoxique murin KB5-C20 (Albert *et al.*, 1982). Différents fragments recombinants de TCR (V $\alpha$ , V $\beta$ , domaine variable *single chain* (scFv : V $\alpha$  relié au V $\beta$  par un linker hydrophile de 26 amino-acides)) ont été successivement exprimés dans des cellules de myélome de souris (Grégoire *et al.*, 1996), purifiés et éventuellement cristallisés.



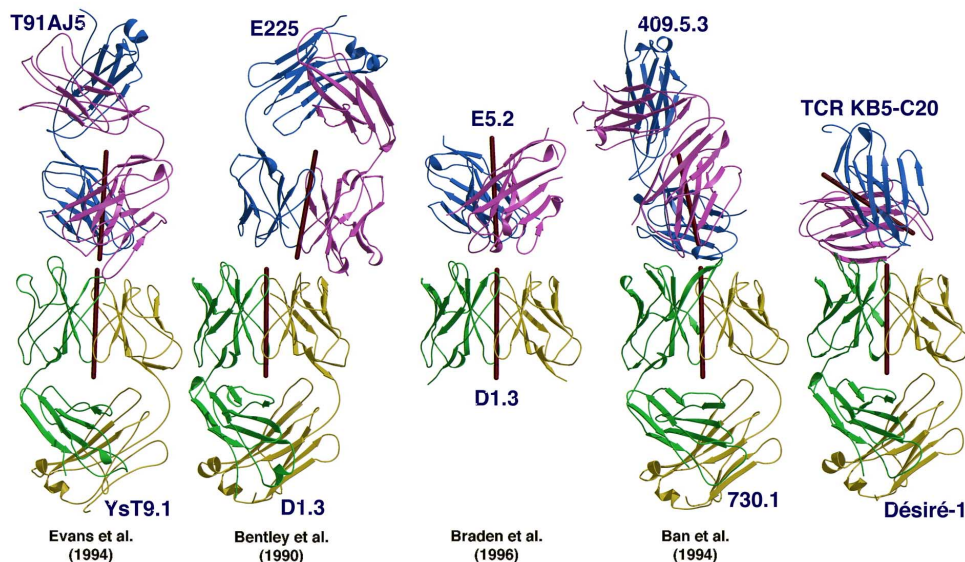
Deux étapes critiques ont été difficiles à franchir : la production en quantité suffisante de protéines et la cristallisation. Ce n'est qu'en avril 96 que les premières données cristallographiques exploitables (3.0 Å de résolution) ont été obtenues à l'ESRF, sur un cristal de complexe entre le TCR scFv recombinant et un anticorps anti-clonotypique Désiré-1

(Grégoire *et al.*, 1991, Hua *et al.*, 1985). Grâce à ces données, la structure a pu être résolue en septembre 1996. Suite à l'enregistrement d'un deuxième jeu de données, la structure, affinée à 2,6 Å de résolution, a été publiée en 97 (Housset *et al.*, 1997, figure A.8).

Cette première structure d'un TCR scFv a pu être comparée aux structures de TCR publiées. D'autre part, le complexe TCR KB5-C20 / Fab Désiré-1 est le premier complexe clonotype/anti-clonotype dont la structure est connue. L'interaction entre le TCR et l'anticorps anti-clonotypique a donc été analysée en détail.

### A.5.1 Interface du complexe clonotype (TCR KB5-C20) / anti-clonotype (Fab Désiré-1)

L'homologie structurale entre TCR et anticorps nous conduit naturellement à comparer ce complexe avec les complexes anticorps-anticorps (Bentley *et al.*, 1990, Ban *et al.*, 1994, Braden *et al.*, 1996, Evans *et al.*, 1994). De tels complexes entre un anticorps (Ab1) et un autre anticorps (Ab2), spécifique du déterminant antigénique (idiotope) d'Ab1 (d'où le terme de complexe idiotype / anti-idiotype), ont permis d'estimer dans quelle mesure Ab2 était capable de « mimer » l'antigène d'Ab1 (Fields *et al.*, 1995b), et d'étudier, sur le plan structural, la notion de réseau idyotopique (Ab1, Ab2, Ab3, ..., Pan *et al.*, 1995).



**Figure A.9 :** Comparaison de quatre complexes idiotype/anti-idiotype avec le complexe clonotype/anti-clonotype KB5-C20/Désiré-1. L'orientation relative des domaines variables du TCR et du Fab est différente de celle observée pour les complexes anticorps/anticorps, avec un angle de 122° entre les deux pseudo-axe 2 des domaines variables.

La surface accessible enfouie à l'interface TCR/Fab est de 1350 Å<sup>2</sup> pour chacun des partenaires. Elle est nettement supérieure à celle impliquée dans les complexes idiotype/anti-idiotype (≈ 900 Å<sup>2</sup>). Cette différence d'interface est liée au fait que le site antigénique de l'anticorps ne fait pas globalement face au site homologue du TCR, contrairement aux complexes idiotype/anti-idiotype (figure A.9), mais interagit avec l'un des côtés du site antigénique, dans un sillon formé à l'interface des domaines Vα et Vβ (figure A.8). Ainsi, seuls 60% des résidus du TCR qui interagissent avec le Fab appartiennent aux régions CDR. Toutefois, le Fab reconnaît effectivement des résidus des CDR3 α et β, confirmant le caractère anti-clonotype de Désiré-1 (figure A.8b). Du fait de cette interaction latérale, le Fab

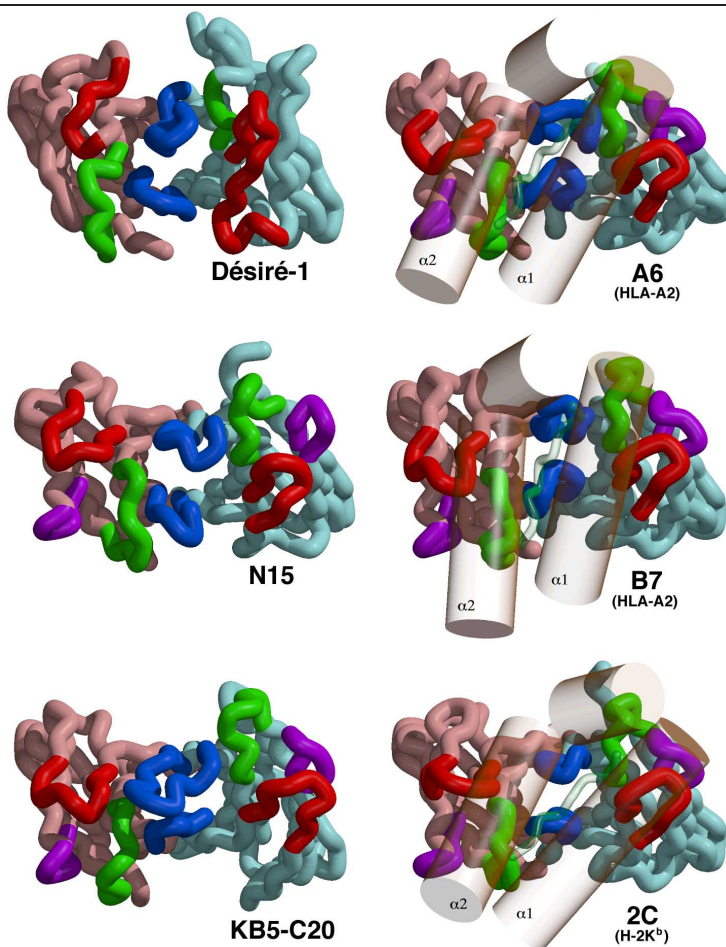


ne peut « mimer » la molécule MHC. Les détails de cette interface sont développés par Mazza *et al.*, 1999.

### A.5.2 Structure du scFv

La structure tridimensionnelle du domaine scFv, malgré la présence d'un *linker* artificiel, est comparable aux domaines variables des autres TCR, confirmant le bon repliement de ce domaine recombinant, déjà suggéré par la bonne affinité de l'anticorps anticlonotypique Désiré-1 pour cette construction (figure A.10). Elle confirmait donc l'intérêt de travailler sur ce type de fragment dans le but d'étudier la structure du site de fixation de l'antigène.

La topologie, commune à tous les TCR dont la structure est connue, est proche de celle des domaines variables des anticorps et ne diffère principalement que par le déplacement du brin  $\beta$  c'' du domaine  $V\alpha$ , du feuillet  $\beta$  interne vers le feuillet  $\beta$  externe. Ce déplacement, appelé « *switch* », fut observé pour la première fois sur le  $V\alpha$  1934.4 (Fields *et al.*, 1995) et associé à une éventuelle dimérisation des TCR. Comme nous le verrons plus loin, cette spécificité topologique peut aussi avoir une origine fonctionnelle pour la reconnaissance du MHC. Le site de reconnaissance de l'antigène est structuralement bien conservé, sauf pour les boucles CDR3, dont la longueur et la conformation sont très variables (figure A.10).

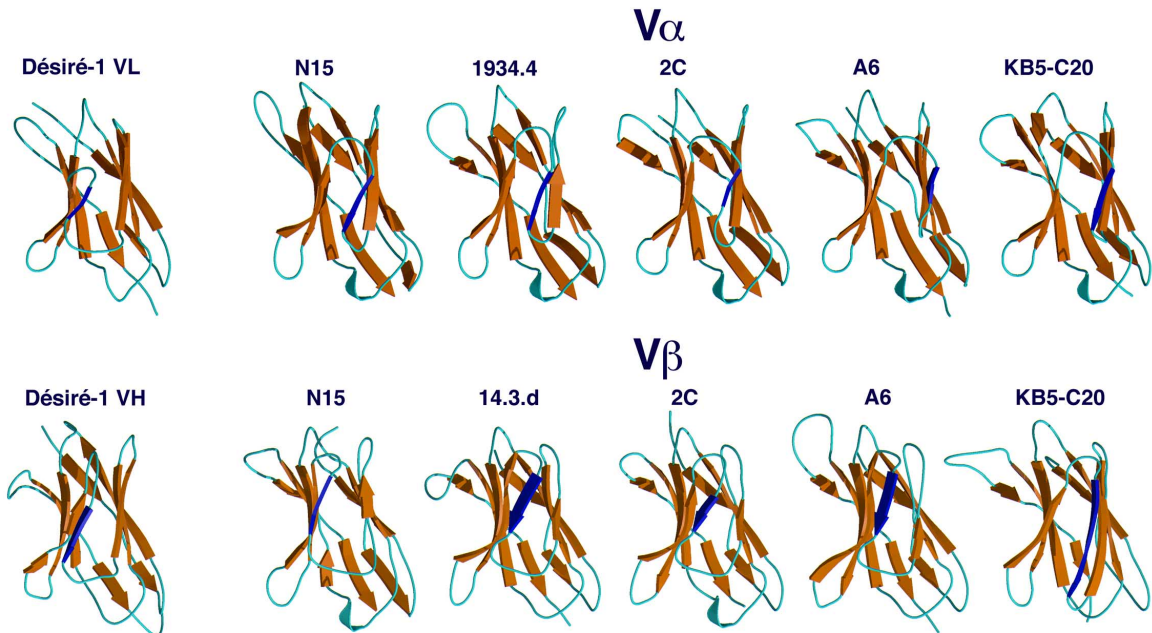


**Figure A.10** : Vue des domaines variables de différents TCR et du Fab Désiré-1. Les domaines  $V\alpha$  (ou  $V_L$ ) sont représentés en rouge pale, les domaines  $V\beta$  (ou  $V_H$ ) en cyan, les boucles CDR1, CDR2, CDR3 et HV4 en vert, rouge, bleu et violet respectivement. Dans les cas où la structure du complexe TCR/pMHC est disponible, les hélices  $\alpha 1$  et  $\alpha 2$  ont été représentées par des cylindres transparents et le peptide en vert translucide. La présence du *switch* sur le domaine  $V\alpha$  modifie la conformation du CDR2 $\alpha$  et ouvre un large sillon sur le coté de l'interface  $V\alpha/V\beta$ . Le *switch* du  $V\beta$  sur KB5-C20, par contre, n'a pas d'effet majeur sur le CDR2 $\beta$ . L'orientation diagonale du TCR par rapport aux hélices  $\alpha$  du MHC est bien mise en évidence.



### A.5.3 Le "switch" du brin $\beta$ c''.

Dans tous les domaines de type Immunoglobuline, le brin  $\beta$  c'' (en bleu, figure 2) appartient au feuillet  $\beta$  interne (formant l'interface  $V_L/V_H$ ). Par contre, tous les domaines  $V\alpha$  ont un brin  $\beta$  c'' "switché" appartenant au feuillet  $\beta$  externe. Les domaines  $V\beta$  connus jusqu'à maintenant présentent tous un brin  $\beta$  c'' non "switché", comparable aux domaines  $V_L$  et  $V_H$  des anticorps. Le domaine  $V\beta$  de KB5-C20 possède, quant à lui, un brin  $\beta$  c'' particulièrement long qui appartient au feuillet  $\beta$  externe, rappelant ainsi la topologie du domaine  $V\alpha$  (figure A.11).



**Figure A.11** : Vue des différents domaines variables  $V\alpha$  et  $V\beta$  dont la structure est connue, présentés dans la même orientation. Le feuillet  $\beta$  interne est à gauche, le feuillet  $\beta$  externe à droite. Le brin  $\beta$  c'' est représenté en bleu. Les domaines  $V_L$  et  $V_H$  du Fab Désiré-1 sont représentés pour illustrer la topologie des domaines variables des anticorps.

Cette topologie unique de ce  $V\beta$  a pu être analysée grâce à l'alignement de séquence basé sur la structure des différents domaines  $V\alpha$  et  $V\beta$  disponibles (figure A.12, Mazza *et al.*, 1998). Cet alignement montre que les domaines  $V\alpha$  ont entre 3 et 4 résidus de moins que les domaines  $V\beta$  dans cette région. Le "switch" du brin c'' apparaît essentiel pour amener la boucle CDR2 au niveau de la surface du site de fixation de l'antigène et permettre l'interaction avec l'hélice  $\alpha 2$  (figure A.13). Il a donc sans doute une origine fonctionnelle.

Pour le  $V\beta$  de KB5-C20, la présence d'un résidu supplémentaire et de la séquence LPGA en jonction des brins c'' et d est sans doute à l'origine du "switch" du brin c''. Les différentes séquences de domaines  $V\beta$  murins ont été classées en trois sous-groupes par Kabat *et al.* (1991). KB5-C20 est le seul représentant du sous-groupe III dont la structure est connue et toutes les séquences de ce sous-groupe possèdent cette séquence LPGA entre les brins c'' et d. Il semble donc probable que tous les  $V\beta$  murins du sous-groupe III présentent cette nouvelle topologie.

$\beta$ -strand	c'	c''	d	
	444444444555555556...6.666666			A6 & N15
	2345678901.234567890...1.23456			
A6	SPELIMSIYs.nGDKEDg--...R.FTAQL			<b>V<math>\alpha</math></b>
N15	APQVLLKSSt.DNKREThq-...G.FHATL			
KB5-C20	GPALLISIRsvsDKKEDg-...R.FTIFF			
2C	GLQLLKYYsgdPVVQGvn...G.FEAEF			
1934.4	GPQFLFRASrdKEKGSsr-...G.FEATY			
	44444444455555555556...6.666666			KB5-C20, 2C & 1934.4
	2345678901234567890...1.23456			
	44444444455555555556666666.666666			KB5-C20 & N15
	234567890123456789012234.56789			
		A		
KB5-C20	qLQWLFLLrspgdkevkslpg-aD.YLATR			<b>V<math>\beta</math></b>
N15	eLKFLIQHyekverDKGflps--R.FSVQQ			
A6	gLRLIHYSVGAGITDQGEVFN-.G-YNVSr			
2C	gLRLIHYSYGAGSTEKGDIPD-.G-YKASr			
14.3.d	gLRLIHYSYGAGSTEKGDIPD-.G-YKASr			
	444444444555555555566666.6666666			A6, 2C & 14.3.d
	2345678901234567890122.3456789			
		A		

**Figure A.12** : Aligement de séquences basé sur les structures disponibles, de la région c'-c''-d des domaines V $\alpha$  et V $\beta$ . Les résidus superposables pour tous les V $\alpha$  ou tous les V $\beta$  sont en lettres capitales. Les résidus superposables pour l'ensemble des domaines V $\alpha$  et V $\beta$  sont surlignés en vert. Les lettres bleues indiquent l'appartenance à un brin  $\beta$ . Deux positions sur les V $\alpha$  (L/I46 et F62) et trois sur les V $\beta$  (L43, L/I46 et Y/F65) sont particulièrement conservées en séquence et en 3D et indiquées en gras. La numérotation est celle définie par Kabat *et al.*, 1991. Les résidus du CDR2 (selon Chotia *et al.*, 1988) sont indiqués en rouge. Cet aligement diffère en quelques positions (brin c'' des domaines V $\alpha$  de A6 et N15, premier résidu (position 63/64) du brin d des domaines V $\beta$ ) des alignements basés uniquement sur la séquence (Chotia *et al.*, 1988, Kabat *et al.*, 1991). Les *gaps* sont représentés par des tirets lorsqu'il s'agit de positions occupées dans d'autres séquences, par des points lorsqu'ils ont été introduits pour refléter la superposition structurale.

