UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

Biotechnologie et signalisation cellulaire, UMR 7242

THÈSE

présentée par :

Aurore Villemin

soutenue le : 20 mai 2014

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline: Aspects moléculaires et cellulaires de la biologie

Generation of transgenic vectors encoding human immunoglobulins, functionality assays and transgenesis in mice

Thèse dirigée par :

Pr. Etienne Weiss, Ecole Supérieure de Biotechnologie, Strasbourg, France

Rapporteurs externes :

Pr. Jean-Louis Boulay, Universität Basel, Switzerland

Pr. Michael Reth, Unversität Freiburg, Germany

Examinateur interne :

Pr. Sylvie Fournel, Université de Strasbourg, France

Examinateurs :

Dr. Elisabetta Traggiai, Novartis Pharma AG, Basel, Switzerland

Dr. André Traunecker, Novartis Pharma AG, Basel, Switzerland

Thèse réalisée pour la société Novartis Pharma AG, au sein du département « Protein Production and Antibodies »

Acknowledgment

This work has been carried out in the last four years at Novartis Pharma AG. I wish to thank my supervisor at Novartis, Dr. André Traunecker, who gave me the opportunity to conduct my PhD thesis in his lab; and I also thank him for allowing me to use his nice pictures that illustrate a large part of this thesis. I also wish to thank Dr. Elisabetta Traggiai and Gilles Sansig for their permanent support, guidance and encouragements.

I would like to express my gratitude to my thesis director, Pr. Etienne Weiss, for its scientific advices and support.

Also, I would like to thank Pr. Sylvie Fournel, Pr. Jean-Louis Boulay and Pr. Michael Reth for having accepted to be member of the jury.

Special thanks go to Emeline Thevenon for her advices on flow cytometry. Thanks also to Martine Marchant and Matthias Wrobel for their assistance on cell sorting, and to Régis Cébe for his advices on protein analyzes.

Furthermore, I would like to thank Jean-François Spetz and Dr. Bernd Kinzel for having undertaken transgenesis in mice.

My gratitude to Thomas Pietzonka from Novartis Institute of Biomedical Research in Basel for welcoming me at Novartis and giving me the wonderful opportunity of working in this dynamic environment.

I thank also all my colleagues and friends at Novartis not mentioned yet for supporting and encouraging me throughout these years and for contributing to the friendly working atmosphere.

Je remercie chaleureusement mes parents, ma sœur, Charline, Julia, Lucie et tous mes amis qui ont toujours été là pour moi.

Résumé en français

Génération de vecteurs codant pour les anticorps humains, analyse de l'expression et transgénèse chez la souris

1. Introduction

Les anticorps : molécules essentielles du système immunitaire

Les anticorps, également appelés immunoglobulines (Ig), sont des glycoprotéines complexes qui jouent un rôle majeur dans le système immunitaire. Ils sont capables de reconnaître et d'éliminer spécifiquement des substances étrangères à l'organisme. Un individu sain présente plusieurs milliards d'anticorps différents pouvant reconnaître spécifiquement autant d'antigènes. Les anticorps sont générés et secrétés par les lymphocytes B. Chaque lymphocyte B produit un seul type d'anticorps de spécificité identique à celle de son BCR (B-cell receptor).

D'un point de vue structural, les anticorps sont composés de deux chaînes lourdes (HC) identiques, auxquelles sont associées deux chaînes légères (LC) identiques aussi. Il existe cinq types de chaînes lourdes chez les mammifères : α, δ, ε, ɣ et μ. Le type de la chaîne lourde définit la classe (ou l'isotype) de l'anticorps, les cinq types précédemment cités sont retrouvés dans IgA, IgD, IgE, IgG et IgM respectivement. Les classes IgA et IgG peuvent être sous-divisées en sous-classes (IgA1, IgA2 et IgG1, IgG2, IgG3, IgG4). Chaque chaîne lourde contient une partie variable (V) et une partie constante (C) ; la partie V est impliquée dans la reconnaissance spécifique d'un antigène alors que la partie C confère des fonctions effectrices aux anticorps. La région variable diffère entre les anticorps produits par différentes cellules B. La région constante est identique entre tous les anticorps du même isotype. De la même façon, les chaînes légères peuvent être de type κ ou λ, et sont composées d'une partie variable et d'une partie constante. La **Figure 1** schématise la structure d'une IgG.

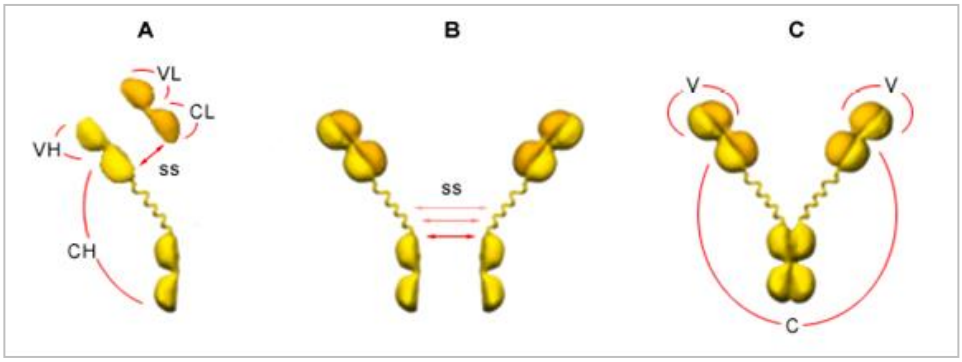

Figure 1: Structure d'une IgG.
A; la chaine lourde (HC) et la chaine légère (LC) sont chacune composées d'une partie variable (respectivement VH et VL) et d'une partie constante (respectivement CH et CL). Les deux chaines sont associées de façon covalente par des ponts disulfures (SS).
B; les deux hétéro-dimères, composés d'une HC et d'une LC, sont associés de façon covalente par des ponts disulfures (SS), et forment ainsi un homo-dimère.
C; un anticorps de type IgG contient deux parties variables (V), composées par VH et VL. Ces parties variables reconnaissent spécifiquement leur antigène apparenté, alors que la partie constant (C) définit les fonctions effectrices de l'anticorps.

Les régions V sont codées par des gènes qui comportent trois types de segments dans le cas de la chaîne lourde : *V* (variable), *D* (diversité) et *J* (jonction), et deux types de segments dans le cas des chaînes légères : *V* et *J*. Ces segments doivent être combinés entre eux par un réarrangement somatique pour former un gène fonctionnel [1]. Le développement des cellules B dans la moelle osseuse correspond à l'acquisition de leur BCR associé au dimère Igα-Igβ. Le BCR est exprimé suite aux réarrangements somatiques V(D)J de la chaine lourde et de la chaine légère.

Les segments de la chaîne lourde réarrangent en premier, générant ainsi un exon *VDJ*. Lors de la transcription du gène, l'ARN messager contient la région *VDJ* recombinée, ainsi que les segments constants de type mu (Cμ). Si le réarrangement de la chaîne lourde est productif, le précurseur de cellule B sera alors capable de l'exprimer et de l'exporter à la surface, en association avec une pseudo-chaîne légère, formant ainsi le pré-BCR ; ce dernier sera alors testé pour ses capacités de signalisation. Après avoir réussi ce premier contrôle, le précurseur de cellule B passe à l'étape suivante, consistant à générer une chaîne légère. La première étape concerne la recombinaison d'un segment *V* et d'un segment *J* pour former un complexe *VJ* auquel est ajouté par la suite un segment constant (Cκ ou Cλ). Si le réarrangement de la chaîne légère a été productif, la cellule exprimera en surface l'immunoglobuline IgM. Avant de quitter la moelle osseuse

pour aller patrouiller dans le corps, les cellules B doivent subir un contrôle qui vise à éliminer les cellules B productrices d'anticorps auto-réactifs. Les cellules qui quittent la moelle osseuse et rejoignent les organes lymphoïdes secondaires sont dites matures. Ces cellules seront amenées à maturer davantage, entre autre en modifiant l'isotype des anticorps qu'elles produisent. La commutation de classe est un processus qui permet de changer l'isotype (classe) des immunoglobulines produites.

Les *loci* de la chaîne lourde et des chaînes légères contiennent plusieurs dizaines de segments différents pouvant recombiner aléatoirement entre eux, ce qui représente une grande source de diversité. A titre d'exemple, le *locus* humain de la chaîne lourde est représenté dans la **Figure 2**.

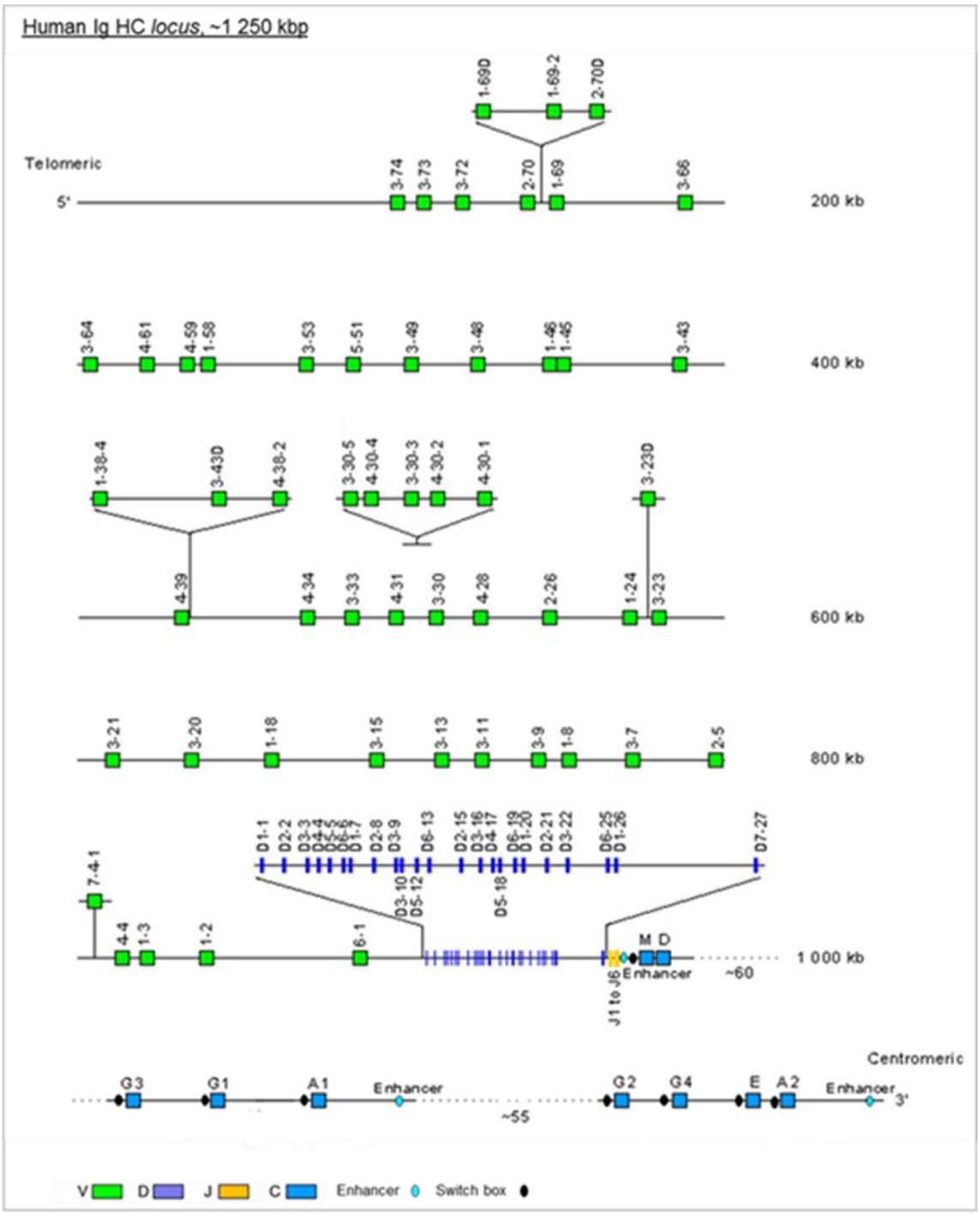

Figure 2 : Le *locus* humain de la chaine lourde s'étend sur 1 250 kilo paires de bases (kbp). Les parties constantes annotées M, D, G3, G1, A1, G2, G4, E et A2 représentent respectivement Cμ, Cδ, Cγ3, Cɣ1, Cα1, Cɣ2, Cɣ4, Cε et Cα2. (Image: Lefranc, M.-P.IMGT®).

Les anticorps thérapeutiques

Depuis la découverte des anticorps monoclonaux en 1975, par George Koehler et César Milstein, leur utilisation en thérapie a été démontrée bénéfique pour de multiples pathologies. Bien que le premier anticorps approuvé en thérapie soit issu de la souris (muromonab-CD3), les anticorps murins ne sont généralement pas bien tolérés par l'homme. Ils sont reconnus comme corps étrangers et donc le système immunitaire engage une réaction afin de les éliminer, ce qui les rend rapidement inopérants. Les anticorps non humains sont fortement immunogènes, ils risquent aussi de provoquer chez le patient un choc anaphylactique aux conséquences irréversibles.

Par conséquent, il est indispensable de générer des anticorps humains ou bien humanisés. De ce fait, des technologies sophistiquées, issues du génie génétique et de la microbiologie ont été développées pour produire des anticorps humanisés. Une de ces méthodes consiste à greffer les parties spécifiques de l'anticorps de souris, impliquées dans la reconnaissance de l'antigène, sur la partie effectrice d'un anticorps humain. Cette opération réduit l'immunogénicité de l'anticorps de souris initial. Une autre méthode repose sur la création de bibliothèques de gènes d'anticorps à partir de cellules B humaines (sélection par phage display ou yeast display). La diversité de telles bibliothèques peut être très importante. En revanche, étant donné que la source de matériel génétique vient de cellules B matures qui ont été sélectionnées pour ne pas réagir contre soi, il y a peu de chance de trouver des gènes d'anticorps destinés à réagir contre des protéines humaines. Ceci représente un facteur limitant non négligeable puisque que la plupart des anticorps thérapeutiques sont dirigés contre des antigènes humains, tels qu'une interleukine, ou un récepteur, ou encore une hormone. Une autre approche permet d'exprimer des anticorps entièrement humains dans une autre espèce, telle que la souris. Cette dernière approche peut palier aux problèmes cités ci-dessus : dans une souris, les molécules de souris sont considérées comme « soi » et les molécules d'autres espèces comme « non-soi » ce qui est également le cas de protéines humaines. La génération d'une telle souris implique plusieurs manipulations génétiques. La production d'anticorps endogènes doit être abrogée par « knockout » génétique. Les gènes d'anticorps humains doivent ensuite être intégrés au génome de la souris par transgénèse [2]. La machinerie de la souris pourra alors être exploitée pour la génération,

la maturation et la sécrétion d'anticorps humains. Ces souris sont alors dites « humanisées » quant à la production d'anticorps.

Parmi toutes les technologies proposées, l'utilisation de souris transgéniques semble être l'approche la plus prometteuse.

2. Problématique

Mon projet de thèse a été de contribuer à la génération de souris transgéniques capables de produire un large répertoire d'anticorps humains dans un but d'applications thérapeutiques. Dans un premier temps, j'ai participé au clonage de l'ensemble des gènes codant pour la chaîne lourde et les chaines légères humaines sous la forme de chromosomes artificiels chez la levure *Saccharomyces cerevisae* (Yeast Artifical Chromosome, YAC). Puis, pour simplifier la manipulation des gènes codant pour la chaîne lourde, j'ai réalisé des vecteurs plasmidiques qui résument l'information génétique contenue dans le *locus* humain de la chaîne lourde. Ces constructions ont ensuite été testées par transfection dans la lignée lymphocytaire 300-19 pour vérifier qu'elles donnent bien lieu à l'expression de chaînes lourdes complètes. Enfin et en parallèle de ce travail, j'ai participé à un programme de réalisation de souris transgéniques générées à partir de ces vecteurs validés.

3. Assemblage des *loci* Ig humains sous forme de YACs

Cette approche vise à cloner dans leur quasi intégralité les trois *loci* d'anticorps humains. Chez l'Homme, les *loci* de la chaîne lourde et des chaînes légères κ et λ se situent sur les chromosomes 14, 2 et 22, et s'étendent sur 1 250 kbp, 1 820 kbp et 1 050 kbp respectivement. Nous avons utilisé la méthode TAR (Transformation Associated Recombination), une méthode de clonage sophistiquée qui a trait à la recombinaison homologue dans la levure [3]. Cette méthode est particulièrement appropriée au clonage de larges fragments d'ADN génomique sous la forme de YAC. L'ensemble des travaux réalisés a permis de cloner la majorité des gènes constituant les

loci humains HC, LC κ et LC λ. Ainsi, le YAC HC obtenu, dont la taille est de 650 kbp, contient 27 segments V (sur 35), 11 D (sur 27), 6 J (sur 6) ainsi que les gènes codant pour les parties constantes de type μ, δ, ɣ3 et ɣ1. Le YAC LC Lambda, dont la taille est de 350 kbp, contient 28 segments V (sur 33) et les 7 J-Cλ. Enfin, le YAC Kappa, dont la taille est de 390 kbp, contient 16 segments V (sur 21), les 5 segments J et la région constante Cκ. Ces trois YACs sont donc parfaitement représentatifs de l'information génétique des anticorps humains. Cependant ces vecteurs de très grande taille présentent plusieurs limites ; en effet, il est difficile de contrôler leur intégrité et leur stabilité, d'autre part, ils sont fragiles et donc délicats à manipuler pour la transgénèse. De ce fait, les travaux de thèse se sont orientés vers la génération de constructions plasmidiques codant pour la chaine lourde des anticorps humains.

4. Constructions codant pour la chaine lourde des anticorps humains

Pour construire des plasmides dont l'organisation et le contenu génétique sont à l'image du *locus* natif de la chaine lourde, nous avons sélectionné des éléments génétiques qui sont essentiels pour l'expression et la diversité de la chaîne lourde des anticorps humains. Il s'agit : 1) des segments V, D et J (région variable) les plus représentés dans les anticorps circulants, 2) des parties qui codent pour différentes régions constantes, et 3) des séquences conservées d'ADN non codant impliquées dans la régulation des différents processus conduisant à l'expression de la chaîne lourde.

Une analyse bio-informatique très minutieuse des séquences d'ADN dans des banques de données a permis de sélectionner un ensemble d'éléments qui permettent de constituer un répertoire restreint mais diversifié. Nous avons sélectionné un total de 13 segments V : six V de la famille 3 (V3), trois V1, deux V2, un V4 et un V6. Les V3 sont les plus représentés dans le génome humain, c'est pourquoi ils représentent quasiment la moitié des segments V sélectionnés. La partie couvrant la région des segments D et J a été entièrement synthétisée. Les segments D sont classés sous 7 familles, les similitudes au sein d'une même famille ont permis de déterminer des séquences consensus. Les segments J ont tous été conservés, sauf J5 qui est une duplication de J4. Concernant la sélection des gènes codant pour la partie constante de la chaîne lourde, nous nous

sommes basés sur le fait que le type μ était celui impliqué dans la différentiation des cellules B, et que le type ɣ était l'isotype le plus représenté dans le sérum humain.

La construction HC Minilocus (78 kbp) est donc composée des 13 segments V sélectionnés, de la région D-J synthétisée et des gènes codant pour les parties constantes de type μ, ɣ3 et ɣ1. Trois autres constructions ont été dérivées des étapes intermédiaires de clonage. Elles sont basées sur 7 segments V et la région D-J synthétisée. Le HC Microlocus Classic (22 kb) a été obtenu après clonage du gène codant pour la partie constante ɣ1 en 3' des éléments V, D et J précédemment cités ; cette construction code pour des chaînes lourdes d'IgG1 (ɣ1 HC). De la même façon, le HC Microlocus Light (21.5 kb) contient le gène codant pour ɣ1 $CH1^-$, cette construction code donc pour des chaînes lourdes IgG sans domaine CH1 (ɣ1 HC $CH1^-$). Les chaînes lourdes de ce type sont exprimées sans être associées à des chaînes légères, à l'image de ce qui est observé chez les camélidés (Heavy Chain only antibodies) [4]. Enfin le HC Microlocus Light shRNA (22 kb) est une construction basée sur le Microlocus Light à laquelle a été ajoutée une séquence codant pour quatre shRNA (small hairpin RNA) visant à réprimer l'expression d'IgM murin. Ce *locus* est destiné à être injecté dans des souris natives (wild type, WT) dans la mesure où l'injection de fragments d'ADN dans les ovocytes de souris HC KO peut s'avérer difficile. De plus, le fait de réprimer l'expression endogène tout en exprimant le transgène en une seule et unique étape de transgénèse permettrait un gain de temps. Comme le HC Microlocus Light, cette construction vise à produire des ɣ1 HC $CH1^-$.

