UNIVERSITE MONTPELLIER II SCIENCES ET TECHNIQUES DU LANGUEDOC

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

Discipline : Bioinformatique **Formation Doctorale :** Interface Chimie-Biologie **Ecole Doctorale :** Sciences Chimiques et Biologiques pour la Santé

présentée et soutenue publiquement par

Quentin KAAS

le 12 décembre 2005

<u>*Titre</u>:*</u>

ANALYSE STRUCTURALE DES RECEPTEURS D'ANTIGENES DANS IMGT ET MODELISATION MOLECULAIRE.

JURY

Mme Marie-Paule Lefranc, Professeur, Université Montpellier II, Directeur de Thèse
M. Laurent Chiche, Directeur de Recherche, CBS, Co-directeur de Thèse
M. Gérard Lefranc, Professeur, Université Montpellier II, Président
Mme Nathalie Colloc'h, Directeur de Recherche, CNRS UMR 6185, Caen, Rapporteur
M. Jean-Luc Pellequer, HDR, SBTN CEA Marcoul, Rapporteur
M. Patrice Marche, Directeur de Recherche, INSERM unité 548, Grenoble

Merci!

Je tiens à remercier tout d'abord les rapporteurs de cette thèse, Nathalie Coloc'h et Jean-Luc Pellequer, ainsi que les membres du Jury, Patrice Marche et le professeur Gérard Lefranc, qui ont accepté d'évaluer et de commenter ce travail.

Je voudrais exprimer toute ma reconnaissance à mes deux directeurs de thèse, le professeur Marie-Paule Lefranc et Laurent Chiche pour le temps qu'ils m'ont accordé, leur aide et leur soutien. Je remercie particulièrement Marie-Paule de m'avoir permis de participer à de nombreux congrès et workshop, qui m'ont conduit jusqu'en Corée.

Un grand merci à tous les membres de l'équipe d'IMGT, passée ou présente, qui m'ont accompagné durant toutes ces journées terriblement ensoleillées ... à l'extérieur : Domi, Nath, Stéphanie, Céline, Séverine, Aude, Oliver, Fabrice, Bertrand, Julien, Manu, Vincent, Vijay ainsi que les sirènes du laboratoire, Chantal, Géraldine, Joumana et Véro et mes copains de thèses Elodie, Wafae, Christelle, Manu et Olivier. Je remercie également Gérard pour son soutien logistique en denrées libanaises.

Je remercie mes parents ainsi que Dorothée, Eva et Jérémie pour leur soutien ô combien constant et leurs encouragements ô combien nécessaires.

Une pensée pour mes amis de Nantes et de l'école de chimie : Elvina, François, Mes Bibis, Anne-Laure, Virginie, Florent, Colette, Marie-Laure, Floppy et tous ceux que j'oublie ici et qui vont m'en vouloir...

Je remercie également les institutions qui m'ont apporté leur soutien financier, le Ministère de la Recherche et l'ARC.

Préface.

Mon projet de thèse, démarré en octobre 2001, sous la direction scientifique de Marie-Paule Lefranc, Professeur à l'Université Montpellier II et Membre Senior de l'Institut Universitaire de France, responsable avec le Professeur Gérard Lefranc du laboratoire d'Immunogénétique Moléculaire à l'Institut de Génétique Humaine (UPR CNRS 1142), et de Laurent Chiche, Directeur de Recherche au Centre de Biochimie Structurale (UMR CNRS 5048), a pour but l'analyse structurale des récepteurs d'antigènes dans IMGT et leur modélisation moléculaire.

A la suite des travaux de thèse de Manuel Ruiz, j'ai poursuivi le développement de la base de données de structures, IMGT/3Dstructure-DB. Elle fait partie intégrante d'IMGT, le système d'information international en ImMunoGénéTique spécialisé dans les immunoglobulines (IG), les récepteurs T (TR), le complexe majeur d'histocompatibilité (MHC), et les protéines apparentées du système immunitaire (RPI, pour *related proteins of the immune system*) de l'homme et des autres vertébrés (Chapitre 1). IMGT/3Dstructure-DB fournit une ressource unique d'expertise de données structurales d'IG, de TR, de MHC, et de RPI. Les IG, les TR, les MHC et les RPI sont des protéines qui ont été largement étudiées au niveau structural. Néanmoins, en raison de la complexité de leur polymorphisme et, pour les IG et les TR, de leur synthèse, une base de données de structures tridimensionnelles spécialisée est nécessaire, en particulier pour identifier de manière experte les gènes et les allèles codant ces protéines ou pour des analyses statistiques poussées.

Avant mon arrivée au laboratoire, la première version IMGT/3Dstructure-DB gérait les structures 3D d'IG complète ou de Fab d'IG. Elle proposait l'identification des gènes et allèles et également la numérotation unique d'un type de domaine protéique. J'ai repris le développement de IMGT/3Dstructure-DB et complètement modifié sa structure afin de pouvoir gérer n'importe quel type de fragment de protéine et de domaine, que ce soit pour les IG, les TR, les MHC ou les RPI (Chapitre 2). Dans la nouvelle organisation d'IMGT/3Dstructure-DB le concept de base est la position IMGT des acides aminés dans un domaine, ce qui a notamment permis la caractérisation des interfaces moléculaires. Les données de références utilisées lors de l'annotation comprennent les noms des gènes et allèles de références, les séquences correspondantes et les termes de description qui sont stockés dans des tables complémentaires, ce qui contribue sensiblement à la réduction de la redondance des données. J'ai développé des programmes de gestion des entrées de la base de données, qui traitent automatiquement les nouvelles structures 3D, et une interface d'annotation semi-manuelle, qui autorise le paramétrage de chacune des étapes. Finalement j'ai amélioré l'interface donnant accès aux informations de la base de données. Elle permet dorénavant la recherche des structures suivant de nombreux critères et l'affichage sous forme d'un tableau synthétique des données sélectionnées. Cette nouvelle interface donne également accès aux outils de visualisation, d'analyse de contacts, ainsi qu'aux fichiers de coordonnées renumérotés selon la numérotation unique IMGT.

La standardisation des données structurales d'IMGT/3Dstructure-DB m'a permis de mettre en place une analyse automatique des structures des récepteurs d'antigènes qui repose sur un grand nombre d'entrées (Chapitre 3). Le nombre de structures des récepteurs d'antigènes et de leurs complexes est recensé. Les structures des domaines protéiques, qui constituent les récepteurs d'antigènes, sont caractérisées par (1) leur variation de composition en acides aminés, (2) leur variation de topologie, (3) la mesure de certains angles caractéristiques, (4) leur variation de conformation et (5) la conservation des contacts entre leurs résidus. Les arrangements des domaines dans les récepteurs sont comparés. Les différentes interfaces moléculaires entre les récepteurs et les antigènes – antigène/IG, peptide/MHC et TR/peptide/MHC – sont analysées. Chaque interface a été le sujet d'une étude de variabilité conformationnelle du complexe ou d'un élément du complexe lors de la liaison à l'antigène, et d'une analyse des contacts entre positions IMGT. Ces données sont particulièrement utiles à la caractérisation des comportements structuraux moyens des IG, TR et MHC.

Par ailleurs, j'ai ensuite mis en place plusieurs protocoles qui permettent notamment l'isolement des chaînes, des domaines et des complexes avec un antigène, la création de graphiques de population et la superposition des différentes informations sur une représentation en deux dimensions des domaines protéiques. Ces protocoles pourront être facilement étendus pour gérer de nouvelles analyses qui seront automatiquement mises à jour en même temps que IMGT/3Dstructure-DB.

La dernière partie de ma thèse concerne la modélisation moléculaire de complexes peptide/MHC et de complexes TR/peptide/MHC (Chapitre 4). Les phénomènes immunologiques impliquent des groupes de complexes peptide/MHC et TR/peptide/MHC dont la population est importante. Pourtant aucune méthode actuelle ne permet de simuler les structures d'un nombre aussi important de complexes. J'ai développé des techniques de modélisation qui s'inscrivent dans ce contexte et permettent de modéliser rapidement les chaînes latérales et la conformation des boucles protéiques. Mon algorithme de prédiction des chaînes latérales, SCHISMo, utilise une hiérarchie de représentations simplifiées et des heuristiques simples. Il a été appliqué à la modélisation des peptides dans les complexes peptide/MHC-II. Ses performances se comparent favorablement aux algorithmes couramment employés, et les objectifs de rapidité et de précision sont atteints. L'algorithme de modélisation de conformations de boucles, LLIPA, utilise une bibliothèque de conformations de fragments de chaîne principale (dipeptides). Cet algorithme a pour but la génération d'un ensemble de conformations de boucles proches de la conformation native, tout en conservant la population de cet ensemble suffisamment réduite. Pour cela la bibliothèque de conformations de chaîne principale a été classée suivant la représentativité des conformations dans des structures expérimentales, et la méthode de recherche de conformations de boucles fait usage de ce classement des conformations. LLIPA a été appliquée avec succès à la modélisation des peptides dans les complexes peptide/MHC-I et à la modélisation des boucles de TR dans les complexes TR/peptide/MHC.

Table des matières

Ρ	réfac	ce.		i
I	ntrod	luction Rôle d Recon Divers	les lymphocytes B et T	5 5 6 6
1	Star	ndardi	sation des données structurales dans IMGT.	8
	1.1	Les co	ncepts d'IMGT-ONTOLOGY	8
	1.2	Struct	ure tridimensionnelle des immunoglobulines	9
	1.3	Struct	ure tridimensionnelle des récepteurs T	12
	1.4	Struct	ure tridimensionnelle des MHC	13
	1.5	Struct	ures 3D et Colliers de Perles IMGT des V-DOMAIN, C-DOMAIN et	
		G-DO	MAIN	15
		1.5.1	V-DOMAIN	15
		1.5.2	C-DOMAIN	15
		1.5.3	G-DOMAIN	18
2	IMO	T/3D)structure-DB.	20
_	2.1	Organ	isation de IMGT/3Dstructure-DB.	$\frac{-}{21}$
		2.1.1	Données d'obtention des structures.	21
		2.1.2	Description des molécules	23
		2.1.3	Caractérisation des positions pour chaque résidu.	24
		2.1.4	Contacts entre résidus et entre domaines	25
		2.1.5	Séquences des gènes et allèles de références.	26
		2.1.6	Termes de DESCRIPTION.	26
	2.2	Admir	nistration de IMGT/3Dstructure-DB	26
		2.2.1	Sélection des nouvelles entrées.	26
		2.2.2	Interface d'annotation des nouvelles entrées.	26
		2.2.3	Etapes d'annotation des nouvelles entrées.	29
	2.3	Interfa	ce utilisateur.	33
		2.3.1	Génération des Colliers de Perles IMGT	34
		2.3.2	Liste des séquences de référence IMGT	34
	2.4	Conclu	usion	35

3	Ana	alyse st	tructurale des récepteurs d'antigènes dans IMGT.	37
	3.1	Nomb	re de structures 3D dans IMGT/3Dstructure-DB. \ldots \ldots \ldots \ldots	. 39
		3.1.1	Nombre de structures 3D d'IG, de TR et de MHC	. 39
		3.1.2	Structures de complexes IG/antigène, pMHC et TR/pMHC	. 39
		3.1.3	Représentativité génétique	. 39
	3.2	Analys	se des domaines	. 40
		3.2.1	Longueur des séquences	. 42
		3.2.2	Composition en acides aminés	. 42
		3.2.3	Distribution des brins dans les feuillets.	. 45
		3.2.4	Angles entre feuillets des V-DOMAINs et C-DOMAINs.	. 47
		3.2.5	Alignements structuraux.	. 48
		3.2.6	Analyse de contacts.	53
	3.3	Organ	isation des domaines dans les récepteurs	. 57
		3.3.1	V-PARTNER.	. 58
		3.3.2	C-PARTNER.	. 62
		3.3.3	Entre le V-PARTNER et le C-PARTNER.	. 63
	3.4	Interfa	ace des complexes IG/antigène	. 64
		3.4.1	Superposition des CDR-IMGT.	. 64
		3.4.2	Positions en contacts	. 64
	3.5	Interfa	ace des complexes peptide/MHC	. 67
		3.5.1	Flexibilité du peptide dans le groove	. 70
		3.5.2	Interaction peptide/MHC	. 70
		3.5.3	Sites de contacts IMGT	. 74
	3.6	Interfa	ace des complexes TR/peptide/MHC	. 79
		3.6.1	Orientation du TR	. 79
		3.6.2	Surface enfouie	. 80
		3.6.3	Positions en interactions	. 82
	3.7	Conclu	usion	. 84
4	Mo	délisati	ion peptide/MHC et TR/peptide/MHC.	86
	4.1	Prédic	tion de la liaison peptide/MHC – Bibliographie	. 88
		4.1.1	Alignements de séquences de peptides	. 88
		4.1.2	Threading	. 90
		4.1.3	Construction de la structure 3D du complexe peptide/MHC	. 91
	4.2	Modél	isation des peptide/MHC-II	. 92
		4.2.1	Approximation de l'espace conformationnel des chaînes latérales .	. 93
		4.2.2	Stratégie de recherche	. 98
		4.2.3	Evaluation des conformations	. 102
		4.2.4	Résultats	. 102
	4.3	Modél	isation des peptide/MHC-I et des boucles du TR – Construction de	
		boucle	s par la méthode LLIPA	. 107
		4.3.1	Prédiction de la conformation des boucles – Bibliographie	. 107
		4.3.2	Algorithme LLIPA	. 111

	 4.3.3 Modélisation de peptides dans les structures RX	. 115
4.4	Conclusion	. 117
Concl	usion.	121
Annez	xe.	137
Public	ations.	169

Publications.

Publications majeures

Kaas Q, Ruiz M and Lefranc MP. IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. *Nucleic Acids Res.* 2004;32 :D208-D210.

Kaas Q and Lefranc MP. Interactive IMGT on line database and tool for the structural analysis of immunoglobulin, T cell receptor, MHC and related proteins. *Focus on immunology research*. Nova Science 2005; (in press)

Kaas Q, Chiche L and Lefranc MP. IMGT unique numbering for standardized contact analysis of immunoglobulin/antigen and T cell receptor/peptide/MHC complexes. *In* : BIOINFO2005, Proceedings of the 2005 International Conference of InCoB, AASBi and KSBI, Lee D., Wong L., Kim D.-W., Tan T.W, Lee K.-H. (KAIST PRESS, Daejon, Korea)

Kaas Q and Lefranc MP. T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. In Silico Biol. 2005; 5:0046.

Lefranc MP, Pommie C, **Kaas Q**, Duprat E, Bosc N, Guiraudou D, Jean C, Ruiz M, Da Piedade I, Rouard M, Foulquier E, Thouvenin V and Lefranc G. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev Comp Immunol.* 2005;29:185-203.

Lefranc MP, Duprat E, **Kaas Q**, Tranne M, Thiriot A and Lefranc G. IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. *Dev Comp Immunol.* 2005;29 :917-938.

Autres publications

Lefranc MP, Giudicelli V, **Kaas Q**, Duprat E, Jabado-Michaloud J, Scaviner D, Ginestoux C, Clement O, Chaume D and Lefranc G. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res.* 2005;33:D593-597.

Lefranc MP, Giudicelli V, Ginestoux C, Bosc N, Folch G, Guiraudou D, Jabado-Michaloud J, Magris S, Scaviner D, Thouvenin V, Combres K, Girod D, Jeanjean S, Protat C, Yousfi-Monod M, Duprat E, **Kaas Q**, Pommie C, Chaume D and Lefranc G. IMGT-ONTOLOGY for immunogenetics and immunoinformatics. *In Silico Biol.* 2004;4 :17-29. Epub 2003 Nov 22.

Gelly JC, Gracy J, **Kaas Q**, Le-Nguyen D, Heitz A and Chiche L. The KNOTTIN website and database : a new information system dedicated to the knottin scaffold. *Nucleic Acids Res.* 2004;32 :D156-D159.

Lefranc MP, Clement O, **Kaas Q**, Duprat E, Chastellan P, Coelho I, Combres K, Ginestoux C, Giudicelli V, Chaume D and Lefranc G. IMGT-Choreography for immunogenetics and immunoinformatics. *In Silico Biol.* 2005;5:45-60.

Introduction.

La description de mes travaux de thèse nécessite la connaissance de quelques éléments d'immunologie qui vont être développés succinctement. Le système immunitaire d'un organisme est un ensemble de mécanismes chimiques et cellulaires ayant pour but la destruction et l'élimination des organismes envahisseurs et de toutes molécules étrangères à l'hôte (invasion bactérienne, fongique, virale, ou tumorale). Des cellules et des molécules spécialisées sont impliquées dans la distinction du *soi* et du *non soi*. Chez les espèces vertébrés à mâchoires (Gnathostomes), les cellules B et T, encore appelées *lymphocytes* B et T, font partie des cellules du système immunitaire.

Rôle des lymphocytes B et T

Les lymphocytes B, qui se développent chez l'homme dans la moelle osseuse (B pour bone), produisent des immunoglobulines (notées IG et encore appelées anticorps) [1]. Ces glycoprotéines se lient aux molécules organiques ou minérales étrangères à l'organisme, appelées antigène (en anglais antigen pour antibody generator). La liaison des anticorps permet le marquage des virus, des toxines bactériennes, et des cellules étrangères afin qu'ils soient détruits (par des cellules phagocytaires ou les molécules du complément). Les immunoglobulines peuvent se trouver soit sous une forme circulante (dans un état de multimères) soit sous forme de protéines membranaires à la surface des lymphocytes B. Les lymphocytes T, qui se développent chez l'homme dans le thymus, sont de deux types. Ceux du premier type expriment le corécepteur CD4. La plupart d'entre eux sont des lymphocytes T auxiliaires qui aident à la différenciation des lymphocytes B. Ceux appartenant au second type expriment le corécepteur CD8 et sont pour la plupart cytotoxiques. Ils identifient et détruisent les cellules du soi infectées par un virus et les cellules tumorales. A la surface de la plupart des cellules de l'organisme, le complexe majeur d'histocompatibilité (MHC) expose des peptides issus de la dégradation des protéines. Les récepteurs T (TR), présents à la surface des lymphocytes T, se lient au complexe peptide/MHC (pMHC) [2]. Le TR fait partie du complexe du récepteur T (TcR), un complexe de surface transmembranaire à plusieurs sous-unités comprenant également des chaînes CD3, un corécepteur CD4 ou CD8, un costimulateur CD28 ou CTLA-4, et une chaîne CD2 [2]. Une liaison entre les TR et les pMHC peut activer la cellule T grâce à des interactions avec les CD3 et d'autres composant du TcR. La cellule T activée secrète des perforines, des protéines perforant les membranes, qui vont détruire la cellule possédant le pMHC. Il existe deux classes de MHC, le MHC de classe I (MHC-I) et le MHC de classe II (MHC-II), qui se lient à des peptides ayant des origines différentes. Les MHC-I lient des peptides endogènes; la liaison au peptide intervient dans le réticulum endoplasmique et le MHC-I est transporté à travers l'appareil de Golgi jusqu'à la membrane plasmique [3]. Les MHC-I peuvent également se lier avec des molécules de nature non protéique, comme un hème [4]. Les MHC-II se lient à des peptides exogènes; la liaison intervient dans des compartiments endosomiques après la dégradation des protéines.

Reconnaissance du soi et développement des lymphocytes B et T

Le système immunitaire doit pouvoir faire la distinction entre les molécules du soi et du non soi. Ce partage est effectué par la *sélection clonale négative* qui élimine les lymphocytes B et T qui réagissent vis-à-vis des molécules et cellules de l'hôte. La sélection clonale négative des lymphocytes B intervient dans la moelle des os. Les lymphocytes T doivent de plus reconnaître les MHC du soi tout en étant capable de détecter les peptides exogènes. Dans le thymus, les lymphocytes T immatures sont en contact avec les cellules exprimant les MHC-I et les MHC-II complexés avec des peptides endogènes. Deux types de sélection clonale interviennent, une sélection négative si la liaison est trop forte, ce qui contribue à empêcher les phénomènes d'auto-immunité, et une sélection positive si la liaison est plus faible, ce qui assure la reconnaissance des MHC du soi.

Diversité génétique des IG, des TR et des MHC.

L'énorme diversité des molécules exogènes nécessite une grande diversité des récepteurs d'antigènes. Le nombre de cellules impliquées dans le système immunitaire est considérable, et chez l'homme la quantité en masse de lymphocytes est comparable à celle des cellules cérébrales. Pour citer une autre statistique frappante, chez l'homme, les IG représentent 20% en poids des protéines du plasma.

Des études biochimiques montrent que les IG sont constituées de deux chaînes lourdes identiques et de deux chaînes légères identiques. Chez l'homme et les autres mammifères, il existe deux types de chaînes légères (lambda et kappa) et cinq types de chaîne lourde (alpha, delta, epsilon, gamma, mu). La séquence des chaînes lourdes est notamment composée de quatre régions, notées V-REGION (variable), D-REGION (diversité), J-REGION (jonction) et C-REGION (constant). Chaque région est codée par un gène appartenant à une famille multigénique. Lors de la maturation des cellules B, le réarrangement combinatoire entre les gènes V et J pour les chaînes légères et V, D et J pour les chaînes lourdes représente une première source de diversité. Au niveau des jonctions entre les gènes V et J d'une part et, V et D-J d'autre part, une diversité très importante est créée par le mécanisme de N-diversité caractérisé par des délétions et des additions aléatoires de nucléotides [1]. Finalement l'apparition de *mutations hypersomatiques* dans la séquence nucléotidique des IG est une source additionnelle de diversité. De façon similaire, les séquences nucléotidiques des chaînes légères se composent de V-REGION, J-REGION et C-REGION (pas de D-REGION). Chaque région est codée respectivement par un gène V, J et C, et les gènes V et J sont soumis aux mêmes mécanismes de diversité combinatoire de réarrangement, de diversité de jonction (mais à un degré nettement moindre que les chaînes lourdes) et de mutations hypersomatiques.

Chaque TR est un hétérodimère formé soit d'une chaîne de type *alpha* (TR-ALPHA) et d'une chaîne de type *bêta* (TR-BETA), soit d'une chaîne de type *gamma* (TRG) et d'une chaîne de type *delta* (TRD). Les chaînes bêta et delta sont composées de V-REGION, D-REGION, J-REGION et C-REGION, comme les chaînes lourdes des IG. Les chaînes alpha et gamma sont composées des V-REGION, J-REGION et C-REGION, comme les chaînes légères des IG. De façon similaire aux IG, chacune de ces régions est codée par un gène appartenant à une famille multigénique et le réarrangement combinatoire V-J ou V-D-J lors de la maturation des lymphocytes T est responsable de la diversité génétique des séquences nucléotidiques. Les mécanismes de N-diversité des jonctions V-J et V-DJ sont sources de diversité mais les séquences de TR ne sont pas soumises à des mutations hypersomatiques.

Les MHC-I sont constitués d'une chaîne lourde, noté *alpha* (I-ALPHA), siège de la diversité génétique, et d'une chaîne légère, notée *bêta-2-microglobuline* (B2M), qui est remarquablement bien conservée même entre espèces. Les MHC-II sont constitués de deux chaînes, *alpha* (II-ALPHA) et *bêta* (II-BETA), qui sont de taille équivalente. Les chaînes alpha de MHC-I et chaque chaîne de MHC-II sont chacune codées par un seul gène appartenant à une famille multigénique. La diversité des MHC au sein d'un organisme est bien moindre que pour les IG et les TR mais il existe une très grande diversité au sein d'une espèce [5]. Ceci est conforme au rôle du MHC puisqu'il sert de marqueur de l'unicité du soi. La description des événements génétiques à l'origine des séquences nucléotidiques et protéiques des IG, des TR et des MHC est complexe. L'ensemble des données de séquences mais aussi d'interactions, de spécificités et la description de la structure tridimensionnelle de ces molécules nécessite un vocabulaire unifié. Une description standardisée permet la comparaison et l'organisation des données, et ainsi la compréhension des mécanismes dont ils sont l'objet¹.

¹«Toute la méthode consiste dans l'ordre et l'arrangement des objets sur lesquels il faut faire porter la pénétration de l'intelligence pour découvrir quelque vérité.» Descartes – Règles pour la direction de l'esprit, règle V

Chapitre 1

Standardisation des données structurales dans IMGT.

Sommaire

1.1	Les concepts d'IMGT-ONTOLOGY	8
1.2	Structure tridimensionnelle des immunoglobulines	9
1.3	Structure tridimensionnelle des récepteurs T	12
1.4	Structure tridimensionnelle des MHC	13
1.5	Structures 3D et Colliers de Perles IMGT des V-DOMAIN,	
	C-DOMAIN et G-DOMAIN.	15
	1.5.1 V-DOMAIN	15
	1.5.2 C-DOMAIN	15
	1.5.3 G-DOMAIN	18

1.1 Les concepts d'IMGT-ONTOLOGY.

Depuis 1989, IMGT, the international ImMunoGeneTics information system¹ (\mathbb{B} [6] [7], http://imgt.cines.fr, spécialisé en ImMunoGénéTiques, rassemble et standardise les données génétiques et structurales sur les IG, les TR, les MHC et les protéines apparentés du système immunitaire, les RPI (RPI pour *Related Proteins of the Immune system*). Les annotations IMGT sont décrites selon les règles de la charte scientifique IMGT, qui est basée sur IMGT-ONTOLOGY [8]. La standardisation IMGT unifie la description des IG, des TR et des MHC au niveau génétique et structural. Elle permet une comparaison directe de ces protéines hautement polymorphes et de leurs complexes, quelque soit le récepteur, le type de chaîne et l'espèce. La charte scientifique d'IMGT fournit le vocabulaire contrôlé ainsi que les concepts et règles d'annotation permettant l'identification, la description, la classification et la numérotation des IG, TR et MHC. Les mots de ce vocabulaire sont écrits en majuscule.

¹Un système d'information est un ensemble de bases de données, d'outils et de ressources internet.

- Le concept d'IDENTIFICATION fournit les mots clés standardisés IMGT indispensables pour l'identification des séquences.
- Le concept de DESCRIPTION fournit les descripteurs standardisés IMGT pour la description des sous-régions structurales et fonctionnelles qui composent les séquences et les structures des IG, des TR et des MHC. Les descripteurs standardisés caractérisent également les assemblages tridimensionnelles de domaines et de chaînes. Ainsi une immunoglobuline formée de deux chaînes kappas et de deux chaînes lourdes, dont les gènes C sont de type mu, est notée IG-MU_KAPPA.
- Le concept de CLASSIFICATION fournit à l'immunologiste et au généticien, une nomenclature standardisée pour chaque locus et chaque espèce. L'ontologie IMGT nomme les gènes et allèles et fournit très simplement leur classification dans la hiérarchie des groupes et sous-groupes de gènes. Ainsi l'allèle IGHV1-2*02 humain, est un gène IGHV1-2 appartenant au sous-groupe IGHV1, qui appartient au groupe IGHV (ensemble des V-GENE du locus de chaîne lourde IGH). En 1999, la nomenclature IMGT des gènes humaines des IG et des TR a été agréée par le comité de nomenclature de HUGO (*HUman Genome Organisation*).
- Le concept de NUMEROTATION fournit la numérotation unique IMGT des séquences des IG et TR, des MHC et des RPI. Les numérotations uniques IMGT sont extrêmement utiles pour la comparaison des séquences et des structures.
- Le concept d'OBTENTION fournit les termes permettant de préciser l'origine des séquences d'IG, de TR et de MHC ainsi que les conditions dans lesquelles ces séquences ont été obtenues.
- Le concept d'ORIENTATION comprend le concept d'orientation génomique et le concept d'orientation des brins d'ADN.

Les vocabulaire standardisé IMGT va être défini et utilisé pour décrire les structures 3D des IG, des TR et des MHC.

1.2 Structure tridimensionnelle des immunoglobulines.

La première structure 3D d'une IG a été déterminée en 1975 par cristallographie par diffraction des rayons X. Des études ultérieures ont révélé la composition en domaines des chaînes lourdes et légères. Les domaines formés par le repliement des chaînes d'IG sont très similaires et se classent en deux catégories, les domaines variables (V-DOMAIN) et les domaines constants (C-DOMAIN). Chaque chaîne légère d'IG est composée d'un V-DOMAIN et d'un C-DOMAIN (Figure 1.1). Chaque chaîne lourde est composée d'un V-DOMAIN et de trois à quatre C-DOMAINs, suivant le type de chaîne lourde. Les V-DOMAINs correspondent à la V-D-J-REGION pour les chaînes lourdes, et pour les chaînes légères à la V-J-REGION. La C-REGION de chaque chaîne se replie en plusieurs C-DOMAINs. La chaîne lourde possède une région charnière flexible qui divise la molécule d'immunoglobuline en trois éléments plus rigides : deux fragments de liaison de l'anticorps (Fab, pour *fragment antigen binding*) et un fragment dit constant (Fc). Le



Figure 1.1: Structure 3D d'une IG, et Colliers de Perles IMGT de V-DOMAIN et C-DOMAIN. (A) Vue de face d'une IG-GAMMA-2A_KAPPA de souris (1igt) avec identification des domaines (description IMGT du domaine suivie de son numéro le long de la chaîne protéique, de D1 à D4). (B) Collier de Perles IMGT du VH (V-DOMAIN) de la chaîne lourde gamma 2A de souris (H-GAMMA-2A) (1igt_B). (C) Collier de Perles IMGT du CH1 (C-DOMAIN) de la chaîne H-GAMMA-2A (1igt_B). Dans les représentations Colliers de Perles, les acides aminés sont indiqués par leur abréviation à une lettre. Les acides aminés hydrophobes et le tryptophane trouvés à une position dans plus de 50% des séquences d'IG et de TR analysés sont colorés en bleu. Toutes les prolines sont présentés en jaune. Pour le V-DOMAIN (A), les CDR-IMGT sont délimités par des acides aminés représentés par des carrés. Les cercles hachurés correspondent aux positions manquantes de cette séquence dans la numéritation unique IMGT des V-DOMAINs et C-DOMAINs. Le CDR1-IMGT est en rouge, le CDR2-IMGT est en jaune et le CDR3-IMGT est en violet. Les flèches indiquent les directions des feuillets bêta et leur désignation dans les structures 3D.

Fab est constitué du V-DOMAIN et du premier C-DOMAIN d'une chaîne lourde et d'une chaîne légère. Le Fc est constitué de quatre ou six C-DOMAINs C-terminaux appartenant aux deux chaînes lourdes. La mise en commun du V-DOMAIN d'une chaîne lourde et d'une chaîne légère appariés dans la structure (fragment variable Fv) constitue le site de liaison de l'anticorps. Le deuxième C-DOMAIN de chaque chaîne lourde gamma est glycosylé².

Le domaine d'IG est une structure de type sandwich bêta, constituée de deux feuillets bêtas antiparallèles reliés par au moins un pont disulfure [9]. Les feuillets des V-DOMAINs possèdent respectivement 4 brins et 5 brins. Les V-DOMAINs possèdent trois boucles qui assurent la liaison avec un antigène. Ces boucles ou CDR (*Complementarity Determining Region*) sont les régions les plus variables en séquence des V-DOMAINs et sont également appelées boucles hypervariables. La boucle CDR3 contient la jonction entre les régions V, (D) et J, si bien qu'elle possède une variabilité de séquence supérieure aux CDR1 et CDR2. Les régions de la séquence des V-DOMAIN and dehors des CDR sont appelées FR (*Framework Region*). Les feuillets des C-DOMAINs possèdent respectivement 3 et 4 brins.

Le domaine de type immunoglobuline est très communément représenté dans la nature. Dans les protéines autres que les IG et les TR, ces domaines sont nommés V-LIKE-DOMAIN ou C-LIKE-DOMAIN, suivant leur similarité avec les V-DOMAINs et C-DOMAINs respectivement. Les IG, les TR et les protéines possédant au moins un V-LIKE-DOMAIN ou un C-LIKE-DOMAIN constituent la superfamille des immunoglobulines (IgSF).

La numérotation unique IMGT est applicable à tous les domaines de type IG des protéines IgSF [10] [11] (Publication 1). Elle permet d'unifier la description de la topologie de ces domaines et elle est utilisée pour la représentation Collier de Perles IMGT [12] . Celle-ci est une représentation à deux dimensions décrivant les boucles de reconnaissances des V-DOMAINs (CDR1-IMGT, CDR2-IMGT et CDR3-IMGT), la composition des brins et des feuillets bêta, et les positions hydrophobes conservées. Les caractéristiques conservées des domaines de type V ou C (V-DOMAIN, V-LIKE-DOMAIN, C-DOMAIN ou C-LIKE-DOMAIN) sont les cystéines 23 et 104 qui forment un pont disulfure, et les autres positions hydrophobes du coeur du domaine, le tryptophane 41 et la position 89. Le V-DOMAIN est constitué de deux feuillets comprenant les brins A, B, E et D, et les brins C, C', C", F et G. Le C-DOMAIN est constitué de deux feuillets comprenant les brins A, B, (D) et E, et les brins C, (D), F et G. L'appartenance du brin D à l'un ou l'autre des feuillets est variable suivant la protéine. Le C-DOMAIN possède un brin CD, non présent dans les V-DOMAINS, qui relie les brins C et D. Avec la numérotation unique IMGT, les régions de séquences suivantes décrivent tous les V-DOMAINs et V-LIKE-DOMAINs : le brin A (positions 1 à 15), le brin B (positions 16 à 26), la boucle CDR1-IMGT des V-DOMAINs (positions 27 à 38), le brin C (positions 39 à 46), le brin C' (positions 47 à 55), la boucle CDR2-IMGT des V-DOMAINs (positions 56 à 65), le brin C" (positions 66 à 74), le brin D (positions 75 à 84), le brin E (positions 85 à 94), le brin F (positions

 $^{^{2}}$ La fonction exacte des glycosylations est difficile à établir. Elles ont des rôles putatifs dans le repliement et dans le transport des protéines, elles empêchent les interactions non spécifiques et permettent d'éloigner les molécules de la membrane.

100 à 104), la boucle CDR3-IMGT des V-DOMAINs (positions 105 à 117) et le brin G (positions 118 à 128).

Les C-DOMAINS et C-LIKE-DOMAINS n'ont pas pas de brins C' et C" mais possèdent un brin transversal CD et quatorze positions d'insertions ont été ajoutées pour tenir compte de longueur plus importante des brins D et E que dans les V-DOMAINS et V-LIKE-DOMAINS. La gestion de la numérotation des positions dans les boucles et les insertions, où les alignements structuraux ne sont plus significatifs, est soumis à un ensemble de règles [10] [11].

1.3 Structure tridimensionnelle des récepteurs T.

Le premier aperçu d'une structure tridimensionnelle de TR a été rapporté en 1994 par Bentley et al. [13] qui ont résolu la structure par diffraction des rayons X d'une chaîne complète de TRB (code 1bec). En 1995, Fields et al [14] ont cristallographié la structure d'un FV-ALPHA_BETA³ (code 1jck). Finalement la structure complète d'un TR-ALPHA_BETA a été résolue en 1996 par Garcia et al [15] (code 1tcr). Les structures 3D révèlent que chaque chaîne de TR est composée d'un V-DOMAIN, d'un C-DOMAIN

³Termes décrivant un récepteur constitué de deux V-DOMAINs, un V-ALPHA et un V-BETA.



Figure 1.2: Structure 3D d'un récepteur T (TR) (1fyt). V-ALPHA et V-BETA sont des V-DOMAINS, et C-ALPHA et C-BETA-1 sont des C-DOMAINS. Les régions transmembranaires et intracytoplasmiques ne sont pas représentés dans les structures 3D. La chaîne TR-ALPHA utilise les allèles suivants : TRAV8-4*05 (V-REGION) et TRAJ48*01 (J-REGION) associé à TRAC*01 (C-REGION). La chaîne TR-BETA-1 utilise les allèles suivants : TRBV28*01 (V-REGION) et TRBJ1-2*01 (J-REGION) associé à TRBC1*01 ou TRBC1*02 (C-REGION).

et d'une partie transmembranaire (Figure 1.2). La partie extracellulaire du TR ressemble donc à un Fab d'IG. Les V-DOMAINs correspondent à la V-J-REGION pour les TRA et TRG, et à la V-D-J-REGION pour les TRB et TRD. La mise en commun des CDR-IMGT des V-DOMAINs des deux chaînes partenaires (TR-ALPHA et TR-BETA d'une part, TR-GAMMA et TR-DELTA de l'autre) constituent le site de liaison du pMHC. La numérotation unique IMGT s'applique aux domaines du TR et elle permet leur comparaison aux domaines d'IG et leur représentation sous forme de Colliers de Perles IMGT.

1.4 Structure tridimensionnelle des MHC.

La première structure de MHC-I a été résolue en 1985 [16]. La première structure de MHC-II date de 1993 [17] et sa détermination fut un travail épique ayant requis l'utilisation concomitante de plusieurs techniques cristallographiques. Les MHC-I sont composés d'une chaîne lourde possédant deux G-DOMAINs (groove domains), d'un C-LIKE-DOMAIN, d'une partie transmembranaire et d'une très courte région cytoplasmique. La chaîne légère de MHC-I est appelée bêta-2-microglobuline (B2M) et possède un unique C-LIKE-DOMAIN.

Les MHC-II sont composés d'une chaîne alpha et d'une chaîne bêta. Chaque chaîne se compose d'un G-DOMAIN, d'un C-LIKE-DOMAIN, d'une partie transmembranaire et d'une très courte régions cytoplasmique.

De manière intéressante, les structures 3D des MHC-I et des MHC-II révèlent une architecture globalement similaire, au total deux G-DOMAINs et deux C-DOMAINs, mais les domaines sont répartis différent entre les chaînes. Chaque G-DOMAIN est composé d'un feuillet bêta antiparallèle de quatre brins et d'une région alpha qui se divise en deux hélices alpha. Les deux G-DOMAINs d'un MHC, qu'ils appartiennent à une même chaîne (domaines G-ALPHA1 et G-ALPHA2 de la chaîne I-ALPHA) ou non (domaine G-ALPHA de la chaîne II-ALPHA et domaine G-BETA de la chaîne II-BETA), s'assemblent pour former un unique feuillet bêta sur lequel repose les deux hélices alpha arrangées symétriquement. Les hélices forment un angle de 40° avec les brins du feuillet bêta. Les deux C-DOMAINs se positionnent sous le feuillet bêta. Le site de liaison du peptide à la forme d'un sillon. Il est constitué par le feuillet bêta (fond du sillon) et le flanc des hélices alpha (parois du sillon). Les analyses de polymorphisme montrent que les positions les plus variables se retrouvent dans ce site de liaison au peptide [4]. Une numérotation unique IMGT des G-DOMAINs et G-LIKE-DOMAINs a été déterminée [18] (Publication 2) de manière très similaire à la numérotation unique IMGT des domaines IgSF [10]. Pour chaque feuillet des G-DOMAINs, le brin A est défini des positions 1 à 14, le brin B des positions 18 à 28, le brin C des positions 31 à 38 et le brin D des positions 42 à 47. Par rapport au canevas commun de la numérotation des positions de tous les G-DOMAINs et G-LIKE-DOMAINs, les domaines G-ALPHA1 et G-ALPHA possèdent une insertion en position 7, et les domaines G-ALPHA2 et G-BETA possèdent une insertion en position 61 ainsi qu'en position 72. Un pont disulfure interne aux G-ALPHA2 et aux G-BETA est établi entre les cystéines aux positions 74 et 11. Les domaines G-ALPHA1 et G-ALPHA possèdent un site de glycosylation à la position 88. La position 15 des domaines G-BETA



Figure 1.3: Structures 3D d'un MHC-I et d'un MHC-II, et Collier de Perles IMGT de G-DOMAIN. (A) Vue de face du MHC-I (10ga) et du MHC-II (1uvq). (B) Vue latérale du MHC-I (10ga) et du MHC-II (1uvq). (C) Collier de Perles IMGT du G-DOMAIN de MHC-I (10ga) et du MHC-II (1uvq). Dans les représentations Colliers de Perles, les acides aminés sont indiqués par leur abréviation à une lettre. Les cercles hachurés correspondent aux positions manquantes de cette séquence dans la numéritation unique IMGT des G-DOMAINs. Les brins bêta A, B, C et D de chaque domaine sont représentés par des rangés de perles et les hélices alpha apparaissent en haut et en bas de la représentation.

est également un site de glycosylation conservé. Les C-LIKE-DOMAINs sont membres des IgSF, et la numérotation unique IMGT des C-DOMAINs leur est applicable.

1.5 Structures 3D et Colliers de Perles IMGT des V-DOMAIN, C-DOMAIN et G-DOMAIN.

Afin de pouvoir mieux appréhender l'analyse des V-DOMAINS, C-DOMAINS et G-DOMAINS, il est nécessaire d'avoir un aperçu de la structure tridimensionnelle de chaque type de domaine. La comparaison avec leur Collier de Perles IMGT respectif permet de se rendre compte des détails structuraux qui ne peuvent être visualisés dans la représentation simplifiée des Collier de Perles IMGT.

1.5.1 V-DOMAIN

Les caractéristiques du Collier de Perles IMGT de V-DOMAIN ont été décrites dans la section 1.2 page 9. L'examen de la structure du V-DOMAIN de la chaîne 12e8_H (Figure 1.4), montre que le 'brin' A est constitué de deux brins, notés A1 (positions 2 à 7) et A2 (positions 8 à 12), qui ne font pas partie du même feuillet. Le brin A2 n'est pas dans le feuillet indiqué sur le Collier de Perles. Les brins B, D et E font partie du feuillet externe et leur appariement⁴ est bien représenté par le Collier de Perles. La structure du feuillet interne comporte une torsion importante. Seules trois positions du brin C", positions 66, 67 et 68, adoptent une conformation de type brin. Ce brin s'apparie avec le haut du brin C'(positions 53, 54 et 55). Le bas du brin C" (positions 69, 70, 71, 72 et 73) n'adopte pas une structure secondaire de type brin. Le bas du brin C' est apparié avec le brin C. Les brins C, F et G sont appariés comme représenté dans le Collier de Perles.

1.5.2 C-DOMAIN

Les caractéristiques du Collier de Perles IMGT de C-DOMAIN ont été décrites dans la section 1.2 page 9. L'examen du C-DOMAIN de la chaîne 12e8_H (Figure 1.5), montre que les feuillets externes et internes ont une composition inversée par rapport aux V-DOMAIN. Contrairement au V-DOMAIN, les brins A1 et A2 sont simultanément dans le feuillet interne, comme il est indiqué sur le Collier de Perles. Les brins B, D et E font partie du feuillet interne et leur appariement est bien représenté par le Collier de Perles. La structure du feuillet externe est très plane. Seules trois positions du brin C, positions 39, 40, et 41, adoptent une conformation de type brin. Ce brin s'apparie avec le haut du brin F (positions 104,105 et 106). Le bas du brin C (positions 42, 43, 44 et 45) n'adopte pas une structure secondaire de type brin mais forme un tour d'hélice alpha. Le brin transversal CD (positions 45A à 45C) n'adopte pas une conformation brin. Les brins F et G sont appariés comme représentés dans le Collier de Perles.

⁴Le terme *appariement* doit être pris dans le sens d'échanges de liaisons hydrogène entre atomes de la chaîne principale.



Figure 1.4: Comparaison de localisation des positions dans les structures du domaine VH (A,B) et de leur Collier de Perles IMGT (C). (A) Feuillet externe du V-DOMAIN (les brins ABED en jaunes sont situés à l'arrière plan) (B) Feuillet interne du V-DOMAIN (les brins GFCC'C" sont situés à l'avant plan en contacts avec l'autre domaine V). L'exemple est 12E8_H dont la séquence du V-DOMAIN correspond aux allèles murins IGHV14S3*01 (V-REGION) et IGHJ3*01 (J-REGION).



Figure 1.5: Comparaison de localisation des positions dans les structures du domaine CH1 (A,B) et de leur Collier de Perles IMGT (C). (A) Feuillet interne du C-DOMAIN (les brins ABED en jaunes sont situés à l'arrière plan) (B) Feuillet externe du C-DOMAIN (les brins GFC sont situés à l'avant plan). L'exemple est 12e8_H dont la séquence du C-DOMAIN CH1 correspond à l'allèle murins IGHG1*02.

1.5.3 G-DOMAIN

Les caractéristiques du Collier de Perles de G-DOMAIN ont été décrites dans la section 1.4 page 13. L'examen des deux G-DOMAINs de la chaîne loga_A (Figure 1.6) montre une bonne correspondance entre le Collier de Perles et la structure 3D pour les brins A, B et C, et les boucles qui les relient. Le brin D n'adopte pas une conformation de brin. La partie en hélice du domaine G-ALPHA1 se compose d'un segment comportant un tour d'hélice et une partie sans structure secondaire clairement identifiable des positions 51 à 56, et d'un segment constitué d'une longue hélice alpha des positions 56 à 85. Les positions 86 à 90 font le lien avec le domaine G-ALPHA2. La partie en hélice du domaine G-ALPHA2 se compose de deux segments repliés en hélice alpha, l'un de 51 à 61 et l'autre de 63 à 85. Les positions 61A et 62 de G-ALPHA2 relient les deux hélices alpha et sont les positions les plus 'élevées' par rapport au plancher.



Figure 1.6: Comparaison de localisation des positions dans les structures des domaines G-ALPHA1 et G-ALPHA2 (A,B) et de leur Collier de Perles IMGT (C). L'exemple 10gt _A dont les séquences de G-DOMAINs correspondent à l'allèle humain HLA-A*0201.

Chapitre 2 IMGT/3Dstructure-DB.

Les intérêts croisés de la recherche dans le domaine de la santé, de la recherche fondamentale en immunologie et d'applications technologiques des IG (catalyseur, etc...) ont conduit à la détermination expérimentale d'un grand nombre de structures 3D d'IG, de TR et de MHC. Une liste des structures 3D d'IG a été réalisées en 2000 [19] (SACS, http://www.bioinf.org.uk/abs/) mais elle ne propose pas la détection des gènes et allèles correspondant. La base de données AAAA a été réalisée en 2001 par A Honneger [20] (http://www.biochem.uniz.ch/antibody/). Elle fournit une numérotation unique des domaines variables très similaire à celle d'IMGT, mais elle ne donne pas non plus l'identification des gènes et allèles. La base de données IMGT/3Dstructure-DB [21] a été conçue pour fournir à la fois des données génétiques et des analyses standardisées des structures 3D des IG. Elle analyse également les structures 3D des TR, des MHC, et des RPI. IMGT/3Dstructure-DB est la base de données d'IMGT spécialisée dans les structures 3D d'IG, de TR, de MHC et de RPI. IMGT/3Dstructure-DB fournit (1) l'identification des gènes et allèles de ces protéines, (2) l'identification des chaînes de chaque molécule, (3) la délimitation et la description des V-DOMAINS, C-DOMAINS, G-DOMAINS, V-LIKE-DOMAINS, C-LIKE-DOMAINs et G-LIKE-DOMAINs, (4) la numérotation des positions des acides aminés dans la numérotation unique IMGT, (5) les représentations en Collier de Perles IMGT des domaines identifiés, (6) les fichiers de coordonnées atomiques renumérotés, (7) les contacts entre résidus¹ et entre domaines. Les données fournies par IMGT/3Dstructure-DB sont conformes aux concepts d'IDENTIFICATION, de DESCRIPTION et de NUMEROTATION d'IMGT-ONTOLOGY. IMGT/3Dstructure-DB est accessible en ligne à l'adresse http://imgt.cines.fr.

¹La notion de *résidu* est ici prise dans le sens très large d'un regroupement d'atomes liées de façon covalente, ce qui est également convenu dans les fichiers de coordonnées de type PDB. Les résidus sont les monomères des polymères organiques, tel que les acides aminés des protéines, les acides nucléiques des ADN ou ARN, ou les oses des polysaccharides. Ils désigneront également ici les molécules isolées telles que les molécules d'eau, les hèmes ou les ions.

Sommaire

2.1	Orga	anisation de IMGT/3Dstructure-DB	21
	2.1.1	Données d'obtention des structures.	21
	2.1.2	Description des molécules	23
	2.1.3	Caractérisation des positions pour chaque résidu.	24
	2.1.4	Contacts entre résidus et entre domaines.	25
	2.1.5	Séquences des gènes et allèles de références	26
	2.1.6	Termes de DESCRIPTION	26
2.2	Adm	inistration de IMGT/3Dstructure-DB	26
	2.2.1	Sélection des nouvelles entrées.	26
	2.2.2	Interface d'annotation des nouvelles entrées	26
	2.2.3	Etapes d'annotation des nouvelles entrées	29
2.3	Inter	rface utilisateur	33
	2.3.1	Génération des Colliers de Perles IMGT	34
	2.3.2	Liste des séquences de référence IMGT	34
2.4	Cone	$\operatorname{clusion}$	35

2.1 Organisation de IMGT/3Dstructure-DB.

IMGT/3Dstructure-DB est géré par le serveur de base de données MySQL, qui est un serveur de type relationnel, rapide et possédant une large communauté d'utilisateurs. Les tables de la base de données sont organisées en 6 groupes (Figure 2.1), les données d'obtention des structures qui retracent l'historique du fichier de coordonnées, les termes de DESCRIPTION d'IMGT, la description des molécules au niveau génétique et structural, la caractérisation des positions pour chaque résidu, les contacts de chaque résidu et de chaque domaine, et les séquences des allèles de références IMGT.

2.1.1 Données d'obtention des structures.

L'identifiant unique dans IMGT/3Dstructure-DB de chaque fichier de coordonnées est le même que celui de la Protein Data Bank (PDB) [22], la base de donnée généraliste des structures 3D des protéines et des acides nucléiques. Chaque fichier de coordonnées est caractérisé par la technique expérimentale d'obtention (RX, RMN, microscopie électronique), la résolution dans le cas des structures RX, la date de dépôt du fichier dans la PDB et les données bibliographiques (table PDB²). La bibliographie est extraite du fichier de coordonnées (table **reference**) et de la base MMDB [23], *Molecular Modeling Database*, qui fait le lien entre les structures 3D connues et la bibliographie recensée dans la base

 $^{^2 {\}rm La}$ même typographie sera employée pour désigner toutes les tables de la base de données, comme par exemple les tables Chain et Domain.



Figure 2.1: Organisation de la base de données IMGT/3Dstructure-DB. Les principales tables sont organisées en 6 groupes. Les relations entre ces tables sont représentées.

MEDLINE Pubmed au NCBI, the National Center for Biotechnology Information (tables PDBPUBMED, PDBPUBMEDauteur et PDBPUBMEDrefauteur).

2.1.2 Description des molécules.

Les molécules sont identifiées (table Molecule) et décrites par leur composition en chaînes (table Chain), la position et l'identification des régions dans chaque chaîne des IG, des TR, des MHC et des RPI (table Regions) avec l'assignement des gènes et allèles correspondant (table RegionAlleles) ainsi que la position et l'identification des domaines des IG, TR, MHC et RPI (tables Domaines, Module et MHCpep).

Table Molecule.

La table Molecule gère le nom attribué à chaque molécule de chaque fichier de coordonnées. Pour les IG, TR, MHC et RPI, le type du récepteur est précisé ainsi que la description IMGT du fragment protéique.

La détermination du nom des protéines constitue une plus value d'annotation importante car dans le fichier de coordonnées le nom est rarement dans un format correct et n'est pas toujours donné car la littérature associée se révèle parfois insuffisante pour le déterminer. Pour identifier une protéine, il faut parfois avoir recours à un alignement de séquence avec la base de données UniProt [24]. Le nom des peptides contient, en plus du nom de la protéine dont il est issu, les positions dans la séquence de référence, l'identifiant de la référence et les mutations. Pour les protéines IG et TR qui n'ont pas de nom, un nom IMGT est créé. Il comprend le type de fragment suivi d'un tiret et d'un nombre (par exemple Fc-1). Les MHC sont nommés d'après les allèles dont ils sont issus. La détermination de la structure complète d'un récepteur est parfois problématique d'un point de vue technique et il convient souvent de ne s'intéresser qu'à un fragment de ce récepteur. A ce jour, par exemple, une seule structure d'IG humaine complète (code 1hzh) est connue. La description IMGT du récepteur donne un nom standardisé aux fragments de protéines ainsi qu'aux chaînes et aux domaines.

Table Chain.

La table Chain gère simultanément 3 séquences pour chaque chaîne : la séquence des résidus extraite des coordonnées atomiques, la séquence de référence de chaque protéine (fournie par le fichier de coordonnées) et la séquence consensus issue de la comparaison des deux premières. En effet, des différences peuvent apparaître entre la séquence contenue dans les banques de données de séquence et la séquence issue de l'observation de la structure de la protéine. Ces différences peuvent provenir de mutations naturelles mais aussi de la technique d'obtention des coordonnées atomiques. Ainsi la structure 3D de certaines parties de la molécule peut ne pas avoir été déterminée (par exemple, dans le cas de structures RX, pour des raisons de flexibilités ou de désordre cristallin). Par ailleurs l'"absence" de certains atomes peut conduire à la mauvaise identification d'un

résidu. Dans le cas d'une divergence sur le type de résidu, la séquence consensus présente le résidu de la séquence extraite des coordonnées atomiques.

Tables Regions et RegionAllele.