Il est par contre intéressant de noter que malgré ce *switch*, la conformation du CDR2 $\beta$  est pratiquement inchangée, quelque soit la conformation du brin c''. Il semble donc que la seule contrainte fonctionnelle soit la position du CDR2, pour permettre l'interaction avec l'hélice  $\alpha$ 1 du MHC (figure A.13). La présence de 3 à 4 résidus supplémentaires (par rapport au V $\alpha$ ) autorise une plus grande variabilité du brin c'', tout en gardant la conformation du CDR2.

La résolution de la structure d'un complexe entre un superantigène (Sag) et un TCR (Li *et al.*, 1998) a récemment montré que l'essentiel des interactions entre le Sag et le TCR s'établissait entre les chaînes latérales du Sag et la chaîne principale du domaine V $\beta$ , notamment dans la région du brin  $\beta$  c''. Ainsi, la topologie du domaine V $\beta$  peut avoir des conséquences directes sur l'interaction TCR/Sag. Dans le cas du Sag de l'entérotoxine B, il semble clair que la topologie du sous-groupe III murin n'est pas du tout favorable.

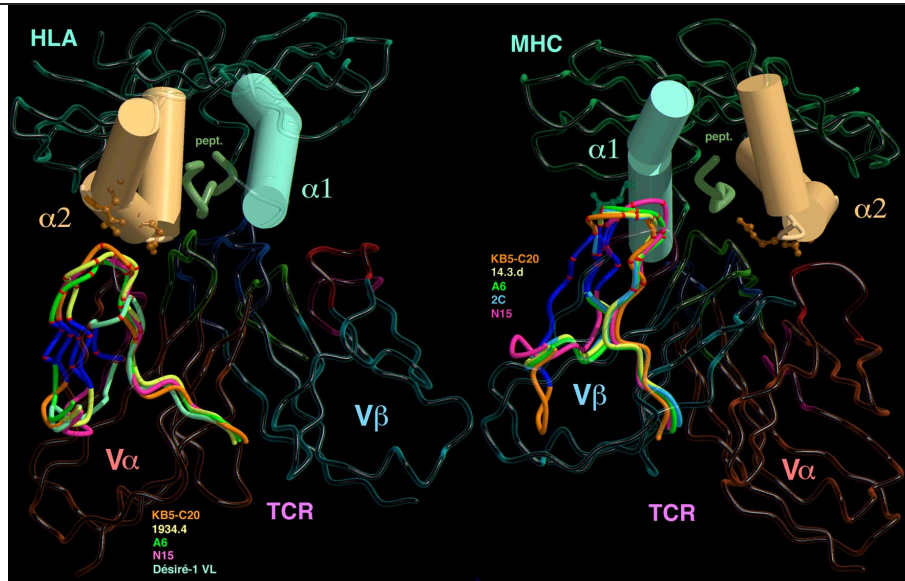


Figure A.13 : Superposition des boucles CDR2 des différents TCR pour les domaines  $V\alpha$  (1934.4, Fields *et al.*, 1995a ; A6, Garboczi *et al.*, 1996 ; N15, Wang *et al.*, 1998 ; 2C, Garcia *et al.*, 1996b) et  $V\beta$  (14.3.d, Bentley *et al.*, 1995), dans le contexte d'un complexe ternaire TCR/pMHC (A6/Tax/HLA-A2 pour le domaine  $V\alpha$ , 2C/dEV8/H-2K<sup>b</sup> pour le  $V\beta$ ). Les CDR2 des domaines  $V_L$  et  $V_H$  du Fab Désiré-1 sont également montrés. La structure des différents CDR2 est remarquablement conservée pour les domaines  $V\alpha$  et  $V\beta$ . La comparaison des CDR2 $\alpha$  avec le CDR2  $V_L$  de Désiré-1 montre que le *switch* du brin  $\beta$  c'' permet au CDR2 d'atteindre le plan du site antigénique et de contacter l'hélice  $\alpha 2$  du MHC.

#### A.5.4 Une boucle CDR3 $\beta$ particulièrement longue

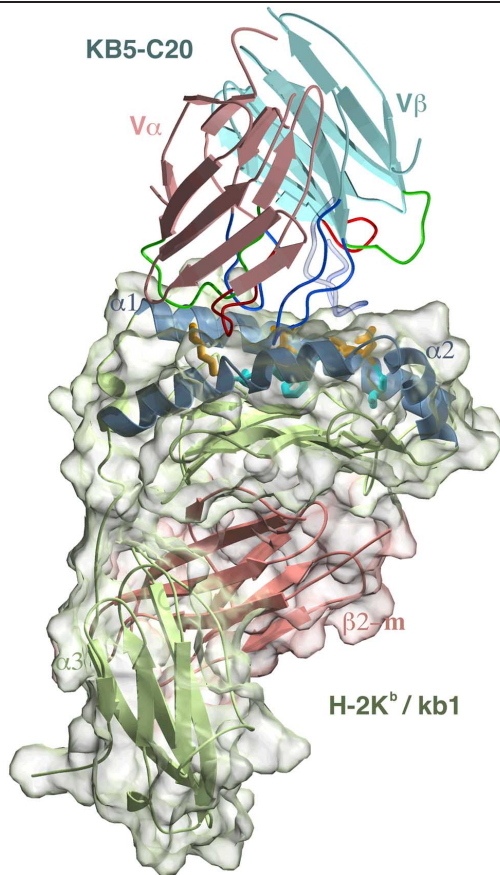
Le TCR KB5-C20 possède une boucle CDR3 $\beta$  de 13 résidus, plus longue que celles des TCR dont la structure est connue (entre 6 et 11 résidus), et parmi les plus longues, pour toutes les séquences connues (Candéias *et al.*, 1991, Pannetier *et al.*, 1993) ; Dans la structure du complexe TCR KB5-C20/Fab Désiré-1, le CDR3 $\beta$  interagit étroitement avec le CDR3 $\alpha$ , par le biais d'une liaison ionique et deux liaisons hydrogène, sans former de poche à l'interface, comme sur les TCR 2C (Garcia *et al.*, 1996) et A6 (Garboczi *et al.*, 1996) (figure A.10). Cette absence de poche peut-être liée à la longueur des CDR3 ou à l'absence du partenaire pMHC. Il sera donc extrêmement intéressant de voir comment un long CDR3 $\beta$  s'adapte pour reconnaître le complexe pMHC.

#### A.5.5 Interaction TCR/peptide/MHC: un modèle du complexe KB5-C20 /kb1/ H-2K<sup>b</sup>

Un premier modèle, relativement rustique, consiste à superposer les structures du scFv TCR (Housset *et al.*, 1997) et de la molécule MHC H-2K<sup>b</sup> en complexe avec le peptide reconnu par KB5-C20 (résultats non-publiés) sur leur homologue dans les structures de complexe ternaire TCR/pMHC déjà connues (figure A.14).

Un tel modèle, construit sur la base du complexe TCR A6/tax/HLA-A2, donne quelques indications sur les interactions entre le domaine  $V\alpha$  et le pMHC. Cette interface est assez bien conservée parmi les trois structures de complexe ternaire déterminées à ce jour, et le modèle permet d'identifier un certain nombre d'interactions conservées entre le  $V\alpha$  et le MHC. De

plus, quelques liaisons hydrogène possibles permettent de prévoir comment la partie N-terminale du peptide est reconnue par ce V $\alpha$ .

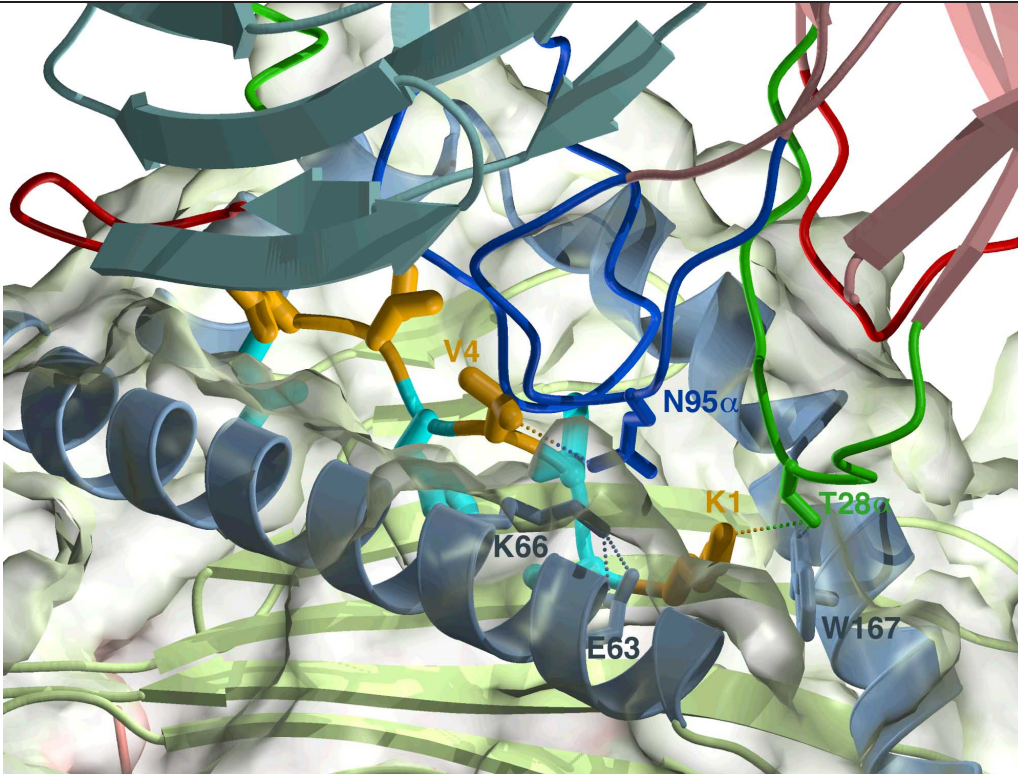


**Figure A.14** : Modèle du complexe TCR KB5-C20 / peptide kb1 / MHC H-2K<sup>b</sup>. Le domaine variable du TCR KB5-C20 a été superposé sur celui du TCR A6, la structure du complexe kb1/H-2K<sup>b</sup> a été superposée sur celle de HLA-A2 (Garboczi *et al.*, 1996). Ce complexe montre que la conformation du CDR3 $\beta$  de KB5-C20 est incompatible avec ce mode général de reconnaissance. Il est donc fort probable que lors de la formation du complexe KB5-C20/pH-2K<sup>b</sup>, la boucle CDR3 $\beta$  s'écarte de la boucle CDR3 $\alpha$ , pour aller interagir avec la partie C-terminale du peptide, peut-être d'une manière analogue à A6, dont la boucle CDR3 $\beta$  est montrée en bleu translucide.

Le modèle permet également d'identifier quelques interactions potentielles avec la partie N-terminale du peptide. Ainsi, la lysine 1 du peptide peut former une liaison hydrogène avec la thréonine 28 du CDR1 $\alpha$  ; la chaîne latérale d'une glutamine du CRD3 $\alpha$  peut former deux liaisons hydrogène avec l'oxygène O $\gamma$ 1 de la thréonine 4 et l'oxygène du groupement carbonyle de la chaîne principale du résidu 2 du peptide (figure A.15). Une interaction similaire est également observée dans le complexe A6/A2, avec une glutamine du CDR1 $\alpha$ . D'autre part, la conformation de la boucle CDR3 du domaine V $\beta$ , telle qu'elle est observée dans la structure du complexe KB5-C20/Désiré-1 est incompatible avec le complexe TCR/peptide/MHC modélisé : une collision entre l'extrémité du CDR3 $\beta$  et la région centrale du peptide semble imposer un changement conformationnel important de cette boucle lors de la reconnaissance du complexe pMHC. Ce changement peut influencer sur l'orientation relative du TCR par rapport au pMHC.

La structure du domaine scFv de KB5-C20 apporte donc de nouvelles informations sur la topologie des domaines variables, et semble indiquer qu'un changement de conformation des CDR3 est très probable lors de l'interaction avec le pMHC, confirmant les observations faites sur un autre TCR (Garcia *et al.*, 1998).





**Figure A.15** : Modèle du complexe TCR KB5-C20 / kb1 / H-2K<sup>b</sup>. Détails des interactions possibles entre le domaine V $\alpha$  et le pMHC.

## A.6 Elucider le lien entre les interactions TCR/pMHC et l'activation de la cellule T

Les structures analysées à l'aune des résultats biologiques et immunologiques apportent déjà une première idée du mode d'interaction entre les TCR d'une part, le MHC et le peptide d'autre part. Cependant, la plasticité du TCR rend difficile toute modélisation, même dans ce contexte restreint de présentation par le MHC. Un exemple est la poche présente à l'interface des boucles CDR3 du TCR 2C, qui semblait idéale pour accueillir une chaîne latérale d'un résidu en position 4 du peptide, et qui reste finalement vide dans le complexe ternaire (Garcia *et al.*, 1996b, 1998). Il va donc sans doute falloir, comme pour les anticorps, un très grand nombre de structures pour permettre d'établir quelques lois générales concernant la reconnaissance du peptide. La suite de notre projet va donc avoir pour objet la détermination de la structure de complexes ternaires KB5-C20/peptide/H-2K<sup>b</sup>, dans le but de mieux cerner les changements de conformation ayant lieu lors de la formation du complexe, et d'aborder de façon systématique la reconnaissance peptidique.

### A.6.1 Bases structurales de la réactivité croisée du TCR

#### A.6.1.1 La réactivité croisée : une propriété essentielle du TCR

Une des caractéristiques essentielles du lymphocyte T est sa capacité à reconnaître plusieurs peptides différents, présentés par la même molécule MHC (Tallquist *et al.*, 1996, Bongrand & Malissen, 1998). Cette propriété, appelée réactivité croisée (*cross-reactivity*), permet au répertoire de lymphocytes T sélectionnés dans le thymus de reconnaître un plus

grand nombre de peptides étrangers. Cette dégénérescence de la spécificité du TCR semble nécessaire pour une protection efficace contre la multitude de pathogènes possibles (Mason, 1998).

En effet, la taille du répertoire en peptides présentables par une molécule de classe I du MHC peut atteindre plusieurs millions. Ce nombre correspond, approximativement, à la possibilité de présenter des peptides de 8, 9 ou 10 acides aminés, tout en conservant 4 résidus ancrés, soit de l'ordre de  $20^6$ , voire plus, si une certaine permissivité se manifeste pour les résidus ancrés. Un individu donné peut être confronté au cours de sa vie à un quelconque de ces peptides antigéniques et doit donc être capable de développer une réponse immunitaire adaptée. Dans le cas où une combinaison  $V\alpha V\beta$  serait dotée d'une spécificité extrême (c'est-à-dire capable de reconnaître avec une affinité suffisante une surface peptide/MHC et une seule) le répertoire en TCR présent dans un individu devrait comporter un nombre de combinaisons  $V\alpha V\beta$  très supérieur à la taille du groupe de peptides présentables (pour être sûr de contenir au moins une structure  $V\alpha V\beta$  complémentaire à la surface peptide/MHC introduite par le pathogène).

Cependant, la taille estimée du répertoire en TCR ne semble pas être très supérieure à la taille du groupe de peptides présentables. En effet, même si en théorie, le nombre de  $TCR\alpha\beta$  possibles grâce à la recombinaison génétique est de l'ordre de  $10^{16}$ , après la sélection intrathymique, cette variété est bien plus faible : on estime le nombre de lymphocytes T différents présents dans le corps à un instant donné à environ  $10^8$ . Ceci a donc conduit à reconsidérer le degré de spécificité atteint par le TCR et à se rendre compte qu'un même lymphocyte T (exprimant une combinaison  $V\alpha V\beta$  et une seule) a en fait la capacité de reconnaître un grand nombre de peptides antigéniques dans le contexte d'une molécule MHC donnée. Ceci est une donnée extrêmement importante sur le plan biologique dans la mesure où elle permet au répertoire de cellules T présent dans un individu donné de reconnaître essentiellement n'importe quel antigène dans une fenêtre de temps compatible avec la vitesse à laquelle un agent infectieux se développe.