Pour ces quatre constructions, nous avons pris garde de conserver les séquences non codantes nécessaires au processus de réarrangement VDJ, ainsi que les séquences promotrice situées en 5' de chaque segment V et les deux principaux activateurs (intronique et 3'). La **Figure 3** résume la génération de ces quatre constructions codant pour la chaîne lourde des anticorps humains.

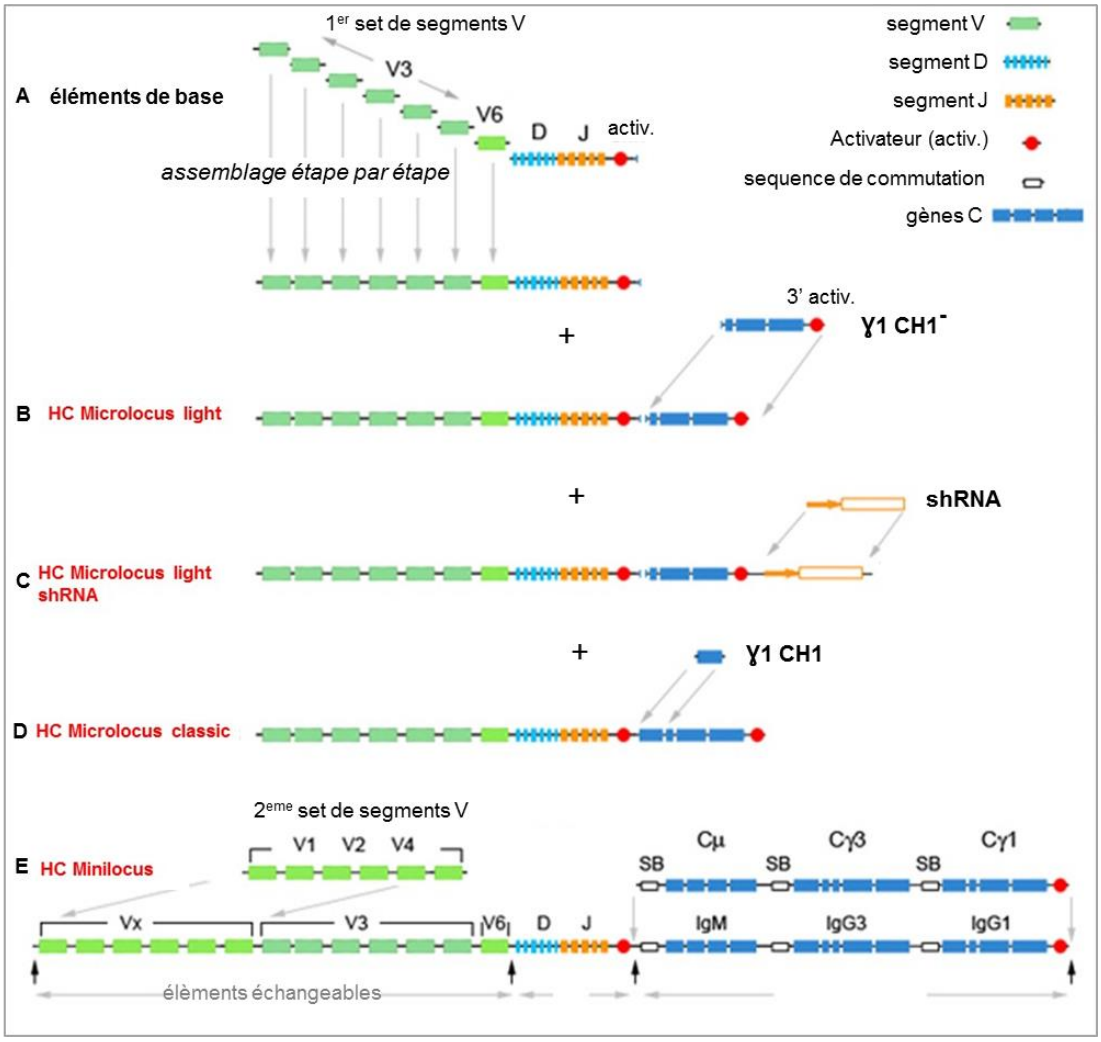

Figure 3: Résumé des constructions codant pour la chaîne lourde des anticorps humains.
Le panneau A représente les éléments de base : six V3, un V6, la région D-J synthétisée et l'activateur intronique. Ceci constitue la base des quatre constructions (*B*, *C*, *D* et *E*). En effet, différents éléments y ont été ajoutés par étape:
- l'ajout du gène Cɣ1 CH1⁻ et de l'élément régulateur 3' activateur (3'-activ.) a conduit à la création du HC Microlocus Light (*B*);
- l'ajout au HC Microlocus Light (*B*) de la séquence codant pour les shRNA a conduit à la construction HC Microlocus Light shRNA (*C*);
- l'ajout au HC Microlocus Light (*B*) de l'exon Cɣ1 CH1 a mené à la création du HC Microlocus Classic (*D*);
- l'ajout du deuxième set de segments V et des gènes codant pour les parties constantes Cμ, Cγ3 et Cγ1 aux éléments de base (*A*) a conduit au HC Minilocus (*E*).

Pour la plupart des constructions, les éléments ont été assemblés en utilisant les méthodes classiques issues de la biologie moléculaire. En utilisant *E. coli*, de grandes quantités de ces vecteurs de transgénèse ont ainsi pu être produites. Ces *loci* HC réduits ont également été pensés pour être facilement modulables et, du fait de leur taille

réduite, permettent de faciliter la transgénèse ou d'optimiser l'expression des gènes d'anticorps humains dans la souris. A l'avenir, on pourrait modifier ces *loci* réduits en échangeant l'un ou l'autre des différents segments ; ainsi, des anticorps recomposés inexistants dans la nature pourraient être générés. Chacune des constructions contient le gène de résistance à la néomycine (Neo^R) dont l'expression est régulée par le promoteur eucaryote de la thymidine kinase du virus de l'herpès (HSVtk). Ainsi ces constructions peuvent être transfectées dans une lignée cellulaire eucaryote et ces dernières peuvent être sélectionnées par leur résistance à l'antibiotique.

5. Analyse *in vitro* de la fonctionnalité des constructions générées

Générer, puis établir des lignées de souris transgéniques afin de produire des anticorps humains est une stratégie au coût important et qui s'inscrit dans la durée. Dans un souci éthique de minimiser le nombre de souris utilisées, mais également dans le but de répondre rapidement à la question essentielle de la fonctionnalité de nos *loci* humains reconstitués, des lymphocytes pro-B 300-19 de souris ont été transformés. De plus, hors du contexte de la souris transgénique, l'utilisation d'un système eucaryote, générant une grande diversité d'anticorps humains à partir de nos *loci*, pourrait avoir des applications pharmacologiques. Le choix d'utiliser la lignée cellulaire 300-19 a été dicté par le fait que ces lymphocytes B, transformés par le virus Abelson de la leucémie de la souris (Abelson murine leukemia virus) sont bloqués au stade de différentiation pro B [5]. Cependant, il a été montré que ces cellules peuvent progresser du stade pro B au stade de cellule B mature, lorsqu'elles sont maintenues en culture [6]. Ainsi cette lignée cellulaire est l'outil idéal pour tester toutes les étapes de la génération des anticorps, incluant le réarrangement, la transcription, la maturation de l'ARNm, l'expression et la sécrétion des chaînes polypeptidiques.

Avant de procéder à la transfection des cellules, nous avons ajouté le gène codant pour la protéine fluorescente mCherry dans la construction HC Minilocus, et ce dans le but de visualiser l'efficacité de transfection (**Figure 4**). La transcription de ce gène est régulée par le promoteur eucaryote CMV.

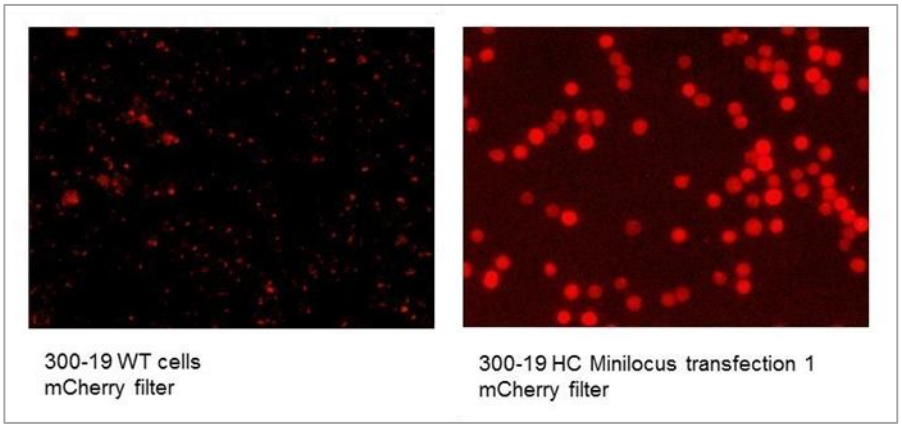

Figure 4: Les cellules 300-19 transfectées avec le HC Minilocus expriment la protéine fluorescente mCherry. Les cellules transfectées (*panneau droit*) sont comparées aux cellules 300-19 WT (*panneau gauche*). Après une excitation à 590 nm, les cellules transfectées émettent une lumière rouge (610 nm). Ces images ont été prises deux semaines après transfection.

Les trois constructions HC Microlocus Classic, HC Microlocus Light et HC Microlocus light shRNA ont été caractérisées au niveau des réarrangements (DJ et VDJ), mais également au niveau de l'expression de chaînes lourdes d'anticorps. Pour chacune des trois constructions intermédiaires, plusieurs réarrangements DJ et VDJ ont été identifiés, montrant une grande diversité dans l'utilisation des segments V, D et J. Par cytométrie en flux (FACS), des chaînes lourdes humaines ont été identifiées dans le cytoplasme et à la surface des cellules. Les cellules exprimant des chaînes lourdes en surface ont été enrichies par FACS (**Figure 5**).

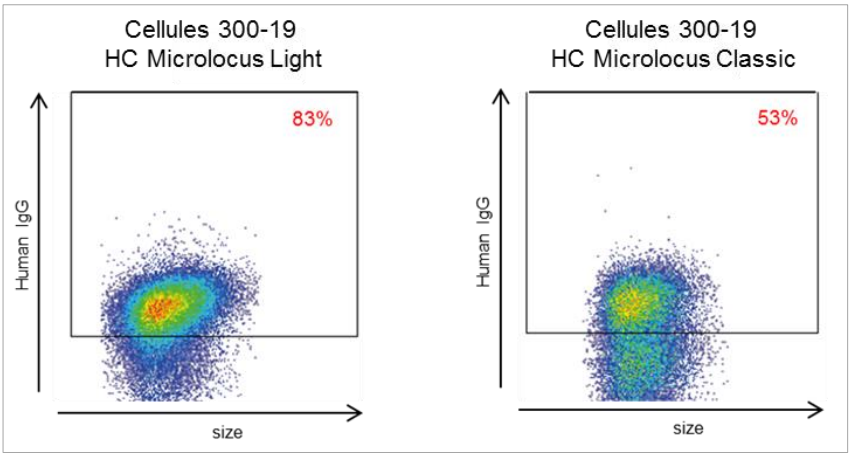

Figure 5: Enrichissement des cellules exprimant ɣ1 HC en surface. Pour HC Microlocus Light (*panneau gauche*) et HC Microlocus Classic (*panneau droit*), l'expression du transgène a été évaluée sur des cellules vivantes. Le seuil de positivité a été déterminé en fonction de cellules 300-19 WT. Le pourcentage de cellules exprimant des chaines lourdes humaines en surface est indiqué en rouge, en haut, à droite de chaque image.

Par la suite, nous avons évalué la présence de chaînes lourdes humaines dans le surnageant de culture en utilisant une méthode immuno-enzymatique (ELISA). Nous avons ainsi pu détecter une concentration d'environ 10 ng/mL de chaînes lourdes d'anticorps humains.

Concernant les cellules HC Minilocus, les analyses par FACS ont été perturbées par l'expression de la protéine fluorescente mCherry. La fonctionnalité complète n'a donc pas pu être mise en évidence mais la détection de plusieurs réarrangements VDJ nous indique que cette construction peut également générer des anticorps humains.

Tous les résultats obtenus à partir des cellules 300-19 transfectées avec les constructions codant pour la chaîne lourde des anticorps humains démontrent que nos vecteurs contiennent toute l'information génétique nécessaire pour produire efficacement des chaînes lourdes humaines dans un système murin.

6. Analyse *in vivo* de la fonctionnalité des constructions générées

Ce succès avec les cellules en culture nous a fortement incités à démarrer un programme de transgénèse dans la souris. Les HC Microlocus Light, HC Microlocus Classic et HC Minilocus ont été injectés dans des ovocytes de souris HC knockout (HC KO), le HC Microlocus Light shRNA a été injecté dans des ovocytes de souris wild type (WT). A l'heure actuelle, seule la transgénèse avec le HC Microlocus Light (dans le background souris HC KO) et le HC Microlocus Light shRNA (dans le background WT) ont mené à la génération de plusieurs animaux transgéniques, permettant ainsi une analyse de l'expression du transgène. Le programme de transgénèse avec les deux autres constructions, HC Microlocus Classic et HC Minilocus, n'a pas permis de générer suffisamment d'animaux transgéniques et doit donc être poursuivi.

6.1. Souris HC Microlocus Light

Les souris HC KO ont été générées par délétion des segments J du *locus* HC endogène. Ces souris sont donc incapables d'exprimer une chaîne lourde et montrent un blocage de la différenciation des cellules B au stade pro B ; par conséquent elles n'ont aucune

cellule B mature dans le sang. L'intégration d'une construction codant pour la chaîne lourde au génome de ces souris devrait donc permettre la reconstitution du pool de cellule B. L'injection du HC Microlocus Light dans des ovocytes issus des souris HC KO a été réalisée avec succès, générant 17 souris transgéniques. Cependant, aucune cellule B n'a pu être détectée dans ces souris, suggérant que les ɣ1 HC CH1$^-$ humaines ne participent pas correctement aux différentes étapes de la génération de cellules B *in vivo*.

6.2. Souris HC Microlocus Light shRNA

La constuction HC Microlocus Light shRNA a été injectée dans des ovocytes de souris WT. La transgénèse a été réalisée avec succès puisque nous avons obtenu 5 souris transgéniques dont une montre l'expression de chaînes lourdes d'anticorps humains dans le cytoplasme et à la surface des cellules B. II a été observé, dans cette souris transgénique, qu'environ 2% des cellules B du sang (CD19+) présentent l'anticorps humain ɣ1 HC CH1$^-$ en surface (**Figure 6**). Nous avons montré que les cellules B qui expriment le transgène, expriment également l'IgM de souris. D'autre part, l'analyse par FACS de l'expression d'IgM de souris ne montre aucune différence entre cette souris transgénique et une souris WT. Ceci suggère que les shRNA n'ont aucun effet sur l'expression endogène d'IgM.

A partir de cette souris transgénique (HC Microlocus Light shRNA #5), une lignée a été établie. Nous avons ainsi pu générer une colonie de 42 souris transgéniques qui, comme la souris #5, expriment des chaînes humaines à la surface des cellules B du sang. Nous avons aussi pu démontrer par ELISA que ces chaînes lourdes d'immunoglobuline humaine sont également secrétées dans le sérum à une concentration d'environ 30 µg/mL.

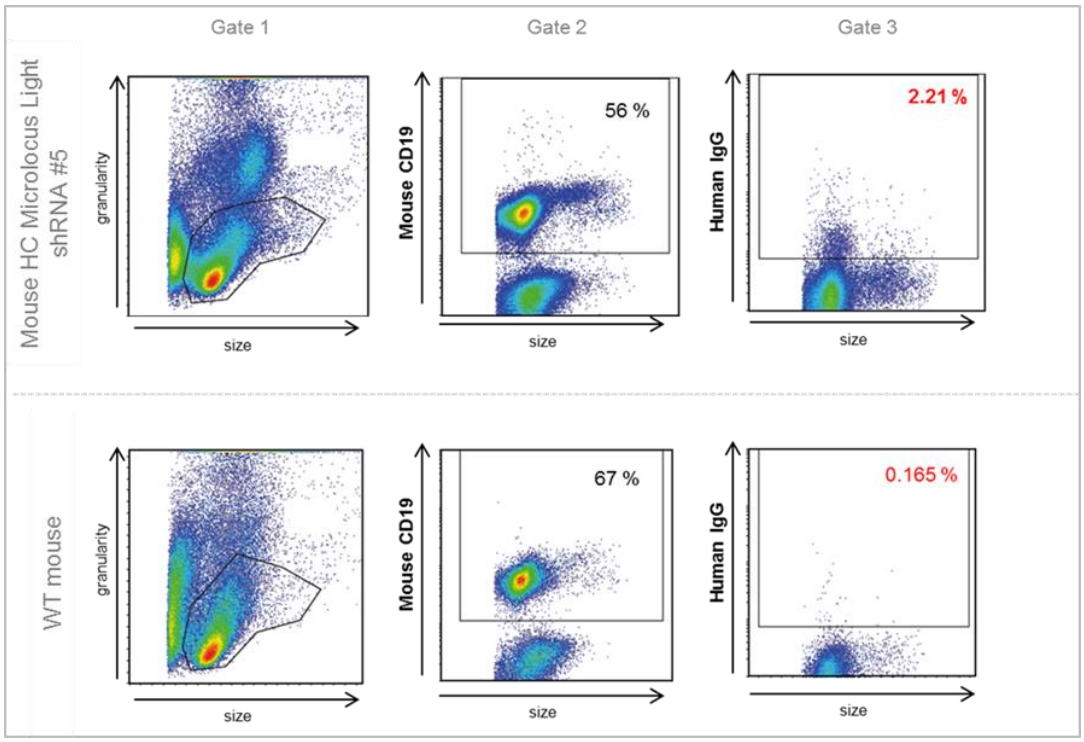

Figure 6: Détection de chaines ɣ1 HC CH1⁻ à la surface des cellules B du sang de la souris HC Microlocus Light shRNA #5 (*panneau supérieur*). En guise de control négatif, la même analyse a été conduite sur les cellules B d'une souris WT (*panneau inferieur*). Les cellules vivantes ont d'abord été sélectionnées en fonction de leur taille et de leur granularité (*Gate 1*), ensuite les cellules B ont été sélectionnées en fonction de l'expression du marquer de surface CD19 (*Gate 2*). Enfin l'expression d'IgG humaine (ɣ1 HC CH1⁻) a été évaluée (*Gate 3*) sur les cellules B précédemment définie. Le pourcentage des cellules B exprimant des ɣ1 HC CH1⁻ humaines est indiqué en rouge.

7. Conclusion et perspectives

Mon travail de thèse a permis de réaliser plusieurs étapes nécessaires pour obtenir, à terme, une large collection d'anticorps monoclonaux humains à partir de souris transgéniques :

- clonage des *loci* des anticorps humains dans des vecteurs de type YAC chez la levure,
- construction de vecteurs plasmidiques contenant l'ensemble de l'information génétique nécessaire pour l'expression hétérologue d'un répertoire de chaînes lourdes d'anticorps humains,
- mise au point d'un système d'analyse par transformation d'une lignée cellulaire permettant de valider la fonctionnalité des constructions codant pour la chaîne lourde des anticorps humains,

- standardisation de multiples techniques d'analyse génomique et de protéomique pour identifier les souris qui ont intégré des transgènes.

L'utilisation de la lignée cellulaire 300-19 pour tester toutes nos constructions n'est pas décrite par ailleurs et a permis de faire un pas en avant décisif dans le cadre de ce projet. Nous avons ainsi pu étudier en détail la fonctionnalité des constructions intermédiaires HC Microlocus Light, HC Microlocus Light shRNA et HC Microlocus Classic. Toutes les étapes de la production des anticorps ont été analysées, incluant les réarrangements somatiques VDJ, la diversité recombinatoire des différents segments, la diversité jonctionnelle entre les segments V et D, et D et J, la transcription, la maturation de l'ARN messager, l'épissage alternatif, l'expression de chaines lourdes à la surface des cellules ainsi que la sécrétion.