La table **Regions** identifie les régions des IG, TR, MHC et RPI. Elle prend en compte le fait que certaines séquences protéiques codées par un gène unique sont composées de plusieurs domaines, et que la séquence de certains domaines est codée par plusieurs gènes. Cette table caractérise la partie de la séquence protéique correspondant à la plus petite unité de séquence entre REGION et DOMAIN. La table **Regions** identifie donc les V-REGION, D-REGION, J-REGION, C-DOMAIN, C-LIKE-DOMAIN, G-DOMAIN, G-LIKE-DOMAIN et aux V-LIKE-DOMAIN.

La table **RegionAllele** identifie les allèles correspondant à chaque portion de séquence de la table **Regions**. Elle contient les données issues du processus d'identification des allèles (pourcentage d'identité de séquence, score de Smith et Waterman, etc...). La cohérence entre les allèles des régions d'une même chaîne est prise en compte grâce au champ **consensus**. En effet plusieurs allèles peuvent correspondre à la même région mais parmi ces allèles certains peuvent ne pas être cohérents avec les allèles des autres régions. Par exemple, les allèles identifiées pour les différents DOMAINs d'une même REGION doivent être identiques.

Tables Domaine, Module et MHCpep.

La table Domain indique le type de domaine (V-DOMAIN ou V-LIKE-DOMAIN, C-DOMAIN ou C-LIKE-DOMAIN, G-DOMAIN ou G-LIKE-DOMAIN) et pour chaque domaine sa description IMGT, sa position dans les 3 types de séquence (stockées dans la table Chain) et sa séquence avec les gaps IMGT correspondant à la numérotation unique IMGT, ainsi que les insertions qu'il possède par rapport au canevas de base de cette numérotation. Pour les domaines de type immunoglobuline, la composition des feuillets en brin est également enregistrée.

La table Module identifie les domaines appariés de manière structurale dans les récepteurs. Ainsi les deux V-DOMAINs qui constituent le site de reconnaissance de l'antigène forment un V-PARTNER, les deux G-DOMAINs qui forment le site de liaison du peptide un G-PARTNER et les C-DOMAINs en contact par leur feuillet interne un C-PARTNER. La table MHCpep a été créée pour gérer les relations entre le G-PARTNER d'un MHC et le peptide antigénique qu'il contient.

2.1.3 Caractérisation des positions pour chaque résidu.

La table **Res idu** définit chaque position par son appartenance à un domaine, un numéro selon la numérotation IMGT et la nature du résidu. Le concept IMGT de *Residu@position* (R@P) désigne le couple (résidu, position) et est caractérisée par le nom du résidu et sa position IMGT dans un contexte structural particulier. Seules les positions ayant au moins un atome avec des coordonnées 3D sont considérées dans la table **Residu**. Les résidus ne présentant aucune coordonnée atomique apparaissent uniquement dans les séquences de la table Chain.

Une position est définie dans **Residu** par son appartenance à un fichier de coordonnées atomiques, à une chaîne et à un domaine (quand un domaine a pu être identifié comme un V-DOMAIN, un C-DOMAIN, un G-DOMAIN, une région charnière ou une région de liaison). Trois numéros désignent une position, le numéro IMGT correspondant à la numérotation unique IMGT du domaine, le numéro IMGT dans le fichier de coordonnées renuméroté (voir section 2.2.3) et le numéro dans le fichier de coordonnées d'origine avant renumérotation.

Les autres propriétés de chaque position sont le nom du résidu qui occupe cette position (avec une abréviation à 3 lettres et à une lettre), sa surface accessible, et pour les acides aminés la structure secondaire et les angles phi et psi. La correspondance entre les abréviations à 3 lettres et le nom des résidus peut être extraite des fichiers de coordonnées (HeteroAtomes) soit être trouvée dans une table générale de correspondance (HeteroAtomesListe).

2.1.4 Contacts entre résidus et entre domaines.

La fonction des IG, des TR et des MHC repose sur leur capacité de reconnaissance moléculaire et la description des interactions de chaque position est donc très importante. Une autre motivation à l'introduction des contacts dans la base de données est la capacité qu'elle donne d'effectuer des comparaisons entre les interfaces grâce à la numérotation unique des positions (dans le chapitre suivant). Les contacts entre atomes dans chaque fichier de coordonnées sont enregistrés au niveau des contacts entre résidus (Rescontacts) et des contacts entre domaines (Domcontacts). Pour les domaines gérés par IMGT, les contacts sont effectués par un résidu à une position précise dans la séquence et la structure 3D de ce domaine.

Les contacts entre résidus ou entre domaines sont caractérisés par le nombre total de contacts entre atomes ainsi que le nombre de contacts non covalents, de liaisons covalentes, de contacts entre atomes chargés, entre atomes non chargés, le nombre de ponts disulfures, de liaisons hydrogène et de liaisons hydrogène par l'intermédiaire d'une molécule d'eau. Les relations entre les différents types de contacts sont les suivantes : (1) le nombre de contacts total est la somme du nombre de contacts non covalents et de liaisons covalentes, (2) le nombre de contacts non covalents est la somme des contacts chargés et non chargés, (3) le nombre de contacts chargés inclut le nombre de liaisons hydrogène (mais pas les liaisons hydrogène utilisant une molécule d'eau intermédiaire), (4) le nombre de liaisons covalentes inclut le nombre de ponts disulfures.

Chaque type de contacts est détaillé en contacts chaîne principale/chaîne principale, chaîne principale/chaîne latérale, chaîne latérale/chaîne principale, et chaîne latérale/chaîne latérale. La notion de chaîne principale et de chaîne latérale n'est prise en compte que pour les acides aminés. Tous les atomes des autres type de résidus sont considérés comme faisant partie de la chaîne principale.

2.1.5 Séquences des gènes et allèles de références.

Les séquences des allèles de références d'IMGT de chaque espèce (table Espece) sont stockées (table Allele) ainsi que la hiérarchie des gènes (table Gene), sous-groupes (table Subgroup) et groupes (table Groupe) correspondants à l'ontologie IMGT.

Dans la table Allele, les propriétés de chaque allèle sont son nom, sa fonctionnalité et son espèce. Les propriétés d'une séquence d'allèle sont sa partialité (séquençage incomplet), sa taille et le fait qu'elle contienne des gaps en accord avec la numérotation unique IMGT. Les séquences des allèles ont été découpées par domaine avant d'être enregistrées dans la base de données (chaque séquence correspond à une entité de la table **Regions**).

2.1.6 Termes de DESCRIPTION.

Les descriptions IMGT des domaines (MainDomDescr et DomDescr), des chaînes (ChainDescr) et des récepteurs (MainReceptorDescr et ReceptorDescr) sont enregistrées ainsi que leurs relations (Dom_ChainDescr et Chain_ReceptorDescr).

2.2 Administration de IMGT/3Dstructure-DB.

IMGT/3Dstructure-DB est maintenue par un processus semi-automatique. La sélection et une première annotation des nouvelles entrées suivent un processus entièrement automatique. L'ensemble des tables d'annotation des fichiers de coordonnées possèdent une copie temporaire où l'annotation initiale est stockée. Les données sont ensuite validées et éventuellement modifiées avant d'être transférés dans les tables définitives. Les programmes de gestion de IMGT/3Dstructure-DB ont été écrits en langage Perl et en langage C.

2.2.1 Sélection des nouvelles entrées.

Les nouveaux fichiers de coordonnées sont extraits chaque semaine de la PDB. Ils sont sélectionnés par identification de mots clés dans le texte du fichier. Une deuxième sélection détecte les domaines de type V, C ou groove par alignement structural, grâce au programme CE [25], entre les molécules du fichier de coordonnées et des domaines représentatifs des C-DOMAINs d'IG et de TR, des C-LIKE-DOMAINs de B2M, des V-DOMAINs d'IG et de TR et des G-DOMAINs de MHC-I et de MHC-II.

2.2.2 Interface d'annotation des nouvelles entrées.

Le programme d'annotation permet une interaction très souple entre l'annotateur et la base de données ainsi que la gestion automatique des nouvelles entrées (en mode *batch*). Les différentes étapes de l'annotation sont exécutables individuellement (chacune est implémentée dans un module Perl différent), sur une sélection de un ou plusieurs fichiers de coordonnées et sur une sélection de une ou plusieurs chaînes. Pour chaque entrée, les

X quentin@imgtapps.igh.cnrs.fr: /home/quentin	_ – ×
ACT : ABCDEFGHIJKLMNOPQRSTUVW Connection à MySql Connection effectuée. Chargement des tables PDB, PDBsequ Chargement de la table Residu [Ph: Chargement de la table Resendect. Chargement de la table Resennect. Chargement de la table Complex [V: Chargement de la table StrucQuat Chargement de la table Charne.	
Chargement de la table Chainconec Chargement de la table Chainsequer Chargement de la table Regions. Chargement de la table RegionAlle Chargement de la table Domaines. Chargement de la table DomReg. Chargement de la table Domcontacts Chargement de la table Domcontacts Chargement de la table Rescontacts Chargement de la table Rescontacts Chargement de la table Module. 3DDB> sel -c /M\$/ 1 Chain sélectionnées. 3DDB> fi Hc ###### == pour 12E8 == ######	
Détermination des gènes. Méthode Fasta Slow Espece : Mouse Dev Der Frag Deb Fin Score Genes Passe 0 0 118 IGKV 1 96 634 IGKV6-13*01(Mouse), Passe 1 0 107 IGKJ 97 107 67 IGKJ5*01(Mouse), Passe 2 0 0 IGKC 108 214 698 IGKC*01(Mouse),IGKC*	:02(Mouse
Au total on a 3 fragments Initial : DIVMTQSQKFMSTSVGDRVSITCKASQNVGTAVAWYQQKPGQSPKLMIYSASNRYTGVPDRFTGSQ FTLTISNMQSEDLADYFCQQYSSYPLTFGAGTKLELKRADAAPTVSIFPPSSEQLTSGGASVVCFL PKDINVKWKIDGSERQNGVLNSATDQDSKDSTYSMSSTLTLTKDEYERHNSYTCEATHKTSTSPIN RNEC IGKV : DIVMTQSQKFMSTSVGDRVSITCKASQNVGTAVAWYQQKPGQSPKLMIYSASNRYTGVPDRFTGSQ FTLTISNMQSEDLADYFCQQYSSYPL IGKJ : TFGAGTKLELK IGKC : RADAAPTVSIFPPSSEQLTSGGASVVCFL	SGTD NNFY KSFN SGTD
PKDINVKWKIDGSERUNGVLNSATDUDSKDSTYSMSSTLTLTKDEYERHNSYTCEATHKTSTSPIN RNEC	KSFN

Figure 2.2: Interface texte d'annotation de IMGT/3Dstructure-DB. La capture d'écran présente une visualisation 3D du fichier de coordonnées 12e8, avec chaque chaîne colorée différemment et les molécules d'eau représentées par des octaèdres blancs. Les régions et allèles de la chaîne 12e8_M ont été identifiées et les séquences correspondantes sont présentées dans des couleurs différentes.

Table 2.1: Modules Perl utilisés par l'interface d'annotation de IMGT/3D structure-DB.

Modules Perl	Fonctions		
PDB.pm	Objet PDB : données d'obtention, de séquences, données sur les positions et des contacts entre positions. Ex- traction des données du fichier de coordonnées et créa- tion d'un objet PDB. Détermination des structures se- condaires. Recherche des références bibliographiques.		
Chain.pm	Objet Chain : définition des régions, des domaines et des contacts entre domaines. Détermination des allèles consensus. Définition des domaines à partir des régions et établissement de la numérotation unique de chaque domaine.		
Moulinette.pm	Gère les différentes étapes de l'annotation et leurs dé- pendances. Recherche des régions et allèles sur une sé- quence.		
Mysql.pm	Echange des données entre les objets Chain et PDB avec IMGT/3Dstructure-DB.		
Aff3D.pm	Affichage 3D (OpenGL) d'un domaine, d'une chaîne, d'une structure 3D, avec colorisation des régions et des CDR.		
Chain_Gaps	Alignement de Smith et Waterman. Alignement de Smith et Waterman modifié entre une séquence de réfé- rence IMGT et une séquence de domaine.		
Contacts.pm	Détermination des contacts entre positions. Identifica- tion et somme des contacts au niveau des interfaces. Dé- termination de la composition des feuillets des domaines de type immunoglobuline.		
Frag.pm	Identification des chaînes appartenant à une même mo- lécule. Description IMGT des domaines, chaînes et frag- ment de récepteurs d'antigènes.		
MetaChain.pm	Différences entre la séquence issue des coordonnées ato- miques et la séquence de référence des protéines. Créa- tion des objets Chain à partir des informations de sé- quence des objets PDB (prise en compte des liaisons co- valentes et des distances entre résidus consécutifs).		
Renum.pm	Renumérotation des positions d'une chaîne. Renuméro- tation d'un fichier de coordonnées.		

données peuvent être chargées et sauvées dans la base de données table par table, dans les tables temporaires ou définitives. Les paramètres de chaque étape d'annotation sont modifiables interactivement. Une visualisation des structures tridimensionnelles et individuellement des chaînes et des domaines permet d'observer le positionnement des régions ou des CDR des V-DOMAINS. Le programme d'annotation a été implémenté en langage Perl. En effet ce langage se prête particulièrement bien au traitement de fichiers textes, ce qui constitue la majeur partie des opérations effectuées. Les données sont organisées dans deux classes d'objet, la classe PDB caractérisant chaque fichier de coordonnées et la classe Chain caractérisant chaque chaîne. Les fonctions ont été regroupées dans différents modules Perl afin de faciliter le maintien du code source et de pouvoir les réutiliser indépendemment dans d'autres programmes (Table 2.1).

2.2.3 Etapes d'annotation des nouvelles entrées.

Lecture du fichier de coordonnées atomiques.

(A) Des expressions régulières permettent d'extraire les coordonnées atomiques de chaque résidu ainsi que la séquence fournie comme référence. Les informations d'obtention (technique expérimentale et référence bibliographique) ainsi que les noms et formules des hétéroatomes sont également extraits.

Définition des chaînes.

 (\mathbf{B}) Les chaînes sont initialement définies à partir des noms de chaînes données dans les coordonnées des atomes.

(C) Les chaînes peuvent être éliminées, renommées ou crées interactivement. L'élimination d'une chaîne intervient par exemple dans le traitement de certains fichiers de coordonnées où un même nom de chaîne est attribué à l'ensemble des molécules d'eau, ce qui n'est pas correct puisqu'il n'existe pas de liaisons covalentes entre elles. La création d'une chaîne intervient par exemple dans le cas où deux chaînes partenaires d'une IG ont été clairement confondues au sein d'un même identifiant de chaîne, et qu'il est nécessaire de les individualiser. Il est aussi parfois utile de créer un nom de chaîne pour un élément chimique isolé, jouant un rôle important (un antigène par exemple), afin de pouvoir le mettre en avant par rapport aux molécules d'eau et aux ions. Les chaînes nouvellement créées sont identifiées par un chiffre (l'identifiant classique est une lettre).

(**D**) Pour chaque chaîne la séquence de résidus dans la partie des coordonnées atomiques est extraite ainsi que la séquence dans la partie des séquences originelles du fichier de coordonnées atomiques. La séquence consensus entre ces deux séquences est déterminée grâce à un alignement de Smith et Waterman [26] et aux règles énoncées dans la partie Chaînes plus haut. (**E**) Les espèces de chaque chaîne sont extraites du fichier de coordonnées grâce à un dictionnaire de correspondance entre les différents noms attribuables à une même espèce, le nom latin, le nom commun ou une partie de ces noms (*Camelus dromedarius*, arabian camel, camel). Table 2.2: Etapes de l'annotation de IMGT/3D structure
DB. Les différentes étapes, désignées par une lettre, sont utilisées dans l'interface d'annotation.

Etapes
Extraction du fichier de coordonnées des informations d'obtentions.
Extraction du fichier de coordonnées des séquences.
Détermination, délétion et création de définitions de chaînes.
Traitement des différences entre la séquence de référence (champs SE QRES) et de la séquence issue des coordonnées.
Extraction de l'espèce de chaque chaîne avec un dictionnaire d'équiva lences entre les différentes appelations.
Détermination des régions par alignement avec les séquences de référenc
Détermination des régions définies implicitements.
Trouver les position des régions dans la séquence issue des coordonnée
Détermination des consensus entre les allèles des régions appartenant la même chaîne.
Détermination du type de récepteur.
Détermination des domaines.
Trouver la position des domaines dans la séquence issue des coordonnées
Détermination des gaps des domaines par alignement avec la séquen- avec gaps de l'allèle de référence.
Détermination des structures secondaires et des angles diedres de chaîne principale.
Renumérotation selon la numérotation unique IMGT de chaque positio
Détermination des contacts et de la composition des feuillets.
Détermination des chaînes appartenant à la même molécule grâce au nombre de contacts.
Détermination des domaines partenaires.
Détermination des complexes.
Détermination de la DESCRIPTION IMGT des domaines, chaînes récepteurs d'antigènes.
Renumérotation des différents champs du fichier de coordonnées.
Identification des régions.

(F) L'identification des gènes et allèles et la délimitation des régions (V-REGION, J-REGION, C-DOMAIN, C-LIKE-DOMAIN, V-LIKE-DOMAIN, G-DOMAIN et G-LIKE-DOMAIN) sont obtenues par alignement, avec le programme FASTA [27], des séquences consensus des chaînes avec le jeu de séquences de références IMGT stocké dans la base de données. (G) Certaines régions sont identifiées implicitement à partir de celles qui viennent d'être positionnées par alignement. Les régions correspondant aux D-REGION³ sont entre une V-REGION et une J-REGION. De même les régions charnières (*Hinge*) sont identifiées entre les deux premiers C-DOMAINs d'une chaîne lourde d'IG. La région de liaison entre les deux V-DOMAINs d'une construction non naturelle scFv est notée comme *Linker*.

 (\mathbf{H}) La position des régions dans la séquence issue des coordonnées atomiques est calculée en prenant en compte les délétions par rapport à la séquence consensus (voir description de la table Chain page 23).

 (\mathbf{I}) Pour les régions consécutives codées par des allèles appartenant au même groupe, les allèles qui sont partagées entre ces régions sont marquées comme 'consensus'.

(J) Le type de protéine (IG, TR, MHC ou RPI) est attribué à chaque chaîne.

Identification des domaines.

(K) Les domaines sont délimités à partir des régions, les V-DOMAINs correspondant à la combinaison des V-D-J-REGION ou des V-J-REGION, et les autres domaines correspondant exactement à une région. La description IMGT de ces domaines est déterminée grâce aux noms des gènes qui les constituent. Suivant la composition en domaines de chaque chaîne, ainsi que la description IMGT de chaque domaine, la description IMGT des chaînes est obtenue par recherche dans les tables de DESCRIPTION.

(L) La position des domaines dans la séquence issue des coordonnées atomiques est calculée en prenant en compte les délétions par rapport à la séquence consensus (voir description de la table Chain page 23).

Renumérotation des positions.

(M) La numérotation des domaines, suivant la numérotation unique IMGT, est réalisée en alignant la séquence de chaque domaine avec la séquence de l'allèle de référence IMGT correspondant. Pour les V-DOMAIN la séquence de référence identifiant la V-REGION est utilisée. L'alignement est effectué par un algorithme de Smith et Waterman modifié qui considère le gap IMGT comme un acide aminé à part entière et qui défavorise la création de gaps dans la séquence de référence (dissymétrie de la gestion des insertions entre les deux séquences). Après alignement, le programme replace les insertions IMGT dans les régions de boucles, en accord avec la numérotation unique de chaque domaine.

 $^{^{3}}$ Les gènes et allèles qui correspondent aux D-REGION ne sont pas clairement identifiables au niveau protéique car elles sont très petites, les allèles de D-GENE sont très nombreux, et le processus de réarrangement imprécis rend leur délimitation très imprécise.

Les insertions de la séquence par rapport au canevas de base de la numérotation unique et les insertions supplémentaires introduites lors de l'alignement sont sauvegardées pour chaque domaine.

 (\mathbf{N}) La structure secondaire de chaque position est identifiée grâce au programme Stride [28]. (\mathbf{O}) La numérotation des positions de chaque domaine est déterminée à partir de la séquence du domaine avec les gaps IMGT, de la numérotation unique IMGT du canevas de base du type de domaine correspondant et les insertions à ce canevas de base du domaine considéré. Les régions hors des domaines sont renumérotées ordinalement.

Contacts atomiques entre résidus et domaines.

 (\mathbf{P}) Les contacts inter atomiques sont déterminés par un programme écrit en C qui rassemble les résultats au niveau de chaque résidu. J'ai écrit un programme d'analyse de contacts inter atomiques qui distingue : les contacts non covalents et les liaisons covalentes hors chaîne protéique et nucléotidique (dont les ponts disulfures), les contacts polaires et non polaires parmi les contacts non covalents (impliquant des atomes polaires) et les liaisons hydrogène parmi les contacts polaires. Par extension les liaisons hydrogène par l'intermédiaire d'une molécule d'eau sont également mesurées quand les coordonnées atomiques des molécules d'eau sont disponibles.

Le programme utilise une distance seuil de contact égale à la somme des rayons de van der Waals et du diamètre moyen d'une molécule d'eau (au total environ 5Å). Les liaisons covalentes sont identifiées par une distance seuil égale à la somme des rayons de van der Waals. Les contacts polaires sont les contacts entre deux atomes polaires.

Un contact polaire est considéré comme une liaison hydrogène quand il vérifie les conditions suivantes : (1) l'un des atomes est potentiellement donneur de liaison hydrogène et l'autre accepteur, (2) la distance entre les atomes est de moins de 3.2Å, (3) l'angle aigu entre le carbone connecté à l'atome donneur, l'atome d'hydrogène et l'atome accepteur est de plus de 140 degrés, et (4) l'angle aigu entre le carbone lié à l'atome accepteur, l'atome accepteur et l'atome donneur est de plus de 90 degrés (les valeurs sont issues de Proteins 2nd Ed., TE Creighton [29]). Deux cas de figure se présentent suivant la possibilité de positionner l'atome d'hydrogène sans ambiguïté grâce aux atomes dont les coordonnées sont connues. Cela concerne les hydrogènes reliés aux azotes des liaisons peptidiques et à l'azote en position epsilon des tryptophane. Dans ce cas l'atome d'hydrogène est construit explicitement, et les angles et distances sont directement évalués. Dans le cas contraire, on suppose la planéité des atomes donneurs, accepteurs, du carbone relié à l'atome donneur et de l'atome d'hydrogène (cas le plus favorable). Des relations simples de géométries permettent de calculer les angles qui nous intéressent sans calculer explicitement la position de l'hydrogène. Cette méthode permet d'identifier les liaisons hydrogène partagées transitoirement entre plus de deux partenaires. Des programmes comme HBPLUS [30] reconstruisent tous les hydrogènes explicitement ce qui ne permet pas de considérer les liaisons hydrogène transitoires utilisant plusieurs atomes accepteurs. Le programme comptabilise les contacts pour chaque résidu. Ces informations sont entrées dans IMGT/3Dstructure-DB et servent à caractériser les interfaces et à identifier les chaînes appartenant au même récepteur et au même complexe.

Le nombre de contacts entre les domaines et les chaînes (pour les ligands sans domaines identifiés) est calculé. Pour les domaines de type immunoglobuline, les liaisons hydrogène entre atomes de la chaîne principale servent à identifier la répartition des brins entre les deux feuillets.

Structures quaternaires et partenaires.

 (\mathbf{Q}) Les chaînes appartenant à une même molécule sont identifiées à partir des analyses de contacts entre domaines (une valeur seuil sur le nombre de contacts est utilisée).

 (\mathbf{R}) Les domaines partenaires sont identifiés dans les IG, TR et MHC grâce à une valeur seuil sur le nombre de contacts entre domaines.

 (\mathbf{S}) Les complexes cristallographiques sont détectés par regroupement des molécules grâce à une valeur seuil sur le nombre de contacts.

(**T**) La composition en chaîne, ainsi que la description IMGT de chaque chaîne, permet de déterminer la description IMGT des IG, TR, MHC et RPI par recherche dans les tables de DESCRIPTION.

Renumérotation des fichiers de coordonnées.

(U) Un fichier de coordonnées renuméroté est créé en utilisant pour chaque position un numéro correspondant à son numéro IMGT dans le domaine. La numérotation IMGT des fichiers de coordonnées suit les règles suivantes, (1) à chaque passage à un nouveau domaine domaine ou une région charnière la numérotation passe aux milliers suivants, (2) la numérotation dans les domaines de type V-DOMAIN, C-DOMAIN et G-DOMAIN suit la numérotation unique des domaines pour les chiffres des centaines, des dizaines et des unités (par exemple le numéro 104 désigne la position 104 du premier domaine et le numéro 2024 désigne la position 24 dans la numérotation unique IMGT du troisième domaine), (3) les numéros d'insertions avec une étiquette alphabétique sont enregistrée en utilisant le champ du fichier de coordonnée correspondant aux insertions de résidus (colonne 27 d'après la description du format PDB), (4) les insertions avec une étiquette numérique sont traduites en étiquette alphabétique (la position '112.1' devient la position '112A'). Tous les champs du fichier de coordonnées sont renumérotées, en dehors des champs REMARK dont le format et la syntaxe n'est pas standardisé et difficilement modifiable automatiquement. Le fichier renuméroté contient à la REMARK 410 les informations de noms de protéines, de définitions des structures quaternaires, de domaines et de régions (avec gène et allèle).

2.3 Interface utilisateur.

L'interface Internet utilisateurs a été décrite dans deux publications, Kaas *et al.* 2004 [21] (Publication 3) et Kaas et Lefranc 2005 [31] (Publication 4), qui peuvent être trouvées dans la partie Publications. Deux éléments qui ne sont pas détaillées dans ces publications vont être décrites ici, la génération dynamique des Colliers de Perles IMGT et l'alignement des séquences de références IMGT sous forme de Protein Display IMGT.

2.3.1 Génération des Colliers de Perles IMGT.

Le programme de génération des Colliers de Perles a été implémenté dans un module Perl. Un dessin au format PostScript est généré, puis converti au format PNG par le programme *convert* de la suite ImageMagick.

Le module comporte des routines dessinant les éléments de base de chaque collier, les brins, les boucles, les hélices et les traits de liaisons. Un canevas spécifique de chaque collier enchaîne les différents éléments de base pour produire le dessin final. La génération du dessin PostScript ne s'effectue qu'à la fin du processus afin de pouvoir géré les chevauchements entre les éléments dans les Colliers de Perles sur deux plans. Les fonctions de dessin des éléments de base tiennent compte des insertions au canevas de base et le positionnement des perles des insertions s'effectue suivant le contexte (par exemple l'orientation du brin, le sommet d'une boucle etc...). Pour ce faire, chaque perle est caractérisée par un numéro ordinal et par son numéro IMGT, et le tracé de chaque élément se fait suivant une direction angulaire. L'exemple le plus visible de cette notion angulaire est le tracé du plancher des G-DOMAINs. Les paramètres passés à la routine principale du tracé permettent de modifier la couleur de certaines perles, d'ajouter des étiquettes à des perles et de tracé des liaisons entre des perles (par exemple pour représenter les liaisons hydrogène).

Les Colliers de Perles IMGT utilisés dans IMGT/3Dstructure-DB sont générés dynamiquement pour chaque domaine. Les dessins sont stockés deux jours sur le serveur puis effacés. La création du fichier PostScript est très rapide et l'étape limitante est la création du fichier au format PNG (moins de 1 seconde en général sur le serveur).

Une interface internet, accessible depuis la page principale de IMGT/3Dstructure-DB, permet de tracer les Colliers de Perles IMGT des V-DOMAINs, C-DOMAINs, et G-DOMAINs à partir d'une séquence d'utilisateur. Cette interface ne génère pas les gaps dans la séquence et ils doivent être préalablement insérés dans la séquence par l'utilisateur. Outre la réalisation de graphiques esthétiques cet outil sert à vérifier que les gaps insérés dans une séquence sont cohérents avec les structures secondaires présentées par le Collier de Perles IMGT. Cette interface permet de choisir le type de coloriage des CDR-IMGT des V-DOMAINs (domaines de chaîne lourde ou légère), le nombre de plan (1 ou 2), la taille du CDR3 pour les V-DOMAINs et d'ajouter des insertions au caneva du Collier de Perles IMGT.

Finalement, la structure du module Perl permet d'ajouter facilement de nouveaux canevas. L'implémentation aisée du caneva permettant de générer des Collier de Perles de knottines, une famille structurale de miniprotéines à trois ponts disulfures qui forment une structure de noeud [32] en est la preuve.

2.3.2 Liste des séquences de référence IMGT

Une interface accessible depuis la page de garde de IMGT/3Dstructure-DB permet d'avoir accès aux séquences d'allèles des tables de références de IMGT/3Dstructure-DB. Cette interface permet de sélectionner les allèles suivant l'espèce, le nom des gènes, des groupes, des sous-groupes et des allèles et le numéro du domaine sur le gène. Plusieurs cribles sur les noms de gènes, de groupes, sous-groupes et allèles peuvent être entrés simultanément dans des champs textes. Les séquences des allèles sont alignées grâce à leur numérotation unique IMGT et aux insertions particulières de chaque allèle. L'affichage de l'alignement se conforme à celui des pages 'Protein displays' (visible dans 'IMGT répertoire' depuis la page de garde de IMGT), où certaines parties de la séquence sont mises en avant par des séparations (des 'blancs') et des changements de couleurs.

2.4 Conclusion

Ma contribution à IMGT/3Dstructure-DB s'est traduite par sa réorganisation totale, par l'écriture de nouveaux scripts d'annotation suffisamment souples pour s'adapter aux différents cas rencontrés et par l'amélioration et le développement de l'interface Internet. Tout ces développements ont constitué la majeur partie de mon travail de thèse. Initialement, la base de données gérait l'identification des gènes et allèles et la numérotation des V-DOMAIN des récepteurs de type Fab et IG. La réorganisation que j'ai effectuée permet désormais de prendre en compte plusieurs domaines dans une même chaîne, des types de domaines différents (V-DOMAIN, C-DOMAIN, G-DOMAIN) et des types de récepteurs différents dont les définitions sont également des données de tables. Les positions sont maintenant décrites individuellement. J'ai introduit la notion de contacts à travers l'espace entre résidus appartenant à un même domaine ou à des domaines, chaînes, récepteurs différents. Ces contacts sont des données importantes pour la caractérisation structurale des interfaces immunologiques.

Parallèlement à la nouvelle organisation des tables, de nouveaux scripts permettant de gérer les différents récepteurs et domaines ont été nécessaires. La version actuelle représente environ 30000 lignes de code et a été l'objet d'un remaniement constant. Par exemple, j'ai écrit un nouveau script de renumérotation qui est basé sur un alignement de Smith et Waterman, utilisant les insertions IMGT comme un nouvel acide aminé et pénalisant fortement l'ouverture de gaps dans la séquence de référence. Ce script me permet d'annoter des séquences avant subit des mutations ou des insertions, ou des séquences dont on ne possède pas encore d'allèles de référence, ce qui était auparavant impossible. J'ai créé un environnement d'annotation qui permet une grande souplesse de sélection de fichiers de coordonnées ou de chaînes, auxquels ont peut appliquer tout ou partie des étapes d'annotation dont les différents paramètres sont ajustables. Cette interface était d'autant plus utile qu'étant le seul annotateur de IMGT/3Dstructure-DB elle a permis un gain de temps indispensable, une fois réalisée. Ce travail d'annotateur consiste essentiellement à déterminer les noms des molécules par recherche bibliographique et également à vérifier que les étapes automatiques se soient correctement déroulées. La gestion des séquences de référence est également à ma charge quand les séquences de références d'une espèce donnée ne sont pas disponibles sur IMGT.

J'ai repris le développement de l'interface Internet en y ajoutant de nouvelles requêtes. Ainsi l'outil IMGT/StructuralQuery permet d'interroger chaque domaine et chaque position d'un domaine. Le script de recherche génère maintenant une requête SQL spécifique des informations demandées. Ce script unifie les recherches, qui étaient auparavant effectuées par des scripts indépendants, ce qui permet notamment de gérer des requêtes effectuées sous forme d'une phrase dans IMGT-StructuralQuery. Les fiches de chaque fichier de coordonnées présentent maintenant l'identification des domaines et des régions sur les séquences, donnent accès à des outils de visualisation et aux références bibliographiques, et proposent les analyses de contacts sous forme de tables. Le nouveau programme de Collier de Perles que j'ai écrit est le successeur d'un programme PostScript qui gérait les V-DOMAIN sans possibilité d'insertions. Le nouveau programme dessine également les C-DOMAINs et les G-DOMAINs, et il permet de colorier des perles en une ou plusieurs couleurs.

Tous ces efforts de standardisation des données de IMGT/3Dstructure-DB permettent maintenant d'offrir une description claire et précise de chaque récepteur d'antigène, ce qui rend possible en particulier, les analyses statistiques automatiques de leurs structures 3D.

Chapitre 3

Analyse structurale des récepteurs d'antigènes dans IMGT.

La standardisation des données structurales employées dans IMGT/3Dstructure-DB, présentée dans le chapitre précédent, a pour objectif de faciliter les comparaisons des structures 3D des récepteurs d'antigènes. Plus précisément, les numérotations uniques IMGT des V-DOMAINS, C-DOMAINS, C-LIKE-DOMAINS et G-DOMAINS ainsi que les termes de DESCRIPTION sont utilisés dans ce chapitre afin de comparer les structures 3D des molécules des IG, des TR et des MHC, ainsi que de leurs interfaces. Toutes les analyses présentées sont automatisées et la mise à jour des tables et figures est disponible sur le site de IMGT, http://imgt.cines.fr. Le nombre de structures 3D de récepteurs d'antigènes déterminées expérimentalement est bien inférieur au nombre d'allèles et même de gènes connus aujourd'hui. L'interprétation des mesures présentées dans ce chapitre doit donc toujours tenir compte de ce déséquilibre important. Elles font apparaître le comportement structural moyen des récepteurs d'antigènes dans l'ensemble des données accessibles ce qui permet d'identifier et de justifier un comportement atypique dans un cas particulier. Dans un contexte plus global, les projets de protéomique structurale vont générer une masse de données structurales toujours plus importante. La standardisation des données structurales apparaît comme un outil essentiel qui permet, comme ici, d'effectuer des analyses à grande échelle des protéines apparentées.

Les effectifs de structures 3D des récepteurs d'antigènes et de leurs complexes sont présentés, ainsi que la représentativité des gènes des IG, TR et MHC humains et de souris. Les trois types de domaines, V-DOMAIN, C-DOMAIN (et C-LIKE-DOMAIN) et G-DOMAIN sont analysés au niveau de leur variation de séquence dans les structures 3D déterminées expérimentalement. La variation de la composition des feuillets des V-DOMAINs et C-DOMAINs est observée. Pour tous les domaines, les variabilités conformationnelles et les variations des interactions atomiques internes sont calculées. Les assemblages des domaines au sein de chaque type de récepteur d'antigène sont analysés, au niveau de l'arrangement des domaines dans les V-PARTNER et C-PARTNER, puis dans l'arrangement des V-PARTNER et C-PARTNER dans le récepteur. Finalement les interfaces entre les récepteurs et l'antigène sont étudiées, à savoir : l'interface IG/antigène, l'interface peptide/MHC et l'interface TR/peptide/MHC [33] (Publication 6). Ces deux dernières interfaces sont particulièrement intéressantes car elles font intervenir la numérotation unique dans chacune des protéines du complexe. Ces analyses sont des exemples simples d'utilisation de IMGT/3Dstructure-DB pour l'étude des récepteurs d'antigènes, beaucoup d'autres mesures et analyses statistiques sont possibles. Par exemple, la nature de l'antigène n'a pas été considérée comme cela été le cas dans d'autres études [34] [35].

Sommaire

3.1	Non	nbre de structures 3D dans IMGT/3Dstructure-DB	39
	3.1.1	Nombre de structures 3D d'IG, de TR et de MHC	39
	3.1.2	Structures de complexes IG/antigène, pMHC et TR/pMHC	39
	3.1.3	Représentativité génétique	39
3.2	Ana	lyse des domaines.	4(
	3.2.1	Longueur des séquences	42
	3.2.2	Composition en acides aminés.	4
	3.2.3	Distribution des brins dans les feuillets.	4
	3.2.4	Angles entre feuillets des V-DOMAINs et C-DOMAINs	4
	3.2.5	Alignements structuraux.	4
	3.2.6	Analyse de contacts	5
3.3	Org	anisation des domaines dans les récepteurs	5
	3.3.1	V-PARTNER	5
	3.3.2	C-PARTNER.	6
	3.3.3	Entre le V-PARTNER et le C-PARTNER	6
3.4	Inte	rface des complexes IG/antigène.	6
	3.4.1	Superposition des CDR-IMGT.	6
	3.4.2	Positions en contacts.	6
3.5	Inte	rface des complexes peptide/MHC	6
	3.5.1	Flexibilité du peptide dans le groove	7
	3.5.2	$Interaction \ peptide/MHC \ \ \ldots $	7
	3.5.3	Sites de contacts IMGT	7
3.6	Inte	${f rface\ des\ complexes\ TR/peptide/MHC\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\$	7
	3.6.1	Orientation du TR	7
	3.6.2	Surface enfouie	8
	3.6.3	Positions en interactions	8
3.7	Con	clusion	8

3.1 Nombre de structures 3D dans IMGT/3Dstructure-DB.

3.1.1 Nombre de structures 3D d'IG, de TR et de MHC.

En octobre 2005, IMGT/3Dstructure-DB recense 1022 fichiers de coordonnées, qui contiennent 683 structures d'IG, 40 structures de TR et 237 structures de MHC. Le nombre de structures 3D présentant un TR est faible car la production en laboratoire de TR soluble reste toujours difficile [36]. Ainsi seules deux structures de TR-GAMMA_DELTA sont aujourd'hui disponibles, une de V-DELTA publiée en 1998 [37], et une structure d'un récepteur complet TR-GAMMA_DELTA publié en 2001 [38] (1hxm). Par contre pour les MHC-I, le développement d'un protocole spécifique de cristallisation en 1992 [39] [40] a permis d'accroître rapidement leur nombre. Les MHC-II sont plus difficiles à cristalliser, et une construction non naturelle, le MHC-II_ALPHA-BETA-PEPTIDE, a été employée pour stabiliser le complexe avec un peptide (1iea et 1ieb [41]). Cette construction comprend une chaîne II-ALPHA, une chaîne II-BETA avec laquelle le peptide est lié de façon covalente grâce à une liaison polypeptidique flexible.

3.1.2 Structures de complexes IG/antigène, pMHC et TR/pMHC.

370 structures 3D présentent un complexe entre une IG et un antigène ce qui constitue 54% des structures contenant une IG. La liaison IG/antigène est en effet très étudiée non seulement pour son rôle immunologique mais aussi pour ses applications technologiques en tant que catalyseur de réactions (ou l'antigène est un intermédiaire réactionnel) et comme stabilisateur de complexe cristallin lors de la production de cristaux pour la diffraction des rayons X. Dans IMGT/3Dstructure-DB, seulement 21 structures de TR/pMHC sont disponibles, ce qui est peu en comparaison des 229 structures 3D de pMHC connues. Cela tient à une faible stabilité des complexes TR/pMHC entraînant une cristallisation difficile. Aucun protocole général de cristallisation pour les TR/pMHC n'a pu être mis au point à ce jour aucune structure de complexe TR/pMHC impliquant un TR-GAMMA_DELTA n'a été déterminée.

3.1.3 Représentativité génétique.

Le Tableau 3.1 montre la représentativité des gènes d'IG, de TR et de MHC dont la structure 3D est connue. Le nombre de structures 3D pour chaque gène IGHV humain est présenté sur la Figure 3.1. Le nombre de structures 3D pour les autres groupes et les autres espèces est disponible en ligne sur les pages de statistiques de IMGT/3Dstructure-DB. Pour les TR, le nombre de structures est faible et le nombre de gènes chez l'homme et la souris est important (47 et 89 gènes respectivement). La plupart des gènes de MHC ont plusieurs structures 3D représentatives mais le nombre de gènes de MHC, par comparaison aux différents gènes d'IG et de TR, est faible, et la diversité génétique des MHC est essentiellement allélique. Le nombre de structures 3D de MHC est loin de refléter la

diversité allélique des MHC (51 allèles sur les 252 allèles de références humains différents au niveau protéique). Le déséquilibre important entre le nombre de structures expérimentales et la diversité génétique ne permet pas de conclure avec certitude que la variabilité structurale observée soit générale. Elle retranscrit uniquement l'état des connaissances actuelles que nous pouvons observer. Par contre la standardisation IMGT permet la mise en place d'une gestion de ces structures 3D avant l'arrivée massive de des données expérimentales, comme le prouve la caractérisation des domaines (section 3.2), des récepteurs (section 3.3) et de leurs complexes avec un antigène (sections 3.4, 3.5 et 3.6).



Figure 3.1: Nombre de structures 3D dans IMGT/3Dstructure-DB pour chaque gène IGHV humain.

3.2 Analyse des domaines.

Afin de comparer les domaines gérés par IMGT/3Dstructure-DB, les domaines présentant moins de 90% d'identité de séquence pour les V-DOMAINs et C-DOMAINs et 98% d'identité de séquence pour les G-DOMAINs, avec la meilleure résolution cristallographique possible (les structures RX sont préférées aux structures RMN) ont été sélectionnés. La constitution des groupes de structures suivant leur similarité de séquence a été effectuée par la procédure suivante : le pourcentage d'identité de séquence entre tous les V-DOMAINs, tous les C-DOMAINs et tous les G-DOMAINs a été déterminé, et l'algorithme de regroupement hclust, utilisant la méthode *complete*, de l'environnement statistique R (http://www.r-project.org) a été utilisé pour constituer les groupes de similarité de séquence. Les domaines présentant plus de 5 positions sans coordonnées atomiques ont été écartés. La structure de l'IG NewM (7fab) a été écartée car elle possède une délétion du brin C". La structure de la protéine R9 (1sjv) a également été écartée car elle présente une modification importante de sa structure due à l'assemblage cristallin (formation de feuillet béta entre unités cristallines).

IMGT/3Dstructure-DB comporte 473 V-DOMAINS et 77 C-DOMAINS ou C-LIKE-DOMAINS de séquences différentes avec un seuil de 90% d'identité de séquence, et 89 G-DOMAINS de séquences différentes avec un seuil de 98% d'identité de séquence. La

Table 3.1: Nombre de gènes fonctionnel	ls IMGT pour	Homo sapiens e	et Mus musculus	[2] $[1]$ et
nombre de gènes possédant au moins un	ne structure 31	D représentative	dans IMGT/3Ds $$	structure-
DB.				

Homo sapier	is		$Mus\ musculus$					
Groupe	Gènes	Structures	Groupe	Gènes	Structures			
IGHV	56	26	IGHV	216	63			
IGKV	51	28	IGKV	120	53			
IGLV	39	12	IGLV	8	1			
IGHJ	6	5	IGHJ	4	4			
IGKJ	5	5	IGKJ	5	4			
IGLJ	7	5	IGLJ	4	3			
IGHC	10	7	IGHC	9	6			
IGKC	1	1	IGKC	1	1			
IGLC	5	3	IGLC	3	1			
TRAV	47	6	TRAV	89	10			
TRBV	66	5	TRBV	26	4			
TRGV	9	1	TRGV	7	0			
TRDV	3	2	TRDV	17	0			
TRAJ	58	6	TRAJ	50	9			
TRBJ	13	4	TRBJ	13	5			
TRGJ	5	1	TRGJ	4	0			
TRDJ	4	1	TRDJ	2	0			
TRAC	1	1	TRAC	1	1			
TRBC	2	2	TRBC	2	2			
TRGC	2	1	TRGC	4	0			
TRDC	1	1	TRDC	1	0			
MHC-Ia	3	3	MHC-Ia	3	3			
MHC-Ib	2	2	MHC-Ib	4	4			
MHC-IIa-A	3	2	MHC-IIa-A	2	1			
MHC-IIa-B	6	3	MHC-IIa-B	2	2			
MHC-IIb-A	2	1	MHC-IIb-A	2	1			
MHC-IIb-B	2	1	MHC-IIb-B	3	1			
B2M	1	1	B2M	1	1			

liste des domaines sélectionnées dans IMGT/3Dstructure-DB est fournie en annexe (page 139).

3.2.1 Longueur des séquences.

Dans le jeu de données sélectionnés, les V-DOMAINs ont une séquence comprise entre 103 et 130 acides aminés, les C-DOMAINs entre 85 et 129 acides aminés et les G-DOMAINs entre 81 et 96 acides aminés (Figure 3.2). La variabilité des longueurs des V-DOMAINs s'explique par les grandes différences de longueur de CDR possibles. Pour les V-DOMAINs des IG et des TR confondu, le CDR1-IMGT à une longueur comprise entre 5 et 12 acides aminés, le CDR2-IMGT une longueur comprise entre 0 et 13¹, et le CDR3-IMGT une longueur comprise entre 5 et 26. La variabilité des longueurs de séquences est également due aux structures incomplètes pour lesquelles la séquence n'est pas connue. L'utilisation de telles structures dans les analyses qui vont suivre n'est pas gênantes, car l'unité de base est la position IMGT.

3.2.2 Composition en acides aminés.

Au sein des domaines sélectionnés, la fréquence de chaque type d'acide aminé ainsi que l'entropie associée a été calculée à chaque position. Les même calculs ont été effectués en regroupant les acides aminés en 3 classes d'hydrophopathies, 5 classes de volume et 11 classes de caractéristiques physico-chimiques [42].

Domaines de type V et C.

Cette analyse permet de retrouver sur notre jeu réduit de données les positions caractéristiques des V-DOMAINs et C-DOMAINs (Figure 3.3). Le coeur hydrophobe est constitué par des résidus strictement préservées, la cystéine 23, le tryptophane 41 et la cystéine 104. La plupart des positions possédant un caractère hydrophobe conservé ont leur chaîne latérale tournée vers l'intérieur du domaine, ce qui le stabilise. Toutes ces positions répondent donc à la définition des topohydrophobes de Poupon et Mornon 1998 [43]. Ainsi sur le Collier de Perles IMGT les positions des brins, B, C, C', D et E sont alternativement marquées comme ayant un caractère hydrophobe conservé. Les V-DOMAINs possèdent des acides aminés conservés à d'autre positions. La glycine 16 et la proline 46 sont conservées et sont caractéristiques des coudes bêtas. Les autres positions occupées par un acide aminé conservé ne peuvent être simplement expliquées par la topologie, la sérine 26, la glutamine 44, la sérine 83, l'acide glutamique 97, l'acide aspartique 98, l'alanine 100, la tyrosine 102, la glycine 119, la glycine 121 et la thréonine 122. Pour les V-DOMAINs d'IG la sérine 7, l'acide glutamique 73 et la sérine 128 sont conservées.

Les positions pouvant être potentiellement glycosylées ne sont pas strictement conservées dans les C-DOMAINs et V-DOMAINs de TR mais elles sont localisées dans la même

¹Le CDR2-IMGT de 13 acides aminés, qui dépassent la longueur de 10 acides résidus prévus dans le canevas de base de la numérotation des V-DOMAINs, appartient à un V-DOMAIN de lama.



Figure 3.2: Répartition de la longueurs des domaines sélectionnés dans IMGT/3Dstructure-DB (moins de 90% d'identité de séquence pour les V-DOMAINs et C-DOMAINs et moins de 98% d'identité de séquence pour les G-DOMAINs). (A) Longueur des 473 V-DOMAINs, longeur des CDR1-IMGT, longueur des CDR2-IMGT et longueur des CDR3-IMGT. (B) Longueur des 81 C-DOMAINs. (C) Longueur des 85 G-DOMAINs.



Figure 3.3: Acides aminés conservés à plus de 80% et positions dont l'hydropathie [42] est conservée à plus de 80% dans les 473 V-DOMAINs (haut) et les 77 C-DOMAINs et C-LIKE-DOMAINs (bas). Rouge = hydrophile, bleu = hydrophobe, jaune = neutre

région. Pour les V-DOMAINs, les positions 22 et 24, aux environs de la cystéine 23, dans les chaînes TR-ALPHA et une position variable dans les brins C" ou D des chaînes TR-BETA sont fréquemment des sites de glycosylation potentiels. Pour les C-DOMAINs, la position 90 dans les chaînes TR-ALPHA et une position à l'extrémité de la boucle FG dans les chaînes TR-BETA sont des sites fréquents de glycosylation potentiels.

Cette analyse de composition sera utilisée par la suite afin d'établir éventuellement des corrélations avec les propriétés particulières d'une position.

G-DOMAIN.

L'acide aspartique 39 est la seule position conservée dans plus de 80% des G-DOMAINs dont la structure 3D est connue. Mais pour chaque type de domaine pris individuellement, G-ALPHA, G-BETA, G-ALPHA1, G-ALPHA2, les positions avec un acide aminé conservé apparaissent tout le long de la structure, aussi bien dans les brins que dans l'hélice (Figure 3.2.2). Un très grand nombre de positions présentent une conservation des caractéristiques d'hydrophobicité. Cette conservation peut être due à la méthode de sélection des domaines et à la diversité essentiellement allélique des domaines, au contraire des IG et TR qui possèdent de nombreux gènes.

3.2.3 Distribution des brins dans les feuillets.

IMGT/3Dstructure-DB fournit la composition des feuillets des domaines IgSF. Elle a été déterminée pour chaque domaine grâce au réseau de liaisons hydrogène impliquant les atomes de la chaîne principale. Une expertise manuelle a également été effectuée pour valider ces résultats et identifier la position des brins ne présentant pas de liaisons hydrogène. En effet, certaines structures à basse résolution ne fournissent que la position des carbones alpha (liga et 1r70), et l'identification automatique de la composition des feuillets basée sur les liaisons hydrogène ne fonctionne pas. Cette étude montre la grande diversité dans les qualités d'affinement des structures RX au sein d'un même fichier de coordonnées atomiques. Ainsi certaines structures RX présentent plusieurs domaines de séquence identiques mais dont le réseau de liaison hydrogène entre brins est différent (par exemple 1jtr ou 1ktk). La topologie de repliement des domaines de type immunoglobuline a été initialement décrite comme un beta-barrel en clef grecque, sous-classe "simple" [44]. Une étude globale des domaines de type immunoglobuline [45] a postulé que les brins B, C. E et F sont toujours arrangés de la même facon, et constituent le coeur structural des domaines IgSF. Bork et al. 1994 [45] ont effectués une étude globale des domaines IgSF dans de nombreuses protéines, en considérant un domaine par famille de protéine. Notre approche s'en distingue par le fait que nous comparons les structures de nombreux domaines appartenant au même type de protéine, IG ou TR, afin d'étudier la variabilité structurale spécifique de ces protéines. Lorsque le brin A est divisé en deux brins distincts, ceux-ci sont notés A1 et A2. Lorsque les deux brins sont dans le même feuillet, ils sont désignés en commun en tant que brin A.

L'analyse des 481 structures de V-DOMAIN montre qu'ils possèdent deux configuration de distribution des brins entre les feuillets, la distribution [A1, B, D, E] [A2, C', C",



Figure 3.4: Acides aminés conservés à plus de 80% et positions dont l'hydropathie est conservée à plus de 80% dans les 28 G-ALPHA1 et les 28 G-ALPHA2 (haut), et dans les 13 G-ALPHA et les 18 G-BETA (bas). Rouge = hydrophile, bleu = hydrophobe, jaune = neutre

F, G] et la distribution [A1, B, C", D, E] [A2, C', F, G]. En effet le brin C" se situe à l'extrémité d'un feuillet et peut changer de feuillet suivant les V-DOMAINs. Il faut noter que la première configuration est largement majoritaire dans l'ensemble des V-DOMAINs étudiés. Elle est majoritaire pour les V-DOMAINs d'IG, et plus particulièrement pour les domaines VH et V-KAPPA, mais pas pour les V-LAMBDA où les deux distributions se rencontrent à parts égales. Pour les V-DOMAINs de TR, les deux distributions de feuillets se rencontrent également à parts égales, le premier groupe étant majoritaire pour les domaines V-BETA et le deuxième groupe étant majoritaire pour les domaines V-ALPHA (initialement remarqué par Fields et al. 1995 [14]). Ponctuellement certaines particularités sont à noter, les VH des fichiers de coordonnées 1hfm, 2hfm et 1p4b présentent un brin A2 dans le feuillet externe, et le V-ALPHA du fichier de coordonnée 1ymm ne possède pas de coordonnées atomiques pour le brin A1, qui doit être de conformation désordonnée. La distribution des brins entre les feuillets des C-DOMAINs des IG adopte une configuration unique, [A, B, D, E] [C, F, G]. La même configuration est adoptée pour les C-DOMAINs de TR. Il faut toutefois noter que le feuillet [C, F, G] des domaines C-ALPHA est pauvrement défini. Bentley et al. 1995 [46] et Garcia et al. 1996 [15] ont notés que le feuillet [C, F, G] n'est pas correctement replié, ses brins ressemblent plus à un succession de petites structures secondaires qu'à un vrai feuillet, et le brin F ne possède pas de liaisons hydrogène avec les autres brins [47]. La composition des feuillets des C-LIKE-DOMAINS des MHC et de la B2M, possèdent la même configuration [A, B, D, E] [C, F, G].

