



HAL
open science

MODELISATION DES REARRANGEMENTS V(D)J AU NIVEAU DU LOCUS TRA/TRD

Maria-Ana Simonet

► **To cite this version:**

Maria-Ana Simonet. MODELISATION DES REARRANGEMENTS V(D)J AU NIVEAU DU LOCUS TRA/TRD. Immunologie. Université Joseph-Fourier - Grenoble I, 2008. Français. NNT: . tel-00516844

HAL Id: tel-00516844

<https://theses.hal.science/tel-00516844>

Submitted on 12 Sep 2010

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

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

Université Joseph Fourier-Grenoble I
UFR de Biologie

THESE

Pour obtenir le grade de
Docteur de l'Université Joseph Fourier
(Arrêté ministériels du 5 Juillet 1984 et du 30 mars 1992)
Spécialité Modélisation des systèmes vivants

Par Maria-Ana Simonet

Soutenue le 22 Décembre 2008

**MODELISATION DES REARRANGEMENTS
V(D)J AU NIVEAU DU LOCUS TRA/TRD**

Devant le jury suivant

Pr Alain DESSEIN	Président du jury
Pr Marie-Paule LEFRANC	Rapporteur
Dr Bastien FERNANDEZ	Rapporteur
Pr Jacques DEMONGEOT	Directeur de thèse
Dr Evelyne JOUVIN-MARCHE	Directeur de thèse

Thèse préparée en co-direction entre
le laboratoire Technique de l'Imagerie Médicale et de la Complexité, CNRS UMR5525
et le laboratoire d'Immunologie Analytique des Pathologies Chroniques, U823
Université Joseph Fourier

Remerciements

La thèse, un rêve de petite fille. J'y arrive enfin, à force de persévérance et de ténacité. En effet comment ne pas être découragé dès les premières heures de cours à la fac où les professeurs nous disent que seulement une vingtaine d'entre nous arriverons à concrétiser leur rêve de passer une thèse et de faire de la recherche. Mais la preuve est que quand on y croit et que l'on ne baisse pas les bras alors tout est possible.

La thèse est un moment privilégié, où l'on apprend à faire de la recherche, à réfléchir, à analyser et à persévérer. L'enthousiasme de découvrir de nouveaux résultats est vraiment une motivation de continuer importante, tout comme la déception d'une manip qui n'a pas fonctionné, qui même si elle amène à passer par des étapes de découragements, permet au final de disposer d'une motivation supplémentaire à réussir.

Les trois années de thèse, à travers les étapes d'enthousiasme et de découragement forgent un entêtement certain à ne pas abandonner ce que l'on aime faire mais aussi de prendre conscience que la recherche n'est pas un long fleuve tranquille. Tout cela n'aurait pas été possible sans la présence des séniors dont l'expertise et les connaissances sont des éléments capitaux.

J'en arrive donc naturellement à remercier toutes les personnes qui m'ont entouré et encadré au cours de ma thèse. Sans eux, ce long et difficile moment n'aurait pas été aussi instructif et agréable.

Je remercie mes directeurs de thèse, Jacques Demongeot et Evelyne Jouvin-Marche, pour m'avoir encadré, pour leur apport scientifique, pour leur confiance, et leur soutien moral. Le seul regret que j'ai est de ne pas avoir pu en apprendre encore davantage à leurs côtés.

Je remercie Patrice Marche pour m'avoir accueilli dans son laboratoire et pour ses discussions scientifiques.

Je remercie Olivier Hansen pour sa collaboration sur le modèle de simulation et sa patience tout au long des étapes de l'élaboration du modèle, pour m'avoir fait partager ces connaissances en modélisation.

Je tiens également à remercier toute l'équipe 8 avec qui j'ai travaillé au cours de mes trois années de thèse et plus particulièrement Aurélie, Ingrid et Christina, pour avoir mis une bonne ambiance dans le bureau et m'avoir soutenu dans les moments difficiles. Les discussions du vendredi par souris interposées vont me manquer. Carine pour les longues discussions scientifiques (ou pas) et pour m'avoir reboosté à plusieurs reprises, et tous les autres pour leur partage scientifique, leur amitié et les moments de craquage collectif que l'on a partagé.

Je remercie également le Pr Dominique Blin pour avoir contribué à me redonné confiance en moi après mon DEA difficile, à travers la collaboration au projet NOESIS.

Sans la famille la thèse n'aurait pas été possible. Avec des parents chercheurs on peut soit avoir envie de rentrer dans ce milieu, soit avoir envie de faire totalement autre chose. Et j'ai eu envie de faire de la recherche. Je n'aurais pas de mots assez forts pour remercier mes parents de leur soutien tout au long de ma thèse, mais aussi dans les années qui l'ont précédée. Je remercie mon papa qui a relu ma thèse ainsi que tous les autres rapports et présentations, à la recherche des erreurs de syntaxe et d'orthographe. Mes petites sœurs aussi pour leur soutien, mais aussi pour les balades à cheval qui permettent de s'aérer la tête et de se changer les idées. Et enfin ma maman, qui m'a toujours poussée à me surpasser à travers ces remarques plus que réaliste tout au long de mon cursus.

Sans les amis ce long chemin n'aurait pas été aussi amusant et agréable. Je remercie donc tous mes amis pour les grandes soirées passées à rigoler mais aussi à discuter plus sérieusement ou à refaire le monde. Merci pour votre soutien, votre patience et vos encouragements.

Sommaire

1.	Introduction	4
2.	Les lymphocytes T	8
2.1.	Origine des lymphocytes T	8
2.2.	La différenciation des lymphocytes T	8
2.2.1.	Les cellules doubles négatives	9
2.2.1.1.	Les cellules DN1	9
2.2.1.2.	Les cellules DN2	9
2.2.1.3.	Les cellules DN3	9
2.2.1.4.	Les cellules DN4	10
2.2.2.	Les cellules ISP	10
2.2.3.	Les cellules DP	11
2.2.4.	Les cellules SP	11
2.3.	La sélection des lymphocytes T	13
2.3.1.	La sélection positive	13
2.3.2.	La sélection négative	13
2.3.3.	La sélection β	14
2.4.	Le TR.....	15
2.4.1.	Les loci du TR	15
2.4.1.1.	Nomenclature des gènes	15
2.4.1.2.	Nomenclature de la chaîne TR α	16
2.4.1.3.	Comparaison du locus TRA/TRD entre l'homme et la souris	16
2.4.2.	Structure des TR	18
2.4.2.1.	Les chaînes du TR	20
3.	La régulation des réarrangements des lymphocytes T	22
3.1.	La recombinaison V(D)J	22
3.2.	Les RSS et la règle des 12/23	23
3.3.	Les facteurs impliqués dans la recombinaison V(D)J	24
3.3.1.	Les enzymes RAG1 et RAG2	24
3.3.2.	HMG1 et 2	25
3.3.3.	ARTEMIS	25
3.3.4.	La protéine kinase ATM	27
3.3.5.	DNAPK	27
3.3.6.	La protéine XRCC4	27
3.3.7.	La ligase IV	28
3.3.8.	Le complexe XLF/Cernunnos	28
3.4.	Contrôle des réarrangements du locus TRA/TRD	29

3.4.1.	Les enhanceurs	29
3.4.1.1.	Enhanceur α	30
3.4.1.2.	Enhanceur δ	30
3.4.2.	Boundary Element Alpha Delta (BEAD).....	31
3.4.3.	T Early α (TEA)	32
3.4.3.1.	TEA et exclusion allélique	33
3.4.4.	Locus Control Region (LCR).....	34
3.4.5.	"Conserved Sequence Block" de la région AJ	34
3.4.6.	L'accessibilité de la chromatine.....	36
3.4.6.1.	Méthylation	36
3.4.6.2.	Acétylation	36
3.5.	La diversité du répertoire	39
3.5.1.	La diversité combinatoire.....	39
3.5.2.	La diversité jonctionnelle	40
3.5.3.	La diversité d'appariement TR $\alpha\beta$ ou TR $\gamma\delta$	40
4.	Modélisation.....	41
4.1.	Dans quel but ?.....	41
4.2.	Etat de l'art de la modélisation en immunologie	41
4.3.	Les différentes manières d'aborder la modélisation	43
4.4.	Les pièges à éviter	43
4.5.	Le modèle existant de recombinaison V(D)J	43
4.5.1.	Le modèle du groupe de Warmflash	43
4.5.2.	Notre modèle	44
5.	Objectifs	46
6.	Article 1	47
7.	Article 2.....	49
8.	Discussion et conclusion de l'article 2.....	54
8.1.	Effet de la vitesse variable et du temps de maturation sur le profil de recombinaison .	54
8.1.1.	Effet d'une vitesse fixe sur le répertoire	54
8.1.2.	Effet d'une vitesse d'ouverture variable sans temps de maturation	55
8.1.2.1.	Effet d'une vitesse d'ouverture variable avec un temps de maturation.....	57
8.1.3.	Validation du modèle : suite.....	58
8.1.3.1.	Proposition pour la seconde vague de recombinaison.....	59
9.	Avancées du modèle.....	60
9.1.	TR α et exclusion allélique.....	60
9.1.1.	Avancées du modèle.....	60

9.2.	Résultats	63
10.	Modélisation des réarrangements VJ au niveau du locus TRA/TRD humain	65
10.1.	Introduction	65
10.2.	Matériels et Méthodes	65
10.2.1.	Mise en place des conditions expérimentales de PCR	65
10.2.1.1.	Conditions de PCR et choix des amorces.....	65
10.2.1.2.	Choix des couples de gènes V et J testés	66
10.2.1.3.	Liste des amorces	67
10.2.2.	Provenance des échantillons.....	68
10.3.	Résultats	68
10.3.1.	Analyse des profils de recombinaison.....	68
10.4.	Modélisation.....	73
10.4.1.	Détermination des vitesses d'ouverture	73
10.4.2.	Calcul des vitesses d'ouverture.....	74
10.4.3.	Résultats	74
10.4.3.1.	Couverture de réarrangement	78
10.4.3.2.	Localisation des Hot Spots de recombinaison.....	78
10.4.3.3.	Fréquences de réarrangement.....	79
10.4.3.4.	Première fenêtre d'ouverture	79
10.4.3.5.	Réarrangements secondaires	80
10.4.3.6.	Profils de recombinaisons et RSS	80
10.5.	Discussion et conclusions.....	82
11.	Discussion et perspectives.....	84
12.	Annexe 1 : Nomenclature des gènes TRAV murin et humain.....	88
13.	Annexe 2 : Séquences des RSS murin et humain et pourcentage d'homologie	90
14.	Bibliographie	100

Introduction

1. Introduction

Le système immunitaire a pour fonction de protéger l'organisme contre les différents pathogènes, les substances étrangères et de maintenir "l'intégrité" de l'organisme en éliminant toute cellule anormale (tumorale, infectieuse, etc...). Il doit s'adapter en permanence aux différentes agressions de l'environnement. Pour cela il a développé une capacité à produire une grande diversité de récepteurs pour reconnaître le plus grand nombre d'antigènes (Ag) à partir d'un nombre fini de gènes.

L'immunité se subdivise en l'immunité innée et l'immunité adaptative. L'immunité innée est mise en place de manière précoce lors d'une agression et elle n'est pas spécifique d'un Ag. Elle permet d'éliminer rapidement l'infection ou du moins de la contenir le temps que le système immunitaire adaptatif soit activé.

La peau et les muqueuses représentent le premier rempart contre les pathogènes extérieurs. La peau va empêcher la majorité des pathogènes de pénétrer dans l'organisme. Les muqueuses, que l'on trouve au niveau de toutes les portes d'entrée naturelles, sont composées de cellules épithéliales qui sécrètent des molécules bactéricides et antivirales. Ces molécules permettent l'élimination des agents pathogènes après leur immobilisation par les cellules ciliées ou le mucus qui les tapisse.

Lorsqu'il y a une blessure ou une dégradation des tissus, l'intégrité de l'organisme est fragilisée. Les pathogènes vont alors pénétrer dans l'organisme, mais ils sont rapidement neutralisés par les cellules du système immunitaire inné postées dans ces tissus. Ces dernières sont principalement composées de macrophages et de cellules dendritiques. Ces cellules dérivent de cellules souches hématopoïétiques et possèdent à leur surface des récepteurs reconnaissant un signal de danger présent à la surface des agents infectieux. Ces récepteurs sont capables d'initier une forte réponse inflammatoire amorcée par une sécrétion d'agents antimicrobiens et de cytokines pro-inflammatoires. L'inflammation provoque une vasodilatation et une augmentation de la perméabilité des capillaires qui permettent le recrutement de macrophages et de neutrophiles qui phagocyteront les pathogènes.

Les cellules de l'hôte présentent à leur surface des marqueurs qui permettent leur reconnaissance par les cellules du système immunitaire : le complexe majeur d'histocompatibilité (CMH). L'expression de molécules du CMH sur des cellules saines les étiquette comme appartenant à l'hôte et les protège de l'activité potentiellement destructrice du système immunitaire. Le CMH est composé de deux classes majeures de molécules, le CMH de classe I (CMH I) et le CMH de classe II (CMH II). L'expression de ces deux types de molécules diffère selon le type cellulaire. Le CMH I est exprimé à la surface de toutes les cellules, tandis que le CMH II l'est uniquement à la surface de certaines cellules, comprenant

les lymphocytes B, les macrophages, les cellules dendritiques et les cellules de l'épithélium thymique.

Le CMH I est exprimé par l'ensemble des cellules de l'organisme. Il présente des peptides du soi et du non soi (peptides viraux, peptides modifiés). C'est l'association CMH I / peptides du non soi qui déclenche l'activation du système immunitaire. Le CMH II a une expression plus limitée, restreinte essentiellement aux cellules présentatrices d'antigènes (macrophage, cellules dendritiques, lymphocytes B). L'association CMH II / peptide viraux a pour conséquence d'activer spécifiquement les lymphocytes T de l'hôte.

L'immunité adaptative se met en place en 3 à 5 jours. Elle est médiée par les lymphocytes T et les lymphocytes B. Ces derniers s'activent lors du contact avec un antigène. Les lymphocytes B et T dérivent de cellules souches hématopoïétiques, mais ils n'agissent pas de manière identique.

Les lymphocytes B (LB) se développent dans la moelle osseuse. Ils possèdent à leur surface des immunoglobulines (Ig) capables de reconnaître un antigène à l'état natif. Les anticorps de surface, ou BR fixent l'antigène en entier. Le complexe Ag/BR est ensuite internalisé et dégradé dans le lysosome. Les peptides sont chargés par des molécules de CMH II et le complexe peptide/CMH est exprimé à la surface des lymphocytes B pour être présenté aux lymphocytes T CD4⁺. L'activation des LB permet aussi leur différenciation en cellules plasmocytes, qui sécrètent des Anticorps (Ac) spécifiques d'un Ag. Les Ig sécrétées se fixent sur les antigènes, les neutralisent et favorisent leur capture par les cellules phagocytaires. Les cytokines produites vont aussi participer à l'activation des lymphocytes T ainsi qu'au recrutement des cellules de l'immunité innée qui vont participer à la phagocytose de l'antigène.

Les lymphocytes T terminent leur développement dans le thymus après avoir migré depuis la moelle osseuse. Ils expriment à leur surface un récepteur de cellule T (TR) qui existe sous deux formes TR $\alpha\beta$ ou TR $\gamma\delta$ avec chacun une spécificité fonctionnelle distincte. Les TR $\alpha\beta$ reconnaissent l'Ag présenté sous forme de peptide dans le contexte du CMH à la surface des cellules de l'hôte. Cette reconnaissance entraîne l'élimination de la cellule reconnue soit par apoptose induite par le lymphocyte T activé, dans le cadre d'une cellule cancéreuse ou infectée par un virus, soit par l'activation de cellules adaptées (lymphocytes B ou cellules phagocytaires) dans le cas de la reconnaissance d'un peptide provenant d'un agent infectieux extracellulaire. Les lymphocytes T exprimant le récepteur TR $\gamma\delta$ sont essentiellement localisés à la surface des épithéliums. Ils protègent l'organisme en répondant à des signaux de danger potentiel ou de stress cellulaire émis par les cellules de l'organisme (Girardi 2006). De plus ils auraient aussi une activité antitumorale et immunorégulatrice (Bonneville and Fournie 2005).

Dans les cellules présentatrices de l'antigène, comme les lymphocytes B ou les cellules dendritiques, l'antigène est dégradé en peptides. Ces derniers sont présentés par le CMH aux TR des lymphocytes T qui présentent un récepteur TR $\alpha\beta$. Il existe deux types de LT exprimant des TR $\alpha\beta$: les CD4⁺ et les CD8⁺. Les CD4⁺ reconnaissent un complexe peptide/CMH de classe II. Ces lymphocytes permettent une amplification de la réponse immunitaire par le biais de leur sécrétion d'interleukines. Les CD8⁺ reconnaissent un peptide présenté par une molécule de classe I du CMH et ont une action directe sur les cellules infectées. L'action conjointe du système immunitaire inné et adaptatif permet d'éliminer la plupart des pathogènes et corps étrangers.

Chaque TR et BR peut reconnaître un nombre limité d'antigènes. Par conséquent, l'expression d'un répertoire étendu de récepteurs lymphocytaires est importante pour assurer la défense de l'hôte. Pour cela, un mécanisme générant une grande diversité s'est développé dans les cellules impliquées dans la réponse immunitaire spécifique. Les gènes codant pour les récepteurs TR et BR des lymphocytes sont localisés de manière discontinue sur le génome en configuration germinale. Leur expression dépend d'un mécanisme permettant la juxtaposition des différents gènes pour former un exon unique. Ce mécanisme, la recombinaison V(D)J, permet à un nombre limité de gènes de produire un grand nombre de protéines différentes. Cette recombinaison est irréversible. En général, un lymphocyte mature qui se divise produit des cellules filles exprimant un récepteur identique. De cette manière, l'activation d'un récepteur par un antigène donné provoque une expansion clonale, ce qui entraîne la production d'un grand nombre de cellules capables de reconnaître et de neutraliser l'antigène responsable de cette activation.

La diversité d'un récepteur peut être engendrée de trois manières différentes :

1. par une diversité combinatoire, qui correspond aux différentes combinaisons de gènes qui peuvent être réalisées,
2. par la diversité jonctionnelle : des paires de bases vont être insérées ou délétées au niveau de la juxtaposition des gènes,
3. par la diversité d'appariement : la diversité d'un récepteur est augmentée car les récepteurs sont formés de deux chaînes (α et β), qui sont produites chacune par le même mécanisme de recombinaison. De par le mécanisme mis en jeu, chaque cellule va produire une chaîne α et une chaîne β différente des cellules voisines. Ce sont les différents appariements possibles entre les chaînes α et β qui traduit la diversité d'appariement.

Par ailleurs, l'expression d'un récepteur unique à la surface des lymphocytes est assurée par le mécanisme d'exclusion allélique. L'expression d'un récepteur ou d'un pré-récepteur à la

surface cellulaire produit l'arrêt de l'expression des gènes réarrangés ou du réarrangement sur les loci codant pour des chaînes non exprimées en surface.

Bien que l'on connaisse le mécanisme qui permette de générer la diversité des récepteurs des lymphocytes, il n'est pas possible, à l'heure actuelle, de quantifier cette diversité, et de dire précisément quels gènes sont recombinaisonnés entre eux et avec quelle fréquence. C'est ici que l'on trouve l'intérêt d'une démarche de modélisation qui permet de simuler les réarrangements V(D)J et d'obtenir les fréquences de réarrangements théoriques. La diversité pouvant être générée au niveau des TR $\alpha\beta$ n'est pas encore connue, que ce soit chez la souris ou chez l'homme. Le logiciel qui a été développé au cours de ma thèse permet de simuler les réarrangements V(D)J au niveau de la chaîne α des TR chez la souris et chez l'homme. Il permet ainsi d'avoir une idée de la diversité combinatoire générée au niveau de la chaîne α du TR. Il permet aussi d'appréhender des paramètres biologiques qui ne sont pas quantifiables expérimentalement.

2. Les lymphocytes T

2.1. Origine des lymphocytes T

Les lymphocytes T se développent à partir de cellules souches hématopoïétiques (CSH), localisées tout d'abord dans le foie fœtal puis dans la moelle osseuse, qui sera le site définitif de production des cellules hématopoïétiques tout au long de la vie. Ces cellules hématopoïétiques vont migrer depuis la moelle osseuse vers le thymus pour y effectuer leur différenciation. Les cellules souches qui arrivent dans le thymus sont à l'origine des lymphocytes T, mais aussi des cellules natural killer (NK), des macrophages et des cellules dendritiques.

2.2. La différenciation des lymphocytes T

L'expression, la différenciation et la sélection des lymphocytes T sont réalisées dans le thymus. Le thymus est un organe immunitaire spécialisé dans la maturation des lymphocytes T. Il se situe au dessus du cœur.

Avant d'être opérationnels et d'aller coloniser l'organisme, les lymphocytes T suivent différentes phases de différenciation et de sélection afin de vérifier que le lymphocyte ne présente pas de risque pour l'organisme (comme par exemple l'élimination de ses cellules saines) (Figure 2).

Au sein du thymus, l'expression de cinq marqueurs membranaires a permis d'identifier les différents stades de maturation des lymphocytes :

1) le marqueur CD44, qui est une protéine impliquée dans les phénomènes d'adhésion et de migration cellulaire (Wu, Kincade et al. 1993) ;

2) le marqueur CD25, qui correspond à la chaîne alpha du récepteur à l'interleukine 2. Il est exprimé de manière constitutive sur les lymphocytes T CD4, et réexprimé de manière transitoire après activation des lymphocytes T en réponse à un antigène ;

3) le marqueur CD4, qui est un co-récepteur associé à la reconnaissance du complexe CMH II/peptide par le TR $\alpha\beta$;

4) le marqueur CD8, qui est le co-récepteur associé à la reconnaissance du complexe CMH I/peptide par le TR $\alpha\beta$;

5) le marqueur CD3 ϵ , qui est un des composants du module CD3 de transduction du signal associé au récepteur TR $\alpha\beta$.

Les lymphocytes vont passer par le stade double négatif (DN) où ils ne vont exprimer ni CD4 ni CD8. Ces cellules deviennent ensuite immatures simple positif (ISP) et vont commencer à réarranger la chaîne TR α . Le stade suivant, appelé double positif (DP) correspond à

l'expression simultanée des molécules CD4 et CD8 à la surface des lymphocytes. A ce stade les thymocytes finissent de réarranger les gènes codant pour leurs récepteurs et expriment ces derniers à leur surface. Il vient ensuite une étape de sélection des TR néo-formés lors de la sélection positive et négative. Enfin, le dernier stade de différenciation avant la sortie des lymphocytes à la périphérie est appelé simple positif (SP). Les cellules T se différencient de manière irréversible en LT CD4 ou LT CD8 et s'apprêtent à coloniser les organes périphériques secondaires.

Dans les prochains paragraphes, j'explicitai la différenciation des lymphocytes T chez la souris. Chez l'homme ces étapes sont presque identiques, et les étapes où l'on peut observer des divergences seront indiquées.

2.2.1. Les cellules doubles négatives

Les cellules doubles négatives pour les marqueurs CD4 et CD8 présentent le stade le plus précoce dans la différenciation intra-thymique des thymocytes. Au cours de cette étape de différenciation, les thymocytes passent progressivement par quatre stades de différenciation suivant l'expression de marqueurs cellulaires spécifiques allant du stade DN1 ($CD44^+CD25^-$) jusqu'au stade DN4 ($CD44^-CD25^-$) en passant par les stades DN2 ($CD44^+CD25^+$) et DN3 ($CD44^-CD25^+$) (Figure 1). Les cellules des stades DN1 et DN2 sont appelées cellules pré-T tandis que les cellules DN3 et DN4 sont appelées cellules pro-T chez la souris.

2.2.1.1. Les cellules DN1

Le stade DN1 est le stade le plus immature des lymphocytes T. Les cellules sont $CD44^+CD25^-$ et $CD3^-$. Elles sont de petite taille et peuvent encore se différencier en lymphocyte T ou lymphocyte B suivant les signaux extérieurs qu'elles reçoivent (Godfrey, Kennedy et al. 1993).

2.2.1.2. Les cellules DN2

Les cellules DN2 ont un phénotype $CD44^+CD25^+$. Elles sont engagées dans la lignée T et produiront majoritairement des cellules $T\gamma\delta$ et $T\alpha\beta$. Cependant, elles gardent encore la capacité de donner des cellules dendritiques, des macrophages et des cellules NK (Balciunaite, Ceredig et al. 2005). Les cellules commencent à réarranger les gènes codant pour les chaînes $TR\gamma$ et $TR\delta$. Elles initient aussi le réarrangement des chaînes $TR\beta$; en effet, le réarrangement de gènes TRDV avec des gènes TRDJ est détectable dans une minorité de thymocytes (Capone, Hockett et al. 1998).

2.2.1.3. Les cellules DN3

Les cellules DN3 sont $CD44^-CD25^+$. A ce stade, les cellules réarrangent majoritairement leur chaîne $TR\beta$ mais elles continuent le réarrangement de leurs chaînes $TR\gamma$ et $TR\delta$ (Capone,

Watrín et al. 1993; Godfrey, Kennedy et al. 1993; Tourigny, Mazel et al. 1997). Elles expriment à la fois le pT α ¹ (pré TR α), CD3e et la chaîne TR β qui sont des composants nécessaires à la formation du complexe pré-TR.

A ce stade les cellules DN3 présentent une hétérogénéité de taille : 70% des cellules sont de petite taille et seulement une minorité d'entre elles expriment la protéine TR β au niveau intracellulaire, tandis qu'elle est exprimée parmi la majorité des 30% de cellules de grande taille restantes. Ces dernières, qui expriment la chaîne invariante pT α associée à la chaîne TR β et au complexe CD3e, sont en transition vers le stade DN4.

C'est lors de cette transition que va pouvoir être exprimé à la surface le complexe pré-TR. L'expression de ce complexe s'accompagne d'un signal intracellulaire provoquant l'arrêt des réarrangements du locus TRB et la survie de la cellule. Les cellules acquièrent la propriété de proliférer fortement (d'un facteur 8 à 10). En absence de ce signal les cellules entrent en apoptose. Ce processus est appelé la sélection β (Wilson, Marechal et al. 2001).

2.2.1.4. Les cellules DN4

A ce stade les cellules sont CD44⁻CD25⁻. Le stade DN4 est le dernier stade double négatif. Le pré-TR est exprimé à la surface et les cellules ont acquis la propriété de proliférer fortement. Si l'on reconstitue les lobes thymiques déplétés en progéniteurs T, il suffit de six jours pour détecter les cellules DP CD4⁺ et CD8⁺. Cela démontre que ces cellules sont au stade double négatif le plus avancé (alors qu'il faut 18 jours pour détecter des DP à partir de cellules au stade DN1). L'expression à la surface de la cellule de la chaîne TR β associée au complexe pré-TR n'est détectable qu'à ce stade de différenciation (Godfrey, Kennedy et al. 1993).

Différence entre la souris et l'homme :

Chez l'homme, la sélection β se produit à un stade plus tardif que chez la souris. En effet, l'expression du pT α débute au stade DN3 avec un pic d'expression au stade DN4. C'est au cours du passage du stade DN4 vers le stade ISP que se produit la sélection β (Dik, Pike-Overzet et al. 2005).

2.2.2. Les cellules ISP

Les thymocytes immatures simples positifs (ISP) expriment modérément le co-récepteur, soit CD4 soit CD8. Ces cellules commencent à réarranger les gènes codant pour la chaîne TR α .

¹ La chaîne pT α est codée par un gène invariant et s'associe avec la chaîne TR β . Le complexe pT α /TR β est nécessaire à la sélection de la chaîne TR β . La chaîne pT α est ensuite remplacée par la chaîne TR α définitive quand cette dernière aura réarrangé à un stade ultérieur.

2.2.3. Les cellules DP

Les thymocytes DP expriment simultanément à leur surface les molécules CD4 et CD8. Ils continuent de réarranger leur chaîne TR α et commencent à exprimer le TR $\alpha\beta$ à leur surface (Pearse, Wu et al. 1989). C'est à ce stade que via leur TR les cellules sont soumises à la sélection positive et négative. Les cellules ayant passé ces étapes de sélection vont pouvoir passer au stade SP.

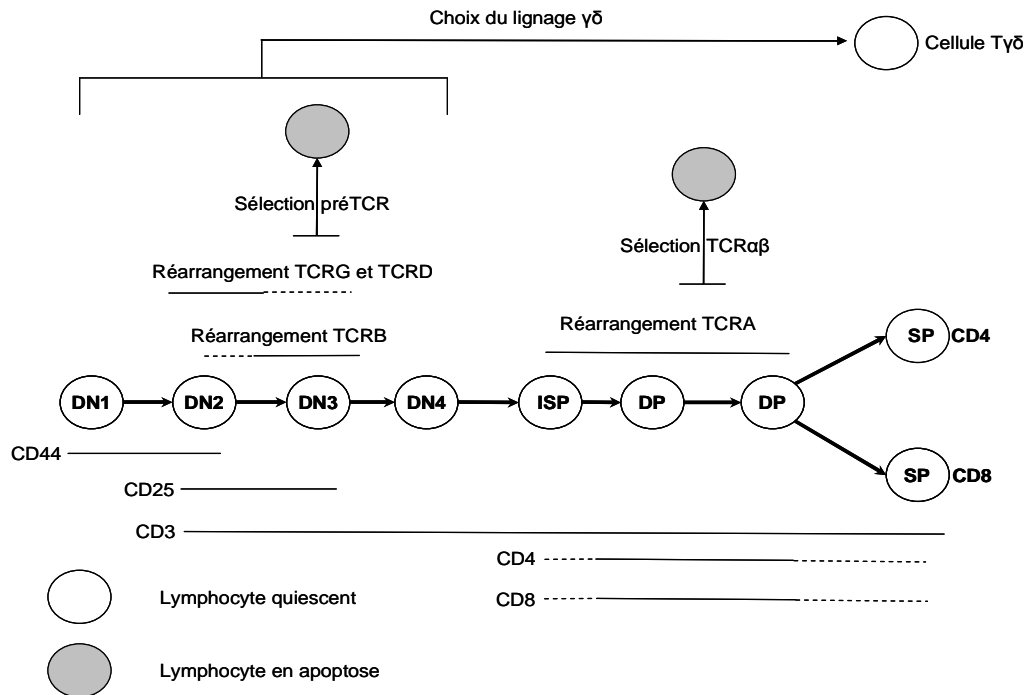


Figure 1 La différenciation intra-thymique des lymphocytes T chez la souris. Les différents stades de maturation et leurs marqueurs sont indiqués. Le réarrangement des gènes est indiqué par des lignes pleines et par des lignes pointillées au stade où ils prennent minoritairement place. Le stade exact de l'engagement des LT dans la lignée $\gamma\delta$ au cours du stade DN est peu clair.

2.2.4. Les cellules SP

Suite aux étapes de sélection positive et négative, les cellules s'engagent dans la lignée CD4 ou CD8, en fonction de la reconnaissance du CMH I ou CMH II par leur TR (Kaye, Hsu et al. 1989). Les cellules entrent en prolifération et effectuent en moyenne six cycles de division (Ceredig 1990; Hare, Wilkinson et al. 1998).

Contrairement à l'adulte, la mise en place du système immunitaire néonatal à la périphérie n'est pas stabilisée et nécessite un apport important en cellules T matures pour constituer le répertoire pré-immun global.

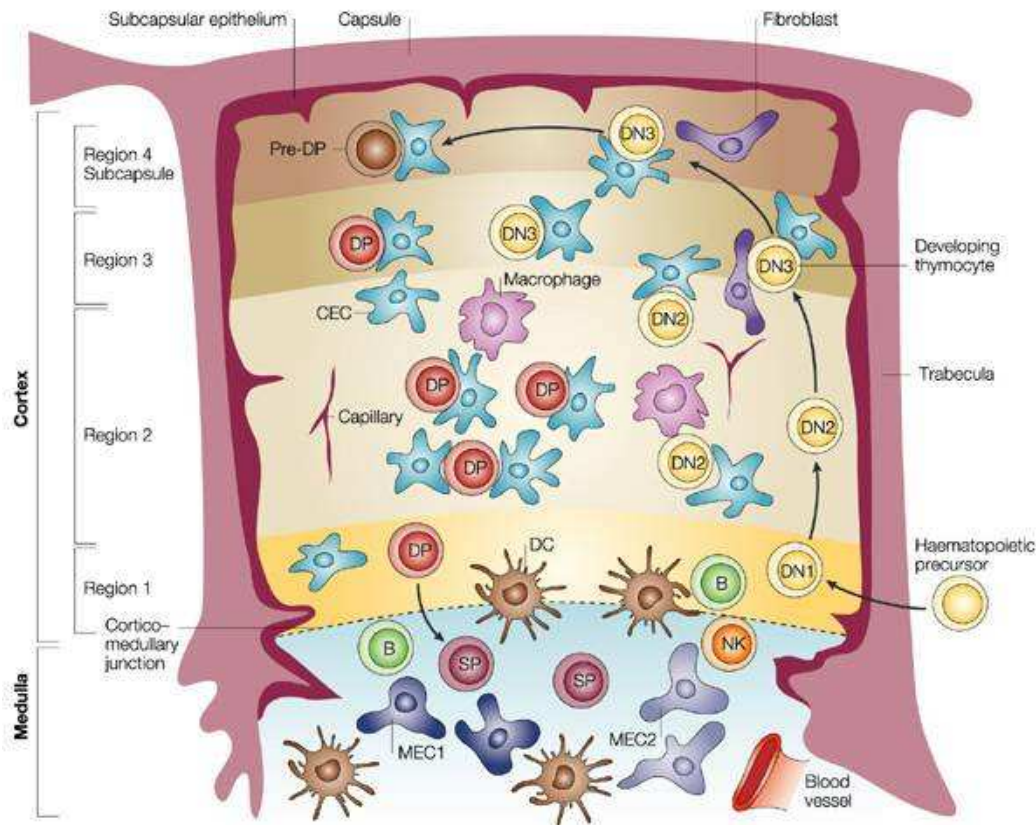


Figure 2 Structure du thymus et cheminement intra-thymique des thymocytes durant leur différenciation.

Le thymus est divisé en deux parties : la médulla et le cortex, qui a été subdivisé en quatre régions. Chaque partie du thymus contient différents sous-types de cellules épithéliales thymiques correspondant aux différents stades de développement des thymocytes. Les cellules souche hématopoïétiques entrent dans le thymus au niveau de la jonction cortico-médullaire. Le stade de différenciation DN (1 à 4) se fait de manière centrifuge, les thymocytes passent de la zone médullaire à la zone corticale. A l'inverse, la transition DN-DP s'accompagne d'une migration centripète, allant du cortex vers la médulla. Les thymocytes subissent la sélection positive par interactions de leur TR avec les molécules CMH de classe I et II exprimées par les cellules corticales de la région 1. Les thymocytes gagnent ensuite la médulla où ils vont subir la sélection négative au contact des cellules présentatrices de l'antigène. Moins de 1% des lymphocytes T survivront à ces deux sélections et gagneront la périphérie. Dans ce schéma, les cellules ISP sont nommées pre-DP. Adapté de (Blackburn and Manley 2004).

2.3.La sélection des lymphocytes T

Chaque individu possède, à la surface de ses cellules, un complexe majeur d'histocompatibilité (CMH) unique. Le CMH présente un antigène du soi ou du non soi à la surface des cellules, et c'est la combinaison CMH + peptide, qui est reconnue par les lymphocytes T.

Afin de gérer au mieux la diversité engendrée par la combinaison CMH + TR, la nature a mis en place un système de sélection des lymphocytes T.

Au stade DP, les thymocytes expriment un TR en surface et interagissent avec le CMH des cellules épithéliales du cortex. Le devenir du TR ne va pas être le même en fonction de sa spécificité au soi. Trois devenirs sont possibles pour le thymocyte : 1) le TR se lie trop faiblement au CMH. Cela indique qu'il ne sera pas capable de reconnaître le "soi", et, faute d'un signal, la cellule enclenche un programme de mort, dit mort par négligence ; 2) le TR se lie de manière modérée au CMH des cellules épithéliales du cortex et/ou des cellules dendritiques thymiques, ce qui permet à la cellule de recevoir un signal et d'être sélectionnée positivement ; 3) le TR se lie de manière trop importante au CMH et aux peptides du soi, ce qui représente un risque pour l'intégrité de la cellule (problème d'auto-immunité). L'organisme ne peut pas prendre le risque de laisser circuler une cellule effectrice, capable de réagir contre les cellules du soi et détruire les tissus sains en périphérie. Ces cellules subissent une sélection négative et sont détruites par apoptose (figure 3).

2.3.1. La sélection positive

La sélection positive a lieu à l'interface cortico-médullaire du thymus. Les thymocytes DP sont encore immatures et expriment à leur surface une faible densité de complexes TR. La sélection se fait sur la quantité de TR exprimés, mais aussi sur la force de liaison du TR au complexe CMH/peptide des cellules corticales. Seules quelques cellules sont capables d'avoir une interaction suffisante pour induire un signal intracellulaire de survie de la cellule. 90% des cellules vont mourir par négligence au cours de cette étape. Les cellules qui auront franchi cette étape expriment plus de TR et de CD5 à leur surface. Le passage de cette étape est aussi marqué par l'expression de CD69 (Anderson, Hare et al. 1999) et par l'arrêt de la transcription des gènes RAG (Kouskoff, Vonesch et al. 1995).

2.3.2. La sélection négative

La migration des cellules dans le thymus est centripète et la sélection négative se déroule dans la médulla et au niveau de la jonction cortico-médullaire. Elle permet d'éliminer les cellules ayant une affinité trop élevée pour le complexe CMH/peptide du soi. Les cellules qui seront

activées par l'activation des peptides du soi à ce niveau seront éliminées par apoptose. A la fin de la sélection négative, les cellules vont exprimer à leur surface le marqueur CD4 ou CD8 suivant la classe du CMH reconnue lors de la sélection. Les cellules progressent ainsi au stade SP.

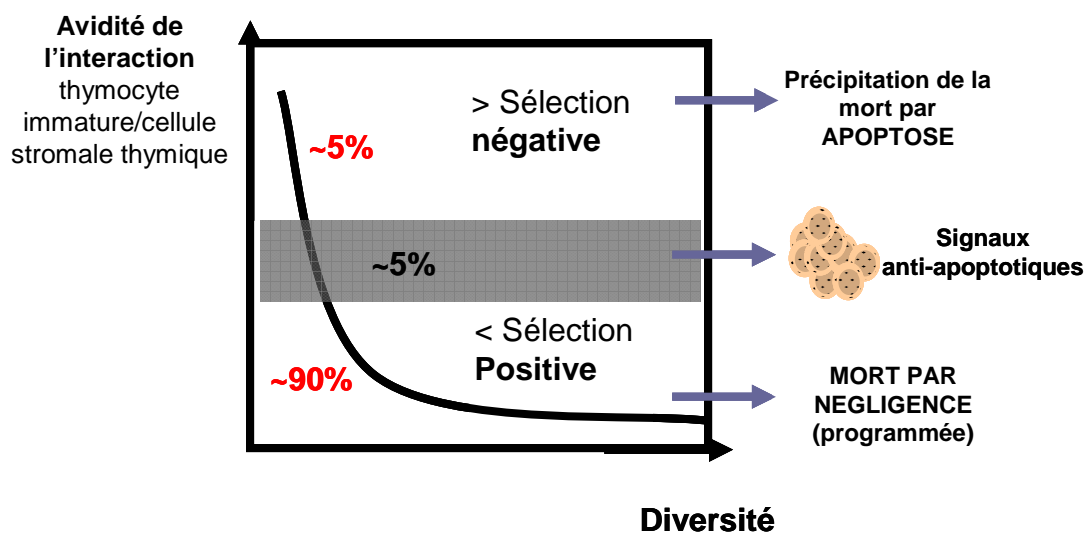


Figure 3 Schématisation du principe de sélections positive et négative des thymocytes. En ordonnée, l'affinité du TR pour le complexe CMH+peptide du soi. En abscisse est représentée la diversité relative du répertoire T. Cette figure est basée à la fois sur la notion de seuil limite d'efficacité du TR, en dessous duquel les cellules sont éliminées par absence de signal de survie et meurent par négligence (sélection positive), et sur la notion de seuil d'intégrité, en dessus duquel les thymocytes sont considérés comme un risque potentiel pour l'organisme et sont éliminés par apoptose (sélection négative). Les pourcentages sont indicatifs de la proportion des cellules sélectionnées. La majorité des cellules sont éliminées au cours de la sélection positive. Environ 5% des thymocytes ont un TR ayant une affinité à la fois suffisante et peu élevée. Ces cellules reçoivent un signal de survie et passent du stade DP au stade SP. Adapté de (Naumov, Naumova et al. 2003).

2.3.3. La sélection β

La sélection β correspond à l'étape de sélection de la chaîne TR β , lors de la formation du complexe pré-TR (composé de la chaîne TR β , du complexe CD3 ϵ et de la chaîne invariante pT α) (voir paragraphe "Les cellules double négatives"). Chez la souris, la sélection β a lieu au passage du stade DN3 vers le stade DN4 (Wilson, Marechal et al. 2001). Chez l'homme, cette étape de sélection se produit à un stade légèrement plus avancé, correspondant au passage du stade DN4 au stade ISP (Dik, Pike-Overzet et al. 2005).

2.4. Le TR

Le TR est essentiel pour la fonction de reconnaissance des peptides antigéniques par les lymphocytes T (Figure 4). Pour appréhender sa complexité et ses propriétés, nous traitons maintenant de l'organisation génomique des gènes codant pour ce récepteur et de sa structure.

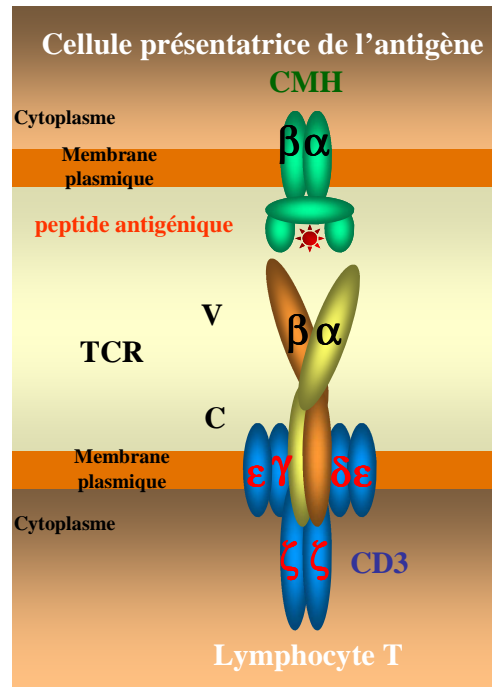


Figure 4 : Représentation schématique d'un TR $\alpha\beta$ associé à un complexe CMH/peptide. La cellule présentatrice de l'antigène (en haut) va présenter à sa surface le peptide antigénique (en rouge) associé au complexe CMH (en vert). Le complexe CMH/peptide antigénique va être reconnu par le lymphocyte T (en bas) à travers son TR $\alpha\beta$ (en jaune pour la chaîne TR α et en orange pour la chaîne TR β) associé au complexe CD3 (en bleu).

2.4.1. Les loci du TR

En position germinale, les gènes codant pour les chaînes du TR sont morcelés et doivent être réarrangés par un processus de recombinaison somatique, la recombinaison V(D)J (voir paragraphe 3.1), pour pouvoir être exprimés. Pour les lymphocytes T, il existe quatre loci différents codant pour les quatre chaînes potentielles du TR ($\alpha\beta$ ou $\gamma\delta$).

2.4.1.1. Nomenclature des gènes

La nomenclature des gènes codant pour les gènes des TR a été modifiée depuis leur découverte. Au départ, ils étaient nommés suivant leur ordre de découverte. Maintenant, ils

sont nommés suivant leur position sur le locus. La nomenclature utilisée dans ce manuscrit est celle proposée par IMGT (<http://imgt.cines.fr/>).

Les gènes codant pour les TR α , TR β , TR γ et TR δ sont notés respectivement TRA, TRB, TRG et TRD.

2.4.1.2. Nomenclature de la chaîne TR α

La dernière nomenclature utilisée (IMGT) ordonne logiquement les familles V et classe de manière plus intuitive les gènes des différents membres d'une famille sur le locus.

Chez l'homme, le gène V1-1 est le plus distal de la région AJ. C'est le gène le plus en amont (du côté 5') de la région V (Figure 5). Plus on se rapproche de la région AJ, plus le numéro de la famille V augmente. Ainsi, le gène V41 est le plus proximal de la région AJ. Les membres d'une même famille ont un nom composé du numéro de la famille suivi du numéro du membre. Par exemple, le gène V8-1 est le plus distal (en 5') des gènes de la famille TRAV8, alors que le gène V8-6 est le plus proximal (en 3') de la région AJ.

Pour la souris, la logique est la même, sauf que la disposition du locus est compliquée par l'apparition de régions dupliquées au cours de l'évolution (Figure 6). Ces gènes sont indiqués par la lettre D. Ainsi le gène TRAV4D4 est l'homologue du gène TRAV4-4. Cette indication permet de savoir que le taux d'homologie entre les régions dupliquées est proche de 100%.

2.4.1.3. Comparaison du locus TRA/TRD entre l'homme et la souris

L'organisation du locus TRA est semblable chez l'homme et la souris. Le locus est situé sur le chromosome 14 pour les deux espèces et s'étend sur environ 1000Kb chez l'homme et 1700Kb chez la souris.

Le locus codant pour la chaîne TR α est le plus complexe car il contient les gènes codant pour la chaîne TR δ . Ainsi, la formation d'une chaîne TR α bloque de manière définitive la possibilité de former une chaîne TR δ .

Le locus TRA contenant les gènes codant les chaînes TR α et TR δ est nommé pour cette raison TRA/TRD. Il est plus complexe chez la souris que chez l'homme, avec 102 gènes V pour la souris contre 54 gènes V pour l'homme. Chez l'homme la majorité des gènes V sont uniques avec 41 familles pour 54 gènes. A l'inverse, chez la souris, la plupart des gènes V sont regroupés par familles selon leur homologie de séquence. Deux gènes V d'une même famille ont une homologie de séquence d'au moins 75%. 23 familles V regroupent 102 gènes, et seuls quelques gènes V sont uniques.

Chez la souris, la différence structurale entre deux régions V est due à des duplications successives, au cours de l'évolution, de petits groupes de gènes V puis à la duplication importante d'un bloc A en un bloc B (Jouvin-Marche, Morgado et al. 1989; Jouvin-Marche, Hue et al. 1990; Gahery-Segard, Jouvin-Marche et al. 1996; Glusman, Rowen et al. 2001).

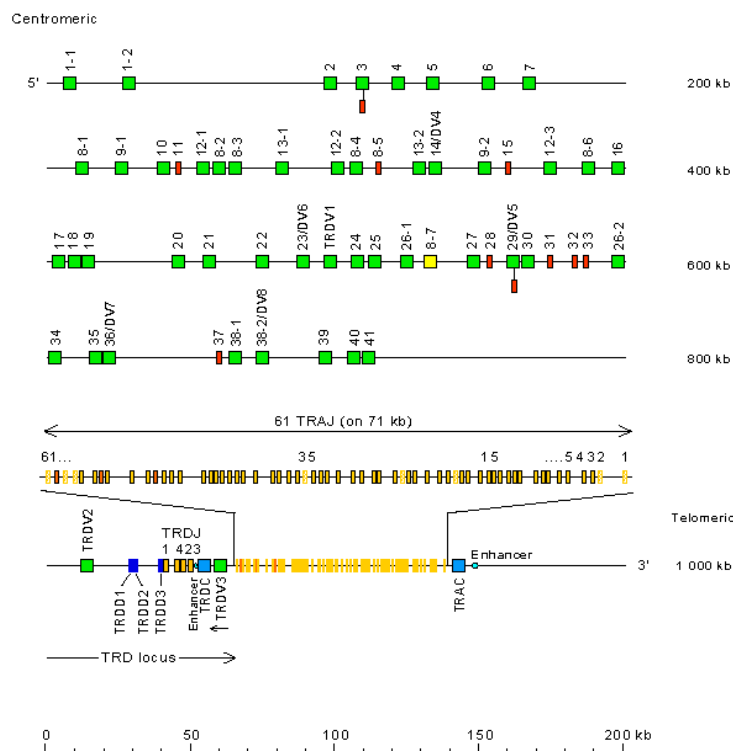


Figure 5 : Carte génétique du locus TRA/TRD humain. Ce locus fait environ 1 Mégabase. Il est composé, de 5' en 3', d'une région AV contenant 54 gènes répartis en 41 familles, d'une région Delta, comprenant 7 gènes DV dont 4 communs avec les gènes AV, 3 gènes DD, 4 gènes DJ et un gène delta constant DC, puis d'une région AJ comprenant 61 gènes, et enfin un alpha constant AC. Pour la région AV, les gènes codants sont représentés en vert, les gènes ORF en jaune, et les pseudogènes en rouge. Pour AJ, les gènes codants sont représentés en jaune foncé, les gènes ORF en jaune clair et les pseudogènes en rouge. Carte d'après IMGT (<http://imgt.cines.fr/>).

La région AJ est très similaire entre l'homme et la souris, avec respectivement 61 et 60 gènes répartis sur environ 60Kb. Certains gènes AJ ne possèdent pas de RSS, de site d'épissage ou de cadre de lecture ouvert correct et sont non fonctionnels. Un certain nombre de ces pseudogènes ont été décrits chez l'homme comme chez la souris. La souris possède 16 AJ pseudogènes AJ (Mancini, Candeias et al. 1999) contre 8 pseudogènes AJ pour l'homme (Fuschiotti, Pasqual et al. 2007). Bien que la région AJ soit similaire entre l'homme et la souris, la région AJ de l'homme présenterait un plus grand nombre de gènes pouvant être reconnus et pouvant être réarrangés avec des gènes AV.

Le locus TRA/TRD contient plusieurs régions régulatrices impliquées dans le contrôle de l'accessibilité à la chromatine. Les sections suivantes présentent les mécanismes contrôlant les réarrangements, les séquences régulatrices associées, ainsi que les différents facteurs et mécanismes de régulation de l'accessibilité des gènes.

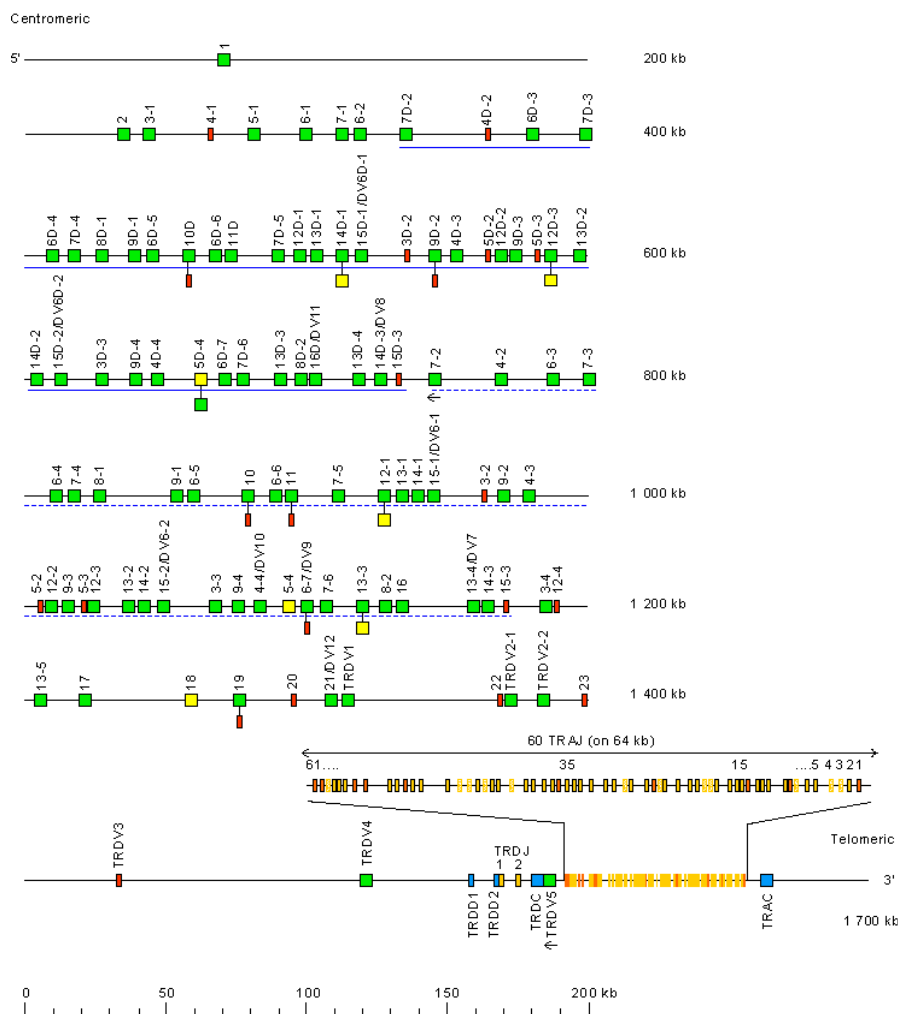


Figure 6 : Carte génétique du locus TRA/TRD murin. Ce locus fait environ 1,7 Mégabases. Il est composé de 5' en 3', d'une région AV contenant 102 gènes répartis en 23 familles, d'une région Delta, comprenant 12 gènes DV dont 4 qui sont commun avec les gènes AV, 2 gènes DD, 2 gènes DJ et un gène delta constant DC, puis d'une région AJ comprenant 60 gènes et, enfin, d'un alpha constant AC. Pour la région AV, en vert sont représentés les gènes codants, en jaune, les gènes ORF et en rouge les pseudogènes. Pour AJ, les gènes codants sont représenté en jaune foncés, les gènes ORF sont représentés en jaune clair et les pseudogènes en rouge. Carte d'après IMGT (<http://imgt.cines.fr/>).

2.4.2. Structure des TR

Chaque récepteur à l'antigène (TR) est composé de deux chaînes polypeptidiques α et β ou γ et δ associées au complexe invariant de transduction du signal CD3 ($\epsilon_2\gamma\delta\zeta_2$) (Figure 7B). Le TR est exprimé de façon clonale à la surface des lymphocytes T (Blumberg, Ley et al. 1990; de la Hera, Muller et al. 1991). Avant d'exprimer ce complexe, les thymocytes en cours de différenciation expriment le complexe CD3 seul puis dans le cadre du pré-TR.