#### A.6.1.2 Approche structurale

Il est donc important de connaître au niveau atomique les caractéristiques permettant à un site de reconnaissance  $V\alpha V\beta$  de déployer une telle réactivité croisée. Une étude récente sur les TCR humain A6 et B7, en analysant systématiquement chaque position du peptide, a montré que le peptide Tax était optimisé pour activer les TCR A6 ou B7 (Hausmann *et al.*, 1999). D'autre part, les structures de quatre complexes A6/peptide/HLA-A2, pour lesquels seul le peptide diffère (Tax (cible de A6), et trois peptides variants synthétisés, qui sont de faibles agonistes ou des antagonistes), sont quasiment identiques (Ding *et al.*, 1999)

Notre démarche, quelque peu différente, consiste à déterminer la structure d'un même TCR murin (KB5-C20) complexé à une série de peptides antigéniques enchâssés dans la même molécule de MHC (H-2K<sup>b</sup>). A l'heure actuelle, nous disposons de cinq peptides : trois d'entre eux sont des peptides naturels, reconnus par KB5-C20. Les deux autres sont des peptides artificiels synthétiques, ayant un pouvoir d'activation proche de 100% pour l'un d'eux, de 20% pour l'autre.

Les structures des complexes binaires entre H-2K<sup>b</sup> et trois des différents peptides viennent d'être résolues et un complexe avec un quatrième peptide est en cours. Nous concentrons nos efforts actuels sur la formation du complexe ternaire TCR/pMHC, plus difficile à obtenir, à cause notamment de la faible affinité du TCR KB5-C20 pour H2-K<sup>b</sup> ( $K_D > 5 \cdot 10^{-6}$  M). Aux concentrations en protéines auxquelles nous travaillons (maximum de 4 mg/ml pour H-2K<sup>b</sup>), la concentration des entités séparées est non-négligeable à l'équilibre, et

la molécule pMHC à tendance à cristalliser seule. Un autre TCR, BM3.3, présentant une affinité plus grande pour H-2K<sup>b</sup>, devrait être testé dans les prochaines semaines.

Ces différents complexes ternaires devraient nous offrir l'opportunité de comprendre la base structurale de la réactivité croisée d'un TCR murin, dans le contexte de présentation de l'antigène par la molécule H-2K<sup>b</sup>. En particulier et comme observé dans le cas d'anticorps résultant de réponses primaires, il sera intéressant de déterminer si des remaniements dans la géométrie des CDR rendent compte de cette plasticité de reconnaissance. La résolution de la structure de ces complexes permettra également de développer un programme de mutagenèse rationnelle permettant de déterminer la contribution énergétique respective des résidus MHC et peptide. Il faut cependant garder à l'esprit que de très petits changements dans la nature des interactions TCR/pMHC peuvent avoir de grandes conséquences quant à l'activation du lymphocyte T. Il y a donc tout lieu de croire, comme l'indique les premiers résultats de Ding *et al.* (1999), que les changements structuraux seront mineurs dans la plupart des cas. Une analyse fine de toutes les interactions sera sans doute nécessaire pour tenter d'établir un lien entre données structurales et activation.

## A.7 Perspectives

Si les structures des différents TCR et complexes permettent de mieux cerner les mécanismes de reconnaissance de l'antigène par le récepteur du lymphocyte T, à l'échelle de l'interface TCR/pMHC, peu d'information existe sur la façon dont les différents complexes et co-récepteurs s'oligomérisent à la surface membranaire, pour finalement transmettre le signal d'activation. L'étude de tels ensembles, incluant tous les acteurs de la reconnaissance antigénique, par exemple par cryo-microscopie électronique, devrait permettre d'avoir une image plus globale de cet évènement clef de la réponse cellulaire. Des travaux préliminaires dans cette direction commencent à émerger (Celia *et al.*, 1999) Pour sa part, la cristallographie devrait sans doute nous permettre de connaître dans peu de temps la structure des quelques pièces manquante de notre puzzle : les domaines CD3, les complexes avec les molécules MHC classe II et CD4, ainsi que les TCR $\gamma\delta$ . Elle permettra également une étude systématique de nombreux TCR, permettant de dégager les règles plus générales de la reconnaissance TCR/antigène.

## A.8 Bibliographie

- Alam, S.M., Davies, G.M., Lin, C.M., Zal, T., Nasholds, W., Jameson, S.C., Hogquist, K.A., Gascoigne, N.R.J and Travers, P.J. (1999). Qualitative and quantitative differences in T cell receptor binding of agonist and antagonist ligands. *Immunity*, **10**, 227-237.
- Alam, S.M., Travers, P.J., Wung, J.L., Nasholds, W., Redpath, S., Jameson, S.C. and Gascoigne, N.R.J. (1996). T cell receptor affinity and thymocyte positive selection. *Nature*, **381**, 616-620.
- Albert, F., Buferne, M., Boyer, C. and Schmitt-Verhulst, A.-M. (1982). Interactions between MHC-encoded products and cloned T-cells. I. Fine specificity of induction of proliferation and lysis. *Immunogenetics*, **16**, 533-549.
- Ban, N., Escobar, C., Garcia, R., Hasel, K., Day, J., Greenwood, A. and McPherson, A. (1994). Crystal structure of an idiotype-anti-idiotype Fab complex. *Proc. Natl. Acad. Sci. USA*, **91**, 1604-1608
- Bentley, G.A., Boulot, G, Riottot, M.M. and Poljak, R.J. (1990). Three -dimensional structure of an idiotope-anti-idiotope complex. *Nature*, **348**, 254-257.
- Bentley, G.A., Boulot, G., Karjalainen, K., Mariuzza, R.A. (1995). Crystal structure of the  $\beta$  chain of a T cell antigen receptor. *Science*, **267**, 1984-1987.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F. Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur. J. Biochem.*, **80**, 319-324.
- Bevan, M.J. (1977). In a radiation chimaera host H-2 antigens determine the immune responsiveness of donor cytotoxic Tcell. *Nature*, **269**, 417-418.
- Bjorkman, P.J., Saper, M.A., Samraoui, B., Benett, W.S., Strominger, J.L. and Wiley, D.C. (1987). Structure of the human class I histocompatibility antigen. *Nature*, **329**, 506-512.
- Bongrand, P. and Malissen, B. (1998). Quantitative aspects of T cell recognition : from within the antigen presenting cell to within the T cell. *BioEssays*, **20**, 412-422..
- Bonniface, J.J., Rabinowitz, J.D., Wülfing, C., Hampl, J., Reich, Z., Altman, J.D., Kantor, R.M., Beeson, C., McConnell, H.M. and Davis M.M. (1998). Initiation of signal transduction through the T cell receptor requires the peptide multivalent engagement of MHC ligands. *Immunity*, **9**, 459-466.
- Boon, T., Cerottini, J.C., van den Eynde, B., van der Bruggen, P. and van Pel, A. (1994). Tumor antigens recognized by T lymphocytes. *Annu. Rev. Immunol.*, **12**, 337-365.
- Braden, B.C., Fields, B.A., Ysern, X., Dall'acqua, W., Goldbaum, F.A., Poljak, R.J. and Mariuzza, R.A. (1996). Crystal structure of an Fv-Fv idiotope-anti-idiotope complex at 1.9 Å resolution. *J. Mol. Biol.*, **264**, 137-151.
- Brady, R.L., Dodson, E.J., Dodson, G.G., Lange, G., Davis, S.J., Williams, A.F. and Barclay, A.N. (1993). Crystal structure of domains 3 and 4 of rat CD4: relation to the NH<sub>2</sub>-terminal domains. *Science*, **260**, 979-983.



- Brocke, S., Gaur, A., Piercy, C., Gautam, A., Gijbels, K., Fathman, C.G. and Steinman, L. (1993). Induction of relapsing paralysis in experimental autoimmune encephalomyelitis by bacterial superantigen.. *Nature*, **365**, 642-644.
- Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L. and Wiley, D.C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*, **364**, 33-39.
- Candéias, S., Waltzinger, C., Benoist, C. and Mathis, D. (1991). The V $\beta$ 17+ T cell repertoire : skewed J $\beta$  usage after thymic selection ; dissimilar CDR3s in CD4<sup>+</sup> versus CD8<sup>+</sup> cells. *J. Exp. Med.*, **174**, 989-1000.
- Célia, H., Wilson-Kubalek, E., Milligan, R.A. and Teyton, L. (1999). Structure and function of a membrane-bound murine MHC class I molecule. *Proc. Natl. Acad. Sci. USA*, **96**, 5634-5639.
- Chothia, C., Boswell, D.R., and Lesk, M.A. (1988). The outline structure of the T-cell  $\alpha\beta$  receptor. *EMBO J.*, **7**, 3745-3755.
- Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M., and Poljak, R.J. (1989). Conformations of immunoglobulin hypervariable regions. *Nature*, **342**, 877-883.
- Claverie, J.-M., Prochnika-Chalufour, A. and Bougueleret, L. (1989) Implications for a Fab-like structure for the T-cell receptor. *Immunol. today*, **10**, 10-14.
- Cole, B.C. and Griffiths, M.M. (1993). Rigerring and exacerbation of autoimmune arthritis by the mycoplasma arthiditis superantigen MAM. *Arthritis Rheum.*, **36**, 994-1002.
- Conrad, B., Weidmann, E., Trucco, G., Rudert, W.A., Behboo, R., Ricordi, C., Rodriguez-Rilo, H., Finegold, D. and Trucco, M. (1994). Evidence for superantigen involvement in insulin-dependent diabetes mellitus aetiology. *Nature*, **371**, 351-355.
- Davis, M.M., Boniface, J.J., Reich, Z., Lyons, D., Hampl, J., Arden, B. and Chien Y. (1998). Ligand recognition by alpha beta T cell receptors. *Ann. Rev. Immunol.*, **16**, 523-544.
- Davis, M.M., and Bjorkman, P.J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature*, **334**, 395-401.
- Ding, Y.H., Smith, K.J., Garboczi, D.N., Utz, U., Biddison, W.E. and Wiley, D.C. (1998). Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino-acid. *Immunity*, **8**, 1-20.
- Ding, Y.H., Baker, B.M., Garboczi, D.N., Biddison, W.E. and Wiley, D.C. (1999). Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. *Immunity*, **11**, 45-56.
- Evans, S.V., Rose, D.R., TO, R., Young, N.M. and Bundle, D.R. (1994). Exploring the mimicry of polysaccharide antigens by anti-idiotypic antibodies. The crystallization, molecular replacement, and refinement to 2.8 Å resolution of an idiotope-anti-idiotope Fab. *J. Mol. Biol.* **241**, 691-705.
- Evavold, B.D., Sloan-Lancaster, J., Wilson, K.J., Rothbard, J.B., Allen, P.M. (1995). Specific T cell recognition of minimally homologous peptides:evidence for multiple endogenous ligands. *Immunity*, **2**, 655-663.
- Fields, B.A., Goldbaum, F.A., Ysern, X., Poljak, R.J. and Mariuzza, R. (1995b). Molecular basis of antigen mimicry by an anti-idiotope. *Nature* , **374**, 739-742

- Fields, B.A., Malchiodi, E.M., Li, H., Ysern, X., Stauffacher, C.V., Schlievert, P.M., Karjalainen, K., and Mariuzza, R.A. (1996). Crystal structure of a T-cell receptor  $\beta$ -chain complexed with a superantigen. *Nature*, **384**, 188-192.
- Fields, B.A., Ober, B., Malchiodi, E.L., Lebedeva, M.I., Braden, B.C., Ysern, X., Kim, J.-K., Shao, X., Ward, E.S., Mariuzza, R.A. (1995). Crystal structure of the V $\alpha$  domain of a T cell antigen receptor. *Science*, **270**, 1821-1824.
- Fremont, D.H., Hendrickson, W.A., Marrack, P. and Kappler, J. (1996). Structure of an MHC class II molecule with covalently bound single peptide. *Science*, **272**, 1001-1004.
- Gao GF, Tormo J, Gerth UC, Wyer JR, Mc Michael AJ, Stuart DI, Bell JI, Jones EY, Jakobsen BK. Crystal structure of the complex between CD8 $\alpha\alpha$  and HLA-A2. *Nature* 1997;**387**:630-634.
- Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E. and Wiley, D. (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature*, **384**, 134-141.
- Garcia, K.C., Teyton, L. and Wilson, I.A. (1999). Structural basis of T cell recognition. *Annu. Rev. Immunol.*, **17**, 369-397.
- Garcia, K.C., Degano, M., Pease, L.R., Huang, M., Peterson, P.A., Teyton, L. and Wilson, I.A. (1998). Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science*, **279**, 1166-1172.
- Garcia, K.C., Degano, M., Stanfield, R.L., Brunmark, A., Jackson, M.R., Peterson, P.A., Teyton, L., Wilson, I.A. (1996b). An  $\alpha\beta$  T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science*, **274**, 209-219.
- Garcia, K.C., Scott, C.A., Brunmark, A., Carbone, F.R., Peterson, P.A., Wilson, I.A. and Teyton, L. (1996a). CD8 enhances formation of stable T-cell receptor/MHC class I molecule complexes. *Nature*, **384**, 577-581.
- Grégoire, C., Malissen, B., and Mazza, G. (1996). Characterization of T cell receptor single-chain Fv fragments secreted by myeloma cells. *Eur. J. Immunol.*, **26**, 2410-2416.
- Grégoire, C., Rebai, N., Schweisguth, F., Necker, A., Mazza, G., Auphan N., Millward, A., Schmitt-Verhulst, A.-M., and Malissen, B. (1991). Engineered secreted T-cell receptor  $\alpha\beta$  heterodimers. *Proc. Natl. Acad. Sci. USA*, **88**, 8077-8081.
- Hare, B.J., Wyss, D.F., Osburne, M.S., Kern, P.S., Reinherz, E.L. and Wagner, G. (1999). Structure, specificity and CDR mobility of a class II restricted single-chain T cell receptor. *Nat. Struct. Biol.*, **6**, 574-581.
- Hausmann, S., Biddison, W.E., Smith, K.J., Ding, Y.H., Garboczi, D.N., Utz, U., Wiley, D.C. and Wucherpfennig, K.W. (1999). Peptide recognition by two HLA-A2/Tax11-19-specific T cell clones in relationship to their MHC/peptide/TCR crystal structures. *J. Immunol.*, **160**, 5389-5397.
- Hedrick, S.M., Cohen, D.I., Nielsen, E.A., Davis, M.M. (1984). Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature*, **308**, 149-153.
- Hong, S.C., Chelouche, A., Lin, R., Shaywitz, D., Braunstein, N.S., Glimcher, L. and Janeway, C.A. Jr. (1992). An MHC interaction site maps to the amino-terminal half of the TCR a chain variable domain. *Cell*, **69**, 999-1009.

- Housset, D., Mazza, G., Grégoire, C., Piras, C., Malissen, B., and Fontecilla-Camps, J.C. (1997). The three-dimensional structure of a T-cell antigen receptor V alpha V beta heterodimer reveals a novel arrangement of the V beta domain. *EMBO J.* **16**, 4205-4216
- Hua, C., Boyer, C., Guimezanes, A., Albert, F. and Schmitt-Verhulst, A.-M. (1986). Analysis of T cell activation requirements with the use of alloantigens or an anti-clonotypic monoclonal antibody. *J. Immunol.*, **136**, 1927-1936.
- Hua, C., Buferne, M., and Schmitt-Verhulst, A.-M. (1985). Lysis of hybridoma cells bearing anti-clonotypic surface immunoglobulin by clonotype-expressing alloreactive cytotoxic T cells. *Eur. J. Immunol.*, **15**, 1029-1032.
- Jores, R., Alzari, P.M., and Meo, T. (1990). Resolution of hypervariable regions in T-cell receptor  $\beta$  chains by a modified Wu-Kabat index of amino acid diversity. *Proc. Natl. Acad. Sci. USA*, **87**, 9138-9142.
- Jorgensen, J.L., Esser, U., de StGroth, B.F., Reay, P.A. and Davis, M.M. (1992). Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics. *Nature*, **355**, 224-230.
- Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M., and Gottesman, K.S. (1991). Sequences of proteins of immunological interest. (National Institutes of Health, Bethesda, M.D.).
- Kappler, J.W, Roehm, N. and Marrack, P. (1987). T cell tolerance by clonal elimination in the thymus. *Cell*, **49**, 273-280.
- Kersh GJ, Allen PM. Structural basis for T cell recognition of altered peptide ligands: A single T cell receptor can productively recognize a large continuum of related ligands. *J Exp Med* 1996;**184**:1259-1268.
- Kersh, G.J., Kersh, E.N., Fremont, D.H. and Allen, P.M. (1998). High- and low-potency ligands with similar affinities for the TCR: the importance of kinetics in TCR signaling. *Immunity*, **9**, 817-826.
- Kolanus, W., Romeo, C. and Seed, B. (1993). T cell activation by clustered tyrosine kinases. *Cell*, **74**, 171-183.
- Leahy, D.J., Axel, R. & Hendrickson W.A. (1992). Crystal structure of a soluble form of the human T cell coreceptor CD8 at 2.6 Å resolution. *Cell*, **68**, 1145-1162.
- Li H, Lebedeva MI, Liera AS, Fields BA, Brenner MB, Mariuzza RA. Structure of the V $\delta$  domain of a human  $\gamma\delta$  T-cell antigen receptor. *Nature* 1998;**391**:502-506.
- Li H, Lebedeva MI, Ward ES, Mariuzza RA. Dual conformations of a T cell receptor V $\alpha$  homodimer : implications for variability in V $\alpha$ V $\beta$  domain association. *J Mol Biol* 1997;**269**:385-394.
- Li, H., Llera, A., Tsuchiya, D., Leder, L., Ysern, X., Schlievert, P.M., Karjalainen, K. and Mariuzza, R.A. (1998) Three dimensional structure of the complex between a T cell receptor b chain and the superantigen staphylococcal enterotoxin B. *Immunity*, **9**, 807-816.
- Lo, D., Ron, Y. and Sprent, (1986). Induction of MHC-restricted specificity and tolerance in the thymus. *Immunol. Res.*, **5**, 221-232.
- Luescher, I.F., Vivier E., Layer, A., Mahiou, J., Godeau, F., Malissen, B., and Romero, P. (1996). CD8 modulation of T-cell antigen receptor-ligand interactions on living cytotoxic T lymphocytes. *Nature*, **373**, 353-356.