3.2.4 Angles entre feuillets des V-DOMAINs et C-DOMAINs.



Figure 3.5: Définition des plans passant par les feuillets d'un V-DOMAIN grâce à deux axes X et Y définis par les carbones alpha de différentes positions. Les même positions sont utilisées pour la définition des plans sur les feuillets des C-DOMAINs. Dans la figure du feuillet externe, les brins jaunes sont situés à l'arrière plan et dans la figure du feuillet externe, les brins jaunes sont situés à l'arrière plan et dans la figure du feuillet externe, les brins jaunes sont situés à l'avant plan.

L'angle de rotation entre les directions des brins des deux feuillets de chaque domaine ainsi que le parallélisme entre les deux feuillets ont été mesurées. Un plan a été défini pour chaque feuillet à partir de quelques positions (Figure 3.5). Chaque plan est défini par deux vecteurs, le premier définissant l'axe X du plan, et l'autre indique la direction de l'axe Y. L'axe Z est déduit de X et Y par un produit vectoriel. L'angle de rotation entre les plans est mesuré par l'angle entre l'axe X du premier et le projeté sur le premier plan de l'axe X du deuxième plan. Le parallélisme entre les plans est mesuré par l'angle formé par les axes Z perpendiculaires aux deux plans.

Le premier plan correspond au feuillet contenant le brin B, le vecteur X est sur le brin B et est défini par les carbones alpha des positions 18 et 25. Le vecteur Y est dans le plan défini par le vecteur X et par le vecteur dont les extrémités sont les coordonnées du carbone alpha de la position 7 et par les coordonnées du carbone alpha de la position 89. Le vecteur Y est orienté suivant le vecteur passant par les carbones alpha des positions 7 et 89.

Le deuxième plan correspond au feuillet contenant le brin F. Le vecteur X est sur le brin F, et est défini par les coordonnées du carbone alpha de la position 100 et par les coordonnées du carbone alpha de la position 105. Le vecteur Y est défini par les coordonnées du carbone alpha de la position 104 et par les coordonnées du carbone alpha de la position 104 et par les coordonnées du carbone alpha de la position 41.

Les positions définissant les différents vecteurs ont été choisies car elles sont partagées par les V-DOMAINs et C-DOMAINs et font partie de brins appartenant toujours au même feuillet (section 3.2.3).

Les angles présentés pour tous les V-DOMAIN (Table 3.2) sont très similaires à ceux des VH, les V-KAPPA et V-LAMBDA pris isolément. Les valeurs pour les V-ALPHA et V-BETA ne sont pas différentes malgré leur représentation beaucoup plus faible.

Par rapport aux V-DOMAINs, la valeur moyenne de l'angle de rotation indique que les brins des C-DOMAIN appartenant à des feuillets différents sont plus parallèles. Par contre au sein des C-DOMAIN des IG, les valeurs sont très similaires à celles rencontrées dans les V-DOMAIN des IG.

Les C-DOMAINS de TR présentent des valeurs très différentes en raison des domaines C-ALPHA dont les feuillets externes sont mal définis.

L'angle de rotation entre les feuillets des C-LIKE-DOMAINS des MHC est inférieur de 14 degrés en moyenne par rapport aux V-DOMAIN. Cet angle de plus est sujet à beaucoup de variabilité, les valeurs s'étendant de 10 à 35 degrés avec un écart type de 7 degrés. Les C-LIKE-DOMAINS de B2M, et de chaînes I-ALPHA, II-ALPHA et II-BETA pris isolément pourraient faire apparaître une plus grande hétérogénéité.

3.2.5 Alignements structuraux.

Les superpositions structurales effectuées dans tout ce chapitre ont toutes été réalisées grâce au programme ProFit (Martin A.C.R., http://bioinf.org.uk/software/profit) implémentant l'algorithme de McLachlan [48] et grâce aux alignements de séquence des domaines fournis par la numérotation unique IMGT. Table 3.2: Valeurs en degrés de l'angle de rotation et des angles entre les plans des feuillets dans les 451 V-DOMAIN d'IG et les 49 C-DOMAIN d'IG sélectionnés dans IMGT/3Dstructure-DB. Pour chaque domain l'écart type (E) de la distribution des angles est fourni.

	Domaine	Nombre de domaines	Angle de rotation	9			Paralléli entre les	sm $ pla$	e ans	
			Moyenne	Е	Min	Max	Moyenne	Е	Min	Max
	V-DOMAIN	473	35	3	22,5	44	20	3	12,5	$31,\!5$
IG	V-DOMAIN	440	35	3	27,5	44	19	3	12,5	$27,\!5$
	VH	261	36	2	29	43	18	2	12,5	$26,\!5$
	V-KAPPA	159	32	2	27,5	38	21	2	16, 5	$27,\!5$
	V-LAMBDA	20	39	3	35	49	21	2	$13,\!5$	25
TR	V-DOMAIN	33	30	3	22,5	$36,\!5$	24	3	$13,\!5$	31,5
	V-ALPHA	15	32	2	24,5	$_{36,5}$	24,5	3	18,5	28
	V-BETA	15	27	2	22,5	31	24	4	$13,\!5$	$31,\!5$
	V-GAMMA	1	32				20			
	V-DELTA	2	31		25,5	36	24		24	$24,\!5$
	C-DOMAIN	77	24	11	-45,5	40,5	23	6	10,5	49
IG	C-DOMAIN	45	29	4	16,5	$40,\!5$	22	4	14,5	36
	CH1	16	28	4	23,5	$_{38,5}$	19	2	14,5	$23,\!5$
	CH2	7	32	4	25	$40,\!5$	24	3	19	31
	CH3	7	29	5	16,5	34,5	24	5	18,5	36
	CH4	1	34				23			
	C-KAPPA	9	27	3	19	31	23	2	20,5	26
	C-LAMBDA	5	28	3	22,5	$_{30,5}$	26	2	23	$_{30,5}$
TR	C-DOMAIN	5	0	29	-45,5	29	32	11	16, 5	49
	C-ALPHA	2	35	11	-45,5	-22	35	3	$_{31,5}$	38
	C-BETA-1	1	17				49			
	C-BETA-2	1	23				26			
	C-GAMMA-1	1	17				29			
MHC	C-LIKE	27	21	$\overline{7}$	10	35	22	6	10,5	$32,\!5$

Domaines de type V et C.

Dans IMGT/3Dstructure-DB, une superposition par paire a été réalisée entre les V-DOMAINS, et entre les C-DOMAINS et C-LIKE-DOMAINS de MHC en alignant toutes les positions ayant le même numéro IMGT sauf les CDR (positions 27 à 38, positions 56 à 65 et positions 105 à 117) et la boucle DE (positions 84 et 85) pour les V-DOMAINS, ainsi que la boucle BC (positions 27 à 38), les positions 42 à 77 (contenant le brin CD), la boucle DE (positions 84.1 à 85.1) et la boucle FG (positions 104 à 118) des C-DOMAINS. La distance entre les positions équivalentes a été enregistrée et la moyenne quadratique de ces distances (RMSD) a été calculée pour chaque position. Les comparaisons ont été effectuées pour les V-DOMAINS et les C-DOMAINS, pour chaque type de récepteur (IG, TR et MHC), et pour chaque type de domaines (VH, V-KAPPA, V-LAMBDA, V-ALPHA, V-BETA).

V-DOMAIN Pour les V-DOMAINs, les variabilités structurales les plus importantes apparaissent au niveau de la jonction entre les brins A1 et A2 (positions 8 et 9), entre les brins C et C' (positions 46 à 49), au niveau du brin C" (positions 66 à 74), aux positions 83 et 84 entre les brins D et E, de même au début (positions 1 et 2) et à la fin du domaine (positions 127 et 128) (Figure 3.6). Au sein des V-DOMAINs des IG, la variabilité est moindre au niveau du brin C", et elle est globalement faible au sein des VH et surtout des V-KAPPA et des V-LAMBDA. La variabilité des positions 8 et 9 est importante pour les VH mais très faible pour les VL (V-KAPPA et V-LAMBDA). En observant de façon plus précise les superpositions structurales on peut effectuer une classification des conformations de la jonction entre les brins A1 et A2. Cette classification avait déjà été effectuée par [49] pour les domaines VH. Les VL adoptent la même conformation correspondant au type IV des VH. Pour les V-DOMAINs de TR, le brin C" est très variable. Les positions 7, 8, 9 et 10 des V-ALPHA présentent plus de variations que les V-BETA, et la partie reliant les brins C et C' (positions 46, 47 et 48). Les V-BETA sont également très variables sur les positions 92 à 96, ce qui n'est pas le cas des V-ALPHA.

C-DOMAIN La comparaison de tous les C-DOMAINs montre que la conservation de la conformation autour du coeur commun est plus faible que pour les V-DOMAINs. Toutes les parties en dehors de ce coeur sont l'objet de variabilité conformationnelle importante : les positions situés avant le premier brin $(1.7 \ a$ 1.1), les positions entre les brins A et B (9 à 19), les positions situés sur le brin CD ainsi que les positions du brin C, les positions de la boucle EF, les positions des brins D et G, et les boucles BC (équivalent au CDR1 des V-DOMAINs) et FG (équivalent au CDR3 des V-DOMAINs) (Figure 3.7). Pour les C-DOMAINs d'IG la variabilité structurale est plus faible. Elle concerne la boucle AB (positions 11 à 18), le brin CD (positions 43 à 77), le brin D, la boucle EF (positions 91 à 101), la boucle FG (surtout positions 109-115) et le brin G. Il faut noter la relative conservation structurale de la boucle BC entre tous les C-DOMAINs d'IG. De même la partie haute du brin E (85.5 à 90) et le haut du brin F (102 à 108) qui sont appariés sont bien conservés. Les C-KAPPA présentent peu de variabilité de leur structure. Les C-LAMBDA présentent un brin G très variable car comme il a été décrit précédemment



Figure 3.6: Moyenne quadratique des distances entre les carbones alpha (RMSD) de chaque position des 469 V-DOMAINS (261 VH, 158 V-KAPPA, 20 V-LAMBDA, 15 V-ALPHA et 15 V-BETA). Les Colliers de Perles IMGT sont coloriés linéairement de 1Å de RMSD et moins (blanc) à 3Å RMSD et plus (noir). Seule une structure de V-GAMMA et deux structures de V-DELTA ont été sélectionnées et leur superposition n'est pas présentée. Les superpositions des CDR-IMGT, qui doit tenir compte de leur longueur, est disponible en ligne sur les pages de statistiques de IMGT/3Dstructure-DB.



Figure 3.7: Moyenne quadratique des distances entre les carbones alpha (RMSD) de chaque position des 43 IG C-DOMAINS (16 CH1, 7 CH2, 7 CH3, 9 C-KAPPA et 5 C-LAMBDA). Les Colliers de Perles sont coloriés linéairement de 1Å de RMSD et moins (blanc) à 3Å RMSD et plus (noir). L'analyse des C-ALPHA, C-BETA-1, C-BETA-2, C-GAMMA-1 et C-DELTA n'est pas présentée car moins de deux structures 3D ont été sélectionnées.

il peut changer de feuillet. Le haut du brin D (positions 82 à 84.4) et une partie de la boucle BC (positions 33 à 36) sont également variable. Les CH1 présentent des variations structurales au niveau des boucles AB (positions 9 à 17), CD (positions 43 à 77), DE (positions 84.4 et 84.5) et EF (positions 91 à 101 et position 114). Le CH2 varie au niveau de la boucle AB (positions 15.2 et 16), de tout le brin CD, de tout le brin B et de la boucle FG. Les C-DOMAINs de TR présentent bien plus de variations structurales que ceux des IG. Toutes les positions en dehors du coeur commun présentent de grandes variabilité. Pourtant les C-ALPHA présentent une bonne conservation de structure en dehors du brin G qui est faiblement structuré et semble pouvoir adopter différentes conformations. Les différences sont donc dues à une variabilité entre les différents types de C-DOMAIN de TR. La superposition des C-LIKE-DOMAINs de MHC montre que la variabilité structurale concerne le brin A, la boucle AB, le brin CD et le brin D. La conformation des boucles BC et FG sont relativement bien conservées, ce qui est important pour les contacts établis avec le plancher du MHC.

Les C-DOMAINS présentent donc globalement plus de variabilité structurale entre les différents types de C-DOMAIN que ne le font les V-DOMAINS. Ce résultat parait étonnant et il est possible que des structures de V-DOMAINS très stables soient requises afin de pouvoir convenablement adapter des séquences de CDR très variables.

G-DOMAIN.

Les G-DOMAINS ont été superposés grâce au programme ProFit en alignant toutes les positions des G-DOMAINS sauf les boucles AB (position 14 à 18) et les boucles CD (position 38 à 42). Les alignements structuraux montrent la bonne conservation du plancher de feuillets bêta. Les valeurs seuils utilisés sont 1Å RMSD et en dessous pour la conservation structurale et 3Å et au-dessus pour la grande variabilité structurale (Figure 3.8). Les premières positions du G-ALPHA (jusqu'à la position 7) ainsi que la première portion de l'hélice (positions 48 à 59) et la dernière portion de l'hélice (positions 84 à 92) présentent une variabilité structurale comparé au reste des positions qui ont une conformation extrêmement bien conservé. Pour le G-ALPHA2, toutes les positions sont extrêmement bien conservées en dehors de la boucle AB (positions 14 à 18) et de la fin du domaine (positions 87-92). Les deux premières positions ainsi que la partie reliant les deux portions de l'hélice (positions 61 et 62) possèdent une variabilité mineure (1,20Å RMSD).

3.2.6 Analyse de contacts.

Les statistiques de contacts internes de chaque domaine ont été calculées pour les différents types de domaines. La conservation des contacts peuvent être consultés sur l'outil en ligne avec un seuil de fréquence réglable et pour différent types de contacts (tous les contacts, les liaisons hydrogène, les liaisons hydrogène par l'intermédiaire d'une molécule d'eau, les contacts polaires et les contacts non polaires) et de niveau de contacts (entre atomes de la chaîne principale, la chaîne latérale avec la chaîne principale, et entre atomes des chaînes latérales).



Figure 3.8: Moyenne quadratique des distances entre les carbones alpha (RMSD) de chaque position des 29 G-ALPHA1, 29 G-ALPHA2, 13 G-ALPHA et 18 G-BETA. Les perles sont coloriées linéairement du blanc au noir, de 1Å de RMSD et moins à 3Å de RMSD et plus.

Analyse de fréquence.

V-DOMAIN Les liaisons hydrogène des V-DOMAINs, conservées dans plus de 80% des cas, correspondent aux liaisons hydrogène entre atomes de la chaîne principale reliant les brins. Deux liaisons hydrogène utilisant des atomes de chaînes latérales sont également conservées, entre la chaîne latérale de la position 6 (une glutamine ou un glutamate) et la chaîne principale de la position 102 (souvent une tyrosine), et entre la chaîne latérale de la position 98 (un acide aspartique). Les contacts conservés pour un CDR donné et une longueur de CDR donnée ont également été calculés. Dans ce cas les contacts avec les positions des autres CDR ont été enregistrés sans tenir compte de la longueur de l'autre CDR, ce qui permet d'avoir un idée des contacts échangés avec les autres CDR. Par exemple, il existe souvent une liaison hydrogène entre la dernière position du CDR1-IMGT et la position 107 du CDR3-IMGT (87% pour une taille de CDR1-IMGT de 6 résidus (position 32) et 71% pour une taille de CDR1-IMGT de 8 résidus (position 34)). Dans le CDR2-IMGT, la position 56 effectue une liaison hydrogène dans la majorité des cas (98% pour une longueur de 8, 95% pour une longueur de 10) avec la position 39.

Les liaisons hydrogène avec une molécule d'eau intermédiaire ne sont jamais conservées.

C-DOMAIN Les liaisons hydrogène conservées des C-DOMAINs, dans 80% des cas étudiés, sont toutes entre atomes de la chaîne principale reliant les brins. Les C-ALPHA ne possèdent de liaisons hydrogène conservées qu'entre les brins A, B, D et E.

G-DOMAIN Les liaisons hydrogène conservées des G-DOMAINs, dans plus de 80% des cas, correspondent aux liaisons hydrogène entre atomes de la chaîne principale reliant les brins et les boucles de l'hélice. On peut noter que cette conservation intervient quel que soit le type de G-DOMAIN.

Pour G-ALPHA1, une liaison hydrogène utilisant des atomes de chaînes latérales est conservée, entre la chaîne latérale de la position 64 (une thréonine) et la chaîne principale de la position 45. Dans le coude AB, la liaison hydrogène entre la chaîne latérale de l'arginine 14 et la chaîne principale du glutamate 19 est conservée, ce qui n'allait pas de soi car la majorité des acides aminés de la boucle sont est conservée mais la superposition structurale a montré que la conformation n'était pas conservée (3,9Å de RMSD à la position 18). Pour G-ALPHA et G-ALPHA2, une liaison hydrogène utilisant des atomes de chaînes latérales sont également conservées, entre la chaîne latérale de la position 29 (une glutamine, un glutamate ou un acide aspartique) et la chaîne principale de la position 4. La chaîne latérale de la position 29 vient ainsi en quelque sorte prolonger les liaisons hydrogènes entre le brin A et le brin B.

Analyse statistiques des contacts des V-DOMAINs des IG.

Avec Fabrice Sarniguet, étudiant du DESS des méthodes statistiques de Montpellier, j'ai réalisé une analyse statistique des contacts internes des V-DOMAINs des IG. Le but étant d'observer si une discrimination entre les VH, les V-KAPPA et les V-LAMBDA



Figure 3.9: Contacts entre positions possédant le plus de contacts entre atomes (A) et contacts permettant de différencier VH et VL (B), et de différencier V-KAPPA et V-LAMBDA (C). Les paires de positions en contacts (positions en rouge reliés par un trait vert) ont été déterminées par une analyse de composante principale de la matrice de contacts entre positions pour chaque type de domaine (voir texte).

pouvant être observée en ayant recours aux densités de contacts internes. Une analyse en composantes principales (ACP) a été réalisée où les coordonnées statistiques caractérisant chaque domaine sont le nombre de contacts effectuée par chacune de ses positions (les différents axes de l'espace sont les positions et les valeurs dans ces directions sont les nombres de contacts). L'analyse permet donc de comparer les différences de densité dans les différentes parties du domaine. Une classification selon la méthode de Ward permet de séparer les deux domaines des chaînes légères des domaines VH, puis de séparer les domaines des chaînes légères V-KAPPA et V-LAMBDA. Une autre ACP a été réalisée en prenant comme coordonnées pour chaque domaine le nombre de contacts atomiques entre paires de positions. Cette analyse permet de conclure (Figure 3.9) que les contacts les plus nombreux interviennent au centre du domaine (21-89, 41-54, 41-89, 42-103, 43-51, 43-102, 77-90, 79-88 et 95-98) ainsi qu'à l'ancrage du CDR2 (55-66). Ces contacts importants sont peu discriminant entre les trois types de domaines. Les contacts suivants permettent par contre de faire la différence entre les domaines VH et les domaines de chaînes légères : 20-88, 52-54, 52-66, 67-78, 69-72, 68-70, 71-75, 71-76, 80-82, 81-83, 81-84, 81-86, 84-86 et 86-88. Ces positions sont situées dans le brin C" et aux extrémités des brins E. C' et D. La position 84 est souvent un acide aminé de petite taille dans les domaines V-KAPPA et V-LAMBDA tandis qu'il est parfois volumineux dans les VH. Les contacts suivants sont plus fréquents pour les domaines V-KAPPA et V-LAMBDA que pour les domaines VH : 8-10, 8-122, 10-123, 12-127, 13-127, 40-42, 41-54, 41-89, 42-54, 43-45, 53-76, 55-66, 55-68, 55-58, 55-58, 55-58, 55-58, 55-58, 55-58, 55-58, 55-58, 55-5867-71, 67-74, 72-75, 72-76, 83-86 et 83-87. Les contacts suivants sont plus fréquents pour le V-KAPPA que pour le V-LAMBDA, 3-26, 8-10, 10-123, 39-87, 83-87, 84-87 et 85-87. Les contacts suivant sont plus fréquents pour le V-LAMBDA que pour le V-KAPPA, 7-9, 7-122, 8-123, 9-123, 26-85, 67-74, 75-92, 84-86, 96-126.

3.3 Organisation des domaines dans les récepteurs.

Les récepteurs sont étudiés au niveau de la formation des domaines partenaires, puis au niveau de la constitution du récepteur par assemblage des partenaires.

142 V-PARTNER et 32 C-PARTNER de séquence différentes ont été sélectionnées avec le même protocole que pour les domaines de séquences indépendantes (section 3.2), en utilisant un seuil de pourcentage d'identité de 90%. Le pourcentage d'identité de séquence est calculé sur la séquence des deux domaines simultanément.

170 fragments Fab d'IG et 5 TR sont également sélectionnés avec le même protocole, le pourcentage d'identité étant calculé sur la séquence des quatre domaines (les deux V-DOMAINs et les deux C-DOMAINs). La liste des structures sélectionnées dans IMGT/3Dstructure-DB est fournie en annexe (page 149 pour les domaines partenaires et page 152 pour les Fab).

Pour deux domaines, le parallélisme entre les feuillets internes et externes au partenaire est mesuré, ainsi que l'angle de rotation entre les feuillets internes et entre les feuillets externes (Figure 3.10). Les mêmes vecteurs que ceux utilisés pour la mesure du parallélisme entre les feuillets d'un même domaine sont employés ici et les calculs sont similaires.



Figure 3.10: Représentation des angles pour la mesure du positionnement de domaines partenaires.

Sur une même chaîne l'angle formé par deux domaines consécutifs est également mesuré à partir de l'angle formé par les premiers vecteurs des feuillets contenant le brin B.

La position relative des V-PARTNER et C-PARTNER est mesurée par le calcul des angles entre deux plans formés par les deux partenaires. Le plan du V-PARTNER est défini par deux vecteurs. Le premier vecteur passe par les carbones alpha des positions 104 de chacun des deux V-DOMAINS (Figure 3.11). Le second vecteur passe par les carbones alpha des positions 19 des deux V-DOMAINS. Le plan du C-PARTNER est défini par un premier vecteur entre les carbones alpha des positions 104 des deux C-DOMAINS et un deuxième vecteur entre les carbones alpha des positions 19 des deux C-DOMAINS.

La surface enfouie par chaque domaine dans son partenaire et dans le récepteur entier est calculée en effectuant la différence entre la surface accessible du domaine isolé et du même partenaire faisant partie du récepteur. La surface accessible de chaque position est fournie par le logiciel Stride [28].

3.3.1 V-PARTNER.

Il n'y a pas de différence observée au niveau des angles entre les V-PARTNER employant un V-KAPPA ou un V-LAMBDA (Tables 3.3 et 3.4). Pour les TR l'angle de parallélisme est similaire à celui des IG, l'angle de rotation des feuillets externes est moins important mais pas celui des feuillets internes. Ces valeurs indiquent que les feuillets externes forment un V dont la partie ouverte serait dirigée vers l'antigène.

La surface enfouie par chaque domaine varie entre 650 et 1100\AA^2 quel que soit le type de domaine et le type de récepteur (Table 3.5). D'après l'étude de Bahadur *et al.* 2004 [50], la surface enfouie à l'interface des homodimères est en moyenne de 3900Å² avec un écart type de 2200Å². Les V-PARTNER ont donc une surface enfouie faible qui est comparable à celle des complexes d'interaction protéine-protéine (1900Å² [50]) ou des complexes cristallins (1600Å² [50]). Table 3.3: Valeurs moyennes (Moy) et écarts types (E) de la distribution des angles entre feuillets internes des V-DOMAIN dans un V-PARTNER d'IG et de TR, et des C-DOMAIN et C-LIKE-DOMAIN dans un C-PARTNER d'IG, de TR et de MHC. 142 structures 3D de V-PARTNER et 32 structures 3D de C-PARTNER avec moins de 90% d'identité de séquence ont été utilisées. Les valeurs pour l'ensemble des domaines de chaque section sont en gras.

	Ang	gles en	tre	feuil	lets in	ternes			
PARTNER	Nb	rotation				$\operatorname{parall}\acute{\mathrm{e}}\mathrm{lisme}$			
		Moy	Е	Min	Max	Moy	Е	Min	Max
VH V-KAPPA	123	47	3	35	55	20	3	10	30
VH V-LAMBDA	9	52	5	40	60	22	2	15	30
	132	47	4	35	60	20	3	10	30
V-ALPHA V-BETA	9	43	4	35	50	8	3	0	15
V-GAMMA V-DELTA	1	48				10			
	10	44	4	35	50	8	3	0	15
CH1 C-KAPPA	9	142	1	135	145	63	2	55	70
CH1 C-LAMBDA	5	135	4	125	145	65	4	60	75
CH2 CH2	1	124				91			
CH3 CH3	5	148	3	140	155	54	3	45	60
CH4 CH4	1	147				52			
	21	140	6	120	155	62	8	45	95
C-ALPHA C-BETA-1	1	122				62			
C-ALPHA C-BETA-2	1	113				61			
	2	117	4	110	125	61	0	60	65
C-LIKE C-LIKE	9	93	7	80	105	53	3	45	60

Table 3.4: Valeurs moyennes (Moy) et écarts types (E) de la distribution des angles entre feuillets externes des V-DOMAIN dans un V-PARTNER d'IG et de TR et des C-DOMAIN et C-LIKE-DOMAIN dans un C-PARTNER d'IG, de TR et de MHC. 142 structures 3D de V-PARTNER et 33 structures 3D de C-PARTNER avec moins de 90% d'identité de séquence ont été utilisées. Les valeurs pour l'ensemble des domaines de chaque section sont en gras.

	Ang	gles en	tre :	feuille	ets ext	ernes			
PARTNER	Nb	rota		$\operatorname{parall\acute{e}lisme}$					
		Moy	Е	Min	Max	Moy	Е	Min	Max
VH V-KAPPA	123	130	4	115	140	43	4	30	55
VH V-LAMBDA	9	138	5	130	150	39	3	30	45
	132	130	4	115	150	43	4	30	55
V-ALPHA V-BETA	9	114	3	105	125	42	6	30	55
V-GAMMA V-DELTA	1	123				42			
	10	115	4	105	125	42	6	30	55
CH1 C-KAPPA	9	67	6	55	75	30	3	25	40
CH1 C-LAMBDA	5	61	5	50	70	32	3	25	40
CH2 CH2	1					43			
CH3 CH3	5	71	4	60	80	16	8	0	30
CH4 CH4	1	65				38			
	21	64	13	10	80	28	9	0	45
C-ALPHA C-BETA-1	1	109				26			
C-ALPHA C-BETA-2	1	120				17			
	2	114	5	105	120	22	5	15	30
C-LIKE C-LIKE	9	29	16	5	60	25	5	15	35

Table 3.5: Valeurs moyennes (Moy) et écarts types (E) de la surface enfouie par chaque domaine dans les V-PARTNER et les C-PARTNER d'IG, de TR et de MHC. 142 structures 3D de V-PARTNER et 32 structures 3D de C-PARTNER avec moins de 90% d'identité de séquence ont été utilisées. Les valeurs pour l'ensemble des surfaces enfouies de chaque section sont en gras.

Surface enfouie par chaque domaine									
PARTNER	Nb	Dom	aine	1		Dom	aine	2	
		Moy	Ε	Min	Max	Moy	Ε	Min	Max
VH V-KAPPA	123	837	68	700	1050	874	70	750	1100
VH V-LAMBDA	9	846	93	650	1000	867	81	700	1000
	132	838	70	650	1050	874	71	700	1100
V-ALPHA V-BETA	9	872	98	750	1050	860	89	700	1000
V-GAMMA V-DELTA	1	852				840			
	10	861	88	750	1050	846	82	700	1000
CH1 C-KAPPA	9	947	73	850	1200	948	51	850	1050
CH1 C-LAMBDA	5	953	44	850	1050	931	26	850	1000
CH2 CH2	1	845				840			
CH3 CH3	5	1169	69	1000	1250	1176	73	1000	1250
CH4 CH4	1	1149				1144			
	21	1006	120	800	1250	1002	120	800	1250
C-ALPHA C-BETA-1	1	1376				1386			
C-ALPHA C-BETA-2	1	1365				1386			
	2	1370	5	1350	1400	1386	0	1350	1400
C-LIKE C-LIKE	9	422	67	350	600	407	72	300	600



Figure 3.11: Définition des plans passant par les feuillets d'un V-DOMAIN grâce à deux axes X et Y définis par les carbones alpha de différentes positions.

3.3.2 C-PARTNER.

Les angles de rotation formés par les deux feuillets internes de chaque domaine présentent deux distributions, l'une s'étend de 80 à 105 degrés et correspond aux C-LIKE-DOMAINs des MHC, l'autre s'étend de 110 à 155 degrés et correspond aux C-DOMAINs des IG et des TR. L'angle de rotation formé entre les deux feuillets externes de chaque domaine s'étend de 10 à 120 degrés et correspond aux C-DOMAINs des IG et des TR et aux C-LIKE-DOMAINs de MHC. Les C-DOMAINs des TR possèdent les angles de rotation les plus élevés, de 105 à 120 degrés, qui sont bien distincts des autres C-DOMAINs. L'angle mesurant le parallélisme entre les feuillets est compris entre 0 et 45 degrés pour les feuillets externes et compris entre 45 et 95 degrés pour les feuillets internes.

La surface enfouie présente deux distributions, la première de 300 à 600Å^2 pour les deux domaines et qui correspond aux C-LIKE-DOMAINs des MHC. La deuxième distribution varie entre 800 et 1400Å^2 avec en moyenne 1040Å^2 et qui correspond aux C-DOMAINs des TR et des IG. Cette distribution se retrouve globalement chez les IG, et la surface enfouie des TR est située entre 1350 et 1400 pour les deux domaines.

3.3.3 Entre le V-PARTNER et le C-PARTNER.

L'angle formé par les domaines VH et CH1, V-KAPPA et C-KAPPA, et V-LAMBDA et C-LAMBDA sont pour la plupart compris entre 35 et 75 degrés (Figure 3.12). L'angle formé par les domaines VH-CH1 possèdent deux pics de forte population, l'un vers 45 degrés et l'autre vers 65 degrés, tandis que les domaines légers n'enregistrent qu'un seul pic vers 60 degrés.

Table 3.6: Surface enfouie en $Å^2$ par chaque domaine d'un récepteur Fab dans 215 Fab sélectionnés dans IMGT/3Dstructure-DB. VL = V-KAPPA + V-LAMBDA, CL = C-KAPPA + C-LAMBDA.

Domaine	Moyenne	Ecart type	Minimum	Maximum
VH	1067	85	850	1300
VL	1181	90	900	1450
CH1	1202	93	1000	1450
CL	1291	94	950	1550

La surface enfouie par les 4 domaines d'un Fab d'IG sont du même ordre de grandeur mais présentent de légères différences (Table 3.6). Le domaine VH possède une surface enfouie inférieure à celle des trois autres domaines. Pour les TR, La surface enfouie par les 4 domaines de TR présentent plus de disparité, les C-ALPHA et C-BETA présentant une surface enfouie supérieure aux V-ALPHA et V-BETA. Ces derniers ont une surface enfouie comparable à celle des domaines d'IG.



Figure 3.12: Angles en degrés formés par les domaines VH et CH1 d'une part et VL (V-KAPPA + V-LAMBDA) et CL (C-KAPPA + C-LAMBDA) d'autre part dans 170 récepteurs Fab extrait de IMGT/3D
structure-DB. L'angle est mesuré entre les ecteurs passant par les carbones alpha des positions 18 et 25 appartenant au brin B.

3.4 Interface des complexes IG/antigène.

De nombreuses études se sont penchées sur l'interaction entre les IG et les antigènes (revues dans [51], [52] et [34]). Les résultats de superpositions des CDR et l'analyse des positions des V-DOMAINS en contact avec l'antigène sont relativement simples et ont valeurs d'exemples quant à la facilité d'emploi de IMGT/3Dstructure-DB pour de telles analyses.

3.4.1 Superposition des CDR-IMGT.

Basé sur le nom usuel des IG (IMGT protein name), les données de 79 IG dont la structure 3D a été déterminée expérimentalement, sans ligand et avec un ou plusieurs antigènes, ont été extraites de IMGT/3Dstructure-DB (liste fournie en annexe page 154).

Les V-DOMAINs ont été superposés grâce au programme ProFit en alignant toutes les positions des V-DOMAINs sauf les CDR (position 27 à 38, positions 56 à 65 et positions 105 à 117) et la boucle DE (positions 84 et 85). La moyenne quadratique des déviations des coordonnées (RMSD) a été calculée pour chaque CDR-IMGT (Figure 3.13). Ces mesures (Figure 3.13) montrent que, quel que soient le type de domaine et le CDR-IMGT, le plus souvent le CDR-IMGT ne subit pas de changements conformationnels à la liaison (entre 0 et 1Å RMSD). Néanmoins on observe de grandes variations atteignant 5Å RMSD pour le CDR3-IMGT du VH. Les CDR2-IMGT et CDR3-IMGT des V-KAPPA et V-LAMBDA sont par contre moins variables avec un RMSD toujours inférieur à 2Å. Les CDR1-IMGT des VH ont un RMSD inférieur à 2,5Å et les CDR1-IMGT des V-KAPPA et V-LAMBDA ont un RMSD inférieur à 5Å. Pour les V-KAPPA et les V-LAMBDA, le CDR dont la conformation est la plus variable est le CDR1-IMGT possède la conformation la plus variable, tandis que les CDR1-IMGT et CDR2-IMGT ont une conformation la plus variable à la liaison.

3.4.2 Positions en contacts.

Les CDR-IMGT sont les principaux éléments de contact avec les antigènes, mais des positions n'appartenant pas aux CDR-IMGT sont également mises à contribution (Figure 3.14, Table 3.7 et Kaas *et al.* 2005 [53] Publication 5). Que ce soit pour les VH ou les VL, les positions plus impliquées sont les 'ancres' du CDR2-IMGT (positions 55 et 66) et la position 40 qui interagissent principalement par l'intermédiaire de leure chaîne latérale. Pour les VL, les positions 67, 68 et 69 sur le brin C", ainsi que les positions 80 et 84 pour le V-KAPPA, et les positions 58 et 118 pour le V-LAMBDA sont parfois impliquées dans la liaison. En effet les petits ligands s'enfouissent profondément au centre du récepteur et font intervenir des positions proches du coeur du domaine.

Les interactions des CDR-IMGT ont été analysés pour chaque longueur de CDR-IMGT et chaque type de domaine. La deuxième partie du CDR1-IMGT crée plus de contacts avec l'antigène que la première, et ce quel que soit le type de domaine (VH, V-KAPPA ou V-LAMBDA) ou la longueur du CDR1-IMGT. Toutes les positions du CDR2-IMGT

Table 3.7: Positions les plus souvent impliquées dans les contacts IG/antigènes et la fréquence de contacts avec l'antigène de ces positions (nombre de structures 3D où les positions sont en contact/ nombre de structures 3D) dans les domaines VH (A), V-KAPPA (B) et V-LAMBDA (C). Quand une seule structure 3D est utilisée, les positions en contacts peuvent être trouvées sur la Figure 3.14.

A - VH

	Longueur		
CDR-IMGT	des CDR	Positions	Fréquence
CDR1-IMGT	8	34	180/260
	9	35	17/27
	10	34	28/30
FR2-IMGT		55	196/317
CDR2-IMGT	7	57	91/113
	8	57	94/182
	10	57 61	16/21
CDR3-IMGT	5	106 107	11/11
	6	115	5/5
	7	107	34/35
	8	107 108	4/5
	9	107	18/19
	10	109 113	21/23
	11	109 110	24/28
	12	107	50/56
	13	113	36/43
	14	107	11/14
	15	112.1	20/22
	16	113	15/19
	17	109	2/2
	18	$112.1\ 113$	2/2
	19	$112 \ 112.1$	5/5
	20	$112.2 \ 112.3$	2/2
	21	$111 \ 111.3 \ 111.4 \ 112.1 \ 112.2$	7/7
		$112.3 \ 112.4$	
	22		1/1
	24	$107 \ 110 \ 112 \ 112.3$	18/18

Fin de table à la page suivante

B - V-KAPPA

CDR-IMGT	Longueur des CDR	Positions	Fréquence
CDR1-IMGT	5	31	10/16
	6	32	96/154
	7	33	6/11
	10	36	15/16
	11	37	52/75
	12	38	17/20
FR2-IMGT		55	91/292
CDR2-IMGT	3	56	120/292
CDR3-IMGT	7		1/1
	8	107	23/24
	9	107	229/255
	10	$107 \ 108 \ 116$	3/5
	11	$110 \ 113$	7/7

C - V-LAMBDA

CDR-IMGT	Longueur des CDR	Positions	Fréquence
CDR1-IMGT	8 9	35	$\frac{1/2}{17/23}$
FR2-IMGT		40	4/26
CDR2-IMGT	3	56	6/26
CDR3-IMGT	9 11 12	107 107 109	${18/23} \ {2/2} \ {1/1}$


Figure 3.13: Moyenne quadratique des distances entre les carbones alpha (RMSD) entre la forme liée et non liée à un antigène des CDR-IMGT de VH, de V-KAPPA + V-LAMBDA. 79 IG différentes possédant des structures en complexe avec un antigène et des structures sans antigène ont été utilisées.

peuvent être impliquées dans la liaison à l'antigène, quels que soient le type de domaine et la longueur du CDR2-IMGT. Le CDR3-IMGT peut également entrer en contact avec l'antigène avec pratiquement toutes ses positions, quels que soient le type de domaine et la longueur du CDR3-IMGT. Même avec des CDR3-IMGT très long (jusqu'à 24 acides aminés dans les données utilisées), les positions en contact se situent aussi bien à la pointe de la boucle que proche de ses ancrages. L'outil de visualisation en ligne fournit des informations utiles sur les positions engagées fréquemment dans une interaction spécifique (liaison hydrogène ou contacts polaires effectués par les atomes des chaînes latérales) avec l'antigène.

3.5 Interface des complexes peptide/MHC.

Les sites de liaisons du MHC-I et du MHC-II ont une structure similaire et les présentations des peptides comportent certains points communs [17]. L'orientation du peptide dans le site de liaison a été sujet à discussions lors des premières publications des structures 3D de peptide/MHC. En effet des études théoriques ont montré que les peptides pourraient se lier dans les deux sens [54] mais actuellement une seule orientation a été VH





La figure continue à la page suivante



Figure 3.14: Implication des positions des domaines VH, V-LAMBDA et V-KAPPA dans la liaisons à l'antigène. Les Colliers de Perles IMGT sont coloriés linéairement depuis l'absence de contacts (blanc) au maximum de contacts rencontrés (bleu). Un nombre différent de structures est utilisé pour les positions dans les CDR-IMGT et les FR-IMGT, et les échelles de couleurs ne peuvent donc être comparées entre elles. La Table 3.7 fournit le nombre de structures utilisées ainsi que l'occurrence maximale rencontrée.

observée expérimentalement. L'extrémité N terminale du peptide étant proche de l'extrémité N terminal de l'hélice de G-ALPHA1 pour les MHC-I et de l'hélice de G-ALPHA pour les MHC-II. Le mode de liaison des peptides présente également des différences, ne serait-ce que par la longueur du peptide présenté.

Peptide/MHC-I Les structures 3D des peptide/MHC-I révèlent que le peptide est complètement enfoui dans le site de liaison. Les extrémités N et C du peptide sont profondément enfouies dans les poches du MHC, tandis que la partie centrale peut sortir du site de liaison [4]. Ce mode de liaison implique que plus le peptide sera long et plus la liaison sera faible, ce qui explique la limitation de la longueur du peptide. Les peptides trouvés dans les structures 3D de complexes peptide/MHC-I ont en majorité une taille de 9 acides aminés de longs. Certaines structures RX (tel 1mhc [55]) présentent une liaison

inhabituelle, avec la partie N-terminale du peptide quittant le site de liaison. La poche à l'extrémité du site est alors occupée par le deuxième acide aminé du peptide.

Peptide/MHC-II Contrairement aux MHC-I, les deux extrémités du site de liaison du MHC-II sont ouvertes, ce qui autorise le peptide à s'étendre au delà du site de liaison [17]. La partie du peptide dans le site de liaison adopte une conformation polyproline-II-like et la partie en dehors du site de liaison est relativement désordonnée. Les peptides se liant au MHC-II sont donc en apparence plus long que ceux se liant au MHC-I, mais seulement 9 résidus occupent réellement le site de liaison. La longueur des peptides en complexe avec un MHC-II est de 6 à 18 acides aminés, lorsque la liaison peptidique des construction MHC-II-ALPHA_BETA-PEPTIDE n'est pas prise en compte. Des modes de liaisons inhabituels sont parfois rencontrés lorsque le peptide ne remplit que partiellement le site de liaison [56].

3.5.1 Flexibilité du peptide dans le groove

Les structures de peptide/MHC de IMGT/3Dstructure-DB (liste en annexe 159) ont été superposées grâce au programme ProFit en alignant toutes les positions des G-DOMAINs sauf les boucles AB (position 14 à 18) et les boucles CD (position 38 à 42). La distance entre les positions équivalentes des peptides a été mesurée entre toutes les paires de structures. La moyenne quadratique (RMSD) entre les positions du peptide a été calculée (Figure 3.15) pour les peptides de MHC-I de longueur 8, 9 et 10, et pour les 9 positions des peptides situées dans le site de liaison du MHC-II. Toutes les structures 3D de peptide/MHC de IMGT/3Dstructure-DB ont été utilisées, sauf les structures 3D 1mhc, 2clr, 1s7s, 1s7t, 1duy et 1s7r qui se lient de manière non usuelle au MHC au niveau de leur partie N-terminale. 114 structures 3D de peptide/MHC-I et 44 structures de peptide/MHC-II ont été superposées. Le RMSD moyen entre les carbones alpha des brins des G-DOMAINs est de 0,75Å pour le MHC-I et 0,84Å pour les MHC-II.

La variabilité conformationnelle des peptides se liant au MHC-I dépend de la longueur du peptide. Le RMSD des atomes de la chaîne principale de ces peptides de MHC est pour la plupart des positions, inférieur à 2.1Å, ce qui avait déjà été noté [57]. Les peptides de 8 résidus se liant au MHC-I ont une conformation de chaîne principale très similaire (RMSD global de 0.8Å) et les peptides de 10 résidus se liant au MHC-I sont plus flexibles (RMSD global de 2.5Å), en particulier au niveau des positions 4 à 7. Les peptides de 9 résidus se liant au MHC-I ont un comportement intermédiaire (RMSD global de 1,8Å), les postions 4 à 7 étant les plus flexibles. Dans le cas des peptides se liant au MHC-II, le RMSD des atomes de la chaîne principale est inférieur à 1,4Å, ce qui souligne le peu de variabilité de la conformation de la chaîne principale du peptide.

3.5.2 Interaction peptide/MHC

Les statistiques des interactions entre la chaîne principale et les chaînes latérales à l'interface entre le peptide et le MHC révèlent que les interactions entre les atomes des



Figure 3.15: Moyenne quadratique des distances entre les carbones alpha (RMSD) de chaque position du peptide pour les peptides de longueur 8, 9 et 10 se liant au MHC-I et pour les peptides se liant au MHC-II. La distance entre les positions du peptides sont calculées après superposition des G-DOMAINs pour toutes les structures 3D étudiées. Le RMSD calculé de cette manière est représenté par les rectangles et la distance maximale observée est représentée par une barre. La RMSD moyen résultant de la superposition des G-DOMAINs est représentée par une ligne grise.

chaînes latérales du MHC avec les atomes de la chaîne principale et des chaînes latérales du peptides sont dominantes (Table 3.8). Adrian *et al.* 2002 [58] ont montré qu'à l'interface entre le peptide et le MHC-I, avec une distance maximale de contact de 3Å les interactions entre les atomes des chaînes latérales du MHC-I et les atomes de la chaîne principale du peptide sont prédominantes et que les distances de contact comprises entre 3,5 et 6Å sont principalement rencontrés pour les interactions entre les chaînes latérales du MHC-I et les chaînes latérales du peptide. Les même statistiques appliquées au MHC-II montre que les interactions à l'interface entre le MHC et le peptide s'effectuent également entre les atomes des chaînes latérales du MHC et les atomes de la chaîne principale et des chaînes latérales du peptide. Il faut noter la plus grande proportion, pour le MHC-II, des interactions entre les chaînes latérales du MHC et la chaîne principale du peptide

Les Tables 3.9 et 3.10 présentent les statistiques des interactions intervenant à l'interface entre le peptide et le MHC. Les liaisons hydrogène conservées du MHC-I apparaissent entre les positions 77 et 84 du domaine G-ALPHA1, et les positions 59 et 70 du domaine G-ALPHA2, et les extrémités du peptide. L'étude de Bouvier et Wiley 1994 [59] souligne l'importance des contacts des extrémités du peptide en analysant l'énergie de liaison provenant des interactions des acides aminés des parties N et C terminales du peptide, ce qui donne une explication du mode de liaison conservé du MHC-II. Néanmoins Achour et al. 2002 [60] montrent que les résidus terminaux ne sont pas toujours requis pour une fixation correcte.

Les liaisons hydrogène conservées interviennent le long de tout le peptide se liant au MHC-II, ce qui est en accord avec la conformation de la chaîne principale très conservée. Les positions du MHC-II qui sont les partenaires de ces liaisons hydrogène ne sont par contre pas forcément conservés. Par exemple, aucune position de MHC-II ne possède de liaison hydrogène strictement conservé avec la position 5 du peptide, et pourtant cette position effectue toujours une liaison hydrogène à l'interface. L'outil d'analyse des interactions conservés entre le peptide et le MHC permet de sélectionner un seuil de représentativité de 40%, ce qui montre que les positions 67 et 70 du G-BETA établissent des liaisons hydrogène avec la position 5 du peptide dans 55% et 41% des structures 3D de peptide/MHC-II. La position 9 du peptide dans les structures de peptide/MHC-II se comporte de façon similaire, elle établit au moins une liaison hydrogène avec le MHC-II dans la plupart des structures avec comme partenaires potentiels les positions 76 et 77 du

	MHC-I	MHC-II
Paires de positions en contacts	57	91
Positions du MHC impliquées	32	45
Paires d'atomes en contacts polaires	68	103
Paires d'atomes en liaison hydrogène	12	16
Paires d'atomes en liaison hydrogène par l'in- termédiaire d'une molécule d'eau	4	3
Paires d'atomes en contacts non polaires	514	692
Chaînes principales du MHC et du peptide	6	63
Contacts entre chaînes latérales du MHC/ principale du peptide	254	305
Contacts entre chaîne principale du $\rm MHC/$ chaînes latérales du peptide	29	93
Contacts entre chaînes latérales du MHC et du peptide	293	334

Table 3.8: Nombre moyen de contacts à l'interface entre peptide et MHC-I, et entre peptide et MHC-II. Les moyennes sont issues des statistiques de contacts aux interfaces de IMGT/3Dstructure-DB.

Table 3.9: Contacts conservés des positions des domaines G-ALPHA1 et G-ALPHA2 du MHC-I avec des peptides de 8, 9, 10, 11, 13 et 14 résidus. Les colonnes suivantes contiennent les positions du peptide en contact dans plus de 80% des interfaces avec les résidus aux positions du MHC-I de la seconde colonne. Les positions en gras signale une liaison hydrogène entre deux positions conservées dans plus de 80% des interfaces. Une croix dans la colonne H indique que le résidu à cette position du MHC-I effectue une liaison hydrogène dans plus de 80% des interfaces.

							Pepti	de		
G-DOMAIN	MHC-I	Η	8	(9	10		11	13	14
G-ALPHA1	5		1		1	1		1	1	1
	7		1 2	1 2	2	$1 \ 2$		1 2	$1 \ 2 \ 3$	$1 \ 2$
	9		5					3	$2\ 3\ 4$	$2 \ 3$
	59		1		1	1		1	1	1
	63		1 2	1 2	2	1 2	1	2	1 2	$1 \ 2$
	66		$1\ 2\ 3\ 4$	234	4	$2\ 3\ 4$	$1 \ 2$	3 4	$2\ 4\ 5\ 6\ 7\ 10$	$2\ 3\ 4$
	70		3 4 5		3		3 4	5	$2 \ 3 \ 4 \ 10$	12
	73		57	678	3	789	$5\ 6\ 7\ 9\ 10$	11	$10 \ 11 \ 12$	$9\ 12\ 13$
	77	Х	678	8 9	9	9 10	10	11	11 12 13	$12 \ 13 \ 14$
	80		8	(9	10	10	11	13	$13\ \ 14$
	81		8	(9	10		11	13	14
	84	Х	8	5 (9	10		11	13	14
G-ALPHA2	7		5				3	5	13	14
0	9		5		3	3	, in the second s	3	2 3	2 3
	26		568	()		5	11	13	14
	33		8	(9	10		11	13	14
	55	Х	8	5)	9 10	10	11	12 13	14
	58		8	8 9	9	9 10	10	11	12 13	13 14
	59	Х	6 7 8	7 8 9	98	3 9 10	9 10	11	11 12 13	11 12 13 14
	63		6		7	8		69	$3 \ 9 \ 10 \ 11$	$5\ 11$
	66		6			6	4	56	3 9 10 11	45 6 7
	67		3		3		$3 \ 4 \ 5$	69	3	3 5
	70	Х	1 2 3	123	3	1 2 3	1	$2\ 3$	1 2 3	1 2 3
	73				1	1		1 2	$1 \ 2 \ 5$	1
	77		1		1	1		1	1	1
	81		1		1	1		1	1	1

Table 3.10: Contacts conservés du peptide avec le MHC-II dans plus de 80% des interfaces. Les colonnes suivantes contiennent les positions du MHC-II en contact dans plus de 80% des interfaces avec la position du peptide de la première colonne. Les positions en gras signale une liaison hydrogène entre deux positions conservées dans plus de 80% des interfaces. Une croix dans la colonne H indique que le résidu de cette position du peptide effectue une liaison hydrogène dans plus de 80% des interfaces.

Position du peptide	Η	G-ALPHA	G-BETA
1	Х	26 34 47 60 61 62	$76 \ 77 \ 81$
2	Х	2662	72A 73 76 77
3		$24 \ 26 \ 62$	73
4	Х	70	$9\ 22\ 70\ 73$
5	Х	70	
6		$9\ 70\ 73\ 74$	7
7	Х	73 77	5963
8		73 77	58 59
9	Х	77 80 84	55 59

G-ALPHA et les positions 5 et 55 du G-BETA.

3.5.3 Sites de contacts IMGT

Les molécules de MHC-I et de MHC-II possèdent des poches de spécificité dans leur site de liaison où les chaînes latérales des peptides peuvent s'ancrer [61]. La spécificité d'un peptide à un MHC particulier est contrôlée par les propriétés physico-chimiques des poches. Inversement l'alignement des séquences de peptides et les structures 3D de pMHC révèlent que l'ancrage de certaines positions du peptide est nécessaire à la liaison d'un allèle particulier de MHC. Néanmoins certaines exceptions ont été relevées [62] [63] [56] et des travaux récents ont montrés qu'une énergie d'interaction forte entre peptide et MHC-I n'était pas nécessaire pour déclencher une réponse immune [63]. Gulukota *et al.* 1997 [64] ont fait la remarque que les motifs de liaisons détectés sur les alignements de peptide sont d'un usage très limité. En effet certains peptides ne présentant pas ce motif peuvent se lier et seulement 30% des peptides possédant le motif se lient effectivement. Pour toutes ces raisons nous avons choisi de ne pas utiliser directement la notion de poche, mais de considérer les de contacts des chaînes latérales du peptide. La comparaison de ces environnements de contacts pour des peptides de MHC-I de longueur différente et des peptides de MHC-II fait apparaître des similitudes.

Les méthodes précédentes de définition des poches [65] utilisaient essentiellement des données de polymorphisme et quelques données structurales (9 structures pour [65]). Notre définition des environnements de contacts repose par contre entièrement sur une définition structurale (114 structures de peptide/MHC-I et 44 structures de peptide/MHC-II).

Le protocole permettant la définition des sites de contacts est basé sur les analyses de contacts de IMGT/3Dstructure-DB. Pour définir les sites de contacts de chaque position du peptide, le schéma de score d'interaction suivant a été appliqué : 40 points pour les liaisons hydrogène directe, 20 points pour les liaisons hydrogène par l'intermédiaire d'une molécule d'eau, 20 points pour les interactions polaires et 1 point pour les interactions non polaires, ce qui correspond globalement aux rapports énergétiques moyens [29]. Pour une structure de peptide/MHC, la position du peptide qui donne le meilleur score d'interaction avec chacune des positions IMGT du MHC est calculée. Le cumule de ces sites de contacts pour toutes les structures 3D peptide/MHC est présenté dans la Table 3.12. Les positions du MHC qui appartiennent à une position du peptide dans moins de 5 structures ont été écartées. Après avoir identifié le premier résidu dans le site de liaison, le même protocole a été appliqué aux 9 résidus dans le site de liaison.

Table 3.11: Définition des sites de contacts de référence IMGT à partir des analyses de contacts de IMGT/3Dstructure-DB. Les sites de contacts IMGT établissent une correspondance entre les environements structuraux des chaînes latérales des peptide de taille différente et présentés par des MHC de classe différente. Ces correspondances ont été obtenues à partir des comparaisons des sites de contacts IMGT de référence (Figure 3.16).