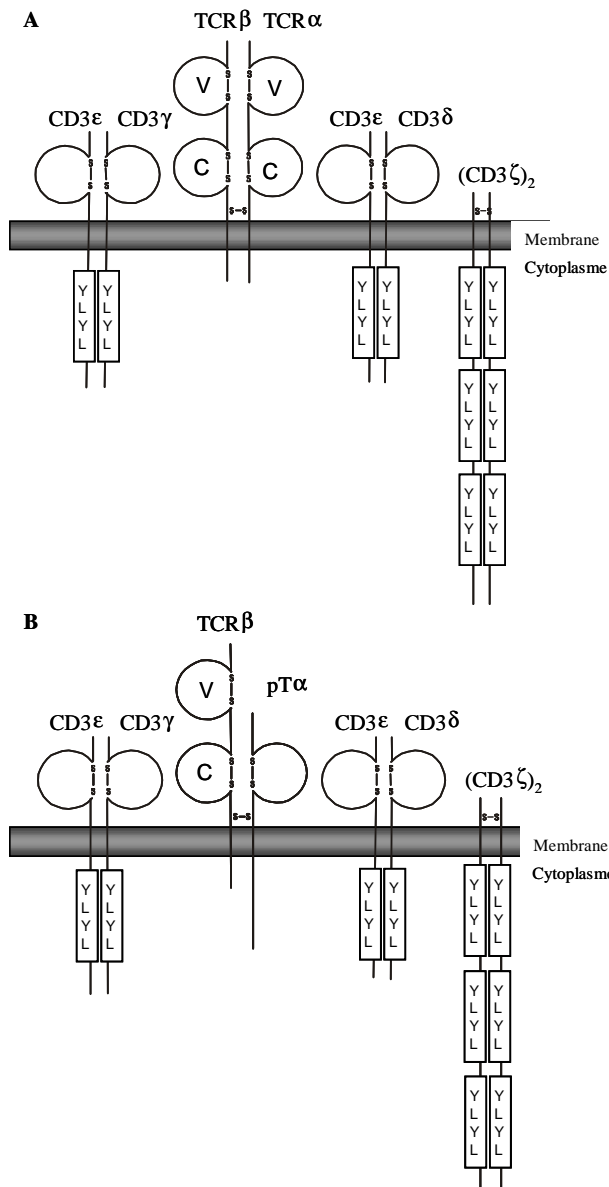


Figure 7 : Représentation schématique des complexes de surface des lymphocytes T. A) TRαβ mature, et B) pré-TR d'un thymocyte au stade pro-T. Les boucles représentent les domaines de type immunoglobuline. Les ponts disulfures sont noté S-S. Les motifs ITAM des chaînes CD3 et leurs résidus tyrosine et lysine sont représentés par des rectangles contenant les lettres YLYL.

Le pré-TR se distingue du TR mature par sa composition. Il consiste en un complexe CD3 ($\epsilon\gamma(\delta)\zeta$), une chaîne invariante pTα et une chaîne clonotypique TRβ (Groettrup, Ungewiss et al. 1993) (Figure 7A). Dans 95% des cas c'est un TRαβ qui est exprimé à la surface des LT. Les chaînes polypeptidiques possèdent un domaine variable V qui assure la reconnaissance du peptide antigénique tandis que le complexe CD3 assure la transduction intracellulaire du signal d'interaction TR/CMH-peptide.

2.4.2.1. Les chaînes du TR

L'hétérodimère $\alpha\beta$ ou $\gamma\delta$ est composé de deux chaînes polypeptidiques reliées entre elles par un pont disulfure. Chacune des deux chaînes est formée de deux domaines en forme de boucle maintenus par des ponts disulfures. Cela donne à l'ensemble une structure de type immunoglobuline.

Le TR est composé d'une large partie extracellulaire en 5' formée de deux domaines de type immunoglobuline : le domaine variable responsable de la reconnaissance de l'antigène et le domaine constant. Le domaine variable des TR possède trois régions hypervariables CDR (Complementary Determining Region). Les régions CDR1 et CDR2 sont codées de façon germinale par les gènes V. La région CDR3 est codée par la juxtaposition des gènes V et J pour les chaînes α et γ et par la juxtaposition des gènes VD et DJ pour les chaînes β et δ .

La région CDR1 interagit directement avec le peptide et le CMH. La région CDR2 interagit avec le CMH et la région CDR3, la plus hypervariable, qui reconnaît directement le peptide (Garcia, Degano et al. 1996) (Figure 8).

La reconnaissance du complexe CMH est assistée par les co-récepteurs invariants CD4 pour les molécules du CMH de classe II et CD8 pour les molécules du CMH de classe I.

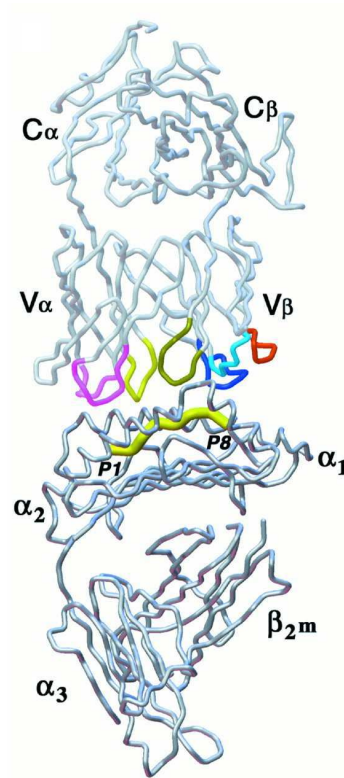


Figure 8 : Structure cristallographique d'un hétérodimère TR $\alpha\beta$ associé à un complexe CMH/peptide. Le TR $\alpha\beta$ se trouve dans la partie supérieure de la figure, les régions CDR des deux chaînes sont indiquées. CDR1 et CDR2 de la chaîne TR α sont indiquées en rose ; CDR1 et CDR2 de la chaîne TR β sont indiquées en bleu ; les CDR3 sont indiquées en jaune. Le complexe CMH/peptide est montré dans la partie inférieure de la figure. Le peptide P1-P8 présent dans la gouttière du CMH est représenté par un tube jaune. D'après (Garcia, Degano et al. 1996)

3. La régulation des réarrangements des lymphocytes T

Pour qu'un lymphocyte soit fonctionnel, il doit exprimer à sa surface un système de reconnaissance spécifique d'un l'antigène. Ce récepteur est mis en place au cours des étapes précoces de la différenciation. Les loci codants pour les chaînes clonotypiques du récepteur TR vont subir des réarrangements pour permettre l'expression d'un récepteur fonctionnel à partir de gènes disposés de manière discontinue sur le génome. Ce mécanisme se nomme la recombinaison V(D)J. La régulation de ce mécanisme est réalisée de manière très précise par la cellule impliquant différents acteurs moléculaires et régulateurs de l'accès à la chromatine.

3.1. La recombinaison V(D)J

La recombinaison V(D)J est un mécanisme commun aux lymphocytes T et B permettant l'assemblage des récepteurs des lymphocytes T (TR) et des immunoglobulines (Ig) à partir des gènes V, D et J préexistants et répartis sur une grande portion germinale au sein des différents loci (Cobb, Oestreich et al. 2006). La recombinaison V(D)J (Figure 9) est basée sur la reconnaissance spécifique de séquences signal de recombinaison (RSS) entre les gènes V(D)J et l'excision de la partie génique intercalée entre les deux gènes qui ont réarrangé.

Chaque gène V, D et J est bordé par une séquence signal de recombinaison (RSS) qui est reconnue par des protéines spécifiques de la recombinaison, RAG1 et RAG2. Les RSS sont constitués de motifs heptamères et nonamères très conservés et séparés par 12 ou 23 pb.

Les protéines RAG, associées aux protéines HMG (High Mobility Group) assemblent des RSS non similaires pour former un complexe synaptique suivant la règle des 12/23, et générer une coupure d'ADN double brin entre les gènes V, D et J et l'heptamère. Cette coupure permet de générer une extrémité codante et une extrémité signal.

Un réarrangement est terminé après la ligation des deux extrémités codantes V et J. Cette étape est précédée par l'action de l'enzyme TdT et d'une nucléase au niveau de la jonction V-J. Une fois réarrangé, le gène néoformé est transcrit puis épissé en ARNm et traduit en protéine membranaire.

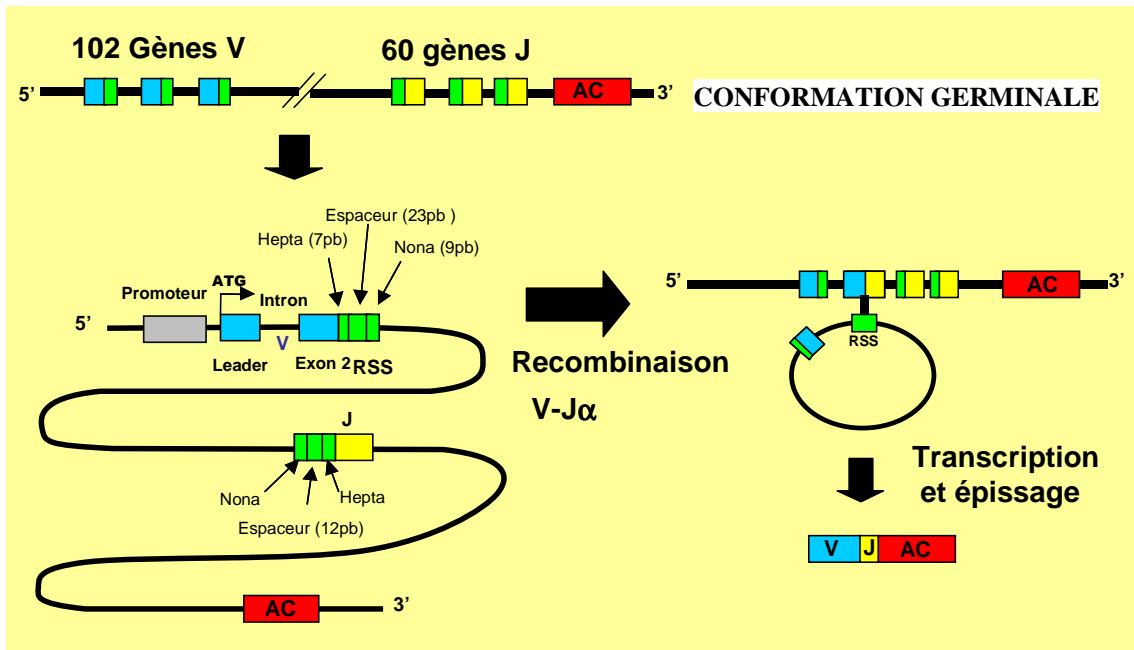


Figure 9 : Représentation schématique du locus TRA/TRD murin. Les gènes V sont représentés en bleu, les gènes J en jaune et les séquences RSS en vert. Un gène V est composé d'un promoteur, d'une séquence leader (exon 1), d'un intron puis d'un exon 2. La séquence RSS des gènes V a la particularité d'avoir un espaceur de 23 pb contrairement aux gènes J qui ont un espaceur de 12 pb. Cette asymétrie de RSS permet de respecter la règle 12/23 qui empêche deux gènes V ayant le même type de RSS de se réarranger entre eux.

3.2. Les RSS et la règle des 12/23

Les RSS sont les substrats géniques de la recombinaison V(D)J. Ces séquences sont constituées d'un heptamère hautement conservé (séquence consensus = CACAGTG) et d'un nonamère moins conservé, riche en AT (séquence consensus = ACAAAAACC). Ces deux éléments sont séparés par une séquence moins conservée de 12 ± 1 pb ou 23 ± 1 pb.

Les RSS sont régis par la règle du 12/23, empêchant ainsi deux gènes du même groupe de recombiner entre eux. Seuls un RSS ayant un espaceur (spacer) de 12pb et un RSS ayant un espaceur de 23pb pourront être reconnus et associés par la recombinaison (Tonegawa 1983).

La séquence des nonamère et heptamère joue un rôle important dans la reconnaissance par la RAG1/2 (Swanson 2002). Les trois premières bases CAC de l'heptamère sont sensibles aux mutations. A l'inverse, des mutations au niveau de l'espaceur, du nonamère ou de la partie 3' de l'heptamère seront moins lourdes de conséquences (Akamatsu, Tsurushita et al. 1994;

Ramsden, Baetz et al. 1994). Cependant, l'ajout ou la délétion de plus d'une base au niveau de l'espaceur réduit fortement la reconnaissance du RSS par les enzymes RAG.

3.3. Les facteurs impliqués dans la recombinaison V(D)J

Les enzymes RAG1 et RAG2 sont les actrices principales de la recombinaison V(D)J. Ce sont elles qui vont reconnaître les régions RSS des gènes V et J et qui vont ainsi pouvoir enclencher la recombinaison. Cependant, elles n'agissent pas seules, mais au sein d'un complexe enzymatique à chacune des étapes du réarrangement (figure 10). Les protéines impliquées dans ce complexe sont les protéines "High Mobility Group" (HMG1 et HMG2), la protéine kinase dépendante de l'ADN (DNAPK) complexée aux hélicases Ku, ou la protéine "X Ray Cross Complementation 4" (XRCC4) associée à la ligase IV, ainsi que les protéines ATM et ARTEMIS et Cernunnos/XLF.

3.3.1. Les enzymes RAG1 et RAG2

Les protéines RAG sont codées par deux gènes contigus chez les vertébrés (Litman, Anderson et al. 1999). Ces protéines reconnaissent les RSS et catalysent le clivage entre la première base de l'heptamère et la séquence codante.

Le promoteur *RAG1* est actif dans les lignées lymphoïde et non lymphoïde chez la souris et chez l'homme, mais l'expression simultanée de *RAG1* et *RAG2* se fait uniquement dans les lignées lymphocytaires B et T.

Cette expression spécifique au lignage se fait sous le contrôle de régions 5' du promoteur *RAG2* sur une distance de -9kb. Pour les cellules B, *RAG2* est sous le contrôle de la région située entre -2kb et -7kb, alors que pour les cellules T, la région s'étend de -7kb à -9kb (Monroe, Sleckman et al. 1999). La spécificité B ou T serait assurée par un double mécanisme de protection de l'intégrité du génome, à la fois : 1) par une activité promotrice différentielle dans ces deux lignées et 2) par un état ouvert/fermé des loci *Ig* ou *TR*.

Les enzymes RAG 1 et RAG 2 jouent un rôle très important dans la recombinaison V(D)J. En effet, la délétion d'une des deux enzymes RAG est suffisante pour supprimer les réarrangements V(D)J (Shinkai, Rathbun et al. 1992). Le complexe RAG1/2 se fixe spécifiquement au nonamère et à l'heptamère du RSS, en association avec les protéines HMG1 et 2. Les enzymes RAG vont cliver l'ADN en amont de l'heptamère, produisant une boucle covalente sur le brin codant et une extrémité franche phosphorylée du côté 5', sur le brin contenant les RSS.

La boucle covalente va ensuite être ré-ouverte de manière asymétrique par les enzymes RAG, permettant ainsi l'addition ou la délétion de nouvelles paires de bases phosphorylées du côté du brin codant. Après le clivage, les enzymes RAG1/2 restent reliées aux brins codant et

signal, au sein d'un complexe composé de protéines ubiquitaires de la réparation, Ku70, Ku80, DNAPK, ainsi que HMG1 et 2.

3.3.2. HMG1 et 2

Les protéines HMG1 et HMG2 sont des protéines ubiquitaires et non spécifiques de la recombinaison V(D)J. En se fixant sur l'ADN au sein du complexe RSS/RAG, elles vont stabiliser ce complexe et amener un changement de conformation de l'ADN, qui permettra aux protéines RAG de réaliser les coupures sur le double brin d'ADN, suivant la règle du 12/23. Il a été montré *in vitro* qu'en cas de manque de HMG1, le 23-RSS est reconnu comme un 12-RSS, et la coupure qui a normalement lieu au niveau 5' de l'heptamère aura lieu 12+7pb du côté du nonamère (Yoshida, Tsuboi et al. 2000). Ces expériences démontrent que les protéines HMG sont nécessaires pour la stabilité du complexe et la fixation de plus grosses protéines, telles que les RAG, en facilitant le déroulement du double brin d'ADN.

Les protéines HMG1 et 2 ont pour rôle : 1) de stabiliser la liaison RSS/RAG ; 2) d'améliorer la présentation des séquences RSS aux enzymes RAG, pour améliorer le clivage de séquences cibles (Yoshida, Tsuboi et al. 2000) ; 3) d'augmenter l'affinité des enzymes RAG pour le nonamère, en favorisant la liaison entre l'heptamère et le nonamère.

3.3.3. ARTEMIS

La protéine ARTEMIS appartient à la famille des métallo- β -lactamases de par son domaine métallo- β -lactamase dans sa région N-terminale (Callebaut, Moshous et al. 2002). De ce fait, elle est activée en présence d'ions Mg^{2+} et inactivée en présence de Mn^{2+} ou Zn^{2+} .

ARTEMIS possède une activité intrinsèque 5'-3' exonucléase sur l'ADN simple brin et les extrémités cohésives (Niewolik, Pannicke et al. 2006), mais acquiert une activité endonucléase lorsqu'elle est complexée et phosphorylée par DNA-PK (Ma, Pannicke et al. 2002). Cette activité endonucléase lui permet d'ouvrir l'épingle à cheveux du brin codant et d'assurer la réparation spécifique de cassures d'ADN (Pannicke, Ma et al. 2004; Cui, Yu et al. 2005; Drouet, Delteil et al. 2005).

La délétion d'ARTEMIS, tout comme la délétion de DNAPK, est moins sévère que la délétion de Ku, XRCC4 ou de la ligase IV (Jeggo and O'Neill 2002). Elle entraîne l'absence de ligation des jonctions codantes et, par conséquent, la "synthèse" d'un ADN réarrangé codant pour une Ig ou pour une chaîne d'un TR, mais ne modifie pas la ligation des jonctions « signal ».

Chez l'homme, une délétion de la protéine ARTEMIS a été décrite chez des patients présentant des symptômes de radiosensibilité proches de ceux observés chez les souris SCID (DNA-PK^{-/-}) (Nicolas, Moshous et al. 1998; Moshous, Callebaut et al. 2001). Il a aussi été

décrit qu'une mutation hypomorphe d'ARTEMIS était accompagnée d'une prédisposition aux lymphomes B, démontrant ainsi l'effet de ce facteur de réparation dans le maintien de l'intégrité du génome (Moshous, Pannetier et al. 2003).

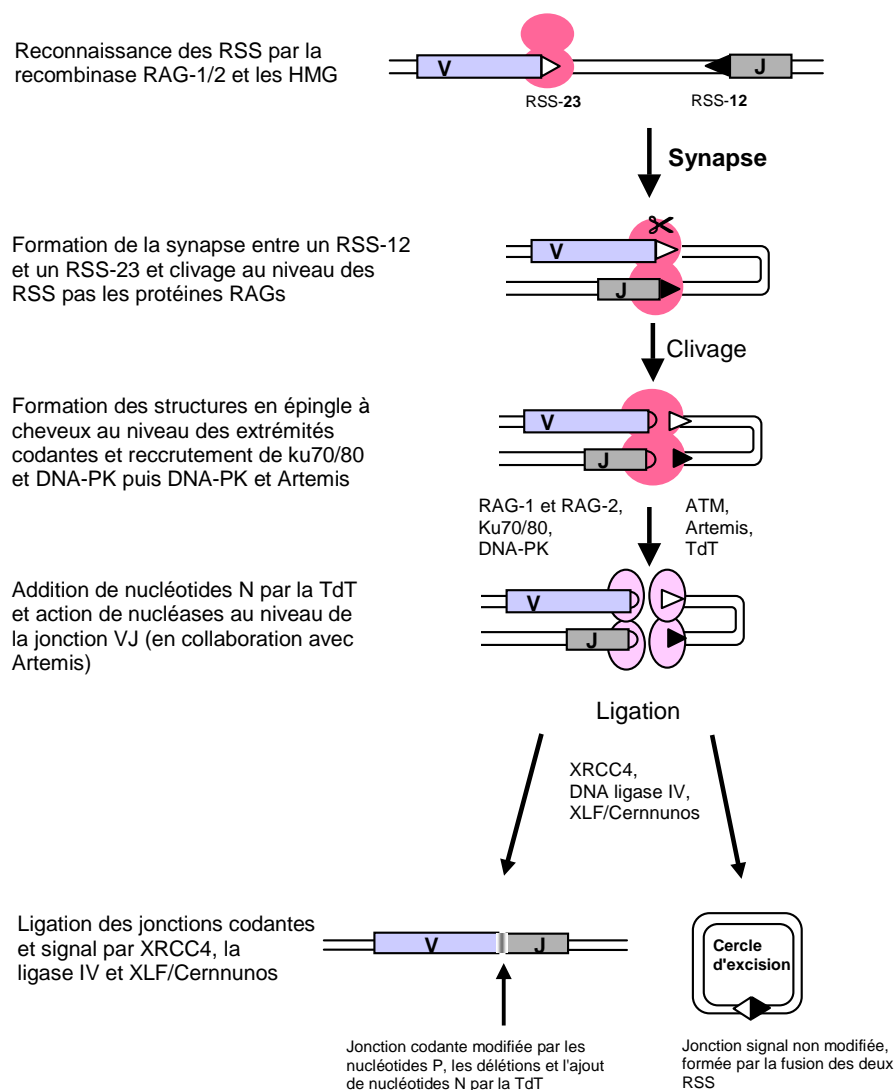


Figure 10 Mécanisme de la recombinaison V(D)J. Ce schéma décrit les principales étapes de la recombinaison V(D)J. Les gènes V et J sont indiqués par des rectangles. Les RSS sont indiqués par des triangles blancs (RSS-23) et noirs (RSS-12). La légende située sur le côté résume les différentes étapes de la recombinaison. Les protéines impliquées sont indiquées entre chacune des étapes. La stœchiométrie exacte des différents membres du complexe n'est pas encore élucidée. Seuls les composants majeurs sont représentés. Adapté de (Krangel 2003).

3.3.4. La protéine kinase ATM

Ataxia-telangiectasia mutated (ATM) est recrutée au niveau des jonctions codantes lors de la recombinaison V(D)J. Son rôle est de participer à la réparation des coupures double brins de l'ADN (Elsa Callen et al. 2007). ATM phosphoryle p53 sur la sérine 15 chez l'homme et la sérine 18 chez la souris (Perkins, Nair et al. 2002). Tout comme les patients dépourvus de cette protéine, les souris *atm*^{-/-} présentent un taux diminué de thymocytes matures et de lymphocytes T périphériques. Ceci indiquerait que le rôle de la protéine ATM est assuré en partie par d'autres protéines, bien que cela traduise aussi un développement intra-thymique défectueux (Barlow, Hirotsune et al. 1996; Elson, Wang et al. 1996; Matei, Gladdy et al. 2007; Vacchio, Olaru et al. 2007).

Cette enzyme est impliquée dans la résolution des intermédiaires de réarrangement pour assurer l'intégrité des recombinaisons dans les cellules B et T.

Les souris déficientes en ATM sont sensibles aux radiations et montrent des prédispositions aux cancers. Elles développent des lymphomes impliquant des translocations au niveau des gènes aux récepteurs des antigènes (TR et Ig) (Petiniot, Weaver et al. 2000). ATM serait donc un facteur enzymatique impliqué dans la protection de l'organisme contre les tumeurs issues de réarrangements aberrants (Perkins, Nair et al. 2002)

3.3.5. DNAPK

Ku70, Ku80 et DNA-PKcs sont les sous-unités de la *DNA dependent protein kinase* (DNA-PK), qui est une enzyme sérine/thréonine protéine kinase nucléaire impliquée dans la réparation des cassures double brin de l'ADN et la recombinaison V(D)J (Gu, Seidl et al. 1997; Downs and Jackson 2004; Drouet, Frit et al. 2006).

Ku70 et Ku80, codées respectivement par XRCC5 et XRCC6, ont une activité hélicase impliquée dans la reconnaissance des dommages de l'ADN. Elles sont recrutées aux extrémités de l'ADN et permettent le recrutement de la protéine DNA-PKcs afin d'activer leur activité kinase. La formation de ce complexe favorise la juxtaposition des extrémités codantes formées par les protéines RAG1 et 2 (Yaneva, Kowalewski et al. 1997; Raval, Kriatchko et al. 2008) et permet la réparation des séquences d'ADN.

DNA-PK pourrait servir de senseur cellulaire de cassures double brin de l'ADN dans le contexte de réparation de cassures double brin d'ADN, incluant celle mise en œuvre par la recombinaison V(D)J (Jeggo, Taccioli et al. 1995; Drouet, Frit et al. 2006).

3.3.6. La protéine XRCC4

La protéine XRCC4 est une des enzymes de réparation de l'ADN. La délétion de cette enzyme est létale dès les stades précoces de l'ontogénèse, indiquant qu'elle possède un rôle

prépondérant (Gao, Sun et al. 1998). Son action au niveau des réarrangements V(D)J s'associe à celles de la protéine DNA ligase IV et de la DNAPK, pour favoriser la ligation des intermédiaires de réarrangement. La délétion de XRCC4 inhibe la résolution des extrémités signal et des extrémités codantes.

3.3.7. La ligase IV

Cette enzyme a pour fonction de relier et unir les extrémités libres de l'ADN. Pour cela elle se lie à XRCC4 par deux motifs enzymatiques de réparation de l'ADN composés de cinq acides aminés BRCAT. Cette liaison protéine-protéine favorise l'activité de résolution des réarrangements. Une défecion en ligase IV provoque une radiosensibilité chez la souris (Barnes, Stamp et al. 1998; Frank, Sekiguchi et al. 1998) et chez l'homme, ainsi qu'un degré d'immunodéficiencie variable (Riballo, Critchlow et al. 1999; O'Driscoll, Cersaletti et al. 2001).

3.3.8. Le complexe XLF/Cernunnos

C'est en 2006, lors de l'étude de facteurs de réparation de cassures double brin de l'ADN, que cette protéine a été découverte (Ahnesorg, Smith et al. 2006). En effet, une lignée cellulaire de patients radiosensibles présentait un défaut de réparation de l'ADN, mais, après avoir vérifié tous les facteurs de réparation connus, il s'est avéré qu'il s'agissait d'un nouveau facteur.

En recherchant de nouvelles protéines pouvant interagir avec XRCC4, il a ainsi été découvert une protéine de 33kDa, le XRCC4-like factor (XLF), encore nommée Cernunnos (Buck, Malivert et al. 2006).

XLF présente une faible homologie avec XRCC4, mais les prédictions lui confèrent un rôle structural similaire. XLF interagit directement avec le complexe XRCC4/ligase IV in vitro et in vivo. De plus, l'action dérégulatrice d'un siARN spécifique de XLF/Cernunnos montre que, si l'on annule l'effet de la XLF, les cellules deviennent radiosensibles et que cela détériore la réparation des cassures double brin d'ADN. Cet effet est réversible, puisqu'il suffit de réintroduire un gène XLF fonctionnel dans des cellules qui ont perdu leur activité XLF pour corriger la radiosensibilité et les défauts de NHEJ. Ceci montre bien le rôle important de XLF dans la réparation des cassures double brin d'ADN.

3.4. Contrôle des réarrangements du locus TRA/TRD

Le locus TRA/TRD est un locus complexe, car il comporte les gènes codant pour la chaîne α et ceux codant pour la chaîne δ . Ainsi les gènes TRDV, TRDD et TRDC sont intercalés entre les gènes TRAV et TRAJ. Un gène variable donné peut se réarranger pour donner, soit une chaîne TR α , soit une chaîne chaîne TR δ . Ce mécanisme est conservé chez la poule (Kubota, Wang et al. 1999), la souris (Chien, Iwashima et al. 1987) et l'homme (Loh, Cwirla et al. 1988).

L'utilisation des gènes du locus évolue au cours de la différenciation du thymocyte. Les gènes codant pour la chaîne δ se réarrangent en premier (Fowlkes and Pardoll 1989), au stade DN2 et DN3 principalement. Contrairement à ce qui se passe sur les locus codant pour les chaînes des IG et des TRb, les réarrangements sont initiés sur les deux chromosomes. Il est probable que les thymocytes au stade DN3, sans avoir réarrangé les gènes codant pour la chaîne delta, initient des réarrangements entre gènes V et gènes J alpha et le plus souvent sur les deux chromosomes, ce qui entraîne une perte définitive pour la cellule de la capacité de produire une chaîne delta. Aussi, deux transcrits sont détectables dans la cellule, montrant que des réarrangements s'effectuent sur chacun des chromosomes en parallèle (Chien, Iwashima et al. 1987; Ito, Bonneville et al. 1989; Jouvin-Marche, Aude-Garcia et al. 1998). Enfin, les gènes AJ détectés dans les réarrangements VJ au cours de l'ontogénie, que soit par l'analyse des réarrangements au niveau génomique, ou celle de transcrits, montre une ouverture progressive du locus des gènes V et J proximaux vers les V et J distaux chez la souris (Rytkonen, Hurwitz et al. 1994; Jouvin-Marche, Aude-Garcia et al. 1998; Mancini, Candeias et al. 1999; Pasqual, Gallagher et al. 2002) et chez l'homme (Fuschiotti, Pasqual et al. 2007).

De ce fait, le mécanisme d'accessibilité des gènes est régulé, les gènes TRD et TRA n'étant pas transcrits au même moment au cours de la maturation des lymphocytes.

3.4.1. Les enhanceurs

Les éléments enhanceurs sont impliqués dans la régulation de la transcription des gènes. Sur les loci codant pour les chaînes TR et Ig, les enhanceurs sont impliqués dans le contrôle de la recombinaison et de la transcription des différents loci (Ho, Yang et al. 1989; Winoto and Baltimore 1989; Bories, Loiseau et al. 1990; Redondo, Hata et al. 1990; Gill, Zaninetta et al. 1991).

De manière générale, la délétion d'un enhanceur dans le cas des Ig ou des TR inhibe la recombinaison et la transcription des gènes qui sont sous son contrôle et perturbe de manière plus ou moins importante la différenciation des cellules B et des thymocytes.

Au niveau du locus TRA/TRD, deux régions régulatrices ont été décrites : l'enhanceur TR α (E α) et l'enhanceur TR δ (E δ) (Bassing, Tillman et al. 2003). Chez la souris, E α se situe en aval du gène TRAC et E δ se situe entre les gènes TRDJ2 et TRDC du locus TRA/TRD (Figure 12). Chez l'homme, E α se situe également en aval du gène AC et E δ se situe entre les gènes TRDJ3 et TRDC. E α et E δ jouent un rôle primordial dans le contrôle lignage spécifique des réarrangements des thymocytes, réarrangements probablement médiés par des différences dans les facteurs cis-activateurs présents (Sleckman, Gorman et al. 1996; Balmelle, Zamarreno et al. 2004).

3.4.1.1. Enhanceur α

Situé en 3' du gène codant pour la région constante de la chaîne TR α , E α est un promoteur puissant dont l'activité couvre plus de 70Kb. Il est spécifique aux lymphocytes T (Ho, Yang et al. 1989; Winoto and Baltimore 1989). Son rôle a été analysé par l'étude de mini locus alpha delta, un substrat artificiel de réarrangement. Avec ce substrat et sous l'influence de E α , la recombinaison V(D)J est détectable à partir de 17 jours de gestation au moment de l'apparition des premiers thymocytes DP et uniquement dans les cellules amenées à présenter un lignage TR $\alpha\beta$ (Lauzurica and Krangel 1994). Ceci correspond au déroulement normal de la différenciation des thymocytes.

E α est impliqué dans l'accessibilité de la région AJ et dans la régulation du taux de transcription des gènes TRA et TRD réarrangés (Sleckman, Bardon et al. 1997). La région centrale de E α de 116pb est nécessaire et suffisante pour initier un réarrangement d'un locus TRA. De fait, un minilocus sous le contrôle de E α initie une recombinaison à partir de 16 jours de gestation, contre 17 pour le gène natif de E α , et uniquement dans les précurseurs TR $\alpha\beta$ (Roberts, Lauzurica et al. 1997).

3.4.1.2. Enhanceur δ

L'activité de E δ dans le contexte du minilocus est détectable à partir de 15 jours gestation, ce qui représente une activité plus précoce que celle de l'E α (Lauzurica and Krangel 1994). Elle n'est cependant nécessaire que dans les cellules immatures. Après l'expression du TR $\gamma\delta$, le réarrangement passe sous le contrôle de E α (Sleckman, Bardon et al. 1997; Hernandez-Munain, Sleckman et al. 1999; Monroe, Sleckman et al. 1999). L'élimination de E δ du locus TRA/TRD chez la souris entraîne la diminution d'un facteur 10 environ du nombre de cellules exprimant le TR $\gamma\delta$ sans affecter le développement des TR $\alpha\beta$ dont le répertoire reste normal et diversifié. La transcription des gènes TRAV-AJ est également normale (Monroe, Sleckman et al. 1999).

3.4.2. Boundary Element Alpha Delta (BEAD)

Devant la faible distance de E δ par rapport au TEA, il a été suggéré l'existence d'un élément "isolateur" alpha delta (BEAD : blocking element alpha delta) empêchant toute action de E δ sur la région AJ (Zhong and Krangel 1997). Les éléments "isolateurs" bloquent l'activité d'un enhanceur pour empêcher l'accessibilité des gènes voisins. Cette fonction est distincte de celle des silenceurs car elle dépend de la localisation entre l'élément enhanceur et le promoteur d'un gène qu'il contrôle, quelle que soit sa distance à l'enhanceur.

Chez l'homme BEAD est située entre TRDV3, à la fin du locus delta, et le TEA (Zhong and Krangel 1997). Chez la souris, BEAD est situé entre E δ et les gènes AJ ce qui restreint l'activité de E δ aux gènes DV, DD, DJ et DC. La séquence BEAD est peu conservée entre l'homme et la souris, et la capacité de blocage de BEAD chez la souris est plus facile. L'inactivation de BEAD chez la souris n'influe ni le développement des thymocytes ni la spécificité temporelle des réarrangements des gènes TRA et TRD (Sleckman, Carabana et al. 2001). D'autres facteurs sont donc certainement mis en jeu pour permettre l'activation de E δ au stade DN sans que son activité affecte la structure chromatinienne de la région AJ.

3.4.3. T Early α (TEA)

L'élément TEA "T Early alpha" est situé en amont de la région AJ. Il se compose d'une séquence conservée à 70% entre l'homme et la souris, et contient plusieurs sites de fixation des facteurs de transcription impliqués dans la régulation des lymphocytes T. Il a été initialement décrit à partir de transcrits germinaux provenant de thymocytes fœtaux humains (de Villartay, Lewis et al. 1987). Les transcrits TEA ont ensuite été mis en évidence chez la souris dans les thymocytes immatures simple positifs (de Villartay and Cohen 1990).

Le TEA est impliqué dans l'accessibilité à la région AJ. Sa délétion dans les souris TEA^{-/-} provoque une diminution importante des transcrits germinaux ainsi que celle de l'utilisation des gènes AJ proximaux entre AJ61 et AJ53, sans affecter la cellularité thymique ni le développement des lymphocytes T $\alpha\beta$ (Villey, Caillol et al. 1996), mais l'accessibilité globale des gènes AJ est très altérée.

Il a été montré que l'activation du TEA durant la transition DN/DP (Wilson, de Villartay et al. 1996) permet une ouverture des gènes AJ à courte distance en influant sur la structure de la chromatine, favorisant ainsi un engagement dans la voie $\alpha\beta$. Cela a été mis en évidence par Mauvieux, qui a montré une inaccessibilité de la région AJ61-AJ53 à la nucléase micrococcale, probablement liée à un défaut d'acétylation de la queue des histones H3 et H4 dans cette région (Mauvieux, Villey et al. 2003).

Le TEA n'est cependant pas indispensable à l'accessibilité de la région AJ car d'autres promoteurs vont prendre le relais. L'équipe de Krangel a ainsi montré que le promoteur TEA permet une ouverture du locus AJ jusque J49, car en l'absence de TEA, seuls les transcrits situés après J49 sont détectés chez la souris TEA^{-/-}. Le promoteur de J49 prend ensuite le relais pour permettre l'accessibilité au reste du locus (Hawwari, Bock et al. 2005). Les autres promoteurs qui entreraient en jeu n'ont pas encore été élucidés.

La question qui reste en suspens est de savoir comment le TEA pourrait réguler l'accessibilité de la recombinaison. L'équipe de Krangel a montré que le TEA génère des transcrits non-codants qui augmentent la distance d'action du TEA. Les chercheurs de la même équipe ont rapporté que les transcrits non-codants du promoteur TEA peuvent activer ou inhiber des promoteurs AJ situés en aval. De plus, la transcription du TEA active les promoteurs AJ situés en position proximale et active l'acétylation et la méthylation des histones favorisant l'accessibilité à la chromatine dans cette région. Au contraire, à une position plus distale, la transcription du TEA inhibe l'activité des promoteurs AJ à travers une interférence transcriptionnelle (figure 11) et inhibe l'accessibilité à la chromatine (Abarrategui and Krangel 2006; Abarrategui and Krangel 2007).

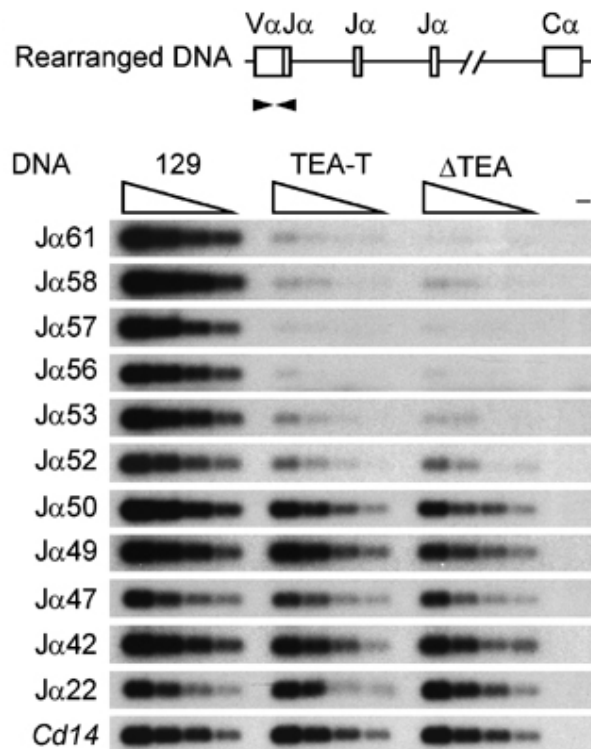


Figure 11 Présentation des réarrangements génomiques $V\alpha J\alpha$ dans des souris TEA-T et Δ TEA. Southern Blot d'amplification par PCR d'ADN génomique des combinaisons de la famille V8 avec différents J. Les produits de PCR sont détectés par sonde J radiomarquée. L'analyse de Cd14 présente un contrôle de charge. (-) désigne un contrôle de PCR sans ADN. En l'absence du TEA (TEA-T ou Δ TEA, les combinaisons VJ avec des J situés entre J61 et J50 sont inhibées. Figure tirée de (Abarrategui and Krangel 2007).

3.4.3.1. TEA et exclusion allélique

Le locus TRA/TRD n'est pas soumis au mécanisme d'exclusion allélique, les deux allèles sont le plus souvent réarrangés. Dans une cellule donnée, il a été montré que deux réarrangements VJ avaient lieu sur des gènes J voisins, mais qu'il n'y avait pas de corrélation du côté des gènes V. Ainsi, l'utilisation des gènes serait synchronisée du côté des gènes J mais pas du côté des gènes V (Malissen, Trucy et al. 1992; Davodeau, Difilippantonio et al. 2001). En l'absence du TEA, les thymocytes perdent le synchronisme de réarrangement VJ qui existait au niveau de la région J (Davodeau, Difilippantonio et al. 2001; Mauvieux, Villey et al. 2001).

Ces résultats montrent l'importance du TEA dans l'initiation des réarrangements du côté 5' de la région AJ.

3.4.4. Locus Control Region (LCR)

Le LCRs (Locus Control Région) fait référence à un groupe d'éléments qui agissent en cis et qui régulent l'accessibilité de la chromatine sur de longues distances. Le LCR du locus TRA/TRD est composé de neuf sites hypersensibles à la DNase I. Ils sont répartis à l'extrémité 3' du locus TRA/TRD sur environ 5kb entre l'exon 2 de la chaîne AC et Dad1, un gène anti-apoptotique et ubiquitaire dont la délétion est létale (Ortiz, Cado et al. 1999). L'étude des différents éléments du LCR a permis de comprendre leur mode de fonctionnement. Ainsi il a été montré que les différents éléments n'avaient pas tous la même fonction. Le site HS1' situé en aval de l'enhanceur alpha ne possède pas de propriété *enhancer-like*, et sert à attribuer la spécificité cellulaire. D'autres, comme HS6, ont été caractérisés pour avoir une activité à la fois ubiquitaire et spécifique de lymphocytes T. D'après Ortiz *et al.* cet élément pourrait jouer un rôle dans la commutation entre l'activité transcriptionnelle de Dad1 et l'activation "tissus spécifique" des réarrangements du locus TRA/TRD (Ortiz, Harrow et al. 2001). Le LCR aurait une activité à distance sur les gènes TR δ et TR α pour induire l'accessibilité de la chromatine. Ainsi, il a été montré que dans les thymocytes, l'activité du LCR est maximale au stade CD4⁺CD8⁺CD25⁺CD44⁻ où débute les réarrangements de la chaîne alpha et diminue au niveau périphérique (Harrow and Ortiz 2005).

3.4.5. "Conserved Sequence Block" de la région AJ

Située sur une région non codante entre les gènes J α 3 et J α 4, la séquence CSB, de 125 pb de long, présente une homologie de 95% entre la souris et l'homme avec seulement six nucléotides de différence (Kuo, Chen et al. 1998). Cette séquence contient plusieurs motifs reconnus par des facteurs de transcription "tissus spécifiques" impliqués dans la formation d'un complexe autour de CSB. Ce complexe protéique évolue au cours de l'ontogénie, et en particulier au cours du passage du stade DN au stade DP. Il a été montré que ces complexes pouvaient interagir avec l'E α , via des protéines telles que GATA-3, Oct1 et Oct2. GATA-3 reconnaît le motif (T/A)GATA(G/A) (Leiden 1993) présent sur les deux séquences régulatrices. Oct1 et Oct2 font partie de la famille des protéines POU. Oct1 est un facteur de transcription ubiquitaire et Oct2 est connue pour son rôle dans la régulation de la transcription des gènes des Ig chez les lymphocytes B (Staudt, Clerc et al. 1988). D'après Kuo *et al.* la fonction de la séquence CSB serait de servir de relais à l'E α pour assurer le contrôle de l'accessibilité de la chromatine sur plus de 70kb de distance (Chen and Kuo 2001).

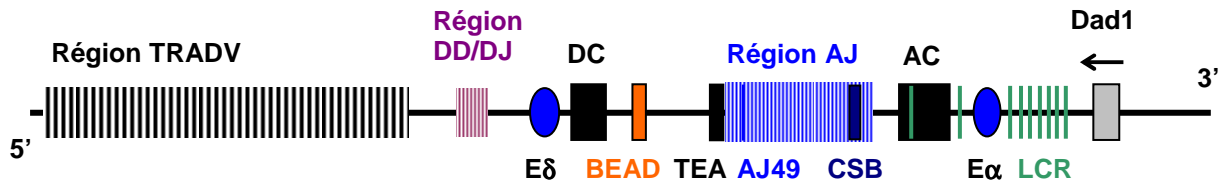


Figure 12 : Organisation du locus TRA/TRD et localisation des séquences régulatrices. De 5' vers 3' : Région TRADV contenant une centaine de gènes V pour la souris et une cinquantaine pour l'humain ; région DD/DJ contenant 2 gènes DD et 2 gènes DJ ; séquence Enhanceur E δ régulant l'accessibilité du locus TRA/TRD sur une distance maximum de 18kb maximum ; DC chaîne constante delta ; séquence isolatrice BEAD empêchant l'activité de E δ sur la région AJ ; promoteur TEA régulant l'accessibilité des 9 premiers gènes de la région 5' de AJ en coordination avec le promoteur AJ49 ; région AJ contenant 41 gènes J ; promoteur AJ49, qui prend le relais du promoteur TEA.

3.4.6. L'accessibilité de la chromatine

L'accessibilité de la chromatine par les complexes enzymatiques de la recombinaison V(D)J est un élément primordial pour que les gènes V, D et J puissent être réarrangés. En dehors des périodes où ils sont transcrits, l'ADN est compacté autour d'un complexe protéique : le nucléosome. La formation de ces superstructures augmente l'état de condensation de l'ADN qui constitue un premier niveau de régulation de l'accessibilité de l'ADN.

Le nucléosome est composé de huit molécules d'histones qui peuvent subir des modifications post-traductionnelles consistant en méthylation, acétylation et ubiquitination. L'ensemble de ces modifications a lieu sur la queue des histones et est appelé le code histone. Ce code histone traduit l'état d'accessibilité de la chromatine en matière de transcription.

3.4.6.1. Méthylation

L'état de méthylation de l'ADN est associé à un état d'inaccessibilité de l'ADN et d'inactivité des gènes. La méthylation de l'ADN a lieu au niveau des cytosines des îlots CpG. Les enzymes responsables sont les cytosines-5-méthyles transférases.

Les gènes fortement transcrits dans un tissu sont hypométhylés, tandis que le même gène se trouvera fortement méthylé dans un tissu où il ne sera pas transcrit. Ainsi, l'étude de la structure de la chromatine associée à un locus Ig ou TR dans un état accessible et actif montre un état d'hypométhylation de l'ADN (Inlay and Xu 2003).

Il a été montré que la méthylation de l'ADN a plusieurs effets. Elle provoquerait une insensibilité à la DNaseI, une la méthylation des histones H3 et une hypoacétylation des histones (Ji, Zhang et al. 2003). Il semble aussi que l'ADN méthylé provoque le recrutement d'histones déacétylases, accroissant la répression de la transcription (Jones, Veenstra et al. 1998).

Au niveau des histones, une méthylation peut être associée à un état d'activation ou d'inactivation transcriptionnelle suivant le résidu méthylé. Par exemple, la diméthylation du résidu lysine 4 de l'histone H3 ou la triméthylation des résidus lysines 4 et 36 de l'histone H3 sont associées à un état d'activation transcriptionnelle, tandis que la méthylation de la lysine 9 de l'histone H3 est associée à un état d'inactivation transcriptionnelle (Jenuwein and Allis 2001; Turner 2002; Grewal and Moazed 2003; Miao and Natarajan 2005).

3.4.6.2. Acétylation

A l'opposé, l'acétylation des histones est associée à un état d'accessibilité à la chromatine. L'acétylation des histones est une modification post-traductionnelle réalisée par les protéines Histone Acétyl Transférase (HAT) qui transfèrent un groupement acétyle sur la lysine, un

acide aminé chargé positivement. Le groupement acétylé étant chargé négativement, il neutralise la charge positive de la lysine et de ce fait perturbe la neutralité de l'ADN. Les charges négatives des phosphates de l'ADN se repoussent et la chromatine devient accessible (Hong, Schroth et al. 1993; Gorisch, Wachsmuth et al. 2005).

Différentes expérimentations ont mis en évidence le rôle activateur de l'acétylation sur l'accessibilité à la chromatine et sur le contrôle des réarrangements (Krangel 2003). *In vitro*, l'inhibition des histones déacétylases induit une augmentation des réarrangements (McBlane and Boyes 2000). Ensuite, l'hyperacétylation des histones H3 sur le mini-locus TR δ ainsi que sur le locus TRA/TRD endogène est corrélée avec le niveau de recombinaison V(D)J. Concernant le locus TRA/TRD, l'hyperacétylation est dépendante de la présence de l'enhanceur alpha (McMurry and Krangel 2000) et du TEA (Abarrategui and Krangel 2007).

La figure 13 illustre l'implication du TEA dans la modification des histones de la région J α . L'acétylation de H3 est associée à un état d'accessibilité de la chromatine. La di-ou triméthylation de H3 K4 et la triméthylation de H3 K36 sont associées à une activité transcriptionnelle. A l'état natif, les histones H3 localisées sur les J61 à J50 sont fortement acétylées ou méthylées sur les résidus K4 et K36. Cela illustre bien que le TEA est associé à une accessibilité des gènes J61 à J50. En effet, la délétion du TEA inhibe l'acétylation de H3, ainsi que la méthylation de H3 K4 et H3 K36 entre J61 et J50, tandis qu'après J50, les résultats s'inversent et montrent une accessibilité des gènes J à partir de J50. Ces résultats illustrent le rôle essentiel du TEA dans l'accessibilité de la région J.

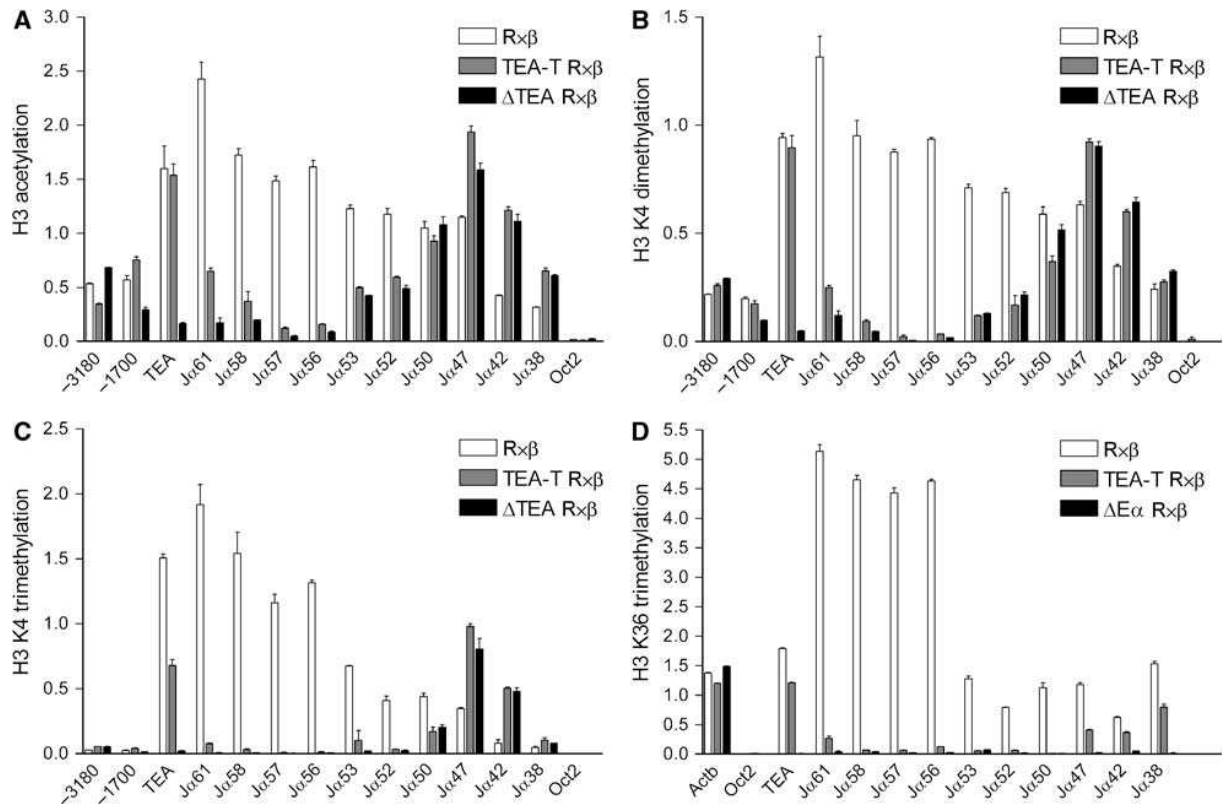


Figure 13 Modifications de la région histone chez les souris Δ TEA. **A)** Analyse de l'acétylation de l'histone H3. **B) et C)** Di et tri-méthylation de la lysine 36 (K36) de l'histone H3. **D)** tri-méthylation de la lysine 36 de l'histone H3. Rxb représente le niveau normal d'acétylation et de méthylation. Oct2 est un témoin négatif, et Actb un témoin positif. Figure tirées de (Abarrategui and Krangel 2007).

3.5. La diversité du répertoire

Un TR doit reconnaître de manière spécifique un nombre limité de peptides antigéniques différents. De ce fait, un large répertoire de récepteurs est requis pour assurer la défense de l'individu contre les pathogènes qu'il est susceptible de rencontrer au cours de sa vie.

Pour cela, le système immunitaire a développé un mécanisme permettant d'assembler des gènes V, D et J positionnés de manière discontinue dans le génome. Ce mécanisme, appelé recombinaison V(D)J est indépendant d'une cellule à l'autre et permet d'obtenir un fragment de gène qui code pour une chaîne TR. Ce système permet, avec un nombre limité de gènes, de générer un grand nombre de récepteurs différents. Chaque cellule utilise une combinaison de gènes selon un mécanisme bien déterminé et obtient une chaîne TR potentiellement unique.

Trois mécanismes majeurs permettent de générer la diversité du répertoire : 1) la diversité combinatoire, qui est générée par les différents mariages possible entre un gène V (un gène D) et un gène J, 2) la diversité jonctionnelle, qui est générée au niveau de la jonction V(D)J et 3) la diversité d'appariement des hétérodimères protéiques $TR\alpha*TR\beta$ ou $TR\gamma*TR\delta$.

3.5.1. La diversité combinatoire

La première étape permettant de générer de la diversité est basée sur le principe de réarrangement V(D)J. Les combinaisons résultantes de l'appariement entre un nombre fixe de gènes V, D et J forment la diversité combinatoire.

Les réarrangements V(D)J ne sont pas aléatoires, mais au contraire dirigés de manière spatio-temporelle (Rytönen, Hurwitz et al. 1996; Jouvin-Marche, Aude-Garcia et al. 1998; Aude-Garcia, Gallagher et al. 2001; Gallagher, Obeid et al. 2001; Pasqual, Gallagher et al. 2002; Fuschioti, Pasqual et al. 2007). Ainsi, un gène V proximal va réarranger de manière préférentielle avec des gènes J proximaux, un gène V distal va se réarranger avec des gènes J distaux et les gènes V et J centraux vont se réarranger entre eux en incluant un peu des gènes proximaux et distaux. Cela rend impossible une estimation précise du répertoire immunitaire par la simple multiplication du nombre de chacun des gènes $nV \times nD \times nJ$.

Même si cette étape ne génère qu'une modeste variabilité de combinaisons possibles par rapport au répertoire maximal estimé à 10^{15} (Davis and Bjorkman 1988), la diversité combinatoire est directement liée au nombre initial de gènes V, D et J initialement disponibles. Les deux autres étapes vont augmenter de manière très importante la diversité du répertoire initial.

3.5.2. La diversité jonctionnelle

La diversité de jonction permet de générer une très grande diversité au niveau de la région CDR3 du récepteur qui est en contact avec le peptide antigénique. Deux mécanismes entrent en jeu. Le premier est dû à l'ajout de nucléotides P (pour palindromiques) issus de la résolution de l'épingle à cheveux par les enzymes RAG1/2 en un point différent de la cassure initiale (Fugmann, Lee et al. 2000).

Cependant cette diversité n'est pas aussi importante que celle générée par le second mécanisme, qui fait intervenir l'enzyme terminale desoxynucléotidyl transférase (TdT). La TdT produit une grande diversité, en ajoutant de manière aléatoire jusqu'à 20 nucléotides N aux extrémités simple brin d'ADN codant, après clivage de l'épingle à cheveux. Des nucléotides peuvent même être éliminés au niveau des jonctions des gènes, par des exonucléases non encore identifiées.

L'ajout ou la délétion de nucléotides peuvent perturber le cadre de lecture dans 2/3 des cas, conduisant à des réarrangements non productifs. Ces réarrangements non productifs ne peuvent pas conduire à une protéine fonctionnelle. Cependant, il existe un mécanisme de réarrangement secondaire permettant à une cellule ayant un réarrangement non productif de tenter un nouveau réarrangement avec les gènes V(D)J encore disponibles au niveau du locus. La propriété d'ouverture concentrique du locus TRA/TRD favorise ce processus en laissant le plus de chances possibles à la cellule, puisque les premiers réarrangements effectués par la cellule ont lieu entre un couple de gènes V et J proches l'un de l'autre (Wang, Huang et al. 1998; Huang and Kanagawa 2001). Ainsi, les réarrangements secondaires permettent la survie d'un grand nombre de cellules qui, à l'issue d'un réarrangement non productif, auraient dû être éliminées.

3.5.3. La diversité d'appariement TR $\alpha\beta$ ou TR $\gamma\delta$.

L'estimation de la diversité d'appariement entre une chaîne TR α et une chaîne TR β est directement obtenue en multipliant le nombre de combinaisons d'une chaîne TR β par le nombre de combinaisons possibles pour la chaîne TR α . La diversité qui en résulte est directement liée au nombre de combinaisons obtenues lors des réarrangements, tout en tenant compte de la diversité jonctionnelle.