Ces résultats *in vitro* nous ont encouragés à entamer un programme de transgénèse avec toutes ces constructions. Les HC Microlocus Light, HC Microlocus Classic et HC Minilocus ont été injectés dans des ovocytes de souris HC knockout (KO), le HC Microlocus Light shRNA a été injecté dans des ovocytes de souris wild type (WT). Le programme de transgénèse avec les constructions HC Microlocus Classic et HC Minilocus doit être poursuivi, car à l'heure actuelle nous n'avons obtenu que très peu d'animaux transgéniques.

La transgénèse avec les deux autres constructions, HC Microlocus Light et HC Microlocus Light shRNA, s'est révélée efficace puisque nous avons obtenu plusieurs animaux transgéniques. Aucune cellule B n'a été détectée dans les souris HC Microlocus Light (background HC KO) ; alors que cette même construction associée à un shRNA (HC Microlocus Light shRNA) dans une souris transgénique au système immunitaire humoral intact, montre l'expression de chaînes lourdes d'anticorps humains dans le cytoplasme et à la surface de cellules B.

L'un des premiers stades de la différenciation des cellules B est l'expression d'une HC fonctionnelle sous la forme d'un récepteur de cellules pré B (pré-BCR). Quand la cellule pré B a accompli cette étape, le pré-BCR est présenté à la surface de la cellule et doit être capable d'initier la signalisation intracellulaire, permettant ainsi à la cellule pré B de passer au stade de différentiation suivant.

Le fait que le transgène soit exprimé dans le background WT mais pas dans le background HC KO nous mènent à l'hypothèse suivante : les ɣ1 HC CH1$^-$ émergeant de ces deux constructions transgéniques ne peuvent qu'être exprimées en parallèle des anticorps endogènes. En fait les ɣ1 HC CH1$^-$ ne seraient pas capables de participer à la différentiation des cellules B, c'est-à-dire que les ɣ1 HC CH1$^-$ engagés dans le pré-BCR des lymphocytes pré B ne seraient pas aptes à transmettre les signaux adéquats permettant à la cellule pré B d'évoluer vers le stade de différenciation suivant. Le développement des cellules B serait donc bloqué, ce qui expliquerait l'absence de cellule B mature dans le sang des souris au background HC KO.

Une étude récente a démontré qu'une glycosylation conservée dans le domaine CH1 d'IgM joue un rôle essentiel dans la capacité du pré-BCR à initier la signalisation intracellulaire [7]. Afin de tester expérimentalement le problème de signalisation des ɣ1 HC CH1$^-$, nous pourrions cloner l'exon μ CH1 en amont du gène ɣ1 CH1$^-$ existant et injecter cette nouvelle construction adaptée dans des ovocytes de souris KO HC. Un résultat positif conduirait au développement des cellules B matures et confirmerait la nécessité de la participation du domaine μ CH1 dans la signalisation du pré-BCR.

Enfin si les ɣ1 HC CH1$^-$ ne sont pas capables d'induire une signalisation, cela implique qu'ils ne sont alors pas soumis aux contrôles visant à éliminer les cellules B productrices d'anticorps auto-réactifs. Ainsi, les molécules de ɣ1 HC CH1$^-$ humaines constitueraient un répertoire d'anticorps vierges, ce qui pourrait conduire à la génération de bibliothèque de chaînes lourdes humaines avec une capacité de reconnaissance d'antigènes très divers et ce dans le contexte de la découverte de nouvelles molécules thérapeutiques.

Références

1. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575-581 (1983).

2. Lonberg, N. *et al.* Antigen-specific human antibodies from mice comprising four distinct genetic modifications. *Nature* **368**, 856-859 (1994).

3. Kouprina, N. & Larionov, V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. *Nature protocols* **3**, 371-377 (2008).

4. Hamers-Casterman, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446-448 (1993).

5. Alt, F., Rosenberg, N., Lewis, S., Thomas, E. & Baltimore, D. Organization and reorganization of immunoglobulin genes in A-MULV-transformed cells: rearrangement of heavy but not light chain genes. *Cell* **27**, 381-390 (1981).

6. Reth, M.G., Ammirati, P., Jackson, S. & Alt, F.W. Regulated progression of a cultured pre-B-cell line to the B-cell stage. *Nature* **317**, 353-355 (1985).

7. Ubelhart, R. *et al.* N-linked glycosylation selectively regulates autonomous precursor BCR function. *Nature immunology* **11**, 759-765 (2010).

Table of contents

Abbreviations and acronyms 6

1. Introduction 9

1.1. Immune system 12

1.1.1. Innate and adaptive immune system 12

1.1.2. Primary barrier: the skin 14

1.1.3. Organs of the immune systems 14

1.1.3.1. Primary organs 14

1.1.3.1.1. Bone marrow 14

1.1.3.1.2. Thymus 15

1.1.3.2. Secondary organs 15

1.1.3.2.1. Spleen 16

1.1.3.2.2. Lymph nodes and lymphatic system 16

1.1.3.2.3. Peyer's patches 17

1.2. Cells of immune system 17

1.2.1. Antigen presenting cells (APCs) 17

1.2.2. B Lymphocytes 18

1.2.2.1. B cell development 18

1.2.2.2. B cell maturation 19

1.2.3. T Lymphocytes 20

1.2.3.1. T cell hematopoiesis 20

1.2.3.2. NK cells 22

1.2.3.3. Cytotoxic T cells 22

1.2.3.4. Helper T cells 23

1.2.4. Interactivity 24

1.3. Molecules of the immune system 26

1.3.1. MHC molecules 26

1.3.1.1. MHC class I 26

1.3.1.2. MHC class II 27

1.3.2. T cell receptor (TCR) 27

1.3.2.1. Structure 27
1.3.2.2. Helper function /Cytotoxic function 28
1.3.3. Antibodies 30
1.3.3.1. Structure 30
1.3.3.1.1. Chains 30
1.3.3.1.2. Domains 31
1.3.3.1.3. Regions 32
1.3.3.1.4. Parts 33
1.3.3.2. Domain folding 33
1.3.3.3. Effector functions 35
1.3.3.4. Antibody isoforms 36
1.3.3.5. BCR 38
1.3.3.6. Non classical antibodies 39
1.4. Molecular biology of Immunoglobulin 40
1.4.1. Antibody genes 40
1.4.1.1. Ig HC *locus* 41
1.4.1.2. Ig LC kappa *locus* 43
1.4.1.3. Ig LC lambda *locus* 44
1.4.2. Generation of antibodies 44
1.4.2.1. V(D)J recombination 44
1.4.2.2. Antibody diversity 46
1.4.2.3. Membrane *versus* secreted Ig, alternative splicing 47
1.4.2.4. Affinity maturation 48
1.4.2.5. Class switch recombination 49
1.4.3. Molecular checkpoints of central and peripheral B cell development 49
1.4.3.1. Phases of B cell development and Ig gene rearrangement 49
1.4.3.2. B cell tolerance 53
1.4.3.3. Peripheral activation 54
1.5. Therapeutic 54
1.5.1. Why antibodies as therapeutic? 54
1.5.1.1. Biomedical 54
1.5.1.2. Safety 55
1.5.1.3. Targets 55

1.5.2. Human and Humanized monoclonal antibodies 57
1.5.3. Phage or yeast display 58
1.5.4. Human monoclonal antibodies produced in humanized non-human species 58
1.6. Production of humanized transgenic mice approaches 59
2. Aim of the project 61
3. Methods and results 62
3.1. Main approach: human Ig *loci* cloning as YAC 62
3.1.1. Genetic material 63
3.1.2. Introduction to TAR cloning 64
3.1.3. Adaptation of TAR cloning method: TAR bridging strategy 68
3.1.4. Human Ig *loci* cloning 69
3.1.4.1. Human Ig LC kappa cloning 69
3.1.4.2. Human Ig LC lambda cloning 74
3.1.4.3. Human Ig HC cloning 77
3.1.5. Human Ig YAC collection 81
3.1.6. YAC purification for transgenesis 81
3.2. Alternative approach: size reduced human Ig HC *loci* 83
3.2.1. Prerequisite to design reduced human Ig HC *loci* 84
3.2.2. The design of the HC Minilocus 84
3.2.3. Choice of Ig genetic elements 85
3.2.3.1. Design of the (D) diversity and (J) junction elements 85
3.2.3.1.1. The D cluster 85
3.2.3.1.2. The J cluster 87
3.2.3.1.3. The synthesized D-J construct 87
3.2.3.2. The V cluster 88
3.2.3.2.1. The variable (V) segments selection 88
3.2.3.2.2. V segments isolation and assembling 90
3.2.3.2.3. The two intermediary constructs 92
3.2.3.3. Cloning (C) constant regions 93

3.2.3.4. Regulatory elements 97
3.2.4. Intermediate constructs 98
3.2.4.1. HC Microlocus Light 99
3.2.4.2. HC Microlocus Classic 101
3.2.4.3. HC Microlocus Light shRNA 103
3.2.5. HC Minilocus 107
3.2.5.1. Assembling V, D and J regions 107
3.2.5.2. Assembling V and C parts by TAR 109
3.2.6. Construct collection 113
3.3. Testing reduced Ig *loci* in the 300-19 cell line 115
3.3.1. 300-19 cell line 115
3.3.2. Transfection of 300-19 cells with reduced HC constructs 116
3.3.3. Human HC expression analysis by FACS 117
3.3.3.1. Method 118
3.3.3.2. Results 118
3.3.4. Positive cell enrichment 122
3.3.4.1. Human ɣ1 HC positive cell panning 122
3.3.4.2. Human ɣ1 HC positive cell sorting by flow cytometry 124
3.3.5. 300-19 ɣ1 HC phenotype characterization 126
3.3.5.1. Method 127
3.3.5.2. Results 128
3.3.6. Human Ig secretion *in vitro* measured by ELISA assay 131
3.3.6.1. Method 131
3.3.6.2. Results 132
3.3.7. Human Ig secretion *in vitro*; ELISPOT assay 133
3.3.7.1. Method 134
3.3.7.2. Results 135
3.3.8. Analysis of secreted human antibodies 136
3.3.8.1. Method 137
3.3.8.2. Results 138
3.3.9. DJ rearrangement analysis 141
3.3.9.1. Method 141
3.3.9.2. Results 144

3.3.10. VDJ rearrangements analysis 148
3.3.10.1. Analysis at DNA level 148
3.3.10.2. Results 150
3.3.10.3. Analysis at RNA level 152
3.3.10.4. Results 153
3.3.10.5. Unusual rearrangements 154
3.3.11. Diversity 155
3.3.11.1. Recombinatory diversity/ Gene usage 155
3.3.11.2. Junctional diversity 156
3.3.12. Alternative splicing 157
3.3.12.1. Method 157
3.3.12.2. Results 159

3.4. Transgenesis in mice 160
3.4.1. Generation of HC KO mice 161
3.4.2. Injections in HC KO background 161
3.4.2.1. HC Microlocus Light transgenesis 161
3.4.2.2. HC Microlocus Classic transgenesis 162
3.4.3. Mice HC Microlocus light shRNA in WT background 162
3.4.3.1. Genotyping 162
3.4.3.2. Detection of human ɣ1 HC by FACS 163
3.4.3.3. shRNA effect 166
3.4.3.4. VDJ rearrangement 167
3.4.3.5. Breeding of Mouse HC Microlocus Light shRNA #5 168
3.4.3.6. Serum analysis of human ɣ1 HC and mouse IgM 169
3.4.3.7. Human γ1 HC expression in progenitor B cell subsets 169

Discussion 176

Conclusion and perspectives 181

Bibliography 184

Abbreviations and acronyms

A

Ab: Antibody
ADCC: antibody-dependent cell mediated cytotoxicity
Ag: Antigen
Ago: Argonaute protein
AID: Activation Induced Cytidine Deaminase
APC: Antigen Presenting Cells

B

BCR: B cell receptor

C

CD: Cluster of Differentiation
cDNA: complementary DNA
CDR: Complementarity-Determining Regions
CEN: Centromere
CLP: Common lymphoid Progenitor
CSM: Complete Supplement Mixture
CTL: Cytotoxic T lymphocytes
DTT: Dithiothreitol

D

dNTP: deoxynucleoside triphosphate

E

ELISA: Enzyme Linked Immunosorbent Assay
ER: Endoplasmic Reticulum
ES cell: Embryonic Stem Cell
Fab: Fragment, antigen binding

F

Fc: Fragment, crystallizable
FCS: Fetal Calf Serum
FSC: Forward Scatter

H

HC: Heavy Chain
HSC: Hematopoietic Stem Cell

I

Ig: Immunoglobulin
IL: Interleukin
IFN: Interferon
ITAMs: Immunoreceptor Tyrosine-based Activation Motifs
ISC: Immunoglobulin Secreting Cell

K

kbp: kilobase pair
kDa: kilodalton

L

LB medium: Lysogeny Broth medium
LC: Light Chain

M

MHC: Major Histocompatibility Complex

N

N: Nucleotide not characterized
NHEJ: non-homologous-end-joining
NK: Natural Killer

O

OD: Optical Density
ORI: Origin of Replication
ORF: Open Reading Frame

P

PolyA: polyadenylation site
PEG: polyethylene glycol
PCR: Polymerase Chain Reaction

R

R: Receptor
RISC: RNA-induced silencing complex
RNAi: RNA interference
RSS: recombination signal sequence
rt: room temperature
RT: Reverse Transcription

S

SDS: Sodium Dodecyl Sulfate
SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SHM: Somatic HyperMutation
shRNA: short hairpin RNA
SLC: Surrogated Light Chain
SSC: Side Scatter

T

TAR: Transformation-Associated Recombination
TCR: T Cell Receptor
TdT: Terminal Deoxynucleotidyl Transferase
TE: Tris-EDTA
TH cells: helper T cells
TM: Transmembrane

W

WT: wild type

Y

YAC: Yeast Artificial Chromosome

1. Introduction

Preface

At the junction between the XIXth and XXth century, Paul Ehrlich and Elie Mechnikov's were both implicated in scientific research leading to basic discoveries in immunology. Ehrlich postulated of creating "magic bullets" to fight against human diseases. Magic bullets reflect indeed the active molecules of the armed arm of the humoral immune system (**Figure 1**).

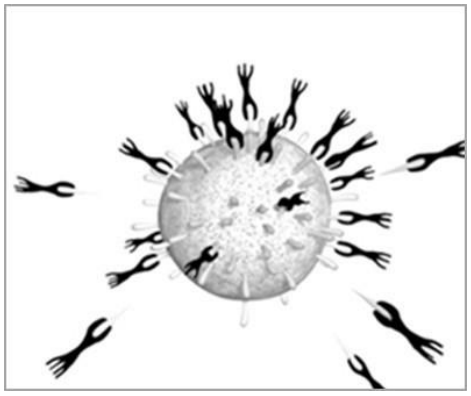

Figure 1: Ehrlich's magic bullets.

Mechnikov identified phagocyte cells, part of the cellular immune system. Ehrlich and Mechnikov were involved in a war of words, each group hunting for evidence that would prove their own theory correct, and their opponent's wrong. In the end, both antagonists were right, because the immune system has, in fact, two defensive arms: the innate immunity and the acquired immunity (also called adaptive immunity). Mechnikov and Ehrlich shared the Nobel Prize in Medicine in 1908.

Research by Gerald Edelman and Rodney Robert Porter in the early 1960s produced fundamental advancement in the understanding of the antibody chemical structure, opening a door for further study [8] (**Figure 2**). For this work, Edelman and Porter shared the Nobel Prize in Physiology or Medicine in 1972.

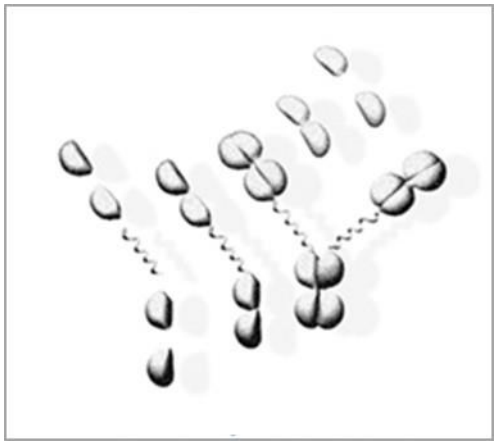

Figure 2: Antibody's structure.

In the late 1940s, George Snell introduced the notion of histocompatibility genes [9]. In 1980, he shared the Nobel Prize in Physiology or Medicine with Baruj Benacerraf and Jean Dausset for their discoveries concerning "genetically determined structures on the cell surface that regulate immunological reactions", termed later the Major Histocompatibility Complex (MHC) molecules.

The Nobel Prize in Physiology or Medicine 1984 was awarded jointly to Georges J.F. Köhler, César Milstein and Niels K. Jerne, "for theories concerning the specificity in development and control of the immune system and the discovery of the principle for production of monoclonal antibodies" [10] (**Figure 3**).

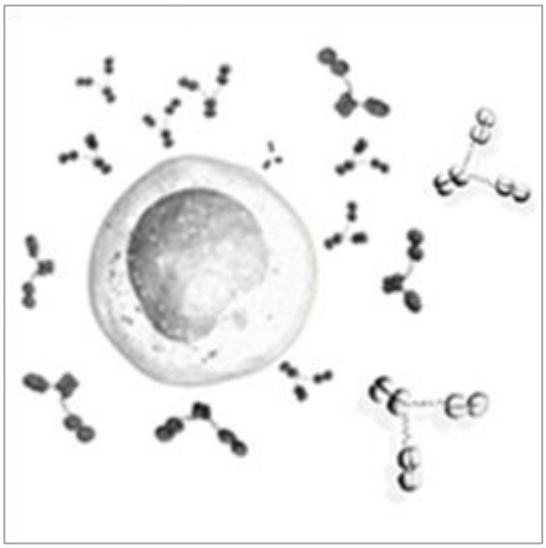

Figure 3: Hybridoma technology.

Susumu Tonegawa, in 1974, shows the first evidences for somatic generation of antibody diversity and in 1976, he described the mechanism of genetic recombination that produce antibody diversity [11] (**Figure 4**). He won the Nobel Prize for Physiology or Medicine in 1987.

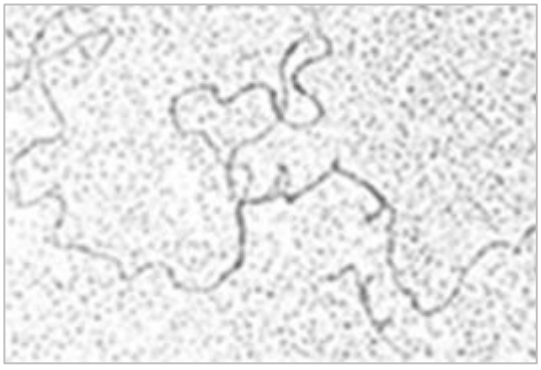

Figure 4: Electromicrogramm showing rearranged HC genes.

Ralph Brinster and collaborator Richard Palmiter pioneered techniques to transfer foreign genes into mammals, by injecting purified DNA into mouse oocyte and showing transmission of the genetic material to subsequent generations. They first utilized these methods to elucidate the activity and function of genes. They developed the first "transgenic mice", and their seminal experiments catalyzed a worldwide revolution in genetic engineering in the 1980s.

Few years later, in 1986, Mario Capecchi and Oliver Smithies were able to genetically modify mouse embryonic stem (ES) cell by homologous recombination. Injection of such modified ES cells into mouse blastocyst enables generation of transgenic mice [12]. They used this technology for disrupting or altering mouse genes at will. In 2007, they received the Nobel Prize in Physiology or Medicine for their work in gene manipulation that led to the development of knockout (KO) mice.

1.1. Immune system

1.1.1. Innate and adaptive immune system

The immune system is a remarkably versatile system evolved to protect animals from invading pathogenic microorganisms, without generating an excessive damage to the own tissues. It can be divided into two major branches: innate and adaptive immune system. The first passive lines of defense against invaders are epithelial barriers, like the skin. After brakeage by injury of the first barrier, phagocytes (such as neutrophils and macrophages) do indeed provide the first active line of defense. They compose the hard core of the innate immunity. They also clear up debris from other invasive cell destructions, like injuries or immune-system battles. Another crucial role of phagocytes is to collaborate with other cells by inducing inflammation to attract other phagocytes to the spot of invasion or by presenting, in the secondary immune organs, to cells of the adaptive immune system what they have engulfed and destroyed, and thereby inducing the activation cascade leading to a specific immunologic response through antibody production.