9
2
3
4
5
6
7
8
9

La Table 3.11 présente la correspondance entre les positions du peptide et les sites de contacts pour les peptides de MHC-I et de MHC-II. Les différents sites de contacts sont définis sur la Figure 3.16. La comparaison des positions impliquées pour les différentes longueurs de peptide de MHC-I et pour les peptides de MHC-II fait clairement apparaître des similitudes qui ont permis de définir les sites de contacts IMGT. Les sites de contacts sont disponibles pour chaque complexe peptide/MHC d'une structure 3D donnée d'IMGT/3D-structure-DB et peuvent être comparés aux sites de contacts IMGT de référence (Figure 3.16).Cette comparaison permet de faire apparaître les contacts atypiques.

Table 3.12: Positions impliquées dans les sites de contacts de structures 3D pMHC (effectifs donné dans la table). Les nombres entre parenthèses désignent le nombre de fois où chaque position des G-DOMAINs du MHC participe (nombre d'occurences) au site de contacts.

MH	C-I (pe	eptide de 8 résidus) [40 structures 3D]	
Р	eptide	G-ALPHA1	G-ALPHA2
C1	1	$5_{(5)} 7_{(2)} 33_{(2)} 58_{(2)} 59_{(29)} 62_{(20)} 63_{(28)}$	$70_{(1)}$ $73_{(26)}$ $77_{(35)}$ $81_{(33)}$
		66(19)	
C3	2	$7_{(26)} \ 9_{(4)} \ 22_{(1)} \ 24_{(23)} \ 45_{(25)} \ 63_{(7)} \ 66_{(6)} \ 67_{(3)}$	$9_{(8)}$ $70_{(1)}$
		70(13)	
C4	3	67(1)	$9_{(7)} \ 24_{(14)} \ 26_{(1)} \ 63_{(14)} \ 66_{(24)} \ 67_{(40)} \ 70_{(38)}$
C5	4	$62 \scriptscriptstyle (1) \ 65 \scriptscriptstyle (1) \ 66 \scriptscriptstyle (10) \ 69 \scriptscriptstyle (2) \ 70 \scriptscriptstyle (5)$	66(4)
C6	5	$7_{(8)} 9_{(35)} 22_{(25)} 24_{(7)} 66_{(1)} 69_{(1)} 70_{(22)} 73_{(2)} 74_{(25)}$	$7_{(38)} \ 9_{(25)} \ 24_{(26)} \ 26_{(17)}$
C9	6	$73_{(2)}$	$58_{(4)}$ 59(18) 61 A(10) 63(26) 66(10)
C10	0 7	$72_{(3)} 73_{(28)} 74_{(4)} 76_{(29)} 77_{(15)} 80_{(7)}$	58(4)
C11	8	$74_{(1)}$ $77_{(25)}$ $80_{(16)}$ $81_{(39)}$ $84_{(23)}$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
			$58_{(1)}$ $59_{(22)}$
MH	C-I (pe	eptide de 9 résidus) [85 structures 3D]	
Р	eptide	G-ALPHA1	G-ALPHA2
C1	1	$5_{(8)}$ $59_{(65)}$ $62_{(41)}$ $63_{(40)}$ $66_{(19)}$	$70_{(1)}$ $73_{(69)}$ $74_{(1)}$ $77_{(79)}$ $80_{(5)}$ $81_{(78)}$
C3	2	$7_{(79)} 9_{(53)} 22_{(6)} 24_{(25)} 25_{(7)} 26_{(7)} 34_{(8)} 36_{(2)}$	9(17) 81(1)
		$45_{(70)}\ 59_{(2)}\ 63_{(44)}\ 66_{(35)}\ 67_{(58)}\ 70_{(14)}$	
C4	3	$7_{(2)} \ 9_{(10)} \ 24_{(1)} \ 45_{(1)} \ 63_{(1)} \ 70_{(1)}$	$7_{(14)} \ 9_{(60)} \ 24_{(28)} \ 26_{(2)} \ 63_{(9)} \ 66_{(17)} \ 67_{(48)}$
			$70_{(81)}$ $71_{(1)}$
C5	4	$62 \scriptscriptstyle (3) \ 65 \scriptscriptstyle (16) \ 66 \scriptscriptstyle (22) \ 69 \scriptscriptstyle (8) \ 70 \scriptscriptstyle (4) \ 73 \scriptscriptstyle (1)$	$9{\scriptstyle (1)} \ 24{\scriptstyle (1)} \ 63{\scriptstyle (1)} \ 66{\scriptstyle (8)} \ 67{\scriptstyle (1)} \ 70{\scriptstyle (1)} \ 73{\scriptstyle (1)}$
C6	5	$9{\scriptstyle (9)} \ 66{\scriptstyle (2)} \ 69{\scriptstyle (2)} \ 70{\scriptstyle (22)} \ 73{\scriptstyle (20)} \ 74{\scriptstyle (19)} \ 77{\scriptstyle (2)}$	$-7{\scriptstyle (19)} \ 9{\scriptstyle (2)} \ 24{\scriptstyle (2)} \ 26{\scriptstyle (8)} \ 59{\scriptstyle (1)} \ 61A{\scriptstyle (2)} \ 63{\scriptstyle (6)}$
			$66_{(31)} \ 67_{(21)} \ 70_{(1)}$
C8	6	$7{\scriptstyle (4)} 9{\scriptstyle (5)} 22{\scriptstyle (4)} 24{\scriptstyle (3)} 65{\scriptstyle (2)} 66{\scriptstyle (7)} 67{\scriptstyle (1)} 69{\scriptstyle (31)}$	$-7{}_{(24)} \ 9{}_{(4)} \ 24{}_{(8)} \ 26{}_{(3)} \ 45{}_{(1)} \ 59{}_{(1)} \ 61A{}_{(6)}$
		$70_{(36)}\ 72_{(1)}\ 73_{(34)}\ 74_{(16)}\ 77_{(2)}$	$62 \scriptscriptstyle (6) \ 63 \scriptscriptstyle (11) \ 64 \scriptscriptstyle (1) \ 66 \scriptscriptstyle (11) \ 67 \scriptscriptstyle (3)$
C9	7	$9_{(3)} \ 11_{(2)} \ 70_{(3)} \ 74_{(3)}$	$5_{(2)} \ 7_{(9)} \ 24_{(11)} \ 26_{(5)} \ 45_{(5)} \ 58_{(8)} \ 59_{(46)}$
			$61A_{(36)} \ 62{}_{(3)} \ 63{}_{(55)} \ 66{}_{(10)} \ 67{}_{(6)}$
C10	8	$72{\scriptstyle (11)}\ 73{\scriptstyle (29)}\ 76{\scriptstyle (62)}\ 77{\scriptstyle (11)}\ 80{\scriptstyle (24)}$	$58_{(30)} \ 61A_{(1)}$
C11	9	$70_{(2)}\ 74_{(8)}\ 77_{(70)}\ 80_{(41)}\ 81_{(81)}\ 84_{(58)}$	$5_{(56)} \ 6_{(3)} \ 7_{(5)} \ 24_{(1)} \ 26_{(65)} \ 27_{(2)} \ 28_{(1)} \ 33_{(85)}$

Fin de table à la page suivante

 $34\scriptscriptstyle (31)\ 55\scriptscriptstyle (84)\ 59\scriptscriptstyle (37)$

MHC	C-I (pe	eptide de 10 résidus) [12 structures 3D]	
Pe	eptide	G-ALPHA1	G-ALPHA2
C1	1	$5{\scriptstyle (1)}\ 58{\scriptstyle (1)}\ 59{\scriptstyle (8)}\ 62{\scriptstyle (8)}\ 63{\scriptstyle (9)}\ 66{\scriptstyle (1)}$	$73{\scriptstyle (10)}\ 76{\scriptstyle (1)}\ 77{\scriptstyle (11)}\ 81{\scriptstyle (11)}$
C3	2	$7_{(9)} \ 9_{(7)} \ 22_{(1)} \ 24_{(2)} \ 25_{(1)} \ 26_{(1)} \ 34_{(1)} \ 36_{(1)}$	$9_{(1)}$
		$45 \scriptscriptstyle (9) \ 63 \scriptscriptstyle (3) \ 66 \scriptscriptstyle (4) \ 67 \scriptscriptstyle (7) \ 70 \scriptscriptstyle (1)$	
C4	3	7(3) 66(3)	$9_{(10)} \ 24_{(7)} \ 67_{(1)} \ 70_{(12)}$
C5	4	$65_{(1)}$ $66_{(3)}$	$73_{(1)}$
C6	5	$9{\scriptstyle (1)} \ 66{\scriptstyle (1)} \ 69{\scriptstyle (2)} \ 70{\scriptstyle (6)} \ 73{\scriptstyle (6)} \ 74{\scriptstyle (6)} \ 77{\scriptstyle (3)}$	$7{\scriptstyle (6)} \ 26{\scriptstyle (5)} \ 59{\scriptstyle (2)} \ 63{\scriptstyle (1)} \ 66{\scriptstyle (2)} \ 67{\scriptstyle (4)} \ 69{\scriptstyle (1)}$
C7	6	$69_{(1)} \ 70_{(2)} \ 73_{(1)}$	$59{\scriptstyle (1)} \ 61A{\scriptstyle (2)} \ 62{\scriptstyle (1)} \ 63{\scriptstyle (3)} \ 66{\scriptstyle (6)} \ 67{\scriptstyle (2)}$
C8	7	$69 \scriptstyle (4) \ 70 \scriptstyle (2) \ 72 \scriptstyle (2) \ 73 \scriptstyle (2) \ 77 \scriptstyle (1)$	$7{}_{(3)} \ 9{}_{(1)} \ 24{}_{(2)} \ 45{}_{(1)} \ 59{}_{(1)} \ 63{}_{(1)} \ 66{}_{(1)} \ 67{}_{(2)}$
C9	8		$58_{(3)} \ 59_{(6)} \ 61A_{(8)} \ 63_{(7)} \ 66_{(3)}$
C10	9	$72{\scriptstyle (1)}\ 73{\scriptstyle (3)}\ 76{\scriptstyle (12)}\ 77{\scriptstyle (4)}\ 80{\scriptstyle (5)}$	58(7)
C11	10	$77 \scriptstyle{(4)} \ 78 \scriptstyle{(1)} \ 80 \scriptstyle{(6)} \ 81 \scriptstyle{(12)} \ 84 \scriptstyle{(10)}$	$5_{(7)} \ 7_{(1)} \ 24_{(2)} \ 26_{(7)} \ 33_{(12)} \ 34_{(2)} \ 55_{(12)} \ 59_{(2)}$
MIIC	а <u>тт</u> Га		
MHC)-11 [4·	4 structures 3D	
Pe	eptide	G-ALPHA	G-BETA
C1	1	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$77_{(35)} \ 80_{(21)} \ 81_{(35)} \ 82_{(5)} \ 84_{(17)} \ 85_{(12)} \ 86_{(1)}$
C2	2		$70{\scriptstyle (1)}\ 72{\scriptstyle (1)}\ 72A{\scriptstyle (39)}\ 73{\scriptstyle (15)}\ 76{\scriptstyle (36)}\ 77{\scriptstyle (8)}$
C3	3	$7_{(17)} \ 7A_{(3)} \ 24_{(39)} \ 26_{(5)} \ 62_{(15)} \ 63_{(3)} \ 66_{(28)}$	
		$67({}^{24})$ $69({}^{9})$ $70({}^{7})$	
C4	4	$7_{(21)}$ $7A_{(4)}$	$8{\scriptstyle (1)} \ 9{\scriptstyle (41)} \ 10{\scriptstyle (3)} \ 11{\scriptstyle (11)} \ 22{\scriptstyle (39)} \ 23{\scriptstyle (2)} \ 24{\scriptstyle (21)}$
			$26_{(1)} 66_{(16)} 67_{(28)} 70_{(35)} 72A_{(1)} 73_{(29)}$
			$74_{(10)}$
C5	5	$66_{(3)} 69_{(4)} 70_{(6)}$	$63_{(4)}$ $66_{(14)}$ $70_{(3)}$
C6	6	$7{\scriptstyle (3)} \hspace{0.2cm} 9{\scriptstyle (38)} \hspace{0.2cm} 69{\scriptstyle (5)} \hspace{0.2cm} 70{\scriptstyle (31)} \hspace{0.2cm} 71{\scriptstyle (4)} \hspace{0.2cm} 73{\scriptstyle (27)} \hspace{0.2cm} 74{\scriptstyle (42)}$	$5_{(7)} \ 6_{(1)} \ 7_{(40)} \ 9_{(3)} \ 24_{(2)} \ 26_{(12)} \ 27_{(1)} \ 33_{(1)}$
		77(7)	67(1)
C9	7		$24_{(11)} \ 26_{(11)} \ 45_{(32)} \ 59_{(33)} \ 61B_{(4)} \ 63_{(33)}$
			$66_{(7)}$ $67_{(12)}$ $70_{(3)}$
C10	8	$72 \scriptscriptstyle (2) \ 73 \scriptscriptstyle (17) \ 76 \scriptscriptstyle (28) \ 77 \scriptscriptstyle (2) \ 80 \scriptstyle (1)$	$58_{(2)} \ 61B_{(2)}$
C11	9	$76_{(3)} \ 77_{(34)} \ 78_{(1)} \ 80_{(32)} \ 81_{(32)} \ 84_{(29)}$	$5_{(26)} \ 26_{(4)} \ 27_{(1)} \ 33_{(12)} \ 34_{(3)} \ 54_{(1)} \ 55_{(36)}$
			$58_{(2)}$ $59_{(11)}$



Figure 3.16: Sites de contacts de références IMGT, de C1 à C11, pour les peptides de 8 acides aminés lié à un MHC-I, les peptides de 9 acides aminés lié à un MHC-I, les peptides de 10 acides aminés lié à un MHC-I, et pour les peptides liés aux MHC-II. Les sites de contacts désignent pour chaque position du MHC à l'interface, la position du peptide avec laquelle il effectue le contact le plus énergétique. Ces figures ont été établies à partir du cumule des sites de contacts (Table 3.12). Pour les positions du MHC participant à plusieurs sites de contacts dans cette table, la participation à un site de contact a été écartée quand son occurrence est inférieure à la moitié de l'occurrence maximale (modifié d'après Kaas et Lefranc 2005 [33], Publication 6).

3.6 Interface des complexes TR/peptide/MHC

Plusieurs études ont tenté de déterminer les règles globales gouvernant la formation des complexes TR/peptide/MHC [66] [67] [68]. Les principes généraux de l'interaction ont été définis à partir des premières structures 3D, le TR s'oriente de façon diagonale par rapport au peptide et sa position par rapport aux extrémités est variable [69]. Sachant que la surface du MHC n'est pas plane, les deux points les plus élevés, situés entre les deux hélices de chaque G-DOMAIN, contraignent stériquement l'approche du TR vers le peptide [39] [70]. Les positions du TR et du MHC varient et jusqu'à présent aucun principe permettant de prédire l'orientation précise du TR n'a pu être découvert [71]. En effet chaque structure 3D de TR/peptide/MHC présente un mode de liaison du CDR3 différent. Les deux CDR3 forment une cavité qui est le site normal de liaison de l'antigène pour les V-PARTNER des IG. Mais ce site n'est pas forcément utilisé dans le cas des TR, et la cavité peut rester inoccupée [70]. Dans le complexe JM22/peptide/HLA-A [71], le CDR3 du V-BETA se place dans une ouverture entre le peptide et le MHC, tandis que dans le complexe 2C/peptide/H2-K1 [72] le CDR3 du V-BETA possède peu de contacts avec le peptide. Généralement la partie centrale des peptides liées au MHC-I est accessible au TR, mais dans la structure loga [71], le peptide expose principalement la chaîne latérale de la position P8 située en bout de peptide. Actuellement 35 structures 3D de complexe TR/peptide/MHC ont été déterminées expérimentalement. Elles ont toutes été utilisées pour cette étude.

3.6.1 Orientation du TR



Figure 3.17: Orientation des TR dans les structures RX par rapport au MHC. Les extrémités des segments sont les projetés sur le plan du plancher des centres des deux V-DOMAINs de chaque TR. Le rond noir au centre de chaque segment symbolise le centre du V-PARTNER. les différents sigles délimitant les segments désignent les différents TR dont le nom est précisée dans la légende.

L'angle formé entre le TR et le MHC a été calculé en définissant deux plans. Un plan

correspondant au plancher du MHC, défini par deux vecteurs dont les extrémités sont les carbones alpha des positions 27 et 21 de chaque G-DOMAIN. Un plan globalement perpendiculaire aux feuillets des V-DOMAINs du TR, défini par deux vecteurs dont les extrémités sont les carbones alpha des positions 47 et 104 de chaque V-DOMAIN. L'orientation entre le TR et le MHC est mesuré par l'angle entre le premier vecteur du plan du MHC et le projeté orthogonal du premier vecteur du plan du TR sur le plan du MHC. Cet angle varie entre 26 et 75 degrés (Table 3.13, et Kaas *et al.* 2005 [53] Publication 5). La position du projeté orthogonal sur le plan du MHC du point à équidistance des positions 104 des V-DOMAINs est également calculé (Figure 3.17). Il avait été postulé que les MHC-I et MHC-II utilisaient des zones angulaires de liaison différentes, mais la structure 10ga [71] qui présente un MHC-I se liant avec un angle de 70 degrés situé dans la zone des MHC-II démontre qu'il existe déjà des exceptions à cette règle.

Table 3.13: Angles en degrés formés par le projeté de la droite passant par le centre des V-DOMAIN du TR et celui de la droite passant par l'axe du sillon du MHC. ID désigne l'identifiant du fichier de coordonnées.

MHC-I		MH	[C-I	MHC-II		
ID	Angle	ID	Angle	ID	Angle	
1ao7	41	1g6r	27	1d9k	65	
1qrn	41	1jtr	27	$1\mathrm{j}8\mathrm{h}$	53	
1 qse	42	$2\mathrm{ckb}$	29	$1 \mathrm{fyt}$	53	
1qsf	43	$1 \mathrm{mwa}$	27			
1 b d 2	57	1 fo 0	42			
1lp9	75	$1 \mathrm{nam}$	42			
10ga	71	1kj2	34			
1 mi 5	45					

3.6.2 Surface enfouie

La surface enfouie du TR et du pMHC est comparable (Table 3.14), ce qui s'explique par la planéité de l'interface d'interaction. L'implication relative des V-DOMAINs à l'interface est variable. Dans 1ao7 le V-ALPHA possède une surface enfouie deux fois plus importante que le V-BETA, par contre dans 1oga le V-BETA possède une surface enfouie deux fois plus importante que le V-ALPHA, et dans 1kj2 la surface enfouie des deux domaines est pratiquement identique. La surface d'interaction entre le TR et le pMHC est une des plus petites parmi les complexes dont la structure 3D est connue [73], ce qui est en accord avec la faible affinité du complexe [74] (K_D compris entre 10⁻⁴ et 10⁻⁶M).



Figure 3.18: Positions du MHC en contact avec les 6 CDR-IMGT dans toutes les structures 3D de complexes TR/peptide/MHC connues.

3.6.3 Positions en interactions

Les analyses de contacts de IMGT/3Dstructure-DB ont été utilisées pour explorer les contacts par paires intervenant aux interfaces entre le TR et le MHC, et entre le TR et le peptide. Certaines positions du MHC et du TR sont impliquées dans toutes les structures connues de pMHC/TR mais le détail des paires de positions en interaction n'est pas conservé. Toutes ces positions appartiennent aux hélices pour le MHC (Figure 3.18). Les positions du TR participant à l'interface font essentiellement partie des CDR mais aussi à la boucle ED, qui est parfois appelé HV4 (*HyperVariable 4*) et aux positions d'ancrages des boucles (positions 26, 66 et 67). Pour chaque structure RX, les paires de positions en interactions montre que le CDR2 du V-ALPHA se place au niveau de l'hélice du MHC-II G-ALPHA2, et que le CDR2 du V-BETA se place au niveau de l'hélice du MHC-II G-ALPHA ou du MHC-I G-ALPHA1. Comme il a été vu précédemment lors de l'examen de la surface enfouie, la contribution relative des deux V-DOMAINs n'est pas constante. Par exemple, dans la structure du complexe

Table 3.14: Surfaces enfouies en $Å^2$ par les V-DOMAIN du TR et les G-DOMAIN du MHC, ainsi que des récepteurs en entier. Elles ont été obtenues en effectuant la différence entre la surface accessibe du domaine ou du récepteur isolément et dans le complexe complet TR/peptide/MHC. ID désigne l'identifiant du fichier de coordonnées.

MHC	C-I					
ID	TR	V-ALPHA	V-BETA	MHC	G-ALPHA1	G-ALPHA2
1ao7	1202,0	781,3	420,7	1141,9	766,3	375,6
$1 \mathrm{qrn}$	1181, 1	737,0	444,1	$1139,\!3$	738,1	401,2
1 qse	$1185,\!8$	742,2	$443,\! 6$	1120,7	724,2	396,5
1qsf	1090,2	790,9	$299,\!3$	$1041,\!4$	$796,\!3$	245,1
$1 \mathrm{bd}2$	$1049,\! 6$	685,1	364,5	$1054,\!3$	$717,\! 6$	336,7
10ga	$943,\!9$	309,2	634,7	$870,\!3$	287,5	$582,\!8$
1 mi 5	$1319,\!0$	737,5	581,5	$1239,\!8$	$726,\! 0$	$513,\!8$
1g6r	$1032,\!3$	$535,\!4$	$496,\!9$	$1059,\!9$	$569,\!4$	490,5
1lp 9	$1254,\!2$	674,2	$580,\! 0$	$1107,\! 0$	$620,\!4$	$486,\! 6$
1jtr	$1151,\!1$	$613,\!9$	537,2	1110,7	$600,\!3$	510,4
$2 \mathrm{ckb}$	$1099,\!4$	621, 1	$478,\!3$	$1162,\! 0$	$653,\!2$	508,8
1mwa	$1151,\!1$	$613,\!9$	$537,\!2$	1110,7	$600,\!3$	510,4
1 fo 0	$797,\!8$	309,9	$487,\!9$	$763,\! 0$	$329,\!3$	433,7
1nam	$1014,\!3$	541,2	$473,\!1$	945,2	$500,\! 6$	$444,\! 6$
1kj2	$1065,\!3$	552,7	$512,\! 6$	$1016,\!3$	489,5	526,8
MHC	C-II					
1 d9k	1220, 1	$723,\!8$	496, 3	1154,7	659,2	495,5
1j 8 h	$1259,\!6$	608,1	$651,\!5$	$1135,\!5$	$532,\!4$	603,1
1fyt	$1270,\!5$	597,7	$672,\!8$	$1181,\!4$	$521,\!4$	660,0

TR/MHC		
Paires de positions en contacts	43	
	V-AL PHA	V-BETA
Positions du TR impliquées	13	11
	G-ALPHA1/G-ALPHA	G-ALPHA2/G-BETA
Positions du MHC impliquées	11	12
Paires d'atomes en contacts polaires	48	
Paires d'atomes en liaison hydrogène	4	
Paires d'atomes en liaison hydrogène par l'in- termédiaire d'une molécule d'eau	2	
Paires d'atomes en contacts non polaires	275	
$\mathrm{TR}/\mathrm{peptide}$		
Paire de positions en contacts	20	
	V-AL PHA	V-BETA
Positions du TR impliqués	6	6
	V-ALPHA	V-BETA
Positions du peptide impliqués	4	4
Paires d'atomes en contacts polaires	25	
Paires d'atomes en liaison hydrogène	4	
Paires d'atomes en liaison hydrogène par l'in- termédiaire d'une molécule d'eau	2	
Paires d'atomes en contacts non polaires	131	

Table 3.15: Nombre moyen de contacts entre le TR et le MHC. Les moyennes sont issues des statistiques de contacts aux interfaces de IMGT/3Dstructure-DB.

LC13/peptide/HLA-B (1m05), les deux V-DOMAINs ont une contribution quasiment symétrique. Par contre, dans le cas du complexe JM22/peptide/HLA-A (1oga), le V-BETA accompli quasiment tous les contacts de l'interface.

Les interactions entre les chaînes latérales sont prépondérantes et les interactions entre les chaînes latérales et la chaîne principale sont également très présentes. Le nombre moyen de liaisons hydrogène est pratiquement identique à l'interface TR/peptide et TR/MHC, ce qui signifie que le TR est aussi spécifique du MHC que du peptide (Table 3.15).

Les principes des interactions TR/peptide/MHC sont loin d'avoir été compris. L'étude des interactions entre le TR et le MHC avec les autres molécules de la synapse immunologique pourraient contribuer à éclaircir ce problème. La compréhension du déclenchement des cellules T fait le sujet de recherches actives [75]. La structure 3D 1jl4 [76], par exemple, est la première structure présentant une interaction entre un MHC et un CD4, un corécepteur qui augmente la sensibilité du complexe TR/pMHC. Le corécepteur NKG2D est également très étudié car il est impliqué dans la prolifération des tumeurs. Finalement la transduction du signal du récepteur T n'est pas encore très claire et la vision actuelle est que la transduction est réalisée soit par un réarrangement conformationnel des C-DOMAINs du TR [47], soit par un mouvement mécanique de la membrane [75].

3.7 Conclusion

Grâce à la standardisation des données structurales introduite dans IMGT/3Dstructure-DB, les comparaisons de structure 3D de récepteurs d'antigènes à grande échelle sont désormais possibles. J'ai ainsi pu réaliser des statistiques sur la représentativités des différents gènes, récepteurs et complexes dans IMGT/3Dstructure-DB. J'ai mis en place des analyses des structures 3D des domaines V, C et G au niveau de leur topologie, de leur composition en acides aminés, de l'arrangement des feuillets dans le domaine, des variations de leur conformation et de leurs contacts internes. L'arrangement des domaines dans les récepteurs a également été le sujet d'analyses statistiques. Les interfaces entre les antigènes et les récepteurs d'antigènes ont été observées par des analyses des conformations (flexibilité du peptide dans le MHC, flexibilité des CDR des IG, orientation du TR sur le MHC), des surfaces enfouies et des interactions.

J'ai mis en place un environnement informatique sous la forme de modules Perl facilitant (1) la sélection de structures 3D dans IMGT/3Dstructure-DB, (2) l'extraction des coordonnées des atomes d'un récepteur, d'une chaîne ou d'un domaine et de l'antigène associé, (3) les alignements de séquences entre domaines, (4) la superposition des structures 3D, (5) la création de graphiques d'effectifs de population et (6) le stockage des données statistiques à jour dans des tables de IMGT/3Dstructure-DB. La sélection des structures 3D utilise des critères de résolution, du nombre de positions sans coordonnées, et de positions clés sans coordonnées, ainsi que l'algorithme de regroupement *hclust*, utilisant la méthode *complete*, de l'environnement statistique R. Les alignements de séquences entre domaines utilisent la numérotation IMGT de chaque position du domaine dans IMGT/3Dstructure-DB. La superposition des domaines utilise le programme ProFit auquel est fourni un alignement de séquences basé sur l'alignement des domaines. Certaines positions, fournies en paramètres, peuvent être exclues de la superposition structurale. Le module de création des Colliers de Perles ainsi que les modules de gestion de IMGT/3Dstructure-DB sont également accessibles depuis cet environnement. Tous ces développements doivent également permettre une écriture simplifiée de nouveaux scripts d'analyse.

J'ai réalisé une interface Internet afin de rendre toutes ces données accessibles. Dans l'avenir, il devrait être possible de coupler directement cette interface Internet avec les modules facilitant les analyses, ce qui permettra à un utilisateur de développer lui même une requête spécifique. La mise à dispositions de ces données standardisées et automatiquement mises à jour présente un intérêt d'autant plus important que le nombre de structures tridimensionelles est en augmentation rapide. Afin de maintenir une cohérence sur la description des données structurales des récepteurs d'antigènes, la base de données IMGT/3Dstructure-DB et ses analyses statistiques associées apparaissent donc comme indispensables car elles fournissent un socle commun. Elles peuvent servir aussi bien au cristallographe désireux de comparer les propriétées d'une nouvelle structure de récepteur d'antigène qu'à un bioinformaticien cherchant à modéliser la structure des complexes impliquant un récepteur d'antigène.

Les données statistiques constituent une source d'information importante pour la construction et l'évaluation de modèles structuraux.

Chapitre 4 Modélisation peptide/MHC et TR/peptide/MHC.

Les chercheurs en immunologie travaillant sur une pathologie spécifique doivent s'intéresser à un grand nombre de peptides, de MHC et de TR. Pour les maladies auto-immunes, l'identification des peptides antigéniques est difficile car des classes entières de protéines sont potentiellement impliquées. De même, de vastes répertoires de cellules T peuvent reconnaître un peptide/MHC donné, si bien que seuls des cas de répertoires restreints sont analysables [77]. Le nombre important de TR et de peptides a étudier rend l'analyse expérimentale systématique de leurs interactions impossible. Nous avons tenté de fournir une solution à ce problème par une approche entièrement structurale reposant sur la modélisation rapide des structures 3D des complexes peptide/MHC et TR/peptide/MHC. Pour cela nous avons prévu un protocole de modélisation (Figure 4.1) qui sélectionne des structures 3D de TR et/ou de MHC dans IMGT/3Dstructure-DB. Ces structures servent ensuite de références pour la construction de modèles de peptide/MHC-I ou de peptide/MHC-II, et également à la construction de modèles de TR avant de constituer un complexe TR/peptide/MHC. Nos efforts se sont concentrés dans un premier temps sur le développement de techniques rapides de génération de conformations grâce à un algorithme de construction des chaînes latérales, SCHISMo, et un algorithme de construction de boucles, LLIPA. Comme le montre la Figure 4.1, l'évaluation des modèles n'a pas encore été développée mais quelques essais de classification des modèles de peptide/MHC-II utilisant l'identification des interactions entre atomes de SCHISMo sont présentés. L'étape de docking de la structure du TR sur celle du peptide/MHC utilisera les représentations simplifiées hiérarchiques de chaînes latérales de SCHISMo, afin de limiter le nombre de centres d'interactions et aussi pour prendre en compte la flexibilité des chaînes latérales (soft docking [78]). La flexibilité des boucles pourra également être considérée par une approche multicopie [79] grâce à l'ensemble de boucles générés par LLIPA lors de la modélisation du TR isolé. Les méthodes prédictives existantes de liaison peptide/MHC vont être d'abord présentées, suivis de la modélisation des structures 3D de peptide/MHC-II par l'algorithme SCHISMo et de la méthode de prédiction des structures 3D de peptide/MHC-I et des CDR de TR par l'algorithme LLIPA. Pour ces deux dernières parties, une bibliographie sera développée conjointement à la description de SCHISMo et LLIPA, suivie par quelques résultats de validation.

A terme ces outils de modélisation devront pouvoir être disponibles sur le serveur de IMGT/3Dstructure-DB, constituant ainsi une approche assez complète des structures 3D de récepteurs d'antigènes, comprenant une base de données, des analyses statistiques et des outils de modélisation.



Figure 4.1: Protocole de modélisation prévu pour les structures 3D de peptide/MHC-I, peptide/MHC-II et TR/peptide/MHC. L'algorithme de modélisation des boucles LLIPA est en bleu. L'algorithme de modélisation des chaînes latérales SCHISMo est en vert. Ce dernier est également utilisé dans la modélisation de LLIPA et pourra être employé dans le docking du TR et du peptide/MHC. Les étapes en noir ne sont pas encore achevées ou développées.

Sommaire

4.1	Préc	liction de la liaison peptide/MHC – Bibliographie \ldots .	88
	4.1.1	Alignements de séquences de peptides	88
	4.1.2	Threading	90
	4.1.3	Construction de la structure 3D du complexe peptide/MHC	91
4.2	Mod	lélisation des peptide/MHC-II \ldots	92
	4.2.1	Approximation de l'espace conformationnel des chaînes latérales	93
	4.2.2	Stratégie de recherche	98
	4.2.3	Evaluation des conformations	102
	4.2.4	$ m R\acute{e}sultats$	102
4.3	Mod	lélisation des peptide/MHC-I et des boucles du TR $-$	
	\mathbf{Con}	struction de boucles par la méthode LLIPA	107
	4.3.1	Prédiction de la conformation des boucles – Bibliographie	107
	4.3.2	Algorithme LLIPA	111
	4.3.3	Modélisation de peptides dans les structures RX	115
	4.3.4	Modélisation des boucles de reconnaissance du TR \ldots	115
4.4	Con	clusion	117

4.1 Prédiction de la liaison peptide/MHC – Bibliographie

Plusieurs serveurs utilisant des algorithmes très différents ont été mis au point pour prédire l'interaction des peptides avec les MHC-I et les MHC-II. Les techniques de prédiction peuvent être classées en 3 grands groupes : les techniques basées sur des alignements de séquences de peptides, celles basées sur l'alignement d'une séquence de peptide sur une structure pMHC (*threading*) et enfin celles s'appuyant sur la construction de la structure 3D du complexe peptide/MHC. Les méthodes de prédiction permettent d'obtenir différents niveaux de détail : (1) prédiction binaire (liaison ou non au MHC), (2) prédiction de groupes discrets d'affinité et (3) prédiction de l'énergie libre d'interaction.

4.1.1 Alignements de séquences de peptides

Positions d'ancrage

Des bases de données spécialisées (SYFPEITHI [80] [81], JenPep [82], MHCPEP [83]) présentent les séquences de peptides connus pour se lier à des allèles de MHC. Ces serveurs proposent des profils de séquences identifiant les acides aminés ou groupes d'acides aminés conservés à chaque position du peptide. Les positions conservées du peptide sont appelées positions d'ancrage. L'utilisation de ces positions d'ancrages est la méthode la plus simple et la plus directe de prédiction binaire de la liaison des peptides à un allèle de MHC

donné. Gulukota *et al.* 1997 [64] font la remarque que l'utilisation des motifs de séquence de peptide est très limitante puisque certains peptides ne présentant pas le motif identifié peuvent néanmoins se lier et déclencher la réponse immune, et inversement que seulement environ 30% des peptides possédant le motif de liaison se lient effectivement.

Matrice d'évaluation dépendant de la position

PSSM Une quantification plus fine de la liaison est obtenue par des matrices d'évaluation de la concordance entre un type d'acide aminé et une position du peptide (PSSM, *Position Specific Scoring Matrix*). Plusieurs méthodes de dérivation des PSSM existent. La plus simple évalue le rapport des fréquences d'un type d'acide aminé à une position donnée dans un groupe de peptides connus pour se lier et dans un groupe de peptide connus pour ne pas se lier [84]. Ce rapport constitue une mesure de la concordance entre un type d'acide aminé et une position du peptide, et la somme de cette mesure sur l'ensemble du peptide permet d'évaluer son affinité. Parmi les serveurs utilisant une PSSM, on peut citer RANKPEP [85] qui dérive ses matrices d'évaluation grâce aux programmes PROFILEWEIGHT [86] ou BLK2PSSM (dans la suite de programmes BLIMPS [87]).

Matrices quantitatives Plusieurs groupes se sont intéressés à l'optimisation de PSSM par rapport à des valeurs expérimentales (la matrice est alors appelée matrice quantitative). Gulukota et al. 1997 [64] utilisent une méthode simple de dérivation de PSSM pour des peptides se liant au HLA-A2. Le score attribué à un acide aminé occupant une position est la moyenne des logarithmes des IC_{50} des peptides qui possèdent le même acide aminé à la même position. Parker et al. 1994 [88] dérivent des PSSM, employés dans le serveur ProPred [89], par minimisation linéaire par rapport au temps de demi-vie d'association. Etant donné le grand nombre de facteurs (acides aminés, positions) et du faible nombre de données expérimentales, cette méthode produit des matrices très différentes suivant le jeu d'alignements employé [90]. La méthode PLS (Partial Least Square) effectue une régression sur une matrice réduite (matrice des vecteurs propres) et est beaucoup plus robuste dans les cas ou les variables sont plus nombreuses que les données. Elle a été employée par Doytchinova et Flower 2003 [90] sur des alignements de peptides de MHC-II HLA-DRB1*0401 de JenPep [82] en corrélation avec les IC_{50} . Le serveur MHCPred [91] prédit quantitativement la liaison d'un peptide en se basant sur des matrices dérivées par PLS. Une autre méthode permettant de gérer le grand nombre de paramètres est l'analyse discriminante itérative (SDA, Stepwise Discriminant Analysis) qui sélectionne les variables les plus informatives. Cette technique a été employée par Mallios 2001 [92] sur des peptides se liant à HLA-DR1 extraits de MHCPEP [83]. L'approche de Udaka et al. 2000 [93] est beaucoup plus expérimentale et consiste à effectuer expérimentalement un scan positionnel en utilisant une librairie combinatoire synthétique. L'affinité (K_B) de chaque peptide complexé avec un MHC H2-Kb, H2-Db et H2-Ld a ainsi été mesurée. En supposant l'indépendance de liaison de chaque position, la contribution à l'affinité de chaque type d'acide aminé à une position donnée peut être déterminée. La matrice quantitative ainsi obtenue fournit des scores globalement corrélés avec l'affinité expérimentale. Pourtant certains complexes présentent une dérive assez importante à cause de la dépendance de liaison entre certaines positions quand elles sont occupées par certains acides aminés.

Classification non supervisée Les méthodes de classification non supervisées permettent également de traiter les cas où il existe un grand nombre de variables mais peu de données. Les différentes approches existantes comprennent les modèles de Markov cachés HMM (Predict [94]), les réseaux de neurones ANN [64] [95] et les machines à vecteurs de supports SVM (SVMHC [96]).

Combinaison de techniques Le recours à plusieurs techniques permet en général d'atteindre une meilleure sélectivité. Gulukota *et al.* 1997 [64] et Altuvia *et al.* 1995 [97] combinent deux approches, les réseaux de neurones et l'utilisation de matrices quantitatives. NhlaPred (http://www.imtech.res.in/raghava/nhlapred/) utilise également une combinaison de PSSM (ComPred) et de réseaux de neurones (ANNPred). Les peptides sont identifiés comme se liant fortement, modérément, faiblement ou ne se liant pas.

Toutes ces méthodes basées sur les alignements de séquences de peptides sont spécifiques d'alignements obtenus pour chaque allèle de MHC, ce qui restreint leur utilité aux allèles les plus étudiés expérimentalement. L'établissement de corrélations entre la nature des acides aminés du peptide et les positions polymorphiques du MHC [98] permet d'avoir une approche plus générale. Le manque de données expérimentales et surtout la supposition de l'indépendance de liaison de chaque position du peptide restent toujours problématique.

4.1.2 Threading

La technique du *threading* consiste à placer dans la structure 3D d'une protéine les acides aminés d'une séquence d'une autre protéine aux positions jugées équivalentes. C'est une technique très rapide car la structure 3D n'est pas reconstruite, seules les informations de proximité issues de la structure 3D sont employées pour évaluer la conformité entre la structure et la séquence. Une matrice de substitution dépendant de la position (PSSM) dans la structure 3D est ainsi calculée.

Altuvia *et al.* 1995 [97] utilisent le potentiel Miyazawa et Jernigan (MJ) [99] [100] qui attribue une valeur fixe aux contacts entre deux types d'acides aminés (le centre des chaînes latérales et une distance maximale de contact de 6,5Å). Betancourt et Thirumalai 1999 [101] dérivent un potentiel basé sur celui de MJ mais qui prend mieux en compte les interactions hydrophiles, ce qui leur permet d'avoir une meilleur sélectivité [102] (serveur PREDEP http ://margalit.huji.ac.il/). MHC-Thread (Brooks 1999 University of Aberdeen, http ://www.csd.abdn.ac.uk/~gjlk/MHC-Thread/) est un serveur utilisant le *threading* sur des HLA-DRB1, mais, la méthode n'ayant pas été publiée, les critères de sélection restent inconnus. Les inconvénients majeurs de la technique du *threading* est de ne pas considérer la flexibilité des molécules ainsi que les variations des contacts causées par la mutation entre acides aminés de taille différente. Ainsi le *threading* employé sur une structure de MHC-I est approximatif sur les positions centrales du peptide [97] car le peptide y présente sa plus grande variabilité structurale (voir 3.5.1).

4.1.3 Construction de la structure 3D du complexe peptide/MHC

Les méthodes de construction explicite des coordonnées des atomes du peptide se divisent en deux catégories, celles qui construisent les chaînes latérales du peptide tout en prenant les coordonnées de la chaîne principale dans une structure expérimentale, et les méthodes qui reconstruisent complètement l'ensemble des atomes en ne se servant que de la position approximative des résidus extrêmes (similaire à la construction d'une boucle). Après construction du modèle, l'énergie d'interaction est évaluée par différentes méthodes afin de pouvoir classer les peptides.

Construction des chaînes latérales La constructions des chaînes latérales peut s'effectuer par optimisation soit dans l'espace continu des angles dièdres [103], ce qui s'avère très long, soit en utilisant des conformères, des conformations discrètes des chaînes latérales. Plusieures approches ont été réalisées en employant des programmes généraux de construction des chaînes latérales. Ogata et al. 2002 [104] fait appel au programme DESIGNER [105] qui utilise la bibliothèque de rotamères de Bower et al. 1997 [106] pour modéliser des peptides dans HLA-A2. Kangueane et al. 2000 [107] et Lee et McConnell 1995 [108] modélisent des peptides dans le MHC-I et le MHC-II grâce au programme CARA de la suite LOOK (Molecular application group, Palo alto, CA) qui implémente une optimisation d'ensemble cohérente (SCEO [109] qui est un recuit simulé¹ de plusieures conformations simultanément). Certains groupes ont dérivé des banques de rotamères spécifiques des MHC-I [110] [57] et ils observent une amélioration de leurs prédictions.

Diverses fonctions évaluant l'énergie d'interaction entre le peptide et le MHC, utilisant des combinaisons entre des termes de solvatation, électrostatique et de mécanique moléculaire, ont été employées. Vasmatzis *et al.* 1996 [103] ont recours à un terme de solvatation [111] et à un potentiel électrostatique de Coulomb. Rognan *et al.* 2000 [110] utilisent la fonction de mécanique moléculaire de CHARMM [112] et un terme de solvatation dépendant de la surface accessible. Schueler-Furman *et al.* 1998 [57] utilisent les termes de van der Waals et électrostatiques du champ de force OPLS [113]. Kangueane *et al.* 2000 [107] n'utilisent pas de fonction d'énergie de mécanique moléculaire mais comptabilisent le nombre de chocs stériques rencontrés et le nombre de résidus hydrophobes exposés.

Construction de tout le peptide Les méthodes construisant entièrement les peptides utilisent diverses méthodes de placement des résidus terminaux. Tong *et al.* 2004 [114] ont recours au placement par mouvement brownien et Rosenfeld *et al.* 1993 [115] au docking par copies multiples. Desmet *et al.* 1997 [116] construisent le peptide entre les

¹Le recuit simulé est une méthode de recherche conformationnelle consistant à 'chauffer' virtuellement les molécules afin de leur donner suffisamment d'énergie pour franchir des barrières énergétiques, puis à refroidir le système. Le mot 'recuit' provient de la technique de sidérurgie consistant à chauffer du métal à blanc puis à le refroidir lentement afin de le rendre plus solide.

résidus d'ancrages par croissance de fragments et les chaînes latérales par l'algorithme *Dead End Elimination* [95]. Sezerman *et al.* 1993 [117] utilisent le programme de fermeture de boucle CONGEN [118] et Tong *et al.* 2004 [114] utilisent l'algorithme implémenté dans MODELLER [119].

Lorsque plusieurs modèles ont été construits, leur classement a été réalisé soit par regroupement des conformations et en considérant la population dans chaque groupe [114], soit en calculant l'énergie d'interaction de type mécanique moléculaire, par exemple avec CHARMM [112] (utilisé par Desmet *et al.* 1997 [116]) ou avec ICM [120] (utilisé par Tong *et al.* 2004 [114]). Un potentiel d'énergie empirique spécifique des interactions peptides-MHC (Fresno) utilisant 5 termes – liaison hydrogène, contacts lipophile, rotationnel, enfouissement des atomes polaires et désolvatation – a été dérivé [121] et semble correctement corrélé avec l'énergie libre expérimentale. Cette fonction d'énergie permet une meilleure classification des peptides que les fonctions d'énergie plus générale (Chemscore, Dock, FlexX, Gold, Pmf et Score) [122].

4.2 Modélisation des peptide/MHC-II - Prédiction de chaînes latérales par la méthode SCHISMo

L'analyse de la conformation du peptide dans le site de liaison du MHC-II a montré que la conformation de la chaîne principale du peptide était très conservée (section 3.5.1 page 70). La modélisation de la structure 3D du peptide dans le MHC-II consiste donc à prédire l'arrangement des chaînes latérales du peptide et du MHC-II. Nous avons développé la technique SCHISMo (*Side Chain prediction using a Hierarchy of Simplified Models*) de modélisation rapide des chaînes latérales qui s'applique à l'obtention d'un grand nombre de structures de peptides dans un MHC-II. Cette technique, de part sa rapidité, peut également être intégrée dans d'autres algorithmes de modélisation (docking, construction de boucles), afin de simuler la flexibilité des chaînes latérales.

SCHISMo emploie une combinaison de représentations simplifiées des protéines [123] [124]. Les angles de valence et la taille des liaisons covalentes sont maintenus à leurs valeurs idéales. Des groupes d'atomes sont représentés par un seul centre d'interaction. L'espace des angles dièdres est parcouru de façon discrète mais sans se limiter aux rotamères. La prédiction de la conformation des chaînes latérales d'une protéine est un problème d'optimisation particulièrement difficile. Depuis plus de 20 ans, de nombreuses méthodes ont été appliquées à ce problème. Les techniques employées peuvent être classées en 3 grands thèmes, les approximations de l'espace des conformations, l'évaluation des conformations et la stratégie de recherche. Pour chacun de ces points, une bibliographie succincte des méthodes existantes est présentée, suivie de l'approche qui a été développée ou choisie. Le nombre de techniques développées par la communauté scientifique est très important mais peu ont abouti à un programme librement accessible et facile d'utilisation. Parmi ceux-ci on peut cité SCAP (http ://honiglab.cmpc.columbia.edu, [125]), SMD [126] et SCWRL [106] [127] qui est de loin le plus utilisé en raison de sa facilité d'utilisation et de sa rapidité.

4.2.1 Approximation de l'espace conformationnel des chaînes latérales

Bibliographie

La détermination de l'ensemble des conformations des chaînes latérales d'une protéine classique dans l'espace continu des conformations est encore au delà de la puissance de calcul actuellement accessible. Les techniques d'exploration dans l'espace continu permettent de déterminer des changements de conformations locales causées par des mutations de chaînes latérales [128]. L'observation des conformations des chaînes latérales dans les structures RX a montré l'existence d'un nombre limité de formes canoniques, appelées rotamères [129], correspondant aux conformations de basse énergie. Une approximation couramment employée est de ne considérer que les rotamères de chaque chaîne latérale, ce qui simplifie énormément la recherche conformationnelle [129] [126] [130] [106] [131] [125]. Pourtant de 5 à 30% des conformations de chaînes latérales suivant le type de l'acide aminé observé dans les cristaux de bonne résolution ne se conforment pas aux structures rotamériques canoniques en raison de contraintes structurales locales [132]. Ces conformations hors rotamères canoniques sont particulièrement importantes quand il s'agit d'établir une corrélation entre les énergies de packing et la stabilité [133] [109]. Mendes et al. 1999 [134] ont développé un modèle de rotamères flexibles avec un ensemble continu de conformations qui se regroupent autour des rotamères 'classiques'. Certaines méthodes hybrides permettent le passage des conformères, utilisés pour la recherche initiale, à une optimisation des conformations dans l'espace continu [135]. Une approximation couramment employée pour simuler le packing des chaînes latérales est que la chaîne principale doit rester fixe. Cette supposition n'est pas toujours étayée. Ainsi Eyal et al. 2003 [136] montrent que, le plus souvent, une mutation fait dévier moins de deux rotamères des chaînes latérales environnantes mais que la chaîne principale subit un déplacement. Pour en tenir compte des approches utilisant des conformations alternatives de la chaîne principale ont été implémentées [137] [138] [139]. Par exemple, Zheng et al. 1994 [139] utilisent un échantillonnage local amélioré (locally enhanced sampling ou LES) et les procédures multicopies.

Approche SCHISMo

Afin de concilier au mieux rapidité de la modélisation et précision des modèles, nous avons choisi de développer une nouvelle méthode utilisant une représentation de plus en plus précise des chaînes latérales protéiques au fur et à mesure que l'algorithme s'approche des modèles optimaux. L'utilisation d'une représentation simplifiée pour la prédiction de la conformation des chaînes latérales a déjà été décrite [140] mais celle-ci s'appuyait sur un seul pseudo atome pour chaque chaîne latérale. De même l'affinement de modèles moléculaires grâce à des représentations de plus en plus précises a déjà été décrite par d'autres auteurs mais jamais de façon hiérarchique. Trois représentations sont implémentées dans SCHISMo : l'utilisation de pseudo atomes pour les chaînes latérales [99] (effet stérique), utilisation de plusieurs centres d'interactions pour les chaînes latérales longues ainsi que pour les centres d'interactions chargés [141] [142] (flexibilité de chaîne latérale et interactions polaires), et finalement la représentation de tous les atomes lourds.

Réduction de l'espace d'exploration Nous nous sommes intéressés à la réduction de l'espace d'exploration des angles dièdres. Pour cela nous avons observé la répartition des atomes en positions delta, epsilon et zêta autour du carbone bêta. La Figure 4.2 présente cette répartition pour un parcours avec un pas constant des angles dièdres. L'atome en position zêta parcourt presque entièrement le volume d'une sphère et l'atome en position de sphère. Parmi ces trois distributions nous avons choisi d'utiliser celle des atomes en position delta qui semble plus facile à décrire. La distribution des angles dièdres cause une répartition inégale des coordonnées explorées. Une exploration plus efficace de l'espace est possible en plaçant d'abord les atomes en position delta uniformément sur leur sphère avant de considérer la position des atomes gamma.

La généralisation de la prise en compte de deux angles dièdres simultanément est la suivante. Soient 3 segments de droites consécutifs. Les 4 sommets sont notés A, B, C et D (Figure 4.3). Les longueurs des segments et les angles directes entre segments consécutifs sont constants. Seuls les deux angles dièdres, dont les axes sont les segments AB et BC, sont variables. Si les coordonnées de A et B sont fixes dans l'espace, les positions occupées par le point C décrivent un cercle dans l'espace (grâce au premier angle dièdre). Les positions du point D décrivent une portion de sphère centrée sur le point B. En effet la distance BD est constante car les distances BC et DC sont constantes et car l'angle de valence BCD est constant. Une bibliothèque de conformères peut être construite en utilisant cette propriété pour positionner les points D avant de calculer les positions des points C. A chaque point D peut correspondre deux points C et C' (sauf pour les points D situé sur le bord de la portion de sphère qui ne correspondent qu'à un seul point C). Les positions des points C et C' se déduisent aisément par un calcul simple des positions des points D (voir Figure 4.3).

Représentation à deux centres Pour la représentation des chaînes latérales avec deux centres au maximum, le premier centre est le carbone bêta et le deuxième est un pseudo atome représentant la moyenne des autres atomes (Figure 4.4). Trois groupes d'acides aminés ont été constitués. Le groupe B : les chaînes latérales comportant uniquement un carbone bêta (alanine) ou un pseudo atome bêta (valine, cystéine, sérine, thréonine). Le groupe G : les chaînes latérales comportant un pseudo atome gamma (leucine, isoleucine, phénylalanine, tyrosine, acide aspartique, asparagine, histidine). Le pseudo atome décrit un cercle dont le centre passe par la droite définie par le carbone alpha et le carbone bêta. Le groupe D : les chaînes latérales comportant un pseudo atome delta (méthionine, acide glutamique, glutamine, arginine, lysine, tryptophane). Le pseudo atome décrit une portion de sphère centrée sur la carbone bêta. Les positions de cet atome sont réparties équitablement sur une sphère comme décrit précédemment. La proline est représentée par un pseudo carbone bêta.



Figure 4.2: Exploration à pas constants de la position des atomes en position delta, epsilon et zêta autour de l'atome en position bêta de la lysine. La première figure présente également la position des atomes gamma (en bleu foncé). Le premier angle dièdre, qui positionne l'atome gamma, ne parcours que quelques degrés afin de pouvoir observer une coupe de la répartition. L'azote de la chaîne principale est en vert et le carbone bêta est en rouge.



Figure 4.3: Détermination des coordonnées de l'atome C dans une chaîne A-B-C-D possédant des longueurs de liaison et des angles de valence de valeur constante. Deux solutions, notées C et C', sont possibles. Les points b, C, C' et d sont coplanaires.



Figure 4.4: Représentation à deux centres pour les acides aminés du groupe D et G (voir texte). Les 5 acides aminés du dessus appartiennent au groupe D et les 6 acides aminés du bas appartiennent au groupe G. Les atomes en blanc représentent le carbone bêta (au centre), l'azote (à gauche) et le carbone de la chaîne principale lié à l'oxygène (vers le bas). Les positions des pseudo atomes sont colorées suivant leur représentativité dans les structures RX, du bleu (jamais observé) au rouge (le plus observé). Afin d'observer la répartition dans l'espace des représentations à deux centres, la représentation correspondant au tryptophane est présentée sous deux angles de vue différents.