4. Modélisation

4.1. Dans quel but ?

Le fonctionnement des organismes vivants est un domaine complexe, dont la compréhension et la maîtrise ne sont pas encore acquises. Afin d'acquérir des connaissances sur le fonctionnement, nous pouvons mettre en œuvre des expérimentations qui nous permettent d'accumuler des données et les connaissances dans un domaine spécifique. Cependant, ces expérimentations peuvent coûter cher en temps et en moyens, et il n'est pas toujours possible de les réaliser. C'est là que l'on trouve l'intérêt de la démarche de modélisation, puisqu'elle va pouvoir se substituer en partie aux expérimentations. D'autre part, une modélisation peut nous aider à comprendre un mécanisme particulier, qui peut s'avérer complexe, tout en s'affranchissant des paramètres superflus et en contrôlant son évolution au niveau temporel. Nous pouvons ainsi construire des hypothèses complexes et les tester, en contrôlant les paramètres mis en jeu.

Dans notre cas, il s'agit de modéliser les réarrangements V(D)J de la chaîne alpha des lymphocytes T. Théoriquement, il pourrait y avoir jusqu'à 6000 recombinaisons pour la souris et 3294 pour l'homme. Cependant, comme il a été montré à plusieurs reprises, les recombinaisons ne sont pas équiprobables, et sont dirigées de manière spatio-temporelle chez la souris comme chez l'homme, avec la présence de hot spots de recombinaisons (Rytönen, Hurwitz et al. 1994; Rytönen, Hurwitz et al. 1996; Pasqual, Gallagher et al. 2002; Fuschiotti, Pasqual et al. 2007).

Pour des raisons de coût et de moyens techniques, il n'est pas possible à l'heure actuelle de tester expérimentalement chacune des combinaisons V-J pouvant avoir lieu, ni de déterminer la fréquence de chaque réarrangement. C'est ici que la mise en place d'une modélisation trouve son intérêt.

D'autre part, la modélisation permet d'évaluer et de tester la valeur de paramètres qu'il n'est pas possible d'appréhender en expérimentation, tels que les vitesses d'ouverture de l'ADN, les temps de maturation entre deux réarrangements ou encore le nombre de réarrangements secondaires, et ainsi de mieux comprendre le fonctionnement de la cellule.

4.2. Etat de l'art de la modélisation en immunologie

La différenciation des lymphocytes T dans le thymus est un phénomène biologique très important au cours duquel les lymphocytes T acquièrent leur fonction de protection de l'organisme.

La modélisation des étapes de différenciation et de sélection, ainsi que de la reconnaissance des antigènes par le système immunitaire est un enjeu important dans la compréhension du

fonctionnement du système immunitaire, mais aussi pour mieux appréhender les interactions des différents types cellulaires qui le composent. Afin de nous aider à mieux comprendre le fonctionnement du système immunitaire, mais aussi pour prévoir comment il peut évoluer, différentes simulations de modèles du système immunitaire ont été réalisées.

D'une manière générale, les simulations de modèles portent sur une étape ou une situation précise et ne cherchent pas à intégrer toutes les connaissances expérimentales sur le domaine. Nous trouvons par exemple des modèles portant sur des domaines très différents, comme la mise en place de la diversité des lymphocytes T, la mise en évidence d'un état de paralysie immunitaire ou encore la reconnaissance des épitopes par les récepteurs des lymphocytes T. *Kaufman et al.* (Kaufman, Andris et al. 1999) proposent ainsi un modèle, dont le but est de comprendre comment un même antigène peut amener, soit à l'activation, soit à l'anergie du système immunitaire. Ils ont montré que le temps, dans la signalisation des différents événements, joue un rôle important sur l'état final du système immunitaire.

D'autres modèles portent sur la mise en place ou l'évolution de la diversité des lymphocytes T. *Buffat et al.* (Buffat and Mary 1992) ont ainsi proposé en 1992 un modèle de recherche automatique composé de quatre modules indépendants, dans le but de modéliser la différenciation des lymphocytes T.

En 2000, à travers un modèle prenant en compte l'état d'un attracteur (point fixe ou cycle limite), *Harada et al.* (Harada and Ikegami 2000) ont étudié l'évolution de la réponse spécifique ou aspécifique du système immunitaire. En 2003, *Wang et al.* (Wang, Krueger et al. 2003) ont proposé un modèle mathématique permettant d'étudier le devenir des lymphocytes T dans le thymus, au cours d'une maturation en situation normale ou en cas d'infection. Ce modèle permet d'étudier les changements intra-thymiques qui apparaissent en cas d'infection par un virus murin. En 2004, la même équipe a proposé un modèle continu (*Wang, Krueger et al. 2004*), pour étudier la différenciation des lymphocytes T et la lymphomagenèse. Ce modèle permet d'étudier les changements du micro-environnement intra-thymique au cours de la maturation des lymphocytes T. Des travaux concernent aussi le développement de la diversité des lymphocytes T $\gamma\delta$ (*Boucontet, Sepulveda et al. 2005*; *Sepulveda, Boucontet et al. 2005*) Des travaux plus récents proposent d'étudier la cinétique de l'activation des lymphocytes T (*Bidot 2006*; *Bidot, Gruy et al. 2008*).

Ces quelques exemples illustrent la grande variété des domaines dans lesquels la modélisation peut nous aider à comprendre les mécanismes et prévoir l'évolution en immunologie.

4.3. Les différentes manières d'aborder la modélisation

Il existe différentes manières d'aborder une modélisation. Une d'elles consiste à traduire le phénomène observé à travers une équation, souvent complexe. C'est ce que nous appelons le modèle mathématique.

Une autre façon d'aborder la modélisation consiste à poser des hypothèses sur le mécanisme que l'on souhaite étudier et à tester ces différentes hypothèses à travers une programmation informatique. C'est ce que nous appelons le modèle mécanistique. Cette approche est parfois appelée "individu-centrée", "multi-agents" ou "événementielle", selon le contexte.

Nous avons choisi cette seconde approche pour aborder la modélisation et la simulation du modèle obtenu, du fait d'une approche complémentaire, proche de la biologie et de ses mécanismes, mise en oeuvre dans nos laboratoires.

4.4. Les pièges à éviter

La modélisation et la simulation d'événements biologiques peuvent être difficiles à mettre en place, et il est facile de "se perdre" dans l'ensemble des paramètres que l'on décide de mettre en jeu, ainsi que dans leur régulation.

Il est en effet possible de vouloir "coller" absolument aux données expérimentales, par exemple en augmentant le nombre de paramètres ou en mettant des contraintes trop précises. Ces modèles ne seront alors applicables que pour un mécanisme très précis et ne pourront pas s'adapter à un changement de la situation dans laquelle ont été observées les données expérimentales.

Dans l'élaboration de notre modèle, nous avons veillé à garder le cadre le plus général possible. Nous avons aussi été attentifs à donner une explication à chaque mécanisme mis en jeu. Cela nous permet d'avoir un modèle qui reste général pour le mécanisme de la recombinaison V(D)J chez la souris, et qui pourrait être adapté de manière simple à d'autres organismes que le modèle murin. Nous avons mis en place un modèle qui réussit à être précis, tout en restant suffisamment flexible, pour rester pertinent dans une situation similaire.

4.5. Le modèle existant de recombinaison V(D)J

4.5.1. Le modèle du groupe de Warmflash

Les technologies actuelles ne permettent pas d'étudier de manière précise toutes les recombinaisons VJ fonctionnelles, ni d'estimer leur fréquence. Pour pallier ce manque, l'équipe américaine de Warmflash *et al.* (Warmflash and Dinner 2006) a proposé un modèle mathématique des réarrangements V(D)J.

Ce modèle est fondé sur un ensemble de paramètres qui décrit la fenêtre d'accessibilité des gènes V et J à un moment donné. Ce modèle propose une couverture du répertoire immunitaire murin, dans laquelle 4% des combinaisons VJ sont peu probables et, dans les 96% restants, 70% sont équiprobables.

Ce premier modèle sur les réarrangements du locus TRAD suit les données observées pour les réarrangements proximaux et distaux, mais n'est pas en adéquation avec des observations faites au niveau central du locus (Rytönen, Hurwitz et al. 1994; Rytönen, Hurwitz et al. 1996; Jouvin-Marche, Aude-Garcia et al. 1998; Aude-Garcia, Gallagher et al. 2001; Pasqual, Gallagher et al. 2002). Actuellement, on sait que les gènes J sont sélectionnés de manière séquentielle, mais leur utilisation n'est pas unimodale. Cela est mis en évidence par la présence de hot spots de recombinaison observés au niveau central du locus (Rytönen, Hurwitz et al. 1994; Villey, Caillol et al. 1996; Jouvin-Marche, Aude-Garcia et al. 1998; Abarrategui and Krangel 2007). Le modèle du groupe de Warmflash ne rend ainsi pas compte des hot spots de recombinaison.

4.5.2. Notre modèle

Le modèle de recombinaisons V(D)J proposé par l'équipe de Warmflash a été publié avant le nôtre. Cependant, notre travail a été mené indépendamment du leur. Le travail concernant la modélisation des réarrangements V(D)J avait été entrepris, dès 2002, à travers la collaboration entre les laboratoires TIMC et IAPC. Un premier modèle de simulation, non publié, a été réalisé en Matlab, mais il ne rendait pas compte des hot spots de recombinaison. Fin 2005, nous avons entrepris l'implantation d'un nouveau modèle qui s'appuyait sur les principes du modèle précédent, mais qui introduisait de nouveaux éléments dans la simulation, afin d'être plus proche du modèle biologique. La programmation dans un langage approprié (Python) a permis de tester rapidement différentes hypothèses. Par ailleurs, la durée réduite de chaque simulation (quelques minutes contre quelques heures pour la version Matlab) a permis de manière plus aisée les ajustements nécessaires des paramètres retenus.

L'objectif principal de notre travail était de trouver les hypothèses permettant de modéliser le mécanisme de recombinaison V(D)J. Avant de débiter la modélisation, deux théories sur l'accessibilité de l'ADN ont été testées. La première théorie est fondée sur la théorie du cliquet brownien, *the brownian ratchet model*. Elle utilise la théorie du mouvement aléatoire pour simuler les recombinaisons V(D)J. Bien qu'il ait été montré que la recombinaison V(D)J n'était pas aléatoire, cette analyse nous a permis de confirmer que la recombinaison V(D)J ne suivait pas un mécanisme fondé sur un mouvement aléatoire.

Nous avons ensuite testé une théorie d'accessibilité par fenêtrages successifs, ou *sequential windowing model* (figure 14). Dans cette hypothèse, les régions V et J se trouvent accessibles

suivant une fenêtre particulière qui dépend de paramètres à définir. Cette théorie permet la simulation de chaînes TR, dont les combinaisons sont dépendantes de la position initiale des gènes V et J sur le locus TRA/TRD. Par ailleurs, cette hypothèse est en accord avec les données expérimentales montrant que l'accessibilité des régions V et J était ordonnée séquentiellement.

Nous avons gardé cette théorie de départ pour le modèle. La réflexion nous a ensuite conduits à émettre des hypothèses concernant les paramètres qui peuvent intervenir dans la mise en place de ces fenêtres d'accessibilité.

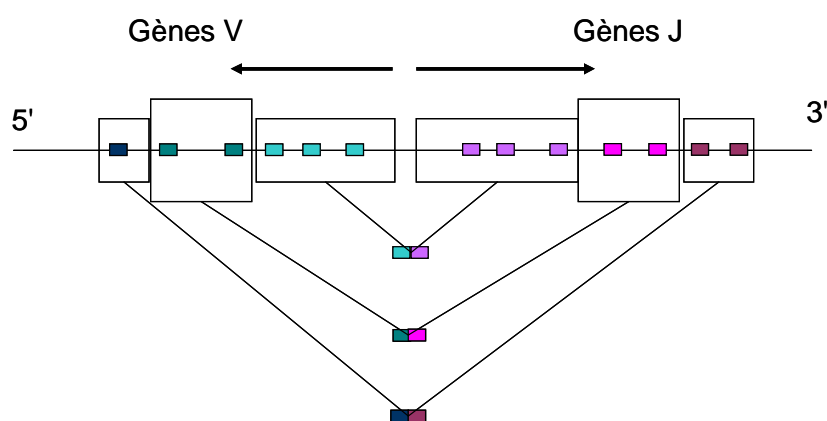


Figure 14 Représentation schématique de la théorie d'ouverture séquentielle. Les gènes V sont représentés par des rectangles verts et les gènes J par des rectangles violets. Les rectangles transparents (regroupant plusieurs gènes V ou J) représentent les fenêtres d'accessibilité des régions V et J. Les flèches représentent la direction de l'ouverture. Un réarrangement ne pourra se faire qu'entre les gènes V et J présents dans une même fenêtre d'accessibilité.

De manière générale, les paramètres pouvant intervenir dans la mise en place des fenêtres d'accessibilité sont la position de départ du mécanisme et la vitesse d'accessibilité à ces régions. La position des gènes V et J le long du locus permet ensuite de déterminer les gènes V et J qui auront la possibilité d'engager un réarrangement.

Nous avons débuté notre modèle en comparant de manière systématique des résultats obtenus par simulation avec des données expérimentales concernant l'utilisation des gènes V et J au cours de l'ontogénie (Aude-Garcia, Gallagher et al. 2001; Pasqual, Gallagher et al. 2002). Par la suite, les résultats de simulation ont été confirmés par les travaux d'autres équipes. Cela concerne notamment la position des hot spots de recombinaison (Rytkonen, Hurwitz et al. 1994) et l'importance du TEA et du promoteur J49 observé par l'équipe de Krangel (Hawwari,

Bock et al. 2005; Jackson, Kondilis et al. 2005; Abarrategui and Krangel 2006; Abarrategui and Krangel 2007).

5. Objectifs

De part leur fonction, les TR sont des éléments clés du système immunitaire. C'est par l'intermédiaire de ces récepteurs que l'organisme se défend contre les agressions extérieures et veille à l'intégrité du soi (élimination de toutes les cellules anormales). La connaissance des règles régissant la mise en place du répertoire $TR\alpha\beta$ et l'estimation de sa diversité sont capitales pour comprendre comment notre organisme assure défense et veille. Jusqu'à présent, l'évaluation du répertoire $TR\alpha\beta$ s'appuyait sur l'analyse de la chaîne $TR\beta$. L'étude de la chaîne $TR\alpha$ en termes de diversité n'est que partiellement élucidée. Ceci est dû en partie à la complexité du locus TRA/TRD, qui comprend les gènes codant pour la chaîne delta et des séquences dupliquées dans la région des gènes V. Les analyses de diversité de la chaîne $TR\beta$ ne donnent qu'une vision partielle du répertoire $TR\alpha\beta$. Plusieurs rapports, notamment d'analyse de complexe TR-CMH-peptide, sont convergents et indiquent que la chaîne $TR\alpha$ est en contact avec des résidus du CMH, voire du peptide, et par conséquent joue un rôle actif dans la reconnaissance du complexe.

La complexité du locus TRA/TRD rend impossible une évaluation expérimentale précise du répertoire complet de la chaîne $TR\alpha$. Afin de pallier ce manque, l'objectif principal de mes travaux a été la mise en place d'une modélisation des réarrangements V(D)J du locus TRA/TRD murin (Article 1), afin de rendre compte de la diversité du répertoire $TR\alpha$.

Par une comparaison systématique des résultats de simulation avec des données expérimentales, nous avons développé, puis validé, un modèle et une simulation des réarrangements des TR alpha chez la souris. Cette approche assistée par ordinateur a permis une estimation fine de la diversité totale des TR générée dans le thymus, en testant des paramètres pouvant influencer le répertoire. Cela nous a permis d'estimer les fréquences de chaque combinaison VJ (Article 2).

Nous avons ensuite réalisé une étude exhaustive des réarrangements VJ chez l'homme, avec l'objectif d'adapter le modèle de simulation murin à l'homme (Partie 3).

6. Article 1

Modeling the Rearrangements of Genes encoding Immune Receptors. Toward a Prediction Tool of Immune Specificity

Maria-Ana Simonet, Olivier Hansen, Florence Thuderoz, Evelyne Jouvin-Marche, Patrice N. Marche
and Jacques Demongeot

Article accepté à CISIS 2009 (Japon, 16-19 mars 2009)

Les lymphocytes T expriment à leur surface des récepteurs ($\alpha\beta$ ou $\gamma\delta$), dont la fonction est de défendre l'organisme. La particularité de ces récepteurs est que les gènes qui codent pour chacune des protéines doivent se réarranger, avant d'être transcrits, puis traduits. Pour la chaîne alpha, l'étape de réarrangement somatique conduit à la juxtaposition d'un gène V et d'un gène J, l'exon VJ nouvellement formé est ensuite transcrit et épissé sur l'exon codant pour la région constante. Le nombre de gènes codant pour la chaîne alpha est connu depuis le séquençage complet du génome. Les gènes codant pour la chaîne alpha appartiennent au locus TRA/TRD situé sur le chromosome 14. Ce locus est complexe, car les gènes qui codent pour la chaîne delta sont intercalés entre les gènes codant pour les régions $V\alpha$ et $J\alpha$ de la chaîne alpha.

Devant la complexité de la région V chez la souris, nous avons décidé de développer un modèle des réarrangements VJ de la chaîne $TR\alpha$. En effet, l'analyse de toute cette région n'est pas possible expérimentalement. Avec le support d'un modèle et de simulations de celui-ci, nous avons pu estimer la diversité du répertoire, estimer les fréquences des combinaisons VJ et tester différentes hypothèses dans la régulation des réarrangements du locus TRA/TRD murin.

Nous présentons ici un travail de synthèse montrant les différentes étapes dans l'élaboration du modèle de recombinaison des TR. Nous montrons comment un modèle de réarrangements aléatoires (Bawnian Ratchet Model) ne permet pas d'expliquer les données expérimentales, tandis qu'un mécanisme fondé sur une ouverture du locus TRA/TRD par fenêtrage successif permet d'aboutir à un modèle en accord avec les données expérimentales (figure 15).

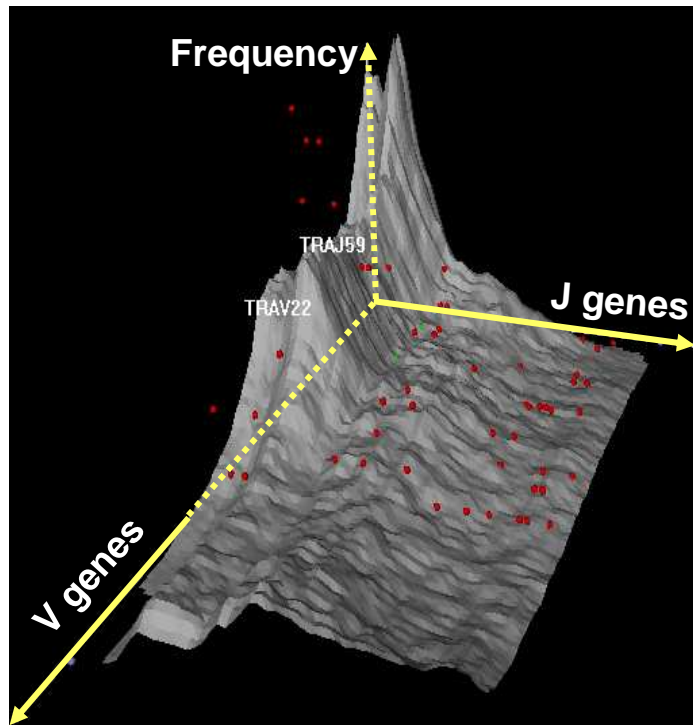


Figure 15 Superposition des données expérimentales et des données issues de la simulation. Les points rouges représentent les points expérimentaux qui se trouvent au dessus des données de simulation (en gris).

Sur les 4900 combinaisons testées, le modèle propose que 96% des combinaisons sont possibles, et seulement 4% sont très improbables. Ces résultats sont aussi en accord avec les données de la littérature et montrent la richesse du répertoire combinatoire VJ.

La validation du modèle est plus longuement explicitée dans l'article 2.

Modeling the Rearrangements of Genes encoding Immune Receptors. Toward a Prediction Tool of Immune Specificity

Maria-Ana Simonet*, Olivier Hansen*, Florence Thuderoz, Evelyne Jouvin-Marche, Patrice N. Marche and Jacques Demongeot

Abstract — This paper describes a general random mechanism of vanishing the central part of a concatenated chain made of characters in an alphabet of cardinal 4. This mechanism allows to obtain chains shorter than the initial one with some constraint for the start and the end of the resulting chain. An application in immunology of this chain shortening procedure will be proposed, describing the genesis of new functional genes from the TCRAD locus on the chromosome 14 in mice and humans encoding immune receptors.

Index Terms — T-Cell Receptor (TCR), V(D)J rearrangements, characters chains shortening, random windowing, mixture of probability distributions, immunologic repertoire

I. INTRODUCTION

THIS paper deals with the modeling of rearrangements of genes encoding immune proteins corresponding to highly variable receptors of antigens. This genetic mechanism is specific of the immune cells, the lymphocytes, and is able to do loops in DNA which are excised to assemble together one V and one J genes together by a specific enzymatic mechanism named V(D)J recombination. This is one of the only examples of DNA rearrangement creating functional “neo-genes” which is shared by genes of immunoglobulin

Manuscript received October 31, 2008. This work has been supported by a research grant from the Rhône-Alpes Region actions in the Clusters 10 and 11, by the European Community Network of Excellence (EC NoE) VPH (Virtual Physiological Human), and in part by the Biotherapeutic program (R06002CC) of the “Pole de Compétitivité Lyon Biopole” and of the Direction Générale des Entreprises, France.

M. A. Simonet is with Laboratories TIMC-IMAG, UMR CNRS 5525 and IAPC, INSERM U 823, Team 8, University J. Fourier, Faculty of Medicine, 38700 La Tronche France (e-mail: Maria-Ana.Simonet@imag.fr).

F. Thuderoz and O. Hansen are with the Laboratory TIMC-IMAG, UMR CNRS 5525 University J. Fourier, Faculty of Medicine, 38700 La Tronche France (e-mail: Olivier.Hansen@imag.fr).

E. Jouvin-Marche and P. Marche are with the Laboratory IAPC, INSERM U 823, Institute Albert Bonniot, University J. Fourier, Faculty of Medicine, 38700 La Tronche France (e-mail: Evelyne.Marche@cirs-dir.fr; patrice.marche@ujf-grenoble.fr).

J. Demongeot is with the Laboratory TIMC-IMAG, UMR CNRS 5525 University J. Fourier, Faculty of Medicine, 38700 La Tronche France (corresponding author: 33-456-520108; fax: 33-476-768844; e-mail: Jacques.Demongeot@imag.fr).

* MAS and OH have contributed equally to this work.

family in vertebrates. In mammals, there are four families of genes encoding T Cell Receptor (TCR), designated α , β , γ and δ , which build the immune repertoire of T lymphocytes. Data used for building and falsifying our model concerns the TCRAD locus (T Cell Receptor Alpha and Delta) of the chromosome 14, which encompasses genes encoding TCR α and δ chains, and come from experimentations in mice [1-5]. Many previous studies (from the seminal paper by N. K. Jerne [6] to the most recent works [7]) have mainly concerned the dynamics of the activation response of the lymphocyte T (stationary or periodic behavior [8], hyper-response [9], immune paralysis [10], specificity of T-cells activation [11],...). We will deal in this paper with the variability of the response based on the potential richness of the TCR repertoire. In a first Section, we will show examples of empirical probability distributions of the rearranged genes from the start of a V gene concatenated with the end of a J gene. Then we will give a realistic model based on a simple progressive windowing mechanism accounting for these data and discuss the model relevance by confronting simulated results to experimental data. Eventually, we will draw some perspectives of application of this methodology in human.

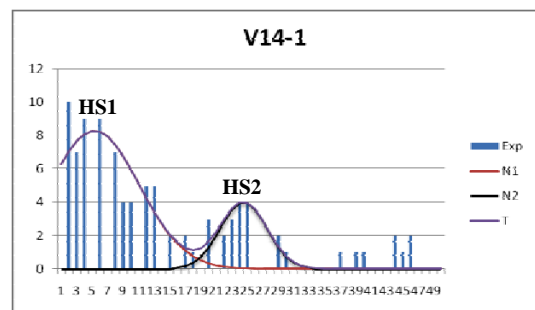


Fig. 1. Empirical probability distributions experimentally observed for rearrangements between the V 14-1 gene and J genes indexed from 1 (start of the J region) to 49 (end of the J region). The blue bars represent the experimental frequencies of rearrangement from Aude-Garcia *et al.* [12], the normal distribution curves N1 and N2, with two modes HS1 and HS2, represent the probability distributions whose mixture (curve T) fits the best these data

II. PROBABILITY DISTRIBUTIONS FOR THE LENGTH OF REARRANGED CHAINS

By using classical techniques like cloning or quantitative PCR, it is possible for a given mice thymocytes to get statistics about the number of TCR rearrangements from the i^{th} V gene of the TCRAD locus (like the 14th, denoted V14-1 on Figure 1), chosen among V located close to the J region, termed proximal V, with the j^{th} gene of the J region (j varying from 1 to 49, numbered from the most distal J of the V region to the most proximal).

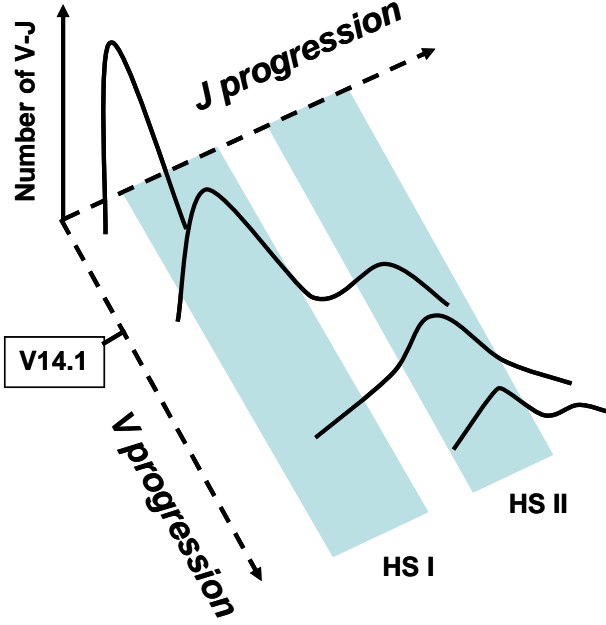


Fig. 2. Family of empirical probability distributions experimentally observed, corresponding to concatenation of the start of a V gene (top left) with the end of a J gene (bottom right).

By repeating the observation of rearrangements for each V gene [1], we obtain the whole set of the possible rearrangements; their empirical frequencies can be exploited in order to draw a 2d-histogram over the plane (i,j) . This histogram shows 3 main features:

- a Poisson like distribution for proximal V genes over the J ones
- a bimodal (Gaussian mixture) distribution for central V genes over the J ones, with two frequencies peaks called Hot Spots (HS1 and HS2 on Figures 1 and 2)
- a Gaussian like distribution for distal V genes.

III. MODELING THE WINDOWING MECHANISM

A. Brownian ratchet

The first step in studying V(D)J rearrangements is to see whether we can find a law as simple as possible to predict them, such as a probabilistic law, related to a pure random mechanism. When we plot, for example, V genes against J ones for the BALB/c (name of an inbred strain of mice) TCRAD locus, we can see Poisson-like as well as Gaussian-

like or Gaussian mixture distribution curves (Figure 2). Thus, we tried first to see whether a simple probabilistic model would fit such curves. The corresponding model makes use of the Brownian Motor Theory (called also Worm-like Chain Theory). In this model, we see DNA as a double stranded string which can be folded to look like a loop with two strands crossing each other. After elimination of the intermediary Delta region (first maturation step), one part of the string is built from the V region and the other from the J one (Figure 3). V and J regions cross each other at a point we call P . A rearrangement is made of the start (resp. end) of the V (resp. J) gene located at the P point. We will suppose that the V and J regions move under constraints like thermal agitation sliding through the *pivot* P . The way they slide is described “click by click”, according to the Brownian Motor Theory. A click is a unit of displacement along one of the V or J loci (or regions) of the DNA string. Each displacement unit is supposed to be a V or a J gene. Thus, one can move along the V region from the most proximal V (V_{prox}) to the most distal V (V_{dist}) with 104 possible consecutive clicks corresponding to the 104 V genes on the V locus. The same mechanism is available for the 49 genes of the J locus. We call pV_{dist} the probability of displacement of the P point towards the distal V gene (V_{dist}) and pV_{prox} the probability of displacement of the P point towards the proximal V gene (V_{prox}) (id. for J region, Figure 3). We call V_{dist} the distance between V_{dist} and P , and V_{prox} the distance between V_{prox} and P (id. for J region). We call V the position of P along the V string, and J the position of P along the J string. If we look the biological results, we observe that proximal V and J genes are used preferentially as compared to distal V and J ones. Thus, we can say that the displacement of the pivot P is proportional to V_{dist} and J_{dist} distances, and inversely proportional to V_{prox} and J_{prox} ones. If we consider that the displacement of P is random and denote α the proportionality factor ensuring that the sum of the considered probabilities is normalized to 1, then we can write the following equations:

pV_{prox} is proportional to J_{dist} and V_{dist} :

$$pV_{prox} = \alpha * J_{dist} * V_{dist}$$

pV_{dist} is proportional to J_{prox} and V_{prox} :

$$pV_{dist} = \alpha * J_{prox} * V_{prox}$$

$$pV_{prox} + pV_{dist} = 1 = \alpha * [(J_{dist} * V_{dist}) + (J_{prox} * V_{prox})]$$

From these relations we can infer the following equations:

$$pV_{prox}(k) = \frac{(V_{dist}(k-1) * J_{dist}(k-1))}{((J_{prox}(k-1) * V_{prox}(k-1)) + (V_{dist}(k-1) * J_{dist}(k-1)))}$$

$$pV_{dist}(k) = \frac{(V_{prox}(k-1) * J_{prox}(k-1))}{((J_{prox}(k-1) * V_{prox}(k-1)) + (V_{dist}(k-1) * J_{dist}(k-1)))}$$

$$pJ_{prox}(k) = pV_{prox}(k) \text{ and } pJ_{dist}(k) = pV_{dist}(k)$$

From these equations, we applied the following algorithm for the displacement of P :

1) for the initial conditions, we randomly chosen:

- V : initial position of P on the V strand
- J : initial position of P on the J strand

For each Brownian click, we made 2 random choices by using 2 random variables $T1$ and $T2$ supposed to be uniform on $[0,1]$:

- $T1$ for pV_{prox} :

- If $T1 \leq pVprox$: P is moved by 1 click towards $Vprox$
 - If $T1 > pVprox$: P is moved by 1 click towards $Vdist$
- $T2$ for $pJprox$:
- If $T2 \leq pJprox$: P is moved by 1 click towards $Jprox$
 - If $T2 > pJprox$: P is moved by 1 click towards $Jdist$

We repeat $T1$ and $T2$ choices 3, 10 and 50 times, corresponding to 3, 10 and 50 Brownian clicks. The position of P at the end of the clicks series is the final rearrangement position. We calculated 10^6 cycles corresponding to 10^6 rearrangements in order to study their repartition on a graph.

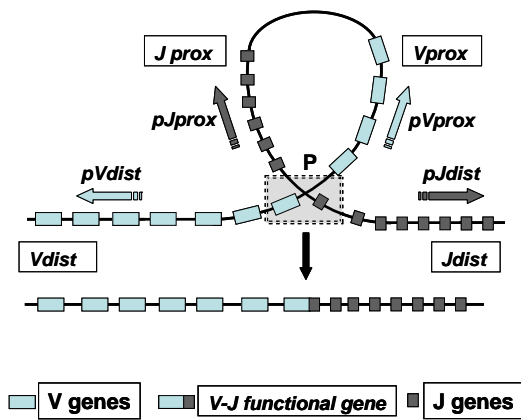


Fig. 3. V and J loci crossing each other at the P point in order to realize rearrangements after the excision of the central loop.

From the curves of Figure 4, obtained for 3 clicks, we can see that rearrangements are possible only for a limited number of clicks. Above 30 clicks for the same V21 or V1 rearrangement, there is no more evolution and rearrangement is fixed. We can also observe that the curves slopes are somehow random and are never so regular than the slopes of the experimental curves (see for example Figure 1).

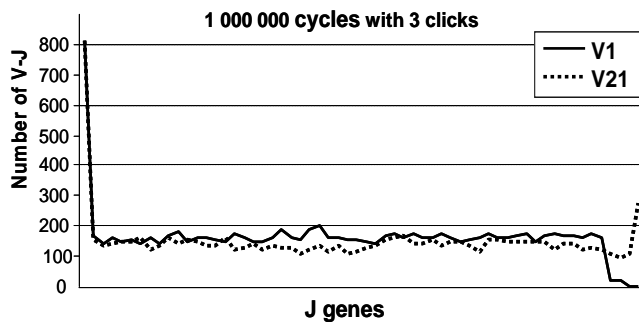


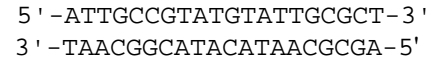
Fig. 4. Simulation rearrangements frequencies obtained for a number of three clicks in the Brownian ratchet model, similar restricted rearrangements frequencies are found with 10, 30 and 50 clicks.

Thus, the Brownian ratchet model is unable to explain experimental curves. Rearrangements come certainly not from

a pure random phenomenon. Then, we tried a more sophisticated model.

B. A realistic progressive windowing model

In the algorithm implementing the model, the DNA double helix is performed by a string made of a strand containing only the four characters: ‘A’, ‘C’, ‘G’ and ‘T’ and by its complementary strand. For example, let us consider a part of a DNA double helix with its two complementary strands, the sense of reading (from the 5' start toward the 3' end):



The sites of V(D)J rearrangements are identified through specific short sequences of 12 or 23 bases (characters) pairs called “RSS” for Recombination Signal Sequences. These RSS are located at the end of each V and the beginning of each J genes. Along the TCRAD locus, the repartition of V and J genes is not uniform (Figure 5).

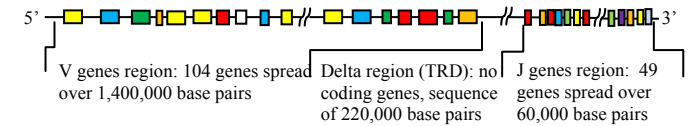


Fig. 5. Schematic representation of the TCRAD locus, with V genes on the left, Delta sequence in the middle, J genes on the right.

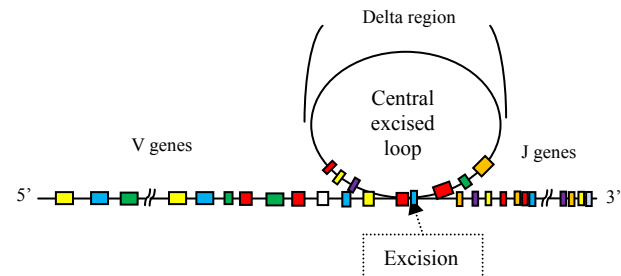


Fig. 6. Mechanism of the maturation

C. Algorithm description

The aim of the realistic new model was to reproduce the global maturation constraints, including the first step of Delta, region excision and after V and J regions progressive elimination by successive rearrangements, with a software algorithm parameterized to find the conditions which reproduce as well as possible the experimental data and as simple as possible.

From the physical positions of V and J genes in the TCRAD locus provided by IMGT® [2], the opening genes parameters are:

- the opening location (the site where the window opening mechanism begins),
- the opening speed for V and J,
- the opening duration.

We will add two more parameters:

- the first maturation time for the TCRD opening and excision (Figure 6),

- an opening Offset, as we suppose a certain rigidity to the DNA chain, so two very close genes cannot be rearranged together (Figure 7).

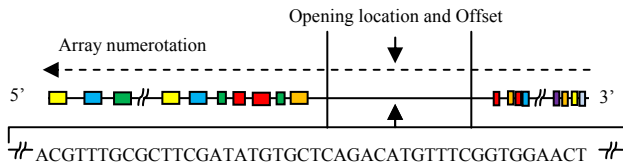


Fig. 7. Initial position

We will use further symbols for variables and parameters of the model defined as follows and valued after a first maturation time as on the Figure 8, where Random denotes a random proportionality coefficient chosen for introducing a variability compatible with the observed inter-individual variations between different individuals and kinds of strains of mice:

$$\begin{aligned} \text{VSpeed} &= \text{VSpeedmin} + \text{Random} \times (\text{VSpeedMax} - \text{VSpeedMin}) \\ \text{WeV0} &= \text{Opening Start} + \text{FirstMaturation} \times \text{VSpeed} \\ \text{JSpeed} &= \text{JSpeedmin} + \text{Random} \times (\text{JSpeedMax} - \text{JSpeedMin}) \\ \text{WeJ0} &= \text{Opening Start} - \text{FirstMaturation} \times \text{JSpeed} \end{aligned}$$

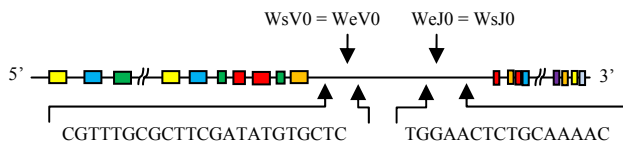


Fig. 8 Iteration of start positions, after a first DNA maturation

Values for variables and parameters of the model after the initial rearrangement are chosen as on the Figure 9:

$$\begin{aligned} \text{OpeningDuration} &= \text{Random} \times (\text{OpeningDurationMax} - \text{Maturation}) \\ \text{VSpeed} &= \text{VSpeedmin} + \text{Random} \times (\text{VSpeedMax} - \text{VSpeedMin}) \\ \text{WeV1} &= \text{WeV0} + \text{Maturation} \times \text{VSpeed} \\ \text{WsV1} &= \text{WeV1} + (\text{OpeningDuration} - \text{Maturation}) \times \text{VSpeed} \\ \text{GeV1} &= \text{WeV1} + \text{Random} \times (\text{WsV1} - \text{WeV1}) \\ \text{JSpeed} &= \text{JSpeedmin} + \text{Random} \times (\text{JSpeedMax} - \text{JSpeedMin}) \\ \text{WeJ1} &= \text{WeJ0} - \text{Maturation} \times \text{JSpeed} \\ \text{WsJ1} &= \text{WeJ1} - (\text{OpeningDuration} - \text{Maturation}) \times \text{JSpeed} \\ \text{GeJ1} &= \text{WeJ1} - \text{Random} \times (\text{WeJ1} - \text{WsJ1}) \end{aligned}$$

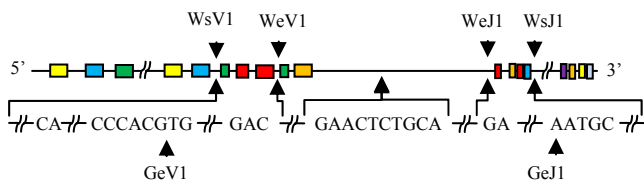
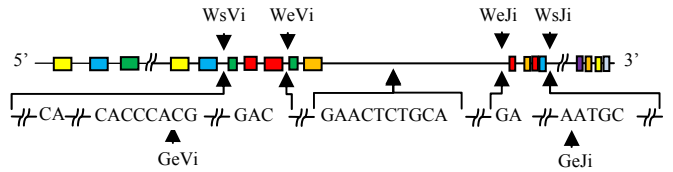


Fig. 9 Positions after a first initial rearrangement

Values for variables and parameters of the model after the i^{th} rearrangement are defined as on the Figure 10:

$$\begin{aligned} \text{OpeningDuration} &= \text{Random} \times (\text{OpeningDurationMax} - \text{Maturation}) \\ \text{VSpeed} &= \text{VSpeedmin} + \text{Random} \times (\text{VSpeedMax} - \text{VSpeedMin}) \\ \text{WeVi} &= \text{WeV}(i-1) + \text{Maturation} \times \text{VSpeed} \\ \text{WsVi} &= \text{WeVi} + (\text{OpeningDuration} - \text{Maturation}) \times \text{VSpeed} \\ \text{GeVi} &= \text{WeVi} + \text{Random} \times (\text{WsVi} - \text{WeVi}) \end{aligned}$$

$$\begin{aligned} \text{JSpeed} &= \text{JSpeedmin} + \text{Random} \times (\text{JSpeedMax} - \text{JSpeedMin}) \\ \text{WeJi} &= \text{WeJ}(i-1) - \text{Maturation} \times \text{JSpeed} \\ \text{WsJi} &= \text{WeJi} - (\text{OpeningDuration} - \text{Maturation}) \times \text{JSpeed} \\ \text{GeJi} &= \text{WeJi} - \text{Random} \times (\text{WeJi} - \text{WsJi}) \end{aligned}$$

Fig. 10. Positions at the i^{th} rearrangement

The excised loop can produce real neo-genes or inefficient non-coding sequences, then the output loop status can lead to:

- a not possible rearrangement (providing a not functional protein) in different cases:
 - biological unproductive sequences (bad triplets congruence),
 - WeVi (resp. WeJi) overpasses the V (resp. J) genes region,
 - the number of allowed rearrangements (5) is over passed,
- an efficient/available rearrangement (leading to a convenient opening triplet reading frame):
 - 33% of the rearrangements are successful and a new string with a new gene is generated (Figure 11).

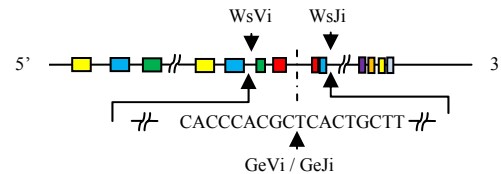


Fig. 11. A new string is generated

When the windows are overlapping the start or the end of the V or J genes regions, the probability of success (33%) is diminished in the following factor (Figure 12)

$$\begin{aligned} \text{PV} &= (\text{WsVi} - \text{GeVi}) / (\text{WsVi} - \text{WeVi}), \text{ if the overlap concerns V region} \\ \text{PJ} &= (\text{WsJi} - \text{GeJi}) / (\text{WsJi} - \text{WeJi}), \text{ if the overlap concerns J region} \\ \text{PV} \times \text{PJ} &= \text{if the overlap concern both V and J regions} \end{aligned}$$

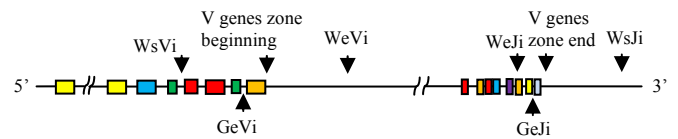


Fig. 12. Success probability moderation

In order to reproduce exactly the same simulation and taking into account the “natural” uncertainty, perturbations are introduced at several points in the algorithm of the model. These perturbations are done by using some random functions. For example, for the variability of the opening speeds, the speeds are not proportionally chosen: an error of 10% is introduced. The first maturation time is not always the same, but randomly chosen between zero and a maximum value. There are two main reasons for the occurrence of the two hot spots of V(D)J recombinations. First, the possible

rearrangements are similar to different random choices of recombination locations respecting the RSS values and being done inside the DNA chain, the first following a Poissonian law like in experiments and the two others are following a Gaussian law. In Figure 13, the choice of successive random numbers allows the appearance of a hot spot.

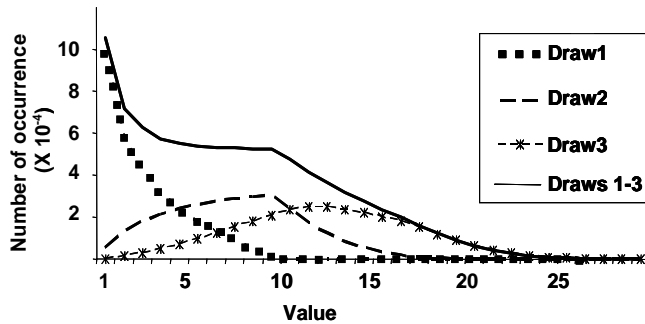


Fig. 13. Three successive draws of random integers were done successively, the first giving an integer draw 1 between 0 and 10 following a Poissonian law. The second and the third follow a Gaussian law, the second giving an integer draw 2 between draws 1 and draw 1 + 10, and the third giving an integer draw 3 between draw 2 and draw 2 + 10, and that 300 000 times. The first, second and third curves were added to build the sum curve.

The analysis of J gene density into the DNA reveals that genes are less dense between J39 and J28. Moreover, we observed a maximum of density at 1.24 for the gene J21, at the place of the second hot spot. Interestingly, we observed the two hot spots when the parameters values of the model were in the range where the opening dynamics allowed more frequent rearrangements in the areas of the maximal gene density.

To conclude, we observed by simulating our model a double mode, contrarily to the Warmflash and Dinner model [12], and this double mode occurred when the parameters were in the range where the two previous circumstances were present.

IV. EMPIRICAL AND THEORETICAL RESULTS COMPARISON

If we consider that the about 100 V genes and 50 J genes have a mean length of 1000 nucleotidic bases the progressive windowing model above explains a variability of rearrangements for about 10^5 neo-genes (which is the estimated richness of the immune repertoire in mammals). This number can be obtained by considering the third part of the total amount of possible choices for the central triplet of the neo-genes, located on the splicing point P , which has about $1000 \times N$ possible different locations on the TCRAD locus, N resulting from the combinatory of the choices of permitted pairs made of one V and of one J gene. In principle, $N = 5000$, but the examination of the VJ rearrangements shows that 96% of the N pairs are observed in simulated results, corresponding to about 4700 pairs with 300 highly infrequent pairs thus explaining an important variability of the VJ repertoire (Figure 14).

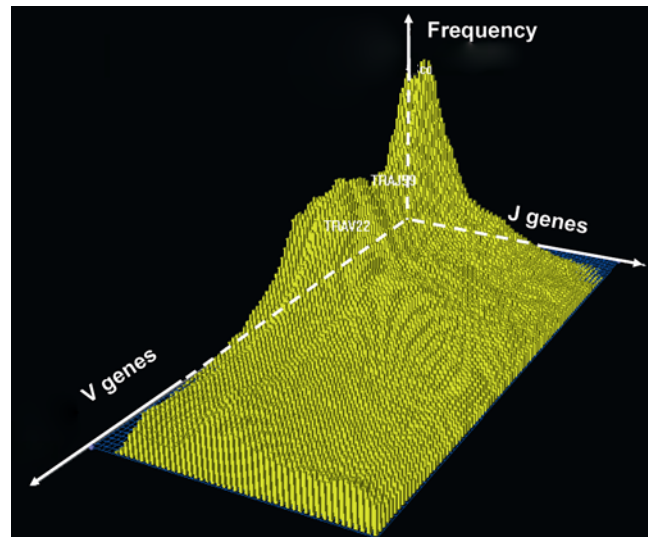


Fig. 14. Repartition of the possible pairs of V and J genes (simulation results)

For the remaining possible pairs, we have confronted the results of the second realistic model with the experimental data (Figure 15). If we represent both empirical and theoretical frequencies by calculating the cubic spline function fitting the best the experimental data by a grey (resp. white) surface on Figure 15 top (resp. middle). We represent in red (resp. green) the experimental observations over (resp. under) the surface corresponding to an under- (resp. over-) estimation by the model, showing that this model accounts well for the real variability of data. In particular, the prediction by the model of the repartition of the proximal, central and distal genes of the J family rearranged with a specific V gene is in a good agreement with the observed rearrangements [1; 12]. For instance the rearrangements of V14, with J61 to J48 are estimated to represent 35% and were actually observed in 37%, with J47 to J24 are estimated to 50% and observed in 46% and with J23 to J1 are estimated to 15% and observed in 17%.

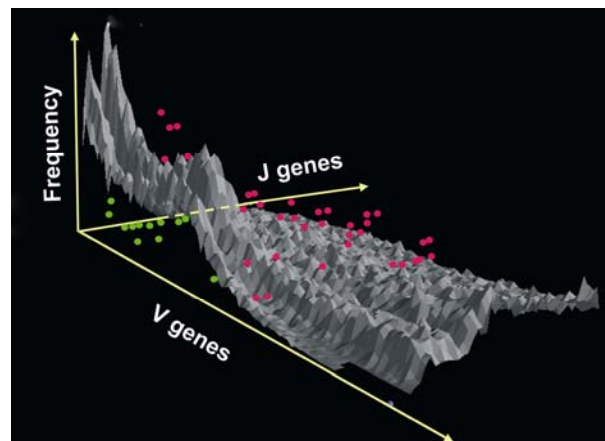


Fig. 15. Confrontation between simulation (grey and surface) and experimental data (red/green points over/under the grey surface)

V. CONCLUSION AND PERSPECTIVES

We proposed in this paper to build a model accounting for the variability of the immune repertoire and we can conclude that the second progressive windowing model accounts well for the experimental observations. The perspectives of this work are double:

- we are now studying the variability of the repertoire in human (by puncturing the thymus in newborn having a thoracotomy for cardiac surgery) and it confirms that for "normal" individuals the same kind of rearrangements repartition is observed in their T lymphocytes of the thymus [14].

- a systematic study could be done for pathologic individuals, particularly for those having been infected by specific infectious diseases (from bacterial ones provoking a normo-, under- or over- response of the immune system to viral ones causing an immune responses or paralysis).

The enrichment of the IMGT data bases using this systematic study in human could allow the definition of an effective immune signature or the identification of profile for an individual in order to prevent in the future abnormal immune responses to specific infections. It could also identify the weight of the possible sources of variability (due to genetic, environmental or aging factors) and hence permit the clustering of a population in "at risk" or "normal" classes in order to improve the global public health policy (vaccination, hygiene improvement,...).

REFERENCES

- [1] N. Pasqual, M. Gallagher, C. Aude-Garcia, M. Loiodice, F. Thuderoz, J. Demongeot, R. Ceredig, P. N. Marche, and E. Jouvin-Marche, "Quantitative and Qualitative Changes in ADV-AJ Rearrangements During Mouse Thymocytes Differentiation : Implication For a Limited TCR ALPHA Chain Repertoire," *J. Exp. Medicine*, vol. 196, pp. 1163-1174, 2002.
- [2] T. P. Baum, N. Pasqual, F. Thuderoz, V. Hierle, D. Chaume, M. P. Lefranc, E. Jouvin-Marche, P. Marche, and J. Demongeot, "IMGT/GeneInfo : enhancing V(D)J recombination database accessibility," *Nucleic Acids Res.*, vol. 32, pp. 51-54, 2004.
- [3] T. L. Messier, J. P. O'Neill, S. M. Hou, J. A. Nicklas, and B. A. Finette, "Normal V(D)J recombination occurs between an RSS with a 12 bp spacer and an RSS with a transposition mediated by V(D)J recombinase in human T lymphocytes," *The EMBO Journal*, vol. 22, pp. 1381-1388, 2003.
- [4] J. P. de Villartay, D. I. Cohen "Gene regulation within the TCR-alpha/delta locus by specific deletion of the TCR-delta cluster," *Res Immunol*, vol. 141, pp. 618-623, 1990.
- [5] I. Abarategui, M. S. Krangel "Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination," vol. 26, pp. 4380-4390, 2007.
- [6] N. K. Jerne, "Towards a network theory of the immune system," *Ann. Immunol.*, vol. 125, pp. 373-389, 1974.
- [7] C. Bidot, F. Gruy, C. S. Haudin, F. El Hentati, B. Guy, and C. Lambert, "Mathematical Modeling of T-Cell Activation Kinetic," *J. Comp. Biol.*, vol. 15, pp. 105-128, 2008.
- [8] M. Bodnar, and U. Fory's, "Periodic Dynamics in a Model of Immune System," *Applicationes Mathematicae*, vol. 27, pp. 113-126, 2000.
- [9] V. A. Folcik, G. C. An, and C. G. Orosz, "The Basic Immune Simulator: An agent-based model to study the interactions between innate and adaptive immunity," *Theor. Biol. Med. Model.*, vol. 4, pp. 39, 2007.
- [10] I. M. Dozmorov, M. O. Kochetkova, G. V. Lutsenko, and I. S. Nikolaeva, "Mathematical modeling of interacting lymphocytes with allogenic hemopoietic stem cells," *Cybernetics and Systems Analysis*, vol. 23, pp. 281-287, 1987.
- [11] K. Harada, and T. Ikegami, " Evolution of Specificity in an Immune Network," *J. Theor. Biol.*, vol. 203, pp. 439-449, 2000.
- [12] C. Aude-Garcia, M. Gallagher, P. N. Marche, E. Jouvin-Marche, "Preferential ADV-AJ association during recombination in the mouse T-cell receptor alpha/delta locus," vol. 52, pp. 224-230, 2001.
- [13] A. Warmflash and A. R. Dinner. 2006. A model for TCR gene segment use. *J. Immunol.*, vol. 177, pp. 3857-3864, 2006.
- [14] P. Fuschiotti, N. Pasqual, V. Hierle, E. Borel, J. London, P. N. Marche, E. Jouvin-Marche, "Analysis of the TCR alpha-chain rearrangement profile in human T lymphocytes," vol. 44, pp. 3380-3388, 2007

7. Article 2

Accurate estimation of TR α repertoire by biological and numerical analysis

Florence Thuderoz, Maria-Ana Simonet, Olivier Hansen, Nicolas Pasqual, Aurélie Dariz,
Thierry Pascal Baum, Vivien Hierle, Jacques Demongeot, Patrice Noël Marche and
Evelyne Jouvin-Marche

Article soumis à Plos Computational Biology

Les données sur le locus TRA/TRD murin montrent que les réarrangements entre les gènes variables (V) et les gènes jonction (J) ne sont pas aléatoires, mais sont au contraire ordonnés et régulés de manière précise au cours du mécanisme de réarrangement V(D)J (Pasqual, Gallagher et al. 2002).

La taille du répertoire immunitaire murin n'est pas connue avec exactitude, mais elle est généralement estimée avec l'hypothèse que les réarrangements VJ sont équiprobables, ce qui conduit à une surévaluation de leur nombre. En effet, certaines combinaisons VJ ont une probabilité très faible d'être rencontrées. Cependant, l'évaluation des combinaisons VJ est difficile à réaliser expérimentalement, car la région variable possède une partie centrale qui a été dupliquée au cours de l'évolution. L'exploration des réarrangements qui peuvent se produire dans cette région du locus TRA/TRD est très difficile, car les différents gènes V dupliqués possèdent une très forte homologie. Par conséquent, il n'est pas possible à l'heure actuelle d'estimer de manière précise toutes les combinaisons VJ, pas plus que leurs fréquences de réarrangement.

Dans ce contexte, la simulation informatique des réarrangements au niveau du locus TRA/TRD nous apporterait une aide significative dans l'estimation des réarrangements V(D)J, mais aussi dans la caractérisation des mécanismes de régulation du locus TRA/TRD.

Pour répondre à ces questions, nous avons développé un modèle dynamique basé sur des observations expérimentales des réarrangements VJ au cours du développement embryonnaire et de la différenciation des lymphocytes T. Notre modèle s'appuie sur la comparaison systématique entre les données expérimentales de notre laboratoire, mais aussi des données de la littérature, et les résultats de simulation. Cette approche permet de valider les hypothèses émises sur les mécanismes de régulation des réarrangements chez la souris, tout en estimant de manière précise la fréquence des combinaisons VJ.

Les données expérimentales nous ont permis de préciser les paramètres dynamiques contrôlant la mise en place des réarrangements, qui constituent la base de notre modèle. Ces données comprennent les vitesses d'ouvertures des régions V et J et la quantification des combinaisons VJ, qui a été utilisée pour la confirmation de notre modèle.