- Madden, D.R. (1995). The three-dimensional structure of peptide-MHC complexes. *Annu. Rev. Immunol.*, **13**, 587-622.
- Manning, T.C., Schlueter, C.J., Brodnicki, T.C., Parke, E.A., Speir, J.A., Garcia, K.C., Teyton, L., Wilson, I.A., and Kranz, D.M. (1998). Alanine scanning mutagenesis of an ab T cell receptor: Mapping the energy of antigen recognition. *Immunity*, **8**, 413-425.
- Mason, D. (1998). A very high level of crossreactivity is an essential feature of the T cell receptor. *Immunol. Today*, **19**, 395-404.
- Matsui, K., Boniface, J.J., Steffner, P., Reay, P.A. and Davis, M.M. (1994). Kinetics of T cell receptor binding to peptide/I-E<sup>k</sup> complexes: correlation of the dissociation rate with T-cell responsiveness. *Proc. Natl. Acad. Sci. USA*, **91**, 12862-12866.
- Matzinger, P. (1994). Tolerance, danger, and the extended family. *Annu. Rev. Immunol.*, **12**, 991-1045.
- Mazza, G., Housset, D., Piras, C., Grégoire, C., Lin, S.Y., Fontecilla-Camps, J.C. and Malissen, B. (1998). Glimpses at the recognition of peptide/MHC complexes by T-cell antigen receptor. *Immunological Rev.*, **163**, 187-196
- Mazza, G., Housset, D., Piras, C., Grégoire, C., Fontecilla-Camps, J.C. and Malissen, B. (1999) Structural features of the interaction between an anti-clonotypic antibody and its cognate T-cell antigen receptor. *J. Mol. Biol.*, **287**,773-780.
- McKeithan, T.W. (1995). Kinetic proofreading in T cell receptor signal transduction. *Proc. Natl. Acad. Sci. USA*, **92**, 5042-5046.
- Meuer, S.C., Hodgdon, J.C., Hussey, R.E., Protentis, J.P., Schlossman, S.F. and Reinherz, E.L. (1983). Antigen-like effects of monoclonal antibodies directed at receptors on human T cell clones. *J. Exp. Med.* **158**, 988-993.
- Novotny, J., Tonegawa, S., Saito, H., Kranz, D.M. and Eisen, H.N. (1986). Secondary, tertiary and quaternary structure of T-Cell-specific immunoglobulin-like polypeptide chains, *Proc. Natl. Acad. Sci. USA*, **83**, 742-746.
- Padlan, E.A. (1990). On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins Struct. Funct. Genet.*, **7**, 112-124.
- Padlan, E.A. (1994). Anatomy of the antibody molecule *Mol. Immunol.*, **31**, 169-217.
- Pan, Y., Yuhasz S.C. and Amzel, L.M. (1995). Anti-idiotypic antibodies: biological function and structural studies. *FASEB J.*, **9**, 43-49.
- Pannetier, C., Cochet, M., Darce, S., Casrouge, A., Zöller, M., and Kourilsky, P. (1993). The sizes of the CDR3 hypervariable regions of the murine T-cell receptor  $\beta$  chains vary as a function of the recombined germ-line segments. *Proc. Natl. Acad. Sci. USA*, **90**, 4319-4323.
- Plaksin, D., Chacko, S., Navaza, J., Margulies, D.H. and Padlan, E.A. (1999). The X-ray crystal structure of a V $\alpha$ 2.6J $\alpha$ 38 mouse T cell receptor domain at 2.5 Å resolution: alternate modes of dimerization and crystal packing. *J. Mol. Biol.*, **289**, 1153-1161.
- Ryu, S.E., Kwong, P.D., Truneh, A., Porter, T.G., Arthos, J., Rosenberg, M., Dai, X.P., Xuong, N.H., Axel, R., Sweet, R.W., et al (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature*, **348**, 419-426.

- Sant'Angelo, D.B., Waterbury, G., Preston-Hurlburt, P., Tim Yoon, S., Medzhitov, R., Hong, S. and Janeway, C.A. Jr. (1996). The specificity and orientation of a TCR to its peptide-MHC class II ligands. *Immunity*, **4**, 367-376.
- Savage, P.A., Boniface, J.J. and Davis, M.M. (1999). A kinetic basis for T cell receptor repertoire selection during an immune response. *Immunity*, **10**, 485-492.
- Sim BC, Travers PJ, Gascoigne NRJ. V $\alpha$ 3.2 selection in MHC class I mutant mice. Evidence for an alternate orientation of TCR-MHC class I interaction. *The Journal of Immunology* 1997;**159**:3322-3329.
- Sun, R., Shepherd, S.E., Geier, S.S., Thomson, C.T., Sheil, J.M., and Nathenson, S.G. (1995). Evidence that the antigen receptors of cytotoxic T lymphocytes interact with a common recognition pattern on the H-2K<sup>b</sup> molecule. *Immunity*, **3**, 573-582.
- Tallquist, M., Yun, T.J., and Pease, L.R. (1996). A single T cell receptor recognizes structurally distinct MHC/peptide complexes with high specificity. *J. Exp. Med.*, **184**, 1017-1026.
- Wang, J., Lim, K., Smolyar, A., Teng, M., Liu, J., Tse, A.G.D., Liu, J., Hussey, R.E., Chishti, Y., Thomson, C.T., Sweet, R.M., Nathenson, S.G., Chang, H-C., Sacchettini, J.C. and Reinherz, E.L. (1998). Atomic structure of an alphabeta T cell receptor (TCR) heterodimer in complex with an anti-TCR fab fragment derived from a mitogenic antibody. *EMBO J.*, **17**, 10-26.
- Weiss, A. and Littman, D. (1994). Signal transduction by lymphocyte antigen receptors. *Cell*, **76**, 263-274.
- Willcox, B.E., Gao, G.F., Wyer, J.R., Ladbury, J.E., Bell, J.I., Jakobsen, B.K. and van der Merwe, P.A. (1999). TCR binding to peptide-MHC stabilizes a flexible recognition interface. *Immunity*, **10**, 357-365.
- Wucherpfennig KW, Strominger JL. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell* 1995;**80**:695-705.
- Wyer, J.R., Willcox, B.E., Gao, G.F., Gerth, U.C., Davis, S.J., Bell, J.I., Anton von der Merwe, P. and Jakobsen, B.K. (1999) T cell receptor and co-receptor CD8 $\alpha\alpha$  bind peptide-MHC independently and with different kinetics. *Immunity*, **10**, 219-225.
- Ysern, X., Li, H. and Maruizza, R.A. (1998). Imperfect interfaces. *Nat. Struct. Biol.*, **5**, 412-414
- Zinkernagel, R.M., Callahan, G.N., Althage, A., Cooper, S., Klein, P.A. and Klein J. (1978). On the thymus in the differentiation of "H-2 self-recognition" by T-cells: evidence for dual recognition? *J. Exp. Med.*, **147**, 882-896.

*Ouvrages d'immunologie utilisés pour la rédaction de la présentation générale du système immunitaire:*

Atlas of Immunology. J.M. Cruse & R.E. Lewis, Springer (Germany) & CRC press LLC (USA). 1999.

Immuno-biology: the immune system in health and disease. C.A. Janeway & P. Travers. Third edition, Current Biology Ltd. (London, San Francisco, New-York). 1997.



*Les figures ont été réalisées avec les logiciels MOLSCRIPT, RASTER-3D, GRASP:*

Kraulis, P.J. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946-950.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK - A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283-291.

Merritt, E.A. & Murphy, M.E.P. (1994). Raster3D Version 2.0 - A program for photo realistic molecular graphics. *Acta Cryst.* **D50**, 869-873.

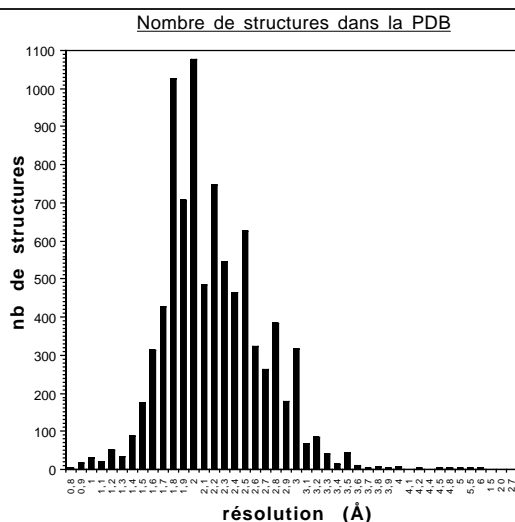
Nicholls, A., Sharp, K.A. and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.*, **11**, 281-296.

## B Les structures de protéines à très haute résolution.

Ce projet est résolument de caractère plus méthodologique. Il y est directement lié aux problèmes d'affinement que se pose le cristallographe, lorsqu'il affine la structure d'une protéine : comment extraire le maximum d'informations fiables des données cristallographiques. Quel modèle atomique utiliser, quelles contraintes appliquer, quels programmes utiliser, quels critères permettent de définir objectivement la convergence d'un affinement, en fonction de la résolution et la précision des données. L'émergence de ce projet vient de l'intérêt pour ces questions concernant l'affinement et l'occasion donnée par des cristaux de toxine de scorpion, de travailler sur des données s'étendant jusqu'à la résolution de 0.96 Å. Que pouvaient apporter de telles données à notre connaissance de la structure ? La collaboration avec l'équipe de C. Lecomte a permis de transférer les méthodes et les bibliothèques de données à très haute résolution acquises sur des peptides, pour l'affinement de la toxine. Cette démarche a permis de déterminer de façon assez convaincante le maximum d'information que l'on pouvait extraire d'un tel jeu de données et d'entrevoir ce que l'on peut gagner avec des résolutions plus élevées. Voici donc en quelque sorte un état de l'art sur la structure des protéines à résolution atomique.

### B.1 Introduction

Des deux méthodes permettant la détermination de structure de macromolécules à l'échelle atomique, la résonance magnétique nucléaire et la cristallographie aux rayons X, la dernière est sans conteste la plus productive, à la fois quantitativement et qualitativement. A l'heure actuelle, plus de 7500 structures de protéines ont été déterminées par cette méthode et déposées dans la *Protein Data Bank* (PDB). Et ce nombre ne cesse de croître rapidement, avec plusieurs structures déposées chaque jour à la PDB. Derrière cette quantité impressionnante d'information, se cache une grande diversité dans la taille des molécules étudiées, la qualité des données de diffraction utilisées et la résolution à laquelle ces structures sont déterminées.



**Figure B.1 :** Nombre de structures de macromolécules dans la *Protein Data Bank*, en fonction de la résolution. Les tranches de résolution sont de 0.1 Å.

La taille des protéines dont la structure est connue varie de quelques dizaines d'acides aminés à plusieurs milliers. La qualité des données cristallographiques obtenues dépend, bien sûr, du cristal mais aussi de la source de rayons X utilisée (générateur à anode tournante,

synchrotron de deuxième ou troisième génération), du détecteur, de la température d'enregistrement, etc... . La résolution est une donnée fondamentale qui va déterminer la quantité d'information contenue dans les données de diffraction, et ce que l'on va pouvoir connaître de la structure. Elle va nous donner une idée de la précision du modèle, et de la validité des paramètres issus de l'affinement.

A l'heure actuelle, la gamme de résolution des structures de protéines présentes dans la PDB va de 27 Å à 0.75 Å avec une moyenne de l'ordre de 2 Å (figure B.1). A cette résolution, le modèle atomique est bien déterminé, avec une précision sur les positions atomiques de l'ordre de 0.15 Å, et un certain nombre de molécules de solvant de structure peuvent être localisées. Il permet déjà de répondre à un grand nombre de questions sur l'architecture de la protéine, sa fonction, les interactions présentes, etc... . Pour beaucoup de cristallographes des macromolécules, cette information est amplement suffisante. Que peut apporter une plus haute résolution ? C'est ce que nous allons essayer de montrer ci-dessous. En effet, nous observons aujourd'hui une convergence entre la cristallographie des petites molécules et celle des macromolécules. Les récents progrès techniques ont permis d'atteindre des résolutions subatomiques pour des protéines comportant plusieurs centaines d'acides aminés et l'on peut désormais, grâce aux ordinateurs actuels, appliquer les méthodes de résolution et d'affinement auparavant réservées aux petites molécules. Des modèles extrêmement précis sont déjà disponibles, et l'enjeu est maintenant de voir comment extraire le maximum d'informations de ces densités électroniques de protéines très précises.

## B.2 Apport d'une résolution atomique ou subatomique.

L'apport de la résolution atomique pour les structures de protéines peut être divisé en deux grandes catégories :

### B.2.1 Les avantages méthodologiques

Il concerne essentiellement les étapes de résolution et d'affinement de la structure. C'est cet aspect qui va techniquement rapprocher de plus en plus le « monde » des macromolécules de celui des petites molécules.

- La possibilité de résoudre la structure *ab initio*, par méthodes directes. La plus populaire pour les macromolécules est la méthode "*Shake-and-Bake*" (Debaerdemacker & Woolfson, 1983, DeTitta *et al.*, 1994), qui a d'abord permis de résoudre des structures de peptides (+ de 100 atomes, Miller *et al.*, 1993), puis des structures de protéines comme la toxine II de scorpion (64 acides aminés, Smith *et al.*, 1997) et plus récemment, la structure du lysozyme (129 acides aminés, Deacon *et al.*, 1998).
- Les méthodes d'affinement et de reconstruction automatique (wARP, Lamzin & Wilson, 1993, Perrakis *et al.*, 1997) sont d'autant plus efficaces que la résolution est élevée. La localisation individuelle des atomes assure en général le succès de ces méthodes.
- La haute résolution permet d'avoir un grand nombre d'observations par rapport au nombre de paramètres (rapport de l'ordre de 5 à 10, voir tableau B.1), ce qui permet une bonne convergence des procédures d'affinement et une meilleure validation statistique. En fonction de la résolution et du degré de désordre, un certain nombre de contraintes stéréochimiques pourront être relâchées.

**Tableau B.1:** Détails identifiables dans la densité électronique en fonction de la résolution. Le nombre de réflexions est donné sur la base d'un cristal contenant 50% de solvant.

Résolution (Å)	nb de réflexions / atomes	nb de paramètres / atomes	Détails "visibles" dans la densité électronique
10	0.1	-	L'enveloppe de la protéine, les domaines
5	0.6	-	Les structures secondaires sont identifiables
3.5	1.7	3	La chaîne polypeptidique et les chaînes latérales volumineuses
3.0	3	3	Toutes les chaînes latérales, affinement possible de la structure
2.4	5	4	L'agitation thermique atomique peut être affinée
1.7	15	4	On commence à "voir" les atomes individuellement
1.2	43	9	Résolution atomique, mouvement anisotrope observable et affinable. Quelques hydrogènes visibles
0.9	100	9-12	Les électrons de valence sont visibles, les atomes d'hydrogène bien définis
0.5	600	9-12	Affinement de la charge atomique,....

## B.2.2 L'intérêt biologique

Même si une structure à 2.0 Å apporte déjà l'essentiel de l'information sur l'architecture de la macromolécule, la très haute résolution permet de clarifier un certain nombre de détails de nature chimique ou structurale, significatifs quant à la fonction biologique :

- La nature des liaisons et des interactions interatomiques peut être identifiée, notamment sur la base d'une estimation précise de la distance interatomique.
- Les atomes d'hydrogène sont visibles dans la densité électronique. Ainsi, l'état de protonation de certains résidus peut être déterminé.
- La stéréochimie peut être déterminée précisément, sans le biais introduit par les contraintes de distance, d'angle, etc., nécessaires à plus basse résolution.
- Les électrons de valence sont observables, la charge atomique peut être estimée dans les cas favorables.