Dans la représentation à deux centres, de même que dans les représentations à plusieurs centres et de tous les atomes lourds, les conformations induisant un choc stérique avec les atomes de la chaîne principale ou avec les autres atomes de la chaîne latérale sont éliminées.

Représentation à plusieurs centres La représentation avec plusieurs centres utilise le carbone bêta et les pseudo atomes ou atomes représentant des groupes chargés ou stériquement importants. Les conformères générés pour la représentation à deux centres servent de base au développement des représentations à plusieurs centres. Par rapport à la représentation précédente, des pseudo bêta sont développés en pseudo gamma, des pseudo gamma sont développés en pseudo delta et des pseudo delta sont développés en pseudo êta. Les différences avec la représentation à deux centres sont les suivantes.

Au sein du groupe B, la sérine est représentée par son carbone bêta et les oxygènes gamma, qui se placent sur un cercle (pseudo gamma).

Au sein du groupe G, le centre du cycle benzénique et l'oxygène en position êta de la tyrosine sont représentés par un pseudo gamma. L'acide aspartique et l'asparagine sont représentés par les deux atomes en position delta (deux oxygènes, et un oxygène et un azote respectivement) qui décrivent une portion de sphère. L'histidine est représentée par son azote en position delta et son azote en position epsilon, qui décrivent tous deux une portion de sphère.

Au sein du groupe D, l'acide glutamique et la glutamine sont décrits par le carbone bêta et leurs deux atomes chargés en epsilon. Les coordonnées dans l'espace de ces atomes chargés sont calculées comme appartenant à des cercles dont le centre appartient à la droite passant par les carbones gamma et delta (qui décrit une portion de sphère). La lysine est représentée par l'azote zêta, qui décrit une portion de sphère autour du carbone delta, qui décrit lui aussi une portion de sphère autour du carbone bêta. Le groupement guanidinium de l'arginine est représenté de façon similaire par un pseudo atome zêta.

Représentation tous atomes Pour la représentation avec tous les atomes lourds, la plupart des atomes peuvent être déduits de la position des atomes dans la représentation à plusieurs centres. Le positionnement de certains atomes nécessitent néanmoins une exploration, comme les atomes des cycles benzéniques de la phénylalanine et de la tyrosine, et les atomes du groupement guanidinium de l'arginine.

Organisation hiérarchique des représentations Les différentes représentations sont construites de manière hiérarchique, c'est-à-dire qu'une conformation de la représentation à deux centres engendre des représentations à plusieurs centres qui engendrent des représentations avec tous les atomes lourds. Toutes les conformations sont ainsi pré-calculées et stockées sous la forme d'un arbre à plusieurs dimensions.

Au sein d'une même représentation les coordonnées des atomes des conformations sont stockées dans différents niveaux de l'arbre qui correspondent aux niveaux de construction des chaînes latérales. Les feuilles correspondent aux atomes ou pseudo atomes les plus éloignés du carbone alpha en suivant l'enchaînement des liaisons covalentes. Le passage entre les différentes représentations s'effectue entre les feuilles de l'arbre de représentation moins complexe et un des niveaux de l'arbre de représentation plus complexe.

Les différentes conformations, à chaque niveau de représentation ont été classées suivant leur représentativité dans 48 structures RX de peptide/MHC ayant une résolution inférieur à 2Å (liste en annexe 167). Rognan et al. 2000 [110] notent que l'utilisation de structures RX du MHC pour la construction des bibliothèques de rotamères permet d'obtenir des résultats bien meilleurs que l'utilisation de bibliothèques de rotamères généralistes. La réduction du nombre de structures RX peut être vue comme un inconvénient car elle limite l'exploration des conformères rares. Cependant la bibliothèque construite fait appel à une représentation simplifiée qui doit permettre de regrouper les conformères structurellement proches, et d'autre part elle ne se limite pas aux conformères rencontrés dans les structures RX. Ces conformations non observées dans les structures RX sont par contre classées aléatoirement. Une solution serait de classer ces conformations à l'aide d'une énergie calculée au moyen d'un potentiel, ce qui n'a pas été tenté. Un encombrement stérique important peut être la cause du rejet des conformations observées expérimentalement et le nombre de conformations pouvant convenir est alors très réduit. En effet les contraintes stériques jouent alors le rôle d'un filtre de sélection qui devrait être suffisamment efficace pour permettre la sélection d'un petit nombre de conformations proches parmi celles qui sont classées aléatoirement. Chaque atome ou pseudo atome de la bibliothèque possède un rayon dérivé des paramètres de mécanique moléculaire de AM-BER [143] et deux propriétés qualitatives : (1) chargé ou non, (2) accepteur ou donneur de liaisons hydrogène ou les deux. Ces paramètres sont utilisés au cours de la construction des chaînes latérales afin d'évaluer leurs interactions.

4.2.2 Stratégie de recherche

Bibliographie

Les stratégies de recherche de la ou des conformations d'énergie minimale sont de deux types, la recherche stochastique et la recherche exhaustive locale. La combinaison de plusieurs algorithmes est souvent employée, un algorithme rapide de sélectivité assez grossière suivi d'un algorithme effectuant une recherche plus fine. Mais, comme il a été dit précédemment aucun algorithme optimal ne peut être utilisé pour une recherche de conformation d'une protéine.

Recherche stochastique Les méthodes stochastiques font appel à : l'algorithme Metropolis Monte Carlo [144], l'algorithme famille-conformationelle Monte Carlo [145] qui utilise les conformations de basse énergie trouvées pour biaiser la recherche, le recuit simulé avec dynamique moléculaire [133], les optimisations de champ moyen [109] [139] [134] [146] [147] qui sont des recuits simulés utilisant plusieurs conformations simultanément, les algorithmes génétiques et évolutifs [126] [135], l'algorithme A* [148] qui sélectionne les conformères en fonction de leur utilisation dans les conformations de basse et de haute énergie. Elagage de l'arbre des possibles et recherche exhaustive locale Les algorithmes stochastiques ne permettent pas d'obtenir à coup sûr la conformation de plus basse énergie [149] et la recherche exhaustive des conformations accessibles des chaînes latérales d'une protéine dépasse encore les capacités de calcul actuelles. Par contre les algorithmes déterministes utilisant des bibliothèques de conformères et des règles simples permettent d'élaguer l'arbre des possibles. Ils sont largement utilisés par les algorithmes de modélisation des chaînes latérales. La méthode la plus utilisée faisant usage de règles simples est la méthode d'élimination de cul-de-sac (dead-end elimination) [116] [95] [150] [149] [151] [148]. Looger et Hellinga 2001 [151] présentent une revue des différents théorèmes d'éliminations des cul-de-sac qui tirent avantage de la décomposition par paire de la fonction d'énergie. Pour identifier le minimum global d'énergie, des critères sur l'énergie sont employés pour détecter les couples (position, conformation) qui ne peuvent faire partie du minimum global. Lorsque l'algorithme ne converge pas, lorsque les critères de sélection ne permettent pas de réduire la recherche à une seule conformation, les conformations restantes sont explorées à l'aide d'autres algorithmes tel que le branch and bound [152] ou le branch-and-terminate [150]. L'algorithme branch and bound est employé pour élaguer l'arbre de recherche de la conformation d'énergie minimum. Cet algorithme permet d'affirmer qu'une branche ne pourra donner un résultat supérieur à un seuil sans devoir calculer les résultats de toutes les feuilles. L'algorithme branch and terminate est une variante de l'algorithme branch and bound où les critères d'élagage sont appliqués séquentiellement à chaque niveau de l'arbre, depuis la racine jusqu'aux feuilles, ce qui est plus rapide. La théorie des graphes a également été employée par Canutescu et al. 2003 [127] pour diviser le problème d'optimisation en plusieurs optimisations locales. Un graphe des positions en interactions est construit et décomposé en sous-graphes biconnexes, par la suite considérés comme des 'super résidus', dans lesquels la recherche de la conformation de plus basse énergie peut être effectuée de manière exhaustive.

Recherche dans l'arbre des SCHISMo

L'algorithme de recherche que nous avons employé est très simple et se base sur une optimisation locale de chaque chaîne latérale utilisant en partie une recherche stochastique. Xiang et Honig 2001 [125] suggèrent que l'efficacité et la précision de la modélisation des chaînes latérales dépend surtout de la précision de la bibliothèque de conformères et très peu de l'algorithme de recherche. Notre méthode est heuristique, c'est-à-dire qu'elle emploie des règles simples permettant d'obtenir rapidement une solution qui doit être proche de la solution optimale sans toutefois l'atteindre à coup sûr.

Construction d'une représentation d'une seule chaîne latérale La construction d'une chaîne latérale simplifiée s'effectue par l'algorithme heuristique suivant (Figure 4.5). Les conformations sont parcourues dans l'ordre décroissant de leur classement. Une conformation présentant une interaction de charges avec un atome externe à la chaîne latérale est retenue si elle ne présente pas de chocs stériques et est directement retournée comme solution. Une conformation présentant une liaison hydrogène avec un atome externe à la chaîne latérale en construction est retenue si il n'y a pas de chocs stériques et elle est retournée comme résultat si les cinq conformations suivantes ne présentent pas d'interaction de charge. Une conformation présentant une interaction de type van der Waals est retenue et retournée comme résultat si aucun choc stérique n'est découvert et si les cinquante conformations suivantes ne sélectionnent pas une conformation présentant une liaison hydrogène ou une interaction de charge, comme décrit précédemment. Si toutes les conformations sont parcourues et que toutes engendrent des chocs stériques, la conformation présentant le moins de chocs stériques est retournée comme résultat. La création des conformations utilise un découpage de l'espace par des boites d'atomes afin d'accélérer la recherche des atomes en choc stérique.

Construction des chaînes latérales du peptide/MHC Pour construire les chaînes latérales du peptide et du MHC, la stratégie suivante est employée (Figure 4.5). La conformation à deux centres la plus proche de la conformation de la structure RX de référence est déterminée pour chaque position du MHC. Les conformations deux centres du peptide sont créées séquentiellement depuis le résidu N-terminal jusqu'au résidu C-terminal. Les chocs stériques sont évalués et sont résolus avec l'algorithme décrit dans le paragraphe suivant. Les chaînes latérales à plusieurs centres engendrées par les chaînes latérales à deux centres sont construites depuis le résidu N-terminal jusqu'au résidu C-terminal. Les chaînes latérales sont reconstruites à nouveau mais en tenant compte de toutes les chaînes latérales à deux centres construites, ce qui permet la mise en place des interaction spécifiques. Les chocs stériques sont évalués et sont résolus par l'algorithme décrit dans le paragraphe suivant. Les chaînes latérales avec tous les atomes lourds engendrées par les chaînes latérales à plusieurs centres sont construites depuis le résidu N-terminal jusqu'au résidu C-terminal. Les chaînes latérales sont reconstruites à nouveau mais en tenant compte de toutes les chaînes latérales multicentres construites, ce qui permet la mise en place des interactions spécifiques. Les chocs stériques sont évalués et sont résolus avec l'algorithme décrit dans le paragraphe suivant.

Résolution des chocs stériques La résolution des chocs stériques est traitée par un algorithme stochastique. Les différentes chaînes latérales sont classées dans trois listes, la liste des chaînes latérales présentant des chocs stériques, la liste des chaînes latérales appartenant à la même boite d'atomes que les chaînes latérales en choc stérique (le voisinage) et les autres chaînes latérales. Un tirage biaisé est effectué en regroupant ces 3 listes, une chaîne latérale en choc stérique ayant un poids de 500, une chaîne latérale dans le voisinage des chaînes latérales en chocs stérique ayant un poids de 150 et les chaînes latérales autres ayant un poids de 1. La chaîne latérale tirée est reconstruite et ses chocs stériques sont évalués. Les 3 listes de chaînes latérales sont réévaluées et l'algorithme s'arrête si la liste des chaînes latérales en chocs stériques est vide ou si plus de 300 chaînes latérales ont été reconstruites. Les valeurs de seuils pour les poids entre groupe et le nombre maximal de chaînes latérales reconstruites ont été trouvées par tâtonnement mais peuvent être aisément modifiées par l'utilisateur.



Figure 4.5: Protocole de construction des chaînes latérales par l'algorithme SCHISMo.

4.2.3 Evaluation des conformations

Bibliographie

Les énergies d'interactions, c'est-à-dire le potentiel de van der Waals et le potentiel électrostatique, peuvent être dérivées des potentiels physiques classiques tel que le champ de force d'AMBER ou de CHARMM [135] [148] [146] [125] [144] [127]. Les potentiels de Lennard-Jones donnent d'excellents résultats pour les groupements enfouis tels que les acides aminés non polaires du coeur de la protéine mais pas pour les chaînes latérales exposées ou les chaînes latérales polaires enfouies [133]. Vasquez 1996 [132] remarque que la prise en compte des liaisons hydrogène offre une faible, voire aucune, amélioration. Liang et Grishin 2002 [144] développent une fonction d'évaluation prenant en compte la surface de contact, les volumes de recouvrement, la dépendance par rapport à la chaîne principale, les interactions électrostatiques et l'énergie de désolvatation. La pondération entre ces différents termes a fait l'objet d'une optimisation afin d'obtenir le plus petit RMS entre la conformation la mieux classée et la structure RX. L'utilisation d'une bibliothèque de rotamères réduit la dépendance à la fonction d'énergie car les bibliothèques ne contiennent que des conformations énergétiquement favorables. L'énergie moyenne de la conformation peut être simulée par le logarithme de la probabilité de la population de rotamères dépendant de la chaîne principale [153] [148]. Par contre les algorithmes utilisant des conformations en dehors des rotamères doivent utiliser un potentiel de torsion [135]. Xiang et Honig 2001 [125] ont montré qu'une fonction d'énergie grossière est suffisante pour déterminer la conformation des chaînes latérales enfouies.

Approche utilisée pour l'évaluation des SCHISMo

Notre procédure de construction des chaînes latérales calcule l'ensemble des interactions réalisées par chaque chaîne latérale. Basé sur ces résultats, un score d'interaction peut être associé à chaque peptide. Ce score est une simple combinaison linéaire du nombre de liaisons hydrogènes, du nombre d'interactions de charge et du nombre de contacts de van der Waals.

$$Score_1 = A.Nb_{LH} + B.Nb_{Charges} + C.Nb_{VdW} + D.Nb_{chocs}$$

Les valeurs des facteurs A, B et C sont 20, 20 et 1, respectivement, qui sont les valeurs utilisées pour l'évaluation des sites de contacts IMGT (section 3.5.3 page 74). Le facteur D a été pris à -50 et le schéma de score est noté (20,20,1,-50). Un autre schéma de score a été essayé, (100, 30, 1, -600), qui pénalise de façon plus importante les chocs stériques et favorise les interactions polaires.

4.2.4 Résultats

Modélisation de structures RX

La méthode de construction des chaînes latérales SCHISMo a été comparée à l'algorithme SCWRL [106] [127]. Les chaînes latérales de 28 structures RX de peptide/MHC
ont été reconstruites par les deux méthodes, les temps de calcul évalués et la moyenne quadratique des distance entre les atomes de la structure RX et des modèles a été calculée. Les résultats sont rassemblés dans le Tableau 4.1.

La méthode SCHISMo obtient une précision comparable à l'algorithme de SCWRL mais avec un temps de calcul inférieur et plus constant. Il faut noter que le chargement de la bibliothèque constitue la majorité du temps d'exécution de SCHISMo, soit environ 2.5s sur 3.0s de temps d'exécution. Une fois la bibliothèque en mémoire, la génération d'un grand nombre de conformation est donc beaucoup plus rapide avec SCHISMo qu'avec SCWRL. Comme décrit précédemment, SCWRL emploie une méthode reposant sur la théorie des graphes pour réduire le problème à la recherche exhaustive du placement des chaînes latérales parmi un nombre réduit d'acides aminés. La réduction n'est pas toujours efficace ce qui peut expliquer les structures pour lesquelles aucun résultat n'est obtenu au bout d'une heure. Une autre raison possible est l'existence de structures cristallines mal résolues. En effet la technique de diffraction des rayons X permet de discerner la localisation approximative des atomes et la structure qui en résulte peut présenter des chocs stériques parfois important entre atomes [154]. SCWRL ne parvient donc pas à réduire entièrement les chocs stériques car cela nécessiterait de déplacer certains atomes de la chaîne principale tandis que les représentations simplifiées de SCHISMo permettent de 'gommer' transitoirement ces chocs stériques. La méthode SCHISMo n'a pas pour objectif de résoudre tous les problèmes de chocs stériques, ce qui permet de traiter ce genre de problème très similaire à ceux rencontrés dans les modèles de moyenne résolution, et la structure obtenue pourra être affinée à une étape ultérieure (par exemple une dynamique moléculaire ou un recuit simulé). L'objectif de rapidité est atteint et les chaînes latérales sont reconstruites avec une précision raisonnable. L'examen des modèles construit par SCHISMo révèle que les chaînes latérales enfouies sont mieux reconstruites que les chaînes latérales exposées au solvant. Les interactions à l'interface sont donc à priori correctement modélisées puisqu'elles font essentiellement intervenir des chaînes latérales enfouies. L'application de la méthode à des techniques de docking n'est pas remise en cause puisqu'elle est suffisamment rapide pour pouvoir être appliquée fréquemment. Les chaînes latérales exposées seront mieux modélisées lors de l'approche du ligand.

Evaluation de la liaison de peptides

La sélectivité de la fonction d'énergie très simple, basée sur le nombre d'interactions, a été évaluée en considérant les complexes entre le MHC-II DRB1*0401 et des peptides générés aléatoirement et connus pour se lier à cet allèle de MHC. Mille séquences de peptides ont été générées aléatoirement et 70 séquences de peptides se liant expérimentalement à DRB1*0401 ont été extraits de la base SYFPEITHI [81]. Les modèles de tous les complexes ont été générés et les énergies sont reportées sur la Figure 4.7.

Une faible sélectivité est observée avec le schéma de score (20, 20, 1, -50). En effet, pour une sélection de 50% des peptides de séquences aléatoires (énergie de 12000), seulement 52% des peptides se liant effectivement sont sélectionnés. Avec le schéma de score (100, 30, 1, -600), qui pénalise de façon plus importante les chocs stériques et favorise les interactions polaires, la sélectivité est un peu meilleure. Pour une sélection de 50% des

Table 4.1: Comparaison de la méthode SCHISMo et de l'algorithme SCWRL pour la prédiction des chaînes latérales à l'aide de la moyenne quadratique des distances (RMSD) entre les atomes de structures RX de pMHC et des modèles construits par la méthode SCHISMo et SCWRL, et des temps de calcul en minutes :secondes. La colonne de droite donne des résultats pour la méthode SCHISMo pour lesquelles SCWRL n'a pu donner de résultats au bout d'une heure de calcul. Les calculs ont été effectués sur un Pentium 1.6GHz.

	$\mathbf{SCHISMo}$		\mathbf{SC}	\mathbf{WRL}		SCHISMo		
ID	RMSD	Temps	RMSD	Temps	ID	RMSD	Temps	
1a6a	$1,\!16$	0:03.07	$1,\!35$	0:21.35	$1 \mathrm{kt2}$	1,73	0:04.48	
1aqd	$1,\!35$	0:03.04	$1,\!12$	0:10.38	$1 \mathrm{ktd}$	$1,\!69$	0:04.72	
$1\mathrm{bx}2$	$1,\!17$	0:02.73	1,41	0:04.89	1lnu	1,73	0:04.13	
1d5m	$1,\!32$	0:03.08	1,22	0:05.70	$1 \mathrm{es0}$	1,59	0:04.55	
1d5x	$1,\!07$	0:02.97	1,25	0:05.45	$1 \mathrm{fne}$	1,71	0:04.68	
1d5z	$0,\!98$	0:03.00	1,27	0:04.90	1fng	1,77	0:04.58	
1d6e	$1,\!12$	0:03.02	1,32	0:06.28	1i3r	1,71	0:04.53	
1 d9k	$1,\!33$	0:02.74	1,46	0:04.71		1,70	0:04.52	
1dlh	$1,\!40$	0:03.06	$1,\!33$	0:04.57		,		
1fv1	$1,\!35$	0:03.07	1, 19	0:11.85				
1fyt	1,28	0:02.73	1,29	0:10.04				
1hqr	$1,\!30$	0:03.06	1,44	0:04.66				
1iak	$1,\!29$	0:03.03	$1,\!32$	0:10.67				
1j8h	$1,\!25$	0:03.04	$1,\!33$	0:05.44				
1jk8	$1,\!36$	0:03.06	1,29	0:04.76				
1jl4	$1,\!32$	0:03.01	1,33	0:05.40				
1jwm	$1,\!23$	0:03.03	$1,\!34$	0:04.87				
1jwu	$1,\!30$	0:03.02	1,31	0:05.31				
1k2d	$1,\!15$	0:03.08	1,27	0:05.14				
1kg0	$1,\!32$	0:03.02	$1,\!37$	0:05.74				
1klg	$1,\!38$	0:03.01	$1,\!37$	0:04.66				
$1 \log 5$	$1,\!25$	0:03.03	1,36	0:04.36				
1r5w	$1,\!34$	0:03.03	1,51	0:05.77				
$1 \mathrm{s}9 \mathrm{v}$	$1,\!34$	0:03.03	$1,\!32$	0:04.40				
1t5w	$1,\!22$	0:03.02	1,29	0:05.27				
1t5x	$1,\!24$	0:03.03	1,26	0:05.29				
1uvq	1,28	0:03.03	1,35	0:05.14				
2seb	1,30	0:03.05	1,48	0:04.73				
	1,25	0 :03.00	1,33	0:06.49				



Figure 4.6: Représentations SCHISMo lors de la construction des chaînes latérales du peptide/MHC-II de 1fyt. En haut la représentation à deux centres (chaînes latérales en rouge). Au milieu représentation multicentres (chaînes latérales en bleu). En bas représentation tous atomes lourds (chaînes latérales en vert) avec une superposition des représentations à deux centres et à plusieurs centres.



Figure 4.7: Energie des modèles peptide/HLA-DRB1*0401 avec 1000 peptides de séquence aléatoire (à gauche) et de 70 peptides se liant expérimentalement (à droite). L'énergie est obtenue à partir d'une combinaison linéaire des interactions rencontrées dans le complexe. Le graphique du bas correspond à l'emploi du schéma de scores suivant : 20 points pour les interactions entre atomes polaires, 20 points pour les liaisons hydrogènes, 1 point pour les interactions de van der Waals, et -50 points pour les chocs stériques. Le graphique du haut correspond à l'emploi du schéma de scores suivant : 100 points pour les interactions entre atomes polaires, 30 points pour les liaisons hydrogènes, 1 point pour les interactions de van der Waals, et -600 points pour les liaisons hydrogènes.

peptides de séquences aléatoires (énergie de 12500), 65% des peptides se liant effectivement sont sélectionnés.

La fonction d'énergie employée est très basique et ce résultat n'est pas surprenant. Une optimisation des paramètres de la combinaison linéaire permettrait d'améliorer sensiblement ces résultats. Finalement, l'utilisation de fonctions plus complexes utilisant des champs de force de mécanique moléculaire et des énergies de solvatation empiriques devraient théoriquement permettre d'obtenir de meilleures prédictions.

4.3 Modélisation des peptide/MHC-I et des boucles du TR – Construction de boucles par la méthode LLIPA

Le peptide lié au MHC-I présente une bonne conservation de la position des résidus extrêmes du peptide dans le sillon du MHC (section 3.5.1 page 70). La conformation de la partie centrale du peptide est par contre beaucoup plus flexible pour les peptides au-delà de 9 résidus. Les peptides de longueur 8 présentent une aussi bonne conservation de la chaîne principale que les peptides liés au MHC-II, et peuvent donc être modélisés avec le même algorithme que dans la section précédente. Les variations de conformation des peptides plus longs présentent toutes les caractéristiques d'une boucle, c'est pourquoi ce problème de modélisation a été considéré en même temps que celui de la conformation des boucles du TR. Des tests préliminaires de modélisation des boucles du TR, réalisés avec le logiciel MODELLER [119], a fait apparaître la nécessité d'un nouvel algorithme de construction de boucles. En effet l'algorithme de MODELLER nécessite plusieurs heures de calculs et la génération de nombreuses conformations alternatives avant d'être sûr de générer au moins une conformation de boucle proche de la conformation native. Notre algorithme de prédiction de boucles, LLIPA (Loop prediction using a LIbrary of Pairs of dihedral Angles), emprunte certaines caractéristiques à l'algorithme SCHISMo. Il utilise une librairie de conformations de chaînes principales de dipeptides avec une exploration efficace de l'espace accessible.

4.3.1 Prédiction de la conformation des boucles – Bibliographie

La prédiction de la conformation de boucles a été source de recherches actives dès les prémisses de la modélisation moléculaire informatique [155]. Les techniques de modélisation de boucles procèdent pour la plupart en 2 ou 3 grandes étapes, la génération des conformations des boucles, éventuellement le regroupement des conformations qui sont très proches et finalement une classification des boucles suivant des critères énergétiques ou suivant un crible. Les méthodes de génération de conformations se classent en trois groupes, (1) la fermeture de petites boucles de 3 ou 4 résidus, (2) la modélisation de boucles basée sur des bases de données de conformations observées expérimentalement et (3) la modélisation ab-initio. Fermeture de boucle Les algorithmes de fermeture de boucles fonctionnent soit en recherchant par un calcul analytique toutes les solutions de fermeture de boucles avec un nombre fini d'angles dièdres (travaux de Go et Scheraga 1970 [155]), soit en utilisant une recherche conformationnelle. Les équations analytiques de Wedemeyer et Scheraga 1999 [156] permettent de calculer les angles dièdres d'une fermeture de boucle à 3 résidus. Il est intéressant de noter, qu'ils trouvent empiriquement que les boucles de 3 résidus (6 degrés de liberté car l'angle oméga est maintenu constant) n'ont au maximum que 8 conformations solutions, bien que les raisons exactes de cette propriété demeurent obscures [156]. La résolution analytique de leur système d'équations nécessite de trouver les racines d'un polynôme de degré 16. Comme il n'existe pas de méthode de résolution exacte d'un tel polynôme dans le cas général, la recherche de solutions est effectuée par la méthode de Newton. Le degré élevé du polynôme ne permet pas d'avoir une bonne précision sur la valeur des angles dièdres, ce qui conduit à de nombreux cas sans solutions [157]. Une solution est d'accorder une certaine variabilité aux angles dièdres et d'effectuer une minimisation finale sous contraintes afin d'obtenir une fermeture exacte [158]. La recherche conformationnelle dans l'espace des angles dièdres nécessite moins de calculs que la résolution analytique car elle procède par une unique optimisation. Le plus souvent, la fermeture est effectuée par une minimisation énergétique [159] ou une recherche type Monte Carlo [160], avec éventuellement un potentiel harmonique de fermeture [161]. Zheng et al. 1993 [162] ont développé une technique de fermeture de boucle par déformation des longueurs de liaisons afin de satisfaire les contraintes de distance de fermeture, puis relaxation des longueurs de liaisons sous contraintes (utilisé notamment par [163]). Canutescu et Jr Dunbrack 2003 [157] appliquent un algorithme utilisé en robotique pour la manipulation de bras mécaniques. La valeur de chaque angle dièdre est successivement calculée de manière à satisfaire au mieux des contraintes de fermeture.

Modélisation utilisant les conformations d'une base de données Les études sur les conformations de boucles dans les bases de données de structures 3D ont montré qu'il n'existait qu'un nombre limité de conformations accessibles pour les petites boucles [164]. Cette observation permet de modéliser les boucles de conformations inconnues à partir des conformations observées expérimentalement [165]. La modélisation utilisant une base de données procède en deux étapes [159], la sélection des conformations de la base de données qui pourraient convenir puis l'adaptation de la conformation pour obtenir une correspondance des extrémités de la boucle avec le reste de la structure. La sélection de conformations fait appel à des critères géométriques et à des critères de séquence. Les critères géométriques concernent les positions d'ancrages de la boucle, telles que les distances entre les atomes des positions d'ancrage ou de plusieurs positions en amont des positions d'ancrage, ou des critères angulaires sur le départ des boucles [166]. Les critères de séquence concernent essentiellement la taille de la boucle mais également la composition en acides aminés pour les bases de données avec une identification des positions structuralement importantes (SDR) [167] [168]. L'application de techniques de threading permet de prendre en compte l'environnement local de chaque position de la boucle [128], comme implémenté dans le serveur FREAD (http://www-cryst.bioc.cam.ac.uk/cgi-bin/coda/fread.cgi). Ces techniques permettent d'obtenir la chaîne principale de la boucle, tandis que les chaînes latérales doivent être construites par des algorithmes dédiés [106] [125]. Cette approche de construction des boucles est limitée par le nombre de conformations présentes dans les bases de données, et tout particulièrement pour la modélisation des grandes boucles. Une solution proposée est de construire les boucles par parties, les flancs de la boucle sont construits grâce à des conformations trouvées dans une base de données, puis la boucle est fermée par un algorithme de fermeture de boucles. Dans l'approche classique, l'algorithme de fermeture de boucles est employé pour parfaire l'adaptation des ancrages au reste de la structure, tandis que dans cette approche pour les longues boucles il est employé pour fermer l'apex de la boucle. D'autres solutions s'appuient sur une approche mixte utilisant des bases de données et de méthodes servant pour la construction ab-initio [169] [170].

Modélisation *ab-initio* **des boucles protéiques** Les méthodes *ab-initio* produisent des conformations de boucles soit par recherche dans l'espace des angles dièdres de la chaîne principale, soit par minimisation ou exploration conformationnelle de boucles générées aléatoirement.

Bruccoleri et Karplus 1987 [118] ont implémenté dans le programme CONGEN une recherche dans l'espace des angles dièdres avec un pas d'exploration constant. D'autres algorithmes utilisent des bibliothèques de rotamères de chaîne principale observés dans des structures RX de protéines [171] [160] [172] au lieu d'explorer exhaustivement les angles dièdres. Avec ces techniques le nombre total de conformations augmente de façon exponentielle avec la taille de la boucle à construire. Pour y remédier CONGEN tronque sa recherche quand sa fonction d'énergie indique qu'un fragment de boucle en construction est d'énergie trop défavorable. DePristo et al. 2003 [160] ne construisent pas toutes les conformations mais seulement les plus probables grâce à un algorithme de planification des tâches, dit Round-Robin, ce qui se révèle plus efficace qu'une exploration exhaustive. Un autre problème de ce type de technique veut que l'utilisation d'une exploration avec un pas discret ou une bibliothèque de conformations peuvent ne pas réussir à trouver une conformation sans choc stérique. Jacobson et al. 2004 [172] utilisent une bibliothèque très fine, avec lors de la construction des angles dièdres, une étape de regroupement par l'algorithme K-means des conformations générées sans choc stérique. Ce regroupement permet d'obtenir un temps de calcul raisonnable tout en n'excluant pas les conformations les plus isolées. Deane et Blundell 2000 [173] développent une approche originale où toutes les conformations de peptides jusqu'à 12 acides aminés de long et sans choc stérique sont précalculées exhaustivement et entrées dans une base de données. Les fragments sont classés suivant la représentativité des fragments protéiques dans les structures RX. Ces techniques de recherche dans les angles dièdres impliquent une discrétisation de cet espace qui ne peut être parcouru de manière continue.

La deuxième grande technique de modélisation *ab-initio* de boucles est la minimisation de conformations générées aléatoirement. Des conformations aléatoires peuvent être obtenues en prenant des conformations de boucles dans une base de données, en choisissant aléatoirement des angles phi/psi [157] ou en plaçant uniformément les atomes de la chaîne principale entre les extrémités de la boucle et en appliquant une dérive aléatoire aux coordonnées cartésiennes des atomes [119]. Certains méthodes se contentent ensuite d'appliquer un algorithme de fermeture de boucles au segment protéique généré aléatoirement [157]. Plus classiquement, la minimisation est effectuée par un protocole de minimisation énergétique utilisant un champ de force complet ou des potentiels simplifiés [154]. Les techniques classiques d'exploration conformationnelle ont également été utilisées : la dynamique moléculaire [174], le recuit simulé [175] [161], et les simulations de Monte Carlo [176] [120] [177]. La méthode des perturbations minimales par "torsion aléatoire" (*random tweak*) permet de modifier simultanément tous les angles dièdres jusqu'à ce que les contraintes de distances de fermeture soient satisfaites. Zheng *et al.* 1994 [139] implémentent la méthode des copies multiples [79] qui permet la minimisation simultanée d'un ensemble de boucles générées aléatoirement. Le reste de la protéine interagit avec l'effet moyen de toutes les copies de boucles, tandis que les boucles n'interagissent qu'avec le reste de la protéine et pas entre elles. Une extension de cet algorithme appliquée aux IG est la modélisation simultanée des 6 CDR, chaque CDR n'interagissant qu'avec l'effet moyen des autres CDR [139].

Classification des conformations produites La dernière étape de la modélisation de boucles est la classification des nombreuses conformations générées grâce à une score ou une énergie. Cette étape est souvent couplée avec une minimisation d'énergie qui permet de régler les légères irrégularités de structure (chocs stériques) provenant par exemple de l'utilisation de valeurs discrètes des angles dièdres. Le problème de la classification des conformations. En effet un algorithme parcourant exhaustivement l'espace des angles dièdres avec un pas suffisamment fin générera assurément une conformation très proche de celle observée expérimentalement. Par contre aucune méthode ne permet, ne serait-ce que théoriquement, de sélectionner la conformation la plus proche de la conformation expérimentale, ce qui s'avère en pratique très subtile. Afin de pouvoir comparer chaque méthode de sélection, des jeux de conformations de boucles ont été créés et partagés sur internet (site Decoys'R us, http://dd.stanford.edu). Les potentiels peuvent se classer en deux grands groupes, les potentiels physiques et les potentiels statistiques.

Les potentiels physiques les plus utilisés sont issus des champs de force de mécanique moléculaire, tels que AMBER, CHARMM ou GROMOS. Aux potentiels de mécanique moléculaire s'ajoutent des techniques permettant de prendre en compte la solvatation des protéines par des méthodes permettant de simuler l'action de l'eau de façon implicite (énergie de solvatation proportionelle à la surface accessible ou méthode Generalized Born). Martin *et al.* 1989 [169] utilisent, pour la modélisation des CDR des IG, une version modifiée des paramètres de mécanique moléculaire de GROMOS ainsi qu'un terme de solvatation dépendant de la surface accessible de chaque atome. van Vlijmen et Karplus 1997 [166] utilisent les paramètres de CHARMM avec un potentiel de van der Waals 9-6, c'est à dire qui pénalise moins les chocs stériques. Jacobson *et al.* 2004 [172] effectuent une minimisation d'énergie des boucles avec le champs de force OPLS [113] et la méthode de solvatation implicite Surface Generalized Born [178]. de Bakker *et al.* 2003 [179] utilisent les paramètres de AMBER et la méthode de solvatation implicite Generalized Born Surface Area GBSA. Ces derniers testent l'utilité des termes de solvatation pour la sélection des boucles et ils concluent qu'ils procurent une amélioration substantielle. Pellequer et Chen 1997 [180] utilisent sur des CDR d'IG le champs de force de CHARMM avec un calcul de la contribution électrostatique par les équations de Poisson Boltzmann de différence finie et la contribution non polaire à la solvatation par un terme proportionnel à la surface accessible. Ils remarquent que le terme de solvatation oriente la sélection des petites boucles vers des boucles non natives.

Les potentiels statistiques peuvent reconnaître implicitement les effets physico-chimiques complexes tels que les effets hydrophobes. L'avantage majeur des potentiels statistiques est qu'ils ne sont pas aussi sensibles que les physiques sur des petits déplacements locaux. Bien que leur validation ne soit pas très claire [181], ils ont été appliqués avec succès à la reconnaissance de repliement, au repliement ab initio et au docking protéine-protéine [182]. Les potentiels basés sur les résidus représentent les résidus par un site d'interaction unique, ce qui est plus approprié au threading. Les potentiels atomiques permettent eux de prendre en compte tous les atomes et ainsi de discriminer des différences de structure plus subtiles [183]. Le potentiel statistiques RAPDF [184] a été dérivé spécifiquement pour la classification des boucles. Il a été utilisé par de Bakker *et al.* 2003 |179| et Deane et Blundell 2000 [173]. de Bakker et al. 2003 [179] ont montré que l'utilisation de RAPDF était beaucoup plus rapide que l'utilisation de AMBER et GBSA mais beaucoup moins performant. Ils concluent que RAPDF peut être utilisé comme une première étape de sélection afin d'éliminer la plupart des conformations. Fiser et al. 2000 [185] utilisent la mécanique moléculaire de CHARMM et une approximation des interactions non liantes et de solvatation par un potentiel statistique par paires. Abagyan et Totrov 1994 [120] utilisent des potentiels statistiques sur les angles dièdres et une technique associée de minimisation dans l'espace des angles dièdres. La prise en compte de la flexibilité des boucles dans le terme entropique est important pour la classification des boucles. Cette flexibilité est difficile à quantifier sans avoir recours à la dynamique moléculaire. Xiang et al. 2002 [186] proposent une méthode originale pour prendre en compte l'entropie conformationnelle dans le calcul de l'énergie en ne considérant que les distances structurales entre les conformations de boucles générées (énergie de colonie). Ils notent une forte corrélation entre leur énergie et le RMSD à la conformation native. En dehors du calcul d'énergie, des méthodes de classifications non supervisées peuvent être utilisées, bien qu'elles ne permettent pas de quantifier la liaison. Reczko et al. 1995 [187] ont entraîné un réseau de neurones sur les séquences et les structures de boucles similaires aux CDR3 de chaîne lourde des IG. Pour 7 boucles sur 20 il obtiennent des boucles prédites avec un RMSD inférieur à 2A.

4.3.2 Algorithme LLIPA

Notre approche de modélisation fait partie des méthodes de modélisation *ab-initio* par recherche dans l'espace des angles dièdres. Elle a pour but la génération d'un ensemble conformationnel de boucles ayant une résolution moyenne. L'ensemble de conformations modélisées doit contenir des boucles suffisamment proches des structures natives mais cet

ensemble doit également être le plus réduit possible au contraire des méthodes de la même catégorie qui génèrent un nombre très important de boucles. La sélection des boucles les plus représentatives est un problème très difficile qui est en partie encore ouvert. La réduction du nombre de boucles apparaît donc comme particulièrement important puisqu'une bonne classification des conformations générées nécessite l'utilisation de fonctions d'énergies gourmandes en temps de calculs [179]. Les boucles sont des éléments par nature flexibles, et leurs coordonnées dans les structures RX résultent souvent d'une moyenne (comme le montre le facteur de température). De plus, les simulations de dynamique moléculaire qui utilisent les structures RX comme conformation de départ enregistrent des déviations importantes des conformations des boucles. L'obtention de boucles avec une résolution moyenne, qui s'écartent légèrement de la structure RX, n'est donc pas à priori absurde.

Bibliothèque de paires d'angles dièdres Une bibliothèque de conformères de la chaîne principale a été construite de la même manière que les bibliothèques de conformères des chaînes latérales. Connaissant les coordonnées du carbones alpha (CA) et du carbone relié à l'oxygène (C) d'un acide aminé , les positions du carbone C de l'acide aminé suivant décrivent une portion de sphère centrée sur le carbone C du premier acide aminé. En effet la liaison peptidique étant considérée comme plane, deux angles dièdres séparent deux C.

Le carbone alpha du deuxième résidu décrit un cercle et sa position peut être calculée en fonction de celle du carbone C (Figure 4.3). Les coordonnées de l'azote et de l'oxygène de la chaîne principale se déduisent de la position des atomes dont les coordonnées sont déjà connues (Figure 4.8).

Les conformations ont été classées suivant leur représentativité dans 250 structures RX de résolution inférieure à 2Å extraites de IMGT/3Dstructure-DB (Figure 4.9) (liste en annexe 168).

Une deuxième bibliothèque de paires d'angles dièdres a été construite à l'identique mais pour la construction dans le sens inverse, de la partie C-terminale vers la partie N-terminale de la boucle. Elle permet de construire à partir des coordonnées du carbone alpha d'un acide aminé, celles de l'azote de l'acide aminé précédent, puis celles du carbone alpha et du carbone C de l'acide aminé précédent.

Construction d'une boucle Une conformation de boucle est obtenue par une construction par récurrence de chaque acide aminé. Pour un acide aminé, le premier atome de la librairie est construit, c'est-à-dire le carbone C dans le sens direct et l'azote dans le sens indirect. La distance entre cet atome et l'extrémité de la boucle à atteindre est calculée, et son incompatibilité avec la fermeture de la boucle cause l'arrêt de la construction par récurrence. Les distances maximales autorisées entre l'extrémité de la boucle et les atomes en constructions sont calculées en construisant un peptide en conformation étendue correspondant aux acides aminés de la boucle. Les deux positions alternatives des autres atomes (N, CA et O dans le sens direct, C, CA et O dans le sens indirect) sont déterminées successivement. Pour chacune, la chaîne latérale SCHISMo à deux centres est construite et les chocs stériques entre les atomes reconstruits et tous les atomes présents



Figure 4.8: Construction de la chaîne principale par la méthode LLIPA, de la partie N-terminale vers la partie C-terminale, et de la partie C-terminale vers la partie N-terminale. L'atome en rouge est le premier à être construit suivi des atomes en bleu.

(ce qui inclut la partie de boucle déjà construite) sont évalués. Si aucun choc stérique n'est détecté, la récurrence se poursuit par la construction de l'acide aminé suivant. A la fin de la récurrence, la conformation de la boucle est enregistrée si (1) les atomes de la chaîne principale du dernier acide aminé sont à moins 0.5Å des atomes correspondants de l'acide aminé d'ancrage et (2) si aucune conformation préalablement enregistrée n'est à moins de 1Å de RMSD sur les carbones alpha.

Comme pour l'algorithme SCHISMo, des boîtes d'atomes sont constituées pour réduire le temps de calcul et aussi pour calculer les proches voisins des chaînes en choc stérique.

Construction des boucles La bibliothèque de conformations de paires d'angles dièdres étant classée suivant la fréquence d'observation des conformations dans les structures RX, le rang dans la bibliothèque est utilisé pour construire en premier lieu les boucles les plus 'probables'. Cette approche est très similaire à celle de l'algorithme de Round Robin utilisé par DePristo *et al.* 2003 [160]. L'algorithme construit les boucles dont la somme des rangs des conformations de chaque acide aminé est égale à une valeur donnée. A chaque itération, la somme des rangs à atteindre est augmentée, et des conformations de plus en plus rares sont utilisées.

Lors de la construction d'une conformation d'un acide aminé, seules les conformations dont les rangs sont compatibles sont parcourues. A cette fin, la somme des rangs des positions précédentes est comparée à la somme à atteindre et aux rangs maximaux des acides aminés encore à construire.

L'algorithme s'arrête quand toutes les conformations ont été explorées ou quand le nombre de conformations générées correspond à une valeur fournie en entrée. Ceci permet la génération d'ensembles conformationnels dont la population maximale est spécifiée par l'utilisateur. En outre les conformations générées sont celles qui utilisent les conformations locales les plus fréquemment observées dans les structures RX.



Figure 4.9: Librairie de conformations pour la construction directe de N vers C pour chaque type d'acide aminé. L'atome blanc au centre représente le carbone C, et l'atome blanc vers le haut (et vers le fond) est l'azote de la chaîne principale. Les positions des atomes C du résidu suivant sont colorés suivant leur représentativité dans les structure RX, du bleu (jamais observé) au rouge (le plus observé).

4.3.3 Modélisation de peptides dans les structures RX

Une méthode classique de modélisation des régions flexibles est la dynamique moléculaire mais elle est mal adaptée au cas des peptide/MHC-I car le peptide se situe dans un environnement stériquement très encombré. Enchâssées dans la crevasse formée par les deux G-DOMAINs, les chaînes latérales du peptide occupent différentes poches. Le passage d'une conformation à une autre plus éloignée nécessite d'extraire les chaînes latérales de ces poches, ce qui est énergétiquement très défavorable. Une méthode de génération *ab-initio* de boucles, tel que LLIPA, devrait par contre permettre de mieux parcourir l'espace conformationnel accessible.

La stratégie mise en place emploie la représentation à deux centres SCHISMo et applique l'algorithme de construction de boucle LLIPA entre les deux acides aminés extrêmes du peptide. L'échantillonnage conformationnel est le problème majeur qui peut être rencontré car le sillon du MHC est très encombré et LLIPA utilise une bibliothèque discrète de conformères. L'utilisation de la représentation simplifiée à deux centres devrait permettre d'éviter les chocs stériques trop nombreux.

Afin de tester la méthode nous avons reconstruit la conformation de tous les peptides de 10 acides aminés se liant au MHC-I de IMGT/3Dstructure-DB. L'encombrement stérique important du MHC réduit grandement le nombre de conformations accessibles et seulement 60 peptides ont été générés par l'algorithme. La génération a été effectuée dans le sens direct, de la partie N-terminale vers la partie C-terminale. Le plus petit RMSD entre les conformations générées et la structure RX, ainsi que le numéro de la conformation de boucle correspondante ont été reportés dans la Table 4.2.

La plupart des ensembles conformationnels générés possèdent une conformation à moins de 2Å de RMSD (carbone alpha et oxygène de la chaîne principale) de la structure RX. Le rang de la meilleure conformation est, de plus, souvent inférieur à 40 (3 cas au-dessus), ce qui confirme que la génération de 60 conformations était suffisante. La méthode réussit donc à produire correctement des conformations de peptides proches des conformations natives.

4.3.4 Modélisation des boucles de reconnaissance du TR

Le même protocole que précédemment est appliqué à la modélisation des boucles hypervariables des TR de 18 complexes MHC/peptide/TR de structures tridimensionnelles connues. Les molécules d'eau ont été retirées car elles ne sont pas compatibles avec l'obtention de modèles de résolution moyenne et peuvent influencer de façon trop importante les conformations générées.

Nous faisons la supposition que l'espace accessible à chaque boucle est moins contraint stériquement que dans le cas de la modélisation du peptide. Par conséquent les boucles doivent présenter plus de conformations de paires de dièdres fréquemment rencontrées dans les structures RX. En conséquence, seulement 20 conformations ont été générées pour chaque CDR-IMGT.

Les Tables 4.3 et 4.4 présentent les meilleurs RMSD ainsi que les performances obtenues pour les calculs des CDR1-IMGT et CDR3-IMGT pour les TR dans les 18 complexes

Table 4.2: Moyenne quadratique des distances (RMSD) entre les carbones alpha et les oxygènes de la chaîne principale de structures RX du peptide et des modèles construits par la méthode LLIPA. Le numéro de génération (rang) est fourni. Soixante conformations ont été générées pour chaque structure RX et le meilleur RMSD est présenté. Les structures RX (identifiant dans la colonne ID), possèdent toutes des peptides de 10 acides aminés.

ID	RMSD	Rang
1bii	2,0	9
1 ddh	2,0	16
1hhh	1,5	31
1i4f	1,6	42
1jf1	1,7	33
1jgd	1,5	17
1n3n	1,5	28
1 qo 3	1,7	16
1qvo	2,2	41
$1 \mathrm{tmc}$	1,8	25
1wbx	1,4	38
1wby	1,5	43
2clr	1,6	23

TR/peptide/MHC. Les boucles de CDR1-IMGT ont une longueur comprise entre 7 et 9 et les RMSD (carbone alpha et oxygène de la chaîne principale) sont le plus souvent inférieurs à 2Å, ce qui rend compte de la proximité des conformations de la chaîne principale avec la structure native. Les temps de calculs sont majoritairement inférieurs à 4 minutes et 70% ont un temps inférieur à 3 minutes. L'objectif de rapidité et de précision semble donc atteint pour les boucles de CDR1-IMGT du TR.

Les RMSD présentés par les CDR3-IMGT sont plus importants que pour les CDR1-IMGT mais les boucles à modéliser sont bien plus grandes. Le RMSD moyen est de 2.6Å, ce qui, compte tenu de la longueur des boucles à modéliser, n'est pas excessif. Plus de la moitié des modélisations ont un temps de calcul inférieur à 5 minutes. L'algorithme LLIPA réussit donc à modéliser la plupart des CDR3-IMGT conformément aux objectifs fixés bien que à quatre reprises le temps de calcul ait été supérieur à 10 minutes et que certains RMSD approchent les 4Å sans toutefois les atteindre. Le comportement de l'algorithme lors de ces dernières modélisations est plus hétérogène en performance. Une paramétrisation supplémentaire est dans doute nécessaire pour régler au mieux le nombre maximal de conformations à générer suivant la taille et l'environnement structural de la boucle.

Plusieurs des meilleures conformations présentent un rang assez élevé. L'utilisation du rang comme critère de sélection relève de la supposition que les interactions de la boucle ne sont pas suffisamment énergétiques pour faire dévier les chaînes principales de leurs conformations d'énergie les plus basses. Les rangs élevés rencontrés dans les grandes



Figure 4.10: Modèle de la boucle CDR3-IMGT du V-BETA de 1d9k_B (en rouge) et la structure RX (CDR3-IMGT en bleu). Le modèle a été sélectionné parmi les vingt conformations qui ont été générées par la méthode LLIPA et le modèle présenté est celui qui est le plus proche de la structure RX (le RMSD des carbones alpha et de l'oxygène de la chaîne principale est de 2.0Å).

boucles tendraient à prouver que leur énergies d'interaction compensent l'utilisation de conformères de chaîne principale d'énergie plus élevée.

4.4 Conclusion

Afin de pouvoir identifier des peptides antigéniques dans un nombre important de protéines et étudier des répertoires de TR spécifiques d'un peptide/MHC, un protocole de modélisation des structures 3D de peptide/MHC et de TR/peptide/MHC a été entrepris. Parmi les différentes étapes de ce protocole, j'ai concentré mes efforts sur le développement de nouveaux algorithmes de modélisation de chaînes latérales et de modélisation de boucles. Ces algorithmes sont nécessaires afin d'atteindre une rapidité suffisante pour la modélisation des structures 3D de grands ensembles de peptide/MHC ou de TR/peptide/MHC.

La technique employée pour la reconstruction des chaînes latérales, SCHISMo, utilise des représentations de chaînes latérales de plus en plus complexes au fur et à mesure de l'affinement du modèle. Pour cela une librairie de conformations de chaînes latérales simplifiées a été construite. Elle présente trois niveaux interconnectés de simplifications Table 4.3: Moyenne quadratique des distances (RMSD) entre les atomes de structures RX du CDR1-IMGT de V-ALPHA et V-BETA et des modèles construits par la méthode LLIPA. Le numéro de génération (rang) et la durée de génération des 20 conformations (en minutes :secondes) sont indiqués. Les calculs ont été effectués sur un Pentium 1,6GHz.

				\mathbf{CDI}	R1-IMGT				
		V-AL	PHA			V-BETA			
ID	Taille	RMSD	Rang	Temps	Taille	RMSD	Rang	Temps	
1ao7	8	12	5	1.26	7	12	11	1.06	
1bd2	8	1.1	9	1 :15	7	1.3	6	1 :21	
1d9k	8	1.3	9	10:39	8	1.5	13	6:21	
1 fo 0	9	$3,\!0$	7	0:39	8	1,7	3	0:50	
1fyt	8	1,8	6	0:41	7	$1,\!5$	14	1:44	
1g6r	8	$2,\!1$	12	3 : 19	7	$1,\!3$	7	2:42	
1j 8 h	8	$1,\!6$	11	1:04	7	1,6	17	2:05	
1jtr	8	1,5	18	1:40	7	1,1	4	3 : 35	
1kj2	8	$0,\!9$	13	4:40	8	1,8	18	2:12	
1lp 9					7	1,5	2	1:45	
1 mi 5	9	2,3	13	1:51	7	$1,\!3$	15	7:12	
1mwa	8	1,5	18	1:55	7	1,1	4	3 : 30	
1nam	9	1,4	4	0:27	8	1,9	6	0 : 55	
10ga									
$1 \mathrm{qrn}$	8	1,7	9	1:27	7	1,5	19	1:40	
$1 \mathrm{qse}$	8	$1,\!6$	16	3:04	7	1,2	14	1 : 09	
1 qs f	8	$1,\!8$	17	1:14	7	$1,\!3$	15	1:57	
2ckb	8	2,3	4	2:36	7	1,2	5	3:42	

et les conformations sont classées par ordre de représentativité dans des structures expérimentales connues. L'algorithme de recherche fait intervenir une étape stochastique afin de réduire le nombre de chocs stériques. L'algorithme a été testé pour l'obtention de structures 3D de peptide/MHC-II qui possèdent la particularité d'avoir un peptide dont la conformation de chaîne principale est très bien conservée dans le site de liaison du MHC (Section 3.5.1 page 70). L'algorithme obtient une précision légèrement supérieure et une vitesse supérieure à SCWRL, l'algorithme rapide le plus couramment utilisé. SCHISMo retourne également un résultat dans les cas ou SCWRL se trouve confronté à un problème de chocs stériques déjà présents dans la structure native, ce qui prouve que SCHISMo est mieux à même de fonctionner avec des modèles de résolution moyenne. D'autre part cet algorithme est destiné à être intégré dans les différentes étapes de la modélisation du TR/peptide/MHC, lors de la construction des boucles et lors du docking.