Les vitesses d'ouverture ont été estimées par l'analyse expérimentale de la détection des réarrangements VJ au cours de l'ontogénie. Les données quantitatives comprennent les profils d'utilisation des gènes J par les familles V1 et V21 situées aux deux extrémités du locus, ainsi que les fréquences de réarrangement des six membres de la famille V14 situés au niveau central de la région V.

Concernant le modèle, la taille de la population à simuler est un paramètre fixé par l'opérateur. Dans cette population, tous les TR vont avoir le même comportement. Une ouverture est initiée au niveau d'un site nommé "site d'ouverture", dont la localisation peut être changée. Dans nos simulations, le TEA est pris comme site d'ouverture (de Chasseval and de Villartay 1993). Une première maturation permet ensuite d'éliminer les gènes codant pour la chaîne delta qui sont situés au centre du locus TRA. Une vitesse et un temps d'ouverture permettent de définir une fenêtre d'accessibilité dans laquelle pourront avoir lieu les réarrangements VJ. Lorsqu'un réarrangement est "productif", la simulation s'arrête et passe à celle du réarrangement d'une autre chaîne TR. Si le réarrangement est "non productif", une seconde phase d'ouverture du locus est initiée, permettant au thymocyte en différenciation d'avoir une nouvelle chance d'exprimer un TR productif. Si le TR est à nouveau déclaré "non productif", ce mécanisme est reconduit, mais il est limité par la taille du locus.

Pour rendre compte des hot spots de recombinaison, nous avons introduit deux éléments qui n'étaient présents, ni dans les premiers travaux de modélisation par notre laboratoire, ni dans le modèle de Warmflash. Le premier est la variabilité de la vitesse d'ouverture, qui permet de moduler les fenêtres d'ouvertures des régions V et J. Le second est un temps de maturation entre deux recombinaisons, au cours duquel des réarrangements ne peuvent pas avoir lieu. Ces deux éléments nous ont permis : 1) de rendre compte des hot spots de recombinaison, 2) d'augmenter la taille du répertoire de la chaîne alpha, ce qui est en accord avec la littérature, 3) de montrer que les recombinaisons VJ n'ont pas toutes les mêmes fréquences de réarrangement et 4) que le nombre de combinaisons VJ préférentielles est quasi-inexistante. Les résultats provenant des différents essais réalisés sans la vitesse variable et/ou sans le temps de maturation sont présentés dans la discussion de l'article.

Résultats :

Détermination expérimentale des paramètres contrôlant le locus

Détermination des vitesses d'ouverture

L'analyse des données expérimentales nous a permis de définir les vitesses moyennes d'ouverture des régions V et J : 13000pb/h et 713pb/h respectivement. Connaissant la taille des deux régions, 1300kb et 64kb pour les régions V et J respectivement, nous obtenons un délai de trois jours pour une ouverture totale.

Une analyse plus approfondie des résultats concernant l'accessibilité des gènes J au cours de l'ontogénie par le gène V19 a montré que l'accessibilité de la région J n'était pas constante au cours de l'ontogénie. En partant de cette observation, nous avons implémenté une vitesse variable dont la valeur est basée sur l'intervalle de confiance à 99,9% de la valeur moyenne de la vitesse expérimentale.

Localisation des hot spots de recombinaison

Différentes études ont montré la présence de hot spots de recombinaison au niveau de l'utilisation de la partie centrale de la région J. Les hot spots correspondent à des zones où les réarrangements ont lieu avec une fréquence plus importante. Sur le locus murin, le premier hot spot est situé entre J59 et J48 et le second se situe entre J31 et J22 (Rytkonen, Hurwitz et al. 1994; Gahery-Segard, Jouvin-Marche et al. 1996; Rytkonen, Hurwitz et al. 1996; Aude-Garcia, Gallagher et al. 2001)

Partie simulation

La figure 17 montre une représentation en deux dimensions des combinaisons VJ. Cette représentation laisse la possibilité de choisir le gène V et/ou J dont on veut étudier le profil de recombinaison. Cette représentation montre : 1) que le profil de recombinaison des gènes V se modifie suivant la position sur le locus, les gènes V proximaux se réarrangeant de manière préférentielle avec les gènes J proximaux et 2) que la présence des hot spots de recombinaison concernent les gènes V situés en position centrale du locus.

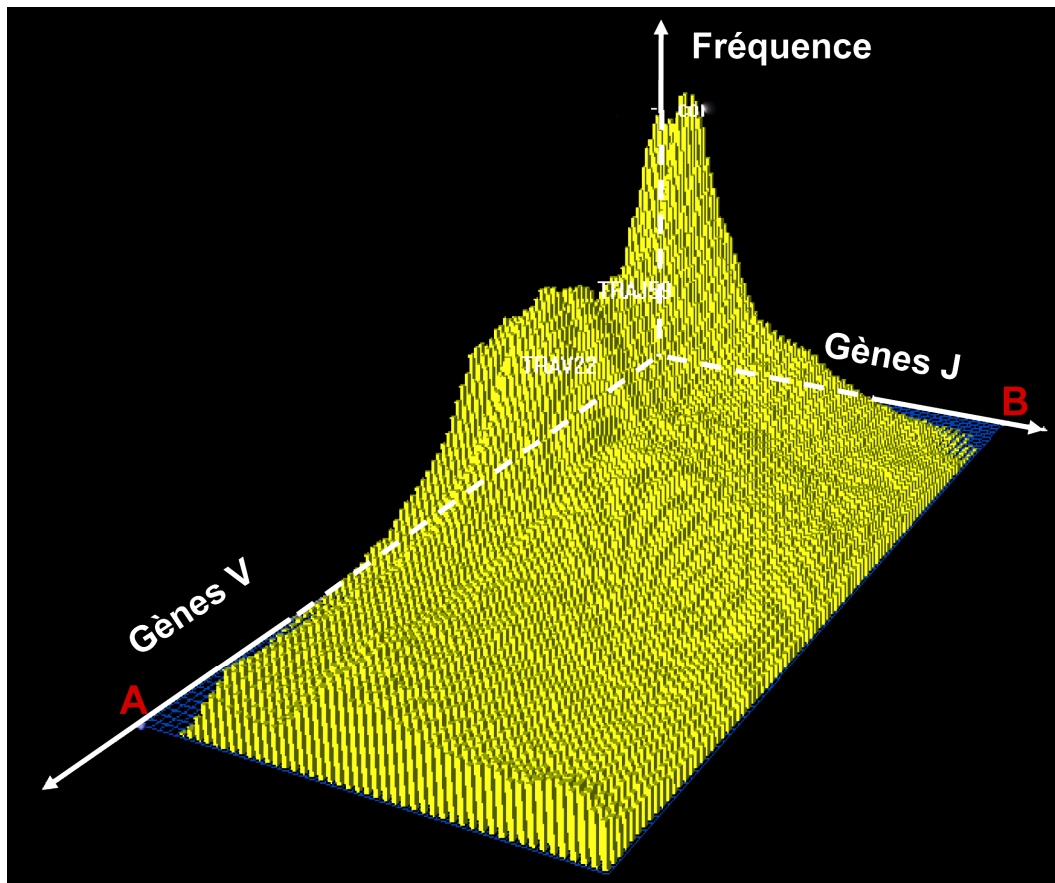


Figure 16 Présentation 3-D des résultats de simulation. L'axe des X représente les gènes J, l'axe des Y les gènes V et l'axe des Z leurs fréquences de réarrangement obtenues par simulation. Les régions A et B correspondent à des zones de réarrangements ultra-minoritaires : A) entre les gènes V distaux et les gènes J proximaux, et B) entre les gènes V proximaux et les gènes J distaux.

La figure 17 montre une représentation en deux dimensions des combinaisons VJ. Cette représentation laisse la possibilité de choisir le gène V et/ou J dont on veut étudier le profil de recombinaison. Cette représentation montre 1) que le profil de recombinaison des gènes V se modifie suivant la position sur le locus ; les gènes V proximaux se réarrangent de manière préférentielle avec les gènes J proximaux et 2) la présence des hot spots de recombinaison pour les gènes V situés en position centrale du locus.

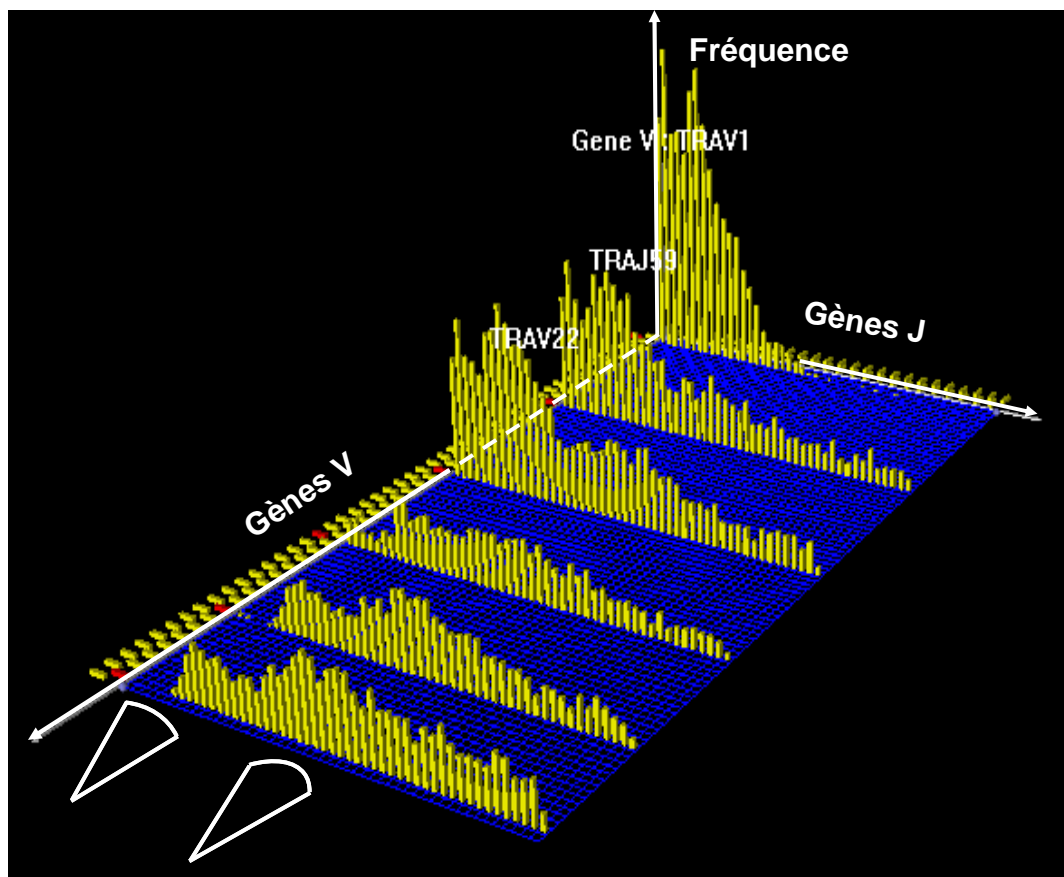


Figure 17 Présentation en 2-D des résultats de simulation. L'axe des X représente les gènes J, l'axe des Y représente les gènes V et l'axe des Z leurs fréquences de réarrangement obtenues par simulation. Les régions des hot spots de recombinaison sont indiquées par les cônes blanc.

Une estimation précise de la diversité du répertoire

La simulation d'un million de réarrangements produit 96% de combinaisons productives avec un maximum de 4700 combinaisons sur les 4900 combinaisons possibles (pseudo-gènes exclus). Il y a donc en moyenne 200 combinaisons très peu probables (4%).

En conclusion, ces travaux apportent une vue plus fine du répertoire des TR α au niveau génomique. Ils permettent d'intégrer et de confronter différentes hypothèses sur l'accessibilité et la régulation du locus TRA/TRD, pour en vérifier la pertinence. De plus, ils permettent de rendre compte de la position des hot spots de recombinaison, en accord avec les données expérimentales.

Numerical modeling of the T cell receptor recombination of the TRA/TRD locus

Florence Thuderoz^{1*}, Maria-Ana Simonet^{1,2*}, Olivier Hansen^{1*}, Nicolas Pasqual^{2,*§}, Aurélie Dariz^{2,3}, Thierry Pascal Baum¹, Vivien Hierle², Jacques Demongeot^{1£}, Patrice Noël Marche^{2,3£} and Evelyne Jouvin-Marche^{2,3£}

¹ CNRS, UMR 5525, Grenoble, France.

² INSERM, U823, Grenoble, France.

³ Université Joseph Fourier-Grenoble I, Faculté de Médecine, Institut Albert Bonniot, UMR-S823, Grenoble, France

* These authors equally contributed to this work; £ These authors equally contributed to this work; § Present address: ImmunID, Commissariat à l'Energie Atomique, Grenoble, France.

Correspondence to:

Patrice Marche (patrice.marche@ujf-grenoble.fr)

Laboratoire IAPC, INSERM U823 Equipe 8, Institut Albert Bonniot, BP170 La Tronche 38042 Grenoble Cedex9, France.

Jacques Demongeot (Jacques.Demongeot@imag.fr)

Laboratoire TIMC-IMAG, CNRS (UMR 5525), Faculté de Médecine de Grenoble, 38706 La Tronche Cedex, France.

Short title: **Estimation of TR α repertoire**

Key words: **TR-Diversity, TRA/TRD locus, rearrangement, simulation, model, repertoire.**

† **Abbreviations:** NCBI: National Center for Biotechnology Information; IMGT: ImMunoGeneTics database; TR: T Cell Receptor; F18: 18 days of Fetal life; D1: first Day of life; V: Variable gene of TRA/TRD locus coding for the mouse TR α chain; J: Junction gene of the TR α chain; P-R: Pseudogene for Rearrangement; RSS: Recombination Signal Sequence; HS: Hot Spot.

Abstract

Background

T-Cell antigen Receptor (TR) repertoire is generated through rearrangements of V and J genes encoding α and β chains. The number and frequency of the different V-J combinations during ontogeny and development of the immune system remain to be precisely established.

Methodology/principal findings

We addressed this issue, by building a model able to account for of V-J α gene rearrangements during thymus development of mice. So we developed, based on experimental data, a numerical model on the whole TRA/TRD locus to estimate how V α and J α genes become accessible to rearrangements. The progressive opening of locus to V-J gene recombinations is modeled through windows of accessibility of different sizes and with different speeds of progression. Furthermore, the possibility of successive secondary V-J rearrangements was introduced in the modeling.

The model points out some unbalanced V-J associations resulting from a preferential access to gene rearrangements and from a non-uniform partition of the accessibility of the J genes, depending on their location in the locus. The model shows that 1 to 3 successive rearrangements are sufficient to explain the use of all the V and the J genes of the locus. Finally, the model provides information on the kinetics of rearrangements and on the frequency of each V-J association.

Conclusion/significance

The model accounts for the essential features of the observed rearrangements on the TRA/TRD locus and may provide a reference for the repertoire of the V-J combinatorial diversity.

Author summary

T lymphocytes express an antigen receptor (TR) at their surface. Each TR is composed of two chains, either α and β or γ and δ , which are encoded by a variable (V), a joining (J) and a constant (C) genes region. The TRA locus, encoding for the α chain, is composed of around 100 V, 60 J genes, and a single C gene. To constitute a functional chain, a V and a J gene rearrange to form a single exon, through a mechanism called V(D)J recombination. The current state of the technology permits the analysis of some V-J combinations, essentially those at the extremities of the locus, but still fails to account for a complete estimation of the V-J combinations.

To get a global view of the V-J recombinations, we have defined a numerical model fitting the available experimental observations and applied it to simulate the dynamics of all V-J combinations. Our model gives new insights on the rules controlling the use of V and J genes, spanning over a 1500 Kb region and provides a dynamic estimation of the total V-J combinations. It provides a useful tool to evaluate repertoires in immune responses.

Introduction

Functional antigen receptors expressed by T lymphocytes (TR) are generated during ontogeny by somatic recombination of gene segments, coding for the variable (V), the joining (J), and the constant (C) segments [1]. The recombination mechanism is largely dependent on both the accessibility of the loci and the RAG enzymatic complex [2-5]. The murine TRA/TRD locus is composite, encoding TR α and δ chains, and encompassed of more than 100 functional V genes [6]. In theory, each of the V genes may target one of the 49 functional J genes. The use of V and J genes during the process of recombination has been widely debated, and the studies support the consensus that V-J combinations are not random with a use of J segment starting at the 5' end (proximal to the V segments) and proceeding to the 3' end [7-13]. The accessibility of the J region is controlled by the TR α enhancer (E α), located at the 3' end of the C gene [14], and by two promoters; i) T early α (TEA), located at the 5' end of the J α region, and ii) J49 located 15Kb downstream of TEA. Both of the promoters are activated by E α [4,5,15]. E α controls all the V to J associations whereas the two promoters are required for the rearrangements of the J genes situated at the 5' end of the J α region. However, the analyses of TEA-deleted alleles and those of the blockade of TEA transcription showed profound alterations in J use and support the hypothesis that the TEA promoter can regulate both positively the promoters located in the first 12 Kb of J genes and negatively the downstream promoters [4,15-17]. One of the characteristics of the TRA/TRD locus is that a primary V-J rearrangement can be replaced by a secondary one. Furthermore, it is suggested that the first V-J association targets the secondary one to a set of J segments located near the J segment involved in the primary rearrangement [5,16]. Concerning the V segments, the rules governing their utilization has not been clearly elucidated. Nevertheless, observations converge to a consensus: the V segments use would progress from proximal V genes, located near the J region, towards the V genes located in the distal region[9,10]. At this point in time, the mechanism involved in the control of V genes accessibility remains unknown.

The current state of the technology permits for the analysis of some V-J combinations, essentially those at the extremities of the locus, but still fails to account for a complete estimation of the V-J combinations. The major difficulty resides in the fact that some V genes are duplicated in several similar copies in the central part of

the V region, making problematic their unambiguous identification by molecular methods [18].

And so, numerical modeling of V-J recombination process may offer valuable issues to overcome the difficulty to access to a global view of TRA repertoire. For instance in the model, if the J genes are chosen in a sequential manner, their use is found unimodal, whereas it is known from experimental data that TRA/TRD locus displays two hot spots of recombination [2-5]. This discrepancy led us to build a new mathematical model, parameterized from the experimental data, on all V and J genes including those in distal, proximal, and central positions. Confrontation between experimental and simulated data makes possible an estimation of dynamical parameters like those relative to the opening speeds and rearrangement frequencies of V and J regions; giving a more accurate estimation of the combinatorial diversity.

This work based on both experimental data and computer assisted analysis of the primary TRA repertoire might serve as a predictive tool to evaluate repertoires in immune responses.

Results

Update of the parameters controlling the mouse TRA/TRD locus utilization by an experimental approach

The goal of building a model representative of the V-J associations was to reproduce the global biological features of T lymphocyte maturation with a software algorithm. This algorithm must be parameterized to find the conditions that reproduce as best as possible the experimental data and should be as simple as possible.

To build the model, we needed information about the number and the position of the V and J genes in the TRA/TRD locus. This information was provided by IMGT (<http://imgt.cines.fr/>) and is summarized in Table 1. In addition, we needed to define parameters such as the opening location (the site where the opening mechanism begins), the opening speed for the access to V and J genes and the opening duration. Two more parameters were added, a first maturation step in order to eliminate the TRD genes and an opening offset as we supposed a certain rigidity of the DNA chain, so that two very close genes cannot rearrange together.

The mouse opening speeds of V and J regions in TRA/TRD locus are highly imbalanced.

The data given in Table 2 were obtained from rearrangements at the genomic level in BALB/c mice during ontogeny of the thymus. Briefly, concerning the J opening speed in fetal day 18 (F18), rearrangements of V19 with J61 to J48 correspond to an opening of the J locus of 14773 bp in 24h, thus the opening speed associated to these 24h period is estimated to 615 bp/h. The same analysis was done on F19, F20 and at day of birth (D0). The data presented in Table 2 show that during ontogeny, the opening speed of the J locus varies slightly, between 375 bp/h and 1150 bp/h, corresponding to a mean opening speed of about 713bp/h \pm 3x396 bp/h with a 99.9 % confidence interval.

In order to determine the "V-speed", single member V families at each V extremity of the TRA/TRD locus were used; for instance, V1 and V2, the most distal from J region, and V19, V20 and V21, located at the extremity the closest to J region were analyzed (see <http://imgt.cines.fr/>). We found rearranged proximal V genes as early as F18, while rearrangements of distal V1 and V2 genes (1300 Kb distant from the proximal V genes) were only detected since D0 [9]. Thus, the entire V region needs

about 3 days to be completely opened. This allows us to estimate that the overall “opening V-speed” is broadly about 18 Kb/h (1300 Kb/(3x24h)) with a 99.9% confidence interval of about $18 \pm 3 \times 5$ Kb/h.

Hotspots of V-J recombinations are observed in the central part of the locus

Due to the multiple V genes duplications in the central part of the V locus, the profiles of the associations of the central V genes have not been clearly elucidated. Nevertheless, it has been reported by Rytönen *et al.* [3,19] that two hot spots (HS) are present in the J locus, one located between J59 to J48 (HS1), which corresponds to the region controlled by TEA, and the second between J31 to J22 (HS2). Fig. 1A shows the analysis of J used by the 6 members of V14 genes family (which are spread from 1145 Kb to 492 Kb in V region) during TR rearrangements. The profiles of J used by proximal V14-1, V14-2, V14-3 (Fig. 1B) and distal V14D-1, V14D-2 and V14D-3 (Fig. 1C) show that the progression of the V and J genes utilization is respected, with a preference for the J genes located at the hot spots HS1 and HS2 [10,20].

Computational modeling approach

Validation of the model by comparing simulated with experimental data

The analysis of the results from the simulations of a first model, based on the Brownian ratchet theory [21], shows that this model is unable to explain experimental results (cf. additional data). A second model, called sequential windowing model, is proposed, which accounts precisely for the experimental data (Fig. 2).

Opening mechanism effects

Simulations were performed using two opening speeds chosen within intervals closed to the experimental 99.9 % confidence intervals for V and J speeds, namely $S_V \in [0.35 \text{ Kb/h}, 34 \text{ Kb/h}]$ and $S_J \in [0.4 \text{ Kb/h}, 1.55 \text{ Kb/h}]$ with respectively, a mean opening speed of about 18 Kb/h for V genes and 1 Kb/h for J genes. The opening location of the simulation was fixed between the V and J genes in order to access directly to the TRAV and TRAJ genes after the first maturation, which eliminated the genes coding for TRD genes (region encompassing TRDV1 to TRDV5).

Validation of the model by comparing simulated with experimental data

The model called sequential windowing model, shows a good consistency with the experimental data (Fig. 2). First, we focused on the modeling for the recombinations involving the proximal and distal V-J. Indeed, simulation data show that the utilization of the V genes decreases from proximal to distal V genes (Fig. 2A), and thus follows the one of the J genes (Fig. 2B) that was previously observed in Pasqual *et al.* [9].

Next, we compared experimental quantifications with simulated outputs for V1 and V21 rearrangements on the same set of 9 J genes, which are scattered along the J region. Both experimental and simulated data show that V1, which is the most distal V, has a low utilization rate of the proximal J genes and uses essentially the middle and distal J genes (Fig. 2C, E and G), and that the V21, which is the most proximal V, is mainly rearranged with the proximal J genes (Fig. 2D, F and H) following a Poisson distribution. In the case of V1, It is important to note that this V is a peculiar one, since it is the most distant V, and it is isolated at 160Kb from the next neighbor V, namely V2. This specific feature of the V1 position in the locus may account for the discrete difference between the simulation and the experimental data for its utilization of the distal J. Nevertheless, the global utilization of V1 is similar in both experimental and simulated data. In conclusion, the observed frequencies of the use of the nine J genes, taken as representative of the J region by V1 and V21, are in good correlation with their simulated frequencies.

Program interface and generated graphics

The simulation program with its user interface (Fig. 3A) provides a 2-dimensional diagram showing a conditional V-J rearrangement distribution for different V from proximal, central and distal positions (Fig. 3B) and a 3-dimensional histogram of V-J rearrangements accounting for the all TR α chain combinational repertoire (Fig. 3C). The 3-dimensional diagram representation shows a non-uniform distribution of the V-J associations in agreement with experimental data. The repertoire distribution takes a “rift” shape resulting of each V-J Gaussian distribution as described in experimental data (Fig. 1). Combinations between the proximal V and J genes represent the major part of the distribution.

The simulation program also provides a 2-dimensional diagram representing the whole V14 family TR α chain repertoire (Fig. 4A and 4B) and a 3-dimensional diagram with the experimental and simulated data on the same diagram (Fig. 5A and 5B). On

Fig. 4A, the two hot spots of recombination are observed just like on Fig. 1A, these results are also in concordance with the experimental data of Rytönen *et al* [3,19] . Moreover, the experimental frequency use of J genes by the different members of the V14 family is correlated with the simulated frequencies: indeed, on the Fig. 5A and Fig. 5B we can see the global repartition of the experimental points over (red) and under (green) the fitted (Fig. 5A, 3-cubic spline fitting) and not fitted (Fig. 5B) surface interpolating the simulated frequencies.

All the control parameter values of the model are in agreement with the experiments and this makes possible the investigation beyond biological aspects of presently experimentally inaccessible parameters such as the total map of the TRA locus repertoire and the average number of secondary rearrangements.

3D representation of the simulated combinatorial repertoire

The V-J associations found by simulation correspond to about 96% (4704 out of 4900) of the potential V-J associations (Fig. 3C). From this figure, we can define two areas in the distribution where V-J combinations rarely occur: the “A” area points associations between proximal V and distal J; the “B” area points associations between distal V and proximal J. These associations are rarely found because of a non-synchronized opening process on the V and J genes involved. An example for the “A” area, when a proximal V is able to rearrange, the J distal region is not opened yet, and later in the opening process, when the distal J genes are able to rearrange, the proximal V genes were mostly deleted by precedent rearrangements, so these V-J associations are strongly under-represented. The same non-synchronized opening explains the under representation of the distal V-proximal J associations, shown on the “B” area.

Strikingly, these two areas correspond to under-represented V-J associations found by experimental quantifications (Fig. 3C) [9].

Influence of the Recombination Signal Sequence (RSS) on the V-J associations

One puzzling question is the degree of influence of RSS in the rearrangement process. No precise *in vivo* information is available for the TRA locus concerning the relations between the frequencies of use of V and J genes in a given rearrangement and the RSS sequences [22,23]. Hence, we ran simulations with and without taking into account RSS scores, available in IMGT. The algorithm for choosing the J genes

with respect to their RSS values has been tested by using a Monte Carlo method, which showed its validity. The 2D graphs of the V14 family repertoire (Fig. 4A and 4B) show that introducing the RSS scores (Fig. 4B) does not drastically affect the shape of the global repertoire distribution.

The contribution of successive rearrangements of the same chromosome on the total V-J associations was evaluated. To this purpose, secondary rearrangements were allowed between V and J genes remaining in the TRA locus after a primary rearrangement. We observed a decreasing contribution of each wave of successive rearrangements: 40% of the V-J associations arise from the first rearrangements, and 33%, 19% and 8% respectively from subsequent rearrangements (Table 3).

The rearrangements of V14 genes were measured from experimental and calculated from simulated data for the combinations with three J panels, corresponding to series of J genes scattered along the J region (Table 4). There is a good concordance between experimental and simulated data: the V14 combinations with J61 to J48 of the first HS represent 37% in the experimental data, and 35% in the simulated data, those with J47 to J24 stands for 46% and 50%, and with J23 to J1 for 17% and 15% of experimental and simulated data respectively. Strikingly, when the simulation of rearrangements is extended to all V, the use of the different J panels are in the same range than those found with V14 genes, 33% with J61 to J48, 51% with J47 to J24 and 16% with J23 to J1.

Robustness of the model

There are two main probabilistic grounds for the occurrence, in the simulation model, of the hot spots observed in the J region utilization. The successive rearrangements were achieved by consecutive random choices of J genes, respecting the progression of their access to recombination and optionally of their RSS score values. The first J gene choices, corresponding to the first rearrangements, follow a Poissonian law whereas, individually, the two other J gene choices follow a Gaussian law (Fig. 6). Altogether, the consecutive random choices of J genes build a multimodal curve of occurrence which allows the appearance of two hot spots (Fig. 6, solid line). The first hot spot results from the Poisson's distribution of the first J choice and the second one from the Gauss' distribution of the subsequent J choices.

The density of J genes in the TRA locus is not uniform: J genes are less dense between J58-J47, J39-J28 and J14-J4, whereas they are denser between J45-J40

and J24-J15 (Fig. 7). It is important to notice that that the maximum of J gene density is reached in the area around J21-J22, which corresponds to the place of the second hot spot. Interestingly, we observed the hot spots when the parameters values of the model were in the range in which the opening dynamics allowed more frequent rearrangements in the areas of maximal gene density.

The algorithm for choosing the J genes with respect to their RSS values has been tested by using a Monte Carlo method, which showed its validity. To be confident that the sampling size used in simulations was sufficient, we checked the representativeness of the repertoire by making sets of simulations ranging from 10^2 to 1.5×10^6 rearrangements. We found that the diversity becomes relevant when the population size is higher than 5×10^5 rearrangements, showing the pertinence of a repertoire calculation based on a 10^6 alpha chains population. A combinatorial diversity graph has been plotted (Fig. 8). Variations of about 5 to 10% of the parameters (opening speed, opening duration and offset) allowed results in concordance with the experimental data. However, larger variations of the parameters introduced major deviations (not shown) of the modeling simulation results in comparison to experimental data.

Discussion

This article focuses on a new approach in order to better reach the entire 3-dimension profile of V-J rearrangements in the TRA/TRD locus. To accomplish that purpose, we have defined a mathematical model fitting experimental observations. Together, the experimental data and the mathematical model made possible for the interpretation of the mouse T-cell alpha chain repertoire characteristics. The evolution of the shapes of the V-J rearrangement frequencies in the simulations, presented in Fig. 9, shows a transient bi-modal shape corresponding to the hot spots of V-J recombinations as observed in the experimental data. Therefore, our model provides a major improvement to previous attempts of simulation of the TRA combinatorial repertoire.

Opening velocities and gene density

Our model is based on the fact that V and J regions are used in a progressive and decreasing manner from 3' to 5' V side and from 5' to 3' J side. Our quantitative approach points out that the physical position of the genes is the main structural parameter governing the genes rearrangement rate. The J (resp. V) region becomes accessible from proximal to distal genes at a mean "opening J- (resp V-) speed" of approximately 1 Kb/h (resp. 18 Kb/h). This difference between V and J regions opening speeds can be related to the gene density, the length in base and the number of the genes in V and J regions differing greatly. Moreover, the V region length is 20 times higher than the J one (~1300 Kb versus ~60 Kb), but the density of J genes is about 10 times greater than the density of the V genes (1 J by Kb versus 0.13 V by Kb). Consequently, the opening speed calculated in terms of genes per hour is 1.4 for V region and 0.83 for J region, determined by the number of genes in the locus times the speed in Kb per hour, then divided by the length of the locus. Therefore, taking into account the gene density of each region, the opening speeds of the V and J regions are nearly identical. Our observations reinforce two putative scenarios which were previously proposed to explain the opening of the J and V regions respectively:

- the J region local service depends on the density of the elements available for the recombination [24] and,

- the V region global service reflects the larger utilization of a V region whose genes are more scattered than in the J one [9].

In addition, the speed calculations also take into account the regulatory elements which control the gene accessibility [25]. According to our model, 70% of the first rearrangements involve J gene between J61 and J48. This progressive opening mechanism provides V-J combinations with a specific pattern; a given V gene is rearranged with a contiguous set of J genes, the distribution of these J genes presenting a Poissonian (for the proximal V genes) or Gaussian (for the distal V genes) shape. The changes in J use are progressive, depending on the V position: more distal is a V gene, more distal and larger is the set of J genes used, and less represented are these V-J associations in the whole V-J repertoire (Fig. 3C).

The sequential windowing model

The sequential windowing Model well fits the experimental data (in particular the J distributions mentioned above). It gives a dynamical approach of the rearrangement mechanism and a first evaluation of the V-J repertoire. The simulation confirmed the opening V- and J-speeds determined from molecular experiments and gave distributions of V-J pairing coherent with the experimental ones. Moreover, a parametric study of different speeds indicates that experimental speeds are the only ones allowing a specific use of V and J regions, correlated with experimental results, e.g., explaining the fact that the curve exhibiting the hot spots of the V-J pairing distributions is near the diagonal (Fig. 9).

Simulated combinatorial repertoire

In as much as all V-J combination frequencies are computable in the framework of our Model, we are able to visualize the whole simulated TRA combinatorial repertoire. A 3D view points out that one central area in the V-J plane contains most of the observed combinations for the whole repertoire (Fig. 3C). Furthermore, two areas, (A and B in Fig. 3C) reveal rarely represented V-J combinations, due to the fact that the V and J genes involved in these combinations are not accessible to rearrangements in the same timing. For the A area, when the proximal J genes are recombining, the distal V genes are still inaccessible, and when they become accessible, the proximal J genes are deleted because of previous rearrangements. Similarly for the B area, the non-synchronized accessibility of the proximal V genes

and the distal J genes explains that their associations are not observed in the simulation.

Improvement given by the model in the understanding of repertoire characteristics

Our model gives the possibility to estimate the V-J association frequencies and the mean of the rearrangement numbers. Indeed, the probability of any combination is given by selecting a specific V-J combination on the 3D graph (Fig. 3C). If all V-J combinations were equiprobable, the probability of one combination should be of about $2 \cdot 10^{-4}$. Table 5 displays the probability of nine V-J combinations selected along the locus. According to the position of the V and J genes on the locus, the probability varies between 0 and 0.2, highlighting the fact that a V-J proximal combination is 25 times more probable than a V-J distal one.

Secondary V-J rearrangements of the TRA/TRD locus are widely accepted [24,26-29]. However, the number of plausible secondary rearrangements remains unknown. Our model predicts that after the first maturation step of the TRA/TRD locus, consisting in the initial rearrangement on TEA eliminating the TRD genes, the first V-J rearrangement of TRA is followed by one or two secondary rearrangements in 92% of cases and by three in 8% of cases (Table 3). The number of the V-J rearrangements decreases after each wave, the fourth rearrangements having a weak contribution.

The number of secondary rearrangements, conditioning the uses of J genes, can be affected by the lifespan of rearranging thymocytes. For instance, ROR γ -deficient mice, displaying a short lifespan of DP thymocytes, rearrange essentially with the most proximal J, and Bcl-xL-transgenic mice, having DP with longer lifespan, rearrange essentially with the most distal J genes [24].

Regarding the Recombination Signal sequence (RSS) influence, our model is able to incorporate the RSS diversity information throughout score. The simulations using these RSS scores show a low quantitative influence (Fig. 5B), and do not change the global profile of the frequencies curves (Fig. 9). In conclusion, the RSS's may only influence local specificities within the accessibility windows, moving across the TRA/TRD locus in a bi-directional way.

It is important to notice that in the mouse TRB locus, RSS of V neither correlate with any particular restriction of J gene nor with any high frequency level V-J rearrangement [30].

Finally, our model integrates i) a realistic estimation of the opening speeds from real data, ii) overlapped windows concordant with the real mechanism, iii) a non uniform partition of the J genes into overlapped sub-regions, whose lengths are estimated from experiments and iv) the introduction of a maturation time in order to take into account for feedback from the post-transcriptional process and the selection of the TR products. Eventually, it would be important to notice that the complete maturation of the TR and their selection is suggested to be one day long and that the lifespan of the DP thymocytes stage is about 3 days [31,32]. It is compatible with the three main rearrangement times deduced from the model simulations. All these remarks are in favor of the realistic character of our model, which accounts for the essential features of the observed rearrangement process in the TRA/TRD locus.

The rearrangement model presented above is a way to explain the creation of the repertoire namely by simulating different scenarios with a large choice of parameters. By using more flexible windows on the V and J genes (allowing their overlap), despite the complex spatio-temporal opening rules of the locus, our model explains the increase of the TR α chain global combinatorial diversity. From the second rearrangement, the overall V locus and almost, the entire J locus are open. Thus, it is possible to increase the diversity of the V-J combinations, and to vary the frequency of each V-J combination by accounting for experimental data. Moreover, our model is able to explain the two hot spots present in the central part of the J locus. We found that only one to two secondary rearrangements were necessary after the first rearrangement, with a weak third secondary rearrangement.

In conclusion, the combination of experimental and mathematical approaches gives new insights on repertoire biases and allows us to define more accurately the TR primary combinatorial repertoire. To better understand multi-integrated genome regulations, the V-J recombination model could be adapted to other TR as well as Ig loci, for an accurate estimation of the V-J combinatorial diversity in their primary repertoires, giving a dynamic vision of the immune diversity generation during thymocytes and B cell differentiation respectively. We studied the explicit rules controlling the use of 1500 Kb TRA/TRD locus, giving a dynamic calculation of the total V-J combinations involved in the TR α chain repertoire in mice. This estimation

procedure of the primary repertoire combinatorial diversity constitutes a reliable tool to evaluate putative inter-individual deviations in mice and opens the possibility to build a similar model in humans in the future.

Materials and Methods

Nomenclature: Official nomenclature for V and J genes is chosen according to the IMGT database (<http://imgt.cines.fr>). NCBI mouse accession numbers are for V region, AE008683-AE008686, and for J region, M64239. Positions of each V and J genes were calculated from these data as previously described [6].

Mouse: BALB/c mice were purchased from Charles-River (L'Abresles, France). Fetal thymi were obtained from timed pregnancies, where fetal day 1 corresponds to the day of detection of a vaginal plug. Thymic lobes from embryonic mice were pooled and mechanically dissociated in PBS before DNA extraction.

Multiplex PCR assay analysis: multiplex PCR assays and quantification analysis were done as described in [9,33,34].

Quantitative PCR: Real time PCR were performed on a Light CyclerTM (Roche diagnostics, Meylan, France). The specificity of the unique amplification product was determined by melting curve analysis and migration on agarose gels followed by southern blotting with the corresponding internal V probe [9].

Description of the computational model:

We have built a minima a series of models assuming progressively (if a step is not available) more and more constraints:

- 1) The only influence of the Brownian thermal agitation for fixing the length of the DNA loop to splice
 - 2) The creation of a first not random window based on the differential opening of both the V and J regions, and the creation of secondary windows if the first is insufficient to fit the experimental data and to explain the non-uniform V-J usage observed [35].
-
- 1) **The Brownian ratchet model** consists in the random choice of the pivotal point for a loop involving V and J regions (see additional data), this pivot moving under

thermal agitation for a certain lapse of time during the DNA double strand splicing before the final excision step by a restriction enzyme, making use of the Brownian Motor Theory [36]: the DNA as a string can be folded to look like a loop with two strands crossing each other at the pivotal point P, one strand being made of the V locus and the other of the J locus. We calculated 1 million of cycles of thermal agitation corresponding to 1 million of rearrangements to study their repartition on a graph.

- 2) **The sequential windowing model** is a specific model of successive windows, each step corresponding to a differential motion of the window extremities, with opening velocity faster on the V side than on the J one. The simulation of the V-J rearrangements in the TRA/TRD locus is based on a computational occurrence discrete model using parameters determined from experiences (Fig. 10). The model consists in dynamical rules depending on constant (structural) and experimental parameters. The constant parameters are physical positions of 1) the 100 V and 49 J functional genes, and 2) the first TRA/TRD locus opening points. The varying parameters (whose sensitivity will be studied by simulation) are 3) the opening speeds of the V and J regions, 4) the opening offset and duration of the maturation phases of a given rearrangement and 5) its opening location. Thus, when a productive recombination occurs at a precise time in the simulation, the corresponding V and J genes are counted as rearranged.

We define variables and dynamical rules of the model as follows:

- 1) **Physical positions of V and J genes in the TRA/TRD locus** are calculated by measuring the genomic distance from the TRAC chain.
- 2) **Opening location** and first maturation characterize the site where the opening mechanism begins. This site is fixed at the TEA location [37].
- 3) **Opening speeds** of the V and J regions calculated above are denoted respectively by S_V and S_J . They are beta-like random variables between a minimum and a maximum.
- 4) N denotes the number of authorized **secondary rearrangements** during the simulation. The **probability to perform a productive rearrangement** at any step k ($1 \leq k \leq N$) is fixed to 1/3 in order to be consistent with an “in frame” rearrangement. If

5) We will call LV_k (respectively LJ_k) the length of the V (resp. J) window of accessible DNA at step k, this window advances from the V (resp. J) proximal to the V (resp. J) distal extremities of the TRA/TRD locus. The LV's verify the equations:

$$LV_1 = S_V (t_0 + \tau_1), LJ_1 = S_J (t_0 + \tau_1),$$

$$LV_k = LV_{k-1} + S_V \tau_k, LJ_k = LJ_{k-1} + S_J \tau_k, \text{ for } k \geq 2,$$

where the opening offset time t_0 denotes a minimal time of opening and τ_k ($k \geq 1$) are random variables uniformly distributed between t_0 and the opening period end, thus avoiding assumptions on the time lag separating two rearrangements.

6) For each rearrangement (productive or not), we define for each gene V_i (resp. J_j) a Boolean variable BV_{ik} (resp. BJ_{jk}) equal at the k^{th} rearrangement to 1 if the gene is open ("accessible"), and to 0 if it is closed ("non-accessible") or if it has been deleted during a previous rearrangement of order i ($i < k$).

7) The RSS score KV_i (resp. KJ_j) represents for each gene V_i (resp. J_j) the homology percentage compared to a consensus sequence. This score basically takes value p , $0 \leq p \leq 1$, if there is p % of identity between the RSS and the consensus proposed by (Glusman, 2001), allowing us to estimate a RSS identity score ranging from 0.3 to 1, where 1 corresponds to a fully consensus RSS (see additional data). The functional status for all J genes is then described (Glusman, 2001). Our RSS score is in agreement with the status of functional versus pseudo V or J gene ($0.3 < \text{pseudo rearrangement score} < 0.65$; $0.65 < \text{functional score} < 1$). According to its non-functional status, a J-pseudo recombination is never found rearranged and consequently the corresponding J-RSS score is assimilated to zero in simulations. RSS score is equal to 1 when simulation is done without taking account the RSS.

We call FV_k (resp. FJ_k) the distribution function (i.e., the relative length) obtained after the k^{th} rearrangement by adding the BV (resp. BJ) variables:

$$FV_k(i) = \sum_{m=1, \dots, i} KV_m BV_{mk} / (\sum_{m=1, \dots, 104} KV_m BV_{mk})$$

$$FJ_k(j) = \sum_{m=1, \dots, j} KJ_m BJ_{mk} / (\sum_{m=1, \dots, 60} KJ_m BJ_{mk})$$

8) At step k , we choose the distribution functions FV_k and FJ_k corresponding to the random variables RV_k (resp. RJ_k) uniform on $[0,1]$ and we then calculate a number NV_k (resp. NJ_k) equal to $\inf(FV_k^{-1}(RV_k))$ (resp. $\inf(FJ_k^{-1}(RJ_k))$). NV_k and NJ_k correspond to the V and J genes to rearrange.

9) **Start and stop the rearrangement mechanism in the simulation.** To construct a model fitting our experimental data, the simulation starts from an equivalent of F18 (18th fetal day) and stops at least at D1 (first neonatal day), corresponding to the TRA/TRD locus set up.

In fact, the rearrangement mechanism does not use such a long time in our model.

The **number of simulated rearrangements** gives the size of the simulated population. In the figures shown in this paper, 1 million of V-J rearrangements have been simulated.

The **simulation Output** is presented in matrix form incremented by the successive V-J productive rearrangements corresponding to the successive individuals of the simulated population. Final results show the total number of V-J combinations available at the end of the whole simulation. It is also possible to display the results for all the family V14 corresponding to the real time PCR's.

Remark :

The first maturation time is fixed before the first rearrangement: its duration value is a constant parameter, called the "first maturation".

References

1. Cobb RM, Oestreich KJ, Osipovich OA, Oltz EM (2006) Accessibility control of V(D)J recombination. *Adv Immunol* 91: 45-109.
2. Jouvin-Marche E, Aude-Garcia C, Candeias S, Borel E, Hachemi-Rachedi S, et al. (1998) Differential chronology of TCRADV2 gene use by alpha and delta chains of the mouse TCR. *Eur J Immunol* 28: 818-827.
3. Rytönen M, Hurwitz JL, Tolonen K, Pelkonen J (1994) Evidence for recombinatorial hot spots at the T cell receptor J alpha locus. *Eur J Immunol* 24: 107-115.
4. Abarrategui I, Krangel MS (2007) Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination. *Embo J* 26: 4380-4390.
5. Villey I, Caillol D, Selz F, Ferrier P, de Villartay JP (1996) Defect in rearrangement of the most 5' TCR-J alpha following targeted deletion of T early alpha (TEA): implications for TCR alpha locus accessibility. *Immunity* 5: 331-342.
6. Baum TP, Pasqual N, Thuderoz F, Hierle V, Chaume D, et al. (2004) IMGT/GeneInfo: enhancing V(D)J recombination database accessibility. *Nucleic Acids Res* 32: D51-54.
7. Thompson SD, Pelkonen J, Rytönen M, Samaridis J, Hurwitz JL (1990) Nonrandom rearrangement of T cell receptor J alpha genes in bone marrow T cell differentiation cultures. *J Immunol* 144: 2829-2834.
8. Thompson SD, Pelkonen J, Hurwitz JL (1990) First T cell receptor alpha gene rearrangements during T cell ontogeny skew to the 5' region of the J alpha locus. *J Immunol* 145: 2347-2352.
9. Pasqual N, Gallagher M, Aude-Garcia C, Loiodice M, Thuderoz F, et al. (2002) Quantitative and qualitative changes in V-J alpha rearrangements during mouse thymocytes differentiation: implication for a limited T cell receptor alpha chain repertoire. *J Exp Med* 196: 1163-1173.
10. Aude-Garcia C, Gallagher M, Marche PN, Jouvin-Marche E (2001) Preferential ADV-AJ association during recombination in the mouse T-cell receptor alpha/delta locus. *Immunogenetics* 52: 224-230.
11. Krangel MS (2003) Gene segment selection in V(D)J recombination: accessibility and beyond. *Nat Immunol* 4: 624-630.
12. Oltz EM (2001) Regulation of antigen receptor gene assembly in lymphocytes. *Immunol Res* 23: 121-133.
13. Davodeau F, Difilippantonio M, Roldan E, Malissen M, Casanova JL, et al. (2001) The tight interallelic positional coincidence that distinguishes T-cell receptor Jalpha usage does not result from homologous chromosomal pairing during ValphaJalpha rearrangement. *Embo J* 20: 4717-4729.

14. Sleckman BP, Bardon CG, Ferrini R, Davidson L, Alt FW (1997) Function of the TCR alpha enhancer in alphabeta and gammadelta T cells. *Immunity* 7: 505-515.
15. Hawwari A, Bock C, Krangel MS (2005) Regulation of T cell receptor alpha gene assembly by a complex hierarchy of germline Jalpha promoters. *Nat Immunol* 6: 481-489.
16. Hawwari A, Krangel MS (2007) Role for rearranged variable gene segments in directing secondary T cell receptor {alpha} recombination. *Proc Natl Acad Sci U S A* 104: 903-907.
17. Mauvieux L, Villey I, de Villartay JP (2003) TEA regulates local TCR-Jalpha accessibility through histone acetylation. *Eur J Immunol* 33: 2216-2222.
18. Lefranc MP (2001) IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res* 29: 207-209.
19. Rytönen MA, Hurwitz JL, Thompson SD, Pelkonen J (1996) Restricted onset of T cell receptor alpha gene rearrangement in fetal and neonatal thymocytes. *Eur J Immunol* 26: 1892-1896.
20. Gahery-Segard H, Jouvin-Marche E, Six A, Gris-Liebe C, Malissen M, et al. (1996) Germline genomic structure of the B10.A mouse Tcra-V2 gene subfamily. *Immunogenetics* 44: 298-305.
21. Bouchiat C, Wang MD, Allemand J, Strick T, Block SM, et al. (1999) Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys J* 76: 409-413.
22. Glusman G, Rowen L, Lee I, Boysen C, Roach JC, et al. (2001) Comparative genomics of the human and mouse T cell receptor loci. *Immunity* 15: 337-349.
23. Livak F, Burtrum DB, Rowen L, Schatz DG, Petrie HT (2000) Genetic modulation of T cell receptor gene segment usage during somatic recombination. *J Exp Med* 192: 1191-1196.
24. Guo J, Hawwari A, Li H, Sun Z, Mahanta SK, et al. (2002) Regulation of the TCRalpha repertoire by the survival window of CD4(+)CD8(+) thymocytes. *Nat Immunol* 3: 469-476.
25. Osipovich O, Milley R, Meade A, Tachibana M, Shinkai Y, et al. (2004) Targeted inhibition of V(D)J recombination by a histone methyltransferase. *Nat Immunol* 5: 309-316.
26. Huang J, Muegge K (2001) Control of chromatin accessibility for V(D)J recombination by interleukin-7. *J Leukoc Biol* 69: 907-911.
27. Rytönen-Nissinen M, Hurwitz JL, Pelkonen S, Levelt C, Pelkonen J (1999) Early activation of TCR alpha gene rearrangement in fetal thymocytes. *Eur J Immunol* 29: 2288-2296.
28. Petrie HT, Livak F, Schatz DG, Strasser A, Crispe IN, et al. (1993) Multiple rearrangements in T cell receptor alpha chain genes maximize the production of useful thymocytes. *J Exp Med* 178: 615-622.
29. Wang F, Huang CY, Kanagawa O (1998) Rapid deletion of rearranged T cell antigen receptor (TCR) Valpha-Jalpha segment by secondary rearrangement in the thymus: role of continuous rearrangement of TCR alpha chain gene and positive selection in the T cell repertoire formation. *Proc Natl Acad Sci U S A* 95: 11834-11839.

30. Wilson A, MacDonald HR, Radtke F (2001) Notch 1-deficient common lymphoid precursors adopt a B cell fate in the thymus. *J Exp Med* 194: 1003-1012.
31. Egerton M, Scollay R, Shortman K (1990) Kinetics of mature T-cell development in the thymus. *Proc Natl Acad Sci U S A* 87: 2579-2582.
32. Huesmann M, Scott B, Kisielow P, von Boehmer H (1991) Kinetics and efficacy of positive selection in the thymus of normal and T cell receptor transgenic mice. *Cell* 66: 533-540.
33. Fuschiotti P, Pasqual N, Hierle V, Borel E, London J, et al. (2007) Analysis of the TCR alpha-chain rearrangement profile in human T lymphocytes. *Mol Immunol* 44: 3380-3388.
34. Mancini SJ, Candeias SM, Di Santo JP, Ferrier P, Marche PN, et al. (2001) TCRA gene rearrangement in immature thymocytes in absence of CD3, pre-TCR, and TCR signaling. *J Immunol* 167: 4485-4493.
35. Gao EK, Kanagawa O, Sprent J (1989) Capacity of unprimed CD4+ and CD8+ T cells expressing V beta 11 receptors to respond to I-E alloantigens in vivo. *J Exp Med* 170: 1947-1957.
36. Zandi R, Reguera D, Rudnick J, Gelbart WM (2003) What drives the translocation of stiff chains? *Proc Natl Acad Sci U S A* 100: 8649-8653.
37. de Chasseval R, de Villartay JP (1993) Functional characterization of the promoter for the human germ-line T cell receptor J alpha (TEA) transcript. *Eur J Immunol* 23: 1294-1298.

Figure legends

Figure 1. Quantification of J region utilization by V14 family. 260 V14

rearrangements were cloned from T lymphocytes, the V14 members and J genes were determined by sequencing [10]. (A) Profile of the J use by the six members of the V14 family. The two arrows indicate the localization of the two hot spots. (B) Profile of the three members the nearest from J genes and (C) J use by the most 5' V14 members.

Figure 2. Validation of the modeling approach by simulation analysis of the V and J regions utilization.

(A) V region utilization: the X axis represents the V region in Kb. The Y axis shows the V gene percentage utilization in simulation. The simulated data sets have been normalized in order to be compared according to the following formula

$$X=(x-\text{average})/\text{Std deviation.}$$

The fixed parameters of the simulation were as follow, one million of alpha chains, ongoing 1 to 4 rearrangements with opening speeds of 18 Kb/h and 1.03 Kb/h for the V and the J region respectively; (B) J region utilization: the X axis represents the J region in Kb; (C) and (D) Amplitude of J region utilization by opposite V genes, V1 (distal) and V21 (proximal). The X axis represents experimental quantification on 9 J genes. The Y axis shows the experimental utilization frequency of 9 J genes by the V1 and V21 genes. (E) and (F) Amplitude of J region utilization in the model. The X axis represents the J genes. The Y axis shows the model frequency utilization by each J genes. (G) and (H) Superposition of experimental and simulated data for the 9 J genes. The X axis represents experimental quantification on 9 J genes. V and J regions utilization from simulated data are similar to experimental data.

Figure 3. Model interface and results. (A) The main user interface window of the simulation program, (B) 2D representation of the rearrangement density, (C) 3D representation of the rearrangements density over all couples of V and J genes.

Figure 4. Representation of rearrangement frequencies of the V14 family. Y axis represents the cumulated frequencies of all V14 genes with the J genes presented

on the X axis, (A) without correction for RSS scores, (B) with correction accordingly to RSS scores. The two red ellipses show the localization of the two hot spots of recombination.

Figure 5. 3-D superposition of V14 family rearrangements. (A) The fitted simulated data and (B) non fitted simulated data are shown in grey shapes. In red are represented the experimental points above the simulation shape. In green are represented the experimental points under the simulation shape.

Figure 6. Simulation of secondary rearrangements. Three successive draws of random integers were done successively, the first giving an integer x_1 between 0 and 10 following a Poissonian law. The second and the third follow a Gaussian law, the second giving an integer x_2 between x_1 and $x_1 + 10$, and the third giving an integer x_3 between x_2 and $x_2 + 10$, and that 300 000 times. The first, second and third curves were added to build the sum curve.

Figure 7. Density and RSS scores of J genes. Values for the density (open squares) and RSS scores (dark circles) were calculated, as described in methods, for each J gene from the four previous and next genes. X axis represents the J genes, the Y axis the density or the RSS score for all J genes.

Figure 8. Simulation of the diversity of V-J combinations.

X axis represents the number of TR tested in the simulation, and Y axis indicates the percentage of the number of the different V-J combinations obtained by the simulation over the total number of possible V-J combinations.

Figure 9. 3D representation of the TRA/TRD locus utilization. The graph shows the rearrangement distribution of 4 V genes with all J genes. The dashed red arrow indicates the decreasing frequency of rearrangements which correlates with the associations of distal V and J genes. A distal V gene is very rarely rearranged with proximal J genes because of the high recombination frequency between proximal V and J genes, leading to the proximal J genes "erosion". 1) the first step governing the

TRA/TRD locus utilization is defined by the opening of the most proximal V-J region, 2) According that T cells can undergo 1 to 3 secondary rearrangements, the second step giving the V and J accessible windows, is defined by the opening speeds of V and J regions. V speed is about 18 Kb/h whereas J speed is 18 fold lower with about 1 Kb/h, 3) J region has two hot spots of rearrangement. HSI is centred on J48 and HSII is centred on J30.

Figure 10. Flow diagram for the algorithm of the sequential windowing model

Tables

Table 1. TRA/TRD locus characterization

	V region	J region
Length (Kb)	1300	64
Number of functional element	104	60
Number of functional element	100	49

Table 2. Variations of the V19 gene rearrangements to J genes during ontogeny

Gestation day [§]	J opening	Opening distance [#]	Maximal opening Speed
F18 to F19	J61 to J48	14773 bp	615 bp/h
F19 to F20	J47 to J20	27618 bp	1150 bp/h
F20 to D0	J19 to J9	8998 bp	375-750 bp/h *
D0	J8 to J2	7940 bp	333 bp/h

[§] Thymus from fetal day 18 (F18) to day of birth (D0), data are analyzed from Fig. 5 in Pasqual et al. [9].