D'une façon plus générale, les apports seront très significatifs dans les domaines où une connaissance très précise de la structure est essentielle (mécanisme réactionnel, états redox, ...). Une estimation précise de la densité électronique devrait ouvrir la voie au calcul du potentiel électrostatique, qui permettra de mieux comprendre les interactions protéine-protéine ou protéine-substrat.

## B.2.3 Les moyens nécessaires

Si l'intérêt pour la très haute résolution est tout juste émergent, c'est que les obstacles techniques à surmonter sont loin d'être négligeables. Un certain nombre de conditions sont primordiales pour pouvoir accéder à la très haute résolution.

- La première des conditions est l'obtention de cristaux de très bonne qualité qui diffractent à résolution atomique. Il n'existe malheureusement pas de « recette » pour obtenir de tels cristaux et il est difficile de prétendre aujourd'hui que toutes les protéines peuvent former de tels cristaux. Toutefois, les progrès faits dans les systèmes de purification et de cristallisation des protéines ont permis d'augmenter les chances d'obtention de tels échantillons.
- Une source de rayons X performante est primordiale pour la réussite de tels enregistrements. Même si les premières structures à très haute résolution ont été réalisées à l'aide de sources classiques (tube scellé, anode tournante) grâce à des cristaux de qualités exceptionnelles, le rayonnement synchrotron est désormais un outil indispensable pour aborder ces études et permet d'obtenir un faisceau fin, intense, peu divergent et de longueur d'onde adaptée.
- Un détecteur bidimensionnel de grande taille (ou comportant un grand nombre de pixels) est aussi indispensable, car le nombre de taches à enregistrer simultanément augmente avec la résolution. Les plaques photosensibles semblent, pour l'instant, les mieux adaptées, mais les nouveaux détecteurs CCD, qui offrent l'avantage d'une lecture très rapide, deviennent très compétitifs.
- La congélation des cristaux est un atout essentiel: en réduisant fortement toute diffusion dans les cristaux, elle limite la dégradation des cristaux due, entre autres, à la diffusion des radicaux libres engendrés par le faisceau de rayons X. Elle prolonge ainsi la durée de vie des cristaux sous le faisceau et permet de réaliser l'enregistrement avec un seul cristal. Les erreurs introduites par le manque d'isomorphisme entre les différents cristaux peuvent être éliminées. Elle présente aussi l'avantage de diminuer l'agitation thermique des atomes (Walsh *et al.*, 1998). Cet avantage est souvent contrebalancé par une augmentation de la mosaïcité, due à la congélation elle-même. L'utilisation de petits cristaux (~100µm), sur lesquels le "flash freezing" est plus efficace, peut paradoxalement être favorable.
- Les progrès incessants des ordinateurs permettent désormais de traiter et d'affiner sans trop de problèmes les structures à très haute résolution. Les temps de calculs nécessaires restent toutefois très importants.
- Les logiciels évoluent également rapidement et permettent des affinements de plus en plus efficaces et rapides. Le nombre croissant de structures de protéines à résolution atomique a rendu nécessaire l'adaptation de programmes jadis développés pour les petites molécules (par exemple, SHELXL-97, Sheldrick & Schneider, 1997), ainsi que la modification de programmes d'affinement de protéines pour la haute résolution (par exemple REFMAC, CCP4, 1994). De nouveaux logiciels, permettant l'utilisation de modèles de densité électronique atomique plus complexes (MOLLY, Hansen & Coppens, 1978, Jelsch *et al.*, en préparation) et des algorithmes plus performants, sont en cours de développement.

#### **B.2.4 Les difficultés majeures**

La difficulté majeure rencontrée lors de l'enregistrement de données de diffraction à très haute résolution réside dans la faiblesse du signal que l'on souhaite enregistrer. Une des raisons fondamentales est que l'intensité diffractée par un cristal est inversement



proportionnelle au volume de la maille élémentaire. Donc, plus la molécule est grande, plus la taille de la maille est grande et plus l'intensité diffractée est faible. C'est sans doute ce qui explique la taille modeste des premières protéines dont la structure a été déterminée à très haute résolution, et réciproquement le besoin actuel de sources puissantes comme les synchrotrons.

Contrairement à un cristal de petite molécule, un cristal de protéine contient une fraction importante de solvant (entre 25 et 80%, avec une moyenne de 50%). La présence de ce solvant pose différents types de problèmes :

- Une fraction significative du volume de la maille est désordonnée (solvant, certains résidus de surface) et donc mal modélisée.
- La présence du solvant permet la diffusion des radicaux libres et limite la stabilité de l'édifice cristallin, ce qui a tendance à se traduire par un moindre pouvoir diffractant au bout d'un certain temps d'exposition aux rayons X.

La mosaïcité des cristaux de protéines doit être réduite au maximum, car elle détériore le rapport signal/bruit et pénalise l'enregistrement des tâches de diffraction à haute résolution (recouvrement des tâches).

L'agitation thermique est aussi un facteur important qui limite la résolution du spectre de diffraction : plus le facteur d'agitation thermique B est élevé, plus l'intensité moyenne  $\langle I \rangle$  sera faible, et ce d'autant plus que la résolution d est élevée :  $\langle I \rangle \propto \exp(-2B / d^2)$ .

## B.3 Etat de l'art

### B.3.1 des pionniers ...

Les premières structures de protéines à résolution atomique sont apparues dans les années 1980 et concernaient des chaînes polypeptidiques relativement courtes: la crambine (Teeter & Hendrickson, 1979), l'insuline (Sakabe *et al.*, 1981), le BPTI (Wlodawer *et al.*, 1984), la rubredoxine (Watenpaugh *et al.*, 1979) furent parmi ces premières structures. Aujourd'hui, le nombre de structures à très haute résolution est encore faible, puisque seulement une cinquantaine de structures à résolution atomique sont déposées dans la PDB (sur plus de 7500). Par contre, leur taille peut aller jusqu'à plus de 500 résidus. Ce faible succès de la haute résolution a sans doute plusieurs causes: un manque de motivation des bio-cristallographes, conforté par le coût important en temps d'enregistrement nécessaire, en nombre de cristaux et en temps de calculs. Le gain apporté par la haute résolution peut paraître mineur par rapport à l'intérêt de la structure elle-même, qui apporte déjà tant d'informations structurales (tableau B.1), et pour laquelle une résolution de 2 Å, voire 3 Å, est amplement suffisante.

Les progrès techniques réalisés depuis le début des années 90, et la plus grande attention attachée à la précision et à la justesse des modèles protéiques (Bränden & Jones, 1990) ont provoqué un regain d'intérêt pour ces études. L'équipe de l'EMBL, à Hambourg, a beaucoup contribué à cette évolution, et plusieurs articles de revue donnent un aperçu des différentes avancées (Dauter *et al.*, 1995, 1997a, Longhi *et al.*, 1998).

### B.3.2 ... aux résultats actuels

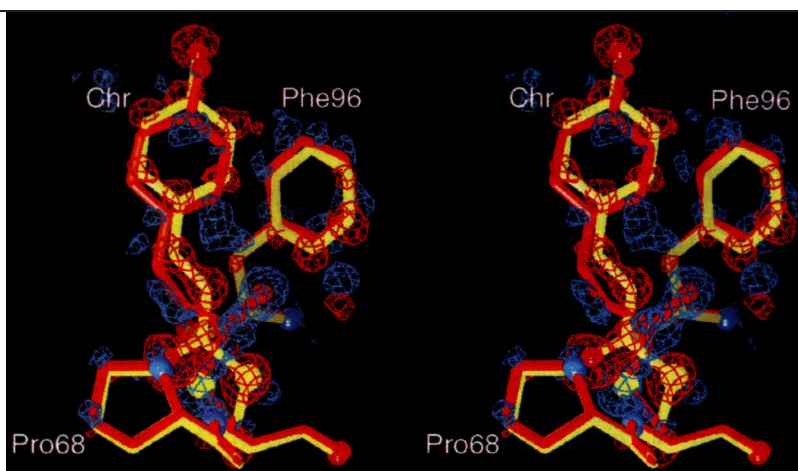
Ces dernières années, un certain nombre de structures ont été affinées à une résolution atomique, et ce jusqu'à très récemment, où deux structures ont été affinées à 0.8 Å de résolution (Genick *et al.*, 1998, Kuhn *et al.*, 1998). Sans reprendre exhaustivement tous ces travaux, voici résumés les apports les plus significatifs de la très haute résolution.

Sur le plan méthodologique, les résultats les plus spectaculaires sont ceux concernant les résolutions par méthode directe "Shake-and-Bake" du lysozyme dans sa forme triclinique, à l'aide de donnée à 0.85 Å de résolution (Deacon *et al.*, 1998) et de la toxine II, à l'aide de données à 0.96 Å de résolution (Smith *et al.*, 1997). Il est important de noter que ces protéines ne contiennent aucun atome lourd (Fe,...), et n'ont donc pas bénéficié de l'aide qu'apporte la présence de tels atomes pour la résolution de la structure.

Sur le plan de la qualité du modèle, les progrès sont multiples et très importants:

Le facteur R, qui indique le degré de désaccord entre les facteurs de structure mesurés et calculés est proche de 10% ou inférieur. L'accord est donc bien meilleur que pour les structures à plus faible résolution (R de l'ordre de 18 - 22%), et la modélisation du mouvement anisotrope des atomes y est pour beaucoup. L'accord est toujours moins bon que pour les petites molécules où le facteur R atteint souvent 2-3 %. Les raisons de cet écart restent encore à préciser, mais l'importante fraction de solvant désordonné et mal modélisé peut être une des sources de ce désaccord.

Ce faible facteur R et le fort rapport entre le nombre de mesures et le nombre de paramètres affinés permettent une très bonne estimation des positions atomiques (erreur typique de 0.02 à 0.03 Å pour la chaîne principale, 0.02 à 0.05 Å pour les chaînes latérales, 0,003 Å pour les atomes plus riches en électrons comme S, Fe, Dauter *et al.*, 1997b). Par comparaison, et pour la chaîne principale, cette erreur est de 0.05 Å à 1.5 Å de résolution et de 0.08 Å à 2.1 Å de résolution (Tickle *et al.*, 1998). Cette précision permet la détermination précise de la stéréochimie de la protéine, indépendamment de contraintes. Par exemple, la géométrie précise d'un cluster Fe-S (Dauter *et al.*, 1997b) ou la distorsion moyenne des liaison peptidique ( $\langle\omega\rangle = 178^\circ$ , Sevcik *et al.*, 1996) ont pu être déterminées.



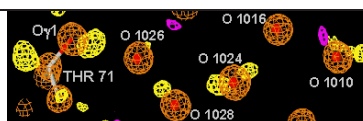
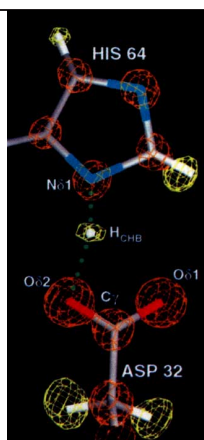
**Figure B.2** : Structure du chromophore de la protéine PYP à 0.85 Å de résolution. En jaune, la conformation non-activée (*dark*), en rouge la conformation photo-activée (*light*). La densité électronique  $F_{dark} - F_{light}$  est contourée à  $+5\sigma$  (bleu) et  $-5\sigma$  (rouge). D'après Genick *et al.*, 1998.

Un grand nombre de conformations alternées peuvent être identifiées et modélisées. Entre 10 et 24 % (cas de la crambine, Teeter *et al.*, 1993) des chaînes latérales peuvent présenter des conformations multiples (Figure B.6 et B.7). De plus, des conformations avec un taux d'occupation assez faible peuvent être construites très précisément. Des données à 0.85 Å de résolution de la protéine PYP, enregistrées immédiatement après un *flash* laser, ont permis d'identifier précisément la conformation activée du chromophore, alors que celle-ci

n'a un taux d'occupation que de 0.25 (Genick *et al.*, 1998). Ainsi, la haute résolution permet de palier l'inefficacité relative du flash laser qui n'active qu'un quart des molécules du cristal, en offrant une densité électronique claire de la liaison thio-ester isomérisée (figure B.2).

La nature chimique des atomes peut être identifiée grâce à la densité électronique (Sevcik *et al.*, 1996). Ceci permet entre autre de déterminer de façon univoque les conformations des résidus comme l'histidine, l'asparagine ou la glutamine.

Une grande partie des atomes d'hydrogène sont "visibles" dans la densité électronique (figure B.3). Ceci permet de déterminer de façon directe l'état de protonation de certains résidus comme l'histidine (Longhi *et al.*, 1997), ou des groupements carboxylate (Kuhn *et al.*, 1998). L'analyse de certaines liaisons hydrogène progresse, comme dans le cas de la subtilisine où pour une liaison hydrogène courte (2.6Å) entre l'aspartate et l'histidine du site actif, l'atome d'hydrogène est localisé à 1.2 Å du N<sup>ε2</sup> de l'histidine et à 1.5 Å du O<sup>δ2</sup> de l'aspartate (Kuhn *et al.*, 1998, figure B.3). Ces informations peuvent être importantes pour la compréhension d'un mécanisme réactionnel.



**Figure B.3 :** Structure de la subtilisine à 0.78 Å de résolution ; Les atomes d'hydrogène sont clairement identifiés par la densité électronique résiduelle  $F_o-F_c$ , contournée à  $+2.5 \sigma$  (jaune). Au-dessus, les atomes hydrogène de molécules d'eau interagissant fortement avec la protéine sont observés ; à gauche, une liaison hydrogène courte (2.6 Å), avec un atome d'hydrogène partagé entre l'histidine et acide aspartique. D'après Kuhn *et al.*, 1998.

L'affinement des facteurs d'agitation thermique anisotropes permet d'obtenir un certain nombre d'informations sur les mouvements corrélés de corps rigides à l'intérieur du cristal (cycles aromatiques, éléments de structures secondaires, etc..., Anderson *et al.*, 1997). Si la prise en compte de l'anisotropie de mouvement améliore sensiblement l'accord entre les  $F_{obs}$  et les  $F_{calc}$ , et traduit un comportement réel des atomes (figure B.7), l'estimation précise de ces facteurs d'agitation pose certains problèmes. En utilisant un modèle sphérique de la densité électronique atomique, le facteur d'agitation thermique anisotrope va prendre en compte tous les aspects non-sphériques de la densité: déplacement atomique, déformation du nuage électronique dû aux liaisons covalentes et aux paires libres, bruit, ...). Pour accéder aux propriétés dynamiques réelles des atomes, il est nécessaire de déconvoluer ces différents termes, et notamment ceux dus aux électrons de valence et aux paires libres, observables à partir de 0.9 Å (Jelsch *et al.*, 1988). Pour ce faire, un modèle plus sophistiqué de la densité électronique atomique doit être employé.

## B.4 Un nouveau modèle de densité électronique atomique

Les résultats précédemment décrits montrent que, sur les structures de protéines à résolution atomique ou subatomique, nous sommes capables d'observer des détails très fins de la densité électronique, correspondant, à peu près, à un électron (atome d'hydrogène, différenciation de type chimique, etc...). L'étape suivante est de passer du qualitatif au quantitatif. Pour ce faire, il faut un modèle de densité électronique atomique qui puisse

prendre en compte la déformation du nuage électronique due aux liaisons covalentes, aux paires d'électrons libres, à la polarisation, etc... . Le modèle sphérique de l'atome neutre, utilisé jusqu'à présent n'est plus suffisant. Un modèle multipolaire de la densité électronique atomique à été développé et appliqué avec succès sur des petites molécules. Son application à des structures de protéines est en cours.

### B.4.1 Un transfert du savoir-faire issu des petites molécules

Il est aujourd'hui possible « d'observer » la distribution des électrons de façon très fine dans les petites molécules (Coppens, 1997; Lecomte 1995 pour revues).

La charge des atomes ainsi que l'écart à la sphéricité de la densité sont observables et quantifiables par l'expérience de diffraction RX à résolution subatomique ( $d=0.5\text{\AA}$ ). La densité électronique non sphérique des atomes (déplacement d'électrons vers les liaisons covalentes et doublets d'électrons libres) est modélisée par des multipôles (Hansen & Coppens, 1978).

$$\rho_{\text{atom}}(r) = \rho_{\text{core}}(r) + P_{\text{val}}\kappa^3\rho_{\text{val}}(\kappa r) + \sum_{l=0, l_{\text{max}}} \kappa'^3 R_l(\kappa' r) \cdot \sum_{m=-l, l} P_{lm} y_{lm}(\theta, \varphi)$$

$\rho_{\text{core}}$  densité électronique sphérique des électrons de cœur

$\rho_{\text{val}}$  densité électronique moyenne des électrons de valence de l'atome libre

$P_{\text{val}}$  population des électrons de valence

$N_{\text{val}}$  population de valence pour l'atome libre

$q = N_{\text{val}} - P_{\text{val}}$  charge atomique

$y_{lm}$  harmoniques sphériques

$R_l$  fonction radiale de type Slater

$P_{lm}$  paramètre de population multipolaire associé à chaque  $y_{lm}$

$\kappa$  et  $\kappa'$  coefficient d'expansion/contraction des termes sphériques et non-sphériques de la densité de valence.