L'algorithme de modélisation de boucles, LLIPA, utilise une bibliothèque de conformations de dipeptides. Cette bibliothèque a été conçue pour parcourir l'espace le plus efficaTable 4.4: Moyenne quadratique des distances (RMSD) entre les atomes de structures RX du CDR3-IMGT de V-ALPHA et V-BETA et des modèles construits par la méthode LLIPA. Le numéro de génération (rang) et la durée de génération des 20 conformations (en minutes :secondes) sont indiqués. Les calculs ont été effectués sur un Pentium 1.6GHz.

				\mathbf{CD}	R3-IMGT			
		V-AL	PHA			V-BETA		
ID	Taille	RMSD	Rang	Temps	Taille	RMSD	Rang	Temps
1ao7	13	3.1	1	2:54	16	3.8	13	4 :28
$1 \mathrm{bd} 2$	12	2.4	18	2:56	15	2.1	16	2:08
1 d9k	12	1.8	5	5:45	15	2.0	6	6:22
1 fo 0	16	3.6	13	15 : 06	14	3.0	15	4:55
1fyt	15	3.1	12	10 : 18	14	2.3	3	1:04
1g6r	12	2.1	4	4:09	11	1.6	10	4:35
1j 8 h	15	3.3	1	8 :20	14	2.3	20	2:43
1jtr	12	1.9	20	5:42	11	1.7	16	4:12
1kj2	13	2.3	11	6:32				
1lp9					13	3.2	16	2:52
1 mi 5	16	3.3	13	14:28	13	3.1	4	2:38
1mwa	12	1.9	20	6:21	11	1.7	16	4:44
1nam	16	2.8	7	12:45	14	2.9	8	5:48
10ga								
1qrn	13	3.1	1	3 : 56	16	3.0	13	2:51
1 qse	13	3.9	20	2:15	16	3.8	9	4:10
1 qs f	12	3.3	10	2:43	16	3.5	6	4:29
$2 \mathrm{ckb}$	12	2.7	15	5:51	16	1.5	17	5:28

cement possible et ses conformations ont été classées suivant leur représentativité dans les structures expérimentales. La recherche conformationnelle construit d'abord les boucles avec les conformations de dipeptides les plus observées avant d'utiliser les moins représentées. L'algorithme a été appliqué à la modélisation des peptides dans les peptide/MHC-I et à la modélisation des boucles du TR dans les TR/peptide/MHC. Pour la modélisation des peptides/MHC-I, l'algorithme réussit à générer des conformations de chaîne latérale le plus souvent à moins de 2Å de RMSD de la structure expérimentale (un cas à 2,2Å). Pour les TR, la modélisation des boucles de petites tailles, les CDR1-IMGT des TR dont la longueur varie entre 7 et 9 acides aminés dans les structures 3D étudiées, produit également le plus souvent des conformations à moins de 2Å de RMSD de la structure expérimentale. Les boucles CDR3-IMGT de TR, dont la longueur s'étale de 11 à 16 acides aminés dans les structures 3D étudiées, obtiennent des RMSD aux environs de 3Å, ce qui est positif compte tenu de la flexibilité des boucles de cette taille. Il faut noter que l'algorithme réussit à trouver un modèle avec un RMSD de 1,5Å pour une boucle de 16 acides aminés.

Afin d'achever le protocole de prédiction de structures 3D de peptide/MHC et de TR/peptide/MHC, deux étapes doivent encore être réalisées. L'évaluation de l'énergie de liaison des peptides au MHC est cruciale afin de pouvoir classer les peptides antigéniques potentiels et également pour classer les différents TR se liant à un même peptide/MHC. Pour les peptide/MHC-II, j'ai tenté d'utiliser une fonction d'évaluation très simple, donnant un poids constant à chaque type d'interaction à l'interface, mais la sélectivité semble très faible. L'utilisation de la fonction d'énergie spécifique, Fresno [121], devrait permettre d'améliorer nettement la sélectivité. Enfin, des fonctions d'énergie plus complexes, par exemple MM-GBSA ou MM-PBSA [188], devraient permettre une classification optimale des peptides pré-sélectionnés grâce à Fresno. Pour les TR/peptide/MHC l'étape de docking du TR sur le complexe peptide/MHC reste à réaliser. Celle-ci pourra être d'abord effectuée grâce à la représentation à deux centres de SCHISMo afin de réduire les chocs stériques intempestifs (soft docking) ainsi qu'une approche multicopie des boucles [139] du TR reconstruites par LLIPA. L'intégration de SCHISMo et LLIPA dans d'autres algorithmes est très facile car elles sont toutes deux implémentées dans des librairies C. L'approche du TR pourra être orientée par les statistiques sur la position relative du TR et du MHC (Section 3.6.1 page 79) et des filtres pourront être construits en se basant sur les analyses de contacts de IMGT/3Dstructure-DB à l'interface TR/peptide/MHC (Section 3.6.3 page 82).

Conclusion.

Au cours de ma thèse j'ai développé la base de données IMGT/3Dstructure-DB, qui rassemble et analyse les structures 3D d'IG, de TR, de MHC et de RPI. Grâce à la standardisation des données de IMGT/3Dstructure-DB, j'ai pu mettre en place un environnement informatique permettant la mesure et l'établissement de statistiques sur les caractéristiques structurales des IG, des TR et des MHC. Finalement j'ai implémenté des méthodes de modélisation applicables aux structures 3D de peptide/MHC et de TR/peptide/MHC, l'utilisation de ces méthodes étant justifiée par les caractéristiques structurales communes de ces complexes.

La base de données IMGT/3Dstructure-DB se conforme aux règles et au vocabulaire de IMGT-ONTOLOGY, ce qui lui confère un haut niveau de standardisation, s'étendant de la description des récepteurs d'antigènes jusqu'aux positions de leurs domaines. Cette standardisation apparaît comme un élément crucial en bioinformatique depuis l'avènement de techniques expérimentales produisant des quantités importantes de données génomiques, protéomiques et bientôt structurales. De plus des programmes d'annotation partiellement automatisés sont nécessaires, car les annotateurs, en nombre souvent réduit, n'ont pas la capacité temporelle de traiter toutes ces informations et car ils doivent pouvoir vérifier, modifier et/ou orienter l'annotation automatique dans certains cas particuliers. L'architecture de IMGT/3Dstructure-DB, simple et naturelle, et son programme d'annotation, souple et entièrement paramétrable, répondent à ces principes. L'annotation automatique apporte également l'avantage de contribuer à la cohérence des données dans la base de données. IMGT/3Dstructure-DB a également permis de développer IMGT-ONTOLOGY, par l'introduction de nouveaux termes de description et par l'amélioration de la numérotation unique des V-DOMAINs, C-DOMAINs et G-DOMAINs. Une description de plus en plus précise des IG, TR et MHC, qui sont des protéines hautement polymorphiques, nécessite en effet un travail d'ajustement constant au vue des nouvelles données de séquence et de structure 3D. L'introduction des RPI dans IMGT/3Dstructure-DB abonde dans ce sens, les structures 3D apparentées devant permettre de dégager les principes structuraux primordiaux gouvernant le repliement des domaines V, C et G.

L'analyse des structures 3D de récepteurs d'antigènes (IG, TR et MHC) par comparaison manuelle de paires de structures 3D est laborieuse et source d'erreurs. J'ai implémenté un environnement informatique qui autorise la comparaison de nombreuses structures 3D simultanément et qui remet automatiquement ces comparaisons à jour à chaque nouvelle version de IMGT/3Dstructure-DB. Ainsi les domaines de type immunoglobuline (domaines V et C) ont été comparés au niveau de leur topologie et de l'arrangement spatial des feuillets. Pour les domaines V, C et G, des alignements structuraux par paires ont mesuré la variabilité structurale à chaque position et les contacts conservés entre les positions ont été identifiés. L'arrangement des paires de V-DOMAINs et des paires de C-DOMAINs, constituant respectivement des V-PARTNER et C-PARTNER, a été évalué par la mesure des angles formés par les feuillets internes et par les feuillets externes. L'arrangement des domaines partenaires dans les récepteurs a été observé de la même manière. Finalement des analyses conformationnelles et de contacts ont été réalisées sur les interfaces des récepteurs d'antigènes. La comparaison des complexes peptide/MHC et IG/antigène permettent de tirer une conclusion quasiment définitive sur les caractéristiques moyennes de ces complexes. Par contre les complexes ternaires TR/peptide/MHC étant peu nombreux, leurs caractéristiques communes restent difficiles à discerner. Une interface Internet présente tous ces résultats de façon synthétique avec la possibilité d'avoir accès à toutes les données. Finalement ces observations globales permettent indirectement de contrôler la cohérence des entrées de IMGT/3Dstructure-DB et de détecter éventuellement des défauts d'annotation.

La raison principale qui a motivé le développement de nouvelles techniques de modélisations de complexes peptide/MHC et TR/peptide/MHC provient de plusieurs demandes qui m'ont été faites par des immunologistes. En effet les techniques actuelles de prédiction de peptides se liant à un MHC ont recours à de nombreuses approximations et se fondent pour la plupart uniquement sur la séquence, ce qui conduit à de nombreuses prédictions erronées. Les techniques de modélisation de structures 3D sont par contre trop lentes pour être applicables à un grand nombre de peptide/MHC. J'ai donc pris la décision de m'intéresser à une technique intermédiaire d'obtention rapide de structures 3D de résolution moyenne, ce qui permet une présélection avant de servir de point de départ à des techniques de minimisation et de dynamique moléculaire. L'algorithme SCHISMo de prédiction de chaînes latérales réussit à fournir une structure 3D de bonne résolution en un temps très court. Une fois sa bibliothèque de chaînes latérales chargée en mémoire, ce qui représente la majeur partie de la durée d'exécution lors de la modélisation d'une seule structure, un nombre important de peptide/MHC-II peuvent être modélisés très rapidement. Ce temps est d'ailleurs tout à fait compatible avec celui d'un service Web et il est envisageable de créer un serveur de modélisation de chaînes latérales qui permettrait à un utilisateur d'obtenir la structure 3D d'un peptide/MHC-II à partir de sa séquence de peptide et du nom des allèles de MHC-II à utiliser (les séquences étant déjà dans IMGT/3Dstructure-DB). L'algorithme LLIPA, bien que relativement rapide, ne permet pas encore un tel développement car la construction des peptide/MHC-I et des boucles de TR nécessitent encore plusieurs minutes. Par contre cet algorithme peut être utilisé de façon assez générale pour obtenir un nombre restreint de conformations de boucles, qui pourraient être classées après minimisation par des logiciels tels que AMBER, CHARMM ou GROMOS. Finalement les deux algorithmes, SCHISMo et LLIPA, que j'ai développés, peuvent très facilement être intégrés dans d'autres algorithmes de modélisation, afin d'utiliser les représentations simplifiées de chaînes latérales et de créer rapidement plusieurs conformations alternatives d'une boucle. Le docking du TR sur un peptide/MHC, qui finaliserait la modélisation des structures 3D de TR/peptide/MHC, pourrait utiliser la représentation à deux centres dans une première étape pour une recherche grossière puis la représentation multicentres et tous atomes pour affiner les modèles. Les conformations multiples des boucles de TR pourront être utilisées simultanément par une approche multicopie afin de simuler leur flexibilité.

La création d'une base de données structurale spécialisée, l'obtention de statistiques sur ses entrées et la création d'algorithmes de modélisation moléculaire adaptés forment un ensemble cohérent qui peut être appliqué à d'autres familles de protéines. La standardisation des descriptions des protéines, de leurs chaînes, domaines et positions est un prérequis essentiel à de telles investigations. IMGT-ONTOLOGY résulte d'un travail minutieux qui s'étale sur plus d'une vingtaine d'années et des démarches similaires, bénéficiant de cette expérience, peuvent être adaptées à d'autres problématiques. Les IG et les MHC font partie des molécules les plus étudiées structuralement et le nombre important de leur structures 3D nécessite une automatisation de leur analyses. L'explosion annoncée du nombre de données structurales engendrera l'apparition de nombreuses bases de données structurales spécialisées et j'espère que les travaux effectués durant ma thèse pourront servir à leur développement.

Bibliographie

- [1] Lefranc M.-P. and Lefranc G. (2001). The immunoglobulin. *FactsBook Academic Press*.
- [2] Lefranc M.-P. and Lefranc G. (2001). The T cell receptor. FactsBook Academic Press.
- [3] Yewdell J.W. and Bennink J.R. (1992). Cell biology of antigen processing and presentation to major histocompatibility complex class I molecule-restricted T lymphocytes. Adv Immunol 52, 1-123.
- [4] Bjorkman P.J. and Parham P. (1990). Structure, function, and diversity of class I major histocompatibility complex molecules. Annual Review in Biochemistry 59, 253-288.
- [5] Parham P., Lomen C.E., Lawlor D.A., Ways J.P., Holmes N., Coppin H.L., Salter R.D., Wan A.M. and Ennis P.D. (1988). Nature of polymorphism in HLA-A, -B, and -C molecules. *Proc Natl Acad Sciences USA* 85, 4005-4009.
- [6] Lefranc M.-P., Giudicelli V., Kaas Q., Duprat E., Jabado-Michaloud J., Scaviner D., Ginestoux C., Clement O., Chaume D. and Lefranc G. (2005). IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* 33, D593-D597.
- [7] Lefranc M.-P., Clement O., Kaas Q., Duprat E., Chastellan P., Coelho I., Combres K., Ginestoux C., Giudicelli V., Chaume D. and Lefranc G. (2005). IMGT-Choreography for Immunogenetics and Immunoinformatics. *In Silico Biol* (in press).
- [8] Giudicelli V. and Lefranc M.-P. (1999). Ontology for immunogenetics : the IMGT-ONTOLOGY. *Bioinformatics* 15, 1047-1054.
- [9] Lesk A.M. and Chothia C. (1982). Evolution of proteins formed by beta-sheets. II. The core of the immunoglobulin domains. J Mol Biol 160, 325-342.
- [10] Lefranc M.-P., Pommie C., Ruiz M., Giudicelli V., Foulquier E., Truong L., Thouvenin-Contet V. and Lefranc G. (2003). IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 27, 55-77.
- [11] Lefranc M.-P., Pommie C., Kaas Q., Duprat E., Bosc N., Guiraudou D., Jean C., Ruiz M., Da Piedade L., Rouard M., Foulquier E., Thouvenin V. and Lefranc G. (2005). IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev Comp Immunol* (in press).
- [12] Ruiz M. and Lefranc M.-P. (2002). IMGT gene identification and Colliers de Perles of human immunoglobulins with known 3D structures. *Immunogenetics* 53, 857-883.

- [13] Boulot G., Bentley G.A., Karjalainen K. and Mariuzza R.A. (1994). Crystallization and preliminary X-ray diffraction analysis of the beta-chain of a T-cell antigen receptor. J Mol Biol 235, 795-797.
- [14] Fields B.A., Ober B., Malchiodi E.L., Lebedeva M.I., Braden B.C., Ysern X., Kim J.K., Shao X., Ward E.S. and Mariuzza R.A. (1995). Crystal structure of the V alpha domain of a T cell antigen receptor. *Science* 270, 1821-1824.
- [15] Garcia K.C., Degano M., Stanfield R.L., Brunmark A., Jackson M.R., Peterson P.A., Teyton L. and Wilson I.A. (1996). An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. *Science* 274, 209-219.
- [16] Bjorkman P.J., Strominger J.L. and Wiley D.C. (1985). Crystallization and X-ray diffraction studies on the histocompatibility antigens HLA-A2 and HLA-A28 from human cell membranes. J Mol Biol 186, 205-210.
- [17] 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.
- [18] Lefranc M.-P., Duprat E., Kaas Q., Tranne M., Thiriot A. and Lefranc G. (2005). IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily MhcSF G-LIKE-DOMAIN. *Dev Comp Immunol* 29, 917-938.
- [19] Allcorn L.C. and Martin A.C. (2002). SACS-self-maintaining database of antibody crystal structure information. *Bioinformatics* 18, 175-181.
- [20] Honegger A. and Pluckthun A. (2001). Yet another numbering scheme for immunoglobulin variable domains : an automatic modelling and analysis tool. J Mol Biol 309, 657-670.
- [21] Kaas Q., Ruiz M. and Lefranc M.-P. (2004). IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. *Nucleic Acids Res* 32, 208-210.
- [22] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. and Bourne P.E. (2000). The Protein Data Bank. *Nucleic Acids Res* 28, 235-242.
- [23] Chen J., Anderson J.B., DeWeese-Scott C., Fedorova N.D., Geer L.Y., He S., Hurwitz D.I., Jackson J.D., Jacobs A.R., Lanczycki C.J., Liebert C.A., Liu C., Madej T., Marchler-Bauer A., Marchler G.H., Mazumder R., Nikolskaya A.N., Rao B.S., Panchenko A.R., Shoemaker B.A., Simonyan V., Song J.S., Thiessen P.A., Vasudevan S., Wang Y., Yamashita R.A., Yin J.J. and Bryant S.H. (2003). MMDB : Entrez's 3D-structure database. *Nucleic Acids Res* **31**, 474-477.
- [24] Bairoch A., Apweiler R., Wu C.H., Barker W.C., Boeckmann B., Ferro S., Gasteiger E., Huang H., Lopez R., Magrane M., Martin M.J. and Natale D.A. (2005). The Universal Protein Resource (UniProt). *Nucleic Acids Res* 33, D154-D159.
- [25] Shindyalov I.N. and Bourne P.E. (1998). Protein structure alignment by incremental combnatorial extension (CE) of the optimal path. *Protein Engineering* **11**, 739-747.
- [26] Smith T.F. and Waterman M.S. (1981). Identification of common molecular subsequences. J Mol Biol 147, 195-197.

- [27] Pearson W.R. and Lipman D.J. (1988). Improved tools for biological sequence comparison. Proc Natl Acad Sciences USA 85, 2444-2448.
- [28] Frishman D. and Argos P. (1995). Knowledge-based protein secondary structure assignment. Proteins 23, 566-579.
- [29] Creighton T.E. (1992). Proteins : Structures and Molecular Properties, 2nd Ed. WH Freeman ISBN 0-7617-7030-X.
- [30] McDonald I.K. and Thornton J.M. (1994). Satisfying hydrogen bonding potential in proteins. J Mol Biol 238, 777-793.
- [31] Kaas Q. and Lefranc M.P. (2005). Interactive IMGT on-line database and tool for the structural analysis of immunoglobulins, T cell receptors, MHC and related proteins of the immune system. *Nova Science* (in press).
- [32] Gelly J.C., Gracy J., Kaas Q., Le-Nguyen D., Heitz A. and Chiche L. (2004). The KNOTTIN website and database : a new information system dedicated to the knottin scaffold. *Nucleic Acids Res* 32, 156-159.
- [33] Kaas Q. and Lefranc M.-P. (2005). T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. In Silico Biology 5, 0046.
- [34] Collis A.V., Brouwer A.P. and Martin A.C. (2003). Analysis of the antigen combining site : correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. J Mol Biol 325, 337-354.
- [35] Pellequer J.L., Chen S., Roberts V.A., Tainer J.A. and Getzoff E.D. (1999). Unraveling the effect of changes in conformation and compactness at the antibody V(L)-V(H) interface upon antigen binding. J Mol Recognit 12, 267-275.
- [36] Fremont H.D., Rees W.A. and Kozono H. (1996). Biophysical studies of T-cell receptors and their ligands. *Curr Opin Immunol* 8, 93-100.
- [37] Li H., Lebedeva M.I., Llera A.S., Fields B.A., Brenner M.B. and Mariuzza R.A. (1998). Structure of the Vdelta domain of a human gammadelta T-cell antigen receptor. *Nature* **391**, 502-506.
- [38] Allison T.J., Winter C.C., Fournie J.J., Bonneville M. and Garboczi D.N. (2001). Structure of a human gammadelta T-cell antigen receptor. *Nature* 411, 820-824.
- [39] Garboczi D.N., Ghosh P., Utz U., Fan Q.R., Biddison W.E. and Wiley D.C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 384, 134-141.
- [40] Stura E.A., Matsumura M., Fremont D.H., Saito Y., Peterson P.A. and Wilson I.A. (1992). Crystallization of murine major histocompatibility complex class I H-2Kb with single peptides. J Mol Biol 228, 975-982.
- [41] Fremont D.H., Hendrickson W.A., Marrack P. and Kappler J. (1996). Structures of an MHC class II molecule with covalently bound single peptides. *Science* 272, 1001-1004.
- [42] Pommie C., Levadoux S., Sabatier R., Lefranc G. and Lefranc M.-P. (2004). IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. J Molec Recog 17, 17-32.

- [43] Poupon A. and Mornon J.P. (1998). Populations of hydrophobic amino acids within protein globular domains : identification of conserved. *Proteins* **33**, 329-342.
- [44] Richardson J.S. (1981). The anatomy and taxonomy of protein structure. Adv Protein Chem 34, 167-339.
- [45] Bork P., Holm L. and Sander C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. J Mol Biol 242, 309-320.
- [46] Bentley G.A., Boulot G., Karjalainen K. and Mariuzza R.A. (1995). Crystal structure of the beta chain of a T cell antigen receptor. *Science* 267, 1984-1987.
- [47] Wang J., Lim K., Smoylar 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 Reinhertz E.L. (1998). Atomic structure of an alpha-beta T cell receptor (TCR) heterodimer in complex with an anti-TCR Fab fragment derived from a mitogenic antibody. *EMBO J* 17, 10-26.
- [48] McLachlan A.D. (1982). Rapid comparison of protein structures. Acta Cryst A38, 871-873.
- [49] Honegger A. and Pluckthun A. (2001). The influence of the buried glutamine or glutamate residue in position 6 on the structure of immunoglobulin variable domain. J Mol Biol 309, 687-699.
- [50] Bahadur R.P., Chakrabarti P., Rodier F. and Janin J. (2004). A dissection of specific and non-specific protein-protein interfaces. J Mol Biol 336, 943-955.
- [51] Davies D.R. and Cohen G.H. (1996). Interactions of protein antigens with antibodies. Proc Natl Acad Sciences USA 93, 7-12.
- [52] Davies D.R., Padlan E.A. and Sheriff S. (1990). Antibody-antigen complexes. Annu Rev Biochem 59, 439-473.
- [53] Kaas Q., Chiche L. and Lefranc M.P. (2005). IMGT unique numbering for standardized contact analysis of immunoglobulin/antigen and T cell receptor/peptide/MHC complexes. BIOINFO2005, Proceedings of the 2005 International Conference of InCoB, AASBi and KSBI, Lee D., Wong L., Kim D.-W., Tan T.W, Lee K.-H. (KAIST PRESS, Daejon, Korea).
- [54] Gopalakrishnan B. and Roques B.P. (1992). Do antigenic peptides have a unique sense of direction inside the MHC binding groove? A molecular modelling study. *FEBS Lett* 303, 224-228.
- [55] Wang C.R., Castano A.R., Peterson P.A., Slaughter C., Lindahl K.F. and Deisenhofer J. (1995). Nonclassical binding of formylated peptide in crystal structure of the MHC class Ib molecule H2-M3. *Cell* 82, 655-664.
- [56] Scott C.A., Peterson P.A., Teyton L. and Wilson I.A. (1998). Crystal structures of two I-Ad-peptide complexes reveal that high affinity can be achieved without large anchor residues. *Immunity* 8, 319-329.
- [57] Schueler-Furman O., Elber R. and Margalit H. (1998). Knowledge-based structure prediction of MHC class I bound peptides : a study of 23 complexes. *Folding* **3**, 549-564.

- [58] Adrian P.E., Rajaseger G., Mathura V.S., Sakharkar M.K. and Kangueane P. (2002). Types of inter-atomic interactions at the MHC-peptide interface : identifying commonality from accumulated data. *BMC Struct Biol* 2, 2.
- [59] Bouvier M. and Wiley D.C. (1994). Importance of peptide amino and carboxyl termini to the stability of MHC class I molecules. *Science* **265**, 398-402.
- [60] Achour A., Michaelsson J., Harris R.A., Odeberg J., Grufman P., Sandberg J.K., Levitsky V., Karre K., Sandalova T. and Schneider G. (2002). A structural basis for LCMV immune evasion : subversion of H-2Db and H-2Kb presentation of gp revealed by comparative crystal structure analyses. *Immunity* 17, 757-768.
- [61] Falk K., Rotzschke O., Stevanovic S., Jung G. and Rammensee H.G. (1991). Allelespecific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 351, 290-296.
- [62] Mandelboim O., Bar-Haim E., Vadai E., Fridkin M. and Eisenbach L. (1997). Identification of shared tumor-associated antigen peptides between two spontaneous lung carcinomas. J Immunol 159, 6030-6036.
- [63] Apostolopoulos V., Yu M., Corper A.L., Teyton L., Pieters G.A., McKenzie I.F.C. and Wilson I.A. (2002). Crystal structure of a non-canonical Low-affinity Peptide Complexed with MHC Class I : A New approach for vaccine design. J Mol Biol 318, 1293-1305.
- [64] Gulukota K., Sidney J., Sette A. and DeLisi C. (1997). Two complementary methods for predicting peptides binding major histocompatibility complex molecules. J Mol Biol 267, 1258-1267.
- [65] Zhang C., Anderson A. and DeLisi C. (1998). Structural principles that govern the peptide-binding motifs of class I MHC molecules. J Mol Biol 281,929-947.
- [66] Vasmatzis G., Cornette J., Sezerman U. and DeLisi C. (1996). TcR recognition of the MHC-eptide dimer : structural properties of a ternary complex. J Mol Biol 261, 72-89.
- [67] Sim B.C., Zerva L., Greene M.I. and Gascoigne N.R. (1996). Control of MHC restriction by TCR Valpha CDR1 and CDR2. *Science* 273, 963-964.
- [68] Kjer-Nielsen L., Clements C.S., Purcell A.W., Brooks A.G., Whisstock J.C., Burrows S.R., McCluskey J. and Rossjohn J. (2003). A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. *Immunity* 18, 53-64.
- [69] Teng M.K., Smolyar A., Tse A.G., Liu J.H., Liu J., Hussey R.E., Nathenson S.G., Chang H.C., Reinherz E.L. and Wang J.H. (1998). Identification of a common docking topology with substantial variation among different TCR-peptide-MHC complexes. *Curr Biol* 8, 409-412.
- [70] 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.

- [71] Stewart-Jones G.B.E., McMichael A.J., Bell J.I., Stuart D.I. and Jones E.Y. (2003). A structural basis for immunodominant human T cell receptor recognition. *Nature Immunology* 4, 657-663.
- [72] Degano M., Garcia K.C., Apostolopoulos V., Rudolph M.G., Teyton L. and Wilson I.A. (2000). A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. *Immunity* 12, 251-261.
- [73] Lo Conte L., Chothia C. and Janin J. (1999). The atomic structure of protein-protein recognition sites. J Mol Biol 285, 2177-2198.
- [74] Tissot A.C., Pecorari F. and Pluckthun A. (2000). Characterizing the functionality of recombinant T-cell receptors in vitro : a pMHC tetramer based approach. *Journal* of *Immunological Methods* 236, 147-165.
- [75] Davis M.M. (2002). A New Trigger for T Cells. Cell 110, 285-287.
- [76] Wang J.H., Meijers R., Xiong Y., Liu J.H., Sakihama T., Zhang R., Joachimiak A. and Reinherz E.L. (2001). Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. *Proc Natl Acad Sciences USA* 98, 10799-10804.
- [77] Hamrouni A., Aublin A., Guillaume P. and Maryanski J.L. (2003). T cell receptor gene rearrangement lineage analysis reveals clues for the origin of highly restricted antigen-specific repertoires. J Exp Med 197, 601-614.
- [78] Fernandez-Recio J., Totrov M. and Abagyan R. (2002). Soft protein-protein docking in internal coordinates. *Protein Science* 11, 280-291.
- [79] Miranker A. and Karplus M. (1991). Functionality maps of binding sites : a multiple copy simultaneous search method. *Proteins* **11**, 29-34.
- [80] Rammensee H.G., Friede T. and Stevanovic S. (1995). MHC ligands and peptides motifs : first listing. *Immunogenetics* 41, 178-228.
- [81] Rammensee H.G., Bachmann J., Emmerich N.P.N., Bachor O.A. and Stevanovie S. (1999). SYFPEITHI : database for MHC ligands and peptide motifs. *Immunogenetics* 50, 213-219.
- [82] Blythe M.J., Doytchinova I.A. and Flower D.R. (2002). JenPep : a database of quantitative functional peptide data for immunology. *Bioinformatics* 18, 434-439.
- [83] Brusic V., Rudy G. and Harrison L.C. (1998). MHCPEP, a database of MHC-binding peptides : update 1997. Nucleic Acids Res 26, 368-371.
- [84] Borras-Cuesta F., Golvano J., Garcia-Granero M., Sarobe P., Riezu-Boj J., Huarte E. and Lasarte J. (2000). Specific and general HLA-DR binding motifs : comparison of algorithms. *Hum Immunol* 61, 266-278.
- [85] Reche P.A., Glutting J.P. and Reinherz E.L. (2002). Prediction of MHC class I binding peptides using profile motifs. *Hum Immunol* 63, 701-709.
- [86] Thompson J.D., Higgins D.G. and Gibson T.J. (1994). Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Comput Appl Biosci* 10, 19-29.

- [87] Henikoff J.G. and Henikoff S. (1996). Using substitution probabilities to improve position-specific scoring matrices. *Comput Appl Biosci* **12**, 135-143.
- [88] Parker K.C., Bednarek M.A. and Coligan J.E. (1994). Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide sidechains. J Immunol 152, 163-175.
- [89] Singh H. and Raghava G.P. (2003). ProPred1 : prediction of promiscuous MHC Class-I binding sites. *Bioinformatics* 19, 1009-1014.
- [90] Doytchinova I.A. and Flower D.R. (2003). Towards the in silico identification of class II restricted T-cell epitopes : a partial least squares iterative self-consistent algorithm for affinity prediction. *Bioinformatics* 19, 2263-2270.
- [91] Guan P., Doytchinova I.A., Zygouri C. and Flower D.R. (2003). MHCPred : A server for quantitative prediction of peptide-MHC binding. *Nucleic Acids Res* **31**, 3621-3624.
- [92] Mallios R.R. (2001). Predicting class II MHC/peptide multi-level binding with an iterative stepwise discriminant analysis meta-algorithm. *Bioinformatics* 17, 942-948.
- [93] Udaka K., Wiesmuller K.H., Kienle S., Jung G., Tamamura H., Yamagishi H., Okumura K., Walden P., Suto T. and Kawasaki T. (2000). An automated prediction of MHC class I-binding peptides based on positional scanning with peptide libraries. *Immunogenetics* 51, 816-828.
- [94] Adams H.P. and Koziol J.A. (1995). Prediction of binding to MHC class I molecules. J Immunol Methods 185, 181-190.
- [95] Desmet J., De Maeyer M., Hazes B. and Lasters I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* 356, 539-542.
- [96] Donnes P. and Elofsson A. (2002). Prediction of MHC class I binding peptides, using SVMHC. BMC Bioinformatics 3, 25.
- [97] Altuvia Y., Schueler O. and Margalit H. (1995). Ranking potential binding peptides to MHC molecules by a computational threading approach. J Mol Biol 249, 244-250.
- [98] Cano P. and Fan B. (2001). A geometric and algebraic view of MHC-peptide complexes and their binding properties. *BMC Struct Biol* 1, 2.
- [99] Miyazawa S. and Jernigan R.L. (1996). Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 256, 623-644.
- [100] Miyazawa S. and Jernigan R.L. (1985). Estimation of effective interresidue contact energies from protein crystal structures : quasi-chemical approximation. *Macromole*cules 18, 534-552.
- [101] Betancourt M.R. and Thirumalai D. (1999). Pair potentials for protein folding : choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Science* 8, 361-369.
- [102] Schueler-Furman O., Altuvia Y., Sette A. and Margalit H. (2000). Structure-based prediction of binding peptides to MHC class I molecules : Application to a broad range of MHC alleles. *Protein Science* 9, 1838-1846.

- [103] Vasmatzis G., Zhang C., Cornette J.L. and DeLisi C. (1996). Computational determination of side chain specificity for pockets in class I MHC molecules. *Mol Immunol* 33(16), 1231-1239.
- [104] Ogata K., Jaramillo A., Cohen W., Briand J.-P., Connan F., Choppin J., Muller S. and and S. J. W. (2002). Automatic sequence design of major histocompatibility complex class I binding peptides impairing CD8 and T cell recognition. J Biol Chem 10, 1281-1290.
- [105] Wernisch L., Hery S. and Wodak S.J. (2000). Automatic protein design with all atom force-fields by exact and heuristic optimization. J Mol Biol 301, 713-736.
- [106] Bower M.J., Cohen F.E. and Dunbrack Jr R.L. (1997). Prediction of protein sidechain rotamers from a backbone-dependent rotamer library : a new homology modeling tool. J Mol Biol 267, 1268-1282.
- [107] Kangueane P., Sakharkar M.K., Lim K.S., Hao H., Lin K., Chee R.E. and Kolatkar P.R. (2000). Knowledge-based grouping of modeled HLA peptide complexes. *Hum Immunol* 61, 460-466.
- [108] Lee C. and McConnell H.M. (1995). A general model of invariant chain association with class II major histocompatibility complex proteins. *Proc Natl Acad Sciences USA* 92, 8269-8273.
- [109] Lee C. (1994). Predicting Protein Mutant Energetics by Self-consistent Ensemble Optimization. J Mol Biol 236, 918-939.
- [110] Rognan D., Stryhn A., Fugger L., Lyngbæk S., Engberg J., Sejer Andersen P. and Buus S. (2000). Modeling the interactions of a peptide-major histocompatibility class I ligand with its receptors. I. Recognition by two GalphaGbeta T cell receptors. J Comput Aided Mol Design 14, 53-69.
- [111] Zhang C., Vasmatzis G., Cornette J.L. and DeLisi C. (1997). Determination of atomic desolvation energies from the structures of crystallized proteins. J Mol Biol 267, 707-726.
- [112] Brooks B.R., Bruccoleri R.E., Olafson B.D., States D.J., Swaminathan S. and Karplus M. (1983). CHARMM : A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J Comput Chem 4, 187-217.
- [113] Jorgensen W.L. and Tirado-Rives J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 110, 1657-1666.
- [114] Tong J.C., Tan T.W. and Ranganathan S. (2004). Modeling the structure of bound peptide ligands to major histocompatibility complex. *Protein Science* 13, 2523-2532.
- [115] Rosenfeld R., Zheng Q., Vajda S. and DeLisi C. (1993). Computing the Structure of Bound Peptides : Application to Antigen Recognition by Class I Major Histocompatibility Complex Receptors. J Mol Biol 234, 515-521.
- [116] Desmet J., Wilson I.A., Joniau M., De Maeyer M. and Lasters I. (1997). Computation of the binding of fully flexible peptides to proteins with flexible side chains. FASEB J 11, 164-172.

- [117] Sezerman U., Vajda S., Cornette J. and DeLisi C. (1993). Toward computational determination of peptide-receptor structure. *Protein Science* 2, 1827-1843.
- [118] Bruccoleri R.E. and Karplus M. (1987). Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* 26, 137-168.
- [119] Sali A. and Blundell T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234, 779-815.
- [120] Abagyan R. and Totrov M. (1994). Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol 235, 983-1002.
- [121] Rognan D., Lauemoller S.L., Holm A., Buus S. and Tschinke V. (1999). Predicting binding affinities of protein ligands from three-dimensional models : application to peptide binding to class I major histocompatibility proteins. J Med Chem 42, 4650-4658.
- [122] Logean A., Sette A. and Rognan D. (2001). Customized versus universal scoring functions : application to class I MHC-peptide binding free energy predictions. *Bioorg Med Chem Let* 11, 675-679.
- [123] Levitt M. (1976). A Simplified Representation of Protein Conformations for Rapid Simulation of Protein Folding. J Mol Biol 104, 59-107.
- [124] Tozzini V. (2005). Coarse-grained models for proteins. Curr Opin Struct Biol 15, 144-150.
- [125] Xiang Z. and Honig B. (2001). Extending the accuracy limits of prediction for sidechain conformations. J Mol Biol 311, 421-430.
- [126] Tuffery P., Etchebest C., Hazout S. and Lavery R. (1991). A new approach to the rapid determination of protein side chain conformations. J Biomol Struct Dyn 8, 1267-1289.
- [127] Canutescu A.A., Shelenkov A.A. and Dunbrack Jr R.L. (2003). A graph-theory algorithm for rapid protein side-chain prediction. *Protein Science* **12**, 2001-2014.
- [128] Topham C.M., McLeod A., Eisenmenger F., Overington J.P., Johnson M.S. and Blundell T.L. (1993). Fragment ranking in modelling of protein structure. Conformationally constrained environmental amino acid substitution tables. J Mol Biol 229, 194-220.
- [129] Janin J. and Wodak S. (1978). Conformation of amino acid side-chains in proteins. J Mol Biol 125, 357-386.
- [130] Dunbrack RL J. and Karplus M. (1993). Backbone-dependent rotamer library for proteins. Application to side-chain prediction. J Mol Biol 230, 543-574.
- [131] Lovell S.C., Word J.M., Richardson J.S. and Richardson D.C. (2000). The penultimate rotamer library. *Proteins* 40, 389-408.
- [132] Vasquez M. (1996). Modeling side-chain conformation. Curr Opin Struct Biol 6, 217-221.

- [133] Lee C. and Levitt M. (1991). Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core. *Nature* **352**, 448-451.
- [134] Mendes J., Baptista A.M., Carrondo M.A. and Soares C.M. (1999). Improved modeling of side-chains in proteins with rotamer-based methods : a flexible rotamer model. *Proteins* 37, 530-543.
- [135] Liu Z., Jiang L., Gao Y., Liang S., Chen H., Han Y. and Lai L. (2003). Beyond the rotamer library : genetic algorithm combined with the disturbing mutation process for upbuilding protein side-chains. *Proteins* 50, 49-62.
- [136] Eyal E., Najmanovich R., Edelman M. and Sobolev V. (2003). Protein Side-Chain Rearrangement in Regions of Point Mutations. *Proteins* 50, 272-282.
- [137] Harbury P.B., Tidor B. and Kim P.S. (1995). Repacking protein cores with backbone freedom : structure prediction for coiled coils. *Proc Natl Acad Sciences USA* 92, 8408-8412.
- [138] Harbury P.B., Plecs J.J., Tidor B., Alber T. and Kim P.S. (1998). High-resolution protein design with backbone freedom. *Science* 282, 1462-1467.
- [139] Zheng Q., Rosenfeld R., DeLisi C. and Kyle D.J. (1994). Multiple copy sampling in protein loop modeling : computational efficiency and sensitivity to dihedral angle perturbations. *Protein Science* 3, 493-506.
- [140] Keskin O. and Bahar I. (1998). Packing of sidechains in low-resolution models for proteins. *Folding* 3, 469-479.
- [141] Wallqvist A. and Ullner M. (1994). A simplified amino acid potential for use in structure predictions of proteins. *Proteins* 18, 267-280.
- [142] Bahar I. and Jernigan R.L. (1997). Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. J Mol Biol 266, 195-214.
- [143] Pearlman D.A., Case D.A., Cladwell J.W., Ross W.R., Cheatham III T.E., DeBolt S., Fergusson D., Seibel G. and Kollman P. (1995). AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comp Phys Commun* 91, 1-41.
- [144] Liang S. and Grishin N.V. (2002). Side-chain modeling with an optimized scoring function. *Protein Science* 11, 322-331.
- [145] Pillardy J., Arnautova Y.A., Czaplewski C., Gibson K.D. and Scheraga H.A. (2001). Conformation-family Monte Carlo : a new method for crystal structure prediction. *Proc Natl Acad Sciences USA* 98, 12351-12356.
- [146] Mendes J., Nagarajaram H.A., Soares C.M., Blundell T.L. and Carrondo M.A. (2001). Incorporating knowledge-based biases into an energy-based side-chain modeling method : application to comparative modeling of protein structure. *Biopolymers* 59, 72-86.

- [147] Mendes J., Baptista A.M., Carrondo M.A. and Soares C.M. (2001). Implicit solvation in the self-consistent mean field theory method : sidechain modelling and prediction of folding free energies of protein mutants. J Comput Aided Mol Des 15, 721-740.
- [148] Glick M., Rayan A. and Goldblum A. (2002). A stochastic algorithm for global optimization and for best populations : A test case of side chains in proteins. *Proc Natl* Acad Sciences USA 99, 703-708.
- [149] Voigt C.A., Gordon D.B. and Mayo S.L. (2000). Trading accuracy for speed : A quantitative comparison of search algorithms in protein sequence design. J Mol Biol 299, 789-803.
- [150] Gordon D.B. and Mayo S.L. (1999). Branch-and-terminate : a combinatorial optimization algorithm for protein design. Structure Fold Des 7, 1089-1098.
- [151] Looger L.L. and Hellinga H.W. (2001). Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable : implications for protein design and structural genomics. J Mol Biol 307, 429-445.
- [152] Eyrich V.A., Standley D.M., Felts A.K. and Friesner R.A. (1999). Protein tertiary structure prediction using a branch and bound algorithm. *Proteins* 35, 41-57.
- [153] Dunbrack Jr R.L. (1999). Comparative modeling of CASP3 targets using PSI-BLAST and SCWRL. Proteins Suppl 3, 81-87.
- [154] Word J.M., Lovell S.C., LaBean T.H., Taylor H.C., Zalis M.E., Presley B.K., Richardson J.S. and Richardson D.C. (1999). Visualizing and quantifying molecular goodness-of-fit : small-probe contact fots with explicit hydrogen atoms. *J Mol Biol* 285, 1711-1733.
- [155] Go N. and Scheraga H.A. (1970). Ring Closure and Local Conformational Deformations of Chain Molecules. *Macromolecules* 3, 178-187.
- [156] Wedemeyer W.J. and Scheraga H.A. (1999). Exact analytical loop closure in proteins using polynomial equations. J Comput Chem 20, 819-844.
- [157] Canutescu A.A. and Jr Dunbrack R.L. (2003). Cyclic coordinate descent : A robotics algorithm for protein loop closure. *Protein Sci* 12, 963-972.
- [158] Coutsias E.A., Seok C., Jacobson M.P. and Dill K.A. (2004). A kinematic view of loop closure. J Comput Chem 25, 510-528.
- [159] Wojcik J., Mornon J.P. and Chomilier J. (1999). New efficient statistical sequencedependent structure prediction of short to medium-sized protein loops based on an exhaustive loop classification. J Mol Biol 289, 1469-1490.
- [160] DePristo M.A., de Bakker P.I.W., Lovell S.C. and Blundell T.L. (2003). Ab initio construction of polypeptide fragments : efficient generation of accurate, representative ensembles. *Proteins* 51, 41-55.
- [161] Collura V., Higo J. and Garnier J. (1993). Modeling of protein loops by simulated annealing. Protein Science 2, 1502-1510.
- [162] Zheng Q., Rosenfeld R., Vajda S. and DeLisi C. (1993). Determining protein loop conformation using scaling-relaxation techniques. *Protein Science* 2, 1242-1248.

- [163] Rosenbach D. and Rosenfeld R. (1995). Simultaneous modeling of multiple loops in proteins. Protein Science 4, 496-505.
- [164] Fidelis K., Stern P.S., Bacon D. and Moult J. (1994). Comparison of systematic search and database methods for constructing segments of protein structure. *Protein Engineering* 7, 953-960.
- [165] Jones T.A. and Thirup S. (1986). Using known substructures in protein model building and crystallography. *EMBO J* 5, 818-822.
- [166] van Vlijmen H.W.T. and Karplus M. (1997). PDB-based protein loop prediction : parameters for selection and methods for optimization. J Mol Biol 267, 975-1001.
- [167] Martin A.C.R. and Thornton J.M. (1996). Structural families loops of homologous proteins : automatic classification, modelling and application to antibodies. J Mol Biol 263, 800-815.
- [168] Chothia C. and Lesk A.M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol 196, 901-917.
- [169] Martin A.C., Cheetham J.C. and Rees A.R. (1989). Modeling antibody hypervariable loops : a combined algorithm. Proc Natl Acad Sciences USA 86, 9268-9272.
- [170] Deane C.M. and Blundell T.L. (2001). CODA : a combined algorithm for predicting the structurally variable regions of protein models. *Protein Science* 10, 599-612.
- [171] Moult J. and James M.N. (1986). An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. *Proteins* **1**, 146-163.
- [172] Jacobson M.P., Pincus D.L., Rapp C.S., Day T.J., Honig B., Shaw D.E. and Friesner R.A. (2004). A hierarchical approach to all-atom protein loop prediction. *Proteins* 55, 351-367.
- [173] Deane C.M. and Blundell T.L. (2000). A novel exhaustive search algorithm for predicting the conformation of polypeptide segments in proteins. *Proteins* **40**, 135-144.
- [174] Tanner J.J., Nell L.J. and McCammon J.A. (1992). Anti-insulin antibody structure and conformation. II. Molecular dynamics with explicit solvent. *Biopolymers* **32**, 23-32.
- [175] Bruccoleri R.E. and Karplus M. (1990). Conformational sampling using hightemperature molecular dynamics. *Biopolymers* 29, 1847-1862.
- [176] Carlacci L. and Englander S.W. (1993). The loop problem in proteins : a Monte Carlo simulated annealing approach. *Biopolymers* 33, 1271-1286.
- [177] Kidera A. (1995). Enhanced conformational sampling in Monte Carlo simulations of proteins : application to a constrained peptide. Proc Natl Acad Sciences USA 92, 9886-9889.
- [178] Gallicchio E., Zhang L.Y. and Levy R.M. (2002). The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. J Comput Chem 23, 517-529.
- [179] de Bakker P.I.W., DePristo M.A., Burke D.F. and Blundell T.L. (2003). Ab initio construction of polypeptide fragments : accuracy of loop decoy discrimination by an all-atom statistical potential and the AMBER force field with the generalized born solvation model. *Proteins* 51, 21-40.

- [180] Pellequer J.L. and Chen S.W. (1997). Does conformational free energy distinguish loop conformations in proteins?. *Biophysical Journal* **73**, 2359-2375.
- [181] Thomas P.D. and Dill K.A. (1996). Statistical potentials extracted from protein structure : how accurate are they?. J Mol Biol 257, 447-469.
- [182] Robert C.H. and Janin J. (1998). A soft, mean-field potential derived from crystal contacts for predicting protein-protein interactions. J Mol Biol 283, 1037-1047.
- [183] Melo F. and Feytmans E. (1997). Novel Knowledge-based mean force potential at atomic level. J Mol Biol 267, 207-222.
- [184] Samudrala R. and Moult J. (1998). An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J Mol Biol 275, 895-916.
- [185] Fiser A., Do R.K. and Sali A. (2000). Modeling of loops in protein structures. Protein Science 9, 1753-1773.
- [186] Xiang Z., Soto C.S. and Honig B. (2002). Evaluating conformational free energies : the colony energy and its application to the problem of loop prediction. *Proc Natl Acad Sciences USA* 99, 7432-7437.
- [187] Reczko M., Martin A.C., Bohr H. and Suhai S. (1995). Prediction of hypervariable CDR-H3 loop structures in antibodies. *Protein Eng* 8, 389-395.
- [188] Chong L.T., Duan Y., Wang L., Massova I. and Kollman P.A. (1999). Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc Natl Acad Sciences USA* 96, 14330-14335.

Annexe.

Liste des domaines : page 139.

Liste des partnenaires : page 149.

Liste des récepteurs : page 152.

Liste des IG se liant à des antigènes différents : page 154.

Liste des structures IG/antigène : page 157.

Liste des peptide/MHC : page 159.

Liste des structures pour la dérivation de la bibliothèque de SCHISMo : page 167.

Liste des structures pour la dérivation de la bibliothèque de LLIPA : page 168.