Adaptive immune responses are carried out by white blood cells called lymphocytes. There are two broad classes of such responses, antibody responses and cell-mediated immune responses, and they are carried out by different classes of lymphocytes, termed B cells and T cells, respectively. The defining characteristics of adaptive immunity are exquisite specificity for distinct molecules and an ability to remember and respond more vigorously to repeated exposure to the same pathogen. In antibody responses, B cells are activated to secrete antibodies, glycoproteins also named immunoglobulins. The antibodies circulate in the bloodstream and permeate the other body fluids, where they bind specifically to the foreign antigen. Antibodies recognize non-self antigens and are highly specific for that one antigen only. Binding of antibody inactivates viruses and microbial toxins by blocking their ability to bind to receptors on host cells. Antibody binding also marks invading pathogens for destruction, mainly by making it easier for phagocytic cells of the innate immune system to ingest them.

In cell-mediated immune responses, the second class of adaptive immune response, T cells are involved with two specific functions:

- one subset, called helper T cells (TH cells), collaborates with the B cells and activates those which are specific for the same antigen. Activated B cells will then pass through a developmental process leading them to develop in either memory B cell and/or plasma cells (*i.e.* immunoglobulin secreting cells); and
- another subset consists of cytotoxic T lymphocytes (CTL) which specifically kill cells infected by a specific pathogen.

Regulatory T cell subpopulation is responsible for modulating immune responses, specifically inhibiting effector T cells in order to control ongoing immune reaction.

Both cellular and humoral immunity are involved in clearing the infection. The result of this defense is the eradication of the pathogen. During the fight, the different cells optimize their weapons and after elimination of the pathogen, they remain in the body and remember the intruder. This leads to long term memory for this particular attack and is the base of the principle of vaccination where inactivated or recombinant antigens are introduced into the body to induce an immune response without symptom of the disease.

Different antigens elicit different type of immune responses. Intracellular organisms such as bacteria, parasites, or viruses which are using host cells to replicate induce a cell-mediated immune response. In contrast extracellular organisms and soluble antigens induce a humoral response.

To summarize, the main specificities of the acquired immune system are:

- **protection**: the immune system protects against disease;
- **tolerance**: the immune system discriminates between self and non self by reacting against every molecule different to the endogenous;
- **specificity**: the immune system can recognize thousands of millions of different antigens, and for each determinant, a specific lymphocyte will be induced; and
- **memory**: the immune system acquires long-term active memory following infection. If the same antigen appears again, a secondary response occurs in a faster and more effective way than in the primary response.

1.1.2. Primary barrier: the skin

The body's own defense against microorganisms begins directly at the skin surface. Special fatty acids from the sebaceous glands and the secretions of certain bacteria belonging to the physiological skin flora inhibit the growth of fungi and bacteria. If the skin is damaged the injured cells recruit non-specific leukocytes which travel from close venules to the site of damage. Those cells leave the blood vessel by extravasation or diapedesis to migrate along the chemokine gradient towards the site of tissue damage or infection.

1.1.3. Organs of the immune systems

1.1.3.1. Primary organs

1.1.3.1.1. Bone marrow

Hematopoiesis, which takes place in bone marrow, is the generation of all circulating blood cells, which all derived from hematopoietic stem cells (HSC) (**Figure 5**). Bone marrow is also the source of self-renewing populations of stem cells. In this immunological organ, common lymphoid progenitors (CLPs), derived from HSCs, give rise to B cell and T cell precursors [13, 14]. Progenitor T cells leave the bone marrow and seed the thymus, where they undergo differentiation to naïve T cells. By contrast, all stages of B cell differentiation, from HSC to naïve B cell, occur in bone marrow. At different stages, B cells are tested to not be auto-reactive and to express a functional B cell receptor (BCR).

In addition to be a primary lymphoid organ where lymphocyte development takes place, the bone marrow is also a place for antibody-secreting cells. After B cells have maturated into plasma cells in the secondary lymphoid organs, they enter the blood stream and migrate to the bone marrow where they may become long-term plasma cell.

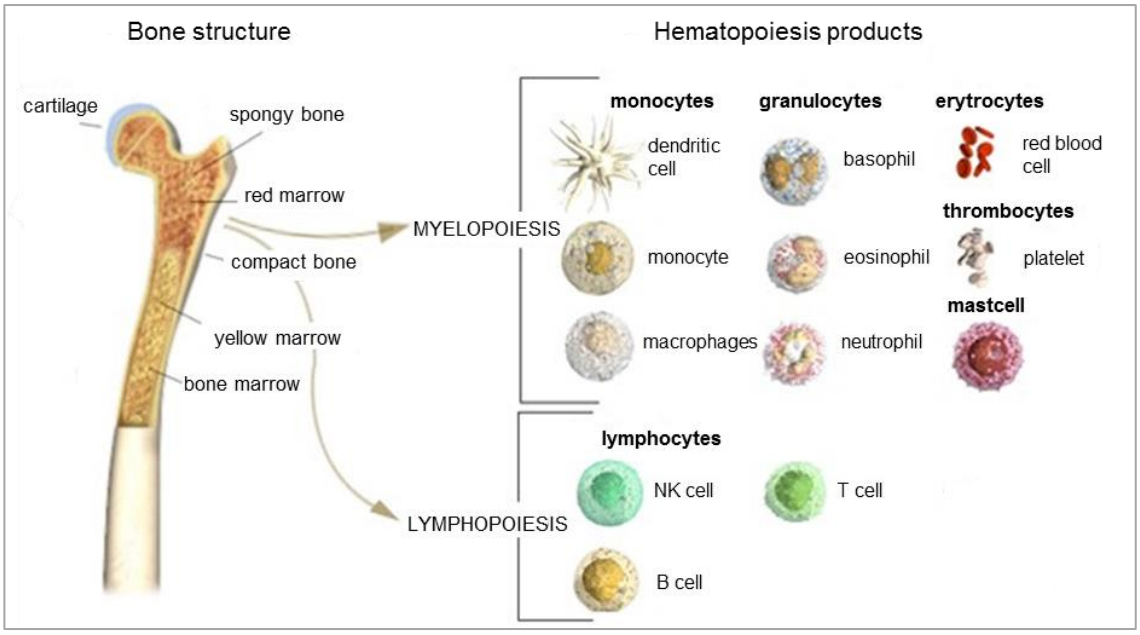

Figure 5: The bone marrow: site of generation of all circulating blood cells.

1.1.3.1.2. Thymus

Once in the thymus, the progenitor T cells, CD4, CD8 or T cell receptor (TCR) negative at this stage, will perform the last differentiation steps and start to mature into naive T cells. The thymus is a school for T cells where they will be checked for a functional TCR, which corresponds to the positive selection, and also to be safe for the body, meaning to not react toward self, which is termed negative selection. These checking points are located in different compartments of the thymus. Many cells will be eliminated by apoptosis and a few survivors can then migrate to periphery where after appropriate activation become helper T cell (TH) or cytotoxic T cells (CTL) depending on their predisposition. The helper T cell will expose CD4 on their surface and the cytotoxic T cell, CD8 as co receptors and as phenotypic markers.

1.1.3.2. Secondary organs

Secondary lymphoid organs are the places colonized by mature and immunocompetent lymphocytes and function in response to antigenic stimuli to mount an efficient response. Antigens are transported to secondary lymphoid organs through lymphatic

vessels *via* antigen-presenting cells (APC). Secondary lymphoid organs include the lymphatic system (lymph nodes), the spleen, and the mucosal associated lymphoid tissue (MALT) including Peyer's patches, the appendix, the tonsils, and selected regions of the body's mucosal surfaces. The secondary lymphoid organs serve two basic functions: they are a site of further lymphocyte maturation, and they efficiently trap antigens for exposure to T and B cells.

1.1.3.2.1. Spleen

The spleen is found in the abdominal cavity behind the stomach. Compared to lymph node, the spleen filters blood rather than lymph. One of its main functions is to bring blood into contact with immune cells. The functional tissue of the spleen is made up of two types of cells: the red pulp, which contains cells named macrophages that remove bacteria, old blood cells, and debris from the circulation; and surrounding regions of white pulp, which contain great numbers of lymphocytes. The splenic artery enters the red pulp through a web of small blood vessels, and blood-borne microorganisms are trapped in this loose collection of cells until they are gradually washed out through the splenic vein. The white pulp contains both B and T lymphocytes. T cells congregate around the tiny arterioles that enter the spleen, while B cells are located in regions called germinal centers, where they are exposed to antigens with the collaboration of helper T cells. This is the process of B cell activation, which leads to antibody-secreting plasma cells and finally to memory B cells.

1.1.3.2.2. Lymph nodes and lymphatic system

The lymph nodes, or lymph glands, are small, encapsulated bean-shaped structures composed of lymphatic tissue. Thousands of lymph nodes are found throughout the body along the lymphatic routes, and they are especially prevalent in areas around the armpits (axillary nodes), groin (inguinal nodes), neck (cervical nodes), and knees (popliteal nodes). The nodes contain lymphocytes, which enter from the bloodstream. The function of collecting antigen from their portals of entry and delivering them to lymph node is performed largely by the lymphatic system. Indeed, the dendritic cell, after capturing

antigens, enters lymphatic vessels, and reaches lymph node, where they can present antigen to a naïve T cells *via* their major histocompatibility complexes (MHC). If the T cell recognizes specifically the antigen, it will become activated and adaptive immune response is initiated. Then, activated helper T cell goes looking for B cells specific for the same antigen and activates them. Activated lymphocytes, carried in the lymph, exit the node through the efferent (outgoing) vessels and eventually enter the bloodstream, which distributes them throughout the body.

1.1.3.2.3. Peyer's patches

Peyer's patches are ovoid bundles of lymphatic tissue made up of unencapsulated lymphatic cells that protect the mucous membranes of the small intestines from infection. They belong to the mucosal immune system. The complete role and function of these mucosal-associated lymphoid tissues are uncertain except that they are colonized by lymphocytes and antigen-presenting cells (APCs).

1.2. Cells of immune system

Many cells from the innate and adaptive immune system are engaged in a complex collaboration to initiate efficiently immune responses to protect the body from pathogens [15]. Interactions between different cell types are mediated by specific molecules which allow communication between the cells. Depending on these interactions a powerful immune reaction can be mounted to recognize and eliminate any invader. These cells, with specific functions, are described below.

1.2.1. Antigen presenting cells (APCs)

An antigen presenting cell (APC) is a cell that capture, process and displays foreign antigens peptides in the context of major histocompatibility complexes (MHC) on their surfaces. Specific T cells recognize these complexes using their T cell receptor (TCR). Antigen presenting cells act as messengers between the innate and the adaptive immune systems. They are present in those tissues that are in contact with external environment

such as the skin and the inner lining of the nose, the lungs, the stomach and the intestines.

Professional APCs are very efficient at internalizing antigen, either by phagocytosis or by receptor-mediated endocytosis. The engulfed antigen is degraded intracellularly and short fragments or peptides are displayed on the cytoplasmic membrane in the context of MHC class II molecules. The combination, peptide-MHC, forms the antigen recognized by the T cell. Several types of professional APC have been identified (**Figure 6**).

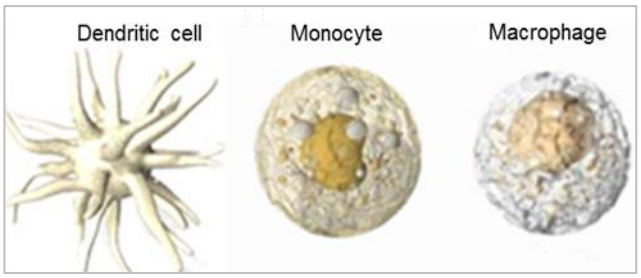

Figure 6: The professional antigen presenting cells.

The dendritic cells have the largest range of antigen presentation and are especially potent helper T cell activators. They have long cytoplasmic projections which effectively increase their surface area. Macrophages are specialized phagocytic cells that attack foreign substances, infectious microbes and cancer cells through destruction and ingestion before MHC presentation to T cells. Monocytes are precursors of macrophages and are actively recruited into inflammatory sites, were they can differentiate into macrophages or dendritic cells.

B cells can internalize the antigen captured with the specific BCR and, after processing, they can present it on surface in the context of MHC class II molecules. They are inefficient APC for most other antigens.

1.2.2. B Lymphocytes

1.2.2.1. B cell development

B cell development occurs through several stages, each stage representing a change in the genome content at the antibody *loci* [16] (**Figure 7**). All steps occur in a precise order (see 1.4.3.1.). After successful heavy and light chain genes rearrangement, an immature B cell express its own BCR encoded by previously rearranged antibody genes. B cell

hematopoiesis generates a very large range of B cells that may be able to recognize, through their BCR, a very large range of different antigens. Before leaving the bone marrow, immature B cells are controlled to not be auto-reactive. B cell with a BCR that recognizes self-antigens will die by apoptosis; this is the negative selection.

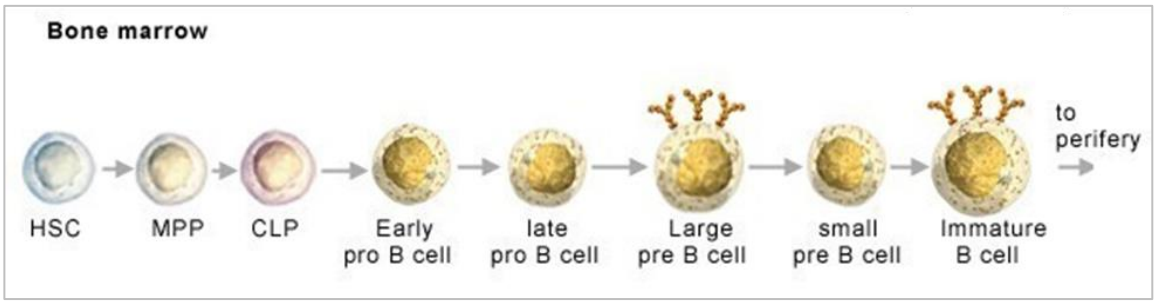

Figure 7: The B cell development stages in bone marrow. (HSC: Hematopoietic Stem Cells; MPP: Multi-Potent Progenitor; CLP: Common Lymphoid Progenitor).

Multiple stages of development between the HSC and immature B lymphocytes have been very thoroughly investigated and characterized, revealing important growth factors and regulatory interactions; but also the central role of the pre-B cell receptor (pre-BCR). The B cell development with the status of IgH and IgL genes (status of BCR) at each stage is further described in the part 1.4.3.1.

1.2.2.2. B cell maturation

Immature B lymphocytes enter the circulation and populate peripheral lymphoid organs and tissues which are site of interaction with foreign antigens. Once a naive B cell binds an antigen specific for its BCR, a maturation process is engaged in the splenic germinal center [17]. Upon activation by specific helper T cell, the B cell proliferates, switches isotype to acquire new functions, performs somatic mutations to optimize the antigen affinity and matures to antibody secreting plasma cells [18]. These effector cells migrate to the site of infection were secreted antibody can neutralize antigens. A subset becomes memory B cells, which reside in tissues and in secondary lymphoid organs (**Figure 8**).

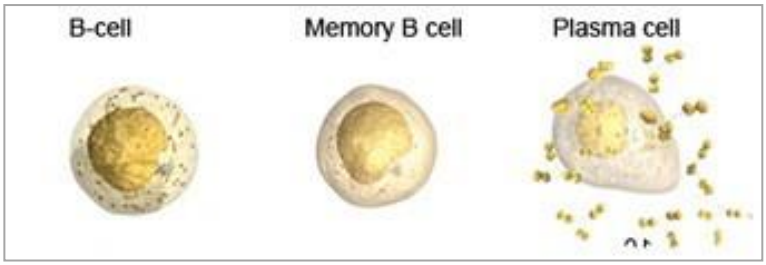

Figure 8: The different B cell subsets.

There are two main populations of B lymphocytes, each of which is composed of distinct subpopulation defined by phenotype and function. A minor population, accounting for around 5% of B lymphocytes, is localized to serous cavities. They are called B-1 B lymphocytes. Although the bone marrow retains the potential to produce B-1 B lymphocytes, they are more efficiently generated from progenitors that arise during embryogenesis.

The predominant population of B lymphocytes, referred to as B-2 B lymphocytes, resides in the spleen and the lymph nodes. B-2 B cells are generated in the bone marrow postnatal life and then migrate to the spleen where, after progressing through several transitional stages, they mature into a major population of follicular and minor population of marginal zone B lymphocytes. Marginal zone B lymphocytes do not circulate and respond to blood-borne pathogens while follicular B lymphocytes respond to protein antigens and, in response to T cell helper, undergo class switching and affinity maturation into B-2 B lymphocytes.

1.2.3. T Lymphocytes

1.2.3.1. T cell hematopoiesis

All T cells originate from hematopoietic stem cells in the bone marrow and become progenitor T cells, which migrate and populate the thymus and expand by cell division to generate a large population of immature thymocytes (**Figure 9**).

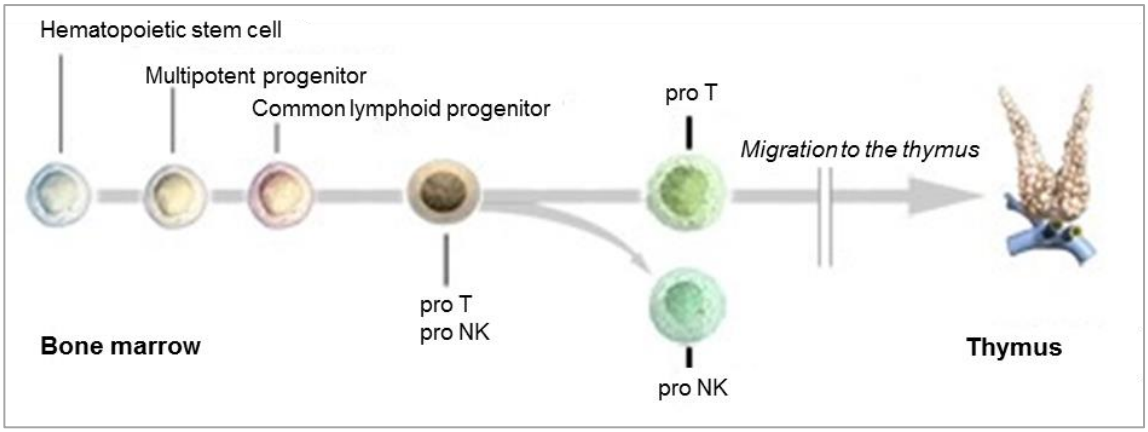

Figure 9: T cell hematopoiesis.

The earliest thymocytes express neither CD4 nor CD8, and are therefore classed as double-negative (CD4-CD8-) cells. At this stage the TCR β chain genes start their recombination, which lead to the expression of β chain at the cell surface. This step allows double-negative cells to mature into CD4 and CD8 double positive cells, and initiates rearrangement of α chain genes. Both chains, α and β, in association with CD3 form the TCR complex. The first step of maturation consists in checking for a functional TCR on cell surface (positive selection). Cells which do pass this exam are eliminated by neglect. Those cells which succeed are subjected to the second control, which eliminates the auto-reactive cells by apoptosis (negative selection). A minority of cells succeeds both controls and leaves the thymus as immature T lymphocytes. At this point they turn off one of the co-receptor and become single positive CD4 or CD8 with predestinated functions [19]. The thymic checking points are located in different compartments of the thymus.

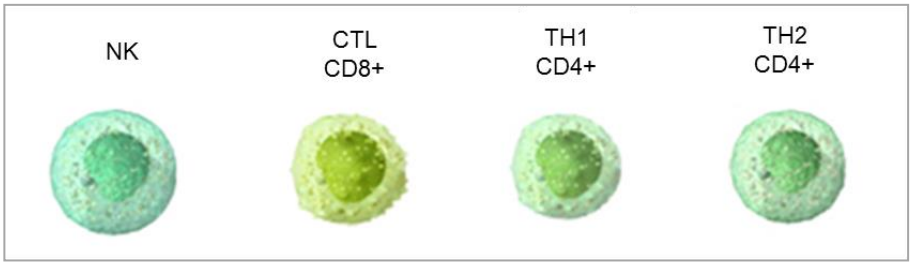

Figure 10: The T cell subsets. NK: Natural Killer cell, CTL: Cytotoxic T lymphocyte, TH: T Helper cell.

1.2.3.2. NK cells

Natural killer cells recognize infected and stressed cells and respond by directly killing these cells.