Dans une étude de densité de charge de petite molécule à résolution subatomique ( $d=0.5\text{\AA}$ ), le nombre de données de diffraction RX est suffisant pour que l'on puisse affiner la valeur de la charge atomique ( $P_{\text{val}}$ ), le coefficient de contraction/expansion de la densité de valence ( $\kappa, \kappa'$ ) ainsi que les coefficients des multipôles ( $P_{lm}$ ) dans une procédure de moindres carrés. On peut, par exemple, accéder ainsi à la population des orbitales d et à l'état fondamental des matériaux engagés dans les liaisons de coordination.

Des études de densité de charge ont été réalisées par l'équipe de C. Lecomte, au LCM3B de l'Université de Nancy, sur de nombreux cristaux de peptides et d'acides aminés afin de construire une banque de charges atomiques et de paramètres multipolaires pour toutes les fonctions chimiques présentes dans les protéines (Pichon-Pesme *et al.*, 1995).

La transférabilité des paramètres multipolaires ouvre la possibilité d'appliquer le modèle de densité électronique non sphérique aux protéines à haute résolution sans ajustement des paramètres concernés (Jelsch *et al.*, 1998). Un nouveau logiciel a été développé à partir de MOLLY avec l'introduction de restrictions de distances et d'angles dans l'affinement par moindres carrés.

### B.4.2 Quelques exemples d'application.

Si l'application de ce formalisme multipolaire et de la transférabilité des paramètres  $P_{lm}$  à un petit peptide a montré l'efficacité de la méthode (Pichon-Pesme *et al.*, 1995), son application à une protéine restait à faire. Au jour d'aujourd'hui, trois études sont achevées ou

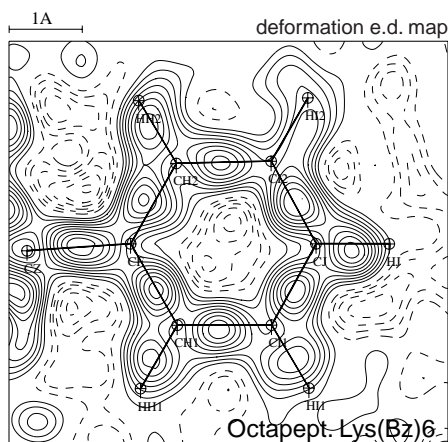
en voie d'achèvement : la structure d'un octapeptide à 0.82 Å de résolution (Jelsch *et al.*, 1998), la structure de la toxine II à 0.96 Å de résolution (Housset *et al.*, 2000) et la crambine à 0.54 Å de résolution (Jelsch *et al.*, 1999). Elles permettent de mieux cerner le champ d'application de cette nouvelle approche (taille du système, facteurs de température, résolution). L'étude de la toxine II, dans laquelle je suis impliqué, sera plus particulièrement détaillée.

#### B.4.2.1 Un octapeptide à 0.82 Å de résolution.

Les paramètres multipolaires ont été transférés à un octapeptide de structure hélicoïdale (Jelsch *et al.*, 1998). Les données de diffraction RX avaient été mesurées par l'équipe de C. Lecomte (Université de Nancy) sans précaution particulière à 0.82 Å de résolution et à une température de 100K. Lors de l'affinement cristallographique avec MOLLY, le transfert des multipôles conduit à une diminution significative des facteurs R résiduels: le facteur R passe de 7.13 % à 6.51 %. L'accumulation des électrons sur les liaisons covalentes et les paires libres de l'oxygène sont clairement visibles, à leur niveau attendu, dans les cartes de déformation de la densité électronique (figure B.4):

$$\Delta\rho_{\text{def}}(\mathbf{r}) = V^{-1} \sum_{\mathbf{h}} [ (k^{-1} |F_{\text{obs}}(\mathbf{h})| \exp(i\varphi_{\text{mult}}) - |F_{\text{sph}}(\mathbf{h})| \exp(i\varphi_{\text{sph}}) ] \exp(-2i\pi\mathbf{h}\cdot\mathbf{r})$$

Cette densité de déformation se définit donc comme la différence entre la densité électronique finale, après affinement multipolaire  $\{ F_{\text{obs}}, \varphi_{\text{mult}} \}$  et la densité électronique calculée avec le modèle sphérique  $\{ F_{\text{sph}}, \varphi_{\text{sph}} \}$ , et représente la partie non-sphérique de la densité électronique.



**Figure B.4** : Octapeptide à 0.82 Å de résolution. Densité de déformation du cycle phényle du résidu Lys(Bz)<sup>6</sup>, après affinement avec MOLLY. Les contours correspondent à  $\pm 0,05 \text{ e}^-/\text{\AA}^3$ . Les contours de densités positives sont représentés par des lignes continues, les contours de densités négatives par des lignes pointillées. La figure montre bien la présence de pics positifs au milieu de chaque liaison du cycle aromatique, indiquant la présence des électrons de valence. La densité de déformation est également bien visible sur les liaisons C-H du cycle. D'après Jelsch *et al.*, (1998).

L'affinement des charges pour neuf types d'atomes différents conduit à des valeurs réalistes. Ainsi dans la liaison peptidique, les charges sont de -0.44, +0.20, +0.42 et -0.65 (e-) pour les atomes N, H, C et O respectivement.

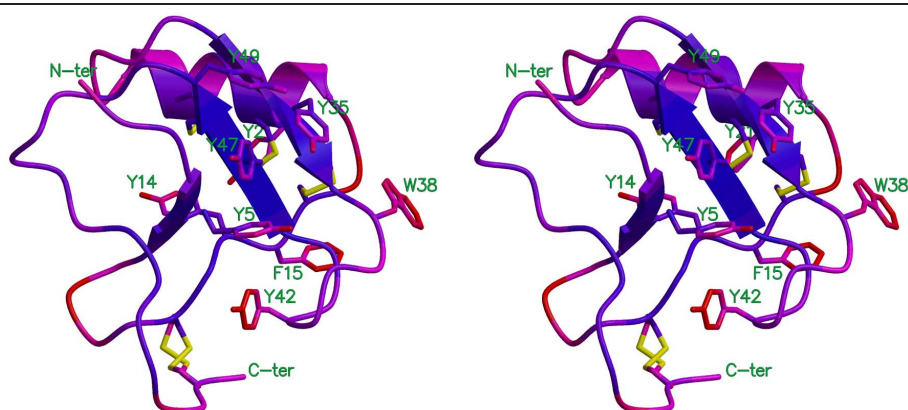
L'estimation du facteur B anisotrope est aussi plus précise, comme cela est illustré par un meilleur score au "rigid bond test" (Hirschfeld, 1976).

Il s'avère donc que la densité électronique de déformation est bien définie dans les cartes calculées par synthèse de Fourier lorsque les données de diffraction sont disponibles à une résolution meilleure que 0.9 Å et que l'agitation thermique des atomes est faible (pour un structure de protéine, c'est à dire  $\sim 2\text{\AA}^2$ ).



### B.4.2.2 Une protéine de 64 acides aminés (toxine II du scorpion *Androctonus australis* Hector) à 0.96 Å de résolution.

La toxine II est une petite protéine de 64 acides aminés extraite du venin du scorpion *Androctonus australis* Hector (figure B.5). Outre ses propriétés toxiques, elle forme facilement des cristaux de très bonne qualité, par évaporation lente de la solution protéique (Miranda *et al.*, 1970, Fontecilla-Camps *et al.*, 1988). Il s'agissait donc d'un bon candidat pour des études à très haute résolution. La structure a été, dans un premier temps, résolue et affinée à 1.8 Å de résolution (Fontecilla-Camps *et al.*, 1988), puis affinée à 1.3 Å de résolution (Housset *et al.*, 1994). Par la suite, des données de diffraction ont été mesurées sur le synchrotron de Hambourg à température ambiante et jusqu'à 0.96 Å de résolution. Ces données, dont les statistiques sont résumées dans le tableau B.2, ont servi de base à différents type d'affinements. Dans un premier temps, la structure a été affinée avec SHELXL-93 et SHELXL-97, puis, après transfert des paramètres multipolaires, l'affinement cristallographique a été poursuivi avec le logiciel MOLLY. Les résultats obtenus ont fait l'objet d'un article soumis à Acta D, dont en voici les grandes lignes.



**Figure B.5 :** Vue stéréographique de la structure de la toxine II. La couleur varie en fonction de l'agitation thermique (de bleu pour les valeurs inférieures à 4 Å<sup>2</sup> à rouge pour les valeurs supérieures à 15 Å<sup>2</sup>). Les résidus aromatiques représentés sont ceux utilisés pour l'analyse TLS (cf. tableau B.4). Les ponts disulfure sont représentés en jaune. La région du feuillet β (résidus 34-37 et 45-49) est la moins agitée, bien stabilisée par deux ponts disulfure.

**Table B.2:** Statistiques des données cristallographiques. Les valeurs pour les dernières coquilles de résolution (1.2-1.11 Å pour le premier cristal, 1.8-1.75 Å et 0.98-0.96 Å pour le second) sont données entre parenthèses.

	résolution (Å)	$R_{\text{merge}}$	$\langle I/\sigma(I) \rangle$	Nb de mesures	nb de réflexions	complétude (%)
Premier cristal (X11, $\lambda=0.92$ Å)	3.5-1.11	0.052 (0.133)	8.3 (4.9)	72171	20089	91 (81)
Second cristal (X31, $\lambda=0.72$ Å)	2.8-0.96 15.8-1.75	0.102 (0.543) 0.091 (0.224)	6.0 (1.3) 4.5 (3.0)	96456 18174	30451 5774	89 (80) 97 (87)
Données rassemblées	10.0 - 0.96	0.073	-	-	32890	90.9 (75.6)

### B.4.2.2.1 Affinement avec SHELXL-97

L'affinement avec SHELXL a été réalisé en trois étapes, afin de bien mettre en évidence la validité du modèle et de l'information interprétée à partir de la densité électronique :

- Etape 1 : Un affinement classique, avec pour paramètres atomiques les coordonnées xyz et un facteur d'agitation thermique isotrope.
- Etape 2 : Même affinement, avec en plus tous les atomes d'hydrogène pris explicitement en compte.
- Etape 3 : Affinement d'un facteur d'agitation thermique anisotrope individuel. Utilisation de l'affinement du type *full matrix*, permettant le calcul de l'incertitude sur l'estimation des différents paramètres.

Les statistiques de l'affinement sont résumées dans le tableau B.3.

**Table B.3** Statistiques d'affinement avec SHELXL-97.

	Etape 1	Etape 2	Etape 3
nb d'atomes de protéine	513	513 + 446 H	560 + 483 H
nb d'atomes de solvant	85	85	121
nb de réflexions utilisées	32847	32827	32827
d (Å)	16.0 - 0.96	16.0 - 0.96	16.0 - 0.96
nb de paramètres	2400	2438	5730
nb de contraintes	2164	2174	2588
<b>facteur-R (F)</b>			
Toutes les données	0.188	0.181	0.118
libre	0.205	0.197	0.136
1.00-0.96 Å	0.319	0.317	0.263
$F > 4\sigma(F)$	0.170	0.163	0.101
libre	0.185	0.178	0.118
<b>facteur-wR (I)</b>			
Toutes les données	0.475	0.465	0.302
Carte résiduelle (Fo-Fc) (e/Å <sup>3</sup> )			
min	-0.52	-0.51	-0.33
max	0.60	0.59	0.34
σ	0.11	0.10	0.07
<b>Stéreo-chimie</b>			
Ecart type par rapport aux valeurs standards :			
distances de liaisons (Å) (σ = 0.013 Å)	0.008	0.009	0.011
distances d'angles(Å) (σ = 0.025 Å)	0.024	0.024	0.026
Contraintes sur les facteurs B anisotropes (Å <sup>2</sup> ) (σ <sub>B</sub> = 0.08 Å <sup>2</sup> )	-	-	0.032

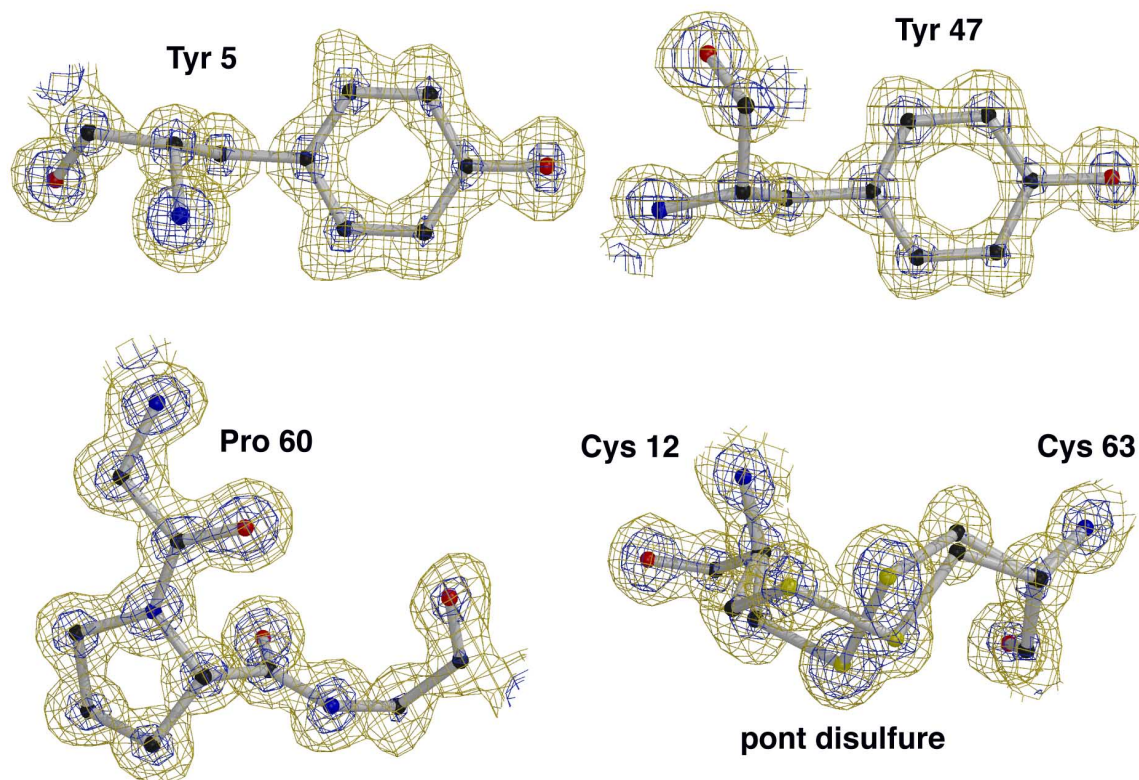
Les facteurs d'agitation thermiques sont donnés avec le même facteur ( $B_{ij} = 8\pi^2 U_{ij}$ ) que les facteurs B, pour faciliter la comparaison avec les autres structures de protéine..

$$R(F) = \frac{\sum |F_{obs} - F_{calc}|}{\sum |F_{obs}|}, wR(I) = \left\{ \frac{\sum [w(F_{obs}^2 - F_{calc}^2)^2]}{\sum [w(F_{obs}^2)]} \right\}^{1/2},$$

$$wR(F) = \left\{ \frac{\sum w(|F_{obs}| - |F_{calc}|)^2}{\sum w(F_{obs}^2)} \right\}^{1/2}, w = 1/\sigma^2(F_{obs}).$$

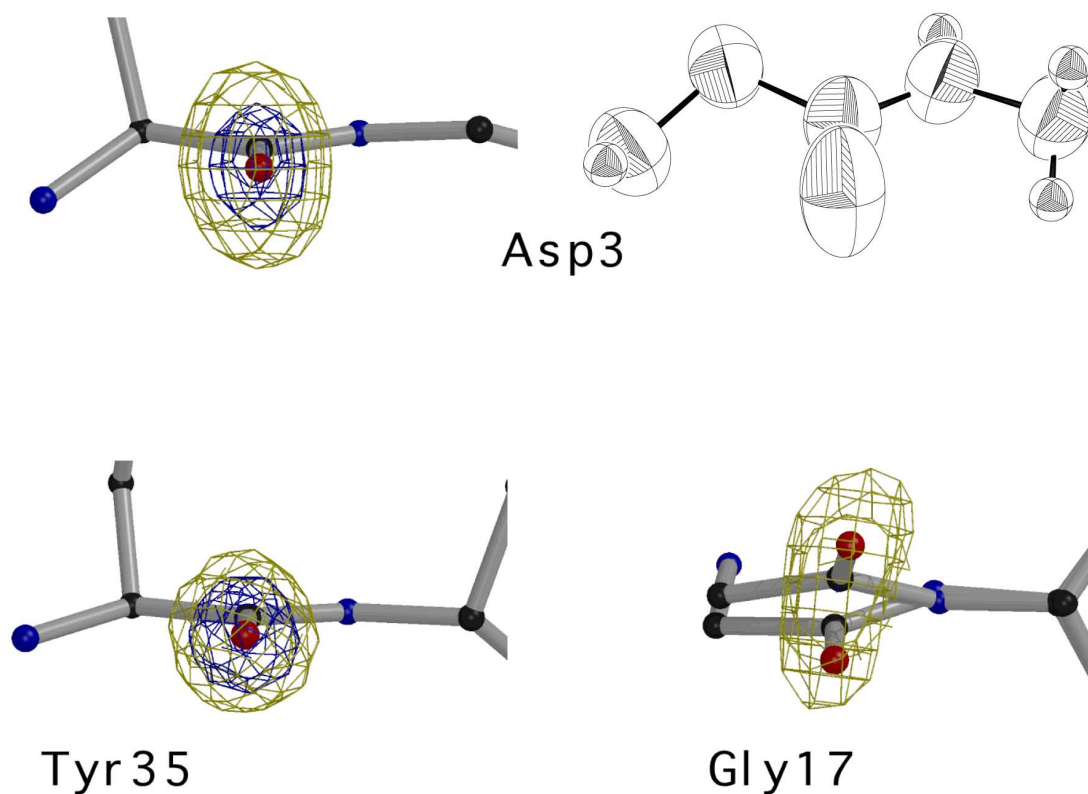
Les enseignements que l'on peut tirer de ces trois étapes sont multiples. A 0.96 Å de résolution, la majorité des atomes a une densité électronique très bien définie (figure B.6) et

les contraintes stéréochimiques ne sont nécessaires que pour quelques chaînes latérales de surface (environ une vingtaine d'atomes sur 513). Les atomes d'hydrogène sont, pour la moitié d'entre eux, visibles dans la densité électronique, et l'amélioration du facteur R, lorsqu'ils sont pris en compte dans le calcul des facteurs de structure, est significative (tableau B.3).