[#] Length of the DNA sequence including in the J opening.

* For F20, the opening speed has been estimated between 375bp/h to 750pb/h depending on the offset of maxima 12h (9000bp/24h or 9000bp/12h).

Table 3. Contribution of each rearrangement in the total V-J rearrangements

Rearrangement	First	Second	Third	Fourth
Percentage	40	33	19	8

Table 4. Comparison of the utilization obtained from experimen and simulation.

	Experiment	Simulation	
	V14 *	V14 [#]	All V [#]

J61 to J48	37%	35%	33%
J47 to J24	46%	50%	51%
J23 to J1	17%	15%	16%

*Frequencies of rearrangements of V14 genes were calculated from the figure 2 in Aude-Garcia *et al.* [10], for the combinations with three J panels, corresponding to series of J genes scattered along the J region.

#Frequencies of rearrangements of V14 genes and of all V genes were calculated from simulated data for the combinations with same series of J genes.

Table 5. V-J combination probability along the TRA locus.

	V21 Proximal	V4-2 Central	V2 Distal
J52 Proximal	0.148	0.050	0
J31 Central	0.002	0.029	0.023
J2 distal	0	0.003	0.007

Figure 1

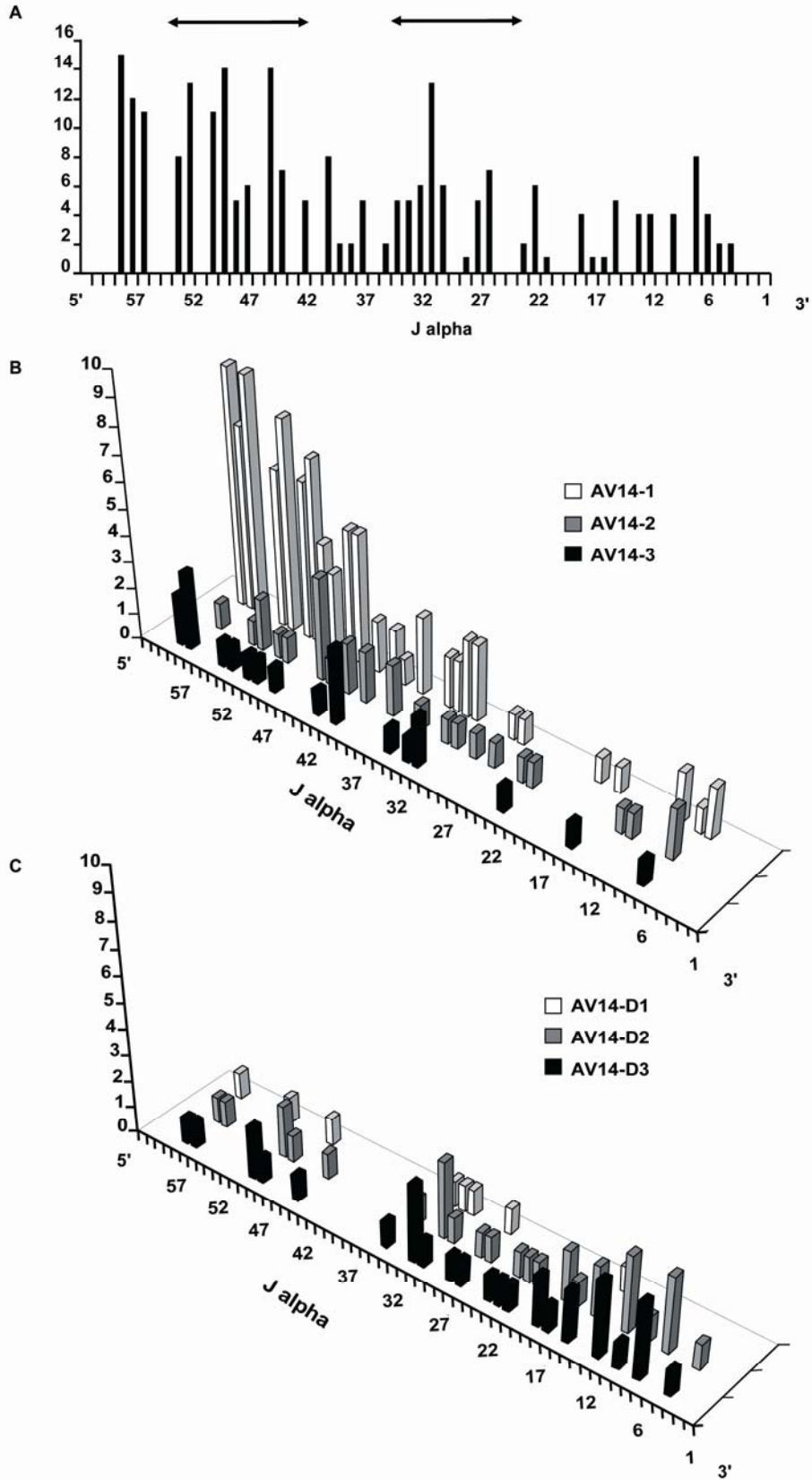


Figure 2

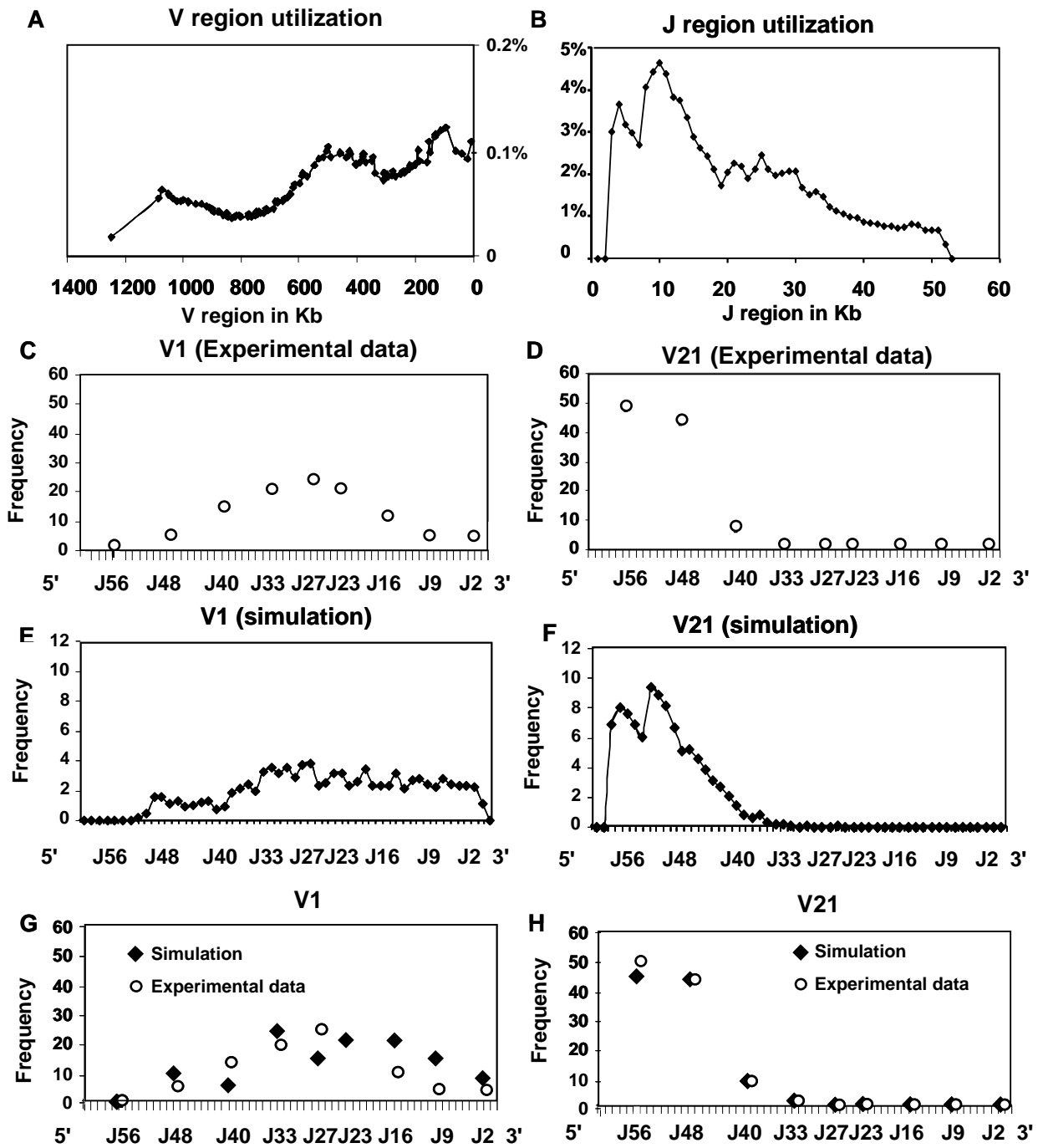
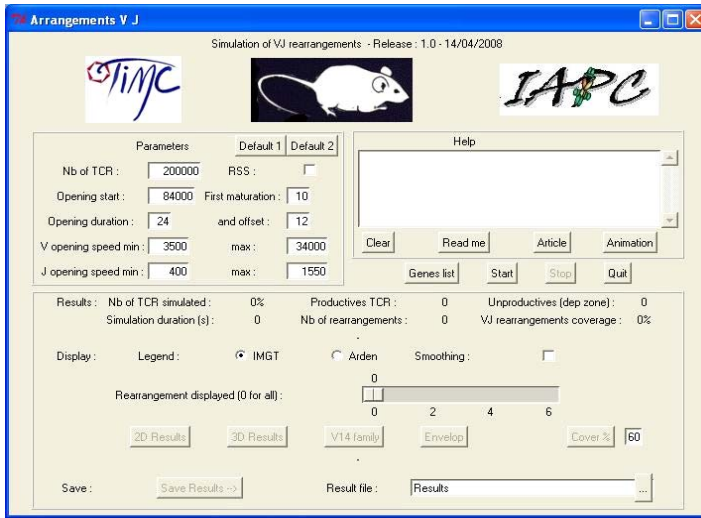
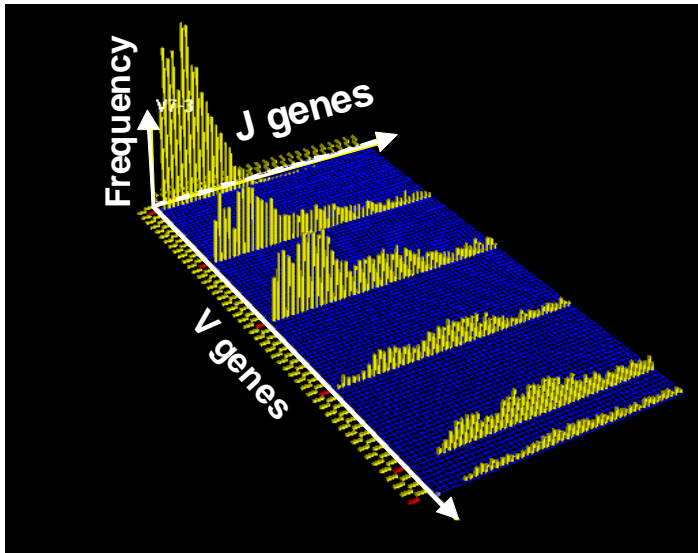


Figure 3

A



B



C

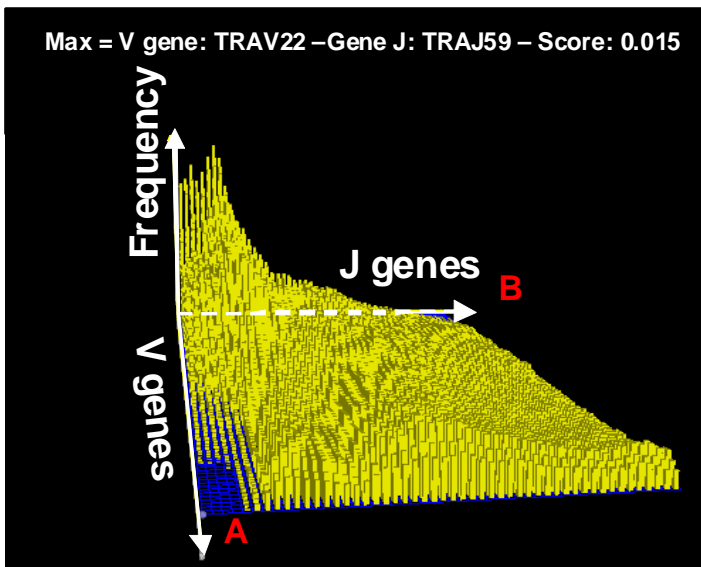


Figure 4

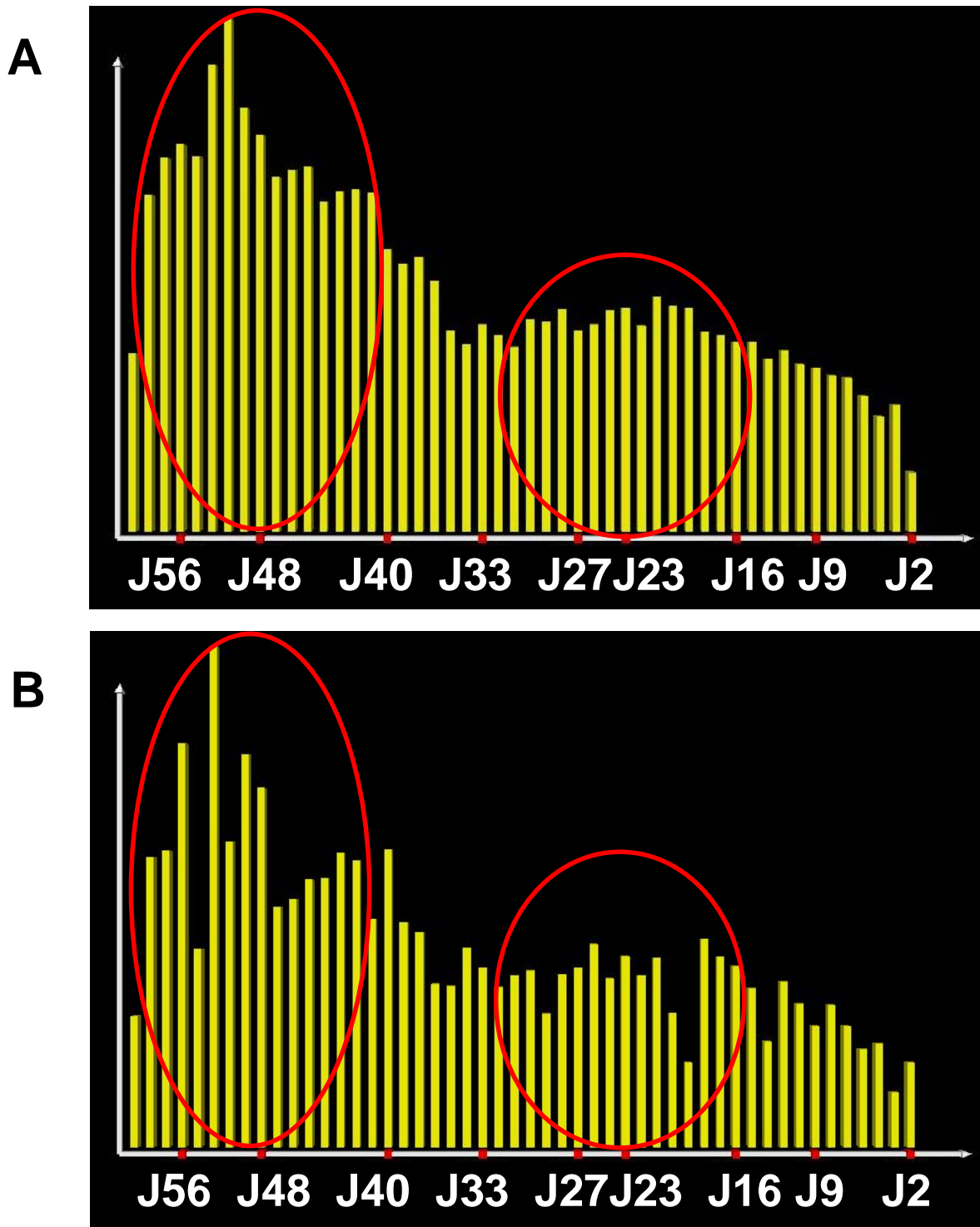


Figure 5

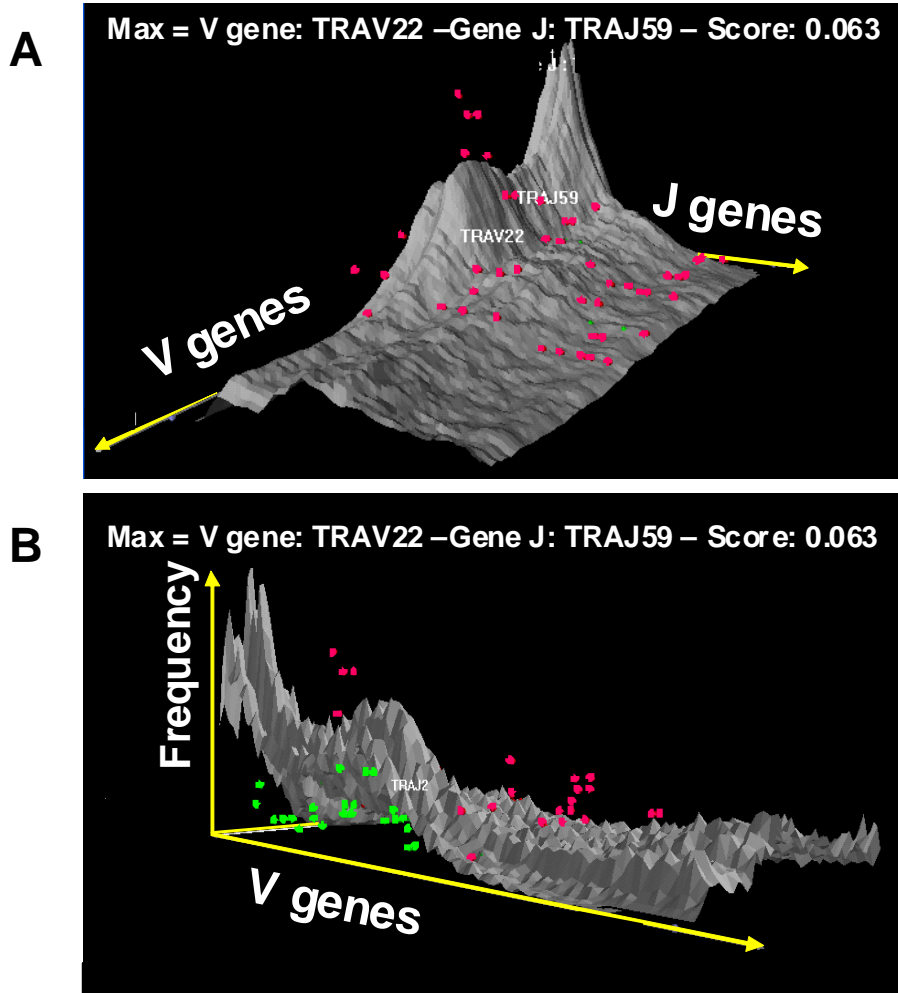


Figure 6

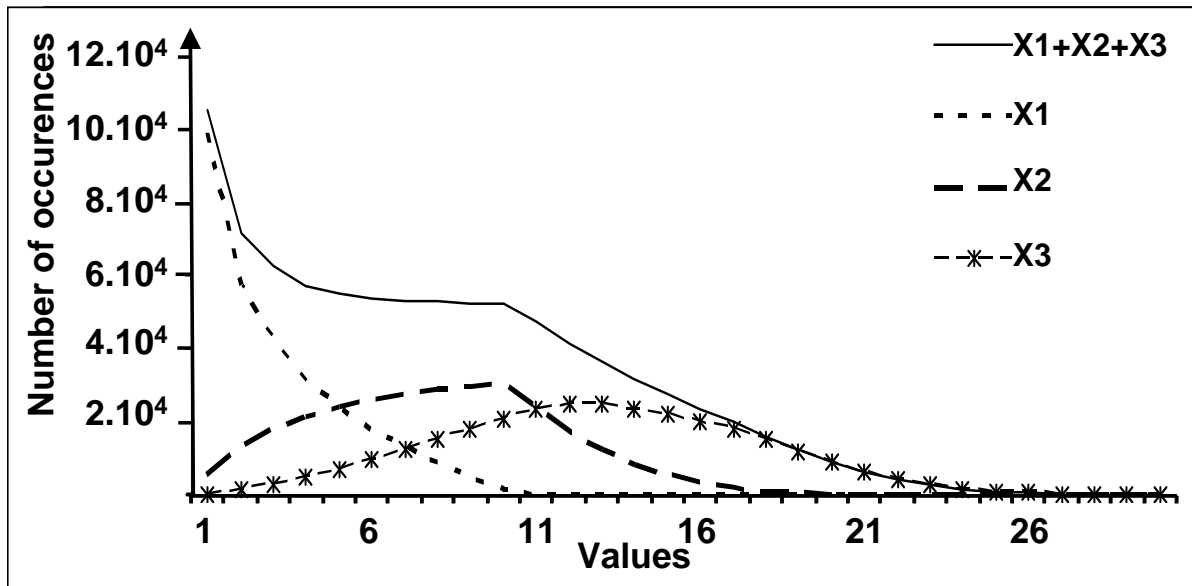


Figure 7

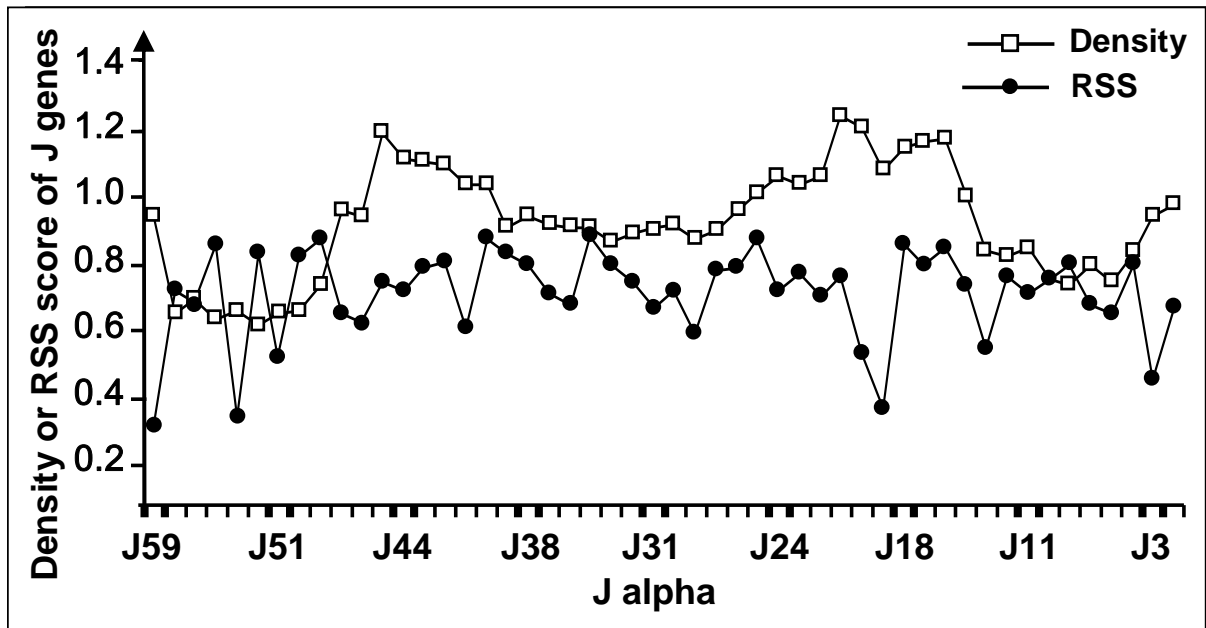


Figure 8

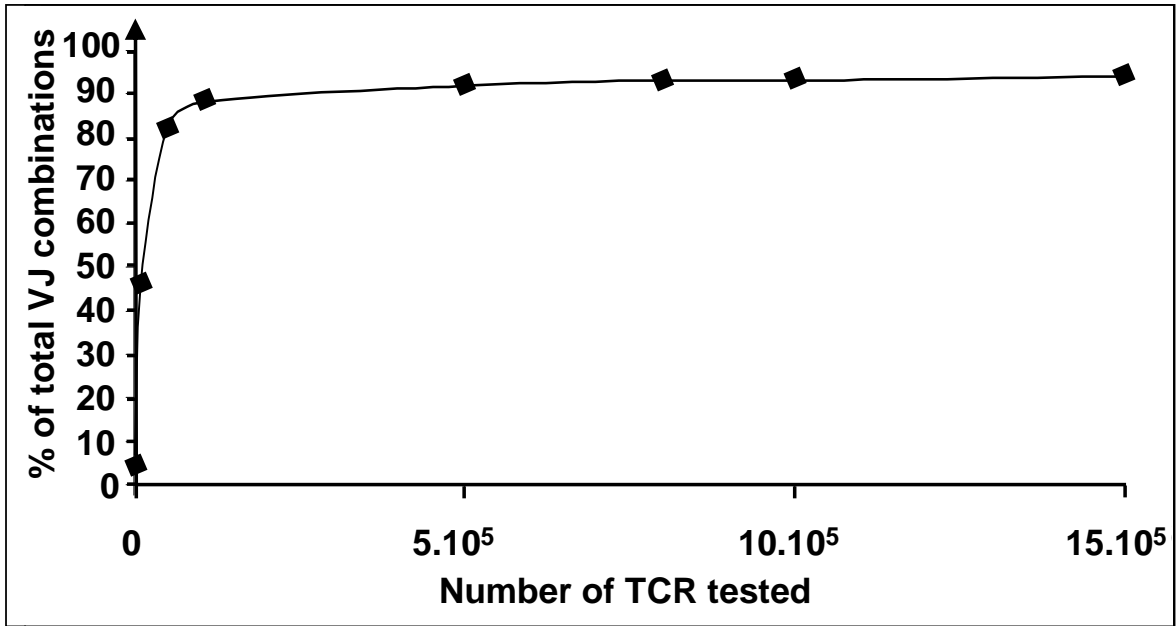


Figure 9

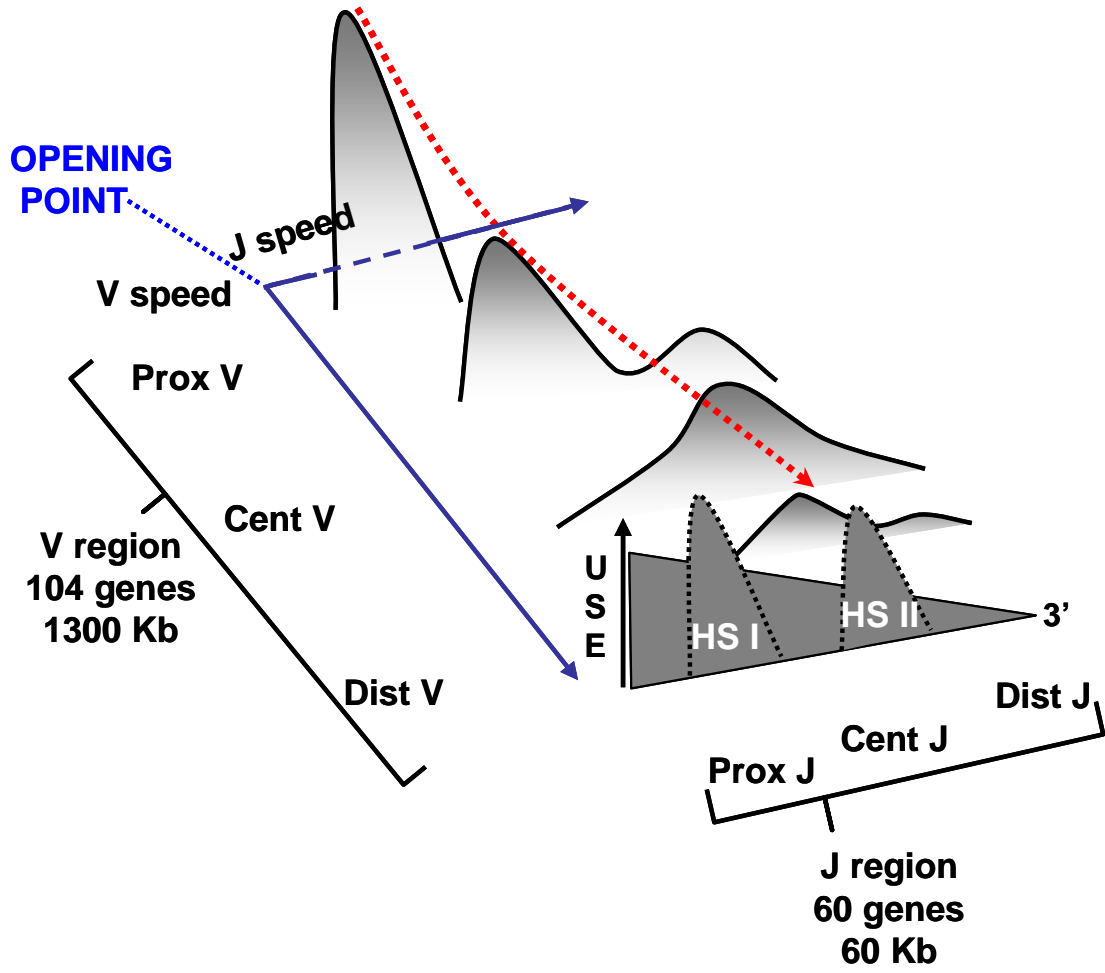
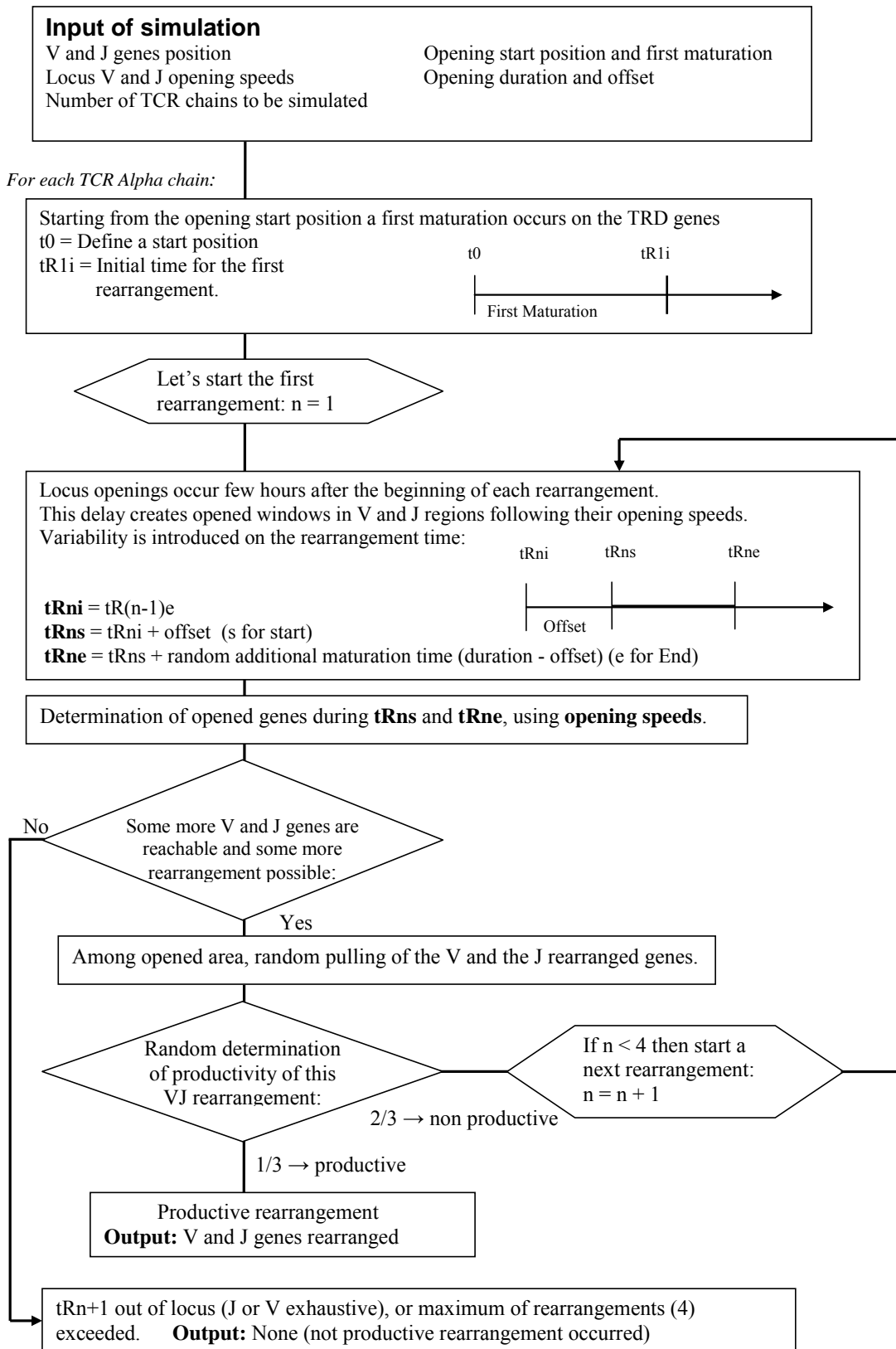


Figure 10



Supporting materials and methods:

Model 1

The first step in studying VJ rearrangements is to see if we can find a law as simple as possible to predict them, such as a probabilistic law, related to a pure random mechanism. When we plot, for example, V1 (resp. V21) against J genes for Balb-c TCR α chains rearrangements, we can see a Gaussian-like (resp. Poisson-like) curve (Fig. S2). Thus, we tried first to see if a simple probabilistic model would fit such a curve. We called the corresponding model the Model 1. Model 1 makes use of the Brownian Motor Theory or of the Worm-like Chain Theory. In this model, we see DNA as a string which can be folded to look like a loop with two strands crossing each other. One strand is made from the V locus and the other from the J locus (Fig. S1). The V and J loci cross each other at a point we call P . We suppose that a rearrangement is made of the V and J genes located at the P point. We will suppose that the V and J loci move under constraints like thermal agitation sliding through the *pivot* P . The way they slide will be described by Model 1, according to the Brownian Motor Theory, “click by click”. A click is a unit of displacement along one of the V or J strands. Each displacement unit is supposed to be a V or a J gene. Thus, one can move along the V locus from the most proximal V (V_{prox}) to the most distal V (V_{dist}) with 104 possible consecutive clicks corresponding to the 104 V genes on the V strand. The same is available for the 60 genes of the J locus. We call $p_{V_{dist}}$ the probability of displacement of the P point towards the distal V gene (V_{dist}) and $p_{V_{prox}}$ the probability of displacement of the P point towards the proximal V gene (V_{prox}) (idem for the J locus, see Fig. S1). We call V_{dist} the distance between V_{dist} and P , and V_{prox} the distance between V_{prox} and P (idem for the J locus). We call V the position of P along the V strand, and J the position of P along the J strand. If we look the biological results, we can see that proximal V and J genes are used preferentially compared to distal V and J genes. Thus, we can say that the displacement of the pivot P is proportional to V_{dist} and J_{dist} distances, and inversely proportional to V_{prox} and J_{prox} distances. If we consider that the displacement of the P point is random and if we call α the proportionality factor, we can write the following equations:

$pVprox$ is proportional to $Jdist$ and $Vdist$: $pVprox = \alpha * Jdist * Vdist$

$pVdist$ is proportional to $Jprox$ and $Vprox$: $pVdist = \alpha * Jprox * Vprox$

$$pVprox + pVdist = 1 = \alpha * [(Jdist * Vdist) + (Jprox * Vprox)]$$

From these relations we can infer the following equations:

$$pVprox(k) = (Vdist(k-1) * Jdist(k-1)) / ((Jprox(k-1) * Vprox(k-1)) + (Vdist(k-1) * Jdist(k-1)))$$

$$pVdist(k) = (Vprox(k-1) * Jprox(k-1)) / ((Jprox(k-1) * Vprox(k-1)) + (Vdist(k-1) * Jdist(k-1)))$$

$$pJprox(k) = pVprox(k) \text{ and } pJdist(k) = pVdist(k)$$

From these equations, we applied the following algorithm for the displacement of P:

Initial conditions: we randomly chose:

- V: initial position of P on the V strand
- J: initial position of P on the J strand

For each Brownian click, we make 2 random choices by using 2 random variables $T1$ and $T2$ supposed to be uniform on $[0,1]$:

- $T1$ for $pVprox$:
 - o If $T1 \leq pVprox$: P is moved by 1 click towards $Vprox$
 - o If $T1 > pVprox$: P is moved by 1 click towards $Vdist$
- $T2$ for $pJprox$:
 - o If $T2 \leq pJprox$: P is moved by 1 click towards $Jprox$
 - o If $T2 > pJprox$: P is moved by 1 click towards $Jdist$

We repeat $T1$ and $T2$ 3, 10 and 50 times, corresponding to 3, 10 and 50 Brownian clicks.

The position of P at the end of the series of clicks is the rearrangement position.

We calculated 1 million of cycles corresponding to 1 million of rearrangements to study their repartition on a graph.

Results

The curves below (Fig. S2) have been obtained by taking as number of clicks successively 3, 10, 30 and 50.

Discussion

From the curves above (Fig. S2), we can see that rearrangements are possible only for a limited number of clicks. Above 30 clicks for the same V21-V1 rearrangement, there is no more evolution and rearrangement is fixed. We can also observe that the curves slopes in are somehow random and do not fit with the slopes of the experimental curve (Fig. S2 of the paper) Thus, the Model 1 is unable to explain experimental curves. Rearrangements come certainly not from a pure random phenomenon. Then, we tried a more sophisticated model, the Model 2, in order to explain experimental curves.

Supporting informations

Figure S1. V and J loci crossing each other at the P point to realize rearrangements

Figure S2. Simulation rearrangements frequencies obtained for an increasing number of clicks (3 to 50)

Figure S1

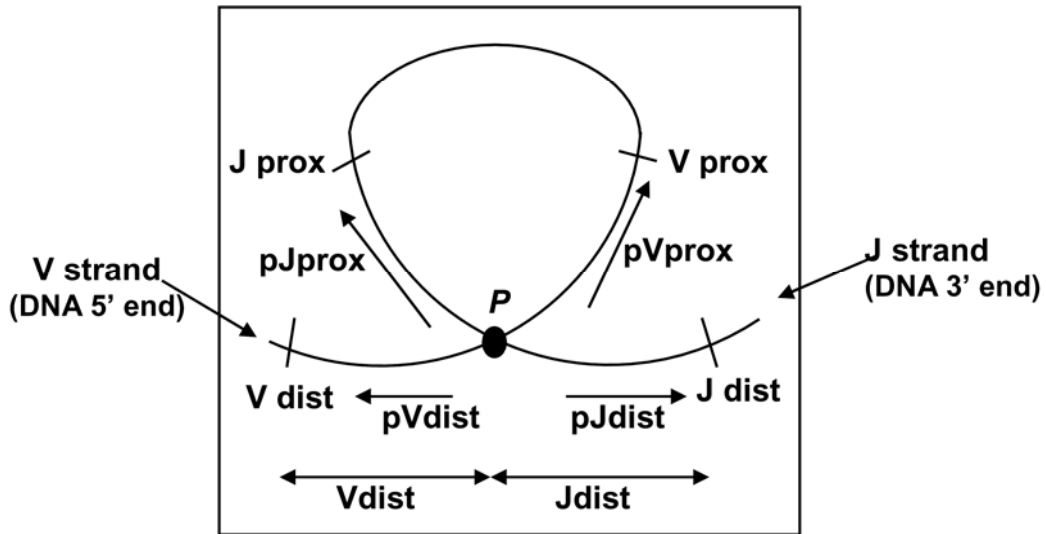
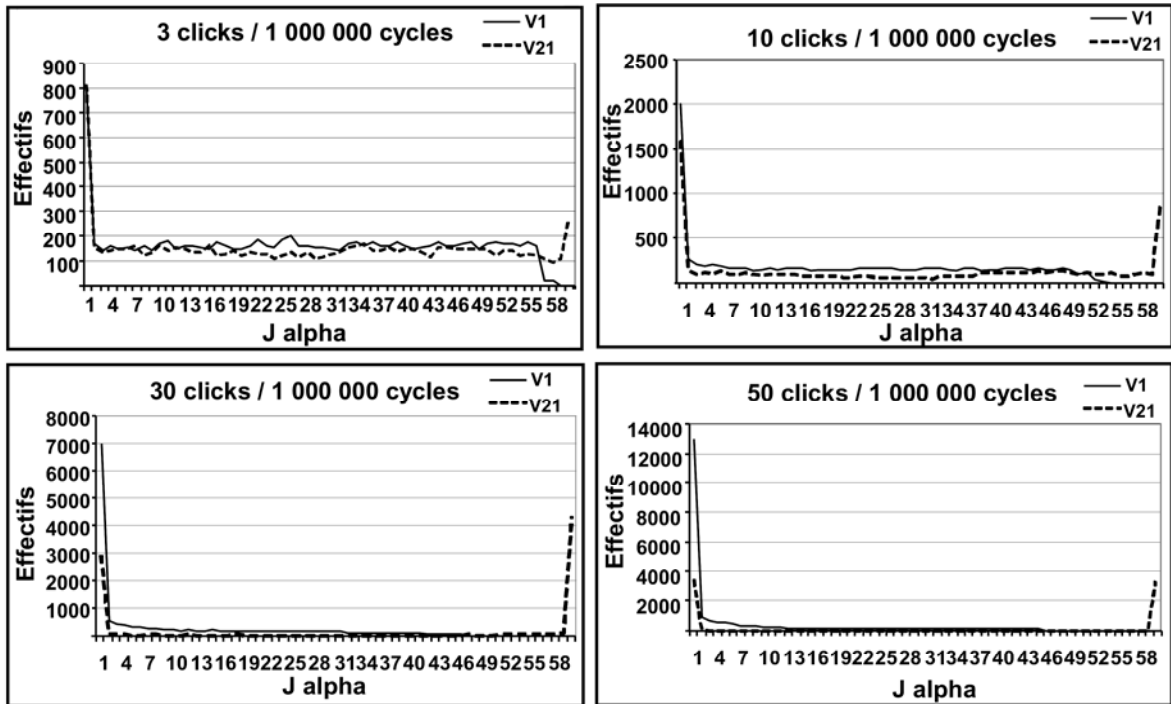


Figure S2



8. Discussion et conclusion de l'article 2

L'implémentation d'un système de modélisation des réarrangements VJ sur le locus TRA/TRD a permis de préciser les valeurs de paramètres intervenant dans la mise en place du répertoire, comme les vitesses d'ouverture et le temps de maturation. De plus, le modèle propose une estimation plus précise de la taille du répertoire combinatoire de la chaîne TR α et des fréquences des réarrangements VJ, ce qui correspond à son objectif.

A travers une méthodologie de comparaison systématique des données expérimentales aux résultats de simulation, nous avons ajusté et validé le modèle de simulation. Le modèle de simulation est en accord avec : 1) les données expérimentales concernant les profils de recombinaison des gènes V1 et V21 localisés respectivement au niveau distal et proximal de la région V, 2) les résultats de recombinaison concernant la famille V14, qui est composée de six membres localisés au niveau central de la région V et 3) rend compte des hot spots de recombinaison VJ observés sur la région J du locus TRA/TRD. Pour ce faire, nous avons, dans un premier temps, ajusté les valeurs des vitesses d'ouverture des régions V et J en modulant leur utilisation non plus comme un paramètre fixe, mais comme un paramètre variable entre un minimum et un maximum. Cet ajustement nous a permis d'augmenter la taille du répertoire observé, mais ne suffit pas pour mettre en évidence les hot spots de recombinaison. Nous avons pour cela rajouté un nouveau paramètre, qui est un temps de maturation entre deux réarrangements, laps de temps au cours duquel une nouvelle combinaison VJ ne pourra pas être engagée. Les étapes montrant l'évolution du modèle de simulation sont présentées ci-dessous.

8.1. Effet de la vitesse variable et du temps de maturation sur le profil de recombinaison

8.1.1. Effet d'une vitesse fixe sur le répertoire

Les premières simulations de notre modèle utilisaient une vitesse d'ouverture fixe pour les régions V et J. Cette vitesse était fixée par rapport aux vitesses d'ouverture moyenne calculées expérimentalement.

La figure 18 montre le profil de recombinaison en 3D obtenu pour un million de recombinaisons testées et une vitesse $S_v = 18\text{Kb/h}$ et $S_j = 1\text{Kb/h}$.

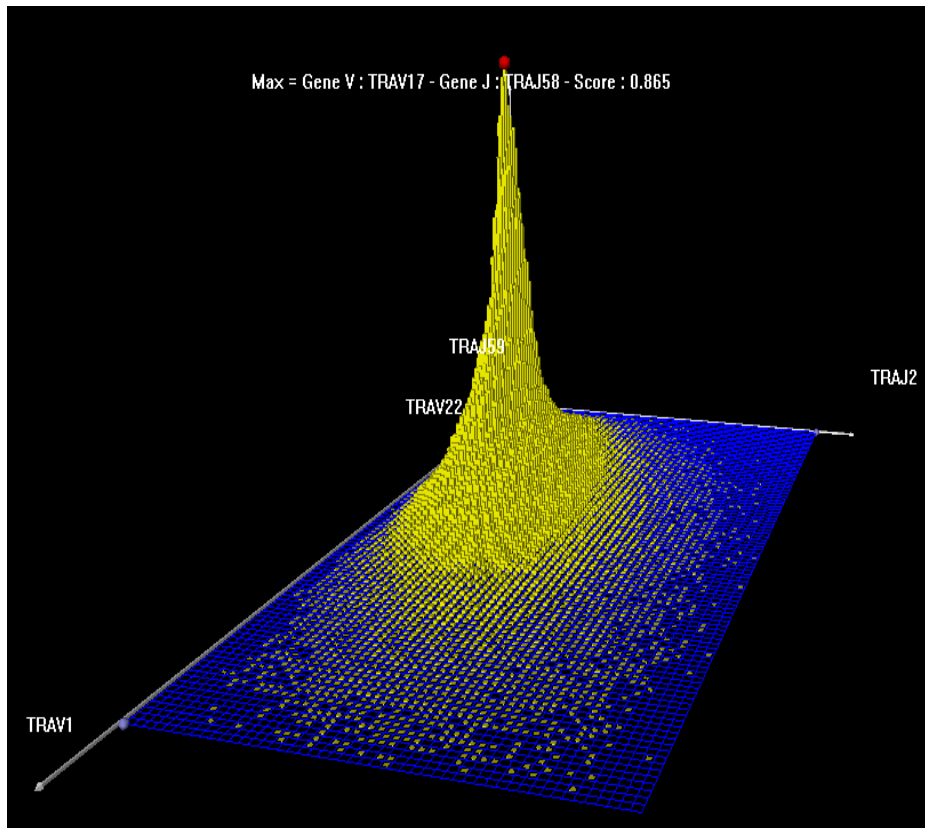


Figure 18 Représentation en 3D du répertoire obtenu par la modélisation à partir de vitesses d'ouverture fixes et sans temps de maturation.

Les résultats des simulations réalisées à partir d'une vitesse d'ouverture fixe des régions V et J, montrent des réarrangements entre les gènes V et J proximaux nettement plus fréquents que les réarrangements entre les gènes V et J distaux. L'observation générale montre également que les réarrangements se placent le long d'une diagonale aux frontières bien définies.

Cependant, sur un million de TR simulés, seulement 72% des combinaisons VJ possibles sont observées. De plus, ces résultats ne rendent compte, ni d'une utilisation plus large des recombinaisons VJ, ni des hot spots de recombinaison.

8.1.2. Effet d'une vitesse d'ouverture variable sans temps de maturation

Afin d'améliorer le modèle de simulation, et suite au réexamen de résultats expérimentaux publiés par notre laboratoire sur l'accessibilité de la région J par différents gènes V au cours de l'ontogénie, nous avons introduit une variabilité dans la vitesse d'ouverture des régions V et J.

Cette vitesse d'ouverture est tirée au sort entre un minimum et un maximum, que nous avons ajustés au cours des simulations. Les vitesses moyennes d'ouverture observées par simulation (dite aussi expérimentation numérique) restent très proches des vitesses moyennes observées par expérimentation biologique.

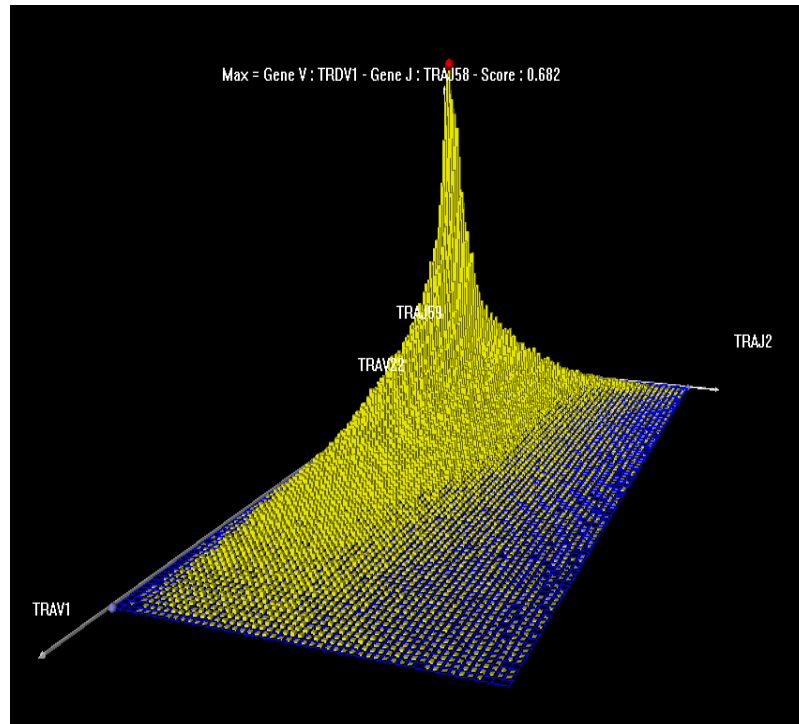


Figure 19 Représentation en 3D du répertoire obtenu par la modélisation à partir de vitesses d'ouverture variables et sans temps de maturation

La figure 19 présente les résultats d'une simulation avec des vitesses variables dont la valeur correspond aux valeurs utilisées dans l'article. La simulation montre à nouveau que tous les réarrangements ne sont pas équiprobables et favorise toujours les combinaisons VJ entre gènes V et J proximaux. Les combinaisons VJ sont maintenant condensées le long d'une région qui comprend tous les gènes V et les gènes J proximaux à centraux.

Le répertoire combinatoire est augmenté de 25%. Il est à présent de 97% de combinaisons VJ observées sur les combinaisons VJ testées. Cependant, l'introduction de ce nouveau paramètre dans la simulation ne permet toujours pas d'observer les hot spots de recombinaison, qui sont un élément important du répertoire immunitaire murin.

8.1.2.1. Effet d'une vitesse d'ouverture variable avec un temps de maturation

Nous avons alors ajouté un temps de maturation entre deux recombinaisons, temps au cours duquel des réarrangements entre les gènes V et J ne pouvaient pas être engagés. Ce nouveau paramètre est basé sur l'hypothèse qu'après un réarrangement, celui-ci doit être testé. Si ce réarrangement conduit à un cadre de lecture ouvert, il faut néanmoins que la chaîne α nouvellement synthétisée puisse s'associer avec la chaîne β et que le TR soit sélectionné, après avoir passé les étapes de sélection négative et positive. D'autre part, même si le cadre de lecture n'est pas respecté, il est probable que le double brin d'ADN doit acquérir une certaine flexibilité pour pouvoir former une boucle de recombinaison.

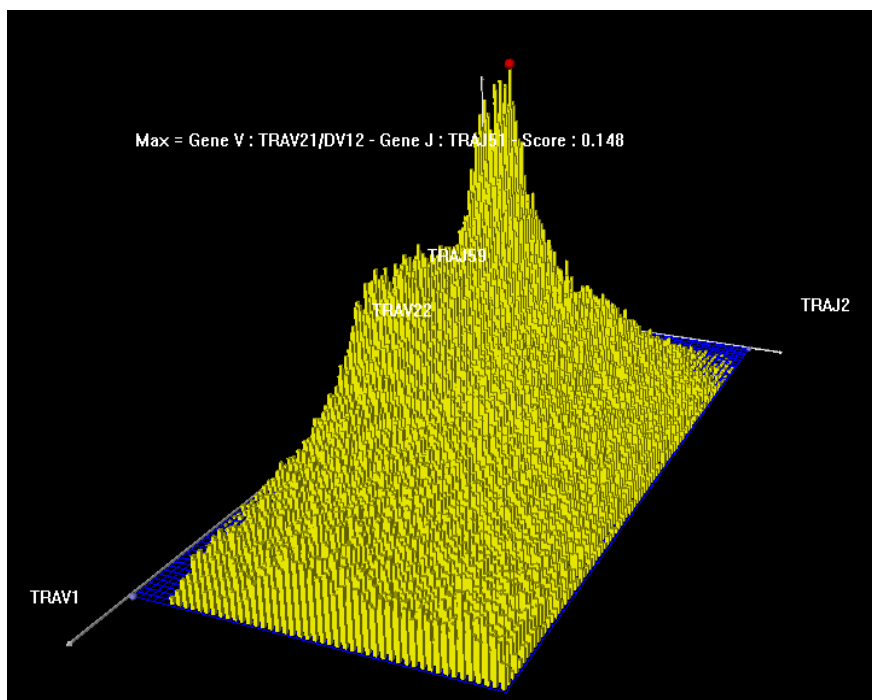


Figure 20 Représentation en 3D du répertoire obtenu par la modélisation combinant des vitesses d'ouverture variables et un temps de maturation

La figure 20 présente les résultats obtenus par simulation, en tenant compte d'une vitesse d'ouverture variable (d'un TR à l'autre) des régions V et J, et d'un temps de maturation entre deux réarrangements successifs. Ces résultats correspondent à ceux décrits dans l'article, où le modèle de simulation a été validé à partir de données expérimentales et a permis de rendre compte de la présence des hot spots de recombinaison. Le nombre de combinaisons VJ observées sur la figure 20 est quasiment identique à celui obtenu sans temps de maturation (96% vs 97%). Cependant, les profils des fréquences des combinaisons VJ sont très différents. Sur la figure 19, les combinaisons VJ sont regroupées sur une bande principale avec une

préférence pour les combinaisons entre les gènes V et J proximaux, et lorsque l'on s'éloigne de cette zone de réarrangements préférentiels, la fréquence des combinaisons VJ diminue de manière drastique. Alors que des combinaisons entre gènes V et J distaux sont observées, leur fréquence de réarrangement est très faible, en comparaison avec les combinaisons entre les gènes V et J proximaux. Sur la figure 20, le nombre de combinaisons VJ est quasiment identique à celui de la figure 19, mais la fréquence de chaque réarrangement varie de manière importante selon la position des gènes V et J le long du locus. Il n'y a plus de zone délimitée, en dehors de laquelle les combinaisons VJ peuvent être observées de manière très peu fréquente. En d'autres termes, quasiment toutes les combinaisons sont possibles, et la zone de gradient de fréquences entre les combinaisons VJ les plus élevées et les plus faibles est moins large.

8.1.3. Validation du modèle : suite

Le modèle de simulation a été validé au cours des simulations réalisées à partir de données expérimentales de notre laboratoire. Nous avons montré que les fréquences de réarrangement obtenues par le modèle de simulation sont en accord avec ces données.