NK cells are classified under lymphocytes and are able to eliminate cells, like tumor cells or virus infected cells. NK cells do not have any antigen specific receptor and are part of the innate immune system. To avoid recognition by cytotoxic T lymphocytes (CTLs), some viruses down regulate the expression of MHC molecules or even produce proteins that pick MHC molecules out of the endoplasmic reticulum. NK cells specifically recognize cells that have no or low MHC class I molecules on the surface and drive them to apoptosis. CTLs and NK cells are complementary to protect the body.

NK cells can also bind to antibody-coated cells by Fc receptor and destroy these cells with a process called antibody-dependent cellular cytotoxicity (ADCC). Once bound to antibody Fc portion, the NK cell releases cytokines such as interferon-ɣ, and cytotoxic granules containing perforin and granzymes. These cytotoxic granules enter the target cell and trigger apoptosis leading to cell death.

1.2.3.3. Cytotoxic T cells

Effector cytotoxic T cells directly kill cells that are infected with a virus or some other intracellular pathogen.

Cytotoxic T lymphocytes (CTLs) express classical alpha-beta TCR and the co-receptor CD8. After education and selection, only few cytotoxic CD8 positive cells will leave the thymus to migrate into secondary immune organs. If they find infected cells that are recognized by testing the MHC class I – peptide complex with their specific TCR and co-receptor CD8, they are then activated. Specific recognition of peptide-MHC class I complex on a target cell by a cytotoxic CD8 T cell leads to the death of the target cell by apoptosis. Cytotoxic T cells can recycle to kill multiple targets (**Figure 11**). Each killing requires the same series of steps, including receptor binding and directed release of cytotoxic proteins, like perforin, stored in lytic granules.

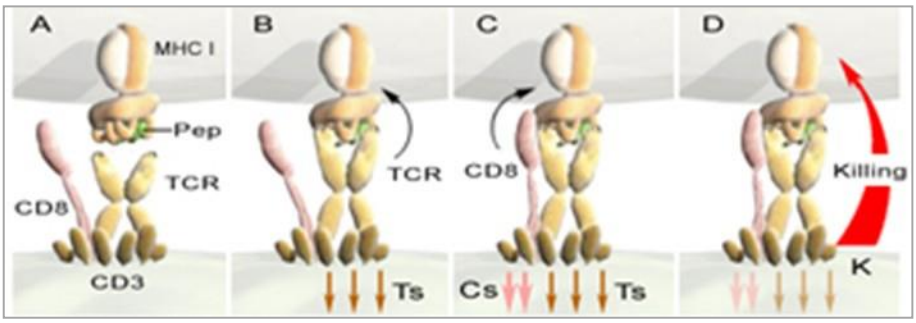

Figure 11: The cytotoxic T lymphocyte action.
A: targeted cell presents peptide (Pep) *via* its MHC I,
B: TCR binds to MHC I and starts to signal (Ts) *via* CD3 associated to TCR,
C: CD8 binds to MHC I and triggers further signaling (Cs),
D: CTL stats the killing process of the target cell (K).

1.2.3.4. Helper T cells

Effector helper T cells help stimulate the responses of other cells; mainly macrophages, B cells, and cytotoxic T cells.

In contrast to cytotoxic T cells, helper T cells are crucial to mount a potent humoral and cellular immune response. Therefore, they stimulate B cells to further maturate to make specific antibodies, which neutralize and eliminate pathogens and their toxic products from body fluid. They also activate cytotoxic T cells that kill infected target cells and thereby interrupt the pathogen life cycle. In addition, the helper T lymphocytes activate macrophages to become more voracious and therewith contribute to the eradication of the pathogen.

Naive helper T cell can only be activated by an antigen-presenting cell to become an effector cell. When activated, a naïve helper T cell differentiates in either TH1 or TH2, two distinct types of effector helper cells. TH1 cells mainly help activate macrophages and cytotoxic T cells, whereas TH2 cells mainly help activate B cells.

By contrast to CTLs, helper T cells recognize specifically the antigen peptide/MHC class II complex on the surface of the APC or B cells. Once in interaction, the accessory molecule CD4 is engaged by binding to the MHC part. That configuration induces intracellular signaling which launch the T cell activation. This results in clonal cell expansion, each new generation recognizing the same antigen complex. If the activated cell is a TH1 type, the activation mediated by cytokines will concern CTL cells and macrophages. In contrast, an activated helper TH2 cell activates exclusively B

lymphocytes bearing on their surface an identical antigen/MHC class II complex than the one which activated previously the TH2 cell. The members of the helper T clone generate chemical that can induce B cell or macrophage/CTLs activation. TH1 cells secrete, among others, IFN-ɣ, which acts on macrophage to increase phagocytosis. TH2 cells secret IL-4, which acts on B cells to stimulate production of antibodies.

1.2.4. Interactivity

Scenario for B cell activation

Germinal centers are sites within secondary lymphoid organs where mature B lymphocytes may become activated after recognition of its cognate antigen and stimulation by specific helper T cells. Activated B cells mutate their antibodies (through somatic hypermutation) to increase specificity against its related antigen; the clone that recognizes the antigen with much higher affinity is selected and preferentially expended. In parallel, the activated B cells also perform antibody class switch. Finally cells proliferate and maturate into plasma or into memory B cell.

B cell activation may occur in a T cell-dependent or T cell-independent manner. Here are described the helper T cell-dependent humoral immune responses to protein antigens. Protein antigens are recognized by specific B and T lymphocytes in peripheral lymphoid organs, and these cell populations come together in these organs to initiate immune response. The sequence of events that lead to the interaction of helper T cells with B cells, and the subsequent events that drive further B cell differentiation, is schematized in **Figure 12**.

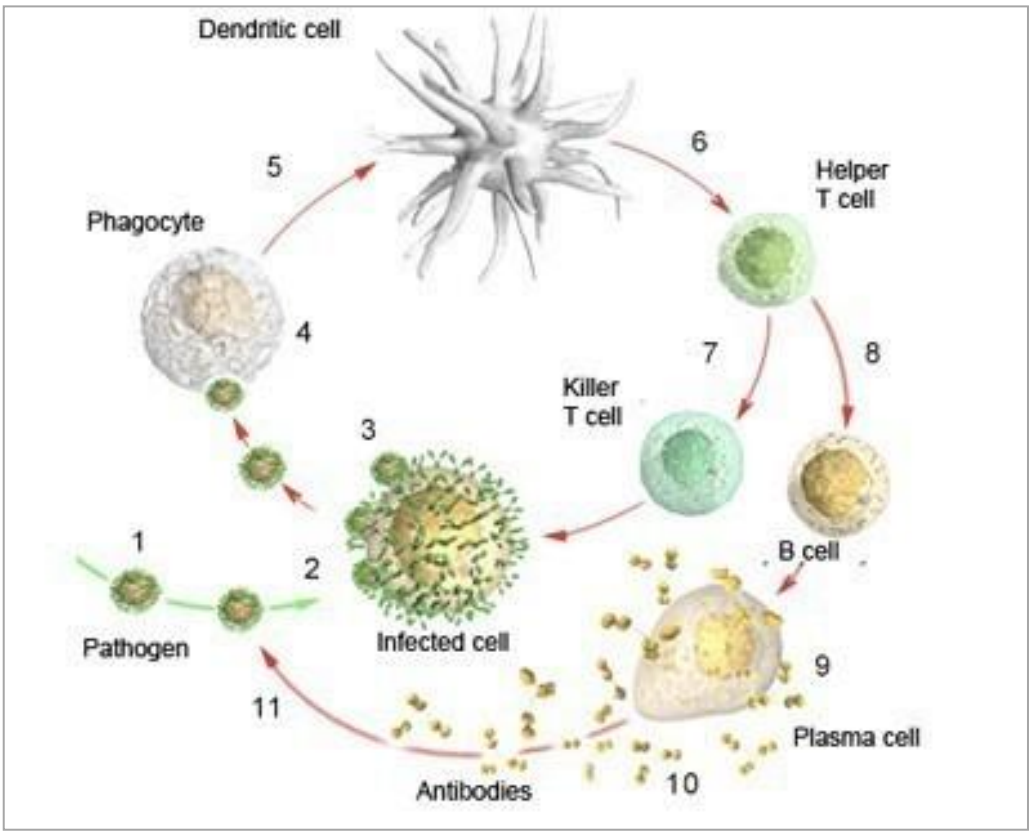

Figure 12: Schematic illustration of T dependent immune response. A pathogen (**1**), for example a virus, starts to infect cells (**2**) in order to multiply (**3**). Such infected cells secrete chemokine, which attract APCs (**4**, **5**). APCs phagocyte (process) infected cells and present peptide on their MHC class II. Dendritic cells migrate to secondary lymphoid organs through lymphatic vessels. MHC class II interacts with TCR complex of helper T cells (**6**) residing in secondary lymphoid organs. T CD4+ cells are then activated and can start to help CD8+ cells (**7**) and specific B cells (**8**). The activated B cells differentiate then in plasma cells (**9**), whereas activated CTLs kill specifically infected cells. Plasma cells produce a large amount of specific antibodies (**10**), which contribute to eliminate pathogen.

1.3. Molecules of the immune system

1.3.1. MHC molecules

1.3.1.1. MHC class I

MHC class I shows to the outer world what has been produced inside the cell [20].

MHC class I is present on all the cells from our body. They consist of a set of molecules inherited from our parents and combinations of them makes that everyone is immunologically different. This makes organ grafts so difficult. MHC class I molecules consist of two polypeptide chains, α chain and β2-microglobulin (**Figure 13**). Both chains are linked non-covalently. The α chain amino end forms a "plateau" exposing two alpha helix that contain the presented peptide with a length of 8 to 10 amino acids. The carboxyl end of the alpha chain contains a transmembrane (TM) region anchoring the molecule in the cell membrane, and interacts with the CD8 co-receptor of cytotoxic T cells. The β2-microglobulin is necessary for cell surface expression of MHC class I and stability of the peptide binding groove. MHC class I presents continuously peptides from proteins produced inside the cell. If infected, the cell will also present peptides from the infecting pathogen which will then be identified by cytotoxic T cells. They will kill the cell before the infection can spread out.

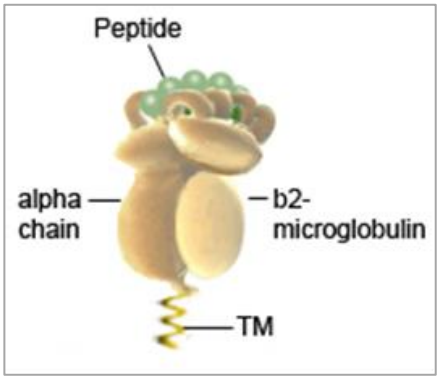

Figure 13: The MHC class I/peptide complex.

1.3.1.2. MHC class II

MHC class II shows to the outer world what has been eaten by the cell [21].

MHC class II is only present on cells of the immune system, particularly on antigen presenting cells like macrophages or dendritic cells from the innate immune compartment but also on B lymphocytes from the acquired immune system. MHC class II is a heterodimer composed of an alpha and a beta chain able to bind peptides of 13 to 24 amino acids (**Figure 14**). MHC class II continuously presents peptides from proteins engulfed by phagocytes or B lymphocytes on the surface to be presented to helper T lymphocytes.

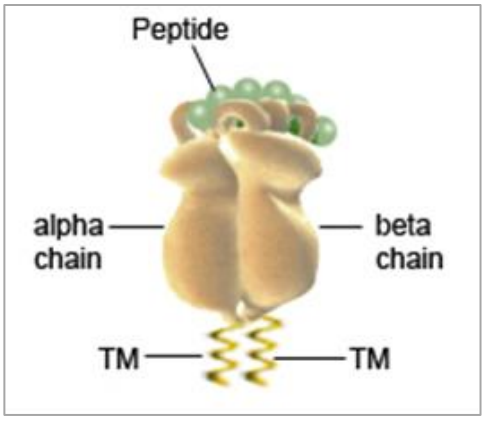

Figure 14: The MHC II/peptide complex.

1.3.2. T cell receptor (TCR)

1.3.2.1. Structure

T lymphocytes have a dual specificity: they recognize polymorphic residues of self MHC molecules and residues of peptide antigens displayed by these MHC molecules, which is responsible for their specificity. The receptor that recognizes these peptide-MHC complexes is called the T cell receptor (TCR) [22]. Like, the B cell receptor, TCR is a clonally distributed receptor, meaning that clones of T cells with different specificities express different TCR. TCR is associated with molecules of CD3 which contribute to

signaling after specific recognition of MHC-peptide complex by the TCR. Together, they form the TCR complex (**Figure 15**).

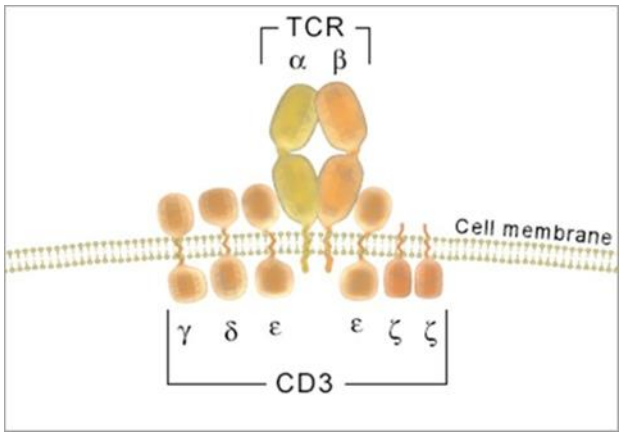

Figure 15: The TCR complex.

The antigen receptor of MHC-restricted CD4+ helper T cells and CD8+ cytotoxic is a heterodimer consisting of two transmembrane polypeptide chains, designated α and β, covalently linked to each other by a disulfide bridge. Each α chain and β chain consists of one Ig-like N-terminal variable domain, one Ig-like constant domain, a hydrophobic transmembrane region, and a short cytoplasmic region. The V region of the TCR α and β chains contains short stretch of amino acids were the variability between different TCRs is concentrated, and these form the hypervariable regions, that specifically recognize peptide-MHC complexes. Each TCR chain is encoded by multiple gene segments that undergo somatic rearrangements during the maturation of the T lymphocytes. The TCR complex has the same conformation on all types of T cells. However, the co-receptor CD4 and CD8 differ according the T cell type.

1.3.2.2. Helper function /Cytotoxic function

The affinity of TCR complex for peptide-MHC complexes on APCs or target cell is usually too low to mediate a functional interaction between the two cells by itself. T cells normally require accessory receptors to help stabilize the interaction by increasing the overall strength of the cell-cell adhesion. Unlike T cell receptors, the accessory receptors do not bind foreign antigens and are invariant. They bind to non-variable parts

of the MHC molecules, far away from the peptide-binding groove. The most important and best understood of the co-receptors on T cells are the CD4 and CD8 proteins, both of which are single-pass trans-membrane proteins with extracellular Ig-like domains.

The **Figure 16** shows that the TCR complex is associated with CD4 on T helper cells; CD4 allows interaction with MHC class II from APC. By contrast, on CTL surface, the TCR complex is associated with CD8, which can interact with MHC class I from target cell

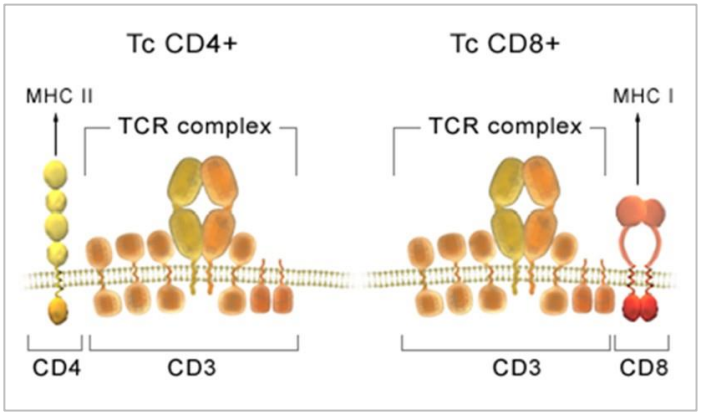

Figure 16: The TCR complex and co-receptors.

In part 1.3.1., was described the processes by which class I and II MHC molecules present peptides derived from intracellular and extracellular protein antigens, respectively. The segregation of CD4+ and CD8+ T cells responses to these different pools of antigens occurs because of the specificities of CD4 and CD8 for different classes of MHC molecules. Most CD4+ class II-restricted T cells are cytokines-producing helper cells and functions in host defense against extracellular pathogens that are ingested by APCs. CD4+ T cell that express the specific TCR against the peptide/MHC II complex presented on APC, become activated and can start to help specific B cells. Most CD8+ class I-restricted T cells are CTLs, which serve to eradicate infections by intracellular pathogens that reside in the cytoplasm of infected cells.

1.3.3. Antibodies

An antibody (Ab), also known as an immunoglobulin (Ig), is a large Y-shaped protein produced by B cells that is used by the immune system to identify and neutralize foreign substances such as bacteria, viruses and toxins. The antibody recognizes a unique part of the foreign target, called an antigen.

1.3.3.1. Structure

1.3.3.1.1. Chains

The basic structure of each immunoglobulin molecule is a unit consisting of 1) two light chains (LC); and 2) two heavy chains (HC). The amino acid sequences of the two LC are identical; so are the sequences of the two HC. There are five types of mammalian Ig heavy chain: α, δ, ε, ɣ and μ. The type of heavy chain defines the class of antibody; the prelisted HC types are found in IgA, IgD, IgE, IgG and IgM respectively. Each heavy chain has two regions, the constant and the variable regions. The constant region is identical in all antibodies of the same isotype. Heavy chains α, δ and ɣ have a constant region composed of three tandem Ig domains, and a hinge region conferring flexibility. Heavy chains ε and μ have a constant region composed of four immunoglobulin domains and no hinge region. The variable region of heavy chain differs in antibodies produced by different B cells, but is the same for all antibodies produced by a single B cell or B cell clone. The variable region of each heavy chain is approximately 110 amino acids long and is composed of a single Ig domain.

In mammals, there are two types of immunoglobulin light chain, which are identified lambda (λ) and kappa (κ). A light chain has two successive domains: one constant domain and one variable domain. The approximate length of a light chain is 215 amino acids. The LC is the same for all antibody produced by a single B cell.

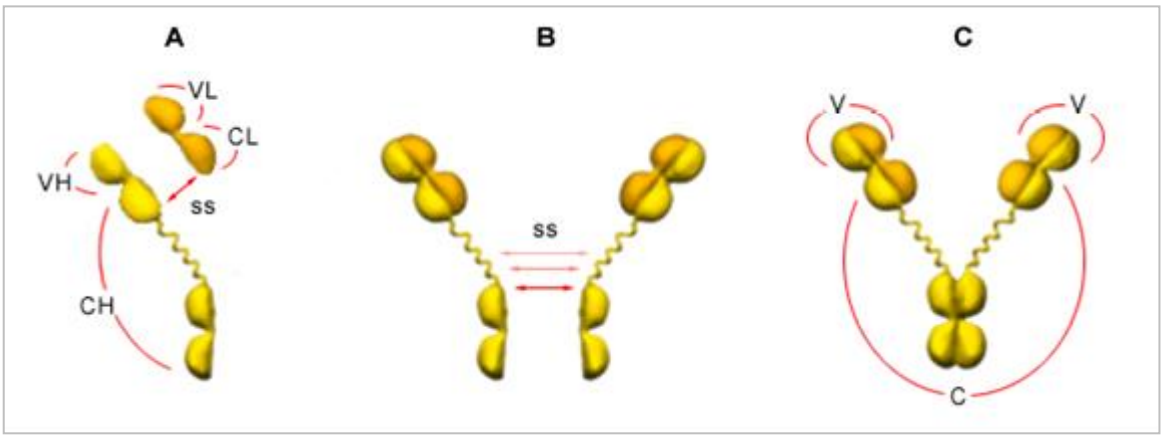

Figure 17: The chain assembly in IgG.
A; HC and LC are each composed of one variable part (VH and VL respectively) and one constant part (CH and CL respectively). Both chains are covalently associated *via* disulfide bridges (ss).
B; both heterodimer, composed of one HC and one LC, are covalently associated *via* disulfide bridges (ss) into a homodimer molecule.
C; functional antibody contains variable parts (V), composed of VH and VL, which specifically recognize antigens; and the constant part (C) which mediates effector functions to the antibody.

1.3.3.1.2. Domains

The domain is the basic unit of protein tertiary structure. It can be defined as the polypeptide chain or that part of it that can fold autonomously into a stable tertiary structure. A number of globular proteins are composed of single domains but many proteins are composed of multiple domains. All the major proteins of the immune system (antibodies, T cell receptors, MHC proteins) are multi-domain proteins.