**Figure B.6 :** Toxine II à 0,96 Å de résolution. Vue de la chaîne latérale des tyrosines 5 et 47, de la proline 60 et du pont disulfure Cys12-Cys63 présentant une double conformation, après affinement avec SHELXL-97. Les taux d'occupation des deux conformations de la cysteine 12-63 sont de 0.59 et 0.41. Cette double conformation est également observée dans toutes les structures de cette toxine précédemment affinées à 1,8 Å et à 1,3 Å de résolution. Elle traduit une flexibilité de la partie C-terminale de la protéine. La densité électronique  $\{2mF_{\text{obs}} - DF_{\text{calc}}, \varphi_{\text{calc}}\}$  (SIGMAA, Read, 1986) est contournée à 1.2  $\sigma$  (en jaune) et 4 $\sigma$  (en bleu). 1  $\sigma = 0.07 \text{ e}^-/\text{\AA}^3$  (Housset *et al.*, 2000).

L'affinement du facteur B anisotrope est tout à fait justifié, comme le montre la forte diminution du facteur R et du facteur R-libre (tableau B.3). Ces facteurs B anisotrope reflètent assez bien le comportement dynamique des atomes, comme le traduit l'analyse TLS de certaines chaînes latérales (tableau B.4) ou la densité électroniques de certains groupements carbonyles de la liaison peptidique (figure B.7). L'incertitude sur l'estimation des paramètres, calculées par la méthode *full matrix* est très bien corrélée avec la qualité de la densité électronique. Pour les régions de faible agitation ( $5\text{-}7 \text{ \AA}^2$ ), l'incertitude de position est de l'ordre 0.015 Å (0.005 Å pour les atomes de soufre) ; l'incertitude sur les facteurs d'agitation thermiques est de l'ordre de 0.2 Å<sup>2</sup> (B) et 0.4 Å<sup>2</sup> (B<sub>ij</sub>).



**Figure B.7** : Mise en évidence du mouvement anisotrope de l'atome d'oxygène de la liaison peptidique. Un déplacement perpendiculaire au plan peptidique est observé dans la majorité des cas. Pour l'Asp3, le déplacement est bien décrit par le facteur B anisotrope dont l'ellipsoïde est montré (en haut à droite). Pour la Tyr 35, l'agitation est très faible et la densité quasi sphérique. Pour la Gly17, l'amplitude du mouvement est telle qu'il a fallu modéliser deux conformations. La densité électronique  $\{2mF_{\text{obs}} - DF_{\text{calc}}, \varphi_{\text{calc}}\}$  est contourée à  $1.2 \sigma$  (en jaune) et  $4\sigma$  (en bleu).  $1 \sigma = 0.07 \text{ e}^-/\text{\AA}^3$  (Housset *et al.*, 2000)

**Table B.4:** Analyse des facteurs d'agitation thermique après affinement avec SHELXL-97 (étape 3) en terme de mouvement de corps rigides, pour les cycles aromatiques. Le facteur  $R_2$  exprime le désaccord entre les termes du tenseur d'agitation thermique  $U_{ij}$  estimés en fin d'affinement et les  $U_{ij}^{TLS}$  correspondant à un parfait mouvement de corps rigide (décrit par les tenseurs  $T$ (translation),  $L$ (rotation),  $S$ (mouvement corrélé de translation et rotation) (Shomaker & Trueblood, 1968). Les  $U_{ij}^{TLS}$  sont obtenus par minimisation de la somme  $\sum_{ij} (U_{ij} - U_{ij}^{TLS})^2$  (programme THMA11, Trueblood, 1990). Pour chacun des résidus, le corps rigide inclut l'atome  $C\beta$  et tous les atomes du cycle aromatique.

Résidu	Sans contrainte sur les $U_{ij}$			Avec contraintes sur les $U_{ij}$		
	$R_2$ (%)	$\langle B \rangle$ ( $\text{\AA}^2$ )	$\langle \Delta Z_B \rangle$ ( $\text{\AA}^2$ )	$R_2$ (%)	$\langle B \rangle$ ( $\text{\AA}^2$ )	$\langle \Delta Z_B \rangle$ ( $\text{\AA}^2$ )
Tyr 5	7.1	8.90	0.95	5.9	8.87	< 0.08
Tyr 14	8.6	9.88	1.03	7.8	9.87	< 0.08
Phe 15	12.2	13.54	3.71	10.0	13.58	< 0.08
Tyr 21	11.1	10.65	2.37	8.0	10.63	< 0.08
Tyr 35	12.2	8.24	1.89	8.0	8.21	< 0.08
Trp 38	17.7	12.38	2.61	14.5	12.38	< 0.08
Tyr 42	14.2	13.90	2.92	12.1	13.77	< 0.08
Tyr 47	8.3	7.76	0.55	7.8	7.73	< 0.08
Tyr 49	7.4	7.35	1.18	6.5	7.35	< 0.08

Critère de la liaison rigide d'Hirschfeld:  $\Delta Z_B = 8\pi^2 * \Delta Z_U < 0.08 \text{\AA}^2$ .

$$R_2 = (\sum_{ij} |U_{ij} - U_{ij}^{TLS}|^2 / \sum_{ij} |U_{ij}|^2)^{1/2}$$

#### B.4.2.2.2 Affinement avec MOLLY

L'affinement a été poursuivi avec MOLLY, en trois étapes :

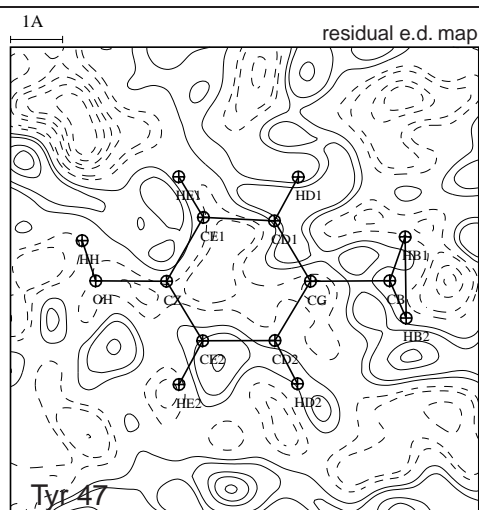
- L'étape 4 est un affinement équivalent à l'étape 3 (modèle sphérique, B anisotrope), dans le but de servir de référence et de bien identifier les différences dues au programme (schéma de pondération, affinement avec les F, contraintes légèrement différentes).
- L'étape 5 est un affinement utilisant le développement multipolaire de la densité électronique atomique.
- L'étape 6 est un affinement de charges des atomes de la liaison peptidique. Considérant que l'étape 5 apporte la meilleure estimation possible des positions atomiques et facteurs d'agitation thermiques, ces paramètres sont fixés à leur valeur obtenue à la fin de cette étape. Les paramètres  $P_{val}$  et  $\kappa$  sont affinés pour chacun des atomes de la liaison peptidique tout en contraignant la distribution de charges à être identique pour les 63 liaisons peptidiques de la toxine. Pour cet affinement, la densité électronique est décrite par les seuls termes sphériques du développement multipolaire.



Table B.5 : Statistiques de l'affinement avec MOLLY.

	Étape 4	Étape 5	Étape 6
Nb de réflexions	24269	24269	24269
$I > 3 \sigma(I)$			
d (Å)	3.3 – 0.96	3.3 – 0.96	3.3 – 0.96
nb de paramètres	4330	4330	15
nb de contraintes	1872	1872	0
Facteur R (F)	0.090	0.091	0.096
Facteur wR(F)	0.093	0.095	0.102
Facteur R libre (F)	0.107	0.108	0.114

L'introduction du modèle multipolaire n'apporte pas d'amélioration significative du facteur R (table B.5). Il semble donc qu'à la résolution de 0.96 Å, et compte tenu de l'agitation thermique de 8 Å<sup>2</sup>, la déformation de la densité de valence contribue trop peu pour observer un effet sur le facteur R. Ceci est d'autant plus vrai que la puissance de l'affinement par moindres carrés va permettre, dans le cas du modèle sphérique, au facteur B anisotrope de prendre en compte une large part de ces électrons de valence. L'efficacité de l'affinement des étapes 3 et 4 se traduit par une carte résiduelle « plate » (figure B.8)

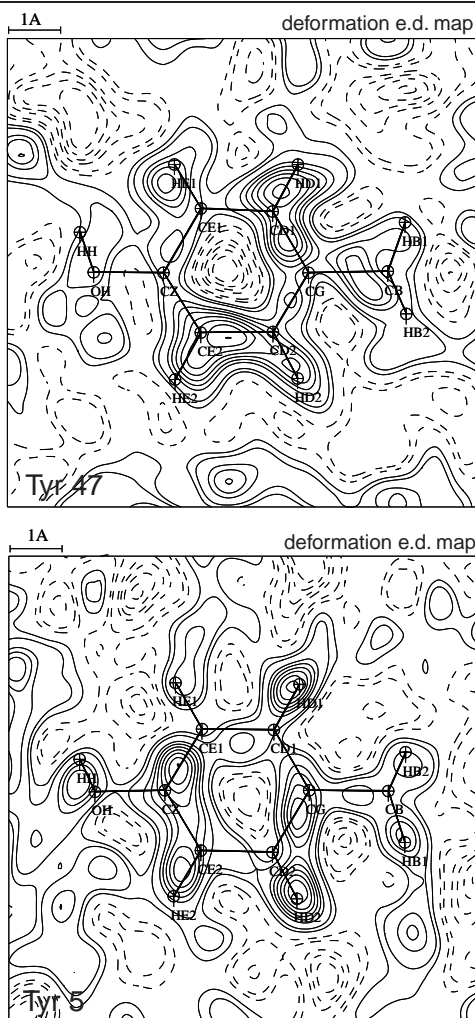


**Figure B.8 :** Toxine II à 0,96 Å de résolution. Carte résiduelle ( $F_o - F_c$ ) de la tyrosine 47, après l'étape 4. Aucun pic significatif n'apparaît sur les liaisons du cycle aromatique. La qualité des phases issues du modèle de densité électronique atomique sphérique et la surestimation des facteurs B anisotropes lors de l'affinement ne permettent pas de mettre en évidence la déformation de la densité de valence.

L'estimation des facteurs d'agitation thermique ne montre pas d'amélioration majeure. Après l'étape 5, le "rigid bond test" n'est toujours pas satisfait et ne montre qu'une très faible amélioration. L'analyse TLS ne montre pas non plus d'apport significatif du modèle multipolaire pour l'estimation des facteurs B. Cela peut s'expliquer par le fait que, pour la toxine, l'incertitude sur les  $U_{ij}$  est de l'ordre de 10% (0.4 Å<sup>2</sup>), valeur similaire au degré de désaccord entre les  $U_{ij}$  et  $U_{ij}^{TLS}$ . Ainsi, la différence entre les deux types d'affinement est sans doute « noyée » dans le bruit.

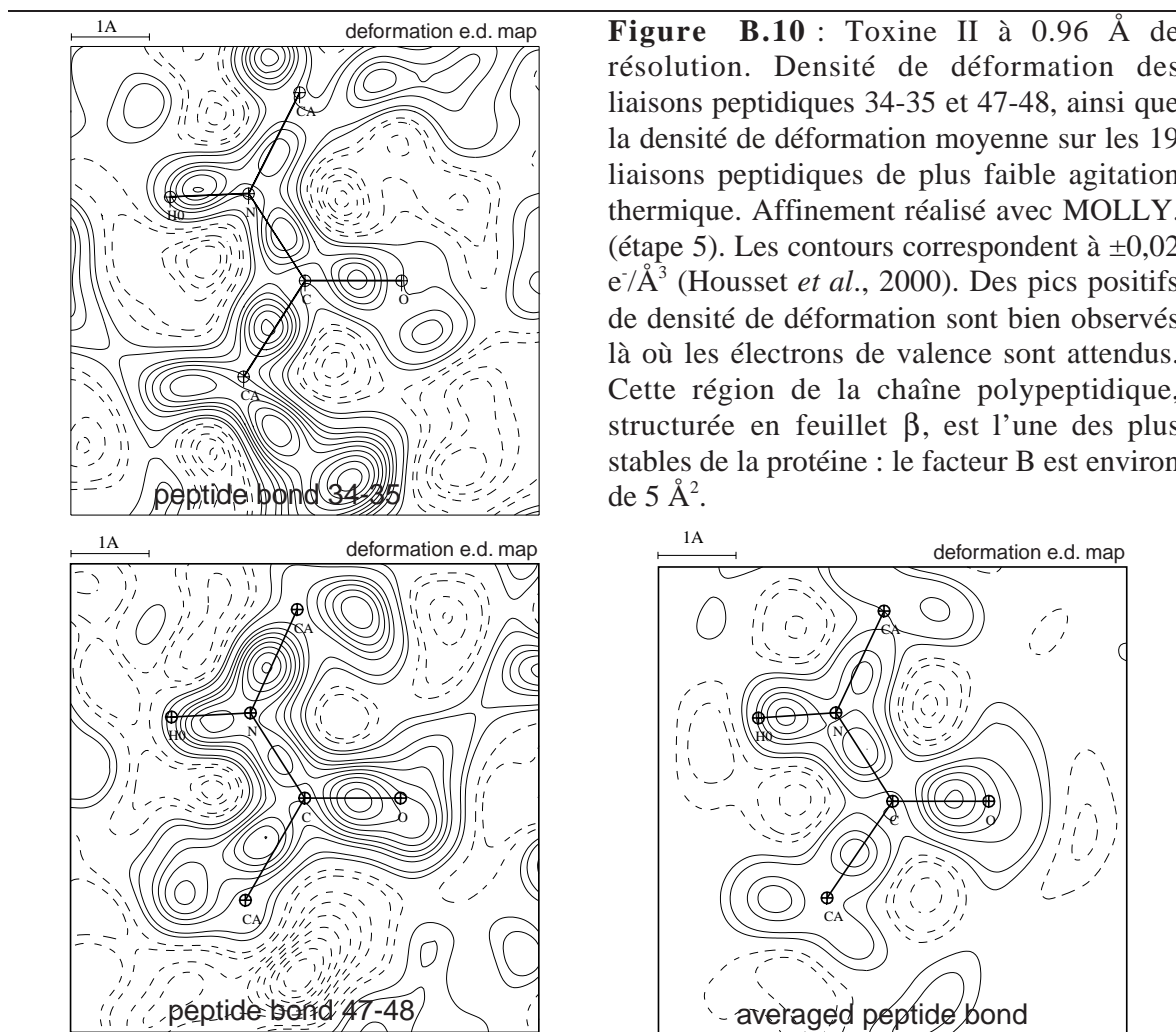
Ainsi, au regard des critères statistiques globaux, on peut donc conclure que ce modèle n'apporte pas d'amélioration tangible. Ceci n'est pas essentiellement dû à la limite de résolution, mais sans doute plus à l'agitation thermique moyenne, 8 Å<sup>2</sup>, faible pour une protéine, mais élevée pour les détails que l'on souhaite observer. Cependant, il est important de noter que dans la structure, le facteur d'agitation varie beaucoup selon les régions. On peut donc tempérer ce constat global décevant en analysant plus en détail le comportement des résidus les mieux ordonnés.