Liste des V-DOMAINS, C-DOMAINS et G-DOMAINS utilisés pour toutes les analyses de domaines. Pour chaque domaine, le type de récepteur (IG, TR ou MHC), sa description IMGT (Domaine) et sa résolution(R) est indiquée.
AIN		Effec	tif total : 473	3		
	Domaine	\mathbf{R}	Chaîne		Domaine	R
IG	V-KAPPA	$1,\!90$	1a2yA	IG	V-KAPPA	1,50
IG	V-KAPPA	2,10	$1a6t_A$	IG	V-KAPPA	2,70
IG	V-KAPPA	2,50	$1ad9_A$	IG	V-KAPPA	$2,\!80$
IG	V-KAPPA	$2,\!80$	$1aif_A$	IG	V-KAPPA	$2,\!90$
IG	V-KAPPA	$2,\!15$	1b0wA	IG	V-KAPPA	$1,\!80$
IG	V-KAPPA	2,74	$1 baf_L$	IG	V-KAPPA	$2,\!90$
IG	V-KAPPA	3,10	1bfo_A	IG	V-KAPPA	$2,\!60$
IG	V-KAPPA	2,10	1bj1_J	IG	V-KAPPA	2,40
IG	V-KAPPA	2,70	1 bww A	IG	V-KAPPA	1,70
IG	V-KAPPA	2,50	1c12A	IG	V-KAPPA	$2,\!60$
IG	V-KAPPA	$1,\!90$	$1c5c_L$	IG	V-KAPPA	$1,\!61$
IG	V-KAPPA	2,40	$1ce1_L$	IG	V-KAPPA	$1,\!90$
IG	V-KAPPA	2,70	1 cic A	IG	V-KAPPA	2,50
IG	V-KAPPA	2,10	1 clz L	IG	V-KAPPA	$2,\!80$
IG	V-KAPPA	$2,\!00$	1 ct 8 A	IG	V-KAPPA	2,20
IG	V-KAPPA	$2,\!00$	$1dbb_L$	IG	V-KAPPA	2,70
IG	V-KAPPA	2,70	$1dfb_L$	IG	V-KAPPA	2,70
IG	V-KAPPA	2,28	$1 dql_L$	IG	V-KAPPA	$2,\!60$
IG	V-KAPPA	2,20	1 dvf C	IG	V-KAPPA	$1,\!90$
IG	V-KAPPA	$2,\!00$	$1 e4x_L$	IG	V-KAPPA	$1,\!90$
IG	V-KAPPA	2,50	$1 eeq_A$	IG	V-KAPPA	1,50
IG	V-KAPPA	$2,\!80$	1ejo_L	IG	V-KAPPA	$2,\!30$
IG	V-KAPPA	$1,\!90$	$1 \mathrm{emt} _\mathrm{L}$	IG	V-KAPPA	2,25
IG	V-KAPPA	$2,\!80$	1 ezv Y	IG	V-KAPPA	$2,\!30$
IG	V-KAPPA	$1,\!87$	$1f58_L$	IG	V-KAPPA	$2,\!00$
IG	V-KAPPA	2,20	$1 fe 8 _L$	IG	V-KAPPA	$2,\!03$
IG	V-KAPPA	$1,\!90$	$1 \mathrm{fj1}_A$	IG	V-KAPPA	$2,\!68$
IG	V-KAPPA	$2,\!00$	$1 \mathrm{fsk} B$	IG	V-KAPPA	$2,\!90$
IG	V-KAPPA	2,20	$1g9m_L$	IG	V-KAPPA	2,20
IG	V-KAPPA	$1,\!95$	$1h0d_A$	IG	V-KAPPA	$2,\!00$
IG	V-KAPPA	2,55	$1hil_A$	IG	V-KAPPA	$2,\!00$
IG	V-KAPPA	$1,\!80$	$1i7z_A$	IG	V-KAPPA	$2,\!30$
IG	V-KAPPA	$1,\!80$	$1i9r_L$	IG	V-KAPPA	3,10
	AIN IG IG IG IG IG IG IG IG IG IG IG IG IG	AIN Domaine IG V-KAPPA IG V-KAPPA <	AINEffectDomaineRIGV-KAPPA $1,90$ IGV-KAPPA $2,10$ IGV-KAPPA $2,50$ IGV-KAPPA $2,50$ IGV-KAPPA $2,15$ IGV-KAPPA $2,74$ IGV-KAPPA $2,70$ IGV-KAPPA $2,70$ IGV-KAPPA $2,70$ IGV-KAPPA $2,70$ IGV-KAPPA $2,70$ IGV-KAPPA $2,70$ IGV-KAPPA $2,00$ IGV-KAPPA $2,00$ IGV-KAPPA $2,00$ IGV-KAPPA $2,20$ IGV-KAPPA $1,90$ IGV-KAPPA $1,90$ IGV-KAPPA $1,95$ IGV-KAPPA $1,80$ IGV-KAPPA $1,80$ IGV-KAPPA $1,80$ IGV-KAPPA $1,80$ IGV-KAPPA </td <td>AINEffectif total : 473 DomaineRChaîneIGV-KAPPA1,90$1a2y_A$IGV-KAPPA2,10$1a6t_A$IGV-KAPPA2,50$1ad9_A$IGV-KAPPA2,80$1aif_A$IGV-KAPPA2,15$1b0w_A$IGV-KAPPA2,74$1baf_L$IGV-KAPPA2,71$1bj1_J$IGV-KAPPA2,70$1bw_A$IGV-KAPPA2,70$1bw_A$IGV-KAPPA2,70$1ct_L$IGV-KAPPA2,70$1ct_L$IGV-KAPPA2,70$1ct_L$IGV-KAPPA2,70$1ct_L$IGV-KAPPA2,00$1ct_A$IGV-KAPPA2,00$1ct_L$IGV-KAPPA2,00$1ct_A$IGV-KAPPA2,00$1ct_L$IGV-KAPPA2,28$1dq_L$IGV-KAPPA2,20$1dv_C$IGV-KAPPA2,50$1eeq_A$IGV-KAPPA2,80$1ej_L$IGV-KAPPA2,80$1ej_L$IGV-KAPPA2,80$1ej_L$IGV-KAPPA2,20$1de__L$IGV-KAPPA2,20$1es__L$IGV-KAPPA2,20$1es__L$IGV-KAPPA2,20$1es__L$IGV-KAPPA2,20$1es__L$IGV-KAPPA2,20$1es__L$IGV-KAPPA2,20$1e$</td> <td>AINEffectif total : 473 DomaineRChaîneIGV-KAPPA1,90$1a2y_A$IGIGV-KAPPA2,10$1a6t_A$IGIGV-KAPPA2,50$1ad9_A$IGIGV-KAPPA2,80$1aif_A$IGIGV-KAPPA2,15$1b0w_A$IGIGV-KAPPA2,74$1baf_L$IGIGV-KAPPA2,70$1by_A$IGIGV-KAPPA2,70$1by_A$IGIGV-KAPPA2,70$1bw_A$IGIGV-KAPPA2,70$1ct_A$IGIGV-KAPPA2,70$1ct_A$IGIGV-KAPPA2,70$1ct_A$IGIGV-KAPPA2,70$1ct_A$IGIGV-KAPPA2,70$1ct_A$IGIGV-KAPPA2,00$1ct_A$IGIGV-KAPPA2,00$1ct_A$IGIGV-KAPPA2,00$1ct_A$IGIGV-KAPPA2,00$1ct_A$IGIGV-KAPPA2,00$1dt_A$IGIGV-KAPPA2,20$1dv_C$IGIGV-KAPPA2,20$1dv_C$IGIGV-KAPPA2,80ezv_YIGIGV-KAPPA2,80ezv_YIGIGV-KAPPA2,90$1fs_B$IGIGV-KAPPA2,90$1fs_B$IGIGV-KAPPA2,90</td> <td>AINEffectif total : 473DomaineRChaîneDomaineIGV-KAPPA1,90$1a2y_A$IGV-KAPPAIGV-KAPPA2,10$1a6t_A$IGV-KAPPAIGV-KAPPA2,50$1ad9_A$IGV-KAPPAIGV-KAPPA2,80$1aif_A$IGV-KAPPAIGV-KAPPA2,15$1b0w_A$IGV-KAPPAIGV-KAPPA2,74$1baf_L$IGV-KAPPAIGV-KAPPA2,10$1bj1_J$IGV-KAPPAIGV-KAPPA2,00$1bbv_A$IGV-KAPPAIGV-KAPPA2,00$1cbc_L$IGV-KAPPAIGV-KAPPA2,00$1cbc_L$IGV-KAPPAIGV-KAPPA2,00$1cbc_L$IGV-KAPPAIGV-KAPPA2,00$1cbL_A$IGV-KAPPAIGV-KAPPA2,00$1cbL_A$IGV-KAPPAIGV-KAPPA2,00$1cbL_A$IGV-KAPPAIGV-KAPPA2,00$1cbL_A$IGV-KAPPAIGV-KAPPA2,00$1cbL_A$IGV-KAPPAIGV-KAPPA2,00$1cbL_A$IGV-KAPPAIGV-KAPPA2,00$1cbL_A$IGV-KAPPAIGV-KAPPA2,00$1cbL_A$IGV-KAPPAIGV-KAPPA2,00$1cbL_A$IGV-KAPPAIGV-KAPPA2,00$1cbL_A$</td>	AINEffectif total : 473 DomaineRChaîneIGV-KAPPA1,90 $1a2y_A$ IGV-KAPPA2,10 $1a6t_A$ IGV-KAPPA2,50 $1ad9_A$ IGV-KAPPA2,80 $1aif_A$ IGV-KAPPA2,15 $1b0w_A$ IGV-KAPPA2,74 $1baf_L$ IGV-KAPPA2,71 $1bj1_J$ IGV-KAPPA2,70 $1bw_A$ IGV-KAPPA2,70 $1bw_A$ IGV-KAPPA2,70 $1ct_L$ IGV-KAPPA2,70 $1ct_L$ IGV-KAPPA2,70 $1ct_L$ IGV-KAPPA2,70 $1ct_L$ IGV-KAPPA2,00 $1ct_A$ IGV-KAPPA2,00 $1ct_L$ IGV-KAPPA2,00 $1ct_A$ IGV-KAPPA2,00 $1ct_L$ IGV-KAPPA2,28 $1dq_L$ IGV-KAPPA2,20 $1dv_C$ IGV-KAPPA2,50 $1eeq_A$ IGV-KAPPA2,80 $1ej_L$ IGV-KAPPA2,80 $1ej_L$ IGV-KAPPA2,80 $1ej_L$ IGV-KAPPA2,20 $1de__L$ IGV-KAPPA2,20 $1es__L$ IGV-KAPPA2,20 $1es__L$ IGV-KAPPA2,20 $1es__L$ IGV-KAPPA2,20 $1es__L$ IGV-KAPPA2,20 $1es__L$ IGV-KAPPA2,20 $1e$	AINEffectif total : 473 DomaineRChaîneIGV-KAPPA1,90 $1a2y_A$ IGIGV-KAPPA2,10 $1a6t_A$ IGIGV-KAPPA2,50 $1ad9_A$ IGIGV-KAPPA2,80 $1aif_A$ IGIGV-KAPPA2,15 $1b0w_A$ IGIGV-KAPPA2,74 $1baf_L$ IGIGV-KAPPA2,70 $1by_A$ IGIGV-KAPPA2,70 $1by_A$ IGIGV-KAPPA2,70 $1bw_A$ IGIGV-KAPPA2,70 $1ct_A$ IGIGV-KAPPA2,70 $1ct_A$ IGIGV-KAPPA2,70 $1ct_A$ IGIGV-KAPPA2,70 $1ct_A$ IGIGV-KAPPA2,70 $1ct_A$ IGIGV-KAPPA2,00 $1ct_A$ IGIGV-KAPPA2,00 $1ct_A$ IGIGV-KAPPA2,00 $1ct_A$ IGIGV-KAPPA2,00 $1ct_A$ IGIGV-KAPPA2,00 $1dt_A$ IGIGV-KAPPA2,20 $1dv_C$ IGIGV-KAPPA2,20 $1dv_C$ IGIGV-KAPPA2,80 ezv_Y IGIGV-KAPPA2,80 ezv_Y IGIGV-KAPPA2,90 $1fs_B$ IGIGV-KAPPA2,90 $1fs_B$ IGIGV-KAPPA2,90	AINEffectif total : 473DomaineRChaîneDomaineIGV-KAPPA1,90 $1a2y_A$ IGV-KAPPAIGV-KAPPA2,10 $1a6t_A$ IGV-KAPPAIGV-KAPPA2,50 $1ad9_A$ IGV-KAPPAIGV-KAPPA2,80 $1aif_A$ IGV-KAPPAIGV-KAPPA2,15 $1b0w_A$ IGV-KAPPAIGV-KAPPA2,74 $1baf_L$ IGV-KAPPAIGV-KAPPA2,10 $1bj1_J$ IGV-KAPPAIGV-KAPPA2,00 $1bbv_A$ IGV-KAPPAIGV-KAPPA2,00 $1cbc_L$ IGV-KAPPAIGV-KAPPA2,00 $1cbc_L$ IGV-KAPPAIGV-KAPPA2,00 $1cbc_L$ IGV-KAPPAIGV-KAPPA2,00 $1cbL_A$

Suite de la table à	la page précéde [.]	nte
---------------------	------------------------------	-----

Chaîne		Domaine	R	Chaîne		Domaine	R
1iai L	IG	V-KAPPA	2,90	1ibg L	IG	V-KAPPA	2,70
ligc L	IG	V-KAPPA	$2,\!60$	ligm L	IG	V-KAPPA	2,30
likf_L	IG	V-KAPPA	2,50	liqd_A	IG	V-KAPPA	2,00
liqw_L	IG	V-KAPPA	2,50	1 it 9_L	IG	V-KAPPA	$2,\!80$
1j05_A	IG	V-KAPPA	1,50	1jfq_L	IG	V-KAPPA	$1,\!90$
1jgu_L	IG	V-KAPPA	$1,\!80$	1 jhl_L	IG	V-KAPPA	2,40
1jnl_L	IG	V-KAPPA	3,00	$1 \mathrm{jps} _\mathrm{L}$	IG	V-KAPPA	$1,\!85$
1jrh_L	IG	V-KAPPA	$2,\!80$	$1 \mathrm{k6q} \mathrm{L}$	IG	V-KAPPA	2,40
$1 \mathrm{kcr} L$	IG	V-KAPPA	$2,\!90$	$1 \mathrm{kcu} \mathrm{L}$	IG	V-KAPPA	2,20
1kcv_L	IG	V-KAPPA	$1,\!80$	$1 \mathrm{kel} \mathrm{L}$	IG	V-KAPPA	$1,\!90$
$1 \mathrm{ken} L$	IG	V-KAPPA	3,50	1l7i_L	IG	V-KAPPA	$1,\!80$
117t L	IG	V-KAPPA	2,10	1 lk 3 L	IG	V-KAPPA	$1,\!91$
11mk_A	IG	V-KAPPA	$2,\!60$	1 lo0 L	IG	V-KAPPA	$2,\!00$
1 lo4 L	IG	V-KAPPA	2,40	$1 \mathrm{mex} L$	IG	V-KAPPA	$1,\!25$
1mhp_L	IG	V-KAPPA	$2,\!80$	1 mim L	IG	V-KAPPA	$2,\!60$
1mju_L	IG	V-KAPPA	1,22	$1 mqk_L$	IG	V-KAPPA	1,28
1 n0 x L	IG	V-KAPPA	$1,\!80$	$1 nak_L$	IG	V-KAPPA	2,57
1nca_L	IG	V-KAPPA	2,50	1 ncw L	IG	V-KAPPA	$1,\!30$
1 nlb L	IG	V-KAPPA	$1,\!60$	$1 \mathrm{nmb} \mathrm{L}$	IG	V-KAPPA	2,50
$1 \mathrm{nsn} \mathrm{L}$	IG	V-KAPPA	$2,\!90$	$10b1_A$	IG	V-KAPPA	$2,\!90$
$10p3_K$	IG	V-KAPPA	1,75	$1 \text{ors} _ A$	IG	V-KAPPA	$1,\!90$
1osp_L	IG	V-KAPPA	$1,\!95$	$1 ots_D$	IG	V-KAPPA	2,51
1p7k_A	IG	V-KAPPA	1,75	$1 \mathrm{psk} \mathrm{_L}$	IG	V-KAPPA	$2,\!80$
$1pz5_A$	IG	V-KAPPA	$1,\!80$	$1q9r_A$	IG	V-KAPPA	$1,\!45$
$1qfw_L$	IG	V-KAPPA	3,50	1 q f w M	IG	V-KAPPA	3,50
1r0a_L	IG	V-KAPPA	$2,\!80$	1r3j_A	IG	V-KAPPA	$1,\!90$
1rhh_A	IG	V-KAPPA	$1,\!90$	$1 \mathrm{rih} \mathrm{L}$	IG	V-KAPPA	2,50
1rjl_A	IG	V-KAPPA	$2,\!60$	$1 \mathrm{rur} \mathrm{L}$	IG	V-KAPPA	1,50
1 rz7 L	IG	V-KAPPA	2,00	$1 rzg_B$	IG	V-KAPPA	$2,\!00$
$1s3k_L$	IG	V-KAPPA	$1,\!90$	$1 { m sbs}_L$	IG	V-KAPPA	$2,\!00$
1 seq L	IG	V-KAPPA	1,78	$1 \mathrm{sy6} \mathrm{_L}$	IG	V-KAPPA	2,10
$1t3f_A$	IG	V-KAPPA	$2,\!00$	1t4kA	IG	V-KAPPA	2,50
1tjg_L	IG	V-KAPPA	$2,\!00$	1 tqb C	IG	V-KAPPA	2,55
$1tzg_L$	IG	V-KAPPA	2,20	1 tzi_A	IG	V-KAPPA	$2,\!80$
$1uac_L$	IG	V-KAPPA	1,70	$1\mathrm{ub6}$ B	IG	V-KAPPA	2,12
$1 \mathrm{um} 5 \mathrm{L}$	IG	V-KAPPA	$1,\!60$	$1 \mathrm{uwe} \mathrm{L}$	IG	V-KAPPA	$2,\!67$
$1 \mathrm{uwg} \mathrm{L}$	IG	V-KAPPA	2,79	1uwx_K	IG	V-KAPPA	2,20
1uz8_A	IG	V-KAPPA	$1,\!80$	$1 v7m_L$	IG	V-KAPPA	2,51

Chaîno		Domaina	Р	Chaîno		Domaina	Б
1vgo I	IC		2 00	1 woi I	IC		1.80
1wt5 C	IC		2,00 2.10	$1 \text{ wej} \underline{L}$ $1 \text{ wet} \underline{\Lambda}$	IC	V KAPPA	1,00
1wtu U	IC	ν-καιτά ν κάρρα	2,10 2.71	1wti_A	IG	ν-ΚΑΓΓΑ V καρρά	1,90 1,70
1 vov I		V KADDA	2,71 1 70	1y111_L 2a77_L	IC	V KADDA	1.80
lyqv_L	IG		1,70	2arr_L Ohm I	IG	V-KAFFA V KADDA	1,00
Zaug_A	IG	V-KAPPA V KADDA	2,50	20ff_L offs: I	IG	V-KAPPA V Kadda	1,90
2CKU_L	IG	V-KAPPA V LADDA	2,20	210J_L	IG	V-KAPPA V V ADDA	1,90
21mm	IG IC	ν-κάρρα νιζαρρα	2,00	21mn	IG IC	V-KAPPA V V ADDA	1,97
32CZ_A	IG	V-KAPPA V LADDA	3,00	3ICt A	IG IG	V-KAPPA	2,40
43C9_A	IG	V-KAPPA	2,20	laqk_L	IG	V-LAMBDA	1,84
Ibjm_A	IG	V-LAMBDA	2,20	IcdU_A	IG	V-LAMBDA	1,90
ljn6_A	IG	V-LAMBDA	2,70	ljvk_A	IG	V-LAMBDA	1,94
llil_A	IG	V-LAMBDA	$2,\!65$	1mcw_W	IG	V-LAMBDA	3,50
1mfa	IG	V-LAMBDA	1,70	1nfd_E	IG	V-LAMBDA	$2,\!80$
1nj9_A	IG	V-LAMBDA	$2,\!35$	1 n l 0 L	IG	V-LAMBDA	$2,\!20$
10aq_L	IG	V-LAMBDA	$1,\!50$	$1q1j_L$	IG	V-LAMBDA	$2,\!50$
1rzf_L	IG	V-LAMBDA	1,70	1 w72 L	IG	V-LAMBDA	$2,\!15$
2 cd0 A	IG	V-LAMBDA	$1,\!80$	2 fb4 L	IG	V-LAMBDA	$1,\!90$
$2mcg_1$	IG	V-LAMBDA	$2,\!00$	$2 \mathrm{rhe}_{-}$	IG	V-LAMBDA	$1,\!60$
8fab_A	IG	V-LAMBDA	$1,\!80$	12e8 H	IG	VH	$1,\!90$
$1a2y_B$	IG	VH	$1,\!50$	$1a4j_B$	IG	VH	$2,\!10$
1a6t_B	IG	VH	2,70	$1ad9_B$	IG	VH	$2,\!80$
1adq_H	IG	VH	$3,\!15$	1ae6 H	IG	VH	3,00
1ai1_H	IG	VH	$2,\!80$	1aif_B	IG	VH	$2,\!90$
1aqk_H	IG	VH	$1,\!84$	$1 \mathrm{axt} \mathrm{H}$	IG	VH	$2,\!15$
$1ay1_H$	IG	VH	$2,\!20$	1b2wH	IG	VH	$2,\!90$
1b4j H	IG	VH	$2,\!90$	1baf H	IG	VH	$2,\!90$
1bbd H	IG	VH	$2,\!80$	1bfo B	IG	VH	$2,\!60$
1bfv H	IG	VH	$2,\!10$	1bj1 H	IG	VH	$2,\!40$
1bln B	IG	VH	$2,\!80$	1bvk B	IG	VH	2,70
1bz7 B	IG	VH	2,50	1bzq K	IG	VH	$2,\!80$
1c12 B	IG	VH	$2,\!60$	1c1e H	IG	VH	1,90
1c5c H	IG	VH	$1,\!61$	1c5d B	IG	VH	2,40
1ce1 H	IG	VH	1.90	1cf8 H	IG	VH	2.70
1cic B	IG	VH	2.50	1clo [–] H	IG	VH	2.10
1clv H	IG	VH	2.50	1cr9 H	IG	VH	2.00
1 ct8 B	IG	VH	2.20	 1d5i H	IG	VH	2,00
1dbb H	IG	VH	2.70	1dee B	IG	VH	2.70
1dfb H	ĪĞ	VH	2.70	1dl7 H	ĪĞ	VH	2.35
-			_,. •	<u>-</u>			_,00

Suite de la table à la page précédente

Chaîne		Domaine	R	Chaîne		Domaine	R
1dlf H	IG	VH	$1,\!45$	1dn0 B	IG	VH	2,28
1dqd H	IG	VH	2,10	1dql H	IG	VH	$2,\!60$
1dqq B	IG	VH	1,80	1dsf H	IG	VH	2,20
1dvf D	IG	VH	1,90	1e4x H	IG	VH	1,90
1e60 H	IG	VH	1,80	1eap B	IG	VH	2,50
1egj H	IG	VH	$2,\!80$	1ehl H	IG	VH	2,40
1ejo H	IG	VH	$2,\!30$	1emt H	IG	VH	2,25
1eo8_H	IG	VH	$2,\!80$	1 etz B	IG	VH	$2,\!60$
1ezv_X	IG	VH	$2,\!30$	1f11B	IG	VH	3,00
1f2x_K	IG	VH	$2,\!10$	$1 \mathrm{f3d} \mathrm{H}$	IG	VH	$1,\!87$
1 f4w H	IG	VH	$2,\!30$	1f58 H	IG	VH	$2,\!00$
1f8t_H	IG	VH	$2,\!20$	1fai_H	IG	VH	2,70
1fbi_H	IG	VH	$3,\!00$	$1 \text{fe8} _ \text{H}$	IG	VH	$2,\!03$
1fgv_H	IG	VH	$1,\!90$	$1 \mathrm{fh5} \mathrm{H}$	IG	VH	$2,\!90$
1fig_H	IG	VH	$3,\!00$	1fj1_B	IG	VH	$2,\!68$
$1 \mathrm{fns} \mathrm{H}$	IG	VH	$2,\!00$	$1 \text{for} \mathbf{H}$	IG	VH	2,75
$1 \mathrm{fpt}_H$	IG	VH	$3,\!00$	$1 \mathrm{frg} \mathrm{H}$	IG	VH	$2,\!80$
$1 \mathrm{fsk} \mathrm{C}$	IG	VH	$2,\!90$	1fvc_B	IG	VH	2,20
$1g9m_H$	IG	VH	$2,\!20$	1gaf_H	IG	VH	$1,\!95$
1ggb_H	IG	VH	$2,\!80$	$1 \mathrm{ghf}_{\mathrm{H}}$	IG	VH	2,70
1gig_H	IG	VH	$2,\!30$	$1 \text{gpo} _ \text{H}$	IG	VH	$1,\!95$
$1h0d_B$	IG	VH	$2,\!00$	$1h3p_H$	IG	VH	$2,\!60$
1hcv	IG	VH	$1,\!85$	1 hi6 B	IG	VH	2,55
1hil_B	IG	VH	$2,\!00$	$1 hyx_H$	IG	VH	$1,\!80$
1i3v_A	IG	VH	$2,\!03$	$1i7z_B$	IG	VH	$2,\!30$
1i8k_B	IG	VH	$1,\!80$	$1i9r_H$	IG	VH	3,10
1iai_H	IG	VH	$2,\!90$	$1ibg_H$	IG	VH	2,70
ligc_H	IG	VH	$2,\!60$	$1igf_H$	IG	VH	$2,\!80$
1igj_B	IG	VH	$2,\!50$	1 igm_H	IG	VH	$2,\!30$
ligy_B	IG	VH	3,20	1ikf_H	IG	VH	2,50
1il1_A	IG	VH	$2,\!20$	1 ind_H	IG	VH	2,20
1iqd_B	IG	VH	$2,\!00$	1 iqw_H	IG	VH	2,50
1it9 H	IG	VH	$2,\!80$	$1j05_B$	IG	VH	1,50
1jgu_H	IG	VH	$1,\!80$	1 jhl_H	IG	VH	$2,\!40$
$1jn6_B$	IG	VH	2,70	1 jnl_H	IG	VH	$3,\!00$
$1 jps_H$	IG	VH	$1,\!85$	1jrh_H	IG	VH	$2,\!80$
1jtp_A	IG	VH	$1,\!90$	$1jv5_B$	IG	VH	2,20
1k6q_H	IG	VH	$2,\!40$	$1 \mathrm{kb5}\mathrm{H}$	IG	VH	2,50

Suite de la table à la page précédente

Chaîne		Domaine	R	Chaîne		Domaine	R
1kcu_H	IG	VH	2,20	$1 \mathrm{kcv}$ H	IG	VH	$1,\!80$
1kel_H	IG	VH	$1,\!90$	$1 \mathrm{ken} \mathrm{H}$	IG	VH	3,50
1kfa_H	IG	VH	$2,\!80$	$1 \mathrm{ktr} \mathrm{H}$	IG	VH	2,70
1kxq_E	IG	VH	$1,\!60$	$1 \mathrm{kxt}$ B	IG	VH	2,00
1kxv_D	IG	VH	$1,\!60$	117i_H	IG	VH	$1,\!80$
1l7t_H	IG	VH	2,10	1lk3_H	IG	VH	$1,\!91$
1lmk_A	IG	VH	$2,\!60$	1 lo0 H	IG	VH	2,00
1lo4H	IG	VH	2,40	1 mcoH	IG	VH	3,20
1mcp_H	IG	VH	2,70	$1 \mathrm{mex}$ H	IG	VH	1,25
$1 \mathrm{mfb} H$	IG	VH	2,10	$1 \mathrm{mhp}\mathrm{H}$	IG	VH	$2,\!80$
1 mim H	IG	VH	$2,\!60$	1 mj 8 H	IG	VH	1,75
1mju_H	IG	VH	1,22	$1 \mathrm{mnu}$ H	IG	VH	2,50
1mqk_H	IG	VH	1,28	$1 \mathrm{mre}$ H	IG	VH	$2,\!30$
1mvf_A	IG	VH	$1,\!65$	1mvu_B	IG	VH	1,78
1 n0 x H	IG	VH	$1,\!80$	$1n4x_H$	IG	VH	1,70
1nak_H	IG	VH	2,57	$1 \mathrm{nbv} \mathrm{H}$	IG	VH	$2,\!00$
$1nc2_D$	IG	VH	2,10	1 nca H	IG	VH	2,50
1ncw_H	IG	VH	$1,\!30$	1 ndg B	IG	VH	$1,\!90$
$1 n f d_F$	IG	VH	$2,\!80$	$1 ngz_B$	IG	VH	$1,\!60$
1nj9_B	IG	VH	$2,\!35$	$1 \mathrm{nl0} \mathrm{H}$	IG	VH	2,20
1nlb_H	IG	VH	$1,\!60$	1 n ld H	IG	VH	$2,\!90$
1 nmb H	IG	VH	2,50	$1 \mathrm{nsn}$ H	IG	VH	$2,\!90$
1oaq_H	IG	VH	1,50	$10b1_B$	IG	VH	$2,\!90$
10hq_A	IG	VH	$2,\!00$	100A	IG	VH	$1,\!80$
$1 \mathrm{op}3 \mathrm{H}$	IG	VH	1,75	$10p9_A$	IG	VH	$1,\!86$
$10pg_H$	IG	VH	$2,\!00$	1ors_B	IG	VH	$1,\!90$
1osp_H	IG	VH	$1,\!95$	$1 ots _C$	IG	VH	2,51
$1 p2 c_B$	IG	VH	$2,\!00$	1 p 4 b H	IG	VH	2,35
$1 p7 k_B$	IG	VH	1,75	$1 pg7_X$	IG	VH	2,50
1pkq_B	IG	VH	3,00	1plg_H	IG	VH	$2,\!80$
1psk_H	IG	VH	$2,\!80$	$1 \text{pz} 5_\text{B}$	IG	VH	$1,\!80$
$1q0x_H$	IG	VH	$1,\!60$	1q1j_H	IG	VH	2,50
$1q72_H$	IG	VH	1,70	$1q9r_B$	IG	VH	$1,\!45$
$1q9w_B$	IG	VH	1,75	1 q d0 A	IG	VH	2,50
1qfu_H	IG	VH	$2,\!80$	$1 q fw_I$	IG	VH	3,50
1 q kz H	IG	VH	$1,\!95$	1 qok A	IG	VH	$2,\!40$
$1r0a_H$	IG	VH	$2,\!80$	$1r3j_B$	IG	VH	$1,\!90$
1 rhh B	IG	VH	$1,\!90$	1ri8_A	IG	VH	$1,\!85$

Suite de la table à la page précédente

Chaîne		Domaine	R	Chaîne		$\operatorname{Domaine}$	R
1rih_H	IG	VH	2,50	1rjc_A	IG	VH	$1,\!40$
1rjl_B	IG	VH	$2,\!60$	$1 \mathrm{rmf} H$	IG	VH	$2,\!80$
1rur_H	IG	VH	1,50	1 rz7 H	IG	VH	$2,\!00$
$1 s_{3k} H$	IG	VH	$1,\!90$	$1 { m sbs}$ H	IG	VH	$2,\!00$
1seq_H	IG	VH	1,78	$1 \mathrm{sm} 3 \mathrm{H}$	IG	VH	$1,\!95$
$1 \mathrm{sy6} \mathrm{H}$	IG	VH	2,10	$1 \mathrm{t} 2 \mathrm{j} \mathrm{A}$	IG	VH	1,50
$1t2q_H$	IG	VH	$1,\!83$	$1t3f_B$	IG	VH	$2,\!00$
$1t4k_B$	IG	VH	2,50	$1t66_D$	IG	VH	$2,\!30$
$1 \text{tet} _ H$	IG	VH	$2,\!30$	$1 \mathrm{tjg}\mathrm{H}$	IG	VH	$2,\!00$
1tqb_B	IG	VH	2,55	$1 \mathrm{txv} \mathrm{H}$	IG	VH	2,75
1 tzgH	IG	VH	2,20	1tzh_B	IG	VH	$2,\!60$
$1tzi_B$	IG	VH	$2,\!80$	$1u0q_A$	IG	VH	$1,\!60$
1u6a_H	IG	VH	$2,\!81$	1 uac H	IG	VH	1,70
$1ub5_A$	IG	VH	$2,\!00$	1uj3_B	IG	VH	2,10
$1 \mathrm{um} 5 \mathrm{H}$	IG	VH	$1,\!60$	1uwe_H	IG	VH	$2,\!67$
1uwg_H	IG	VH	2,79	$1 \mathrm{uwx}$ H	IG	VH	2,20
1uz8B	IG	VH	$1,\!80$	$1v7m_H$	IG	VH	2,51
$1 \text{ves} _ A$	IG	VH	2,18	$1 \mathrm{vge} \mathrm{H}$	IG	VH	$2,\!00$
1 w72 H	IG	VH	$2,\!15$	1wej_H	IG	VH	$1,\!80$
1 wt5 A	IG	VH	2,10	$1 x 9 q_A$	IG	VH	1,50
1xgy_H	IG	VH	2,71	1xiw_D	IG	VH	$1,\!90$
$1yc7_A$	IG	VH	$1,\!60$	1 yed B	IG	VH	$3,\!10$
1yee_H	IG	VH	2,20	1yej_H	IG	VH	$1,\!85$
1yjd_H	IG	VH	2,70	$1 ynl_H$	IG	VH	1,70
1yqv_H	IG	VH	1,70	$1yuh_B$	IG	VH	$3,\!00$
$1yy8_B$	IG	VH	$2,\!00$	1za3B	IG	VH	$3,\!35$
1ztx_H	IG	VH	2,50	25c8 H	IG	VH	$2,\!00$
$2a77$ _H	IG	VH	$1,\!80$	2 a dg B	IG	VH	2,50
2ai0_I	IG	VH	2,20	$2 \mathrm{brr} \mathbf{H}$	IG	VH	$1,\!95$
2 ck 0 H	IG	VH	2,20	2 fb4 H	IG	VH	$1,\!90$
2fbj_H	IG	VH	$1,\!95$	2 gfb B	IG	VH	$3,\!00$
2h1p_H	IG	VH	2,40	2hrp_H	IG	VH	2,20
2jel H	IG	VH	2,50	2pcp B	IG	VH	2,20
32c2B	IG	VH	$3,\!00$	43c9B	IG	VH	2,20
6fab H	IG	VH	$1,\!90$	8fab B	IG	VH	$1,\!80$
1ac6 A	TR	V-ALPHA	$2,\!30$	1ao7 D	TR	V-ALPHA	$2,\!60$
1b88 A	TR	V-ALPHA	2,50	1bd2 D	TR	V-ALPHA	2,50
1d9kA	TR	V-ALPHA	3,20	1h5bA	TR	V-ALPHA	$1,\!85$

Suite de la table à la page précédente

Chaîne		Domaine	\mathbf{R}	Chaîne		Domaine	\mathbf{R}
1j8h_D	TR	V-ALPHA	2,40	$1 \mathrm{jtr}_A$	TR	V-ALPHA	$2,\!40$
1 kb5 A	TR	V-ALPHA	$2,\!50$	$1 lp9_E$	TR	V-ALPHA	$2,\!00$
1nam_A	TR	V-ALPHA	2,70	1 n f d A	TR	V-ALPHA	$2,\!80$
loga_D	TR	V-ALPHA	$1,\!40$	$1u3h_A$	TR	V-ALPHA	$2,\!42$
$2ak4_D$	TR	V-ALPHA	$2,\!50$	$1ao7_E$	TR	V-BETA	$2,\!60$
$1 \mathrm{bd2}\mathrm{-E}$	TR	V-BETA	$2,\!50$	1bec_{-}	TR	V-BETA	1,70
$1j8h_E$	TR	V-BETA	2,40	$1 \mathrm{jtr}_B$	TR	V-BETA	$2,\!40$
$1 \mathrm{kb5}B$	TR	V-BETA	$2,\!50$	$1 \rm kgc_E$	TR	V-BETA	$1,\!50$
$1 \mathrm{ktk} \mathrm{E}$	TR	V-BETA	3,00	$1 lp9_F$	TR	V-BETA	$2,\!00$
1 nam B	TR	V-BETA	2,70	1 n f d B	TR	V-BETA	$2,\!80$
1oga_E	TR	V-BETA	$1,\!40$	$1u3h_B$	TR	V-BETA	$2,\!42$
$1 \mathrm{ymm}_{\mathrm{E}}$	TR	V-BETA	3,50	$2ak4$ _E	TR	V-BETA	$2,\!50$
1hxm_A	TR	V-DELTA	3,12	1 tvd A	TR	V-DELTA	$1,\!90$
1hxm_B	TR	V-GAMMA	3,12				
C-DOMA	IN		Effect	tif total : 77			
Chaîne		Domaine	R	Chaîne		Domaine	R
1bfo A	IG	C-KAPPA	2.60	1c5c L	IG	C-KAPPA	1.61
1c5d A	IG	C-KAPPA	2.40	1kcv L	IG	C-KAPPA	1.80
1miu L	IG	C-KAPPA	1.22	1 a 72 L	IG	C-KAPPA	1.70
1t4k A	IG	C-KAPPA	2,50	1 um 5 L	IG	C-KAPPA	$1,\!60$
2hmi C	IG	C-KAPPA	$2,\!80$	1agk L	IG	C-LAMBDA	1.84
1 nc 2 C	IG	C-LAMBDA	2.10	1nfd E	IG	C-LAMBDA	2,80
$1 \text{pg7}^{-} \text{W}$	IG	C-LAMBDA	2,50	1 q0 x L	IG	C-LAMBDA	1.60
1dee B	IG	CH1	2,70	1h0d B	IG	CH1	2,00
1h3p H	IG	CH1	$2,\!60$	1kcv H	IG	CH1	1,80
1ncw H	IG	CH1	$1,\!30$	1 n f d F	IG	CH1	2,80
1ngz B	IG	CH1	$1,\!60$	10ts C	IG	CH1	$2,\!51$
1 pz 5 B	IG	CH1	$1,\!80$	1q72 H	IG	CH1	1,70
1sy6 H	IG	CH1	$2,\!10$	1uj3 B	IG	CH1	2,10
1wej H	IG	CH1	$1,\!80$	1ynl H	IG	CH1	1,70
2fbj H	IG	CH1	1,95	2pcp B	IG	CH1	2,20
1e4k A	IG	CH2	3,20	1hzh H	IG	CH2	2,70
1i1c A	IG	CH2	2,70	1igt B	IG	CH2	2,80
1igv B	IG	CH2	3.20	100v A	IG	CH2	$2,\!60$
10w0 A	IG	CH2	$3,\!10$	1hzh H	IG	CH3	2,70
1i1c A	IG	CH3	2,70	1igt B	IG	CH3	2,80
1igy B	IG	CH3	$3,\!20$	100v A	IG	CH3	$2,\!60$
<u> </u>			,	_			<i>'</i>

Suite de la table à la page précédente

Chaîne		Domaine	R	Chaîne		Domaine	R
10w0 A	IG	CH3	3,10	1pfc	IG	CH3	$3,\!12$
1f6a_B	IG	CH4	3,50	1 e f x A	MHC	C-LIKE	$3,\!00$
1fne_A	MHC	C-LIKE	$1,\!90$	1 fne B	MHC	C-LIKE	$1,\!90$
$1 \mathrm{frt} \mathbf{A}$	MHC	C-LIKE	4,50	$1 fv1_A$	MHC	C-LIKE	$1,\!90$
1fv1_B	MHC	C-LIKE	$1,\!90$	1jk8_B	MHC	C-LIKE	$2,\!40$
1jpf_A	MHC	C-LIKE	$2,\!18$	$1k8i_A$	MHC	C-LIKE	3,10
1k8i_B	MHC	C-LIKE	3,10	1kjv_A	MHC	C-LIKE	$1,\!48$
1kjv_B	MHC	C-LIKE	$1,\!48$	$1 \mathrm{kpr}_A$	MHC	C-LIKE	$2,\!80$
$1ld9_A$	MHC	C-LIKE	2,40	$1lk2_B$	MHC	C-LIKE	$1,\!35$
1mhc_A	MHC	C-LIKE	2,10	1 nez A	MHC	C-LIKE	2,10
10ga_A	MHC	C-LIKE	$1,\!40$	$1 qo 3_A$	MHC	C-LIKE	$2,\!30$
1r3h_A	MHC	C-LIKE	$2,\!50$	$1t0n_A$	MHC	C-LIKE	$1,\!80$
$1u3h_C$	MHC	C-LIKE	2,42	$1u3h_D$	MHC	C-LIKE	$2,\!42$
1uvq_A	MHC	C-LIKE	$1,\!80$	$1 vgk_A$	MHC	C-LIKE	$2,\!06$
1ydp_A	MHC	C-LIKE	$1,\!90$	1 zhl_B	MHC	C-LIKE	1,50
$1 jtr_A$	TR	C-ALPHA	2,40	$10ga_D$	TR	C-ALPHA	$1,\!40$
1bec	TR	C-BETA-1	1,70	$10ga_E$	TR	C-BETA-2	$1,\!40$
1hxm_B	TR	C-GAMMA-1	3,12				
G-DOM	AIN		Effec	tif total · 89			
Chaîne	1111	Domaine	R	Chaîne		Domaine	R
1es0 A	MHC	G-ALPHA	2.60	1f3i A	MHC	G-ALPHA	3.10
1fne A	MHC	G-ALPHA	1.90	liak A	MHC	G-ALPHA	1.90
1ik8 A	MHC	G-ALPHA	2.40	1il4 A	MHC	G-ALPHA	4.30
1k2d A	MHC	G-ALPHA	2,20	1muj A	MHC	G-ALPHA	2.15
1r5i A	MHC	G-ALPHA	2.60	1s9v A	MHC	G-ALPHA	2.22
1sje A	MHC	G-ALPHA	$2,\!45$	1uvq A	MHC	G-ALPHA	1,80
2seb A	MHC	G-ALPHA	2,50	1agd A	MHC	G-ALPHA1	2,05
1bii A	MHC	G-ALPHA1	2,40	1e27 A	MHC	G-ALPHA1	2,20
1efx A	MHC	G-ALPHA1	$3,\!00$	1frt A	MHC	G-ALPHA1	4,50
1fzm A	MHC	G-ALPHA1	$1,\!80$	1hsb A	MHC	G-ALPHA1	1,90
1i4f A	MHC	G-ALPHA1	$1,\!40$	1im 9 A	MHC	G-ALPHA1	2,80
1juf A	MHC	G-ALPHA1	2,00	1 k5n A	MHC	G-ALPHA1	1,09
1k8d A	MHC	G-ALPHA1	$2,\!30$	1kjm A	MHC	G-ALPHA1	$2,\!35$
1kjv_A	MHC	G-ALPHA1	$1,\!48$	1kpr_A	MHC	G-ALPHA1	$2,\!80$
1 d9 A	MHC	G-ALPHA1	$2,\!40$	1lk2_A	MHC	G-ALPHA1	$1,\!35$
1m60_A	MHC	G-ALPHA1	$1,\!60$	1mhc_A	MHC	G-ALPHA1	2,10
1nez_A	MHC	G-ALPHA1	$2,\!10$	1r3hA	MHC	G-ALPHA1	2,50

Chaîne		Domaine	R	Chaîne		Domaine	R
1t1w_A	MHC	G-ALPHA1	2,20	1vgk_A	MHC	G-ALPHA1	$2,\!06$
1 w72 A	MHC	G-ALPHA1	$2,\!15$	$1 x7q_A$	MHC	G-ALPHA1	$1,\!45$
1xh3_A	MHC	G-ALPHA1	$1,\!48$	1xr9_A	MHC	G-ALPHA1	1,79
1ydp_A	MHC	G-ALPHA1	$1,\!90$	$2bvp_A$	MHC	G-ALPHA1	$1,\!35$
$1e27_A$	MHC	G-ALPHA2	2,20	1 efx A	MHC	G-ALPHA2	3,00
$1 \mathrm{frt} A$	MHC	G-ALPHA2	4,50	1fzk_A	MHC	G-ALPHA2	1,70
$1hsb_A$	MHC	G-ALPHA2	$1,\!90$	$1i4f_A$	MHC	G-ALPHA2	$1,\!40$
$1k5n_A$	MHC	G-ALPHA2	$1,\!09$	$1 \mathrm{k8d} \mathrm{A}$	MHC	G-ALPHA2	$2,\!30$
1kjm_A	MHC	G-ALPHA2	$2,\!35$	1kjv_A	MHC	G-ALPHA2	$1,\!48$
1kpr_A	MHC	G-ALPHA2	$2,\!80$	$1ld9_A$	MHC	G-ALPHA2	$2,\!40$
1lk2_A	MHC	G-ALPHA2	$1,\!35$	1 m 05 A	MHC	G-ALPHA2	$1,\!90$
1m60_A	MHC	G-ALPHA2	$1,\!60$	$1 \mathrm{mhc} A$	MHC	G-ALPHA2	$2,\!10$
1 nez A	MHC	G-ALPHA2	$2,\!10$	1 qo 3 A	MHC	G-ALPHA2	$2,\!30$
1qqd_A	MHC	G-ALPHA2	2,70	1t1wA	MHC	G-ALPHA2	$2,\!20$
$1 \mathrm{tmc} A$	MHC	G-ALPHA2	2,30	1 vgk A	MHC	G-ALPHA2	$2,\!06$
1 w72 A	MHC	G-ALPHA2	$2,\!15$	1 wbx A	MHC	G-ALPHA2	$1,\!90$
1x7q_A	MHC	G-ALPHA2	$1,\!45$	$1xh3_A$	MHC	G-ALPHA2	$1,\!48$
1xr9_A	MHC	G-ALPHA2	1,79	1ydp_A	MHC	G-ALPHA2	$1,\!90$
2bvp_A	MHC	G-ALPHA2	$1,\!35$	1a6a_B	MHC	G-BETA	2,75
$1d5z_B$	MHC	G-BETA	$2,\!00$	$1d9k_D$	MHC	G-BETA	3,20
$1 \text{es} 0 _ \text{B}$	MHC	G-BETA	$2,\!60$	1f3j_B	MHC	G-BETA	3,10
1fv1_B	MHC	G-BETA	$1,\!90$	1iak_B	MHC	G-BETA	$1,\!90$
1jk8_B	MHC	G-BETA	2,40	$1k8i_B$	MHC	G-BETA	3,10
1klu_B	MHC	G-BETA	$1,\!93$	$1 \mathrm{ktd}$ B	MHC	G-BETA	$2,\!40$
1lnu_B	MHC	G-BETA	$2,\!50$	$1r5v_B$	MHC	G-BETA	$2,\!50$
$1 \mathrm{s}9 \mathrm{v}$ B	MHC	G-BETA	$2,\!22$	$1u3h_D$	MHC	G-BETA	$2,\!42$
1uvq_B	MHC	G-BETA	$1,\!80$	$1 \mathrm{ymm}_B$	MHC	G-BETA	3,50
2iad_B	MHC	G-BETA	2,40				

Suite de la table à la page précédente

Liste des V-PARTNER et C-PARTNER utilisés lors de l'analyse de l'arrangement des domaines dans les V-PARTNER et C-PARTNER. Le type de récepteur (IG ou TR), la description IMGT (Partner) et la résolution cristallographique (R) sont indiquées.

V-PAI	RTN	IER	Effecti	f total :	142	
Chaîne		Partner	R	Chaîne		Partner R
12e8	IG	VH_V-KAPPA	$1,\!90$	15c8	IG	VH_V-KAPPA 2,50
1a2y	IG	VH_V-KAPPA	1,50	1a5f	IG	VH_V-KAPPA 2,80
1a6t	IG	VH_V-KAPPA	2,70	1ad0	IG	VH_V-KAPPA 2,50
1ad9	IG	VH_V-KAPPA	$2,\!80$	1ai1	IG	VH_V-KAPPA 2,80
1aif	IG	VH_V-KAPPA	$2,\!90$	1axt	IG	VH_V-KAPPA 2,15
1b2w	IG	VH_V-KAPPA	$2,\!90$	1baf	IG	VH_V-KAPPA 2,90
1bbj	IG	VH_V-KAPPA	3,10	1bfo	IG	VH_V-KAPPA 2,60
$1 \mathrm{bfv}$	IG	VH_V-KAPPA	2,10	1bln	IG	VH_V-KAPPA 2,80
1bvk	IG	VH_V-KAPPA	2,70	1bz7	IG	VH_V-KAPPA 2,50
1c12	IG	VH_V-KAPPA	$2,\!60$	1c1e	IG	VH_V-KAPPA 1,90
1c5d	IG	VH_V-KAPPA	2,40	1 ce 1	IG	VH_V-KAPPA 1,90
1cic	IG	VH_V-KAPPA	2,50	1clo	IG	VH_V-KAPPA 2,10
1 ct 8	IG	VH_V-KAPPA	2,20	1d5i	IG	VH_V-KAPPA 2,00
1dbb	IG	VH_V-KAPPA	2,70	1dee	IG	VH_V-KAPPA 2,70
1 dn 0	IG	VH_V-KAPPA	2,28	1dqd	IG	VH_V-KAPPA 2,10
1 ds f	IG	VH_V-KAPPA	2,20	1dzb	IG	VH_V-KAPPA 2,00
1e60	IG	VH_V-KAPPA	$1,\!80$	1eap	IG	VH_V-KAPPA 2,50
1egj	IG	VH_V-KAPPA	$2,\!80$	1eo8	IG	VH_V-KAPPA 2,80
1ezv	IG	VH_V-KAPPA	$2,\!30$	1f58	IG	VH_V-KAPPA 2,00
1fgv	IG	VH_V-KAPPA	$1,\!90$	1fj1	IG	VH_V-KAPPA 2,68
$1 \mathrm{fns}$	IG	VH_V-KAPPA	$2,\!00$	$1 \mathrm{fsk}$	IG	VH_V-KAPPA 2,90
1fvc	IG	VH_V-KAPPA	2,20	1 g9 m	IG	VH_V-KAPPA 2,20
1gpo	IG	VH_V-KAPPA	$1,\!95$	1hyx	IG	VH_V-KAPPA 1,80
1i7z	IG	VH_V-KAPPA	$2,\!30$	1i8i	IG	VH_V-KAPPA 2,40
1i9r	IG	VH_V-KAPPA	3,10	1iai	IG	VH_V-KAPPA 2,90
1ibg	IG	VH_V-KAPPA	2,70	1igc	IG	VH_V-KAPPA 2,60
1igm	IG	VH_V-KAPPA	$2,\!30$	1ikf	IG	VH_V-KAPPA 2,50
1iqw	IG	VH_V-KAPPA	2,50	1it9	IG	VH_V-KAPPA 2,80
1j05	IG	VH_V-KAPPA	1,50	1jhl	IG	VH_V-KAPPA 2,40
$1 \mathrm{jps}$	IG	VH_V-KAPPA	$1,\!85$	1k6q	IG	VH_V-KAPPA 2,40
1kcu	IG	VH_V-KAPPA	2,20	$1 \mathrm{kel}$	IG	VH_V-KAPPA 1,90
$1 \mathrm{ken}$	IG	VH_V-KAPPA	3,50	117i	IG	VH_V-KAPPA 1,80
117t	IG	VH_V-KAPPA	2,10	1lk3	IG	VH_V-KAPPA 1,91

Chaîne		Partner	R	Chaîne		Partner	R
$1\mathrm{lo}0$	IG	VH_V-KAPPA	$2,\!00$	1lo4	IG	VH_V-KAPPA	$2,\!40$
$1 \mathrm{mcp}$	IG	VH_V-KAPPA	2,70	1mhp	IG	VH_V-KAPPA	$2,\!80$
$1 \mathrm{mim}$	IG	VH_V-KAPPA	$2,\!60$	1mqk	IG	VH_V-KAPPA	$1,\!28$
$1 \mathrm{mvu}$	IG	VH_V-KAPPA	1,78	1 n0 x	IG	VH_V-KAPPA	$1,\!80$
1n4x	IG	VH_V-KAPPA	1,70	1ncw	IG	VH_V-KAPPA	$1,\!30$
1nlb	IG	VH V-KAPPA	$1,\!60$	$1 \mathrm{ob1}$	IG	VH V-KAPPA	$2,\!90$
$1 \mathrm{op} 3$	IG	VH_V-KAPPA	1,75	$1 \mathrm{opg}$	IG	VH_V-KAPPA	$2,\!00$
$1 \mathrm{osp}$	IG	VH_V-KAPPA	$1,\!95$	1 ots	IG	VH_V-KAPPA	$2,\!51$
1p7k	IG	VH_V-KAPPA	1,75	$1 \mathrm{psk}$	IG	VH_V-KAPPA	$2,\!80$
1q9w	IG	VH V-KAPPA	1,75	1qfu	IG	VH V-KAPPA	$2,\!80$
1qfw	IG	VH_V-KAPPA	$3,\!50$	1qfw	IG	VH_V-KAPPA	3,50
$1 \mathrm{qyg}$	IG	VH V-KAPPA	$1,\!81$	1r0a	IG	VH V-KAPPA	$2,\!80$
$1\mathrm{rhh}$	IG	VH_V-KAPPA	$1,\!90$	$1 \mathrm{rih}$	IG	VH_V-KAPPA	$2,\!50$
1rjl	IG	VH_V-KAPPA	$2,\!60$	1rur	IG	VH_V-KAPPA	$1,\!50$
1rz 7	IG	VH_V-KAPPA	$2,\!00$	1rzg	IG	VH_V-KAPPA	$2,\!00$
1 s3 k	IG	VH_V-KAPPA	$1,\!90$	$1 \mathrm{sbs}$	IG	VH_V-KAPPA	$2,\!00$
1seq	IG	VH_V-KAPPA	1,78	1 sy6	IG	VH_V-KAPPA	$2,\!10$
1t2q	IG	VH V-KAPPA	$1,\!83$	1 t 4 k	IG	VH V-KAPPA	2,50
1tjg	IG	VH_V-KAPPA	$2,\!00$	1ub6	IG	VH_V-KAPPA	$2,\!12$
$1\mathrm{um}4$	IG	VH V-KAPPA	$1,\!80$	1uwx	IG	VH V-KAPPA	$2,\!20$
1uz8	IG	VH V-KAPPA	$1,\!80$	1vge	IG	VH V-KAPPA	$2,\!00$
1wej	IG	VH V-KAPPA	$1,\!80$	$1 \mathrm{wt5}$	IG	VH V-KAPPA	$2,\!10$
1xgy	IG	VH_V-KAPPA	2,71	1yej	IG	VH_V-KAPPA	$1,\!85$
1yqv	IG	VH V-KAPPA	1,70	1yy8	IG	VH V-KAPPA	$2,\!00$
1za3	IG	VH_V-KAPPA	$3,\!35$	1ztx	IG	VH_V-KAPPA	$2,\!50$
2ai0	IG	VH V-KAPPA	$2,\!20$	2brr	IG	VH V-KAPPA	$1,\!95$
2ck 0	IG	VH_V-KAPPA	$2,\!20$	2hrp	IG	VH_V-KAPPA	$2,\!20$
32c2	IG	VH_V-KAPPA	$3,\!00$	43c9	IG	VH_V-KAPPA	$2,\!20$
$6 {\rm fab}$	IG	VH V-KAPPA	$1,\!90$	1aqk	IG	VH V-LAMBDA	$1,\!84$
1etz	IG	VH_V-LAMBDA	$2,\!60$	1jn6	IG	VH_V-LAMBDA	2,70
$1 \mathrm{mfb}$	IG	VH_V-LAMBDA	$2,\!10$	1nj9	IG	VH_V-LAMBDA	$2,\!35$
1q1j	IG	VH_V-LAMBDA	2,50	1 w72	IG	VH_V-LAMBDA	$2,\!15$
2fb4	IG	VH V-LAMBDA	$1,\!90$	8fab	IG	VH V-LAMBDA	$1,\!80$
1ao7	TR	V-ALPHA_V-BETA	$2,\!60$	$1 \mathrm{bd} 2$	TR	V-ALPHA_V-BETA	$2,\!50$
1jtr	TR	V-ALPHA V-BETA	$2,\!40$	$1 \mathrm{kb5}$	TR	V-ALPHA V-BETA	2,50
1lp9	TR	V-ALPHA_V-BETA	$2,\!00$	1nam	TR	V-ALPHA_V-BETA	2,70
10ga	TR	V-ALPHA V-BETA	$1,\!40$	1u3h	TR	V-ALPHA_V-BETA	2,42
2ak4	TR	V-ALPHA_V-BETA	2,50	1hxm	TR	V-GAMMA_V-DELTA	3,12
		—				—	

аъ		תר		e , , 1	22		
$C-P_{1}$	ARINI	ER	Effecti	t total	: 32		
ID		Partner	\mathbf{R}	ID		Partner	\mathbf{R}
1 dn 0	IG	CH1_C-KAPPA	2,28	$1 \mathrm{kcv}$	IG	CH1_C-KAPPA	$1,\!80$
1lk3	IG	CH1_C-KAPPA	$1,\!91$	1q72	IG	CH1_C-KAPPA	1,70
1t4k	IG	CH1_C-KAPPA	2,50	$1 \mathrm{um} 5$	IG	CH1_C-KAPPA	$1,\!60$
1za6	IG	CH1_C-KAPPA	$2,\!80$	2fbj	IG	CH1_C-KAPPA	$1,\!95$
2hmi	IG	CH1_C-KAPPA	$2,\!80$	1adq	IG	CH1_C-LAMBDA	3,15
1etz	IG	CH1_C-LAMBDA	$2,\!60$	1 n f d	IG	CH1_C-LAMBDA	2,80
1 pg7	IG	CH1_C-LAMBDA	2,50	1q0x	IG	CH1_C-LAMBDA	$1,\!60$
100v	IG	CH2_CH2	$2,\!60$	1hzh	IG	CH3_CH3	2,70
1i1c	IG	CH3_CH3	2,70	1igt	IG	CH3_CH3	2,80
1igy	IG	CH3_CH3	3,20	1 ow 0	IG	CH3_CH3	3,10
1f6a	IG	CH4 CH4	3,50	$1 \mathrm{fne}$	MHC	C-LIKE_C-LIKE	$1,\!90$
1fv1	MHC	C-LIKE_C-LIKE	$1,\!90$	$1 \mathrm{hdm}$	MHC	C-LIKE_C-LIKE	2,50
1jk8	MHC	C-LIKE_C-LIKE	2,40	1jpf	MHC	C-LIKE_C-LIKE	2,18
1k8i	MHC	C-LIKE_C-LIKE	$3,\!10$	1kjv	MHC	C-LIKE_C-LIKE	$1,\!48$
1u3h	MHC	C-LIKE_C-LIKE	$2,\!42$	1zhl	MHC	C-LIKE_C-LIKE	1,50
1jtr	TR	$C\text{-}ALPHA_C\text{-}BETA\text{-}1$	2,40	10ga	TR	$\operatorname{C-ALPHA_C-BETA-2}$	$1,\!40$

Suite de la table à la page précédente

Liste des structures de Fab d'IG, IG, TR et MHC utilisées lors de l'analyse des récepteurs. Pour chaque structure la résolution cristallographique (R) est indiquée.