Antibodies are composed of several compact polypeptides domains composed of approximately 100-110 amino acids and stabilized by one intra-domain disulfide bridge. They are exclusively composed of seven beta shits. One refers to antibody domains which are largely used in nature not only for antibodies but to build up many other molecules containing immunoglobulin domains alone or in combination with other kind of folding.

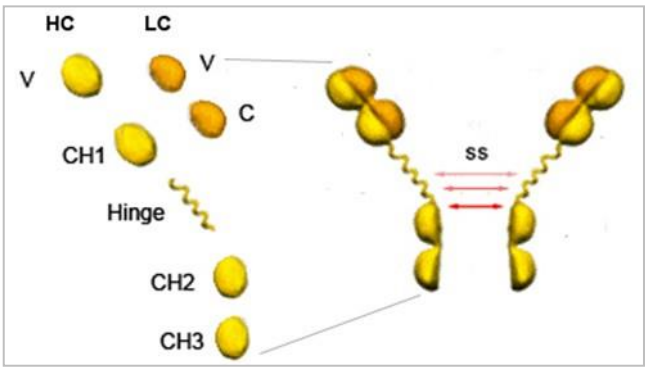

Figure 18: The IgG domains. The ɣ HC constant part is composed of domains CH1, CH2, CH3, and a hinge region. The LC constant part is composed of only one domain; either Cκ or Cλ. Heavy and light chain have one variable domain.

1.3.3.1.3. Regions

The N-terminal region of these chains composes the variable antigen binding site (V region); whereas the C-terminal part is called constant region (C region).

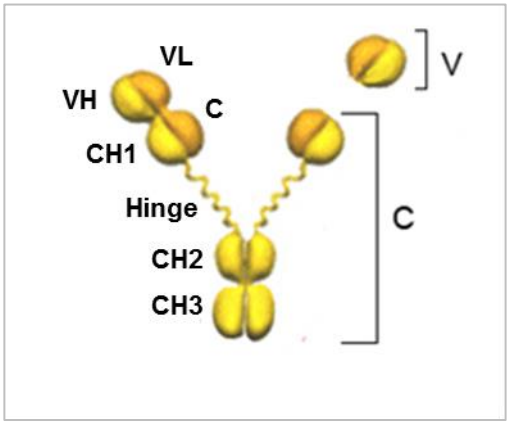

Figure 19: The IgG regions. V: Variable antigen binding site is composed of VH and VL; C: Constant region is composed of CH1, Hinge, CH2, CH3 from heavy chain and C from light chain.

While the V region binds and neutralizes the specific target by aggregation, the constant region carries effector functions that can activate the complement cascade to kill the bacterium directly or interact with phagocyte cell surface Fc receptor to mediate phagocytosis, which will ultimately result in destruction of the invading microbe. The Fc receptors are isotype-specific, which gives greater flexibility to the immune system. Another mode of action involving Fc receptor is Antibody-Dependent Cell-mediated Cytotoxicity (ADCC), a mechanism of cell-mediated immune defense whereby an

effector cell of the immune system actively lyses a target cell, whose membrane-surface antigens have been bound by specific antibodies. Classical ADCC is mainly mediated by natural killer (NK) cells. Not every immunoglobulin isotype will mediate all effector functions (See 1.3.3.4.).

1.3.3.1.4. Parts

Proteolytic treatment of IgG by Porter, Nisonoff and others in the early sixties allowed cleavage of the molecule with papain (**Figure 20**) in two fragments: one that retained antigen binding (Fab: Fragment, antigen-binding) and one that did not (Fc: Fragment, crystallizable). Cleavage of the IgG with pepsin (**Figure 20**) instead produced a bivalent Fab (Fab2) but resulted in degradation of the Fc. These studies suggested a molecule with three functional domains (two antigen-binding and one non antigen-binding), a hypothesis that later proved to be correct.

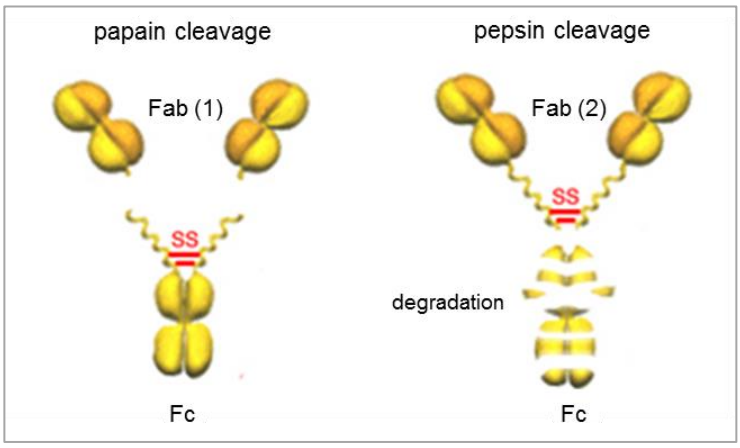

Figure 20: Antibody fragmentation with papain (*left panel*) and pepsin (*right panel*) treatment. Fab (1): Fragment, antigen-binding; Fab (2): bivalent fragment, antigen-binding; Fc: Fragment, crystallizable; SS: disulfide bridge.

1.3.3.2. Domain folding

The protein folding is the physical process by which a polypeptide made of amino acid residues fold into its characteristic and functional three-dimensional structure. Two main types of secondary structure have been identified; the alpha helix and the beta strand or beta sheet (**Table 1)**.

Table 1: The four major classes of protein folding

all α	composed mainly of α-helices (example: the four-helical bundle)
all β	composed mainly of β-sheets (example: antibodies and TCRs)
α/β	composed of α-helices and β-sheets that alternate along the chain
α + β	composed of α-helices and β-strands that tend to segregate

As seen in part 1.3.3.1.2., immunoglobulin heavy and light chains are composed of a series of discrete protein domains. These protein domains all have a similar folded structure. Within this basic three-dimensional structure, there are distinct differences between V and C domains. The structural similarities and differences can be seen in the scheme of **Figure 21**. Each domain is constructed from two β sheets, which are elements of protein structure made up of strands of the polypeptide chain (β strands) packed together; the sheets are linked by a disulfide bridge and together form a roughly barrel-shaped structure, known as a β barrel.

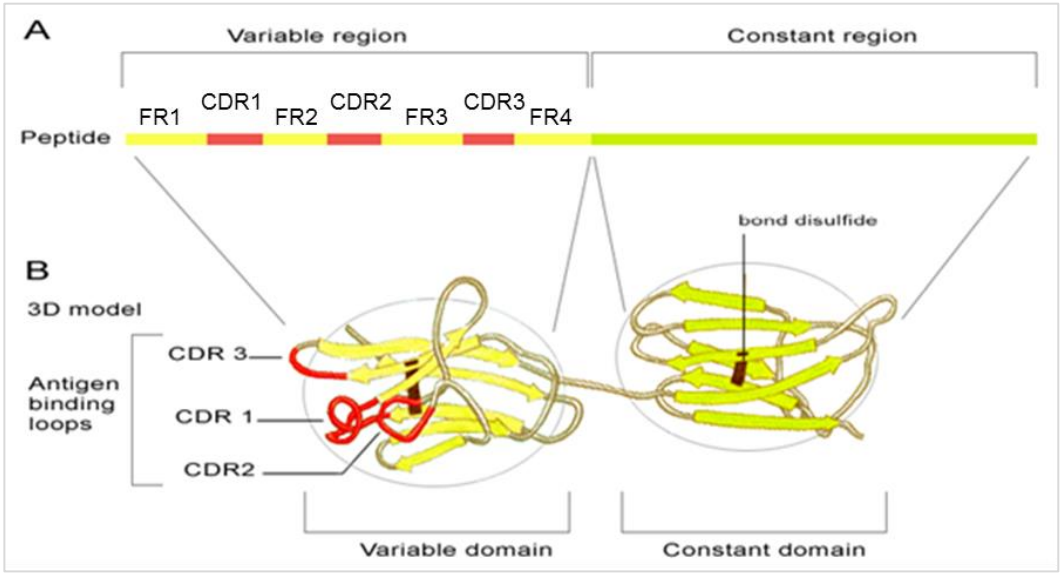

Figure 21: Antibody light chain structure. **A**: representation of the linear polypeptide chain, with complementarity-determining regions (CDRs) in red and framework residues (FRs) in yellow; **B**: 3D model showing domain folding and CDRs forming antigen-binding site.

The most variable parts of the V regions are limited to several small hypervariable regions or complementarity-determining regions (CDRs). The light- and heavy chain V regions contain three CDRs. The CDRs come together at the amino-terminal end of the antibody molecule to form the antigen-binding site, which determines specificity. The invariant regions of amino acids, between the CDRs, make up about 85% of the V regions and are designated as framework residues (FRs).

1.3.3.3. Effector functions

While the variable parts of the antibody recognize specifically the antigen, the constant portion contains several effector functions:

Complement fixation

When antibodies binds to the surface antigen of a pathogen, mostly bacteria, the complement cascade is activated. This results in lysis of the targeted cells.

Binding to various cell types (ADCC)

Phagocytic cells, lymphocytes (mainly NK cells), platelets, mast cells, and basophils have Fc receptors to which antibodies can bind. This binding can activate the cells to perform cytotoxic function.

Transport mechanisms

Some classes of antibodies can bind to specific receptors on placental trophoblasts, which results in transfer of the immunoglobulin across the placental barrier. As a result, the transferred maternal antibodies provide immune protection to the fetus and newborn until his own immune system takes over. A similar mechanism uses the milk as transfer agent of antibodies from mother to offspring. Other antibodies like IgA are actively transported to mucus.

Flexibility

The hinge domain of antibody is a flexible stretch of amino acids that joins the Fab arms to the Fc piece. The hinge element may differ by the length and the flexibility from one isotype to another. The flexibility of the hinge region in IgG and IgA molecules allows the Fab arms to adopt a wide range of angles, permitting binding to epitopes spaced at variable distances and angles.

Serum half life

The Ig domains confer stability to the antibody which endows the molecules with a long serum half-life.

Other interactions

Non immune associated functions such as binding protein A and G from *Streptococcus* is also associated to antibodies.

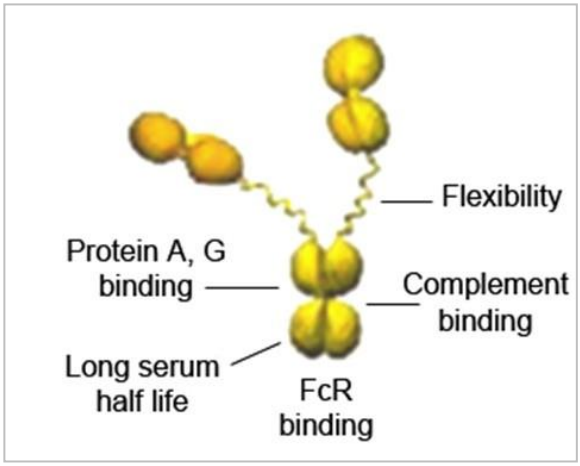

Figure 22: The effector functions of IgG.

1.3.3.4. Antibody isoforms

In human, five distinct classes of immunoglobulin molecules are recognized namely IgG, IgA, IgM, IgD, and IgE, respectively encoded by the constant exons Cγ, Cα, Cμ, Cδ and Cε. They differ in size, charge, amino acid sequence and carbohydrate content. The IgG class can be subdivided into four subclasses (IgG1, IgG2, IgG3, IgG4) and the IgA into two subclasses (IgA1, IgA2).

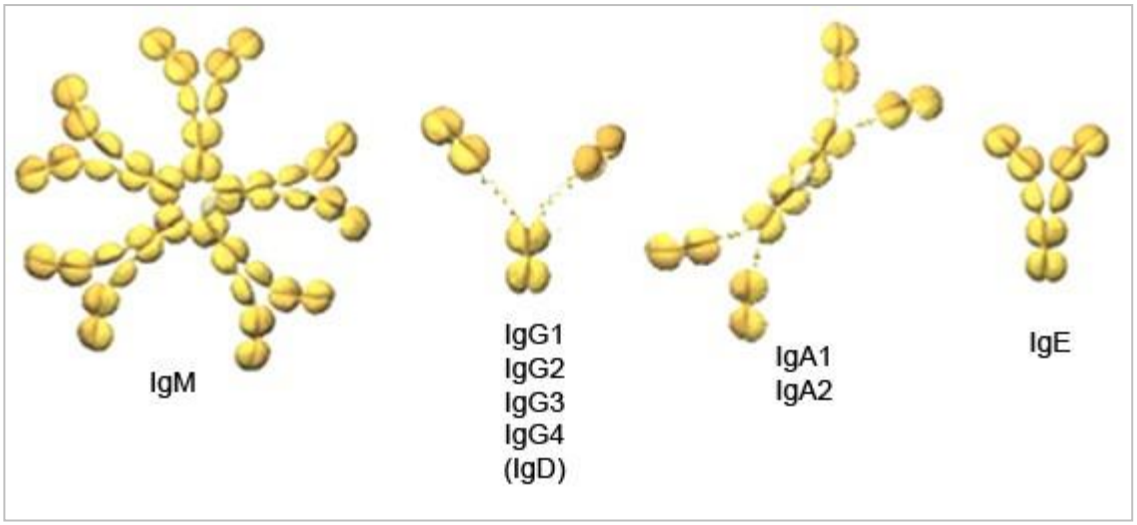

Figure 23: Representation of the different antibody isotypes.

IgM is a pentamer with a mass of 970 kDa and account for approximately 10% of the serum immunoglobulin pool. A transmembrane monomeric form *m*IgM is present as an antigen-specific receptor on immature and mature B cells.

IgG is the predominant immunoglobulin in normal human serum, this isotype accounts for 70-75% of the total serum immunoglobulin pool and consists of a monomeric four chains molecule. IgG1 is the most representative subclass.

IgA is a monomer, but in most mammals serum IgA is predominantly polymeric (occurring mostly as a dimer). In human serum, IgA accounts for around 15 to 20%. This form is actively transported into mucus (tears, saliva colostrum) and secretions (genitourinary tract, gastrointestinal tract prostate, respiratory epithelium).

IgD, accounting for less than 1% of the serum immunoglobulin pool, is an antigen-specific receptor on mature B cells, in fact a transmembrane monomeric (*m*IgD) form is present as an antigen-specific receptor on mature B cells (double positive B cells have both *m*IgM and *m*IgD receptors [23], though on any individual B cell these receptors will have the same antigen specificity).

IgE serum levels are very low relative to the other immunoglobulin isotypes. IgE-antigen complex binds to specific receptors (FcR-IgE) on basophiles and mast cells and thereby activate them. Afterward this cells release chemokines to attract other cells of the immune system. This reaction can eventually induce inflammation, asthma or anaphylactic chocks.

Table 2: Antibody isotype associated functions (Table adapted from Cellular and Molecular Immunology, edition 6, Abbas *et al*.).

Property	IgM	IgG1	IgG2	IgG3	IgG4	IgA1	IgA2	IgD	IgE
Heavy chain	μ	γ1	γ2	γ3	γ4	α1	α2	δ	ε
Forms (mer)	penta	mono	mono	mono	mono	di	di	mono	mono
Mol.Wt (kDa)	970	146	146	170	146	385	386	184	188
Half life (days)	10	21	20	7	21	6	6	3	2
Human serum Conc.(mg/mL)	1.5	9	3	1	0.5	3	0.6	0.03	0.00005
Complement activation	+++	+++	+	+++	-	-	-	-	-
Macrophage (Fc Receptor binding)	-	+	-	+	-	-	-	-	-
External secretions	mucus	milk	milk	milk	milk	mucus	mucus	-	-
Placental transfer	-	+	+	+	+	-	-	-	-

1.3.3.5. BCR

At the heart of the capability of B lymphocytes to generate specific responses to a huge number of antigens, is the B cell receptor (BCR) complex. The BCR is the surface Ig receptor of B cells. An association of antibody monomers of diverse forms (depending on the differentiation or maturation stage) and other surface molecules implicated in intracellular signaling (Igα and Igβ), forms the BCR complex (**Figure 24**).

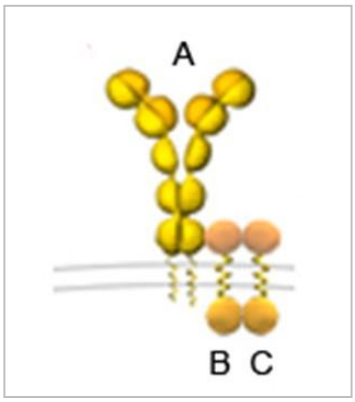

Figure 24: The B cell receptor, comprising 2 types of components: an antigen recognition unit constituted by a membrane immunoglobulin (A), and the signal transmission unit formed by the co-receptors Ig-alpha (B) and Ig-beta (C).

To detect its cognate antigen, each B cell contains up to 120 000 BCR complexes on its cell surface. The BCR is involved in signaling to the cell the *status quo* at the surface

during B cell differentiation and tolerance, and particularly in the presence of specific antigens. The BCR is the membrane form of the secreted antibodies. The two forms are generated in a modulated way depending on the differentiation or maturation stage of the cell by alternative splicing.

Transmembrane receptors have to convert an extracellular signal into an intracellular signal. In other words, antigen binding has to be translated into a structural alteration of the receptor that is transduced across the membrane and results in the activation or deactivation of intracellular signaling molecules (Igα and Igβ). The mechanism of signal transduction by the B lymphocyte antigen receptor complex is still investigated. It is thought that the BCR delivers activating signals to a B cell when two or more receptor molecules are brought together, or cross-linked, by multivalent antigens. This would induce phosphorylation of the tyrosine in the immunoreceptor tyrosine-based activation motifs (ITAMs) of Igα and Igβ. By contrast to the crosslinking theory, in 2000, the oligomeric organization of BCR complex was introduced. Each subunit in the oligomer consists of a 1:1 complex of the membrane Igα–Igβ heterodimer and the transmembrane-bound immunoglobulin molecule (*m*Ig). Such dynamic complex should alternate between a closed and an open structure [24]. According to this model, the oligomer complex stays closed in absence of stimulation, whereas upon antigen binding the oligomer complex open (distribution geometry changes) which lead to an activation signal. In absence of signal during a certain time, the cell dies by neglect.

The BCR has two key roles in B cell activation. First, antigen-induced conformation changes of cell surface receptors followed by intracellular signaling that initiates the process of activation of the engaged B cell. Second, the surface receptor bound antigen is internalized by endosomal vesicles, and processed into peptides after fusion with lysosomes. Finally the antigen peptides are presented in the context of MHC class II on the B cell surface to the helper T cells.

1.3.3.6. Non classical antibodies

Theoretically, absence of HC or LC expression leads to arrest of B cell development by apoptosis. However, some species produce HC-only antibodies (HCAb) as part of their normal B cell development and repertoire. In 1993 unusual IgG were discovered in

camel serum (*Camelus dromaderius*). These unconventional molecules have a molecular weight of around 90 kDa, indicating that they lack IgG CH1 domain. This was confirmed by further analysis. The missing CH1 might be replaced by an extended hinge [4]. Nguyen and colleagues demonstrated that these camel HCAbs undergo antigen-mediated selection and affinity maturation, and that their variable domains are subject to somatic hypermutation [25]. Because their variable domain is two times smaller than conventional antibodies, they are thought to recognize unusual epitopes. However, the details of generation of HC only antibodies are not well understood.

The antigen-binding site, a single variable domain, resembles V_H of conventional antibodies. However, differences in FR2 and CDR3 prevent pairing with a variable LC, and hydrophilic amino acids increase solubility. Human V_H segments of family 3 show similarities with V segments used to generate Heavy chain only antibodies in *camelidae* [26, 27].

Transgenic mice producing such antibodies were generated. The transgenic *loci* were composed of two llama V_H segments linked to the human HC D and J segments, followed by Cγ2 and Cγ3, both without CH1 exon. A rescue of B cell development was shown in mice bearing such transgenes [28].

In 1995 followed the discovery of new antigen receptor (IgNAR) in cartilaginous fish (Chordata).