Les travaux de l'équipe de Krangel ont montré que le promoteur TEA contrôle l'accessibilité de la chromatine de la région J. Le TEA permet l'ouverture de la chromatine de J61 à J49 et inhibe l'ouverture au-delà du J49. Suite à l'activation du J49, les J en aval de J49 deviennent accessibles (Abarrategui and Krangel 2007).

La Figure 21 illustre le profil de réarrangement de la famille V14 lors du premier réarrangement. Ces résultats montrent que 70% des réarrangements observés lors de ce premier réarrangement sont localisés entre J61 et J49.

Si nous confrontons ce résultat aux observations expérimentales de l'équipe de Krangel, nous constatons qu'ils correspondent à la première fenêtre d'accessibilité décrite par cette équipe.

Nous n'avons pas paramétré le modèle pour obtenir ce résultat, mais nous l'avons obtenu directement avec les valeurs des paramètres actuellement utilisés par le modèle. Ce résultat montre une fois de plus que le modèle est en accord avec les données expérimentales.

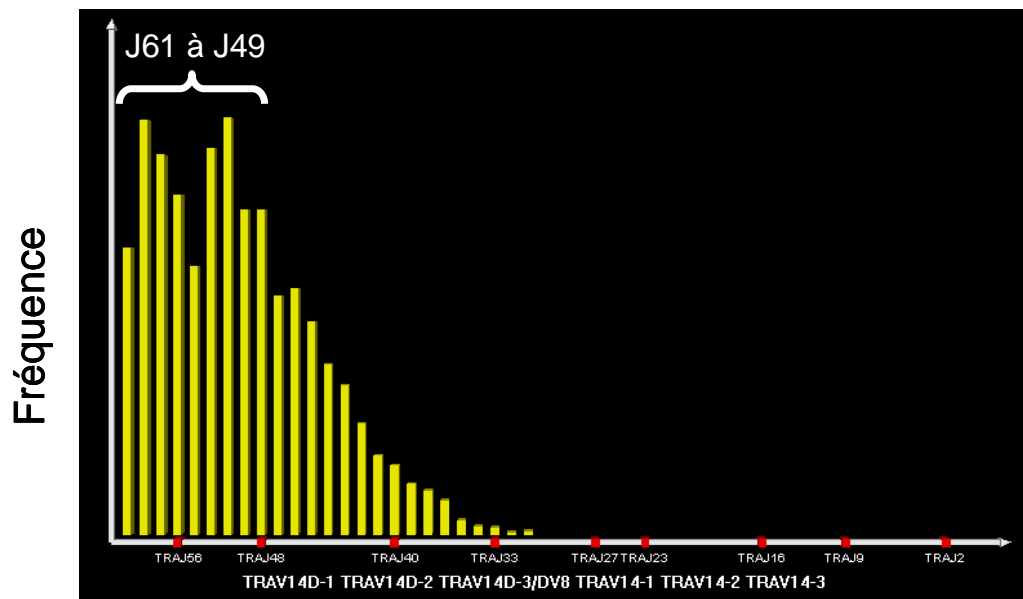


Figure 21 Profil de réarrangement obtenu lors de la première recombinaison pour la famille V14. Nous avons annoté la zone où les réarrangements J sont les plus utilisés (J61 à J49).

8.1.3.1. Proposition pour la seconde vague de recombinaison

L'accessibilité du reste de la région J n'est pas encore élucidée. Il a été montré que le promoteur J49 prend le relais pour l'ouverture de la région J, mais on ne sait pas encore quelle est sa distance d'action, ni s'il existe d'autres promoteurs relais sur le reste du locus.

Concernant les autres fenêtres d'accessibilité, notre modèle montre que, pour le second réarrangement, 80% des combinaisons VJ ont lieu entre les gènes J49 et J24. Pour le troisième et le quatrième réarrangement, notre modèle ne montre pas d'utilisation préférentielle de J, mais au contraire une utilisation uniforme (qui exclut quasiment les J proximaux compris entre J61 et J48).

9. Avancées du modèle

9.1. TR α et exclusion allélique

Au cours de leur différenciation intra-thymique, les lymphocytes T vont réarranger les gènes codant pour leurs chaînes de TR. Du fait que les cellules sont diploïdes, chaque cellule T pourrait exprimer à sa surface quatre combinaisons différentes de TR $\alpha\beta$. Cependant, des études ont montré que les gènes codant pour la chaîne β sont soumis à l'exclusion allélique, empêchant la cellule d'exprimer deux chaînes β à sa surface (Malissen, Trucy et al. 1992; Davodeau, Peyrat et al. 1995; Schlimgen, Reddy et al. 2008). En effet, l'expression d'une chaîne TR β à la surface d'un précurseur de lymphocytes T empêche le second chromosome d'engager le réarrangement d'une chaîne TR β .

Concernant la chaîne TR α , de nombreuses études ont montré qu'un thymocyte réarrange ses deux chromosomes en parallèle chez la souris (Malissen, Trucy et al. 1988; Letourneur and Malissen 1989; de Villartay and Cohen 1990; Malissen, Trucy et al. 1992) et chez l'homme (Davodeau, Difilippantonio et al. 2001).

A travers l'étude de clones de lymphocytes T murins et humain, *Davodeau et al.* (Davodeau, Difilippantonio et al. 2001) ont montré qu'il y avait une co-localisation des gènes J utilisés par les deux chromosomes chez la souris et chez l'homme. La distance moyenne des J réarrangés sur chaque chromosome est très proche entre l'homme et la souris dans les cellules T : $10,1 \pm 9,4$ J et $7,1 \pm 6,7$ J respectivement.

Contrairement à la région J, la région V ne présente pas de corrélation dans la position allélique des deux gènes V réarrangés. De plus, la distance intra-allélique de deux gènes V réarrangés est comparable à une distribution aléatoire. Enfin, il n'y a pas de corrélation dans la distance intra-allélique des gènes V et J réarrangés.

9.1.1. Avancées du modèle

Dans l'élaboration de notre programme de simulation, nous avons modélisé chaque chaîne α de manière indépendante l'une de l'autre, sans prendre en compte le fait que les utilisations des gènes J sur les deux chromosomes évoluent probablement de manière interdépendante.

Pour remédier à cet oubli, nous avons considéré que les réarrangements se faisaient toujours de manière individuelle, mais par paire (figure 22) : c'est la notion d'inclusion allélique, qui traduit l'inverse de l'exclusion allélique.

Si le premier réarrangement est "productif", il est enregistré, comme c'était déjà le cas auparavant. S'il est "non productif", alors un premier réarrangement est engagé sur le second

chromosome, donnant au thymocyte une nouvelle possibilité d'avoir un réarrangement fonctionnel. Les deux différences avec la version précédente sont :

1) nous considérons que les deux chromosomes s'ouvrent de manière synchronisée, suggérant que les mêmes gènes V et J sont accessibles sur le second chromosome. Du fait qu'il n'a pas été rapporté de corrélation dans la sélection des gènes V sur les deux chromosomes, nous laissons la possibilité au thymocyte de choisir un V dans toute la portion de la région V déjà rendue accessible.

2) de manière à tenir compte de la co-localisation des gènes J utilisés par un même thymocyte, le nouveau J sera choisi à plus ou moins cinq kilobases du J utilisé sur le premier chromosome, ce qui correspond en moyenne à $10J\alpha$ autour du J qui a été utilisé sur le premier chromosome.

Si le réarrangement sur le second chromosome est "productif", alors la simulation s'arrête. Sinon, le programme conduit à l'étape suivante, qui correspond à un temps de maturation pour permettre l'accès à une nouvelle fenêtre de gènes V et J.

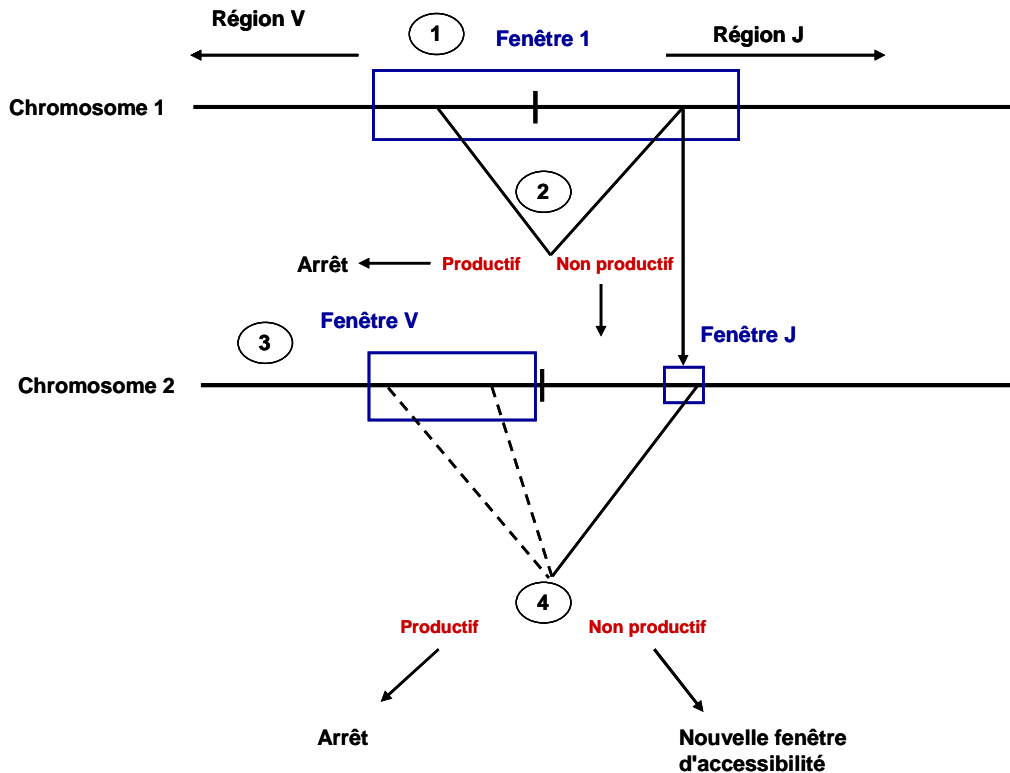


Figure 22 Etapes de modélisation de l'inclusion allélique. 1) Lors du réarrangement des gènes V et J, une première fenêtre d'ouverture est mise en place, donnant accès à un nombre limité de gènes V et J la possibilité de se recombiner. 2) Un premier réarrangement VJ est initié. 3) Si ce réarrangement est productif, le programme de modélisation s'arrête. Sinon, un premier réarrangement est initié sur le second chromosome. 4) L'accessibilité à la région J est restreinte aux gènes J se localisant à 5Kb autour du gène J utilisé sur le chromosome 1. L'accessibilité à la région V est permise sur la même fenêtre d'ouverture que celle utilisée sur le chromosome 1. 5) Si le réarrangement est productif, le programme de modélisation s'arrête, sinon une nouvelle fenêtre d'ouverture est mise en place à partir des derniers V et J utilisés sur chaque chromosome. Les fenêtres d'accessibilité sont représentées en bleu. Les traits en pointillés indiquent que, sur le second chromosome, le gène V peut être choisi tout au long de la fenêtre d'ouverture.

9.2. Résultats

Les résultats de la simulation montrent un profil très similaire à ceux proposés sans l'inclusion allélique. La figure 23 montre que tous les réarrangements ne sont pas équiprobables. Nous constatons de même, que des réarrangements sont peu probables (entre les gènes V proximaux et les gènes J distaux, et inversement).

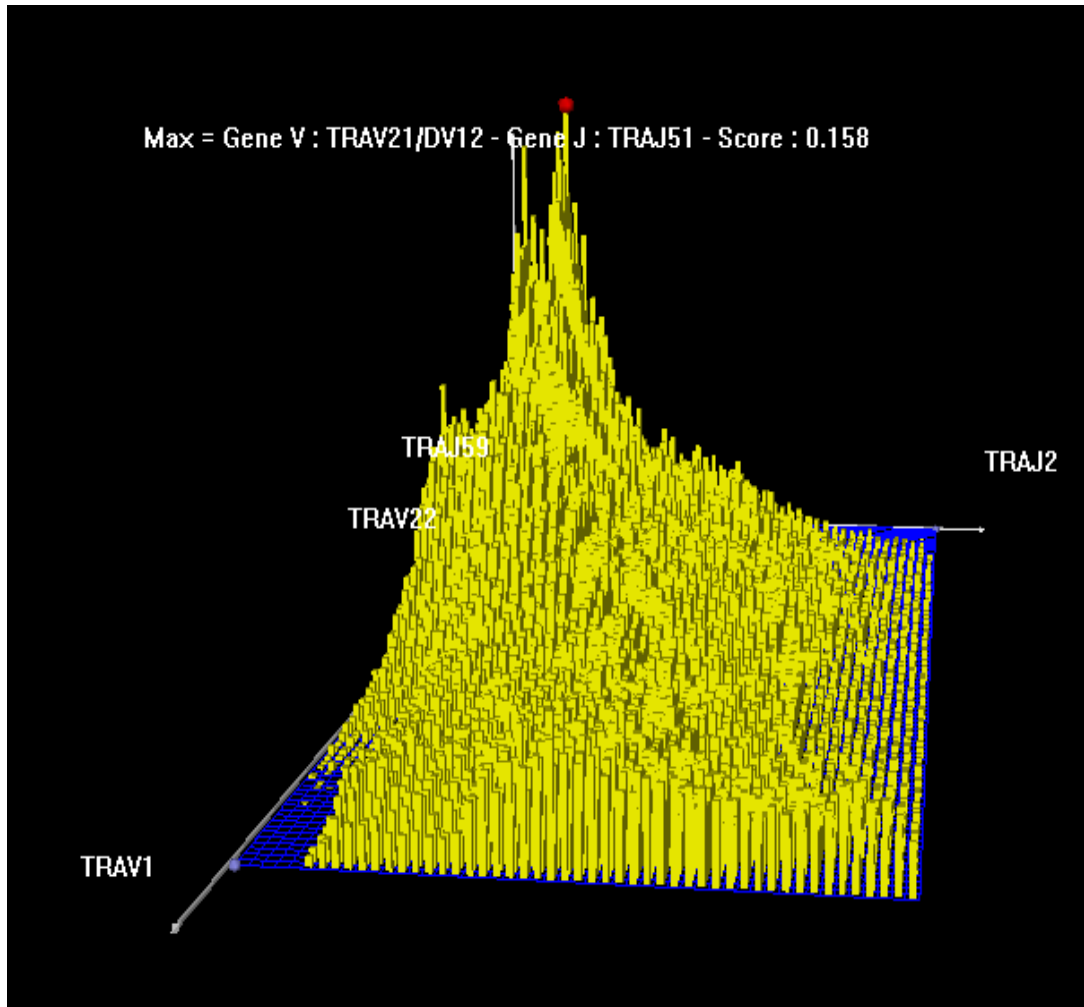


Figure 23 Représentation 3-D des résultats de simulation tenant compte de l'inclusion allélique

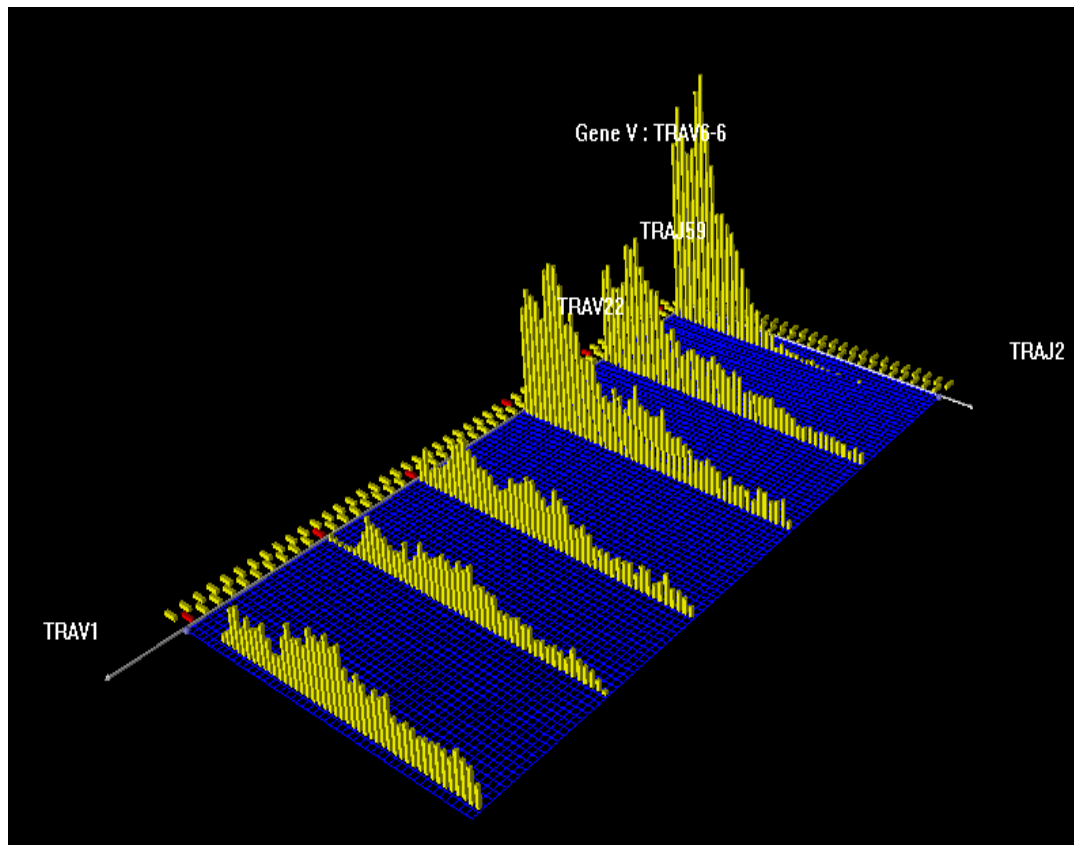


Figure 24 Représentation 2-D des résultats de simulation tenant compte de l'inclusion allélique

Sur la figure 24, la présentation en deux dimensions montre que les hot spots de recombinaison sont toujours visibles, et correspondent aux données biologiques.

La synchronisation des gènes J permet d'augmenter de 1% le répertoire combinatoire VJ observé et amène à 97% de recombinaison par rapport à la version précédente qui était de 96%. Cela porte à 4756 combinaisons possibles sur les 4900 testées (pseudo-gènes exclus).

Dans la précédente version du modèle, si le premier réarrangement n'était pas productif, le modèle passait à la simulation d'une nouvelle chaîne TR α .

Dans l'ensemble, la prise en compte de la synchronisation des gènes J ne modifie pas le profil des réarrangements. Dès que l'on augmente le nombre de recombinaisons, une compensation s'établit et les variabilités sont faibles.

Cela n'est néanmoins plus valable à partir du second réarrangement. Cependant l'inclusion allélique mise en place ne montre pas de changement flagrant dans le profil d'expression des combinaisons VJ. Dans l'ensemble, les fréquences des différentes combinaisons ne sont guère modifiées et sont en accord avec les données expérimentales.

10. Modélisation des réarrangements VJ au niveau du locus TRA/TRD humain

10.1. Introduction

La diversité des TR α joue un rôle important dans la reconnaissance du peptide antigénique. Il n'existe pas à l'heure actuelle de technique permettant de quantifier et de prédire les fréquences de toutes les combinaisons VJ. En effet, la complexité du locus TRA/TRD, en particulier de la région V, ne permet pas une quantification de toutes les combinaisons VJ (~3200) par une approche expérimentale, d'où l'intérêt de mettre en place un modèle de réarrangement, permettant la simulation des combinaisons VJ.

Les loci TRA/TRD murin et humain présentent de nombreux points communs. Cependant, si les deux régions J sont très bien conservées, la situation est différente pour les deux régions V. Les différences concernent le nombre de gènes V et la taille du locus. Chez la souris, il y a moins de familles de gènes V (22 familles V versus 41 chez l'homme), mais le nombre de membres, c'est-à-dire le nombre de gènes par famille est beaucoup plus important (3 membres au maximum pour les familles V chez l'homme versus plus de 10 membres pour les familles V alpha chez la souris) et, par conséquent, il y a plus de gènes V chez la souris qui sont répartis sur une distance plus importante (600 kb chez l'homme versus plus de 1400 kb chez la souris). Pour développer un modèle de simulation des réarrangements du locus TRA/TRD humain, nous avons dû enrichir les données de la littérature par une approche expérimentale, fondée essentiellement sur la quantification des combinaisons VJ par PCR quantitative en temps réel au niveau génomique. Chez l'homme, il est impossible de suivre expérimentalement l'ouverture du locus au cours de l'ontogénie, comme cela a été fait chez la souris. Pour pallier ce manque, nous avons décidé de quantifier environ 3% des réarrangements VJ.

10.2. Matériels et Méthodes

10.2.1. Mise en place des conditions expérimentales de PCR

10.2.1.1. Conditions de PCR et choix des amorces

Les séquences des amorces ont été choisies pour que chaque amorce soit spécifique du gène V ou J. En utilisant BLAST, nous avons vérifié la spécificité de chaque amorce, c'est-à-dire que l'amorce n'hybride aucun autre gène du génome.

Les amorces ont été choisies en répondant aux critères suivants :

1. les températures optimales d'amplification sont voisines. En effet, les conditions d'amplification seront identiques pour toutes les PCR.
2. le rendement d'amplification de chaque couple d'amorces V-J est de l'ordre de 90%.

3. Chaque couple d'amorces amplifie un seul amplicon. Cela est vérifié par gel d'agarose, par analyse des courbes de fusion et par séquençage.
4. la taille de chaque amplicon est sensiblement constante, afin de pouvoir comparer les résultats de PCR entre eux.

Compte tenu des remarques, des critères et du choix des amorces, l'ensemble des PCR quantitatives ont été réalisées dans les conditions suivantes : dénaturation : 95°C, 10 minutes ; amplification (40 cycles) ; dénaturation : 94°C, 15 secondes ; hybridation : 67°C, 7 secondes ; élongation : 72°C, 7 secondes.

10.2.1.2. Choix des couples de gènes V et J testés

Nous avons quantifié les réarrangements de 9 gènes V α , combinés à 10 gènes J α différents. Les gènes J α ont été choisis en regard de ceux analysés chez la souris et en tenant compte des J localisés au niveau des fenêtres d'accessibilité, et de ceux localisés dans les régions des hot spots de recombinaison. Pour la région V, les 9 V étudiés sont répartis sur l'ensemble de la région V, en privilégiant les V situés près de la région codant pour la chaîne delta du TR. Nous avons aussi choisi des gènes V et J, dont le RSS est conservé ou quasiment identique à la séquence consensus. Quelques V et J ont des RSS moins conservés, mais nous les avons choisis, car leur position était stratégique.

10.2.1.3. Liste des amorces

Les gènes V et J sont nommés suivant la nomenclature de la base de données ImMunoGeneTics (IMGT) (<http://imgt.cines.fr/>).

Numéros d'accèsion : pour la région V, AE000658-AE000662 ; pour la région J, M94081.

Amorce	Séquence 5' → 3'	Sens/Anti-sens
hTRAV2	TCTCTTCATCGCTGCTCATCCTCC	S
HTRAV7	GGGACCCCAGCAGGGAGACGTTGCC	S
hTRAV 10	GCATCTGACGACCTTCTTGGT	S
hTRAV 16	AGAGTGACTCAGCCCGAGAAG	S
hTRAV 21	TGCCTCGCTGGATAAATCATCAGG	S
hTRAV 30	GCCGTGATCCTCCGAGAAGGGG	S
hTRAV35	GGCTGGGAAGTTTGGTGATATAGTGTC	S
hTRAV38	AGCAGCCAAATCCTTCAGTCTCAA	S
hTRAV40	AAGACAAAACCTCCCCATTGTGAAATA	S
hTRAV41	GCCCTCCTGAAAATGTGTAAAGAAATGT	S
hTRAJ5	CTGTCTCTGCAATGATGAAATGGCC	A
hTRAJ10	CCACTTTTAGCTGAGTGCCTGTCCC	A
hTRAJ24	GGTCCCTGCTCCAAACTGC	A
hTRAJ33	CGCCCCAGATTAAGTATAGTTGCT	A
hTRAAJ41	TGCCCCGAGACCTGATAACCAA	A
hTRAAJ47	GGGTTGCCTTCGAGAGCGTTAATC	A
hTRAAJ48	AGCACTTGACGGCAGCAGC	A
hTRAAJ53	CTTCCCCACTCCCTTCAAACCTTAC	A
hTRAJ56	ACTGGGCAGGAGATTCGGTTAT	A
hTRAAJ61	ACTTGCTGAGTTTCATGATTCCTC	A
G3DPHup	AGCAATGCCTCCTGCACCACCAAC	S
G3DPHdo	CCGGAGGGGCCATCCACAGTCT	A

Tableau 1 Amorces utilisées en PCR quantitative sur ADN génomique de thymus humain

1.1.1. Standardisation et présentation des résultats

90 combinaisons VJ ont été quantifiées, pour chaque échantillon, représentant 3% des combinaisons VJ potentielles de la chaîne TR alpha humaine. De manière à comparer les résultats, la quantification d'un gène de ménage, le gène G3PDH, a été réalisée systématiquement. Chaque combinaison VJ a été ensuite standardisée par rapport au gène G3PDH. Pour cela, nous avons soustrait le cycle de sortie du produit d'amplification de G3PDH au cycle de sortie du produit d'amplification de la combinaison VJ d'intérêt.

10.2.2. Provenance des échantillons

Les ADN génomiques ont été extraits à partir de déchets opératoires de thymus de quatre enfants sains, âgés entre un mois et un an (Fuschiotti, Pasqual et al. 2007).

10.3. Résultats

10.3.1. Analyse des profils de recombinaison

Les résultats des 90 combinaisons testées sont présentés sous la forme d'histogrammes pour deux échantillons (figure 25) et la moyenne et l'écart-type des valeurs obtenues pour chaque combinaison, sur les 4 échantillons, sont présentés dans la figure 26. De manière générale, nous avons appelé gènes V proximaux, les gènes V proches de la région J. Les gènes V distaux correspondent aux gènes les plus éloignés de la région J. Le même raisonnement est utilisé pour les gènes J. La quantité des produits détectés pour chaque combinaison est variable d'un échantillon à l'autre. Cette variabilité est liée au fait que, vu que les combinaisons VJ sont supérieures en potentiel au nombre de cellules T présentes à un instant donné dans le thymus, chaque individu a un répertoire spécifique. A cela, il convient d'ajouter que les amorces ne sont pas strictement spécifiques, à la base près, des gènes V et J d'un individu donné (la séquence des différents gènes n'a pas été déterminée et vu les mutations alléliques, les séquences des amorces ne sont pas les homologues parfaites des gènes à amplifier). Malgré ces variations, les profils de réarrangement, d'un échantillon à l'autre, sont relativement constants (figure 27). De manière générale, les familles V41 et V40 réarrangent avec les gènes J situés entre J61 et J24, avec une préférence pour les J situés autour de J61, J53 et J48. Les familles V38 et 35 réarrangent avec les J situés entre J61 et J10 et on note à nouveau une utilisation préférentielle autour de J48. Dans l'ensemble, les combinaisons testées pour V38 sont plus fréquentes que celles observées avec V41, V40 et V35, mais il est à souligner que la famille V38 regroupe deux membres et que, par conséquent, cela influence les fréquences des combinaisons détectées (tableau 2), Pour les V situés dans la partie centrale, comme V30, V21 et V16, un profil de fréquence d'association VJ relativement

constant est observé. Néanmoins, les combinaisons semblent plus fréquentes avec J53 et J48 d'une part, et J24 et J10 d'autre part. Les familles V7 et V2, localisées au niveau distal du locus, s'associent avec les J situés au niveau distal, donc par conséquent J24, J10 et J5.

JV	V41	V40	V38	V35	V30	V21	V16	V7	V2
61	0,0183	0,0224	0,0172	0,0084	0,0026	0,0077	0,0051	0,0005	0,0002
56	0,0018	0,0089	0,0435	0,0112	0,0043	0,0128	0,0113	0,0014	0,0004
53	0,0178	0,0136	0,0626	0,0321	0,0085	0,0115	0,0199	0,0002	0,0001
48	0,0206	0,0268	0,0358	0,0145	0,0081	0,0055	0,0197	0,0006	0,0003
47	0,0067	0,0195	0,0209	0,0087	0,0069	0,0054	0,0097	0,0007	0,0006
41	0,0005	0,0064	0,0159	0,0107	0,0046	0,0079	0,0039	0,0021	0,0016
33	0,0122	0,0088	0,0163	0,0088	0,0050	0,0063	0,0080	0,0014	0,0062
24	0,0041	0,0060	0,0099	0,0178	0,0046	0,0054	0,0172	0,0234	0,0304
10	0,0008	0,0011	0,0105	0,0032	0,0120	0,0075	0,0391	0,0594	0,0377
5	0,0002	0,0000	0,0006	0,0000	0,0029	0,0056	0,0017	0,0054	0,0115

Tableau 2 Fréquences des combinaisons VJ testées expérimentalement. Les fréquences des combinaisons sont calculées à partir de la quantification de l'ensemble des combinaisons ($F_{VJ} = Q_{VJ} / \sum Q_{ViJi}$) où F représente la fréquence d'une combinaison VJ et Q sa quantification.

Le tableau 2 résume les fréquences obtenues pour chaque combinaison VJ testée. Nous observons un gradient de fréquence de combinaison montrant les combinaisons préférentielles entre les gènes V et J proximaux et les gènes V et J distaux. Concernant la famille V41 par exemple, les fréquences de combinaisons avec les gènes J5 ou J10 sont entre 90 et 20 fois moins importantes que celles observées avec les gènes J proximaux J61 ou J53.

D'une manière générale les fréquences des recombinaisons de V30 avec l'ensemble des J sont plus faibles que celles observées pour les autres V. Il est à noter que la séquence RSS de V30 présente une homologie de 91% avec la séquence consensus.

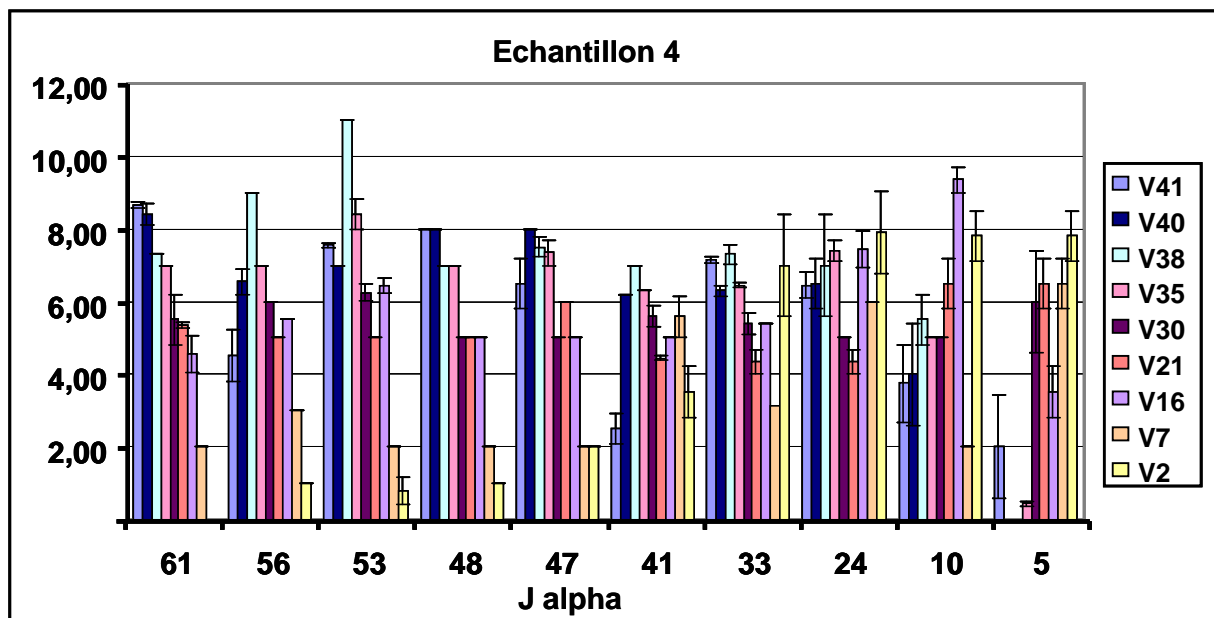
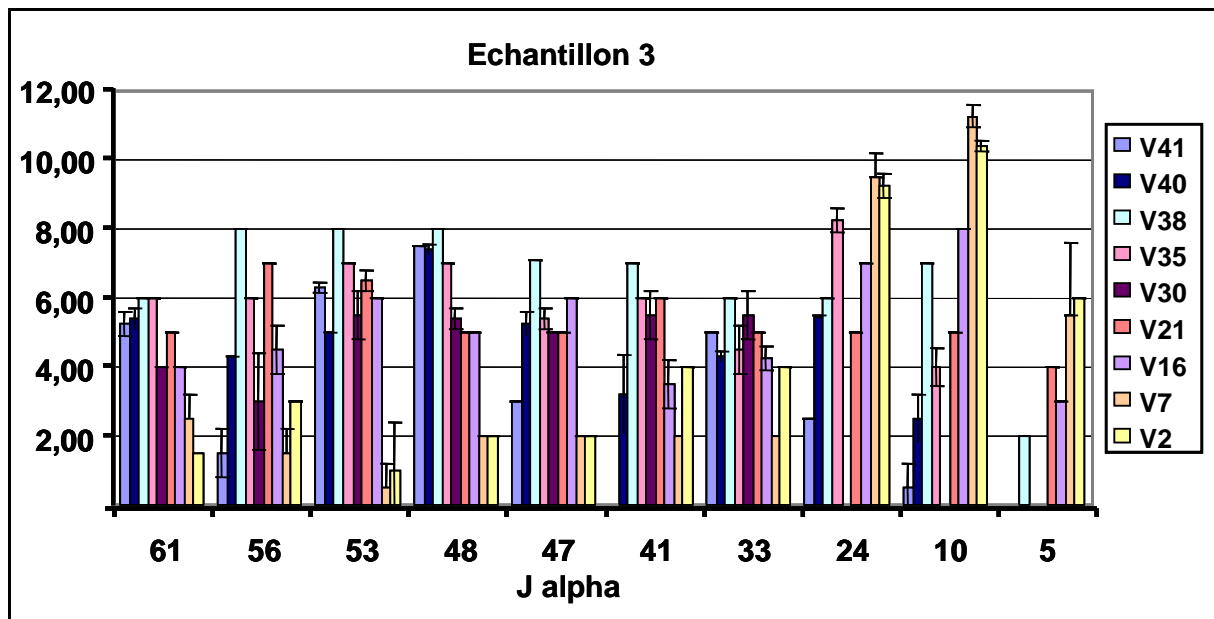


Figure 25 Profils de recombinaison $VJ\alpha$ observés sur deux échantillons. L'axe des abscisses représente les différents gènes J alpha testés, du plus proximal (J61) au plus distal (J5). L'axe des ordonnées traduit, en unités arbitraires, l'abondance relative des recombinaisons VJ observées par PCR.

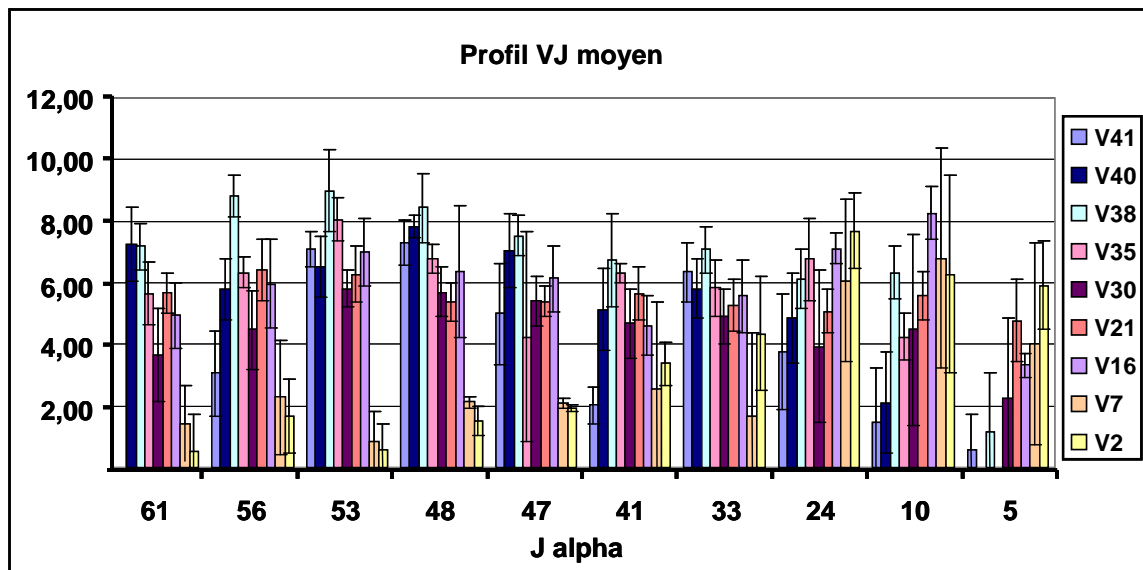


Figure 26 Profil de recombinaison VJ moyen obtenu pour les quatre échantillons. L'axe des abscisses représente les différents J alpha testés, du plus proximal (J61) au plus distal (J5). L'axe des ordonnées traduit, en unités arbitraires, l'abondance relative des recombinaisons VJ observées par PCR.

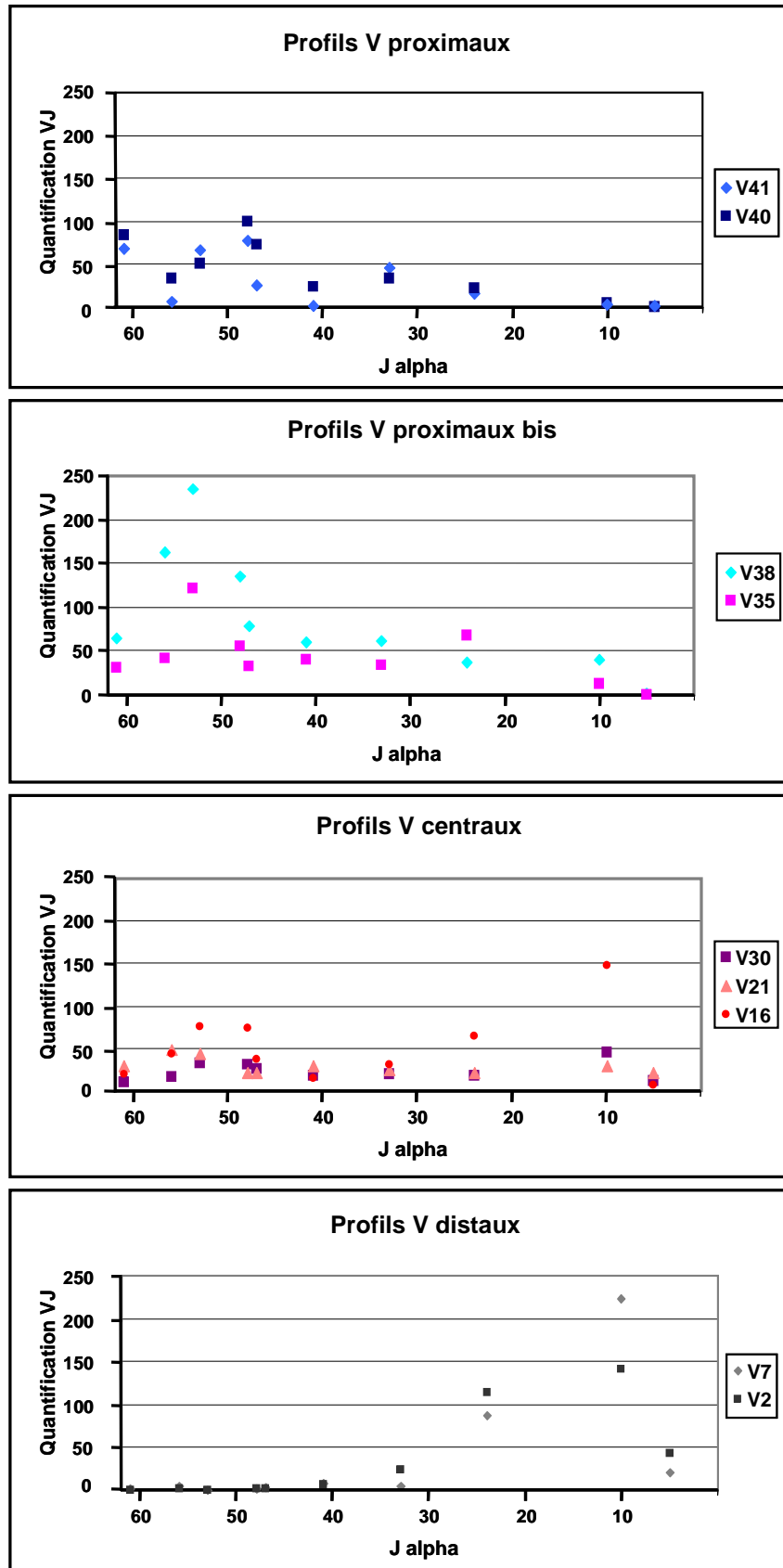


Figure 27 Profils de distribution des familles $V\alpha$ après quantification des combinaisons VJ. La quantification des combinaisons VJ est calculée pour un rendement de PCR à 90%, suivant la formule $1,8^n$, où n représente la sortie de cycle.

Dans l'ensemble, les profils de réarrangement obtenus confirment les résultats précédents publiés par le laboratoire en utilisant la PCR multiplex (Fuschiotti, Pasqual et al. 2007) et montrent des combinaisons plus fréquentes entre gènes V et J proximaux, de même qu'entre gènes V et J distaux. Cela indique que par PCR quantitative et malgré l'utilisation d'un grand nombre d'amorces, un minimum de biais a été introduit. De surcroît, par rapport aux résultats de PCR multiplex qui reflètent l'utilisation qualitative des réarrangements d'un gène donné V avec les gènes J, l'approche présentée dans ces travaux permet de proposer un profil quantitatif de la diversité combinatoire du locus TRA/TRD humain, profil qui servira de base pour développer le modèle de simulation des réarrangements TR α humains.

10.4. Modélisation

10.4.1. Détermination des vitesses d'ouverture

Comme précisé dans l'introduction, la différence essentielle entre le locus murin et humain réside au niveau de la taille de la région V. Celle-ci est deux fois plus petite chez l'homme. En revanche, la région J est similaire en taille et en nombre de gènes J. Dans un premier temps, nous avons testé si les vitesses d'ouverture estimées pour les régions V et J du locus TRA/TRD murin pouvaient être utilisées pour la simulation des réarrangements VJ alpha humains. Cette approche simpliste donne des résultats aberrants. Les résultats montrent une distribution inattendue des combinaisons VJ, à savoir une augmentation des combinaisons entre V distaux (V15 à V1-1) avec les gènes J proximaux (J61 à J 30), ce qui n'est pas observé sur les résultats expérimentaux (figure 26 et figure 27). En conclusion, cette première analyse montre la nécessité d'adapter les vitesses d'ouverture des régions V et J à l'homme.

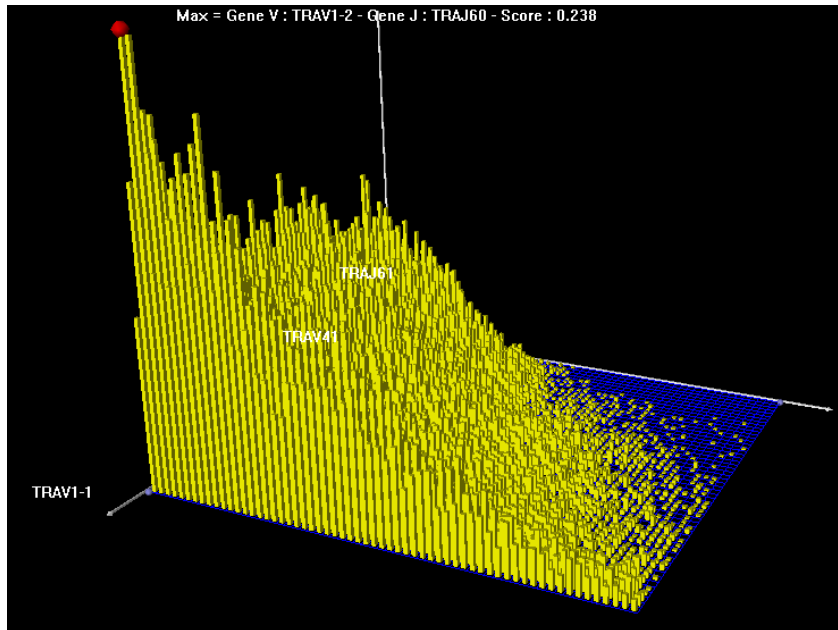


Figure 28 Représentation 3-D des combinaisons VJ chez l'homme par transposition directe des paramètres murins.

10.4.2. Calcul des vitesses d'ouverture

Nous avons donc choisi de commencer les simulations des réarrangements VJ chez l'homme, en réduisant de manière proportionnelle la vitesse en fonction de la taille de la région V, ce qui nous amène à une vitesse S_v comprise entre 2500pb/h et 14000pb/h.

La région J étant de même taille que chez la souris, nous avons gardé les mêmes vitesses d'ouverture pour débiter la modélisation.

10.4.3. Résultats

Les figure 29 et figure 30 présentent les profils des réarrangements VJ obtenus par simulation. La figure 29 permet d'avoir une visualisation complète du répertoire simulé à travers la présentation en trois dimensions des résultats. La figure 30, à travers une visualisation en deux dimensions, montre que les profils obtenus pour les gènes V40, V30, V15 et V2 sont cohérents avec les profils expérimentaux présentés à ceux de la figure 27. Nous avons ensuite superposé les résultats expérimentaux et ceux provenant de la simulation (figure 31). Les résultats préliminaires montrent que les points résultant de la simulation (en gris) suivent globalement les variations des points expérimentaux (en rouge et en vert). Il y a cependant encore quelques améliorations à apporter, notamment du fait que les J situés entre J61 et J56 sont mal représentés par la modélisation (tableau 3). Cependant, les profils de recombinaison VJ simulés sont similaires aux données expérimentales (figure 29, figure 30 et figure 31).

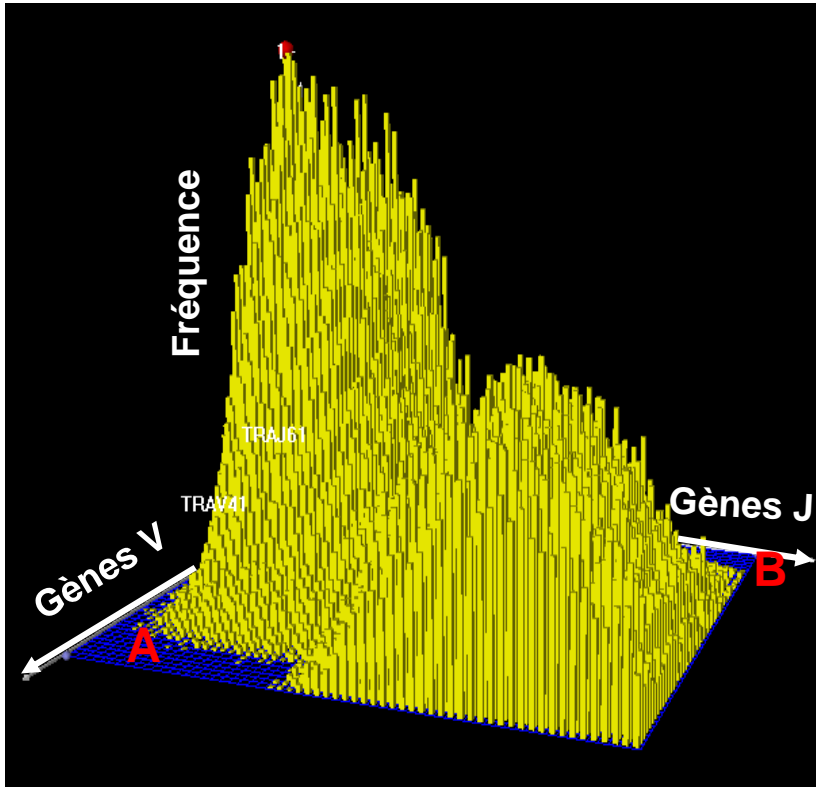


Figure 29 Présentation en 3-D des combinaisons VJ chez l'homme.

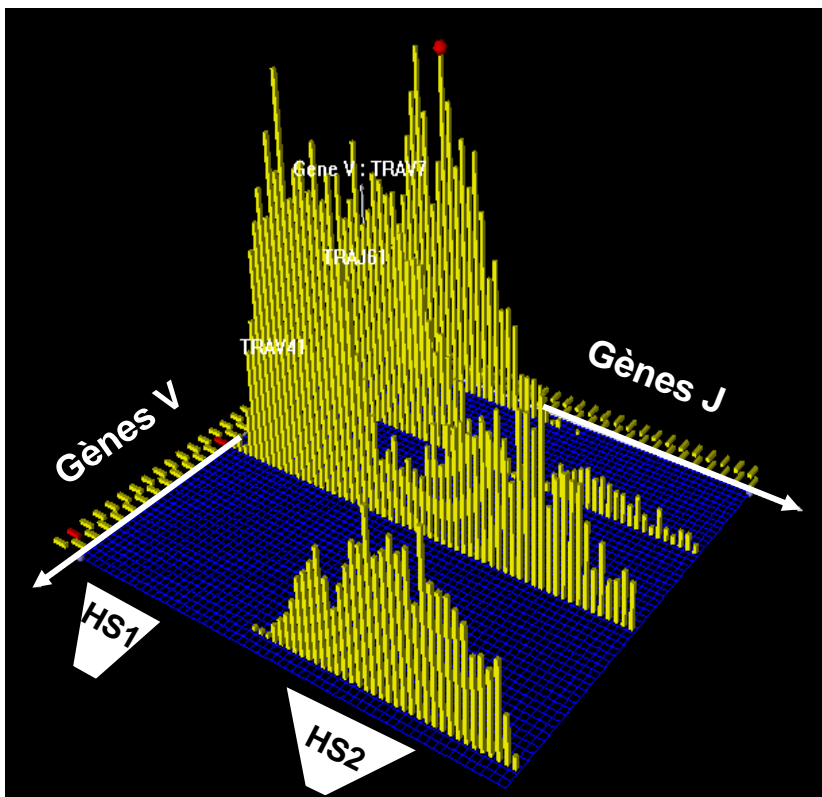


Figure 30 Représentation en 2D des combinaisons VJ pour les familles V40, V30, V16 et V2. La localisation des deux zones de hot spots est indiquée.

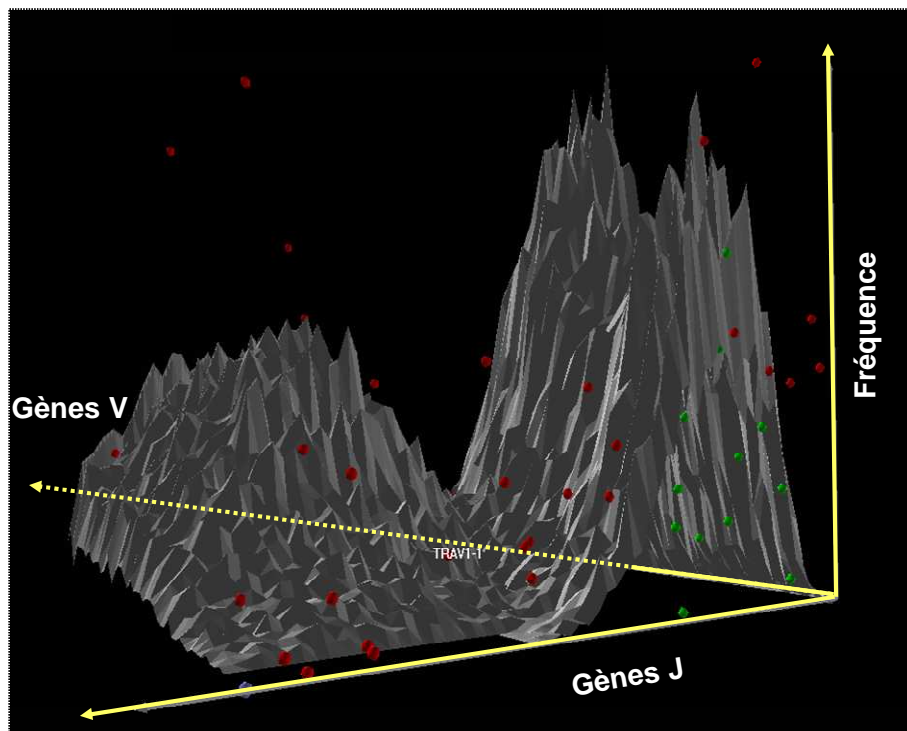
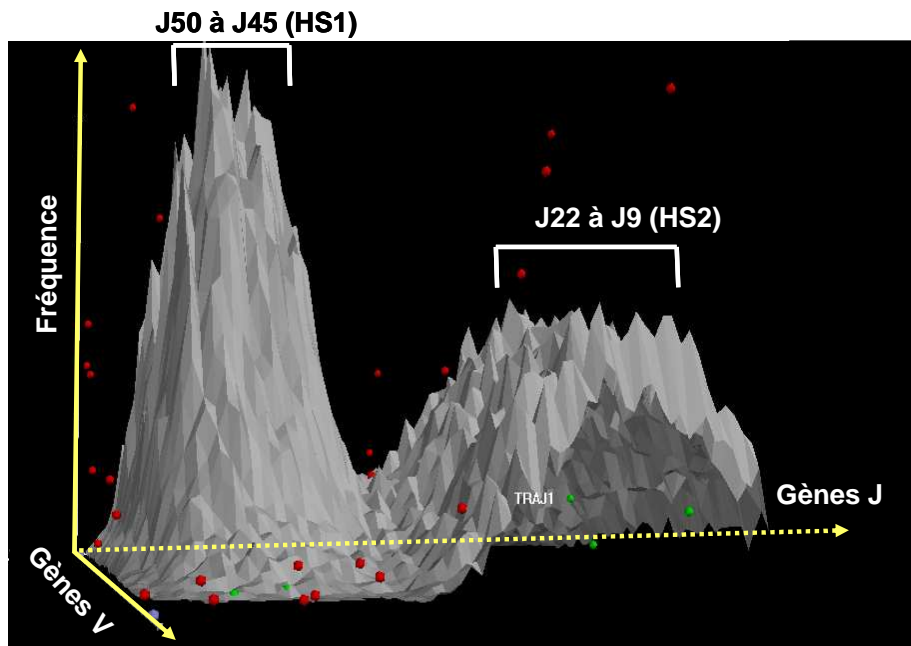


Figure 31 Superposition des résultats de simulation et des points expérimentaux. Les points rouges représentent les expérimentaux au dessus des résultats de simulation (gris) et les points verts les points au dessous. Deux régions de recombinaisons préférentielles sont indiquées sur la figure du haut.