Pour les deux cycles tyrosine les mieux ordonnés (facteur B  $\sim 6 \text{ \AA}^2$ ), la carte de densité électronique de déformation, calculée à la fin de l'étape 5, montre clairement la présence d'électrons de valence (figure B.9).



**Figure B.9 :** Toxine II à 0,96 Å de résolution. Densité de déformation des deux tyrosines les mieux ordonnées: Tyr5 et Tyr47, après affinement avec MOLLY. Les contours correspondent à  $\pm 0,02 \text{ e}^-/\text{\AA}^3$  (Housset *et al.*, 2000). La densité de déformation montre un pic positif sur un certain nombre de liaisons des cycles aromatiques et sur toutes les liaisons C-H. L'hydrogène du groupement hydroxyle de la tyrosine 5 est visible mais peut avoir plusieurs positions. Les cartes de densité sont moins "claires" que pour l'octapeptide mais permettent tout de même d'observer un certain nombre d'électrons de valence, malgré l'agitation thermique plus élevée.

Ce qui est également vrai pour les liaisons peptidiques les plus stables (facteurs B de l'ordre de  $5 \text{ \AA}^2$ , figure B.10). Il apparaît donc que localement, pour les régions les plus ordonnées et les mieux définies dans la densité électronique, ce modèle permet une meilleure analyse de la densité électronique (les électrons de valence des liaisons conjuguées sont observables).



**Figure B.10** : Toxine II à 0.96 Å de résolution. Densité de déformation des liaisons peptidiques 34-35 et 47-48, ainsi que la densité de déformation moyenne sur les 19 liaisons peptidiques de plus faible agitation thermique. Affinement réalisé avec MOLLY. (étape 5). Les contours correspondent à  $\pm 0,02 \text{ e}^-/\text{Å}^3$  (Housset *et al.*, 2000). Des pics positifs de densité de déformation sont bien observés là où les électrons de valence sont attendus. Cette région de la chaîne polypeptidique, structurée en feuillet  $\beta$ , est l'une des plus stables de la protéine : le facteur B est environ de  $5 \text{ Å}^2$ .

L'affinement des charges atomiques des atomes de la liaison peptidique est qualitativement satisfaisant (tableau B.6). Le relatif succès de cet affinement vient sans doute du fait que l'on palie le manque d'information lié à la limite de résolution par une moyennation sur 63 liaisons peptidiques. Cette dernière étape confirme donc bien la présence de l'information dans les données, même si son niveau est faible.

**Tableau B.6** : Charges nettes obtenues pour chacun des 8 atomes de la liaison peptidique, à la fin de l'étape 6.

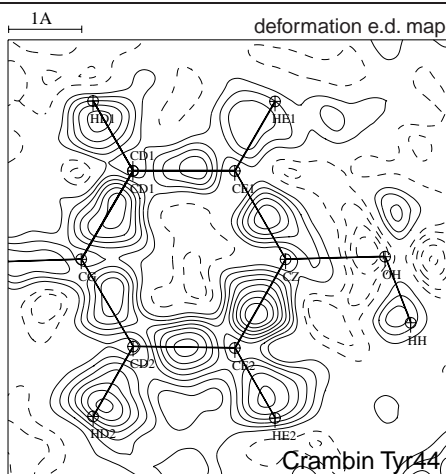
Atome	Type	Charges nettes
N	N-H	-0.20(2)
H	H-N	+0.07(2)
O	O=C	-0.11(2)
C	C=O	+0.20(2)
C	C $\alpha$	-0.23(2)
H	H $\alpha$	+0.04(1)
C	C(sp <sup>3</sup> )	-0.10(2)
C	C(sp <sup>2</sup> )	+0.28(3)

Si l'on tient compte du fait que les données de la toxine II ont été enregistrées à la température ambiante, on peut prévoir un apport bien plus important pour des structures dont les données sont enregistrées à basse température (110 K) et jusqu'à des résolutions de 0.8 Å ou mieux.

Ces tests donnent donc une bonne indication du champ d'application de la méthode. Et quelques protéines, dont de très bonnes données sont désormais disponibles, pourront faire avantageusement l'objet d'une telle étude.

#### B.4.2.3 La crambine à 0.54 Å de résolution.

Un travail est en cours dans l'équipe de C. Lecomte, sur la crambine, une protéine de 46 acides aminés dont les données de diffraction ont été mesurées au synchrotron de Hambourg à 100K jusqu'à 0.54 Å de résolution (Teeter *et al.*, 1997). Il s'agit, à l'heure actuelle, des données de diffraction à la plus haute résolution jamais acquise sur une protéine. Les facteurs d'agitation thermique sont remarquablement bas, de l'ordre de 2 à 4 Å<sup>2</sup> pour la plupart des atomes. La présence d'une hétérogénéité de séquence dans le cristal et le nombre élevé de chaînes latérales de surface possédant plusieurs conformations sont les seuls « défauts » de ce cas d'école. Pour la structure de la crambine, la densité électronique de déformation est encore plus nette (Figure B.11), Pour les liaisons peptidiques, la densité de déformation est comparable à celle obtenue sur des petits peptides, par calculs *ab initio*.



**Figure B.11** : Crambine à 0.54 Å de résolution. Densité de déformation du cycle aromatique de la tyrosine 44, après affinement avec MOLLY. Les contours correspondent à  $\pm 0,05 \text{ e}^-/\text{\AA}^3$ . La plus haute résolution et la plus faible agitation thermique permettent de bien mettre en évidence tous les électrons de valence du cycle. L'atome d'hydrogène du groupement hydroxyle présente, comme pour la tyrosine 5 de la toxine, deux positions diamétralement opposées. (Jelsch *et al.*, 1999).

## B.5 Limites actuelles

La première des limites est la disponibilité de cristaux diffractant à très haute résolution. Ils sont encore peu nombreux, mais les travaux récemment publiés montrent qu'un certain nombre de structures de protéines affinées à une résolution de l'ordre de 0.8 Å devraient voir le jour dans les prochains mois. La qualité des données, notamment traduite par le rapport  $I / \sigma(I)$  ou le  $R_{\text{merge}}$  dans la dernière coquille de résolution, est un paramètre tout aussi fondamental que la résolution extrême du jeu de données.

Les efforts à faire portent essentiellement sur la qualité des cristaux (maîtrise de la cristallisation, cryoprotection et congélation). Ces efforts devraient permettre de diminuer le degré de désordre dans les cristaux de protéines.

La modélisation du solvant sera aussi une des difficultés à surmonter, notamment pour l'estimation du potentiel électrostatique.

## B.6 Perspectives

La disponibilité de nouvelles lignes de lumière sur les synchrotrons de troisième génération, et les nouveaux détecteurs vont rendre les mesures bien plus abordables. Ainsi, on peut penser qu'un certain nombre de questions de biologie structurale que posent les protéines redox, les enzymes ou les complexes protéine-substrat pourront trouver des réponses grâce à ces densités électroniques très précises.

## B.7 Définitions et abréviations:

résolution atomique: Données de diffraction complètes à mieux que 1.20 Å de résolution

résolution sub-atomique: Données de diffraction complètes à mieux que 0.90 Å de résolution

$F_{\text{obs}}$ : Facteur de structure observé

$F_{\text{calc}}$ : Facteur de structure calculé à partir du modèle atomique

$F_{\text{sph}}$ : Facteur de structure calculé à partir du modèle atomique sphérique

$F_{\text{mult}}$ : Facteur de structure calculé à partir du modèle atomique multipolaire

$\phi_{\text{sph}}$ : phase calculée à partir du modèle atomique sphérique

$\phi_{\text{mult}}$ : phase calculée à partir du modèle atomique multipolaire

Facteur R =  $\sum_{\mathbf{h}} (|F_{\text{obs}} - F_{\text{calc}}|) / \sum_{\mathbf{h}} F_{\text{obs}}$



## B.8 Bibliographie

- Anderson, D.H., Weiss, M.S., Eisenberg, D. (1997). Charges, hydrogen bonds, and correlated motions in the 1 Å resolution refined structure of the mating pheromone Er-1 from *Euplotes raikovi*. *J. Mol. Biol.*, **273**, 479-500.
- Bränden, C.I. and Jones, T.A. (1990). Between objectivity and subjectivity. *Nature*, **343**, 687-689.
- CCP4, Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. (1994). *Acta Cryst.*, **D50**, 760-763.
- Coppens P. (1997). X-ray charge densities and chemical bonding. International Union of Crystallography (IUCr). Edited by Oxford University Press.
- Dauter, Z, Lamzin V.S. & Wilson K.S., (1995). Proteins at atomic resolution. *Curr. Opin. Struct. Biol.*, **5**, 784-790.
- Dauter, Z, Lamzin V.S. & Wilson K.S., (1997a). The benefits of atomic resolution. *Curr. Opin. Struct. Biol.*, **7**, 681-688.
- Dauter, Z., Wilson, K.S., Sieker, L.C., Meyer, J. and Moulis, J.M. (1997b). Atomic resolution (0.94 Å) structure of *Clostridium acidurici* ferredoxin. Detailed geometry of [4Fe-4S] clusters in a protein. *Biochemistry*, **36**, 16065-16073.
- Deacon, A.M., Weeks, C.M., Miller, R. and Ealick, S.E. (1998). The Shake-and-Bake structure determination of triclinic lysozyme. *Proc. Natl. Acad. Sci. USA*, **95**, 9284-9289.
- Debaerdemaeker, T. and Woolfson, M.M. (1983). On the application of phase relationships to complex structures. XXII. Techniques for random phase refinement. *Acta Cryst.*, **A39**, 193-196.
- DeTitta, G.T., Weeks, C.M., Thuman, P., Miller, R. and Hauptman, H.A. (1994). Structure solution by minimal-function phase refinement and Fourier filtering. I. Theoretical basis. *Acta Cryst.*, **A50**, 203-210.
- Fontecilla-Camps, J. - C., Habersetzer-Rochat, C. & Rochat, H. (1988). Orthorhombic crystals and three-dimensional structure of the potent toxin II from the scorpion *Androctonus australis* Hector. *Proc. Natl. Acad. Sci. USA*, **85**, 7443-7447.
- Genick, U.K., Soltis, M., Kuhn, P., Canestrelli, I.L., Getzoff, E.D. (1998). Structure at 0.85 Å resolution of an early protein photocycle intermediate. *Nature*, **392**, 206-209.
- Hansen N.K. & Coppens P. (1978). Testing aspherical atom refinements on small-molecule data sets. *Acta Cryst.*, **A34**, 909-921.
- Hirshfeld F.L., (1976). Can X-ray data distinguish bonding effects from vibrational smearing? *Acta Cryst.* **A32**, 239-244.
- Housset D., Habersetzer-Rochat C., Astier J.P., Fontecilla-Camps J.C. (1994). The crystal structure of toxin II from the scorpion *Androctonus australis* Hector refined at 1.3 Å resolution. *J. Mol. Biol.*, **238**, 88-103.
- Housset, D., Benabicha, F., Pichon-Pesme, V., Jelsch, C., Maierhofer, A., David, S., Fontecilla-Camps, J.C. and Lecomte, C. (2000) Towards the charge density study of proteins : a room temperature scorpion toxin structure at 0.96 Å resolution as a first test case. *Acta Cryst.*, **D56**, 151-160.

- Jelsch C, Pichon-Pesme V., Lecomte C. & Aubry A. (1998), Transferability of multipole charge-density parameters: application to very high resolution oligopeptide and protein structures. *Acta Cryst.*, **D54**, 1306-1318.
- Jelsch, C., Teeter, M. M., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (1999) Ultra high resolution structure of crambin: first charge density analysis of a protein. Invited talk, session Macromolecules at high resolution : Refinement and validation, XVIIIth IUCR Congress, Glasgow, UK, 4-11 August 99.
- Kuhn, P., Knapp, M., Soltis, S.M., Ganshaw, G., Thoene, M. and Bott, R. (1998). The 0.78 Å structure of a serine protease: *Bacillus lentus* subtilisin. *Biochemistry*, **37**, 13446-13452
- Lamzin, V. S. and Wilson, K. S. (1993) . Automated refinement of protein models. *Acta Cryst.*, **D49**, 129-149.
- Langs D.A. (1988). *Science* 241, 188-191.
- Lecomte, C. (1995). 261-302. *Advances in Molecular Structure Research*, vol 1, JAI Press Inc. I. and M. Hargittai Eds. Greenwich, CT, USA.
- Longhi, S., Czjzek, M. & Cambillau, C. (1998). Messages from ultrahigh resolution crystal structure. *Curr. Opin. in Struct. Biol.*, **8**, 730-737
- Longhi, S., Czjzek, M., Lamzin, V. , Nicolas, A. and Cambillau, C. (1997). Atomic resolution (1.0 Å) crystal structure of *Fusarium solani* cutinase: stereochemical analysis. *J. Mol. Biol.*, **268**, 779-799.
- Miller, R., DeTitta, G.T., Jones, R., Langs, D.A., Weeks, C.M. and Hauptman, H.A. (1993). On the application of the minimal principle to solve unknown structures. *Science*, **259**, 1430-1433.
- Miranda, F., Kupeyan, C., Rochat, H. & Lissitzky, S. (1970). Purification of animal neurotoxins: isolation and characterization of eleven neurotoxins from the venoms of the scorpions *Androctonus australis* Hector. *Eur. J., Biochem.*, **17**, 477-484.
- Perrakis, A. ,Sixma,T. K. ,Wilson, K.S. and Lamzin V. S. (1997). wARP: improvement and extension of crystallographic phases byweighted averaging of multiple-refined dummy atomic models. *Acta Cryst.*, **D53**, 448-455.
- Pichon-Pesme V., Lecomte C. & Lachekar H. (1995). On building a databank of transferable experimental electron density parameters: application to polypeptides. *J. Phys. Chem.*, **99**, 6242-6250.
- Read, R. J. (1986). Improved fourier coefficient for maps using phases from partial structures with errors. *Acta Cryst.*, **A42**, 140-149.
- Sakabe, N., Sasaki, K. and Sakabe, K. (1981). Insulin structure at 1.1 Å resolution and its dynamic behavior with anisotropic temperature factors. *Acta Cryst.*, **A37** (supplement), 50-53.
- Sevcik, J., Dauter, Z., Lamzin, V.S. and Wilson, K. (1996) Ribonuclease from *Streptomyces aureofaciens* at atomic resolution. *Acta Cryst.*, **D52**, 327-344.
- Sheldrick G.M. and Schneider, T. (1997). SHELXL: high-resolution refinement. *Methods in Enzymology*, Vol. 277, pp 319-343, Academic Press.
- Shomaker, V. & Trueblood, K. N. (1968). On the rigid-body motion of molecules in crystals. *Acta Cryst.*, **B24**, 63-76.

- Smith, G.D., Blessing, R.H., Ealick, S.E., Fontecilla-Camps, J.C., Hauptman, H.A., Housset, D., Langs, and Miller, R. (1997). Ab initio structure determination and refinement of a scorpion protein toxin. *Acta Cryst.*, **D53**, 551-557.
- Teeter, M.M and Hendrickson, W.A. (1979). Highly ordered crystals of the plant seed protein crambin. *J. Mol. Biol.*, **127**, 219-223.
- Teeter, M.M., Lamzin, V.S., Dauter, Z. and Wilson, K. (1997). Crambin crystals at ultra-high resolution of 0.67 Å resolution and 100K. American Crystallographic Association, Annual Meeting Abstract Book, p67.
- Teeter, M.M., Roe, S.M. and Heo, N.H. (1993). Atomic resolution (0.83 Å) crystal structure of the hydrophobic protein crambin at 130 K. *J. Mol. Biol.*, **230**, 292-311.
- Tickle, I.J., Laskowski, R.A. and Moss, D.S. (1998). Error estimates of protein structure coordinates and deviations from standard geometry by full-matrix refinement of gammaB- and betaB2-crystallin. *Acta Cryst.*, **D54**, 243-252.
- Trueblood, K.N. (1990). THMA11. Univ. of California at Los Angeles, USA.
- Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V.S . & Wilson, K. S. (1998). *Acta Cryst.* D54, 522-546.
- Watenpaugh, K.D., Sieker, L.C. and Jensen, L.H. (1979). The structure of rubredoxin at 1.2 Å resolution. *J. Mol. Biol.*, **131**, 509-522.
- Wlodawer, A., Walter, R., Huber, R. and Sjolín, L. (1984). Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. *J. Mol. Biol.*, **180**, 301-329.