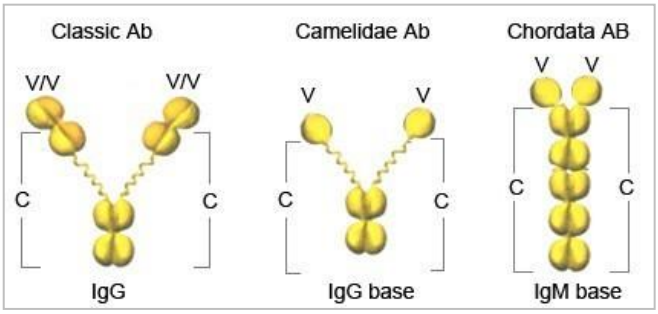

Figure 25: The unconventional antibody forms.

1.4. Molecular biology of Immunoglobulin

1.4.1. Antibody genes

The light chain Ig *loci* contain a set of V gene segments plus their associated leader (L) sequences, the J gene segments, and the C_L exons, while the heavy chain *locus* contains

the V (plus L), D, and J gene segment clusters [29, 30], and the C_H exons. Ig proteins cannot be produced from the string of multiple V, D, and J gene segments; indeed they need to rearrange to generate variable exons using V(D)J recombination (see 1.4.2.1.). Ig *loci* are filled with important regulatory DNA sequences in addition to the sequences that contain the amino acid coding information. These control sequences regulate V(D)J recombination, govern transcription initiation, and control isotype switching. Enhancers constitute one such class of control elements, and specific enhancers buried within the Ig loci have proved critical for the optimal expression of the Ig genes.

1.4.1.1. Ig HC *locus*

Located on chromosome 14, at band 14q32.33, the human HC *locus* covers 1 250 kilobase pairs (kbp) composed of specific gene segment clusters: V cluster = 41-46 V segments, D cluster = 25 D segments, J cluster = 6 J segments, C cluster = 9 C exons. Each C exon, which corresponds to a C isotype (μ, γ3, ɣ1, α1, ɣ2, ɣ4, ε or α2), is made up of clustered exons. Each of them encodes a corresponding extracellular domain in the constant region of the final Ig heavy chain protein. Furthermore Cα, Cδ and Cɣ also have a separate exon coding for the hinge. All C exons have one or more exons coding for membrane-bound immunoglobulin. The Cμ, Cγ3, Cɣ1, Cα1, Cɣ2, Cɣ4, Cε and Cα2 exons are preceded by sequences, called switch box, allowing class switching (see. 1.4.2.5.).

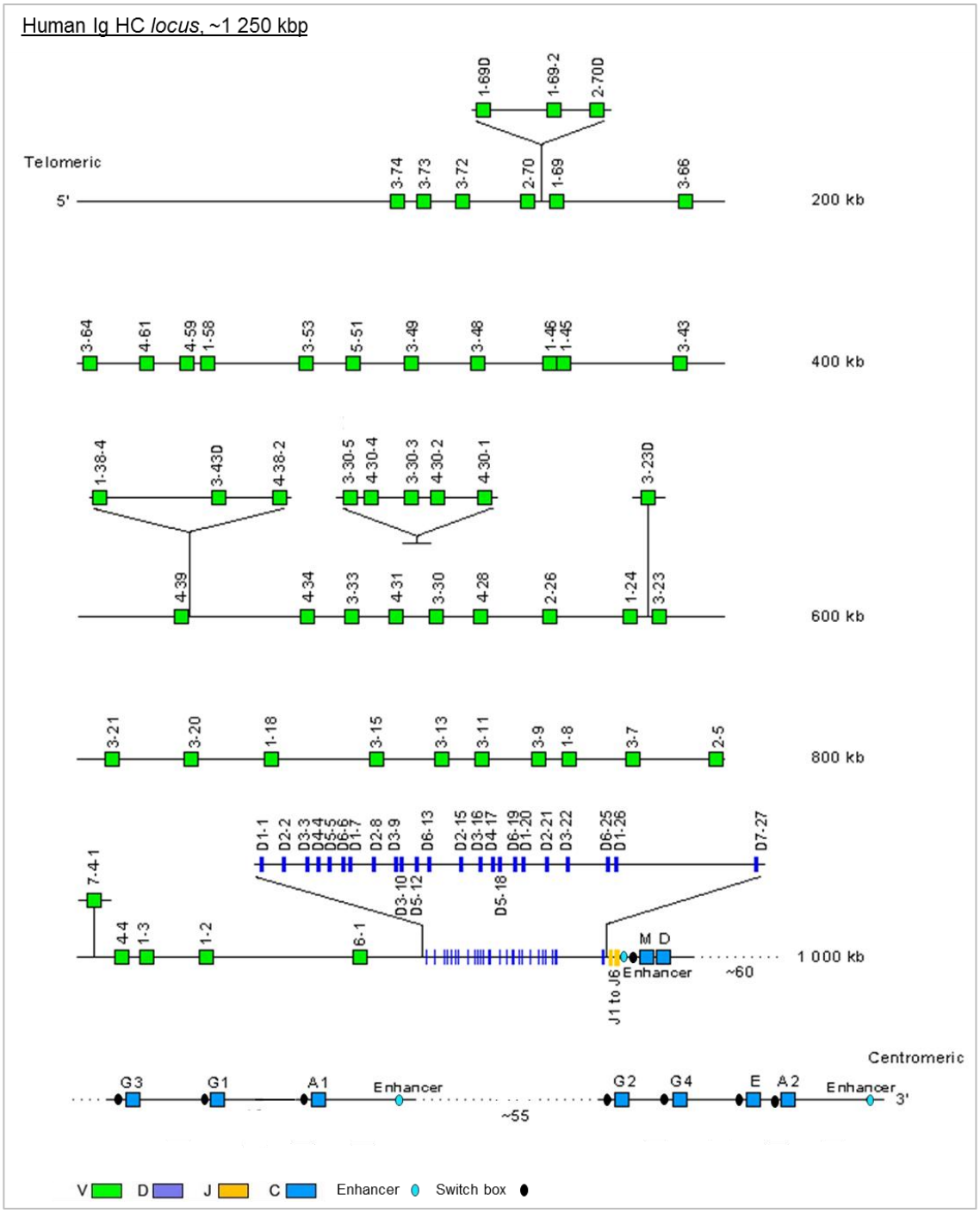

Figure 26: The human Ig HC *locus*. Constant parts annotated M, D, G3, G1, A1, G2, G4, E and A2 stand for Cμ, Cδ, Cγ3, Cɣ1, Cα1, Cɣ2, Cɣ4, Cε and Cα2. The boxes representing the gene segments are not to scale. Coding sequences for constant domains and hinge are not shown. (Picture: Lefranc, M.-P.IMGT®).

1.4.1.2. Ig LC kappa *locus*

Located on chromosome 2, on the short arm, at band 2p11.2, the LC kappa *locus* covers 1 850 kbp containing 35 V and 5 J segments [31], and one C exon. The V cluster has been duplicated; note that these V segments are encoded in reverse sense. They are fully functional and are rearranged by gene inversion.

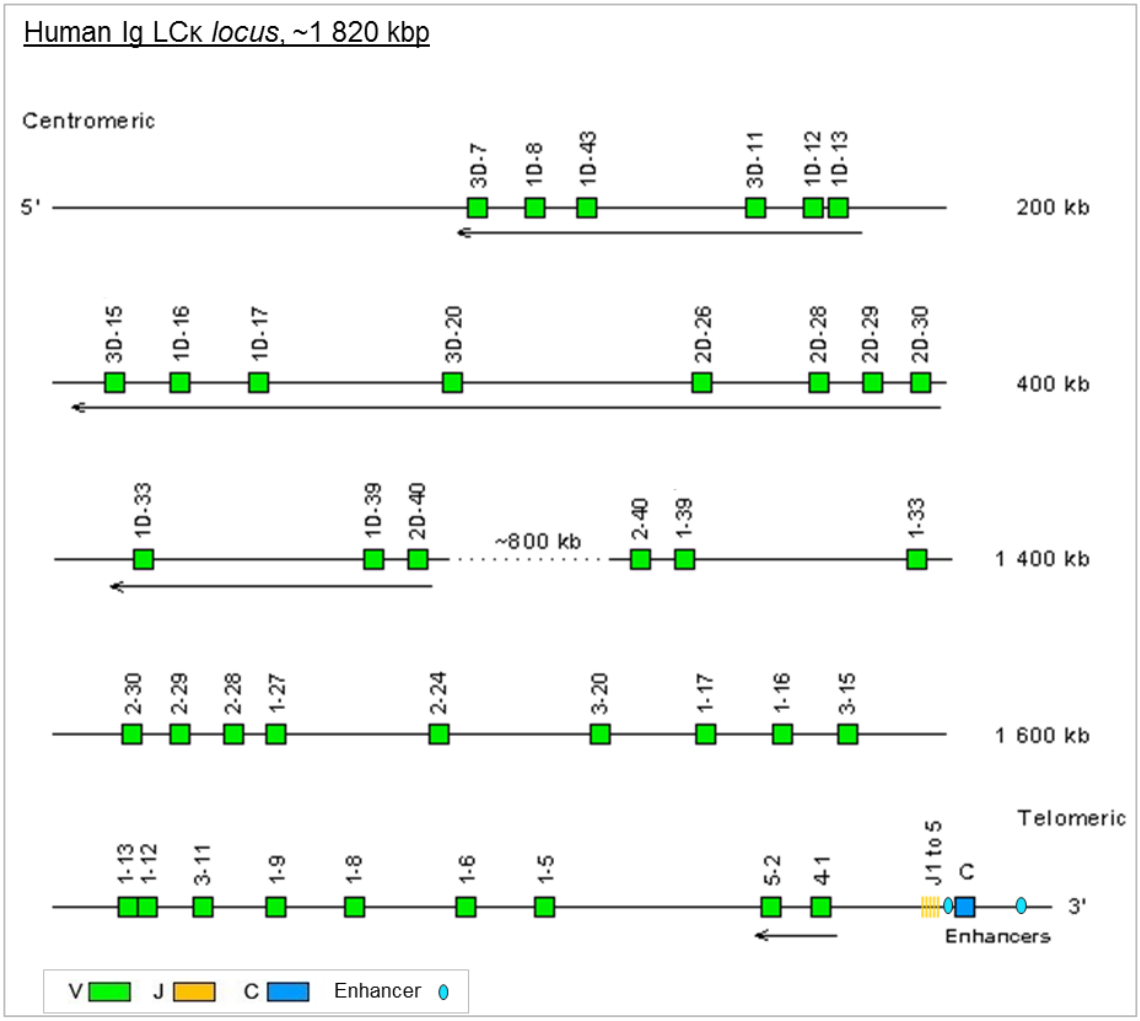

Figure 27: The human Ig LC kappa *locus*. The V segments of the distal duplicated V cluster are designated by the same numbers as the corresponding genes in the proximal V cluster, with the letter D added. Arrows show the V gene segments whose orientation is opposite to that of the J-J cluster. The boxes representing the gene segments are not to scale. (Picture: Lefranc, M.-P.IMGT®).

1.4.1.3. Ig LC lambda *locus*

Located on chromosome 22, on the long arm, at band 22q11.2, the LC lambda *locus* covers 1050 kbp and contains 32 V and 7 J - C segments.

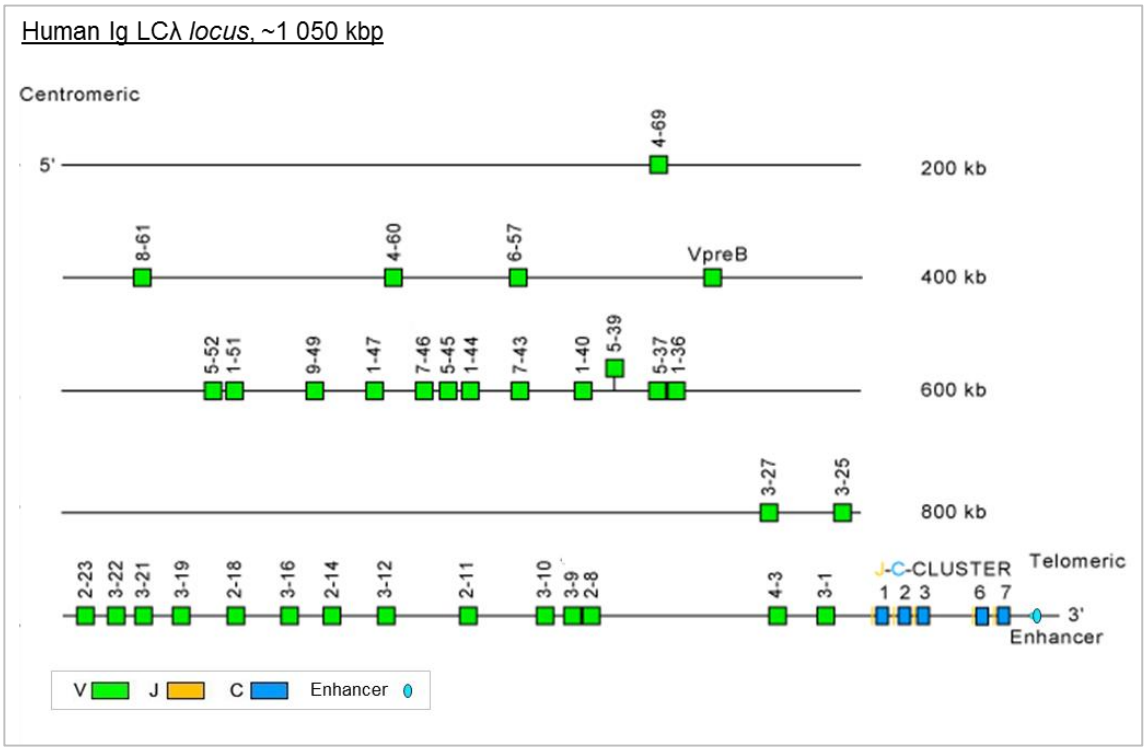

Figure 28: The human Ig LC Lambda *locus*. The boxes representing the gene segments are not to scale. (Picture: Lefranc, M.-P.IMGT®).

1.4.2. Generation of antibodies

1.4.2.1. V(D)J recombination

Functional BCR are created only in developing B cells following DNA rearrangement events that bring randomly chosen V, (D), and J gene segments into contiguity [32, 33, 34]. Rearrangement of immunoglobulin genes represents a special kind of non-homologous DNA recombination. This is mediated by the coordinated activities of several enzymes. Some enzymes are only found in developing lymphocytes while others are ubiquitous DNA double-strand break repair enzymes.

Critical lymphocytes-specific factors that mediate V(D)J recombination, recognize defined DNA sequences called recombination signal sequences (RSSs) :

- A signal sequence found in 3' of V_H, V_L and D_H gene segments consists of a heptamer (CACAGTG), followed by a spacer of non-conserved sequence (12 or 23 bp), and then a nonamer (ACAAAAACC).
- Immediately upstream (in 5') of a germline J_L, J_H and D_H segments is a corresponding signal sequence of a nonamer (GGTTTTTGT) and a heptamer (CACTGTG), again separated by an unconserved sequence (12 or 23 bp) (**Figure 29**) [1].

Heptamer and nonamer sequences given above are the most representative but it is important to note that there are variations. The 12- and 23-nucleotides spacers roughly correspond to one or two turns of a DNA helix respectively.

A specific enzyme introduces double-strand breaks in the DNA, cleaving between the heptamer of the RSS and the adjacent V, D or J coding region. This process is mediated by the protein products of the two recombination-activating genes (RAG-1 and RAG-2):

- a RAG-1-RAG-2 complex recognizes the RSS, bringing a 12-RSS and a 23-RSS together into a synaptic complex;
- the RAG proteins initiate cleavage by introducing a nick in the area between the coding region and its flanking RSS;
- the RAG proteins then convert this nick into a double-strand break;
- the hairpinned coding end must be opened before the joining step, and usually undergoes further processing, addition or deletion of nucleotides, resulting in an imprecise junction, this contributes to antibody diversity. Randomly inserted nucleotides at the junction are called N nucleotides. The addition of these nucleotides is aided by an enzyme called Terminal deoxynucleotidyl transferase (TdT) [35, 36, 37]. The two ends are joined by non-homologous-end-joining (NHEJ) factors.

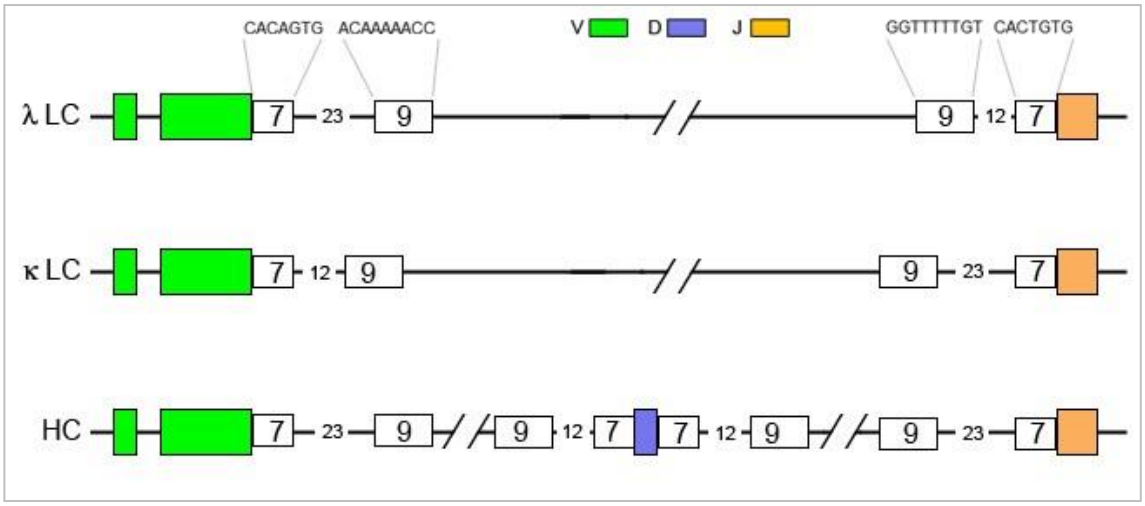

Figure 29: Heptamer – nonamer and 12 or 23 spacers. Inspired from Tonegawa, 1983.

1.4.2.2. Antibody diversity

The variable region of immunoglobulin heavy chain is encoded by three separate gene segments coding respectively from 5' to 3': variable region segments (V_H), diversity regions (D) and joining segments (J_H) genes. Within the variable domain, CDR1 and CDR2 are found in the variable (V) region of the polypeptide chain, and CDR3 is the product of the rearrangement including the D and the junctional N nucleotides. Thus, CDR3 is the most variable CDR.

Because the variable region of heavy chain is encoded in three gene segments and the joining can occur in any combinations, the diversity generated by the HC rearrangement is higher than for the light chain, which is the product of two gene segments only: V_L to J_L.

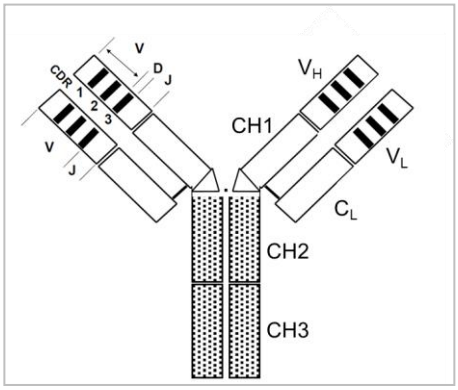

Figure 30: Twelve CDRs region per monomeric four chains molecule. Each black segment accounts for a CDR.

The antibody repertoire refers to the entire set of antibodies produced, usually in reference to a particular individual. B cells may produce antibodies that recognize more than 10^{10} different antigens. Such an impressive diversity comes from:

- use of different variable (V), diversity (D) and joining (J) gene segments in the heavy chain (recombinatory diversity);
- use of different V and J gene segments in the light chain (recombinatory diversity);
- differential pairing of heavy and light chain rearrangements (combinatory diversity);
- somatic hyper mutation (affinity maturation); and,
- addition of non-germline encoded nucleotides at junctions of gene segments (N nucleotide addition), and exonucleolytic removal of nucleotides at these junctions (junctional diversity).

1.4.2.3. Membrane *versus* secreted Ig, alternative splicing

Immunoglobulins of all heavy chain isotypes can be produced either in secreted form or as a membrane-bound receptor. Each heavy chain C exon has two exons (TM1 and TM2) that encode the transmembrane region and cytoplasmic tail of the transmembrane form, and a secretion-coding (S) sequence that encodes the carboxy-terminus of the secreted form.