FAB			Effecti	f total :	171				
ID	R	ID	R	ID	R	ID	R	ID	R
12e8	$1,\!90$	15c8	2,50	1a3r	2,10	1a5f	$2,\!80$	1a6t	2,70
1 a d 0	2,50	1ad9	$2,\!80$	1adq	$3,\!15$	1ae6	$3,\!00$	1 ahw	$3,\!00$
1ai1	$2,\!80$	1aif	$2,\!90$	1aqk	1,84	$1 \mathrm{axt}$	$2,\!15$	1b2w	$2,\!90$
$1 \mathrm{b}4 \mathrm{j}$	$2,\!90$	$1 \mathrm{baf}$	$2,\!90$	$1 \mathrm{bbj}$	3,10	1bfo	$2,\!60$	1bln	$2,\!80$
1 bz7	2,50	1c12	$2,\!60$	1c1e	$1,\!90$	1c5d	2,40	1 ce 1	$1,\!90$
$1 \mathrm{cic}$	2,50	1 clo	2,10	1cly	2,50	1 clz	$2,\!80$	1 ct 8	2,20
1d5i	$2,\!00$	1 dbb	2,70	$1 \mathrm{dee}$	2,70	1 dn 0	2,28	1dqd	$2,\!10$
1dqq	$1,\!80$	1e6j	$3,\!00$	1eap	2,50	1egj	$2,\!80$	$1\mathrm{ehl}$	$2,\!40$
1eo8	$2,\!80$	$1 \mathrm{etz}$	$2,\!60$	1f11	$3,\!00$	1f3d	$1,\!87$	1 f4 w	$2,\!30$
1f58	$2,\!00$	$1\mathrm{f8t}$	2,20	1fai	2,70	1fbi	$3,\!00$	1fig	3,00
1fj1	$2,\!68$	1 fl 5	2,10	$1 \mathrm{fns}$	$2,\!00$	$1 \mathrm{fpt}$	$3,\!00$	1frg	$2,\!80$
$1 \mathrm{fsk}$	$2,\!90$	1fvd	2,50	$1 \mathrm{g}9 \mathrm{m}$	2,20	1 gaf	$1,\!95$	1gig	$2,\!30$
1h0d	$2,\!00$	1 hyx	$1,\!80$	1i7z	$2,\!30$	1i9r	3,10	1iai	$2,\!90$
1ibg	2,70	1igc	$2,\!60$	1igf	$2,\!80$	1igi	2,70	1ikf	$2,\!50$
1ind	2,20	1iqw	2,50	1it9	$2,\!80$	1jgl	$2,\!15$	1jgu	$1,\!80$
1jn 6	2,70	$1 \mathrm{kb5}$	2,50	$1 \mathrm{kcu}$	2,20	$1 \mathrm{kel}$	$1,\!90$	$1 \mathrm{ken}$	3,50
$1 \mathrm{kfa}$	$2,\!80$	117i	$1,\!80$	117t	2,10	1lk3	$1,\!91$	1lo0	$2,\!00$
1 lo 4	$2,\!40$	$1 \mathrm{mcp}$	2,70	$1 \mathrm{mfb}$	2,10	1mhp	$2,\!80$	$1 \mathrm{mim}$	$2,\!60$
$1 \mathrm{mju}$	1,22	$1 \mathrm{mnu}$	2,50	1 n0 x	$1,\!80$	$1 \mathrm{nbv}$	$2,\!00$	1nc2	$2,\!10$
1 ncw	$1,\!30$	1nj9	$2,\!35$	1nlb	$1,\!60$	1 ob 1	$2,\!90$	$1 \mathrm{opg}$	$2,\!00$
1orq	3,20	$1 \mathrm{ors}$	$1,\!90$	$1 \mathrm{osp}$	$1,\!95$	1 ots	2,51	1 p2c	$2,\!00$
1 p7k	1,75	$1 \mathrm{pg7}$	2,50	$1 \mathrm{plg}$	$2,\!80$	$1 \mathrm{psk}$	$2,\!80$	1q0x	$1,\!60$
1q1j	2,50	1q9r	$1,\!45$	1qfu	$2,\!80$	$1 \mathrm{qkz}$	$1,\!95$	1 qyg	$1,\!81$
1r0a	$2,\!80$	1r24	$3,\!10$	1r3j	$1,\!90$	$1 \mathrm{rhh}$	$1,\!90$	$1 \mathrm{rih}$	$2,\!50$
1rjl	$2,\!60$	$1 \mathrm{rmf}$	$2,\!80$	1rur	1,50	1 rz7	$2,\!00$	$1 \mathrm{rzg}$	$2,\!00$
1rzi	$2,\!90$	1 s 3 k	$1,\!90$	$1 { m sbs}$	$2,\!00$	1 seq	1,78	1 sm 3	$1,\!95$
1 sy 6	$2,\!10$	1t2q	$1,\!83$	1 t 4 k	2,50	1t66	$2,\!30$	1tjg	$2,\!00$
$1 \mathrm{txv}$	2,75	$1 \mathrm{tzh}$	$2,\!60$	1u6a	$2,\!81$	$1\mathrm{ub5}$	$2,\!00$	1uj 3	$2,\!10$
$1 \mathrm{um4}$	$1,\!80$	1 uwx	2,20	1uz 8	$1,\!80$	$1 \mathrm{vge}$	$2,\!00$	1 w 72	$2,\!15$
1 wej	$1,\!80$	$1 \mathrm{xgy}$	2,71	1 yed	$3,\!10$	1 yej	$1,\!85$	1yjd	2,70
$1 \mathrm{ynl}$	1,70	1 y q v	1,70	1 yuh	3,00	1yy 8	$2,\!00$	1za3	3,35
1ztx	2,50	2a1w	2,70	$2\mathrm{brr}$	$1,\!95$	2 ck 0	2,20	2 fb4	$1,\!90$
2fbj	$1,\!95$	2h1p	2,40	2hrp	2,20	2jel	2,50	2pcp	2,20
32c2	$3,\!00$	3 h fm	3,00	4fab	2,70	$6 {\rm fab}$	$1,\!90$	8fab	$1,\!80$

IG			Effecti	f total	: 3				
ID	R	ID	R	ID	R				
1hzh	2,70	1igt	2,80	1igy	3,20				
TR			Effecti	f total	: 5				
ID	R	ID	\mathbf{R}	ID	R	ID	R	ID	\mathbf{R}
1jtr	$2,\!40$	1lp9	$2,\!00$	1oga	$1,\!40$	1qrn	2,80	2ak4	2,50
MH	C-I		Effecti	f total	: 12				
ID	R	ID	\mathbf{R}	ID	R	ID	R	ID	\mathbf{R}
1frt	$4,\!50$	1i4f	1,40	1k5n	$1,\!09$	$1\mathrm{k8d}$	$2,\!30$	1kjv	$1,\!48$
1kpr	$2,\!80$	1lk 2	$1,\!35$	$1 \mathrm{mhc}$	$2,\!10$	1nez	2,10	1r3h	2,50
1vgk	2,06	1ydp	1,90						
MH	C-II		Effecti	f total	: 7				
ID	R	ID	\mathbf{R}	ID	R	ID	R	ID	\mathbf{R}
1f3j	3,10	$1 \mathrm{fne}$	$1,\!90$	1hdm	$2,\!50$	1jk8	2,40	1k8i	3,10
1klu	$1,\!93$	1u3h	2,42						

Liste des IG utilisés lors de l'analyse de ma flexibilité des CDR lors de la liaison IG/antigène. Les identifiant des structures (ID) avec une étoile correspondent aux structures ne présentant pas d'antigène (forme apo).

IG	IDs
10E5	1txv 1 ty 3
10G6D6	1jn6* $1jnh$
13B5	1e60* 1e6j
1696	$1n4x^*$ $1jp5$
17-Ia	$1 for^* 1 rvf$
17/9	1hil $*$ 1ifh
17E12E5	1jnl $*$ 1 jnn
19D9D6	$1nlb^* \ 1n64$
19G2	$1\mathrm{ub6}^*$ 1fl3
1D4	1 jgv* 1 jgu
21D8-chimeric	$1c5b^* \ 1c5c$
26-10	1igi* 1igj
28	2hmi 1n5y
2F5	$2f5a^* 1tjg 2f5b 1u91$
39-A11	1a4j* 1a4k
3D6	$1dfb^*$ 10be
3F4	$1 cr 9^* 1 cu 4$
4-4-20	1flr 4fab
4155	1bfv* 1cfv
43C9	43c9* 43ca
48G7	1gaf* 1aj7
4C4	lezv 1ejo
50.1	$1 \mathrm{ggb}^*$ $1 \mathrm{ggi}$
57-2	$1jhk^* 1jgl$
58.2	1f58 $3f58$ $2f58$
59.1	lacy lail
5C8	15c8*35c825c8
5G9	$1 \mathrm{fgn}^* 1 \mathrm{ahw}$
77	$117t^*$ $117s$
7G12	$1 \text{ngy}^* 1 \text{n7m} 1 \text{ngx}$
8F5	$1 \mathrm{bbd}^* \ 1 \mathrm{a} 3 \mathrm{r}$
9B1	1 q 0 y 1 q 0 x
	1 0 1

9D9	1104* 1103 1100 1102
AL2 6E > 7 P9 > G	1h80* 1h8s
AZ-28 precursor-chimeric	1d5i* 1d6v
Anti-Testosterone	1i9i* 1i9i
B1-8	1a6u* 1a6v 1a6w
B13I2	ligf* 2igf
Bv04-01	$1 \text{nbv}^* 1 \text{cbv}$
C219	$1ap2^* 2ap2$
CETUXIMAB	1yy8*1yy9
Cab-CA05	$1f2x^*$ 1g6v
Campath-1H	1bey* 1ce1
Cb41	1cfq* 1hh9 1hh6 1hi6 1cfn
	1cft 1cfs 1bog
Cha255	1ine 1ind
Cnj206	$2 \mathrm{gfb}^* 1 \mathrm{kno}$
D1.3	1cic* 1vfb
D2.3	1yeh* 1yec 1yej 1yek 1yef
	1kn4 1kn2 1yei 1yeg
D3H44	$1 \mathrm{jpt}^* 1 \mathrm{jps}$
D44.1	$1 m lb^* 1 m lc$
DNA-1	$1i8m^* 1xf2$
Db3	2dbl 1dbk 1dbj 1dba 1dbb
	$1 \mathrm{dbm}$
E8	$1 qbl^* 1 wej$
F11.2.32	$1 m f 2^* 2 hrp$
Fab131	$2 ck0 \ 3 ck0$
Fab28	1j50 1r0a
HULYS11	$1bvl^*$ $1bvk$
Hemagglutinin neutralising IG	2vis 2vit 2vir

IG

Hyhel-10

Hyhel-63

Lnkb-2

M82G2

MS5-393

MS6-12 McPC603

Hyhel-5

Suite de la table à la page précédente

IDs

Suite de la table à la page suivante

1hfm* 3hfm

3hfl 2iff 1bql

1dqm* 1dqj 1f8t* 1f90

 $1 qyg^* 1q72$

1mie^{*} 1mj7

1mju^{*} 1mjj

 $1mcp^* 2mcp$

IG IDs				
Mcg	$1dcl^*$ 1mck 1mcn 1mcl			
	Imcd Imcj Imci Imci Imce			
M_{22} 19119	1mcc Inco Incq Inci Iaoj			
	ımnu impa			
NIG9	$lngq^{*}$ $lngp$			
NC10	1a14 $1nmc$			
NC41	$1nca \ 1ncb$			
NC6.8	$1 cgs^* 1 ynl 2 cgr$			
PC282	$1 \mathrm{kcv}^* \ 1 \mathrm{kcs}$			
PC287	$1 m kcu^{*} \ 1 m kc5$			
S-20-4	$1 f4 w^* 1 f4 x$			
S25-2	1q9k* 1q9v 1q9r 1q9w 1q9q			
SPE-7	10aq* 10ax 10au 10az			
Sya/J6	$1m71^*$ $1pz5$			
Tp7	1ay1* $1bgx$			

Liste des structures d'IG utilisées pour le calcul des contacts IG/antigène.

1ar1	3 h fl	$3 \mathrm{hfm}$	1acy	1a3r	1a6w	1a6v	1 ahw
1a2y	1a4k	1a3l	1ai1	1afv	1aj7	$1 \mathrm{bgx}$	$1 \mathrm{bog}$
$1 \mathrm{tzg}$	1bql	1bvk	1c08	1c12	1c5c	$1 \mathrm{cbv}$	1 ce 1
1cf8	$1 \mathrm{cfs}$	$1 \mathrm{cft}$	$1 \mathrm{cfv}$	1 ct 8	1 cu 4	1 d6 v	1dba
1dbj	1dbk	1 dee	1dl 7	1dqj	1dzb	1f3d	1f3r
1f4x	1f58	1f90	1fbi	$1 \mathrm{fe8}$	1fig	1fj1	1fl3
1flr	$1 \mathrm{fns}$	$1 \mathrm{fpt}$	$1 \mathrm{fsk}$	1e4x	1 e 6 j	1eap	1egj
$1\mathrm{ehl}$	1ejo	$1\mathrm{em}4$	1etz	1ezv	1g7h	1g7i	1g7j
1g7l	$1 \mathrm{g7m}$	1g9n	1ggi	1 h8 s	$1 \mathrm{hez}$	1 hh6	1hh9
1 hi 6	$1 \mathrm{him}$	1hin	1hyx	1hyy	1i3g	1i7z	1i8i
1i8k	1i9j	1i9r	1ibg	1ic4	1ic5	1ic 7	1igj
1iif	1ine	1iqd	1j1o	1j1p	1j1x	1j5o	1jgl
1jgu	1jhl	1jnh	1jnn	1jp 5	1jps	1jrh	1k4c
1k4d	1kb9	$1 \mathrm{kc5}$	$1 \mathrm{kcs}$	$1 \mathrm{kfa}$	1kn4	$1 \mathrm{kcr}$	$1 \mathrm{ken}$
1kip	1kiq	1kir	$1 \mathrm{kn} 2$	1kno	$1 \mathrm{ktr}$	117s	1lk3
$1\mathrm{lo}0$	1lo2	1lo3	$1 \mathrm{mhh}$	$1 \mathrm{mlc}$	1mpa	10au	1n5y
1n6q	1ncc	1ngp	1nma	1nmc	$1 \mathrm{nsn}$	10ak	$1 \mathrm{osp}$
1qle	1qkz	1qnz	$1 \mathrm{rvf}$	$1 \mathrm{sm}3$	1vfb	$1 \mathrm{wej}$	1yec
1yed	1yee	1yef	1yeg	1 yuh	1yei	1yej	1 yek
25c8	$2 \mathrm{cgr}$	2ap2	$2 \mathrm{bfv}$	2f58	2h1p	$2 \mathrm{hmi}$	2hrp
2mcp	$2\mathrm{pcp}$	2vir	2vis	2vit	35c8	3f58	3fct
43ca	4fab	1kyo	1dbb	$1 \mathrm{nmb}$	$1 \mathrm{bj} 1$	1qfu	1mhp
1n64	$1 \mathrm{nby}$	1nbz	1ncw	1ngx	1ob1	1orq	1 ors
1 ots	1 ott	1otu	1p84	1pkq	2f5b	$1 \mathrm{g}9 \mathrm{m}$	1ifh
1cz8	1eo8	2 h fm	$2 \mathrm{mpa}$	1 gc1	1baf	1ikf	1 mj7
1tet	2igf	$1 \mathrm{cfn}$	$1 \mathrm{dbm}$	$1 \mathrm{hys}$	1ind	1nca	1ncb
$1 \mathrm{ncd}$	2dbl	1 e4 w	$1 \mathrm{mex}$	$1 \mathrm{mh5}$	$1 \mathrm{mjj}$	$1 \mathrm{mvu}$	1 nak
2 ck 0	$3 \mathrm{ck} 0$	1nc2	1nc4	1nl 0	10ax	10ay	10az
1 p2c	1q1j	1q9q	1q9r	1q9t	1q 9 v	1q9w	1ru9
1rua	1ruk	1rul	$1 \mathrm{rum}$	1rup	1rzk	1ua6	1uac
$1 \mathrm{v7m}$	1 v7n	1 n0 x	1 p 4 b	1 q0 x	1 q 0 y	1s78	$1\mathrm{ub}5$
$1\mathrm{s}5\mathrm{h}$	1t66	1 pz 5	1uj3	$1 \mathrm{tzh}$	1tzi	1a0q	1rjl
1tjg	$1 \mathrm{tjh}$	1tji	1 txv	1 ty 3	$1 \mathrm{ty} 5$	1 ty 6	$1 \mathrm{ty} 7$
$1\mathrm{u}8\mathrm{h}$	1u8i	1u8j	1u8k	1u8l	1u8m	1u8n	1u8o
1u8p	1u8q	1u91	1u92	1u93	1u95	1 w72	$1 \mathrm{frg}$
1yjd	$1 \mathrm{seq}$	1 svz	1n8z	1r0a	1fdl	2iff	1ndg

1ndm	1h0d	1n7m	1nd0	$1 \mathrm{ngw}$	1qfw	1r3i	1r3j
1r3k	1r3l	1riu	1riv	1t03	$1 \mathrm{tpx}$	1tqb	$1 \mathrm{tqc}$
1uz8	1a14	$1 \mathrm{xf} 2$	1yqv	1yy9	1yyl	1yym	2bob
2boc	1c1e	1obe	1q72	1rzj	2jel	1uwx	1xgp
1xgq	1xgr	1xgt	1xgu	1xgy	1ynk	1ynl	1za3
1ztx	$2\mathrm{brr}$						

Suite de la table à la page précédente

Liste des structures de pMHC-I et de pMHC-II. Pour chaque structure, la séquence du peptide, le ou les allèles du MHC et le nom et l'espèce du peptide sont indiqués. Pour les pMHC-II la séquence des neufs acides aminés situés dans le site de liaison est fournie.

Chaîne du	Séquence du	Allèle du MHC	Fanàga at nom du nantida
peptide	peptide	MIIO	Espèce et nom du peptide
Peptides s	e liant au MHC-I	de taille de 8 (40	structures)
1kj2 P	KVITFIDL	H2-K1b	[Mouse] GTP-binding protein 1 peptide 161-
<u> </u>			168 pKB1 (O08582)
1kj3 P	KVITFIDL	H2-K1b	[Mouse] GTP-binding protein 1 peptide 161-
• <u> </u>			168 pKB1 (O08582)
1agd C	GGKKKYKL	HLA-B*0801	[HIV-1] Gag peptide P17 23-30 (P03349)
1agb C	GGRKKYKL	HLA-B*0801	[HIV-1] Gag peptide P17 23-30 (P03349),
0 _			K3>R
1 agf C	GGKKRYKL	HLA-B*0801	[HIV-1] Gag peptide P17 23-30 (P03349),
0 _			K5>R
1 agc C	GGKKKYQL	HLA-B*0801	[HIV-1] Gag peptide P17 23-30 (P03349),
0 _	-		K7>Q
1age C	GGKKKYRL	HLA-B*0801	[HIV-1] Gag peptide P17 23-30 (P03349),
0 _			K7>R
1rjy P	SSIEFARL	H2-K1b	[Herpes simplex virus] Glycoprotein B peptide
_			448-505 (P06436)
1rk0 P	SSIEFARL	H2-K1b	[Herpes simplex virus] Glycoprotein B peptide
_			448-505 (P06436)
1t0m P	SSIEFARL	H2-K1b	[Herpes simplex virus] Glycoprotein B peptide
_			448-505 (P06436)
1t0n P	SSIEFARL	H2-K1b	[Herpes simplex virus] Glycoprotein B peptide
			448-505 (P06436)
1rjz_P	SEIEFARL	H2-K1b	[Herpes simplex virus] Glycoprotein B peptide
			448-505 (P06436), S2>E
1rk1_P	SEIEFARL	H2-K1b	[Herpes simplex virus] Glycoprotein B peptide
			448-505 (P06436), S2>E
1 n 59 P	AVYNFATM	H2-K1b	[LCMV] Glycoprotein G1 peptide 33-41
			(P07399), C9>M
$1 \mathrm{s7q} \mathrm{C}$	AVYNFATM	H2-K1b	[LCMV] Glycoprotein G1 peptide 33-41
			(P07399), C9>M
$1 \mathrm{s7s} \mathrm{C}$	ALYNFATM	H2-K1b	[LCMV] Glycoprotein G1 peptide 33-41
			(P07399), V3>L C9>M
$1 \mathrm{s7t} \mathrm{C}$	AVFNFATM	H2-K1b	[LCMV] Glycoprotein G1 peptide 33-41
			$(P07399), Y4{>}F C9{>}M$
$1e28_C$	TAFTIPSI	HLA-B*5101	[HIV-1] Immunodominant epitope KM2 pep-
			tide 282-289 (P12499)
$1 \text{lk}2_P$	GNYSFYAL	H2-K1b	[Mouse] Insulin receptor beta-subunit peptide
			423-430 (P15208)

Chaîne du peptide	Séquence du peptide	Allèle du MHC	Espèce et nom du peptide
lg7q_P	SAPDTRPA	H2-K1b	[Human] Mucin 1 transmembrane peptide 180-187 (NM_002456)
1jtr_P	EQYKFYSV	H2-K1b	[Mouse] NADH-ubiquinone oxidoreductase MLBO subunit peptide 61-68 (O62425)
$1 \log_P$	EQYKFYSV	H2-K1b	[Mouse] NADH-ubiquinone oxidoreductase MLRO subunit peptide 61-68 (Q62425)
1lek_P	EQYKFYSV	H2-K1b	[Mouse] NADH-ubiquinone oxidoreductase MLRO subunit peptide 61-68 (Q62425)
1mwa_P	EQYKFYSV	H2-K1b	[Mouse] NADH-ubiquinone oxidoreductase MLRQ subunit peptide 61-68 (Q62425)
2ckb_P	EQYKFYSV	H2-K1b	[Mouse] NADH-ubiquinone oxidoreductase MLRQ subunit peptide 61-68 (Q62425)
1a1n C	VPLRPMTY	HLA-B*3501	[HIV-1] Nef peptide 78-85 (P03407)
1bqh_C	RGYVYQGL	H2-K1b	[Stomatitis] Nucleocapsid protein VSV8 pep- tide 52-59 (P11212)
1fzj_P	RGYVYQGL	H2-K1b	[Stomatitis] Nucleocapsid protein VSV8 pep- tide 52-59 (P11212)
1fzm_P	RGYVYQGL	H2-K1b	[Stomatitis] Nucleocapsid protein VSV8 pep- tide 52-59 (P11212)
1nam_P	RGYVYQGL	H2-K1b	[Stomatitis] Nucleocapsid protein VSV8 pep- tide 52-59 (P11212)
$2 \mathrm{mha}\mathrm{E}$	RGYVYQGL	H2-K1b	[Stomatitis] Nucleocapsid protein VSV8 pep- tide 52-59 (P11212)
2vaa_P	RGYVYQGL	H2-K1b	[Stomatitis] Nucleocapsid protein VSV8 pep- tide 52-59 (P11212)
$1 \text{osz} _ \text{C}$	RGYLYQGL	H2-K1b	[Stomatitis] Nucleocapsid protein VSV8 pep- tide 52-59 (P11212), V4>L
1kpu_P	RGYVYQGL	H2-K1b	[Stomatitis] Nucleocapsid protein peptide 52- 59 (P11212)
1p1z_P	SIINFEKL	H2-K1b	[Chicken] Ovalbumin peptide 257-264 (P01012)
1p4l_P	SIINFEKL	H2-K1b	[Chicken] Ovalbumin peptide 257-264 (P01012)
1vac_P	SIINFEKL	H2-K1b	[Chicken] Ovalbumin peptide 257-264 (P01012)
1g6r P	SIYRYYGL	H2-K1b	[Chimeric] Superantagonist peptide SIYR
1fo0 P	INFDFNTI	H2-K1b	[Mouse] pBM1 peptide
1nan_M	INFDFNTI	H2-K1b	[Mouse] pBM1 peptide
Peptides se l 1g7p_P	liant au MHC-I de SRDHSRTPM	taille de 9 (85 st H2-K1b	ructures) [Yeast] Alpha-glucosidase MAL32 peptide 438-446 (P38158)
1vad_P	SRDHSRTPM	H2-K1b	[Yeast] Alpha-glucosidase MAL32 peptide 438-446 (P38158)
1kjm_P	AQFSASASR	RT1-A	[Synthetic] B6 Peptide

Chaîne du peptide	Séquence du peptide	Allèle du MHC	Espèce et nom du peptide
$1 \mathrm{s9w} \mathrm{C}$	SLLMWITQC	HLA-A*0201	[Human] Cancer/testis antigen 1B peptide 157-165 (P78358)
$1 \mathrm{s9x}_{\mathrm{C}}$	SLLMWITQA	HLA-A*0201	[Human] Cancer/testis antigen 1B peptide 157-165 (P78358) C9>A
$1s9y_C$	SLLMWITQS	HLA-A*0201	[Human] Cancer/testis antigen 1B peptide 157-165 (P78358), $C9>S$
1kjv_P	NPRAMQALL	RT1-A	[Rat] DA41 peptide 450458 (BAA92267)
1 m 05 E	FLRGRAYGL	HLA-B*0801	[EBV] EBNA-3A peptide 193-201 (P12977), I9>L
1mi5_{C}	FLRGRAYGL	HLA-B*0801	[EBV] EBNA-3A peptide 193-201 (P12977), I9>L
$1a9b_C$	LPPLDITPY	HLA-B*3501	[EBV] EBNA-3C peptide 98-106 (P03204)
$1a9e_C$	LPPLDITPY	HLA-B*3501	[EBV] EBNA-3C peptide 98-106 (P03204)
1cg9_{C}	LPPLDITPY	HLA-B*3501	[EBV] EBNA-3C peptide 98-106 (P03204), H4>P
1a1m C	TPYDINQML	HLA-B*5301	[HIV-2] Gag peptide P16 182-190 (P18095)
1 uxs_C	RRRWRRLTV	HLA-B*2705	[EBV] Gene terminal protein peptide 236-244 (P13285)
1uxw_C	RRRWRRLTV	HLA-B*2709	[EBV] Gene terminal protein peptide 236-244 (P13285)
1 fg 2 C	KAVYNFATC	H2-D1b	[LCMV] Glycoprotein G1 peptide 33-41 (P07399)
$1 \mathrm{ffn}_{\mathrm{C}}$	KAVYNFATM	H2-D1b	[LCMV] Glycoprotein G1 peptide 33-41 (P07399), C9>M
$1n5a_C$	KAVYNFATM	H2-D1b	[LCMV] Glycoprotein G1 peptide $33-41$ (P07399) C9>M
$1s7u_C$	KAVYNFATM	H2-D1b	[LCMV] Glycoprotein G1 peptide $33-41$ (P07399) C9>M
$1s7r_C$	KAVYNLATM	H2-K1b	[LCMV] Glycoprotein G1 peptide 33-41 (P07399) $F_{0} > L_{0} > M$
$1s7v_C$	KAVYNLATM	H2-D1b	[LCMV] Glycoprotein G1 peptide $33-41$ (P07399) $F6>L C9>M$
1ffo_C	AAVYNFATM	H2-D1b	(107355), $10 > 10 < 0 > M$ [LCMV] Glycoprotein G1 peptide 33-41 (P07399) $K1 > 4 C9 > M$
$1 \mathrm{ffp}_C$	SAVYNFATM	H2-D1b	(107399), $K1 > K C9 > M$ [LCMV] Glycoprotein G1 peptide 33-41 (P07300) $K1 > S C0 > M$
1 s7wC	KALYNFATM	H2-D1b	(107399), $K1>5$ C9>M [LCMV] Glycoprotein G1 peptide 33-41 (P07300) V3>L C0>M
1s7x_C	KAVFNFATM	H2-D1b	(107399), $\sqrt{3} \ge 1 C g > M$ [LCMV] Glycoprotein G1 peptide 33-41 (P07300) $\sqrt{4} \ge C 0 \ge M$
1 inq_C	SSVVGVWYL	H2-D1b	[Mouse] H13A minor histocompatibility anti- ren pentide 51-59 (AAB81862)
1juf_C	SSVIGVWYL	H2-D1b	[Mouse] H13A minor histocompatibility anti- ran pontide 51 50 (AAB\$1962) V4> I
1kpr P	VMADRTVII	HIA F*0103	gen pepude 51-59 (AAD61605), $V4>1$ [Human] HI A B*0714 pontide (HI A01040)
1mhe P	VMAPRTVII	HLA_E*0101	[Human] HLA-B*0714 peptide (HLA01049) [Human] HLA-B*0714 peptide (HLA01040)
1ktl P	VTAPRTLLL	HLA-E*0103	[Human] HLA-B*1313 peptide (HLA01858)

Chaîne du peptide	Séquence du peptide	Allèle du MHC	Espèce et nom du peptide
1im9 C	QYDDAVYKL	HLA-Cw*0401	[Human] HLA-Cw4 specific peptide
laad C	QYDDAVYKL	HLA-Cw*0401	[Human] HLA-Cw4 specific peptide
1m6o_C	EEFGRAFSF	HLA-B*4402	[Human] HLA-DPA1*0201 peptide 46-54
$1n2r_C$	EEFGRAFSF	HLA-B*4403	(HLA00504) [Human] HLA-DPA1*0201 peptide 46-54 (HLA00504)
1syv_C	EEFGRAFSF	HLA-B*44	[Human] HLA-DPA1*0201 peptide 46-54 (III A 00504)
1wbz_P	SSYRRPVGI	H2-K1b	[Influenza A virus] I RNA-directed RNA poly- merase subunit P1 peptide 703-711 (P21426)
$1e27_C$	LPPVVAKEI	HLA-B*5101	[HIV-1] Immunodominant epitope KM1 pep- tide 742-750 (P24740)
$1 {\rm efx}_{\rm C}$	GAVDPLLAL	HLA-Cw*0304	[Human] Importin alpha-2 subunit peptide 204-212 (P52292)
1alo C	KPIVQYDNF	HLA-B*5301	[Plasmodium] LSA-1 1786-1794 (Q25893)
1jge [–] C	GRFAAAIAK	HLA-B*2705	[Synthetic] M9 peptide
1k5n C	GRFAAAIAK	HLA-B*2709	[Synthetic] M9 peptide
1hhi_C	GILGFVFTL	HLA-A*0201	[Influenza A virus] Matrix protein M1 peptide 58-66 (Q66PA1)
$10ga_C$	GILGFVFTL	HLA-A*0201	[Influenza A virus] Matrix protein M1 peptide 58-66 (Q66PA1)
$1 jht_C$	ALGIGILTV	HLA-A*0201	[Human] Melan-A peptide 27-35 (NP 005502), A2>L
1 w 72 C	EADPTGHSY	HLA-A*0101	[Human] Melanoma-associated antigen 1 pep- tide 161-169 (P43355)
$1 \mathrm{qew} C$	FLWGPRALV	HLA-A*0201	[Human] Melanoma-associated antigen 3 pep- tide 271-279 (P43357)
1kpv_P	FAPGNYPAL	H2-K1b	[Sendai] Nucleocapsid protein peptide 324-332 (P04857)
$1ce6_C$	FAPGNYPAL	H2-D1b	[Sendai] Nucleocapsid protein peptide 324-332 (P04857)
lfzk_P	FAPGNYPAL	H2-K1b	[Sendai] Nucleocapsid protein peptide 324-332 (P04857)
lfzo_P	FAPGNYPAL	H2-K1b	[Sendai] Nucleocapsid protein peptide 324-332 (P04857)
$2 vab_P$	FAPGNYPAL	H2-K1b	[Sendai] Nucleocapsid protein peptide 324-332 (P04857)
$1 q l f_C$	FAPSNYPAL	H2-D1b	[Sendai] Nucleocapsid protein peptide 324-332 (P04857) C4>S(O CleNac)
1bz9_C	FAPGVFPYM	H2-D1b	(104357), 04>5(0-0)(1040) [Sendai] Nucleocapsid protein peptide 324-332 (P04857) N5 \sim V V6 \sim E A8 \sim V 10 \sim M
li7r_C	FAPGFFPYL	HLA-A*0201	(104037), 10 > 7 10 > 7 Ao > 1 L9 > M [Sendai] Nucleocapsid protein peptide 324-333 (D04857) N5 > F V6 > F A8 > V
1 hoc C	ASNENMETM	H2-D1b	[Influenza A virus] Nucleoprotein peptide 366- 374 (P18277)

Chaîne du peptide	Séquence du peptide	Allèle du MHC	Espèce et nom du peptide
1jpg_C	FQPQNGQFI	H2-D1b	[LCMV] Nucleoprotein peptide 396-4 04 (M20869)
1ldp P	APAAAAAAM	H2-Ld-2	[Melanogaster] Peptide
1hsa C	ARAAAAAAA	HLA-B*2705	[Human] Peptide
1 d9 C	YPNVNIHNF	H2-Ld-2	[Synthetic] Peptide
1 qr1 C	IISAVVGIL	HLA-A*0201	[Human] Receptor tyrosine-protein kinase erbB-2 peptide 654-662 (P04626)
1eez_C	ILSALVGIL	HLA-A*0201	$ \begin{array}{c c} [{\rm Human}] & {\rm Receptor} & {\rm tyrosine-protein} & {\rm kinase} \\ {\rm erbB-2} & {\rm peptide} & 654\text{-}662 & ({\rm P04626}), & {\rm I2{>}L}, \\ {\rm L5{>}V} \end{array} $
1eey_C	ILSALVGIV	HLA-A*0201	[Human] Receptor tyrosine-protein kinase erbB-2 peptide 654-662 (P04626), $I2>L$, L5>V, $L9>V$
$1 \mathrm{hhj}_{\mathrm{C}}$	ILKEPVHGV	HLA-A*0201	[HIV-1] Reverse transcriptase (POL) peptide 476-484 (P03366)
$1 p7q_C$	ILKEPVHGV	HLA-A*0201	[HIV-1] Reverse transcriptase (POL) peptide 476-484 (P03366)
$1i1f_C$	FLKEPVHGV	HLA-A*0201	[HIV-1] Reverse transcriptase (POL) peptide 476-484 (P03366), I1>F
li1y_C	YLKEPVHGV	HLA-A*0201	[HIV-1] Reverse transcriptase (POL) peptide 476-484 (P03366), I1>Y
$1 \mathrm{k8d}_{\mathrm{P}}$	ILMEHIHKL	H2-Q9	[Mouse] Self 60s ribosomal protein peptide 137-145 (P84099)
1b0g C	ALWGFFPVL	HLA-A*0201	[Human] Self peptide P1049
1 lp9 C	ALWGFFPVL	HLA-A*0201	[Human] Self peptide P1049
$1i7t_C$	ALWGVFPVL	HLA-A*0201	[Human] Self peptide P1049, F $>5V$
$1i7u_C$	ALWGFVPVL	HLA-A*0201	[Human] Self peptide P1049, F>6V
1sys_C	EEPTVIKKY	HLA-B*4403	[Human] Sorting nexin 5 peptide 257-265 (Q9Y5X3)
$1ao7_C$	LLFGYPVYV	HLA-A*0201	[HTLV-1] Tax peptide 11-19 (Q82235)
$1 \mathrm{bd2}\mathrm{C}$	LLFGYPVYV	HLA-A*0201	[HTLV-1] Tax peptide 11-19 (Q82235)
1 duz C	LLFGYPVYV	HLA-A*0201	[HTLV-1] Tax peptide 11-19 (Q82235)
$1 \mathrm{hhk}\mathrm{C}$	LLFGYPVYV	HLA-A*0201	[HTLV-1] Tax peptide 11-19 (Q82235)
$1 \mathrm{im} 3 \mathrm{C}$	LLFGYPVYV	HLA-A*0201	[HTLV-1] Tax peptide 11-19 (Q82235)
$1 qrn_C$	LLFGYAVYV	HLA-A*0201	[HTLV-1] Tax peptide 11-19 (Q82235), P6>A
$1 qse_C$	LLFGYPRYV	HLA-A*0201	[HTLV-1] Tax peptide 11-19 (Q82235), V7>R
$1qsf_C$	LLFGYPVAV	HLA-A*0201	[HTLV-1] Tax peptide 11-19 (Q82235), Y8>A
1 of 2 C	RRKWRRWHL	HLA-B*2709	[Human] Vasoactive intestinal polypeptide receptor 1 peptide 400-408 (P32241)
$1 \mathrm{ogt}_{\mathrm{C}}$	RRKWRRWHL	HLA-B*2705	[Human] Vasoactive intestinal polypeptide receptor 1 peptide 400-408 (P32241)
$1 \mathrm{hhg}_{\mathrm{C}}$	TLTSCNTSV	HLA-A*0201	[HIV-1] gp120 envelope protein peptide 192- 200 (P03375)
1q94_C	AIFQSSMTK	HLA-A*1101	[HIV-1] reverse transcriptase (POL) peptide 325-333 (P03366)

Suite	de	la	table	à	la	page	précéd	lente
-------	----	----	-------	---	----	------	--------	-------

Chaîne du peptide	Séquence du peptide	Allèle(s) du MHC	Espèce et nom du peptide
1akj_C	ILKEPVHGV	HLA-A*0201	[HIV-1] reverse transcriptase (POL) peptide 476-484 (P03366)
1 jgd C	RRLLRGHNQY	HLA-B*27	[Human] HLA-B*2707 peptide S10R 107-116 (HLA00228) G1>B
$1n3n_I$	SALQNAASIA	H2-D1b	[Mycobacterium bovis] HSP60 epitope peptide
Peptides se	e liant au MHC-I de	taille de 10 (12 struc	ctures)
1wbx_C	SQLKNNAKEI	H2-D1b	[Influenza A virus] Hemagglutinin peptide 468-477 (P26140)
$1 \mathrm{jf1}_\mathrm{C}$	ELAGIGILTV	HLA-A*0201	$ \begin{array}{c c} [\mathrm{Human}] & \mathrm{Melan-A} & \mathrm{peptide} & 26\text{-}35 \\ (\mathrm{NP} \ 005502), \ \mathrm{A2>L} \end{array} $
$1i4f_C$	GVYDGREHTV	HLA-A*0201	[Human] Melanoma-associated antigen 4 pep- tide 230-239 (P43358)
lavo C	QVPLRPMTYK	HLA-A*1101	[HIV-1] Nef peptide 77-86 (P03407)
1hhh C	FLPSDFFPSV	HLA-A*0201	[HBV] Nucleocapsid protein peptide 18-27
1tmc C	EVAPPEYHRK	HLA-A*6801	[Human] Peptide
1wby_C	SSLENFRAYV	H2-D1b	[Influenza A virus] Polymerase subunit pep- tide 224-233 (Q8QLZ6)
1bii_P	RGPGRAFVTI	H2-D1d	[HIV-1] gp120 envelope protein peptide 311- 320 (P03375)
$1 ddh_P$	RGPGRAFVTI	H2-D1d	[HIV-1] gp120 envelope protein peptide 311- 320 (P03375)
1qo3_P	RGPGRAFVTI	H2-D1d	[HIV-1] gp120 envelope protein peptide 311- 320 (P03375)
Peptides se	e liant au MHC-II		
1s9v_C	PFPQPELPY	HLA-DQA1*0501 HLA-DQB1*0201	[Wheat] Alpha/beta-gliadin peptide 78-88 (P02863), Q8>E
$2 \mathrm{seb}\mathrm{E}$	MRADAAAGG	HLA-DRA*0101 HLA-DRB1*0401	[Human] Collagen alpha 1(II) peptide 1168- 1179 (P02458), Q1>A 7Q>A L12>A
$1r5w_E$	IAYFKAATK	H2-EAck H2-EB1k	[Moth] Cytochrome c peptide 96-107 (P00037), L7>F E9>A S10>A
$1r5v_E$	IAYPKAATK	H2-EAck H2-EB1k	[Moth] Cytochrome c peptide 96-107 (P00037), L7>P E9>A S10>A
$1h15_C$	YHFVKKHVH	HLA-DRA*0101 HLA-DBB5*0101	[Human] DNA Polymerase peptide 628-641 (P03198)
$1 sjh_C$	VIPMFSALS	HLA-DRA*0101 HLA-DBB1*0101	[HIV-1] Gag peptide P17 167-179 (P03349)
1sje_C	VIPMFSALS	HLA-DRA*0101 HLA-DRB1*0101	[HIV-1] Gag peptide P17 167-182 (P03349)
1lnu_B	AQKAKANKA	H2-AAb H2-ABb	[Mouse] H-2 Class II Histocompatibility Anti-
1aqd_C	WRFLRGYHQ	HLA-DRA*0101 HLA-DRB1*0101	

Suite de la table à la page précédente

Chaîne du peptide	Séquence du peptide	Allèles du MHC	Espèce et nom du pentide
peptide	peptide	MIIIO	Lispece et nom du peptide
$1a6a_C$	MRMATPLLM	HLA-DRA*0101 HLA DRB1*0301	[Human] HLA-DR antigens associated inva-
1muj_C	MRMATPLLM	H2-AAb H2-ABb	[Human] HLA-DR antigens associated inva-
1dlh_C	YVKQNTLKL	HLA-DRA*0101	[Influenza] Hemagglutinin HA1 peptide 322-
$1 \mathrm{fyt}_{\mathrm{C}}$	YVKQNTLKL	HLA-DRB1*0101 HLA-DRA*0101	334 (P03437) [Influenza] Hemagglutinin HA1 peptide 322-
1dlh_C	YVKQNTLKL	HLA-DRB1*0101 HLA-DRA*0101	334 (P03437) [Influenza] Hemagglutinin HA1 peptide 322-
$1 \mathrm{fyt}_C$	YVKQNTLKL	HLA-DRB1*0101 HLA-DRA*0101	334 (P03437) [Influenza] Hemagglutinin HA1 peptide 322-
1hxy_C	YVKQNTLKL	HLA-DRB1*0101 HLA-DRA*0101	334 (P03437) [Influenza] Hemagglutinin HA1 peptide 322-
1j8h C	YVKQNTLKL	HLA-DRB1*0101 HLA-DRA*0101	334 (P03437) [Influenza] Hemagglutinin HA1 peptide 322-
1jwm C	YVKQNTLKL	HLA-DRB1*0401 HLA-DRA*0101	334 (P03437) [Influenza] Hemagglutinin HA1 peptide 322-
1jws C	YVKQNTLKL	HLA-DRB1*0101 HLA-DRA*0101	334 (P03437) [Influenza] Hemagglutinin HA1 peptide 322-
1jwu C	YVKQNTLKL	HLA-DRB1*0101 HLA-DRA*0101	334 (P03437) [Influenza] Hemagglutinin HA1 peptide 322-
1ko0 D	YVKONTLKL	HLA-DRB1*0101 HLA-DRA*0101	334 (P03437) [Influenza] Hemagglutinin HA1 peptide 322-
11 ₀₅ C	VVKONTLKL	HLA-DRB1*0101 HLA-DRA*0101	334 (P03437) [Influenza] Homogolutinin HA1 poptide 322
1.5. 0		HLA-DRB1*0101	334 (P03437)
Irbi_C	YVKQNILKL	HLA-DRA*0101 HLA-DRB1*0101	[Influenza] Hemaggiutinin HAT peptide 322- 334 (P03437)
1jk8_C	EALYLVCGE	HLA-DQA1*0302 HLA-DQB1*0302	[Human] Insulin B chain peptide 35-48 (P01308), F48>G
lf3j_P	RHGLDNYRG	H2-AAd H2-AB- NOD	[Chicken] Lysozyme peptide 29-42 (P00698)
1iak_P	DYGILQINS	H2-AAk H2-ABk	[Chicken] Lysozyme peptide 68-80 (P00698)
1 bx2 C	VHFFKNIVT	HLA-DRA*0101 HLA-DRB1*1501	[Human] Myelin basic protein MBP peptide 117-230 (P02686)
$1 fv1_F$	FKNIVTPRT	HLA-DRA*0101 HLA-DRB5*0101	[Human] Myelin basic protein MBP peptide 119-236 (P02686)
1k2d_P	GGASQYRPS	H2-AAu H2-ABu	[Human] Myelin basic protein MBP peptide 131-142 (P02686), L1>S D2>R V3>G M4>G Q8>R
$1 uvq_C$	LPSTKVSWA	HLA-DQA1*0102 HLA-DQB1*0602	[Human] Orexin peptide 1-13 (O43612)
1d9k_P	HRGAIEWEG	H2-AAk H2-ABk	[Chicken] Ovotransferrin peptide 150-165 (P02789), T1>G L2>N L3>S

Suite de la table à la page précédente

Chaîne du peptide	Séquence du peptide	Allèles du MHC	Espèce et nom du peptide
1jl4_C	HRGAIEWEG	H2-AAk H2-ABk	[Chicken] Ovotransferrin peptide 150-165 (P02789), T1>G L2>N L3>S
1t5wC	YSDQATPLL	HLA-DRA*0101 HLA-DRB1*0101	[Yeast] Regulatory protein MIG1 peptide 448- 460 (P27705), S1>A S2>A S3>Y L5>D S6>O T7>A
1t5x_C	YSDQATPLL	HLA-DRA*0101 HLA-DRB1*0101	[Yeast] Regulatory protein MIG1 peptide 448- 460 (P27705), S1>A S2>A S3>Y L5>D S6>O T7>A
1klu_C	IGTLNAAKV	HLA-DRA*0101 HLA-DBB1*0101	[Human] Triosephosphate isomerase peptide 21-36 (P60174)
$1 \rm klg_C$	IGILNAAKV	HLA-DRA*0101 HLA-DRB1*0101	[Human] Triosephosphate isomerase peptide 21-36 (P60174), T6>I
1kt2 B	IAYLKQATK	H2-EAck H2-EB1k	[Mouse] pMHC100
$1 \mathrm{ktd} \mathbf{B}$	IAYLKQASA	H2-EAck H2-EB1k	[Mouse] pMHC101
$2iad_B$	TQGVTAASS	H2-AAd H2-ABd	[Mouse] pMHC114
$1 \mathrm{es0}$ B	IAPVFVLLE	H2-AAd H2-AB- NOD	[Mouse] pMHC71
1fne B	ITAFNDGLK	H2-EAck H2-EB1k	[Mouse] pMHC74
lfng_B	ITAFNEGLK	H2-EAck H2-EB1k	[Mouse] pMHC75
1i3r_B	ITAFNEGLK	H2-EAck H2-EB1k	[Mouse] pMHC84
1iao_B	SQAVHAAHA	H2-AAd H2-ABd	[Mouse] pMHC86
1 iea_B	ITAFNEGLK	H2-EAck H2-EB1k	[Mouse] pMHC87
$1ieb_B$	VNHFIAEFK	H2-EAck H2-EB1k	[Mouse] pMHC88

Liste des structures de MHC utilisées pour le classement de la librairie de SCHISMo.

1mfa	1c5c	1a2y	1bec	$1 \mathrm{bww}$	1dlf	1eeq	1eeu
1efq	1g7j	1g7p	1g7q	1i4f	$1 \mathrm{k5n}$	1kgc	1kjv
1l6x	2dlf	2rhe	1kxv	1kxq	1kpu	1kpv	1 mqk
1mvf	1n4x	1ncw	1nlb	10ga	10p3	1lk2	1m60
$1 \mathrm{mex}$	1 mj 8	1mvu	1q72	1j05	1oaq	1n2r	$1 \mathrm{ogt}$
1q9o	1q9q	1q 9 r	1q9t	1q 9 v	1q9w	1ruk	1rup

Liste des structures extraites de IMGT/3Dstructure-DB utilisées pour la dérivation de la librairie de LLIPA.

1mfa	12e8	1c5c	$1 \mathrm{atm}$	$1 \mathrm{aag}$	$1 \mathrm{aqk}$	1a6w	1a6v
1a2y	1a3l	$1\mathrm{b}0\mathrm{w}$	$1\mathrm{bec}$	1bww	1c1e	1 cd0	1 ce 1
1dgx	$1\mathrm{dh4}$	$1\mathrm{dh}5$	$1\mathrm{dh}6$	$1\mathrm{dh7}$	1dh8	1dh9	1 dha
1dho	1 dhq	$1 \mathrm{dhu}$	$1 \mathrm{dhv}$	1dhw	1 dhz	1dlf	1 dqq
1duz	1 dv f	1dx2	1dx3	1f3d	1f3r	$1 \mathrm{fgv}$	$1 \mathrm{flr}$
1fne	$1 \mathrm{fng}$	$1 \mathrm{fv1}$	1fvb	1fvw	$1 \mathrm{fzj}$	1 fzk	$1 \mathrm{fzm}$
1fzo	1 e4 w	1e4x	1e6o	$1 \mathrm{eeq}$	1eeu	1efq	$1\mathrm{ek}3$
$1\mathrm{em}4$	1g7h	1g7i	1g7j	1g7m	1g7p	1g7q	1g 84
1gaf	$1 \mathrm{gpo}$	$1 \mathrm{bwm}$	1h5b	1h8n	$1 \mathrm{hfm}$	1hou	$1 \mathrm{hsb}$
1hyx	1hyy	1i4f	1i7u	1i8k	1iak	1ieh	1iga
1ige	1iif	1j1o	1j1p	1j1x	1jf1	1jfq	1jgu
1jgv	$1 \mathrm{jps}$	1jpt	1jtp	1jvk	$1 \mathrm{k5n}$	$1 \mathrm{kcv}$	$1 \mathrm{kel}$
1kgc	$1 \mathrm{kiq}$	1kjv	1klu	$1 \mathrm{kn} 2$	$1 \mathrm{kn4}$	1l6q	116x
1l7i	1 lds	$1 \log$	11k3	1lve	$1 \mathrm{maj}$	$1 \mathrm{mak}$	$1 \mathrm{mig}$
1obe	$1 \mathrm{osp}$	$1 \mathrm{qac}$	$1 \mathrm{qgc}$	$1 \mathrm{qkz}$	1qnz	$1 \mathrm{rog}$	$1 \mathrm{roh}$
1roi	1roj	$1 \mathrm{rok}$	$1 \mathrm{rol}$	$1 \mathrm{sm} 3$	1 tvd	1 v fa	$1 \mathrm{vfb}$
1vhp	$1 \mathrm{wej}$	$1 \mathrm{wtl}$	1yec	1yei	1yej	2cd 0	2dlf
2fb4	2fbj	2fvb	2fvw	2 h fm	2ige	$2\mathrm{imn}$	2loi
2psk	$2 \mathrm{rhe}$	$6 {\rm fab}$	8fab	1kxv	1kxq	1jgd	1kpu
1kpv	1lgv	$1 \mathrm{mqk}$	$1 \mathrm{mvf}$	1n4x	$1 \mathrm{nby}$	$1 \mathrm{nbz}$	1ncw
1ndg	1ngx	$1 \mathrm{nlb}$	$1 \mathrm{oga}$	$1 \mathrm{op} 3$	1ors	1lk 2	$1 \mathrm{m} 05$
$1 \mathrm{m} 6 \mathrm{o}$	$1 \mathrm{mex}$	$1 \mathrm{mie}$	1 mj 8	$1 \mathrm{mju}$	1mvu	10p9	1 pz 5
1q72	$1 \mathrm{qyg}$	1j05	10aq	10au	1n2r	1 ogt	100
1q9k	1q9o	1q 9 q	1q 9 r	1 q9 t	1q9v	1q9w	1rua
1ruk	$1 \mathrm{rul}$	$1\mathrm{rum}$	1rup	1ruq	1rur	1rzf	1ua6
1uac	1uvq	1 n0 x	1 q0 x	1s7q	1s7s	1 sjv	1pew
1s3k	1 pw3	1 sq2	1xfp	1 p7k	1rjy	$1 \mathrm{syv}$	$1 \mathrm{t0n}$
$1\mathrm{um}4$	$1\mathrm{um}5$	$1\mathrm{um}6$	1uxs	1uxw	$1 \mathrm{xh}3$	1xiw	1r70
1rhh	1ri8	1rjc	$1 \mathrm{wbx}$	1seq	1 svz	$1 \mathrm{fnl}$	1g9e
1hcv	1i3u	1jnj	1n7m	1ngz	1r3j	$1 \mathrm{shm}$	1 t 6 v
$1 \mathrm{t7v}$	1uz8	1ydp	2fcb	1lqv	1t2j	1 tvb	$1 \mathrm{tvh}$
1x9q	1xr9	1yqv					

Publications.