J/V	V41	V40	V38	V35	V30	V21	V16	V7	V2
61	0,0000	0,0008	0,0021	0,0048	0,0008	0,0018	0,0020	0,0000	0,0000
56	0,1022	0,0973	0,1014	0,1021	0,0796	0,0909	0,0907	0,0000	0,0000
53	0,1947	0,1833	0,1802	0,1861	0,1827	0,1649	0,1395	0,0000	0,0000
48	0,2248	0,2510	0,2481	0,2375	0,2288	0,2106	0,1960	0,0105	0,0000
47	0,2236	0,2418	0,2334	0,2489	0,2330	0,1814	0,1743	0,0168	0,0000
41	0,1983	0,1840	0,1982	0,1710	0,1660	0,1567	0,1496	0,0063	0,0000
33	0,0553	0,0418	0,0365	0,0332	0,0369	0,0397	0,0534	0,0737	0,0568
24	0,0012	0,0000	0,0000	0,0060	0,0235	0,0608	0,0725	0,3621	0,2988
10	0,0000	0,0000	0,0000	0,0091	0,0293	0,0534	0,0665	0,3474	0,3975
5	0,0000	0,0000	0,0000	0,0012	0,0193	0,0397	0,0554	0,1832	0,2469

J/V	V41	V40	V38	V35	V30	V21	V16	V7	V2
61	0,2203	0,1977	0,0738	0,0723	0,0432	0,1017	0,0373	0,0051	0,0018*
56	0,0220	0,0786	0,1866	0,0971*	0,0726*	0,1698	0,0835*	0,0142	0,0051
53	0,2146*	0,1194	0,2684	0,2779	0,1425	0,1525*	0,1470*	0,0019	0,0013*
48	0,2480	0,2358	0,1535	0,1256	0,1364	0,0723	0,1454	0,0058	0,0038
47	0,0812	0,1720	0,0898	0,0754	0,1167	0,0716	0,0715	0,0069	0,0063
41	0,0056	0,0562	0,0683	0,0927	0,0774	0,1039	0,0286	0,0220	0,0177
33	0,1470	0,0776	0,0698	0,0763	0,0840	0,0834	0,0589*	0,0147	0,0693
24	0,0493	0,0526	0,0423	0,1544	0,0772	0,0718*	0,1271	0,2464	0,3413
10	0,0094	0,0100	0,0451	0,0280	0,2012	0,0993	0,2882	0,6258	0,4236
5	0,0026*	0,0000*	0,0024*	0,0002	0,0487	0,0736	0,0126	0,0572	0,1297*

Tableau 3 Fréquences de recombinaisons VJ obtenues par simulation (en haut) et expérimentalement (en bas). Les fréquences sont données pour chaque famille de gènes V à partir du nombre de combinaisons obtenues en simulation ou de la quantification des combinaisons VJ pour les données expérimentales. * indique les valeurs empiriques qui sont considérées comme égales à travers un test de comparaison d'une valeur empirique à une valeur simulée.

Points à améliorer dans le modèle

En ce qui concerne V41 et V40, les résultats expérimentaux montrent une importante diminution de leur combinaison avec J56. Il est à noter que ce résultat se retrouve pour V41 et V40, alors que leurs RSS respectifs présentent d'importantes différences de leur score.

Il serait intéressant de voir si V40 et V41 présenteraient un positionnement préférentiel sur J61 et J58 (J60 et J59 étant des pseudogènes). Ce positionnement pourrait alors signaler au second chromosome de commencer à se réarranger, mais sur une fenêtre d'accessibilité plus grande, ce qui expliquerait que V41 et V40 sont aussi observés dans des combinaisons avec des J plus centraux. De plus, le positionnement sur J61 serait quasi-exclusif de V41 et V40, qui sont localisés à moins de 6Kb l'un de l'autre, car il n'est presque plus observé à partir de V38. Il serait également intéressant de voir comment se comporte V39, qui est localisé entre V40 et V38, avec J61, J56 et J53.

10.4.3.1. Couverture de réarrangement

La simulation d'un million de réarrangements VJ produit 90% de combinaisons productives, avec un maximum de 2965 combinaisons sur les 3294 combinaisons possibles ($54V \times 61J$). Il y a en moyenne 329 combinaisons peu probables, soit environ 10%. Le pourcentage important de combinaisons VJ peu probables chez l'homme par rapport à la souris (4% de combinaisons peu probable) n'est pas encore expliqué.

Tout comme chez la souris, il existe deux régions où les combinaisons VJ sont peu probables, les régions A et B sur la figure 29, qui correspondent aux réarrangements entre les gènes V distaux et gènes J proximaux, et inversement.

10.4.3.2. Localisation des Hot Spots de recombinaison

Chez la souris, la localisation des HS est décrite (Rytönen, Hurwitz et al. 1994; Aude-Garcia, Gallagher et al. 2001) et nous avons montré qu'une augmentation de la densité des gènes J correspondait aux régions de forte recombinaison. La présence des HS chez l'homme n'a pas encore été étudiée. Pour tenter d'élucider ce point, nous avons, comme chez la souris, regardé si la distribution des J était ou non homogène. Les résultats sont présentés sur la figure 32.

La figure 32 met en évidence deux régions où les gènes J sont plus fréquents, par conséquent la densité des gènes J plus élevée. La première se situe entre J48 et J45 et la seconde entre J22 et J12, ce qui nous donne une indication sur la localisation potentielle des HS. Il est à noter que le positionnement des HS n'est pas une variable de notre modèle, mais il nous permet de valider les paramètres essentiels, qui sont les mêmes que pour la souris.

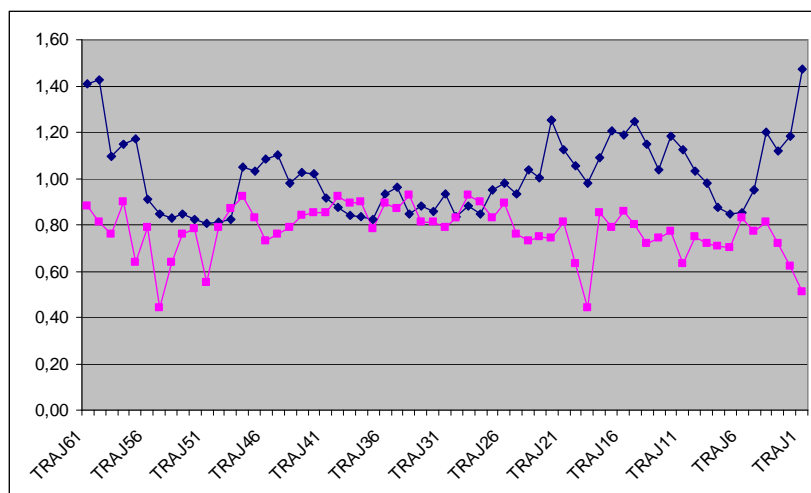


Figure 32 Superposition graphique de la densité des gènes J (en bleu) et de leur RSS (en rose). L'axe des ordonnées représente la valeur des RSS de chacun des gènes J (donnée entre 0 et 1) et la densité des gènes J.

10.4.3.3. Fréquences de réarrangement

Les fréquences de réarrangement obtenues par le modèle pour chacune des combinaisons VJ testées montrent que les réarrangements ne sont pas équiprobables (tableau 4). Si cela avait été le cas, la fréquence de chaque combinaison aurait été de 3.10^{-4} . Le tableau 4 montre les fréquences de réarrangement d'un ensemble de douze combinaisons VJ, sélectionnées tout au long du locus. Suivant la position des gènes V et J, les fréquences de recombinaison varient entre 0,034 et 0 dans l'exemple de combinaisons sélectionnées.

J/V	V40	V20	V1-1
J60	0,00323	0,00350	0,00000
J57	0,02370	0,03420	0,00000
J26	0,00054	0,02477	0,02773
J1	0,00000	0,00539	0,01454

Tableau 4 Exemple de fréquences de combinaisons VJ obtenues par simulation. Les fréquences des combinaisons VJ sont données par rapport aux combinaisons VJ totales. Compte tenu du nombre important de simulations faites, les fréquences supérieures à 0 sont significativement différentes de 0.

10.4.3.4. Première fenêtre d'ouverture

Les résultats de la simulation montrent que 47% des combinaisons VJ sont localisées entre J61 et J48 au cours de la première recombinaison, et que cela monte à 70%, si l'on agrandit la fenêtre d'utilisation de J48 à J44. Ce résultat propose que la première fenêtre d'ouverture soit légèrement plus grande que chez la souris. En effet, lors de la première recombinaison chez la souris, 70% des combinaisons obtenues étaient comprises entre J61 et J48.

10.4.3.5. Réarrangements secondaires

Les réarrangements secondaires sont largement décrits (Wang, Huang et al. 1998; Huang and Kanagawa 2001; Guo, Hawwari et al. 2002), mais leur nombre n'est pas encore connu. Notre modèle prédit qu'après une première maturation permettant l'élimination des gènes D, un premier réarrangement est réalisé suivi de deux (dans 99% des cas) ou trois (dans 1% des cas) réarrangements supplémentaires. Le tableau 5 montre la répartition des réarrangements, selon qu'ils ont été générés lors du premier, deuxième, troisième ou quatrième réarrangement. Le résultat montre une contribution décroissante, après chaque réarrangement, et une très faible contribution du quatrième réarrangement.

Réarrangement	1	2	3	4
Contribution (%)	56	39	4	1

Tableau 5 Contribution des réarrangements

10.4.3.6. Profils de recombinaisons et RSS

Chez la souris, nous avons montré que, si l'on tient compte de la valeur des RSS dans la modélisation, les profils de recombinaison obtenus ne sont pas modifiés de manière drastique. Nous avons réalisé une simulation en tenant compte des RSS chez l'homme. Les résultats (figure 33 et figure 34) montrent que les profils de recombinaisons ne sont pas modifiés de manière importante, si l'on prend en compte les RSS dans la simulation des réarrangements VJ. Ces résultats sont à comparer avec les figure 29 et figure 30. Il est certain que les gènes V et J, qui ont une séquence RSS peu conservée, seront *de facto* moins fréquents.

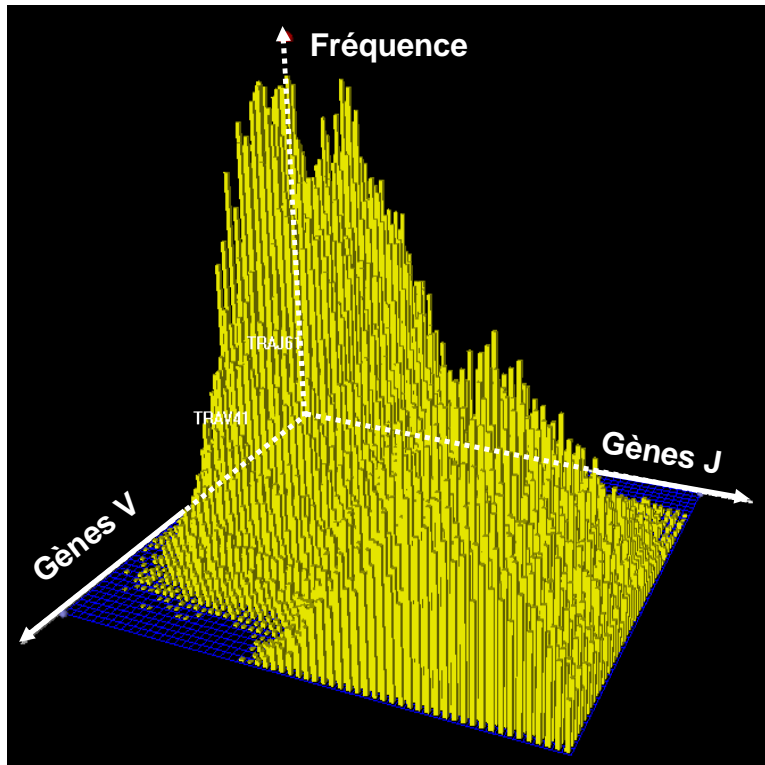


Figure 33 Représentation 3-D des résultats de simulation avec les RSS

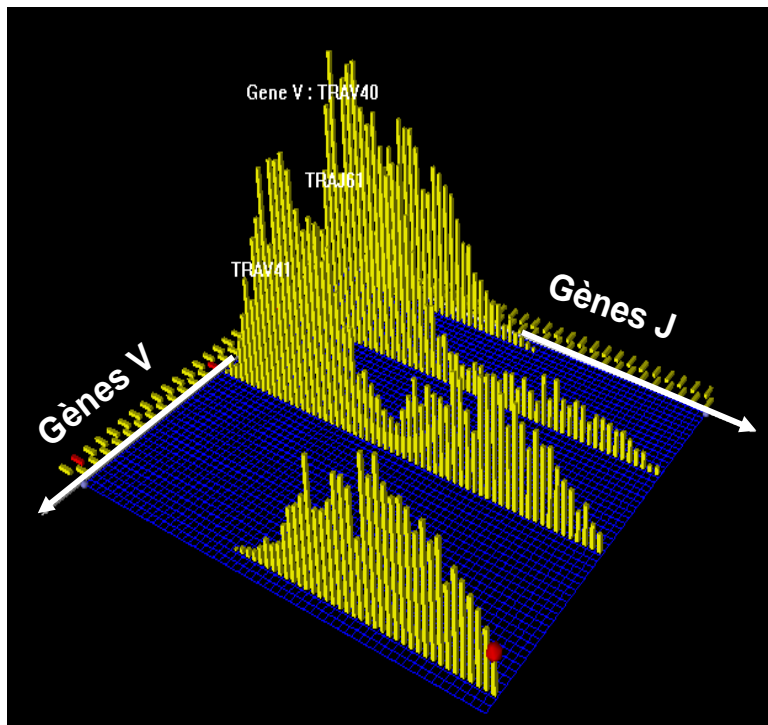


Figure 34 Représentation 2-D des résultats de simulation avec les RSS pour les familles V40, V30, V16 et V2.

10.5. Discussion et conclusions

Nous avons réalisé une analyse des profils des réarrangements VJ de thymocytes humains sains, à travers une analyse par PCR quantitative sur ADN génomique. Cela a permis de quantifier un ensemble de combinaisons VJ et de dresser une vue d'ensemble des profils de combinaisons VJ du locus TRA/TRD humain, ce qui n'était pas encore décrit avec autant de précision. Les résultats confirment ceux rapportés par Fuschiotti *et al.* (Fuschiotti, Pasqual et al. 2007) en particulier : **1)** le profil des combinaisons VJ n'est pas aléatoire, ce qui réduit de manière importante le répertoire combinatoire, **2)** un V donné utilise de manière préférentielle un ensemble de J donnés, et les profils varient suivant la position du gène V sur le locus, **3)** les réarrangements sont observés préférentiellement entre les gènes V et J proximaux plutôt qu'aux extrémités du locus, ces trois premiers points confirmant les résultats présentés dans la papier de Fuschiotti *et al.*, et **4)** le profil de recombinaison ne dépend pas uniquement de la valeur des RSS, mais est aussi corrélé à la densité des gènes sur le locus.

Les profils des combinaisons VJ humaines sont en accord avec ceux obtenus chez la souris (Pasqual, Gallagher et al. 2002). Les résultats suggèrent que les fréquences des combinaisons VJ sont plus importantes en général avec J53 et J48. Ceci est en concordance avec les résultats publiés chez la souris, où le premier hot spot est situé autour de J48, ce qui montrerait que le mécanisme de régulation de l'accessibilité de la région J est conservé entre la souris et l'homme, notamment par l'intermédiaire du TEA (Abarrategui and Krangel 2007).

Comme chez la souris, nous montrons que les réarrangements VJ sont observés préférentiellement avec les J localisés dans des régions denses en gènes J. Des travaux récents ont montré que les combinaisons VJ (Hawwari and Krangel 2007), mais aussi la fixation d'un complexe enzymatique sur l'ADN (Chakalova, Debrand et al. 2005), permettent des modifications épigénétiques de l'ADN, notamment en favorisant l'acétylation des histones, ce qui a pour effet de rendre accessibles les gènes V et J en aval des gènes impliqués dans la combinaison VJ.

Au niveau de la simulation, les résultats humains montrent : **1)** une première fenêtre d'accessibilité va de J61 à J44, ce qui correspond à quatre J plus éloignés que chez la souris, **2)** la présence de deux hot spots de recombinaison, le premier localisé entre J50 et J45 et le second qui comprend une région plus étendue de J22 à J9, et **3)** 90% des combinaisons possibles, avec 10% de combinaisons improbables, ce qui est beaucoup plus important que pour la souris, où nous avons observé uniquement 4% de combinaisons improbables.

La couverture combinatoire prédite par le modèle est beaucoup plus importante que celle proposée par Fushiotti *et al.* (Fuschiotti, Pasqual et al. 2007), qui a estimé les combinaisons VJ à 44-56% des combinaisons VJ possibles. Il est à noter que cette estimation est basée 1) sur une technique de PCR multiplexe qui est une technique qualitative et non quantitative et 2) sur un nombre plus faible de combinaisons VJ. Les combinaisons VJ peu fréquentes au départ et donc non détectables par PCR multiplexe sont quantifiables par PCR quantitative. Ainsi, même si la fréquence d'une combinaison est très faible, elle est comptabilisée comme probable par notre technique. Ceci permet d'expliquer la différence avec l'estimation des combinaisons VJ proposée par Fushiotti *et al.* (Fuschiotti, Pasqual et al. 2007). Cette différence dans l'estimation du répertoire montre toute la difficulté rencontrée pour estimer les combinaisons possibles et tout l'intérêt d'un modèle et de simulations de ce modèle.

En conclusion, les résultats préliminaires de simulation chez l'homme montrent que le modèle de réarrangement VJ de la chaîne TR α développé chez la souris et adapté aux données du locus TRA/TRD humain présente des résultats prometteurs, mais qu'il reste encore à adapter plus finement la valeur des différents paramètres, en ajustant notamment la valeur des vitesses d'ouverture ou du temps de maturation. Une étude plus poussée par l'analyse de nouvelles combinaisons VJ, mais aussi l'utilisation de nouveaux témoins permettra de consolider le modèle chez l'homme et d'affiner l'évolution des profils de réarrangement VJ au niveau expérimental.

Discussion
et
perspectives

11. Discussion et perspectives

Les travaux de cette thèse avaient pour objectif l'implémentation d'un modèle des réarrangements V(D)J du locus TRA/TRD chez la souris, puis d'adapter ce modèle pour estimer la fréquence des réarrangements VJ alpha humains. Le locus TRA/TRD est relativement conservé entre l'homme et la souris, la grande différence étant que la région variable est plus complexe chez la souris, du fait des duplications dans la partie centrale de la région V. La complexité du locus TRA/TRD chez la souris a rendu impossible l'estimation de la diversité et des fréquences des combinaisons VJ. Pour résoudre ce problème, nous avons développé un modèle des réarrangements VJ du locus TRA/TRD.

L'article 1 montre les différentes étapes de la réalisation du modèle et de son algorithme d'implémentation pour effectuer des simulations, ainsi que les hypothèses de départ qui ont été testées concernant l'accessibilité des régions V et J.

Les travaux sur le locus TRA/TRD murin montrent que les réarrangements entre les gènes V et J ne sont pas aléatoires, mais, au contraire, régulés de manière précise (Thompson, Pelkonen et al. 1990; Thompson, Pelkonen et al. 1990; Aude-Garcia, Gallagher et al. 2001; Davodeau, Difilippantonio et al. 2001; Pasqual, Gallagher et al. 2002; Krangel 2003). La théorie du cliquet brownien, qui fut la première testée, montre qu'une approche aléatoire des réarrangements VJ ne permet pas d'expliquer les résultats expérimentaux. L'approche par fenêtrage successif, seconde théorie testée, permet de rendre compte des données expérimentales ; la validation de cette seconde approche est détaillée dans l'article 2. Le modèle de simulation a permis d'évaluer différents paramètres contrôlant l'ouverture des régions V et J.

Le locus TRA/TRD a aussi pour particularité de contenir les gènes codant pour la chaîne delta, intercalée entre les gènes V et J de la chaîne TR α . Au cours de l'ontogénie, les réarrangements de la chaîne delta précèdent ceux de la chaîne alpha (Pearse, Wu et al. 1989; Capone, Watrin et al. 1993; Godfrey, Kennedy et al. 1993; Tourigny, Mazel et al. 1997). Afin de prendre en compte ce décalage temporel dans le modèle, une première maturation permet l'élimination des gènes codant pour la chaîne delta. Certains gènes V δ se retrouvent intercalés dans la région codant pour les V α ou peuvent être utilisés pour les chaînes TR δ ou TR α . Cela ne pose pas de problème dans l'interprétation de la simulation chez l'homme, car, chez lui, les gènes V δ les plus utilisés sont ceux situés en amont des premiers gènes V spécifiques à la chaîne alpha, la combinaison V γ 9V δ 2 représentant jusqu'à 5% des PBL chez l'adulte (Haas, Pereira et al. 1993; Hayday 2000).

Il est maintenant connu que les thymocytes peuvent effectuer plusieurs réarrangements successifs au niveau du locus TRA/TRD (Couez, Malissen et al. 1991; Padovan, Casorati et al. 1993; Petrie, Livak et al. 1993; Hayball, Jones et al. 1996; Wang, Huang et al. 1998; Huang and Kanagawa 2001; Guo, Hawwari et al. 2002). Cependant, certaines questions restent en suspens, notamment en ce qui concerne le nombre de réarrangements secondaires successifs qu'un thymocyte peut réaliser. Les valeurs des paramètres implémentés dans le modèle amènent à proposer que la cellule peut réaliser jusqu'à trois réarrangements secondaires, si le réarrangement initial n'est pas productif. Entre deux réarrangements, un temps de maturation a été introduit et correspond au délai pour tester si la chaîne TR α nouvellement réarrangée est fonctionnelle et si elle sera sélectionnée. Actuellement, la durée de cette étape n'est pas connue. Cependant la durée de vie d'un thymocyte au stade DP a été estimée à trois jours (Petrie, Livak et al. 1993; Wang, Huang et al. 1998; Huang and Kanagawa 2001; Guo, Hawwari et al. 2002).

Deux paramètres qui n'ont pas été introduits dans le modèle de Warmflash *et al.* (Warmflash and Dinner 2006) jouent un rôle très important dans notre modèle : une vitesse d'ouverture variable des régions V et J et un temps de maturation entre deux réarrangements successifs. Ces paramètres permettent : **1**) d'être en accord avec les données observées pour les combinaisons impliquant des gènes V localisés aux extrémités du locus, **2**) d'expliquer la présence des hot spots de recombinaison au niveau central de la région J (Rytkonen, Hurwitz et al. 1994; Aude-Garcia, Gallagher et al. 2001), qui n'ont pas été modélisés dans (Warmflash and Dinner 2006) et **3**) d'augmenter la couverture combinatoire des réarrangements VJ en passant de 70% à 96% de combinaisons VJ possibles chez la souris. Ces résultats constituent une avancée importante dans la modélisation des réarrangements VJ car, outre le fait de proposer un résultat plus en accord avec les données expérimentales, ils permettent également d'ajuster les fréquences théoriques des combinaisons VJ proposées par le modèle. Les deux paramètres nouveaux permettent aussi d'avoir une certaine flexibilité dans les fenêtres d'accessibilité aux régions V et J. Le précédent modèle proposé par le groupe de Warmflash, bien que fondé sur une ouverture par fenêtrages successifs, n'explique pas les hot spots de recombinaison. Les fenêtres proposées par Warmflash sont beaucoup plus rigides que dans notre modèle et découpent le locus TRA/TRD en cinq fenêtres comprenant chacune 37V et 13J. Deux réarrangements successifs ne peuvent être réalisés dans une même fenêtre de recombinaison, ce qui implique que les gènes de deux fenêtres différentes ne pourront pas se réarranger entre eux. Par conséquent, le fenêtrage restreint réduit la diversité dans le répertoire

simulé et privilégie certaines combinaisons par rapport à d'autres. Il est à noter que les données de la littérature n'ont pas rapporté de combinaisons privilégiées.

Nous avons montré également une relation entre la densité des gènes J et la localisation des hot spots. En effet, chez la souris, les hot spots se retrouvent au niveau des régions riches en gènes J. L'accessibilité des gènes J est favorisée au niveau des régions à forte concentration en gènes. De ce fait, alors que la séquence RSS est importante (Tonegawa 1983; Akamatsu, Tsurushita et al. 1994; Ramsden, Baetz et al. 1994), la densité des gènes influe aussi sur les fréquences de recombinaison observées. Des travaux récents proposent que les combinaisons VJ (Hawwari and Krangel 2007), ainsi que la fixation d'un complexe enzymatique sur l'ADN (Chakalova, Debrand et al. 2005), dirigeraient les réarrangements secondaires, en rendant accessibles, par acétylation des histones, les gènes V et J directement en aval du réarrangement primaire.

Les lymphocytes T peuvent exprimer jusque deux chaînes TR $\alpha\beta$ différentes à leur surface. La chaîne TR β sera identique (Padovan, Casorati et al. 1993; Davodeau, Peyrat et al. 1995), mais la chaîne TR α peut être différente. Cela est expliqué par le phénomène dit "d'inclusion allélique", qui traduit le phénomène inverse de l'exclusion allélique observé pour la chaîne TR β , où seulement un seul des deux chromosomes est réarrangé à la fois (Khor and Sleckman 2002; Mostoslavsky, Alt et al. 2004; Jackson, Kondilis et al. 2005; Jackson and Krangel 2006; Schlimgen, Reddy et al. 2008). Le locus TRA/TRD est réarrangé en parallèle sur les deux chromosomes, donnant à la cellule des chances supplémentaire d'avoir au moins un récepteur fonctionnel (Malissen, Trucy et al. 1992; Davodeau, Difilippantonio et al. 2001; Lacorazza and Nikolich-Zugich 2004). De plus, il a été observé une synchronisation entre les deux chromosomes dans l'utilisation des gènes J, alors que l'utilisation des gènes V semble plus aléatoire (Malissen, Trucy et al. 1992; Davodeau, Difilippantonio et al. 2001). Dans une version plus avancée de notre modèle, nous avons implémenté le mécanisme d'inclusion allélique du locus TRA/TRD chez la souris. Les résultats obtenus montrent que les profils de recombinaison restent similaires.

Les résultats de simulation montrent une grande diversité combinatoire chez la souris avec 96% des combinaisons VJ possibles soit un maximum de 4700 combinaisons VJ possibles sur les 4900 possibles (100V*49J), soit 200 combinaisons peut probables, soit environs 4%. Chez l'homme, la diversité combinatoire est légèrement inférieure. Elle est estimée à 90% des combinaisons VJ possibles, soit un maximum de 2965 combinaisons sur les 3294 combinaisons possibles (54V*61J). Il y a en moyenne 329 combinaisons peu probables, soit

environ 10%. La différence dans le pourcentage de combinaisons peu probable entre l'homme et la souris ne s'explique pas encore.

La chaîne TR α joue un rôle majeur dans la reconnaissance des antigènes. Connaître et prédire les fréquences d'apparition d'une combinaison VJ constitue un enjeu médical important, car cela permettrait de définir une signature immunitaire individuelle, mais aussi de diagnostiquer une maladie de l'immunité, à condition de voir une modification des réponses immunitaires liées à des infections spécifiques. L'analyse de quinze combinaisons VJ par PCR quantitative sur ADN génomique de thymus et de PBL n'a pas montré de grande différence entre les répertoires thymique et circulant (Fuschiotti, Pasqual et al. 2007). Il serait intéressant de poursuivre cette analyse, en prenant plus de combinaisons VJ, afin de voir si ce qui a été observé au niveau des gènes V41, V40 et V1 se retrouve sur le reste du répertoire. Ce travail permettrait d'adapter le modèle au répertoire VJ des lymphocytes circulant. Dans le futur, il serait intéressant de réaliser une étude comparative des combinaisons VJ de patients atteints par exemple d'hépatite C ou B. Cette analyse pourrait permettre la mise en évidence de profils d'expression spécifiques de l'évolution de la pathologie, et, à l'avenir, d'adapter les prescriptions aux patients en fonction de l'évolution prédite de leur état. Une étude plus poussée par l'analyse de nouvelles combinaisons VJ, mais aussi de nouveaux témoins, permettra de consolider la modélisation chez l'homme et d'affiner l'évolution des profils de réarrangement VJ au niveau expérimental.

La souplesse du modèle développé pour le TR α murin, notamment dans le paramétrage des variables d'entrée, a permis de développer très rapidement un modèle du TR α humain. Il est possible d'envisager d'utiliser le modèle sur d'autres TR (Sepulveda, Boucontet et al. 2005), aussi bien que sur les loci des immunoglobulines, pour une estimation plus fine des répertoires primaires des TR ou des Ig, ce qui donnerait une vision dynamique de la génération de la diversité immunitaire au cours de la différenciation des lymphocytes T et B respectivement.

Annexes

12. Annexe 1 : Nomenclature des gènes TRAV murin et humain

Nb	SOURIS	TRAV gene name	
	IMGT	Arden, B. et al. [4]	Lee, I.Y. et al.[3]
1	DV105		DV105
2	DV104		DV104
3	DV102		DV102
4	DV101		DV101
5	TRAV23		TRAV23P
6	TRAV22		TRAV22P
7	TRAV21/DV12	AV6S1/DV6S2	TRADV21
8	TRAV20	AV16S2P	TRAV20P
9	TRAV19	AV20S1	TRAV19
10	TRAV18		TRAV18
11	TRAV17	AV9S2	TRAV17
12	TRAV13-5		TRAV13-5
13	TRAV12-4		TRAV12-4P
14	TRAV3-4		TRAV3-4
15	TRAV15-3		TRAV15-3P
16	TRAV14-3	AV2S3	TRAV14-3
17	TRAV13-4/DV7	DV10S7	TRADV13-4
18	TRAV16	AV17S3	TRAV16
19	TRAV8-2		TRAV8-2
20	TRAV13-3	AV10S5	TRAV13-3
21	TRAV7-6	AV1S5	TRAV7-6
22	TRAV6-7/DV9	DV4S8	TRAV6-7
23	TRAV5-4	AV13S2	TRAV5-4
24	TRAV4-4/DV10	ADV11S5	TRADV4-4
25	TRAV9-4	AV3S1	TRAV9-4
26	TRAV3-3	AV5S1/AV5S2	TRAV3-3
27	TRAV15-2/DV6-2	DV7S4	TRDV15-2
28	TRAV14-2	AV2S1/AV2S9	TRAV14-2
29	TRAV13-2	AV10S1	TRAV13-2
30	TRAV12-3	AV8S2/AV8S10	TRAV12-3
31	TRAV5-3		TRAV5-3P
32	TRAV9-3	AV3S7	TRAV9-3
33	TRAV12-2	AV8S13	TRAV12-2
34	TRAV5-2		TRAV5-2P
35	TRAV4-3	AV11S4	TRAV4-3
36	TRAV9-2		TRAV9-2
37	TRAV3-2		TRAV3-2P
38	TRAV15-1/DV6-1	ADV7S2	TRADV15-1
39	TRAV14-1	AV2S2/AV2S7	TRAV14-1
40	TRAV13-1	AV10S8	TRAV13-1
41	TRAV12-1	AV8S4/AV8S7/AV8S9/	TRAV12-1
42	TRAV7-5		TRAV7-5
43	TRAV11	AV14S2	TRAV11
44	TRAV6-6	AV4S4	TRAV6-6
45	TRAV10	AV15S1	TRAV10
46	TRAV6-5		TRAV6-5
47	TRAV9-1		TRAV9-1
48	TRAV8-1		TRAV8-1
49	TRAV7-4	AV1S4	TRAV7-4
50	TRAV6-4		TRAV6-4
51	TRAV7-3	AV1S1	TRAV7-3
52	TRAV6-3		TRAV6-3
53	TRAV4-2	AV11S2	TRAV4-2
54	TRAV7-2		TRAV7-2
55	TRAV15D-3		TRAV15D-3P
56	TRAV14D-3/DV8	DV2S8	TRADV14D-3
57	TRAV13D-4	AV10S3	TRAV13D-4
58	TRAV16D/DV11	AV17S1/ADV17S2	TRADV16D
59	TRAV8D-2		TRAV8D-2
60	TRAV13D-3		TRAV13D-3
61	TRAV7D-6		TRAV7D-6
62	TRAV6D-7	AV4S9	TRAV6D-7
63	TRAV5D-4	AV13S1	TRAV5D-4
64	TRAV4D-4	AV11S1/AV11S6	TRAV4D-4
65	TRAV9D-4	AV3S2/AV3S5	TRAV9D-4
66	TRAV3D-3		TRAV3D-3
67	TRAV15D-2/DV6D-2	DV7S5	TRADV15D-2
68	TRAV14D-2		TRAV14D-2
69	TRAV13D-2	AV10S6	TRAV13D-2
70	TRAV12D-3	AV8S6/AV8S14	TRAV12D-3
71	TRAV5D-3		TRAV5D-3P
72	TRAV9D-3	AV3S3	TRAV9D-3
73	TRAV12D-2	AV8S3/AV8S5	TRAV12D-2
74	TRAV5D-2		TRAV5D-2P
75	TRAV4D-3	AV11S3/AV11S7	TRAV4D-3
76	TRAV9D-2	AV3S6	TRAV9D-2
77	TRAV3D-2		TRAV3D-2P
78	TRAV15D-1/DV6D-1	ADV7S1/DV7S6	TRADV15D-1
79	TRAV14D-1	AV2S4	TRAV14D-1
80	TRAV13D-1	AV10S2/AV10S9	TRAV13D-1
81	TRAV12D-1	AV8S1/AV8S8/AV8S1	TRAV12D-1
82	TRAV7D-5		TRAV7D-5
83	TRAV11D	AV14S1	TRAV11D
84	TRAV6D-6	AV4S3/AV4S7/AV4S1	TRAV6D-6
85	TRAV10D		TRAV10D
86	TRAV6D-5	AV4S2	TRAV6D-5
87	TRAV9D-1	AV3S4/AV3S8	TRAV9D-1
88	TRAV8D-1	AV18S1	TRAV8D-1
89	TRAV7D-4	AV1S2/AV1S8	TRAV7D-4
90	TRAV6D-4		TRAV6D-4
91	TRAV7D-3	AV1S3/AV1S6	TRAV7D-3
92	TRAV6D-3	AV4S6/AV4S12	TRAV6D-3
93	TRAV4D-2		TRAV4D-2P
94	TRAV7D-2	AV1S7	TRAV7D-2
95	TRAV6-2	AV4S5	TRAV6-2
96	TRAV7-1		TRAV7-1
97	TRAV6-1	AV4S11	TRAV6-1
98	TRAV5-1		TRAV5-1
99	TRAV4-1		TRAV4-1P
100	TRAV3-1	AV5S3	TRAV3-1
101	TRAV2	AV12S1	TRAV2
102	TRAV1	AV19S1	TRAV1

Tableau 6 Nomenclature IMGT des gènes TRAV murin, et comparaison avec les nomenclatures précédentes. Les gènes sont classés suivant leur position sur la région V. En haut se trouvent les gènes V les plus proches des J (proximaux). En bas se trouvent les gènes V les plus éloignés (distaux) des gènes J. En gris sont indiqués les gènes V qui ont été retrouvé dans des transcrits TRDV. On voit que certain V sont commun au locus alpha et delta.

Nb	HUMAIN	TRAV gene name	
	IMGT	Boysen et al. [2]	Arden et al. [3]
1	TRDV3		
2	TRDV2		
3	TRDV1		
4	TRAV41	41S1	19S1
5	TRAV40	40S1	31S1
6	TRAV39	39S1	27S1
7	TRAV38-2/DV8	hADV38S2	14S1-ADV14S1
8	TRAV38-1	38S1	14S2
9	TRAV37	37S1	
10	TRAV36/DV7	hADV36S1	28S1-DV28S1
11	TRAV35	35S1	25S1
12	TRAV34	34S1	26S1
13	TRAV26-2	26S2	4S1
14	TRAV33	33S1	
15	TRAV32	32S1	
16	TRAV31	31S1	
17	TRAV30	30S1	29S1
18	TRAV29/DV5	hADV29S1	21S1-ADV21S1
19	TRAV28	28S1	
20	TRAV27	27S1	10S1
21	TRAV8-7	8S7	
22	TRAV26-1	26S1	4S2
23	TRAV25	25S1	32S1
24	TRAV24	24S1	18S1
25	TRAV23/DV6	hADV23S1	17S1-ADV17S1
26	TRAV22	22S1	13S1
27	TRAV21	21S1	23S1
28	TRAV20	20S1	30S1
29	TRAV19	19S1	12S1
30	TRAV18	18S1	
31	TRAV17	17S1	3S1
32	TRAV16	16S1	9S1
33	TRAV8-6	8S6	1S3
34	TRAV12-3	12S3	2S2
35	TRAV15	15S1	
36	TRAV9-2	9S2	22S1
37	TRAV14/DV4	hADV14S1	6S1-ADV6S1
38	TRAV13-2	13S2	8S2
39	TRAV8-5	8S5	
40	TRAV8-4	8S4	1S2
41	TRAV12-2	12S2	2S1
42	TRAV13-1	13S1	8S1
43	TRAV8-3	8S3	1S4
44	TRAV8-2	8S2	1S5
45	TRAV12-1	12S1	2S3
46	TRAV11	11S1	
47	TRAV10	10S1	24S1
48	TRAV9-1	9S1	
49	TRAV8-1	8S1	1S1
50	TRAV7	7S1	
51	TRAV6	6S1	5S1
52	TRAV5	5S1	15S1
53	TRAV4	4S1	20S1
54	TRAV3	3S1	16S1
55	TRAV2	2S1	11S1
56	TRAV1-2	1S2	7S2
57	TRAV1-1	1S1	7S1

Tableau 7 Nomenclature IMGT des gènes TRAV humain, et comparaison avec les nomenclatures précédentes. Les gènes sont classés suivant leur position sur la région V. En haut se trouvent les gènes V les plus proches des J (proximaux). En bas se trouvent les gènes V les plus éloignés (distaux) des gènes J. En gris sont indiqués les gènes V qui ont été retrouvé dans des transcrits TRDV. On voit que certain V sont commun aux loci alpha et delta.

13. Annexe 2 : Séquences des RSS murin et humain et pourcentage d'homologie

Tableau 8 Compilation des RSS des gènes TRAJ murins : Sortie in silico des séquences signal de recombinaison à partir des numéros d'accèsion GenBank AE000662. Un RSS de gène J est composé d'un heptamère (7pb), d'un spacer (12pb) et d'un nonamère (9pb). La séquence consensus est indiquée (en inversé) en tête de colonne correspondante. Le pourcentage d'homologie de séquence avec le consensus est indiqué (en bleu) dans la colonne précédant la séquence de l'heptamère, du spacer et du nonamère. Le score global d'homologie de séquence du RSS par rapport au consensus général est indiqué (en rouge) dans la dernière colonne.

	Heptamère		Spacer		Nonamère		Score RSS
	% homologie	cactgtg	% homologie	aaag-----	% homologie	-gtttttgt	
P-TRAJ1	28,6%	aggactg	75,0%	taaggaagcatt	25,0%	gtgttactg	32,5%
TRAJ2	57,2%	tgcactg	100,0%	aaaggtgtcacc	75,0%	ggcttctgt	67,4%
P-TRAJ3	28,6%	tggactg	100,0%	aaagaccttacc	50,0%	ggcttcttc	45,6%
TRAJ4	85,7%	tactgtg	100,0%	aaagcacctttc	75,0%	aattcttgt	79,7%
TRAJ5	57,2%	tagggtg	50,0%	actgtgctgtac	87,5%	ggtatttgt	65,1%
TRAJ6	100,0%	cactgtg	75,0%	aaaggctttcct	37,5%	ggtttcatc	67,9%
TRAJ7	71,4%	cagagtg	75,0%	gaagcacttact	100,0%	cgtttttgt	79,5%
P-TRAJ8	85,7%	cacagtg	50,0%	atggattaatgt	50,0%	ccaccttgt	63,0%
TRAJ9	100,0%	cactgtg	25,0%	cacaggacaaat	75,0%	ccatttttgt	75,1%
TRAJ11	71,4%	tgctgtg	50,0%	ggagaggtttgc	87,5%	catttttgt	71,2%
TRAJ12	100,0%	cactgtg	0,0%	ctgactaagaaa	87,5%	tgtttttga	76,3%
TRAJ13	71,4%	tacagtg	100,0%	aaagacctcact	75,0%	ccatttttgt	73,6%
P-TRAJ14	42,9%	gaccacg	25,0%	cagacagcacgg	87,5%	catttttgt	55,4%
TRAJ15	100,0%	cactgtg	50,0%	agggcctcgttt	62,5%	agaatttgc	73,9%
TRAJ16	100,0%	cactgtg	25,0%	ggtggaagagat	100,0%	agtttttgt	84,6%
TRAJ17	100,0%	cactgtg	25,0%	ttggcttcagat	87,5%	tgtttttgc	79,9%
TRAJ18	100,0%	cactgtg	100,0%	aaagggggctgg	75,0%	gatttgtgt	85,8%
P-TRAJ19	57,2%	aaatctg	25,0%	tttgacagaggac	24,9%	tgggggtgac	37,6%
P-TRAJ20	71,4%	gcctgtg	25,0%	cagaatgaaata	50,0%	aggtttga	53,3%
TRAJ21	85,7%	cattgtg	75,0%	aatggtgctaac	75,0%	gctttctgt	76,1%
TRAJ22	85,7%	cacagtg	0,0%	tgttgagcaaat	87,5%	ggttcttgt	70,2%
TRAJ23	85,7%	cacagtg	50,0%	cagggtatgtaa	87,5%	tgtttttga	77,3%
TRAJ24	85,7%	cacagtg	50,0%	agacgtgtttgt	75,0%	ccatttttgt	72,5%
P-TRAJ25	85,7%	cactgta	0,0%	tgttgagacaag	75,0%	tggttttga	65,4%

TRAJ26	100,0%	cactgtg	75,0%	taagaggaaacc	87,5%	agtttttgc	87,0%
TRAJ27	100,0%	cactgtg	50,0%	aatagcacggag	75,0%	ggttattgc	78,7%
TRAJ28	71,4%	ttctgtg	100,0%	aaagaaaggaaa	87,5%	agtttttgc	78,3%
P-TRAJ29	100,0%	cactgtg	50,0%	taggaagatggt	24,9%	ggggatgag	59,6%
TRAJ30	85,7%	cacagtg	50,0%	tatggtcccaat	75,0%	cgtttttggg	72,5%
TRAJ31	57,2%	tgatgtg	100,0%	aaaggcaagaga	75,0%	ggtttcagt	67,4%
TRAJ32	71,4%	ggctgtg	75,0%	aaggctctgcag	87,5%	agttattgt	74,8%
TRAJ33	71,4%	gtctgtg	75,0%	taaggtttttgt	100,0%	tgtttttgt	79,5%
TRAJ34	100,0%	cactgtg	50,0%	agaacatgatat	100,0%	ggtttttgt	88,2%
TRAJ35	100,0%	cactgtg	0,0%	tgtagagtcagc	100,0%	agtttttgt	81,1%
P-TRAJ36	85,7%	cactatg	50,0%	actgggcagaaa	62,5%	tgctattga	67,8%
TRAJ37	42,9%	tggggtg	100,0%	aaagtgcagcat	100,0%	agtttttgt	70,9%
TRAJ38	85,7%	aactgtg	100,0%	aaagctctctat	75,0%	catttctgt	79,7%
TRAJ39	100,0%	cactgtg	50,0%	tgagctggagat	87,5%	ggtttttgc	83,4%
TRAJ40	100,0%	cactgtg	75,0%	agagacacagaa	87,5%	ggtttatgt	87,0%
P-TRAJ41	100,0%	cactgtg	25,0%	tcaagaaggacg	37,5%	gaggttgtt	60,8%
TRAJ42	85,7%	gactgtg	75,0%	aaggcccaaat	87,5%	ggttattgt	80,9%
TRAJ43	85,7%	tactgtg	25,0%	tagagtgtgtat	100,0%	agtttttgt	78,5%
TRAJ44	85,7%	cacagtg	50,0%	tatggagtgtct	75,0%	ggtttctga	72,5%
TRAJ45	71,4%	cagggtg	75,0%	caaggcttgctt	87,5%	ggtttatgt	74,8%
P-TRAJ46	71,4%	gacagtg	25,0%	gctgttattgga	75,0%	gttttctgt	62,8%
TRAJ47	71,4%	ggctgtg	75,0%	acaggagtttga	62,5%	agttttgtc	65,2%
TRAJ48	100,0%	cactgtg	75,0%	aatggcttgcca	87,5%	ggattttgt	87,0%
TRAJ49	85,7%	cacagtg	50,0%	tgaggttattgt	100,0%	ggtttttgt	82,1%
TRAJ50	71,4%	tgctgtg	100,0%	aaagggttgatt	87,5%	gttttttgt	78,3%
P-TRAJ51	42,9%	tgacagag	75,0%	aaaacctccac	62,5%	gtctgttgt	53,0%
TRAJ52	71,4%	tacagtg	100,0%	aaaggcccgtgt	100,0%	tgtttttgt	83,1%
TRAJ53	71,4%	ggctgtg	100,0%	aaagcctcccca	100,0%	tgtttttgt	83,1%
P-TRAJ54	57,2%	cagagag	0,0%	gggctccctcaa	24,9%	aacctggct	34,0%
P-TRAJ55	14,3%	gttagca	0,0%	gggtagctccct	75,0%	tggttttga	34,8%
TRAJ56	85,7%	cattgtg	75,0%	agagtcccgtgt	100,0%	agtttttgt	85,7%
TRAJ57	57,2%	gggtgtg	75,0%	aaggcagtgtgt	87,5%	agtatttgt	68,6%
TRAJ58	57,2%	tgacagtg	100,0%	aaagcccttcag	87,5%	agtttttgc	72,2%
P-TRAJ59	28,6%	gcccatac	75,0%	caaggacagaca	24,9%	tttgggatt	32,5%
P-TRAJ60	85,7%	cactctg	0,0%	ctcaggaccag	37,5%	actctaagg	51,1%
P-TRAJ61	71,4%	ttctgtg	100,0%	aaagggtgctcac	100,0%	ggtttttgt	83,1%

Tableau 9 Compilation des RSS des gènes TRAV murins : Sortie in silico des séquences signal de recombinaison à partir des numéros d'accèsion GenBank AE000658, AE000659, AE000660, AE000661. Un RSS de gène V est composé d'un heptamère (7pb), d'un spacer (22/23pb) et d'un nonamère (9pb). La séquence consensus est indiquée (en inversé) en tête de colonne correspondante. Le pourcentage d'homologie de séquence avec le consensus est indiqué (en bleu) dans la colonne précédant la séquence de l'heptamère, du spacer et du nonamère. Le score globale d'homologie de séquence du RSS par rapport au consensus général est indiqué (en rouge) dans la dernière colonne.

	Heptamère		Spacer		Nonamère		Score RSS
	% homologie	CACAGTG	% homologie	--TC,C/G,--A,G/C,-- ,A/C,A/C,CTGCAGCCTG	% homologie	ACACAAAACC	
V1	100,0%	CACAGTG	35,3%	ACTATGCTAACCTCTGCAGCTGC	77,8%	ACAGAAATT	0,74
V2	100,0%	CACAGTG	53,0%	AGACAGCTGGCTCTGCAGCTGTG	22,2%	CAGTATCCT	0,59
V3-1	100,0%	CACAGTG	88,2%	CCCCCAGGCACCTGCAGTCTGC	88,9%	ACAAAAACC	0,93
P-V4-1P	71,4%	AATAGTG	70,6%	CTCCCAGCAAGGCTAGAGCCTGG	55,6%	CTCCAAACT	0,65
V5-1	100,0%	CACAGTG	100,0%	CTCCCAGGCACCTGCAGCCTGT	66,7%	CTTCAAACC	0,88
V6-1	100,0%	CACAGTG	58,8%	ACAGAAACCACAGGGGACCTGA	77,8%	GCACAAACT	0,81
V7-1	100,0%	CACAGTG	100,0%	CTCCCAGGCACCTGCAGCCTGT	77,8%	ATGCAAACC	0,92
V6-2	100,0%	CACAGTG	58,8%	GTAGAAACCACAGGGGAACTGA	77,8%	GCACAAACT	0,81
V7D-2	100,0%	CACAGTG	100,0%	ATCCCAGGCACCTGCAGCCTGC	100,0%	ACACAAAACC	1,00
P-V4D-2P	100,0%	CACAGTG	70,6%	CTCCCAGCAAGGCTAGAGCCTGG	66,7%	CCTCAAACC	0,80
V6D-3	100,0%	CACAGTG	53,0%	GCAGAAACTGCAGGGGACCTGA	55,6%	GCATAAAGT	0,71
V7D-3	100,0%	CACAGTG	94,1%	CTTCCCAGACACCTGCAGCCTGT	77,8%	ATGCAAACC	0,90
V6D-4	100,0%	CACAGTG	53,0%	ACAGAAACTGTAGGGGAGCTGA	77,8%	GCACAAACT	0,79
V7D-4	100,0%	CACAGTG	88,2%	CTCTCCAGGCACCTGCAGGCTGC	88,9%	ACCCAAAACC	0,93
V8D-1	100,0%	CACAGTG	70,6%	TGTGGCGGGCACCTGCAGCCCTG	88,9%	ACACAAAAC	0,88
V9D-1	100,0%	CACAGTG	47,1%	TGTGGGGCTGCAGGGGAGCTGA	100,0%	ACACAAAACC	0,86
V6D-5	100,0%	CACAGTG	53,0%	GCAGAAACCACAAGGGGAGCTGA	77,8%	GCACAAACT	0,79
V10D	85,7%	CACACTG	88,2%	TTTCTTAGGCACCTGCAGCCTAT	66,7%	CCTCAAACC	0,79
V6D-6	100,0%	CACAGTG	53,0%	GCAGAAACTGCAGGGGAGCTGA	88,9%	GCACAAACC	0,83
V11D	71,4%	CACCATG	82,4%	CTCCCAGAAACTTACAGCCAGT	77,8%	ACCCAAAACC	0,77
V7D-5	100,0%	CACAGTG	100,0%	CTCCCAGGCACCTGCAGCCTGC	77,8%	ACCTAAAACC	0,92
V12D-1	100,0%	CACAGTG	58,8%	AGGGAGACTGCAGGGGAAGCTGC	77,8%	ACATGAACC	0,81
V13D-1	100,0%	CACAGTG	94,1%	CTCCCACACACCTGCAGCCCGC	100,0%	ACACAAAACC	0,98
V14D-1	100,0%	CACAGTG	76,5%	CTCTCCAGGCACCTGTAGCCCAT	88,9%	ACCCAAAACC	0,90

V15D-1	100,0%	CACAGTG	53,0%	TTTGAAGTGACAGCACAAGCTGA	77,8%	ACAAAAACT	0,79
P-V3D-2P	100,0%	CACAGTG	70,6%	TCCTTCAGGCACGTCTCTGCTTGT	77,8%	ACACTGACC	0,84
V9D-2	100,0%	CACAGTG	47,1%	TGTGGGGCTGCAGGGGAGCTGA	88,9%	ACACAAACA	0,82
V4D-3	100,0%	CACAGTG	76,5%	CTCCCAGCAAGGCTGGAGCCTGG	66,7%	CCTCAAACCT	0,82
TRAV5D-2P	100,0%	CACAGTG	29,4%	TTCCATGGCTCTAGTAGCCTGTC	33,3%	CTCAAACCT	0,57
V12D-2	100,0%	CACAGTG	58,8%	AGGGAGACTGCAGGGGAAGCTGC	77,8%	ACATGAACC	0,81
V9D-3	100,0%	CACAGTG	41,2%	TGTGGGGCTGCAGGGGAGCCGA	33,3%	ATCTACAAT	0,60
TRAV5D-3P	85,7%	CACAGGG	35,3%	ACCTGCAGTCCATCTGTCTCAA	88,9%	ACACAAACA	0,73
V12D-3	100,0%	CACAGTG	58,8%	AGGGAGACTGCAGGGGAAGCTGC	77,8%	ACATGAACC	0,81
V13D-2	100,0%	CACAGTG	100,0%	CTCCCCACACACCTGCAGCCTGT	100,0%	ACACAAACC	1,00
V14D-2	100,0%	CACAGTG	82,4%	CTCTCCAGGCACCTGCAGCCCAT	88,9%	ACCCAAACC	0,91
V15D-2	85,7%	CACGGTG	47,1%	TTTGAAGTGATAGCACAAGCTGA	77,8%	ACAAAAACT	0,72
V3D-3	100,0%	CACAGTG	100,0%	CTCCCCAGGCACCTGCAGCCTGC	88,9%	ACATAAACC	0,96
V9D-4	100,0%	CACAGTG	41,2%	TGTGGGGCTGCAGGGGAGCCGA	77,8%	ACAAAAACA	0,76
V4D-4	100,0%	CACAGTG	70,6%	CTCCCAGCAAGGCTGGAGCCCGG	66,7%	CCCCAAACT	0,80
V5D-4	85,7%	CACACTG	88,2%	CTCCACGGGCACCTGCAGTCTGT	77,8%	CCTCAAACC	0,84
V6D-7	100,0%	CACAGTG	53,0%	ACAGAAACTGCAGGGGAGCTGA	77,8%	GCACAAACT	0,79
V7D-6	100,0%	CACAGTG	100,0%	CTCCCCAGGCACCTGCAGCCTGC	77,8%	ACCCAAACA	0,92
V13D-3	100,0%	CACAGTG	94,1%	CTCCCCACACACCTGCAGCCCGC	100,0%	ACACAAACC	0,98
V8D-2	100,0%	CACAGTG	70,6%	TGTGGCGGGCACCTGCAGCCCTG	100,0%	ACACAAACC	0,92
DV16D	100,0%	CACAGTG	17,7%	GCAGAAGTATCTACAAGAGCCCC	88,9%	ACAAAAACC	0,74
13D-4	100,0%	CACAGTG	94,1%	CTCCCCACACACCTGCAGCCCGC	100,0%	ACACAAACC	0,98
DV14D-3	100,0%	CACAGTG	82,4%	CTCTCCAGGCACCTGCAGCCCAT	88,9%	ACCCAAACC	0,91
TRAV15D-3P	42,9%	AGCATAA	29,4%	CTTAGAACATCTTCACAGGTATA	11,1%	ATCGTTTTTA	0,28
TRAV7-2	100,0%	CACAGTG	35,3%	CTCCCCAGGCACCTGCAGCCTGT	77,8%	ATGCAAACC	0,74
V4-2	100,0%	CACAGTG	76,5%	CTCCCAGCAAGGCTGGAGCCTGG	66,7%	CCTCAAACG	0,82
V6-3	100,0%	CACAGTG	53,0%	GCAGAAACTGCAGGGGAGCTGA	55,6%	GCATAAAGT	0,71
V7-3	100,0%	CACAGTG	94,1%	CTTCCCAGACACCTGCAGCCTGT	77,8%	ATGCAAACC	0,90
V6-4	100,0%	CACAGTG	47,1%	ACAGAAACTGTAGGGGAGCTGA	77,8%	GCACAAACT	0,78
V7-4	100,0%	CACAGTG	82,4%	CTCTCCAGGCACCTGCAGGCTGC	88,9%	ACCCAAACC	0,91
V8-1	85,7%	CACAGTA	70,6%	TGTGGCGGGCACCTGCAGCCCTG	88,9%	ACACAAAAC	0,83
V9-1	100,0%	CACAGTG	47,1%	TGTGGGGCTGCAGGGGAGCTGA	88,9%	GCACAAACC	0,82
V6-5	100,0%	CACAGTG	47,1%	GCACAAACCACAGGGGAGTTGA	77,8%	GCACAAACT	0,78
V10	85,7%	CACACTG	88,2%	TTTCTAGGCACCTGCAGCCTAT	66,7%	CCTCAAACCT	0,79
V6-6	100,0%	CACAGTG	53,0%	GCAGAAACTGCAGGGGAGCTGA	88,9%	GCACAAACC	0,83
V11	71,4%	CACCATG	82,4%	CTCCCCAGAAACTTACAGCCAGT	77,8%	ACCCAAACC	0,77

V7-5	100,0%	CACAGTG	100,0%	CTCCCCAGGCACCTGCAGCCTGC	77,8%	ACCTAAACC	0,92
V12-1	85,7%	CACAGCG	58,8%	AGGGAGACTGCAGGGGAAGCTGC	55,6%	ACTTGAACCT	0,67
V13-1	100,0%	CACAGTG	94,1%	CTCCACACACACCTGCAGCCCGC	88,9%	ACACATAACC	0,94
V14-1	100,0%	CACAGTG	82,4%	CTCTCCAGGCACCTGCAGCCCAT	88,9%	ACCCAAAACC	0,91
TRADV15-1	100,0%	CACAGTG	58,8%	TTTGAAGTGACAGCACAAGCTGA	77,8%	ACAAAAACT	0,81
P-V3-2P	100,0%	CACAGTG	64,7%	TCCTTCAGGCACGTCTGCTTGC	77,8%	ACACTGACC	0,82
V9-2	100,0%	CACAGTG	47,1%	TGTGGGGCTGCAGGGGAGCTGA	88,9%	ACACAAAACA	0,82
V4-3	100,0%	CACAGTG	76,5%	CTCCAGCAAGGCTGGAGCCTGG	66,7%	CCCCAAACT	0,82
P-V5-2P	100,0%	CACAGTG	35,3%	TTCCATGGCTCTAGTAGCCTGTC	33,3%	CTCAAACCT	0,58
V12-2	100,0%	CACAGTG	58,8%	AGGGAGACTGCAGGGGAAGCTGC	77,8%	ACATGAACC	0,81
V9-3	100,0%	CACAGTG	47,1%	TGTGGGGCTGCAGGGGAGCTGA	88,9%	ACACAAAACA	0,82
V5-3P	85,7%	CACAGGG			33,3%	ATCTACAAT	0,43
V12-3	100,0%	CACAGTG	58,8%	AGGGAGACTGCAGGGGAAGCTGC	77,8%	ACATGAACC	0,81
V13-2	100,0%	CACAGTG	100,0%	CTCCCCACACACCTGCAGCCTGT	100,0%	ACACAAAACC	1,00
V14-2	100,0%	CACAGTG	82,4%	CTCTCTAGGCACCTGCAGCCCAT	88,9%	ACCCAAAACC	0,91
V15-2	85,7%	CACGGTG	47,1%	TTTGAAGTGATAGCACAAGCTGA	77,8%	ACAAAAACT	0,72
V3-3	100,0%	CACAGTG	100,0%	CTCCCCAGGCACCTGCAGCCTGC	88,9%	ACATAAACC	0,96
V9-4	100,0%	CACAGTG	47,1%	TGTGGGGCTGCAGGGGAGCTGA	88,9%	ACACAAAACA	0,82
DV4-4	100,0%	CACAGTG	76,5%	CTCCAGCAAGGCTGGAGCCTGG	66,7%	CCCCAAACT	0,82
V5-4	85,7%	CACACTG	94,1%	CTCCATGGGCACCTGCAGCCTGT	77,8%	CCTCAAACC	0,85
V6-7	100,0%	CACAGTG	53,0%	GCACAAACCACAAGGGGAGCTGA	77,8%	GCACAAACT	0,79
V7-6	100,0%	CACAGTG	100,0%	CTCCCCAGGCACCTGCAGCCTGC	77,8%	ACCCAAAACA	0,92
V13-3	100,0%	CACAGTG	94,1%	CTCCCCACACACCTGCAGCCCGA	100,0%	ACACAAAACC	0,98
V8-2	100,0%	CACAGTG	70,6%	TGTGGCGGGCACCTGCAGCCCTG	100,0%	ACACAAAACC	0,92
V16	100,0%	CACAGTG	17,7%	GCAGAAGTATCTACAAGAGCCCC	88,9%	ACAAAAACC	0,74
V13-4	100,0%	CACAGTG	88,2%	CTCCCCACACACCAGCAGCCCGC	100,0%	ACACAAAACC	0,97
V14-3	100,0%	CACAGTG	82,4%	CTCTCCAGGCACCTGCAGCCCAT	88,9%	ACCCAAAACC	0,91
V15-3P	0,0%	TTTTAGC	23,6%	TTACAGACTGGAAAATTCAATGA	11,1%	ATTATTTTT	0,10
V3-4	100,0%	CACAGTG	100,0%	CTCCCCAGGCACCTGCAGCCTGC	88,9%	ACATAAACC	0,96
TRAV12-4P	71,4%	TACTGTG	35,3%	CTGTGGGGACACAGGGAGGCTGC	55,6%	AGGGAAACT	0,56
V13-5	100,0%	CACAGTG	100,0%	CTCCCCAGGCACCTGCAGCCTGT	100,0%	ACACAAAACC	1,00
V17	100,0%	CACAGTG	41,2%	CCTGAGGCTGCAGGGGAGCCGA	77,8%	GCGCAAACC	0,76
V18	85,7%	CACACTG	70,6%	CTCCTGTGGCCTAGGCTGCCTGC	55,6%	CCCCAAACC	0,71
V19	71,4%	CACGGTA	76,5%	CTCTGTGCATGCCTGCAGCCTCT	77,8%	ACCCAAAACC	0,75
P-V20P	100,0%	CACAGTG	82,4%	TTCTGTAGGCATCTGCAGTCTGG	55,6%	GTGCAAAACA	0,79
DV21	85,7%	CACACTG	53,0%	TGACAGACAGGGCTGCACCTGTA	22,2%	CAAAATCTG	0,54

TRDV1	100,0%	CACAGTG	70,6%	GTCTACAGTCAGCCACAGGCTGT	66,7%	CTCCAAACC	0,80
TRAV22P	85,7%	CACGGTG	35,3%	ACCCTCAGGCACCTGCAGCCTTT	77,8%	ACCTAAACC	0,69
TRDV2-1	100,0%	CACAGTG	35,3%	ACGCAGGCAACAGAGAGGGTCTA	88,9%	ACAGAAACC	0,79
TRDV2-2	85,7%	CACGGTG	35,3%	ACACAGGCAACAGAGAGGGTCTA	88,9%	ACAGAAACC	0,73
P-V23P	100,0%	CACAGTG	41,2%	CTCTCCTGACGTTGTTGGTCTGT	77,8%	ACCTAAACC	0,76
TRDV3P	100,0%	CACAGTG	29,4%	CCAAGGACACAGTGGGAGCTGT	88,9%	ACAAAACC	0,77
deltaREC1	100,0%	CACAGTG	88,2%	GTGCACAAGCACCTGCACCCTGT	77,8%	GCATAAACC	0,89
TRDV4	85,7%	CACGGTG	35,3%	CTACAGTTCCGCTTCTGATCCCT	44,5%	TCAAGAAAG	0,57

Tableau 10 Compilation des RSS des gènes TRAJ humains : Sortie in silico des séquences signal de recombinaison à partir des numéros d'accèsion GenBank NG_001332. Un RSS de gène J est composé d'un heptamère (7pb), d'un spacer (12pb) et d'un nonamère (9pb). La séquence consensus est indiquée (en inversé) en tête de colonne correspondante. Le pourcentage d'homologie de séquence avec le consensus est indiqué (en bleu) dans la colonne précédant la séquence de l'heptamère, du spacer et du nonamère. Le score global d'homologie de séquence du RSS par rapport au consensus général est indiqué (en rouge) dans la dernière colonne.

	Heptamère		Spacer		Nonamère		Score RSS
	% homologie	CACTGTG	% homologie	AAAG-----A-	% homologie	GGTTTTTGT	
61	71,4%	TCCTGTG	100,0%	AAAGGTGCCAC	100,0%	GGTTTTTGT	0,88
P-60	85,7%	CACTATG	80,0%	AAAGGGCTGGG	77,8%	GCTTTTGGT	0,81
P-59	71,4%	TCCTGTG	80,0%	AAAGGCAGCAGC	77,8%	AGTTTATGT	0,76
58	85,7%	CACAGTG	100,0%	AAAGCCCCTCAG	88,9%	GGTTTTTGC	0,90
57	42,9%	GGGGGTG	80,0%	AAAGCAGTCTGT	77,8%	AGTATTTGT	0,64
56	85,7%	CATTGTG	40,0%	AGATCCTCGTGT	88,9%	AGTTTTTGT	0,79
P-55	42,9%	GTTAGTG	20,0%	GGATGGATCCCT	55,6%	TGATTTGGA	0,44
54	42,9%	TGTGGTG	80,0%	AAAGCCTCGTGC	77,8%	AGTTTCTGT	0,64
53	71,4%	GGCTGTG	80,0%	AAAGCCTTCTGT	77,8%	TGTTTCTGT	0,76
52	57,2%	TGCAGTG	100,0%	AAAGGCCTCCAG	88,9%	GGTTCTTGT	0,78
P-51	42,9%	AGCAGGG	80,0%	AAAACCTCCCTGA	55,6%	AGTCGTCGT	0,55
50	71,4%	GGCTGTG	100,0%	AAAGGTTTGGAT	77,8%	AGTTATTGT	0,79
49	85,7%	CACAGTG	60,0%	TGAGCTTCCCTAT	100,0%	GGTTTTTGT	0,87
48	100,0%	CACTGTG	80,0%	AATGACTTAGAA	88,9%	GGTTTTTGC	0,92
47	85,7%	CGCTGTG	60,0%	AGAGGAGTTTGA	88,9%	TGTTTTTGT	0,83
46	57,2%	AGCCGTG	100,0%	AAAGCTGCTGAC	77,8%	TGTTTCTGT	0,73
45	71,4%	CAGAGTG	80,0%	AAAGGGTTGGCC	77,8%	AGTTTATGT	0,76
44	85,7%	CACAGTG	40,0%	TATGAAGCATCT	88,9%	GGTTTCTGT	0,79
43	85,7%	TACTGTG	40,0%	TAGAGCATGTAT	100,0%	GGTTTTTGT	0,84
42	85,7%	GACTGTG	100,0%	AAAGCCCATAG	77,8%	GATTATTGT	0,85
41	100,0%	CACTGTG	40,0%	TTAGGGAGAACG	88,9%	GTTTTTTGT	0,85
40	100,0%	CACTGTG	80,0%	AGAGACACATAA	88,9%	GGTTTATGT	0,92
39	100,0%	CACTGTG	60,0%	TCAGCCGAAGAT	88,9%	GGTTTTTGC	0,89
38	85,7%	GACTGTG	100,0%	AAAGCTTTCTAT	88,9%	GGTTTTGGT	0,90
37	57,2%	TAGAGTG	100,0%	AAAGTACAGCAT	88,9%	AGTTTTTGT	0,78

36	100,0%	CACTGTG	60,0%	ACTGGGCAGAAA	88,9%	TGTTTTTGT	0,89
35	85,7%	CATTGTG	60,0%	AAAAGAATGAGC	100,0%	GGTTTTTGT	0,87
34	100,0%	CACTGTG	60,0%	AGATCTCAGTAT	100,0%	GGTTTTTGT	0,93
33	71,4%	GTCTGTG	60,0%	TAAGGTTTTTGT	100,0%	GGTTTTTGT	0,81
32	85,7%	GACTGTG	80,0%	AAGGCTCTGAAG	77,8%	GGTTAGTGT	0,81
31	71,4%	TGCTGTG	100,0%	AAAGGCAGGAAG	77,8%	GGTTTCAGT	0,79
30	85,7%	CACAGTG	60,0%	TATGGTCCCAAT	88,9%	AGTTTTTGT	0,83
29	100,0%	CACTGTG	60,0%	TATGGAGGAAAT	100,0%	GGTTTTTGT	0,93
28	85,7%	CTCTGTG	100,0%	AAAGAAAGGAAA	88,9%	GGTTTTTGC	0,90
27	85,7%	GACTGTG	60,0%	AATAGCACTAAA	88,9%	GGTTATTGC	0,83
26	100,0%	CACTGTG	60,0%	TGAGCCAGAAA	88,9%	GGTTTTTGC	0,89
P-25	85,7%	CACTATG	20,0%	TGCTGAGATAAT	88,9%	GGTTTTTGA	0,76
24	85,7%	CACAGTG	60,0%	AGAGGTGTTTGT	66,7%	CCATTTTGT	0,73
23	85,7%	CACAGTG	40,0%	CAGGATATGTAA	77,8%	TGTTTTTGA	0,75
22	71,4%	CATAGTG	20,0%	TGTTGAGCAAAT	100,0%	GGTTTTTGT	0,74
21	85,7%	CATGGTG	80,0%	AATGCCAATAAA	77,8%	ATTTTTTGT	0,81
20	100,0%	CACTGTG	20,0%	TAGGCGACCTCG	44,5%	AGGTTTGTG	0,63
P-19	57,2%	AGATGTG	40,0%	TTGCGAGAGGAC	33,4%	TGAGGTGAT	0,44
18	85,7%	CATTGTG	100,0%	AAAGGGGACCAG	77,8%	GGTTCATGT	0,85
17	85,7%	CATTGTG	40,0%	TGGGCCTTAAAT	88,9%	GGTTTTTGC	0,79
16	100,0%	CACTGTG	20,0%	GGTACAATAGAT	100,0%	GGTTTTTGT	0,86
15	100,0%	CACTGTG	40,0%	AGGGCCTCATTT	77,8%	GGTATTTGC	0,80
14	71,4%	TGCTGTG	60,0%	CAGGCAGCACAG	77,8%	CATTTTTGT	0,72
13	71,4%	TACAGTG	100,0%	AAAGGCAGGCAT	66,7%	TCATTTTGT	0,74
12	100,0%	CACTGTG	20,0%	CTGACTAAGAAA	77,8%	TGTTTTTGA	0,77
11	57,2%	TATAGTG	40,0%	ATGGGGATTTGC	77,8%	CATTTTTGT	0,63
10	100,0%	CACTGTG	60,0%	TGAGGCATCAAA	55,6%	AGTTTATTG	0,75
9	100,0%	CACTGTG	20,0%	CGCAGTGCAAAT	66,7%	CCATTTTGT	0,72
8	71,4%	CAGAGTG	80,0%	ACAGAGTTATGT	66,7%	CCATTTTGT	0,71
7	85,7%	CACAGTG	40,0%	GTAATACACTTA	66,7%	GAGGTTTTT	0,70
6	100,0%	CACTGTG	80,0%	AAAGGCTGTCCT	66,7%	GGTTTTATC	0,83
5	71,4%	CAGGGTG	60,0%	ACTGTGATGTAC	88,9%	GGATTTTGT	0,77
4	85,7%	GATTGTG	80,0%	AAAGCACCATCT	77,8%	AGTTCTTGT	0,81
3	85,7%	CACAGTG	80,0%	AAAGACCTTACC	55,6%	GGTTATCTC	0,72
P-2	57,2%	TACGGTA	60,0%	AATGGTGTACC	66,7%	AGTTTGTGC	0,62
P-1	28,6%	GGGCATG	40,0%	AATAGAGACACG	77,8%	GGATTCTGT	0,51

Tableau 11 Compilation des RSS des gènes TRAV humains : Sortie in silico des séquences signal de recombinaison à partir des numéros d'accèsion GenBank NG_001332. Un RSS de gène V est composé d'un heptamère (7pb), d'un spacer (22/23pb) et d'un nonamère (9pb). La séquence consensus est indiquée (en inversé) en tête de colonne correspondante. Le pourcentage d'homologie de séquence avec le consensus est indiqué (en bleu) dans la colonne précédant la séquence de l'heptamère, du spacer et du nonamère. Le score global d'homologie de séquence du RSS par rapport au consensus général est indiqué (en rouge) dans la dernière colonne.

	Heptamère		Spacer		Nonamère		Score RSS
	% homologie	CACAGTG	% homologie	-T---,C/G,AG-,C/G,AC- CTG-A,A/G,C/G,CTGT	% homologie	AC,A/C,CAAACC	
V1-1	100,0%	CACAGTG	56,3%	ACTATGAGGCCTCCTTAACTGTG	55,6%	CCAAAATTC	0,72
V1-2	85,7%	CACGGTG	62,5%	ACTATGAGGCCTCTTTAGCTGC	55,6%	ACCAAATT	0,69
V2	85,7%	CACAGAG	56,3%	GCAGGGAACCCATGAAGAGCTGA	77,8%	ACAGAAACA	0,75
V3	85,7%	CACACTG	56,3%	ATAGGGGCTGCAGGGGGAGCAGA	88,9%	ACACAAACT	0,80
V4	100,0%	CACAGTG	50,0%	AGACAGATGGGCCTGCACCTGTG	33,3%	CCGTTTTCC	0,63
V5	85,7%	CACATTG	93,8%	CTTCTCAGGCACCTGTATCCTGT	100,0%	ACCCAAACC	0,93
V6	85,7%	CACAGTA	75,0%	GTGCCCTGGCAGCTGCTTCCTGC	88,9%	ACCCAAACT	0,84
V7	85,7%	CACAGTA	93,8%	CTCCCTAGGCACCTGCAACCTGT	77,8%	ATCCAAACA	0,85
V8-1	100,0%	CACAGTG	62,5%	TCTGGGACTGCAAAGGGAGCTGA	88,9%	ACACAAACT	0,86
V9-1	100,0%	CACAGTG	68,8%	ACAGGGACTGCAGGGGAAGCTGA	77,8%	GCACAAACT	0,84
V10	85,7%	CACTGTG	87,5%	CTCCACAGGCACTTGAAGCCAGT	77,8%	ATGCAAACC	0,84
P-V11	28,6%	CAGTAGT	50,0%	GTCTCCCCAGCACCTGCAGCCTG	66,7%	TACCATAAC	0,48
V12-1	100,0%	CACAGTG	93,8%	CTCCCCAGACACCTGCAGTCTGT	100,0%	ACCCAAACC	0,99
V8-2	100,0%	CACAGTG	50,0%	CTTGAGACTGCAGGAGAGCTGA	33,3%	ACACAAGCC	0,63
V8-3	100,0%	CACAGTG	56,3%	CCTGGGTCTGCAGGGGGAGCTGA	88,9%	ACACAAACT	0,85
V13-1	85,7%	CACATTG	87,5%	CTTCCCAGGCACCTGTACCCGT	100,0%	ACACAAACC	0,92
V12-2	100,0%	CACAGTG	93,8%	CTCCCCAGACACCTGCAGTCTGT	100,0%	ACCCAAACC	0,99
V8-4	100,0%	CACAGTG	50,0%	CTTGAGACTGCAGGAGAGGTGA	33,3%	ACATAAACC	0,63
P-V8-5	100,0%	CACAGTG	43,8%	CCTGGGTCTGCAGGGGAGCTGA	33,3%	ACACAAACT	0,61
V13-2	85,7%	CACATTG	81,3%	CTTTCCAGGCATCTGTAACCATC	100,0%	ACCCAAACC	0,90
V14DV4	100,0%	CACAGTG	43,8%	ACAGAACTGTGCGAGGGAGGTGT	77,8%	ACAAAAGCC	0,77
V9-2	100,0%	CACAGTG	62,5%	ACAGGGACTGCAGGGGGAGCTGA	77,8%	GCACAAACT	0,82
P-V15	85,7%	CACAGGG	25,0%	TCCCCAAGCACCTGCAGCCTGT	66,7%	ACCACAACC	0,63
V12-3	100,0%	CACAGTG	93,8%	CTCCCCAGACACTTGCAGTCTGT	100,0%	ACCCAAACC	0,99

V8-6	100,0%	CACAGTG	43,8%	CCTGAGACTGCAGGAGAGCTGA	44,5%	ACACAAACC	0,65
V16	85,7%	CACAGTA	37,5%	GCTGGTTTTTGAAGGAAGCAGA	55,6%	ACACAAACC	0,62
V17	100,0%	CACAGTG	93,8%	TTCCCCAGGAACCTGCAGCCTCT	88,9%	ACGCAAACC	0,95
V18	100,0%	CAGAGTG	50,0%	GGAGGGACTGCAGCGAGAGCCCA	88,9%	GCACAAACC	0,83
V19	100,0%	CACAGTG	56,3%	AGATGGGTGCCTGTGGGAGCCCT	88,9%	ACAAAAACC	0,85
V20	85,7%	CACAGCG	93,8%	TTCCCCAGGCACCTGCAACTTGT	77,8%	ATCAAAACC	0,85
V21	100,0%	CACAGTG	81,3%	CACAACAGGCACCTGCAACCAAT	88,9%	ACCCAAACT	0,91
V22	100,0%	CACAGTG	93,8%	CTCCCCAGGCACCTGCGGCCTGT	100,0%	ACACAAACC	0,99
V23DV6	100,0%	CACAGTG	100,0%	CTCCCCAGGCACCTGAAGCCTGT	100,0%	ACCCAAACC	1,00
TRDV1	100,0%	CACAGTG	37,5%	TTTGAAGTGATAGTAAAAGCAAA	88,9%	ACAAAAACC	0,80
V24	100,0%	CACAGTG	87,5%	CTGTTTCAGGCACCTGCAGCCCAT	88,9%	ACGCAAACC	0,93
V25	100,0%	CACAGTG	87,5%	CTCCCCAAGCACCTGCTGCCTGT	66,7%	CTCCAAATC	0,85
V26-1	85,7%	CACACTG	50,0%	GGACAGATGGGGCTGCACCTGTG	33,3%	CAATATCTC	0,57
V8-7	85,7%	GACTGTG	56,3%	CCTGGGACTGCAGGAGGAGCTGA	88,9%	ACACAAACT	0,80
V27	100,0%	CACAGTG	87,5%	CTCTTGAGGCACCTGCTGCCTGC	100,0%	ACCCAAACC	0,97
P-V28	85,7%	CACTGTG	93,8%	CTCTTCAGACACCTGCAGCCTAT	66,7%	ACATGAAAC	0,81
V29DV5	100,0%	CACAGTG	93,8%	CTCTCCAGGCACCTGCAGCCCGT	88,9%	ACTCAAACC	0,95
V30	100,0%	CACAGTG	81,3%	ATACCCAGGCCTCCAAGACCTGT	88,9%	ACTCAAACC	0,91
P-V31	28,6%	CACTGTG	43,8%	AAGAACATGTTAGAAGAGCCTT	33,3%	ACAAAAAGA	0,34
P-V32	14,3%	CACAGAA	25,0%	CTCTTCAGGCACCTGCAACCTGT	55,6%	ACTCAAACC	0,32
P-V33	14,3%	CACAGAA	37,5%	GTAGAAATGCAGTGGAAGATAA	55,6%	ACAAAAACC	0,36
V26-2	100,0%	CACAGTG	56,3%	GGACAGATGGGGCTGCAGCTGTG	33,3%	CAATATCTC	0,64
V34	85,7%	CACAGCG	87,5%	ATCTTCAGGCCTCTATCAGCTGT	77,8%	CTCCAAACC	0,84
V35	100,0%	CACAGTG	93,8%	CTCCCCAAACACCTGCAGCCTGT	77,8%	ACTCAAACC	0,90
V36DV7	100,0%	CACAGTG	93,8%	CTCCCTAGTCACCTGCAGCCTGT	66,7%	ACTCAAATT	0,86
P-V37	14,3%	CACAGTG	25,0%	CCCACAGTCACCTGCACCCGGT	55,6%	ACCTAAAGC	0,32
V38-1	85,7%	CACAATG	37,5%	AGATGAGCAGCAGGGAGAGGCTT	88,9%	ACAGAAACC	0,75
V38-							
2DV8	100,0%	CACAGTG	31,3%	AGACAAGCAACAGGGAGAGGCTT	88,9%	ACAGAAACC	0,78
V39	100,0%	CACAGTG	68,8%	CTCCCCTGACGCCACCAGTCTGT	100,0%	ACCCAAACC	0,92
V40	85,7%	CACTGTG	43,8%	TTAAAAGCACAGTGGGAGCTAT	44,5%	ACAAAAACC	0,60
V41	100,0%	CACAGTG	93,8%	CTCCCCAGGCACCTGGAGCCCGT	77,8%	ACCTAAACT	0,90
TRDV2	71,4%	CACCCTG	62,5%	CTGCAGCTCTACTTCTGAGCAGC	77,8%	TCAAAAACC	0,72

Bibliographie

14. Bibliographie

- Abarrategui, I. and M. S. Krangel (2006). "Regulation of T cell receptor-alpha gene recombination by transcription." Nat Immunol 7(10): 1109-15.
- Abarrategui, I. and M. S. Krangel (2007). "Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination." Embo J 26(20): 4380-90.
- Ahnesorg, P., P. Smith, et al. (2006). "XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining." Cell 124(2): 301-13.
- Akamatsu, Y., N. Tsurushita, et al. (1994). "Essential residues in V(D)J recombination signals." J Immunol 153(10): 4520-9.
- Anderson, G., K. J. Hare, et al. (1999). "Positive selection of thymocytes: the long and winding road." Immunol Today 20(10): 463-8.
- Aude-Garcia, C., M. Gallagher, et al. (2001). "Preferential ADV-AJ association during recombination in the mouse T-cell receptor alpha/delta locus." Immunogenetics 52(3-4): 224-30.
- Balciunaite, G., R. Ceredig, et al. (2005). "The earliest subpopulation of mouse thymocytes contains potent T, significant macrophage, and natural killer cell but no B-lymphocyte potential." Blood 105(5): 1930-6.
- Balmelle, N., N. Zamarreno, et al. (2004). "Developmental activation of the TCR alpha enhancer requires functional collaboration among proteins bound inside and outside the core enhancer." J Immunol 173(8): 5054-63.
- Barlow, C., S. Hirotsune, et al. (1996). "Atm-deficient mice: a paradigm of ataxia telangiectasia." Cell 86(1): 159-71.
- Barnes, D. E., G. Stamp, et al. (1998). "Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice." Curr Biol 8(25): 1395-8.
- Bassing, C. H., R. E. Tillman, et al. (2003). "T cell receptor (TCR) alpha/delta locus enhancer identity and position are critical for the assembly of TCR delta and alpha variable region genes." Proc Natl Acad Sci U S A 100(5): 2598-603.
- Bidot, C. (2006). Modélisation mathématique de la réponse lymphocytaire T spécifique à une infection virale. Saint-Etienne, Ecole Nationale Supérieure des Mines de Saint-Etienne: 272.
- Bidot, C., F. Gruy, et al. (2008). "Mathematical modeling of T-cell activation kinetic." J Comput Biol 15(1): 105-28.
- Blackburn, C. C. and N. R. Manley (2004). "Developing a new paradigm for thymus organogenesis." Nat Rev Immunol 4(4): 278-89.

- Blumberg, R. S., S. Ley, et al. (1990). "Structure of the T-cell antigen receptor: evidence for two CD3 epsilon subunits in the T-cell receptor-CD3 complex." Proc Natl Acad Sci U S A 87(18): 7220-4.
- Bonneville, M. and J. J. Fournie (2005). "Sensing cell stress and transformation through Vgamma9Vdelta2 T cell-mediated recognition of the isoprenoid pathway metabolites." Microbes Infect 7(3): 503-9.
- Bories, J. C., P. Loiseau, et al. (1990). "Regulation of transcription of the human T cell antigen receptor delta chain gene. A T lineage-specific enhancer element is located in the J delta 3-C delta intron." J Exp Med 171(1): 75-83.
- Boucontet, L., N. Sepulveda, et al. (2005). "Mechanisms controlling termination of V-J recombination at the TCRgamma locus: implications for allelic and isotypic exclusion of TCRgamma chains." J Immunol 174(7): 3912-9.
- Buck, D., L. Malivert, et al. (2006). "Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly." Cell 124(2): 287-99.
- Buffat, L. and J. Y. Mary (1992). "Automatic search for model to simulate the differentiation of T lymphocytes within the thymus." Acta Biotheor 40(2-3): 205-20.
- Callebaut, I., D. Moshous, et al. (2002). "Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family." Nucleic Acids Res 30(16): 3592-601.
- Capone, M., R. D. Hockett, Jr., et al. (1998). "Kinetics of T cell receptor beta, gamma, and delta rearrangements during adult thymic development: T cell receptor rearrangements are present in CD44(+)CD25(+) Pro-T thymocytes." Proc Natl Acad Sci U S A 95(21): 12522-7.
- Capone, M., F. Watrin, et al. (1993). "TCR beta and TCR alpha gene enhancers confer tissue- and stage-specificity on V(D)J recombination events." Embo J 12(11): 4335-46.
- Ceredig, R. (1990). "Intrathymic proliferation of perinatal mouse alpha beta and gamma delta T cell receptor-expressing mature T cells." Int Immunol 2(9): 859-67.
- Chakalova, L., E. Debrand, et al. (2005). "Replication and transcription: shaping the landscape of the genome." Nat Rev Genet 6(9): 669-77.
- Chen, M. L. and C. L. Kuo (2001). "A conserved sequence block in the murine and human T cell receptor Jalpha loci interacts with developmentally regulated nucleoprotein complexes in vitro and associates with GATA-3 and octamer-binding factors in vivo." Eur J Immunol 31(6): 1696-705.
- Chien, Y. H., M. Iwashima, et al. (1987). "A new T-cell receptor gene located within the alpha locus and expressed early in T-cell differentiation." Nature 327(6124): 677-82.
- Chien, Y. H., M. Iwashima, et al. (1987). "T-cell receptor delta gene rearrangements in early thymocytes." Nature 330(6150): 722-7.
- Cobb, R. M., K. J. Oestreich, et al. (2006). "Accessibility control of V(D)J recombination." Adv Immunol 91: 45-109.

- Couez, D., M. Malissen, et al. (1991). "Each of the two productive T cell receptor alpha-gene rearrangements found in both the A10 and BM 3.3 T cell clones give rise to an alpha chain which can contribute to the constitution of a surface-expressed alpha beta dimer." Int Immunol 3(7): 719-29.
- Cui, X., Y. Yu, et al. (2005). "Autophosphorylation of DNA-dependent protein kinase regulates DNA end processing and may also alter double-strand break repair pathway choice." Mol Cell Biol 25(24): 10842-52.
- Davis, M. M. and P. J. Bjorkman (1988). "T-cell antigen receptor genes and T-cell recognition." Nature 334(6181): 395-402.
- Davodeau, F., M. Difilippantonio, et al. (2001). "The tight interallelic positional coincidence that distinguishes T-cell receptor Jalpha usage does not result from homologous chromosomal pairing during ValphaJalpha rearrangement." Embo J 20(17): 4717-29.
- Davodeau, F., M. A. Peyrat, et al. (1995). "Dual T cell receptor beta chain expression on human T lymphocytes." J Exp Med 181(4): 1391-8.
- de Chasseval, R. and J. P. de Villartay (1993). "Functional characterization of the promoter for the human germ-line T cell receptor J alpha (TEA) transcript." Eur J Immunol 23(6): 1294-8.
- de la Hera, A., U. Muller, et al. (1991). "Structure of the T cell antigen receptor (TCR): two CD3 epsilon subunits in a functional TCR/CD3 complex." J Exp Med 173(1): 7-17.
- de Villartay, J. P. and D. I. Cohen (1990). "Gene regulation within the TCR-alpha/delta locus by specific deletion of the TCR-delta cluster." Res Immunol 141(7): 618-23.
- de Villartay, J. P., D. Lewis, et al. (1987). "Deletional rearrangement in the human T-cell receptor alpha-chain locus." Proc Natl Acad Sci U S A 84(23): 8608-12.
- Dik, W. A., K. Pike-Overzet, et al. (2005). "New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling." J Exp Med 201(11): 1715-23.
- Downs, J. A. and S. P. Jackson (2004). "A means to a DNA end: the many roles of Ku." Nat Rev Mol Cell Biol 5(5): 367-78.
- Drouet, J., C. Delteil, et al. (2005). "DNA-dependent protein kinase and XRCC4-DNA ligase IV mobilization in the cell in response to DNA double strand breaks." J Biol Chem 280(8): 7060-9.
- Drouet, J., P. Frit, et al. (2006). "Interplay between Ku, Artemis, and the DNA-dependent protein kinase catalytic subunit at DNA ends." J Biol Chem 281(38): 27784-93.
- Elson, A., Y. Wang, et al. (1996). "Pleiotropic defects in ataxia-telangiectasia protein-deficient mice." Proc Natl Acad Sci U S A 93(23): 13084-9.
- Fowlkes, B. J. and D. M. Pardoll (1989). "Molecular and cellular events of T cell development." Adv Immunol 44: 207-64.

- Frank, K. M., J. M. Sekiguchi, et al. (1998). "Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV." Nature 396(6707): 173-7.
- Fugmann, S. D., A. I. Lee, et al. (2000). "The RAG proteins and V(D)J recombination: complexes, ends, and transposition." Annu Rev Immunol 18: 495-527.
- Fuschiotti, P., N. Pasqual, et al. (2007). "Analysis of the TCR alpha-chain rearrangement profile in human T lymphocytes." Mol Immunol 44(13): 3380-8.
- Gahery-Segard, H., E. Jouvin-Marche, et al. (1996). "Germline genomic structure of the B10.A mouse Tcra-V2 gene subfamily." Immunogenetics 44(4): 298-305.
- Gallagher, M., P. Obeid, et al. (2001). "Both TCR alpha and TCR delta chain diversity are regulated during thymic ontogeny." J Immunol 167(3): 1447-53.
- Gao, Y., Y. Sun, et al. (1998). "A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis." Cell 95(7): 891-902.
- Garcia, K. C., M. Degano, et al. (1996). "An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex." Science 274(5285): 209-19.
- Gill, L. L., D. Zaninetta, et al. (1991). "A transcriptional enhancer of the mouse T cell receptor delta gene locus." Eur J Immunol 21(3): 807-10.
- Girardi, M. (2006). "Immunosurveillance and immunoregulation by gammadelta T cells." J Invest Dermatol 126(1): 25-31.
- Glusman, G., L. Rowen, et al. (2001). "Comparative genomics of the human and mouse T cell receptor loci." Immunity 15(3): 337-49.
- Godfrey, D. I., J. Kennedy, et al. (1993). "A developmental pathway involving four phenotypically and functionally distinct subsets of CD3-CD4-CD8- triple-negative adult mouse thymocytes defined by CD44 and CD25 expression." J Immunol 150(10): 4244-52.
- Gorisch, S. M., M. Wachsmuth, et al. (2005). "Histone acetylation increases chromatin accessibility." J Cell Sci 118(Pt 24): 5825-34.
- Grewal, S. I. and D. Moazed (2003). "Heterochromatin and epigenetic control of gene expression." Science 301(5634): 798-802.
- Groettrup, M., K. Ungewiss, et al. (1993). "A novel disulfide-linked heterodimer on pre-T cells consists of the T cell receptor beta chain and a 33 kd glycoprotein." Cell 75(2): 283-94.
- Gu, Y., K. J. Seidl, et al. (1997). "Growth retardation and leaky SCID phenotype of Ku70-deficient mice." Immunity 7(5): 653-65.
- Guo, J., A. Hawwari, et al. (2002). "Regulation of the TCRalpha repertoire by the survival window of CD4(+)CD8(+) thymocytes." Nat Immunol 3(5): 469-76.
- Haas, W., P. Pereira, et al. (1993). "Gamma/delta cells." Annu Rev Immunol 11: 637-85.

- Harada, K. and T. Ikegami (2000). "Evolution of specificity in an immune network." J Theor Biol 203(4): 439-49.
- Hare, K. J., R. W. Wilkinson, et al. (1998). "Identification of a developmentally regulated phase of postselection expansion driven by thymic epithelium." J Immunol 160(8): 3666-72.
- Harrow, F. and B. D. Ortiz (2005). "The TCRalpha locus control region specifies thymic, but not peripheral, patterns of TCRalpha gene expression." J Immunol 175(10): 6659-67.
- Hawwari, A., C. Bock, et al. (2005). "Regulation of T cell receptor alpha gene assembly by a complex hierarchy of germline Jalpha promoters." Nat Immunol 6(5): 481-9.
- Hawwari, A. and M. S. Krangel (2007). "Role for rearranged variable gene segments in directing secondary T cell receptor {alpha} recombination." Proc Natl Acad Sci U S A 104(3): 903-7.
- Hayball, J. D., C. M. Jones, et al. (1996). "A T cell clone with three potential TCR alpha chain rearrangements expresses only one receptor combination at the cell surface." Mol Immunol 33(15): 1177-81.
- Hayday, A. C. (2000). "[gamma][delta] cells: a right time and a right place for a conserved third way of protection." Annu Rev Immunol 18: 975-1026.
- Hernandez-Munain, C., B. P. Sleckman, et al. (1999). "A developmental switch from TCR delta enhancer to TCR alpha enhancer function during thymocyte maturation." Immunity 10(6): 723-33.
- Ho, I. C., L. H. Yang, et al. (1989). "A T-cell-specific transcriptional enhancer element 3' of C alpha in the human T-cell receptor alpha locus." Proc Natl Acad Sci U S A 86(17): 6714-8.
- Hong, L., G. P. Schroth, et al. (1993). "Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA." J Biol Chem 268(1): 305-14.
- Huang, C. and O. Kanagawa (2001). "Ordered and coordinated rearrangement of the TCR alpha locus: role of secondary rearrangement in thymic selection." J Immunol 166(4): 2597-601.
- Inlay, M. and Y. Xu (2003). "Epigenetic regulation of antigen receptor rearrangement." Clin Immunol 109(1): 29-36.
- Ito, K., M. Bonneville, et al. (1989). "Different gamma delta T-cell receptors are expressed on thymocytes at different stages of development." Proc Natl Acad Sci U S A 86(2): 631-5.
- Jackson, A., H. D. Kondilis, et al. (2005). "Regulation of T cell receptor beta allelic exclusion at a level beyond accessibility." Nat Immunol 6(2): 189-97.
- Jackson, A. M. and M. S. Krangel (2006). "Turning T-cell receptor beta recombination on and off: more questions than answers." Immunol Rev 209: 129-41.

- Jeggo, P. and P. O'Neill (2002). "The Greek Goddess, Artemis, reveals the secrets of her cleavage." DNA Repair (Amst) 1(9): 771-7.
- Jeggo, P. A., G. E. Taccioli, et al. (1995). "Menage a trois: double strand break repair, V(D)J recombination and DNA-PK." Bioessays 17(11): 949-57.
- Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science 293(5532): 1074-80.
- Ji, Y., J. Zhang, et al. (2003). "A multistep mechanism for the activation of rearrangement in the immune system." Proc Natl Acad Sci U S A 100(13): 7557-62.
- Jones, P. L., G. J. Veenstra, et al. (1998). "Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription." Nat Genet 19(2): 187-91.
- Jouvin-Marche, E., C. Aude-Garcia, et al. (1998). "Differential chronology of TCRADV2 gene use by alpha and delta chains of the mouse TCR." Eur J Immunol 28(3): 818-27.
- Jouvin-Marche, E., I. Hue, et al. (1990). "Genomic organization of the mouse T cell receptor V alpha family." Embo J 9(7): 2141-50.
- Jouvin-Marche, E., M. G. Morgado, et al. (1989). "Complexity, polymorphism, and recombination of mouse T-cell receptor alpha gene families." Immunogenetics 30(2): 99-104.
- Kaufman, M., F. Andris, et al. (1999). "A logical analysis of T cell activation and anergy." Proc Natl Acad Sci U S A 96(7): 3894-9.
- Kaye, J., M. L. Hsu, et al. (1989). "Selective development of CD4+ T cells in transgenic mice expressing a class II MHC-restricted antigen receptor." Nature 341(6244): 746-9.
- Khor, B. and B. P. Sleckman (2002). "Allelic exclusion at the TCRbeta locus." Curr Opin Immunol 14(2): 230-4.
- Kouskoff, V., J. L. Vonesch, et al. (1995). "The influence of positive selection on RAG expression in thymocytes." Eur J Immunol 25(1): 54-8.
- Krangel, M. S. (2003). "Gene segment selection in V(D)J recombination: accessibility and beyond." Nat Immunol 4(7): 624-30.
- Kubota, T., J. Wang, et al. (1999). "Characterization of an avian (*Gallus gallus domesticus*) TCR alpha delta gene locus." J Immunol 163(7): 3858-66.
- Kuo, C. L., M. L. Chen, et al. (1998). "A conserved sequence block in murine and human T cell receptor (TCR) Jalpha region is a composite element that enhances TCR alpha enhancer activity and binds multiple nuclear factors." Proc Natl Acad Sci U S A 95(7): 3839-44.
- Lacorazza, H. D. and J. Nikolich-Zugich (2004). "Exclusion and inclusion of TCR alpha proteins during T cell development in TCR-transgenic and normal mice." J Immunol 173(9): 5591-600.

- Lauzurica, P. and M. S. Krangel (1994). "Temporal and lineage-specific control of T cell receptor alpha/delta gene rearrangement by T cell receptor alpha and delta enhancers." J Exp Med 179(6): 1913-21.
- Leiden, J. M. (1993). "Transcriptional regulation of T cell receptor genes." Annu Rev Immunol 11: 539-70.
- Letourneur, F. and B. Malissen (1989). "Derivation of a T cell hybridoma variant deprived of functional T cell receptor alpha and beta chain transcripts reveals a nonfunctional alpha-mRNA of BW5147 origin." Eur J Immunol 19(12): 2269-74.
- Litman, G. W., M. K. Anderson, et al. (1999). "Evolution of antigen binding receptors." Annu Rev Immunol 17: 109-47.
- Loh, E. Y., S. Cwirla, et al. (1988). "Human T-cell-receptor delta chain: genomic organization, diversity, and expression in populations of cells." Proc Natl Acad Sci U S A 85(24): 9714-8.
- Ma, Y., U. Pannicke, et al. (2002). "Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination." Cell 108(6): 781-94.
- Malissen, M., J. Trucy, et al. (1992). "Regulation of TCR alpha and beta gene allelic exclusion during T-cell development." Immunol Today 13(8): 315-22.
- Malissen, M., J. Trucy, et al. (1988). "A T cell clone expresses two T cell receptor alpha genes but uses one alpha beta heterodimer for allorecognition and self MHC-restricted antigen recognition." Cell 55(1): 49-59.
- Mancini, S., S. M. Candeias, et al. (1999). "TCR alpha-chain repertoire in pTalpha-deficient mice is diverse and developmentally regulated: implications for pre-TCR functions and TCRA gene rearrangement." J Immunol 163(11): 6053-9.
- Matei, I. R., R. A. Gladdy, et al. (2007). "ATM deficiency disrupts TcrA locus integrity and the maturation of CD4+CD8+ thymocytes." Blood 109(5): 1887-96.
- Mauvieux, L., I. Villey, et al. (2001). "T early alpha (TEA) regulates initial TCRVAJA rearrangements and leads to TCRJA coincidence." Eur J Immunol 31(7): 2080-6.
- Mauvieux, L., I. Villey, et al. (2003). "TEA regulates local TCR-Jalpha accessibility through histone acetylation." Eur J Immunol 33(8): 2216-22.
- McBlane, F. and J. Boyes (2000). "Stimulation of V(D)J recombination by histone acetylation." Curr Biol 10(8): 483-6.
- McMurry, M. T. and M. S. Krangel (2000). "A role for histone acetylation in the developmental regulation of VDJ recombination." Science 287(5452): 495-8.
- Miao, F. and R. Natarajan (2005). "Mapping global histone methylation patterns in the coding regions of human genes." Mol Cell Biol 25(11): 4650-61.

- Monroe, R. J., B. P. Sleckman, et al. (1999). "Developmental regulation of TCR delta locus accessibility and expression by the TCR delta enhancer." Immunity 10(5): 503-13.
- Moshous, D., I. Callebaut, et al. (2001). "Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency." Cell 105(2): 177-86.
- Moshous, D., C. Pannetier, et al. (2003). "Partial T and B lymphocyte immunodeficiency and predisposition to lymphoma in patients with hypomorphic mutations in Artemis." J Clin Invest 111(3): 381-7.
- Mostoslavsky, R., F. W. Alt, et al. (2004). "The lingering enigma of the allelic exclusion mechanism." Cell 118(5): 539-44.
- Naumov, Y. N., E. N. Naumova, et al. (2003). "A fractal clonotype distribution in the CD8+ memory T cell repertoire could optimize potential for immune responses." J Immunol 170(8): 3994-4001.
- Nicolas, N., D. Moshous, et al. (1998). "A human severe combined immunodeficiency (SCID) condition with increased sensitivity to ionizing radiations and impaired V(D)J rearrangements defines a new DNA recombination/repair deficiency." J Exp Med 188(4): 627-34.
- Niewolik, D., U. Pannicke, et al. (2006). "DNA-PKcs dependence of Artemis endonucleolytic activity, differences between hairpins and 5' or 3' overhangs." J Biol Chem 281(45): 33900-9.
- O'Driscoll, M., K. M. Cerosaletti, et al. (2001). "DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency." Mol Cell 8(6): 1175-85.
- Ortiz, B. D., D. Cado, et al. (1999). "A new element within the T-cell receptor alpha locus required for tissue-specific locus control region activity." Mol Cell Biol 19(3): 1901-9.
- Ortiz, B. D., F. Harrow, et al. (2001). "Function and factor interactions of a locus control region element in the mouse T cell receptor-alpha/Dad1 gene locus." J Immunol 167(7): 3836-45.
- Padovan, E., G. Casorati, et al. (1993). "Expression of two T cell receptor alpha chains: dual receptor T cells." Science 262(5132): 422-4.
- Pannicke, U., Y. Ma, et al. (2004). "Functional and biochemical dissection of the structure-specific nuclease ARTEMIS." Embo J 23(9): 1987-97.
- Pasqual, N., M. Gallagher, et al. (2002). "Quantitative and qualitative changes in V-J alpha rearrangements during mouse thymocytes differentiation: implication for a limited T cell receptor alpha chain repertoire." J Exp Med 196(9): 1163-73.
- Pearse, M., L. Wu, et al. (1989). "A murine early thymocyte developmental sequence is marked by transient expression of the interleukin 2 receptor." Proc Natl Acad Sci U S A 86(5): 1614-8.

- Perkins, E. J., A. Nair, et al. (2002). "Sensing of intermediates in V(D)J recombination by ATM." Genes Dev 16(2): 159-64.
- Petiniot, L. K., Z. Weaver, et al. (2000). "Recombinase-activating gene (RAG) 2-mediated V(D)J recombination is not essential for tumorigenesis in Atm-deficient mice." Proc Natl Acad Sci U S A 97(12): 6664-9.
- Petrie, H. T., F. Livak, et al. (1993). "Multiple rearrangements in T cell receptor alpha chain genes maximize the production of useful thymocytes." J Exp Med 178(2): 615-22.
- Ramsden, D. A., K. Baetz, et al. (1994). "Conservation of sequence in recombination signal sequence spacers." Nucleic Acids Res 22(10): 1785-96.
- Raval, P., A. N. Kriatchko, et al. (2008). "Evidence for Ku70/Ku80 association with full-length RAG1." Nucleic Acids Res 36(6): 2060-72.
- Redondo, J. M., S. Hata, et al. (1990). "A T cell-specific transcriptional enhancer within the human T cell receptor delta locus." Science 247(4947): 1225-9.
- Riballo, E., S. E. Critchlow, et al. (1999). "Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient." Curr Biol 9(13): 699-702.
- Roberts, J. L., P. Lauzurica, et al. (1997). "Developmental regulation of VDJ recombination by the core fragment of the T cell receptor alpha enhancer." J Exp Med 185(1): 131-40.
- Rytkonen, M., J. L. Hurwitz, et al. (1994). "Evidence for recombinatorial hot spots at the T cell receptor J alpha locus." Eur J Immunol 24(1): 107-15.
- Rytkonen, M. A., J. L. Hurwitz, et al. (1996). "Restricted onset of T cell receptor alpha gene rearrangement in fetal and neonatal thymocytes." Eur J Immunol 26(8): 1892-6.
- Schlimgen, R. J., K. L. Reddy, et al. (2008). "Initiation of allelic exclusion by stochastic interaction of Tcrb alleles with repressive nuclear compartments." Nat Immunol 9(7): 802-9.
- Sepulveda, N., L. Boucontet, et al. (2005). "Stochastic modeling of T cell receptor gamma gene rearrangement." J Theor Biol 234(2): 153-65.
- Shinkai, Y., G. Rathbun, et al. (1992). "RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement." Cell 68(5): 855-67.
- Sleckman, B. P., C. G. Bardon, et al. (1997). "Function of the TCR alpha enhancer in alphabeta and gammadelta T cells." Immunity 7(4): 505-15.
- Sleckman, B. P., J. Carabana, et al. (2001). "Assessing a role for enhancer-blocking activity in gene regulation within the murine T-cell receptor alpha/delta locus." Immunology 104(1): 11-8.
- Sleckman, B. P., J. R. Gorman, et al. (1996). "Accessibility control of antigen-receptor variable-region gene assembly: role of cis-acting elements." Annu Rev Immunol 14: 459-81.

- Staudt, L. M., R. G. Clerc, et al. (1988). "Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif." Science 241(4865): 577-80.
- Swanson, P. C. (2002). "A RAG-1/RAG-2 tetramer supports 12/23-regulated synapsis, cleavage, and transposition of V(D)J recombination signals." Mol Cell Biol 22(22): 7790-801.
- Thompson, S. D., J. Pelkonen, et al. (1990). "First T cell receptor alpha gene rearrangements during T cell ontogeny skew to the 5' region of the J alpha locus." J Immunol 145(7): 2347-52.
- Thompson, S. D., J. Pelkonen, et al. (1990). "Nonrandom rearrangement of T cell receptor J alpha genes in bone marrow T cell differentiation cultures." J Immunol 144(7): 2829-34.
- Tonegawa, S. (1983). "Somatic generation of antibody diversity." Nature 302(5909): 575-81.
- Tourigny, M. R., S. Mazel, et al. (1997). "T cell receptor (TCR)-beta gene recombination: dissociation from cell cycle regulation and developmental progression during T cell ontogeny." J Exp Med 185(9): 1549-56.
- Turner, B. M. (2002). "Cellular memory and the histone code." Cell 111(3): 285-91.
- Vacchio, M. S., A. Oлару, et al. (2007). "ATM deficiency impairs thymocyte maturation because of defective resolution of T cell receptor alpha locus coding end breaks." Proc Natl Acad Sci U S A 104(15): 6323-8.
- Villey, I., D. Caillol, et al. (1996). "Defect in rearrangement of the most 5' TCR-J alpha following targeted deletion of T early alpha (TEA): implications for TCR alpha locus accessibility." Immunity 5(4): 331-42.
- Wang, F., C. Y. Huang, et al. (1998). "Rapid deletion of rearranged T cell antigen receptor (TCR) Valpha-Jalpha segment by secondary rearrangement in the thymus: role of continuous rearrangement of TCR alpha chain gene and positive selection in the T cell repertoire formation." Proc Natl Acad Sci U S A 95(20): 11834-9.
- Wang, G., G. R. Krueger, et al. (2003). "A simplified and comprehensive computational model to study the behavior of T cell populations in the thymus during normal maturation and in infection with mouse moloney leukemia virus." In Vivo 17(3): 225-8.
- Wang, G., G. R. Krueger, et al. (2004). "A continuous model studying T cell differentiation and lymphomagenesis and its distinction with discrete models." Anticancer Res 24(3a): 1813-6.
- Warmflash, A. and A. R. Dinner (2006). "A model for TCR gene segment use." J Immunol 177(6): 3857-64.
- Wilson, A., J. P. de Villartay, et al. (1996). "T cell receptor delta gene rearrangement and T early alpha (TEA) expression in immature alpha beta lineage thymocytes: implications for alpha beta/gamma delta lineage commitment." Immunity 4(1): 37-45.

- Wilson, A., C. Marechal, et al. (2001). "Biased V beta usage in immature thymocytes is independent of DJ beta proximity and pT alpha pairing." J Immunol 166(1): 51-7.
- Winoto, A. and D. Baltimore (1989). "Developmental regulation of the TCR alpha delta locus." Cold Spring Harb Symp Quant Biol 54 Pt 1: 87-92.
- Winoto, A. and D. Baltimore (1989). "A novel, inducible and T cell-specific enhancer located at the 3' end of the T cell receptor alpha locus." Embo J 8(3): 729-33.
- Wu, L., P. W. Kincade, et al. (1993). "The CD44 expressed on the earliest intrathymic precursor population functions as a thymus homing molecule but does not bind to hyaluronate." Immunol Lett 38(1): 69-75.
- Yaneva, M., T. Kowalewski, et al. (1997). "Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies." Embo J 16(16): 5098-112.
- Yoshida, T., A. Tsuboi, et al. (2000). "The DNA-bending protein, HMG1, is required for correct cleavage of 23 bp recombination signal sequences by recombination activating gene proteins in vitro." Int Immunol 12(5): 721-9.
- Zhong, X. P. and M. S. Krangel (1997). "An enhancer-blocking element between alpha and delta gene segments within the human T cell receptor alpha/delta locus." Proc Natl Acad Sci U S A 94(10): 5219-24.

Résumé :

Les lymphocytes T expriment à leur surface des récepteurs (TR) composés de chaînes $\alpha\beta$ ou $\gamma\delta$, chargés de reconnaître des peptides antigéniques. Une chaîne d'un TR α donnée est le résultat de l'assemblage de gènes variable (V), jonction (J) et constant (C) sous le contrôle d'un mécanisme de recombinaison spécifique dénommé « recombinaison V(D)J ». De part le nombre important de gènes V et J dispersés sur plus de 1 méga base et des segments V dupliqués chez la souris, il est quasiment impossible d'établir la diversité combinatoire VJ exacte et surtout leurs fréquences par les techniques actuelles de biologie moléculaire. Aussi pour appréhender ces questions, nous avons développé un modèle de simulation des réarrangements des combinaisons VJ codant pour la chaîne TR α chez la souris. Le modèle implémenté est basé sur des fenêtres d'accessibilités flexibles, des vitesses d'ouverture variables et inclue aussi un temps de maturation entre deux réarrangements. Il permet de rendre compte de la dynamique et de la variabilité de la chaîne α en termes de répertoire (nombre et fréquence des combinaisons). Il propose que 2 à 3 réarrangements sont suffisant pour utiliser l'ensemble des segments J et permet aussi de donner un profil global de l'ensemble des combinaisons VJ. En parallèle, une analyse des réarrangements VJ a été réalisée chez l'homme. Cette analyse permet de visualiser les profils des combinaisons VJ expérimentaux et permettra de valider l'adaptation du modèle de simulation des réarrangements des chaînes TR α chez la souris aux chaînes TR α chez l'homme. Le modèle pourra servir d'outils pour analyser les variations entre répertoire sain et répertoire altéré ou modifié dans le cas de pathologies ou de reconstruction du répertoire immunitaire après une greffe de moelle osseuse.

Mots clés : Diversité combinatoire des TR, locus TRA/TRD, réarrangement V(D)J, simulation, modèle murin, modèle humain

T lymphocytes express at their surface an antigen receptor composed by $\alpha\beta$ or $\gamma\delta$ chains. A TR α chains are encoded by a variable (V), a joining (J) and a constant (C) segments which are under the of the control of a specific recombination mechanism called "V(D)J recombination". The VJ combinatorial diversity is unknown and the current state of molecular technology does not allow us to perform an analysis of all putative VJ combination or to estimate the frequencies of the functional VJ in mice. To overcome this difficulty we defined a mathematical model fitting experimental data. This model gives new insights on the rules controlling the use of the V and the J genes and provides a dynamic calculation of the VJ combinations. The model proposes an accessibility of the TRA/TRD locus by successive windows of different sizes and with different speed of progression. Furthermore, a possibility of successive secondary rearrangements was introduced. In parallel, an experimental analysis of the VJ combination has been performed in humans. From this analysis, the VJ combination profiles are calculated and used to validate our simulation program. In the future, the model may be use to analysis the variations between sound or altered repertoire.

Key words: TR-combinatorial diversity, TRA/TRD locus, V(D)J rearrangements, simulation, murine model, human model.