The events that dictate whether a heavy chain mRNA will result in a secreted or transmembrane immunoglobulin occur during processing of the primary transcript. Studies of IgH gene and cDNA structure have provided an explanation for how a single heavy chain gene can encode both, the soluble form of the HC secreted by plasma cells and the alternative membrane-bound immunoglobulin displayed as BCR on the surface of B cells [38, 39]. Each heavy chain C gene has two potential polyadenylation sites (polyA). The sequence encoding the 20 C-terminal residues of secreted form (S) is derived from DNA contiguous with the last constant domain exon, whereas in the transmembrane mRNA, the last constant domain exon derives from two exons, TM1 and TM2, lying about 1.5 kbp further 3' (**Figure 31**).

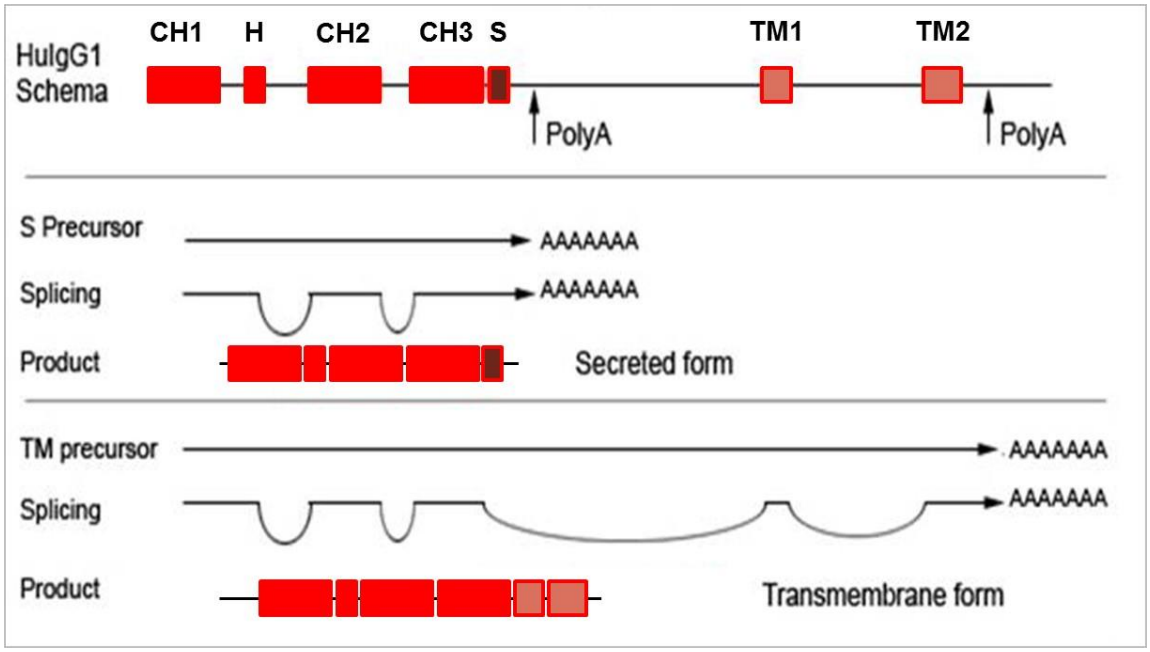

Figure 31: The principle of alternative splicing of Cɣ1. The S exon specifies a domain allowing secretion of the immunoglobulin from the B cell, while heavy chain transmembrane and cytoplasmic domains are encoded by the TM1 and TM2 exons.
In the upper panel, Cɣ1 in germline configuration is shown. In the middle panel, the primary transcript is cleaved and polyadenylated (AAA) at the first polyadenilation site (PolyA), eliminating the TM exons and giving rise to the secreted form of the heavy chain. In the lower panel, the primary transcript is cleaved and polyadenylated (AAA) at the second polyadenilation site. Splicing between a splice-site located between the Cɣ1 CH3 exon and the S sequence, and a second splice-site at the 5′ end of the TM exons, results in removal of the S sequence and joining of the TM exons to the Cɣ1 CH3 exon. This generates the transmembrane form of the heavy chain.

1.4.2.4. Affinity maturation

Affinity maturation is the result of somatic mutations of Ig genes V region followed by a selection of the B cells producing the antibodies with the highest affinities. The molecular mechanism for improving the affinity of the BCR is called somatic hypermutation (SHM) performed by the enzyme Activation Induced cytidine Deaminase (AID). This enzyme is a DNA deaminase that converts C residues to U residues at hotpots for mutation; this means in and around CDRs regions. The U residue is then replaced by T residue during DNA replication, thus generating a common type of C to T mutation. Otherwise, the U may be excised by uracil N-glycosilase, and the abasic site repaired by an error-prone repair process, thus generating all types of substitutions at

each site of AID-induced cytidine deamination. This mechanism operates in T cell dependent antibody responses during proliferative expansion of the responding B cells in a particular microenvironment, the germinal center.

1.4.2.5. Class switch recombination

Mature B cells express IgM and IgD on cell surface [23]. After B cell activation, some of the progeny of activated IgM- and IgD-expressing cells perform heavy chain isotope switching, leading to the production of antibodies with heavy chain of different classes, such as ɣ, α, and ε [33, 34].

Mature B lymphocytes are able to specifically alter their Ig isotype expression in response to extracellular stimuli with a highly regulated deletional recombination process called isotype switch recombination. Switch recombination breakpoints predominantly map to large (1-10 kbp) G-rich and highly repetitive switch regions (or switch boxes) that are located directly upstream of immunoglobulin heavy chain constant region exons. All of switch regions of downstream isotypes include pentamers similar to GAGCT and GGGGT embedded in large repeat units. DNA is nicked and broken at two selected switch regions by the activity of a series of enzymes, including AID. Thus class switch recombination allows the generation of different immunoglobulin isotypes which display different effector functions crucial for a proper defense against pathogens (see **Table 2**, page 38).

1.4.3. Molecular checkpoints of central and peripheral B cell development

1.4.3.1. Phases of B cell development and Ig gene rearrangement

The common lymphoid progenitor has the potential to differentiate into B or T lymphocytes (see 1.1.3.1.1.). **Figure 32** shows the main steps of central B cell development and the BCR evolution at each step. The pre-BCR is expressed after productive VDJ rearrangement of the heavy chain. Signaling through pre-BCR initiates LC gene recombination and induces cell proliferation. If the small pre B cells achieve VJ

rearrangement and express a functional LC, correctly associated to the pre-existing HC, they may differentiate to immature B cells.

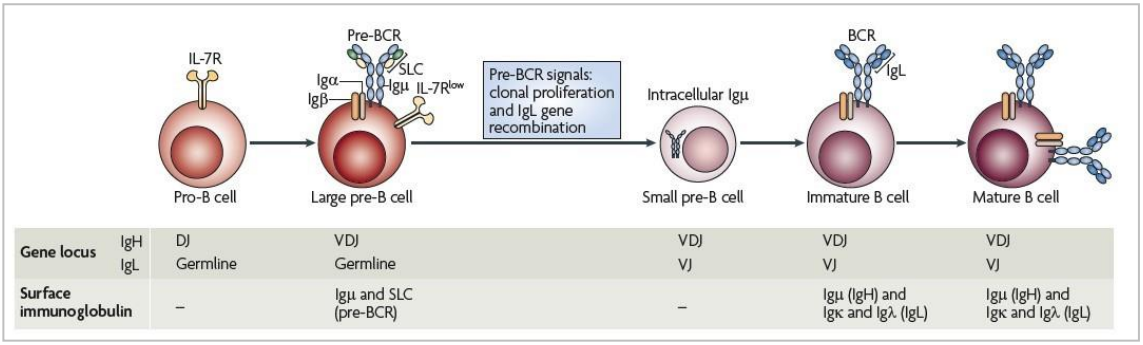

Figure 32: BCR development through B cell hematopoiesis [40].
(IL‑7R, interleukin‑7 receptor; SLC, surrogate light chain).

- From CLP to pro-B

The maturation of CLP into pro-B lymphocytes is dependent of the expression of *λ5*, *RAG1/2*, and *VpreB* genes. D to J_H recombination is initiated in early pro-B lymphocyte stage. The assembly of the pre-BCR is a key element on the way to becoming a B cell. As previously stated, expression of pre-BCR elements, Igα and Igβ, occurs early and before IgHC gene recombination is completed, and probably before its initiation. The DJ_H rearrangement may simply be the consequence of stepwise expression of each BCR component or, rather, may evidence a need for Igα/Igβ signaling at these early stages.

- From pro-B to pre-B

V_H to DJ_H recombination begins during the pro-B stage. Those progenitor B lymphocytes that successfully perform this recombination, resulting in an in-frame coding region, are then able to present IgHC (Igμ) on the cell surface, in association with Igα/Igβ heterodimer and surrogated LC (SLC) proteins (λ5 and VpreB) [41]. The generation of this pre-BCR marks the latest pro-B or the pre-B stage and signal generated through it triggers changes in gene expression. This event allows progression through the pro-B to pre-B developmental checkpoint. Signals generated through pre-BCR and IL-7R initiates multiple rounds of mitosis.

IgHC gene recombination is an “error-prone” process, and it has been estimated that only one-third of pro B cells undergoes productive IgHC recombination and mature into

pre-B cells. Pre-B cells are defined by the expression of Igμ protein in their cytoplasm and the pre-BCR on the cell surface. The scheme in **Figure 33** describes the pre-BCR associated with the co-receptor Igα/Igβ. VpreB and λ5 constitute the surrogated LC (SLC), without which the IgμHC could not be expressed on the cell surface. Indeed, it was shown that the three CH1 cysteines play a central role in immunoglobulin folding, assembly, and secretion. Unassembled IgHC are retained intracellularly by delayed folding of the CH1 and irreversible interaction of binding immunoglobulin protein (BiP) with this domain. BiP is a molecular chaperone located in the lumen of the endoplasmic reticulum (ER) [42].

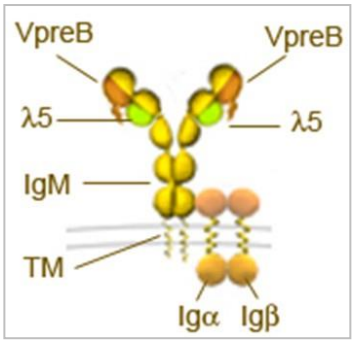

Figure 33: The pre-BCR complex. The receptor is composed of Igμ chain dimer as transmembrane form (TM), SLC elements (VpreB and λ5) and co-receptors Igα and Igβ.

The pre-BCR marks an important checkpoint in B-cell development. Successful generation of a pre-BCR complex is required for developmental progression [43]. Signaling *via* the pre-BCR provides rapid feedback about the functionality of the recombined HC gene, so only pre-B cells that express a signaling-competent receptor can further maturate. Disruption in assembly, expression, and therefore, disruption in signaling capability results in developmental arrest and cell death by neglect. This is called the positive selection.

Signaling through pre-BCR:

- induces down-regulation of λ5 gene expression,
- inhibits V_H to DJ_H recombination at the second IgH allele (allelic exclusion is an integral step of B cell differentiation and a mechanism to ensure that BCRs of a single specificity are expressed on the cell surface), and
- induces cell proliferation and IgL genes recombination.

- From pre-B to immature B cells

The pre-B to immature B lymphocytes transition is the final developmental checkpoint during bone marrow B lymphopoiesis. The transition to the small pre-B lymphocyte stage is associated with loss of proliferation, cessation of *λ5* and *VpreB* expression, reactivation of *RAG1/2* genes, and conventional IgL gene rearrangement and expression. An ordered model has been proposed for the regulation of immunoglobulin light chain, in which LC rearrangement is first performed on Vκ and Jκ segments, on both alleles and if the cell fails this recombination, Vλ and Jλ are used [44]. It has been also proposed a stochastic model where Igλ rearrangement is independent from Igκ rearrangements [45, 46], but considering the ratio between Igκ and Igλ existing in nature the order model is more suitable to explain allelic exclusion.

The fully assembled form of the BCR, composed of IgH and IgL complex associated with Igα and Igβ, is then expressed on the surface of a newly produced B lymphocyte. At this stage, immature B lymphocytes acquire the ability to recognize polymorphic antigen and become subject to negative and positive selection. Of the 2.10^7 immature B cells that develop daily in mouse bone marrow, only 10% migrate to the periphery and only a third of these progresses to the mature B lymphocytes pool. This impressive loss of B lymphocytes is due to negative selection of self-reactive B lymphocytes that directs them to apoptosis [47] (see 1.4.3.2.).

- From immature to mature B cell

Negative selection eliminates potential auto-reactive immature B cells and positive selection may promote their maturation. In mature B cells, the antigen receptor normally consists of two isotypes, membrane IgM and IgD (*m*IgM, *m*IgD).

This developmental evolution through these different stages depends also on expression of different cell surface markers, *i.e.* receptors for different cytokines which may promote evolution to the next stage (Surface phenotypic description of different mouse B cell subsets will be presented later, in the part 3.4. Transgenesis in mice).

1.4.3.2. B cell tolerance

Experimental studies have led to the concept that during B cell maturation and activation, there are multiple mechanisms repressing cells that recognize self-antigen. Central tolerance takes place in bone marrow, during antigen-independent phase of B cell development; while peripheral tolerance occurs in secondary lymphoid organs (spleen, lymph node), during the antigen-dependent phase.

Central tolerance

Immature B lymphocytes that recognize self-antigen with high affinity either change their specificity or are deleted. Change of their specificity, a process called receptor-editing, involves the reactivation of the *RAG1/2* genes and further rearrangement of LC gene [48]. Therefore, additional light chain V-J recombination occurs and leads to the production of new Ig LC, giving chance to the cell to express a different BCR that is not self-reactive anymore. Generally, receptor editing is targeted at self-reactive κ LC genes. VJ_κ exons encoding the variable domains of auto-reactive light chains are deleted and replaced by new VJ_κ exons or by λ LC rearrangements. Furthermore, editing may represent a diversification mechanism for the immune system, *i.e.* a way to create additional specificities. At the DNA level, the new VJ_κ exon may be generated by the recombination of a V segment upstream of the original V segment, and J segment located downstream of the originally rearranged J segment. Around 25% of mature B cells in mice show evidence of κ chain editing. If editing fails to eliminate auto-reactivity, the immature B cells may die by apoptosis. This negative selection is responsible for maintaining B cell tolerance to self-antigens that are present in the bone marrow. Because this process may fail, auto-reactive immature B cell may migrate to the periphery, but another process, called peripheral tolerance, will track them.

Peripheral tolerance

Mature B cells that recognize self-antigens in peripheral tissues may be rendered functionally unresponsive or die by apoptosis [49]. These auto-reactive cells are less able to migrate into lymphoid follicles than are normal naive B cells. This is due to the reduction of expression of the chemokine receptor CXCR5 that normally brings naive B

lymphocytes into the follicle. Because cells excluded from follicle do not receive necessary survival signals, they die.

1.4.3.3. Peripheral activation

Mature B lymphocytes are able to recognize protein antigens as well as microbial constituents such as bacterial polysaccharides. Antibody response to protein antigen, defined as thymus dependent responses, requires antigen specific T cell help. Upon binding to the BCR, the antigen is internalized by receptor-mediated endocytosis, digested, and complexed with MHC II molecules on the B cell surface. Activated helper T cells that recognize the peptide-MHC complex, provide to B cells the correct help, specifically *via* CD40 molecule and cytokines. B cells proliferate extensively and differentiate to plasma or memory B cells.

Notably B cells are also able to generate a response defined as thymus independent, in this case B cell is activated directly by the antigen without the requirement of T cell help; either the extensive conformation changes of B cell receptors (which occurs when a B cell binds repeated epitopes on bacterial cells) provides the signals, or the binding of antigen to a family of pattern recognition receptors, namely Toll like receptors.

1.5. Therapeutic

1.5.1. Why antibodies as therapeutic?

1.5.1.1. Biomedical

Since the first publication by Kohler and Milstein on the production of mouse monoclonal antibodies (mAbs) by hybridoma technology, mAbs have had a profound impact on medicine by providing an almost limitless source of therapeutic and diagnostic reagents. Therapeutic use of mAbs has become a major part of treatments in various diseases including transplantation, oncology, autoimmune, cardiovascular, and infectious diseases [50]. Therefore, mAbs now comprise the majority of recombinant proteins currently in the clinic, and as a group, mAbs represent one of the largest classes of drugs in development.

1.5.1.2. Safety

Many therapeutic applications for mAbs would require repeated administrations, especially for chronic diseases such as autoimmunity or cancer. Because mice are convenient for immunization and recognize most human antigens as foreign, mAbs against human targets with therapeutic potential have typically been of murine origin.

Although the first monoclonal antibody (mAb) to be approved as a human therapeutic was a mouse antibody, muromonab-CD3 (OKT3, anti-CD3), murine mAbs have inherent disadvantages as human therapeutics. First, they require more frequent dosing to maintain a therapeutic level of mAb because of a shorter circulating half-life in humans than human antibodies [51]. More critically, repeated administration of murine immunoglobulin (Ig) creates the likelihood that the human immune system will recognize the mouse protein as foreign, generating a human anti-mouse antibody (HAMA) response. At best, a HAMA response will result in a rapid clearance of the murine antibody upon repeated administration, rendering the therapeutic useless. Worst, a HAMA response can cause a severe allergic reaction [52]. These issues became driving forces for the evolution of mAbs production technology: the potential utility of therapeutic mAbs was obvious, but the initial execution was unsatisfactory. Attempts were made to increase efficacy and decrease immunogenicity through two parallel paths: 1) the production of mAb chimeras derived from both human and mouse DNA, and 2) the production of fully human mAbs.

1.5.1.3. Targets

The following table regroups the therapeutic mAbs approved or in review in the EU or US. This information dates from February 2013.

Table 3: therapeutic monoclonal antibodies.

International non-proprietary name	Trade name	Type	Indication first approved	First EU (US) approval year
Muromonab-CD3	Orthoclone Okt3	Anti-CD3; Murine IgG2a	Reversal of kidney transplant rejection	1986* (1986#)
Abciximab	Reopro	Anti-GPIIb/IIIa; Chimeric IgG1 Fab	Prevention of blood clots in angioplasty	1995* (1994)
Rituximab	MabThera, Rituxan	Anti-CD20; Chimeric IgG1	Non-Hodgkin's lymphoma	1998 (1997)
Basiliximab	Simulect	Anti-IL2R; Chimeric IgG1	Prevention of kidney transplant rejection	1998 (1998)
Daclizumab	Zenapax	Anti-IL2R; Humanized IgG1	Prevention of kidney transplant rejection	1999 (1997); #
Palivizumab	Synagis	Anti-RSV; Humanized IgG1	Prevention of respiratory syncytial virus infection	1999 (1998)
Infliximab	Remicade	Anti-TNF; Chimeric IgG1	Crohn disease	1999 (1998)
Trastuzumab	Herceptin	Anti-HER2; Humanized IgG1	Breast cancer	2000 (1998)
Gemtuzumab ozogamicin	Mylotarg	Anti-CD33; Humanized IgG4	Acute myeloid leukemia	NA (2000#)
Alemtuzumab	MabCampath, Campath-1H	Anti-CD52; Humanized IgG1	Chronic myeloid leukemia	2001 (2001)
Adalimumab	Humira	Anti-TNF; Human IgG1	Rheumatoid arthritis	2003 (2002)
Tositumomab-I131	Bexxar	Anti-CD20; Murine IgG2a	Non-Hodgkin lymphoma	NA (2003)
Efalizumab	Raptiva	Anti-CD11a; Humanized IgG1	Psoriasis	2004 (2003); #
Cetuximab	Erbitux	Anti-EGFR; Chimeric IgG1	Colorectal cancer	2004 (2004)
Ibritumomab tiuxetan	Zevalin	Anti-CD20; Murine IgG1	Non-Hodgkin's lymphoma	2004 (2002)
Omalizumab	Xolair	Anti-IgE; Humanized IgG1	Asthma	2005 (2003)
Bevacizumab	Avastin	Anti-VEGF; Humanized IgG1	Colorectal cancer	2005 (2004)
Natalizumab	Tysabri	Anti-a4 integrin; Humanized IgG4	Multiple sclerosis	2006 (2004)
Ranibizumab	Lucentis	Anti-VEGF; Humanized IgG1 Fab	Macular degeneration	2007 (2006)
Panitumumab	Vectibix	Anti-EGFR; Human IgG2	Colorectal cancer	2007 (2006)
Eculizumab	Soliris	Anti-C5; Humanized IgG2/4	Paroxysmal nocturnal hemoglobinuria	2007 (2007)
Certolizumab pegol	Cimzia	Anti-TNF; Humanized Fab, pegylated	Crohn disease	2009 (2008)
Golimumab	Simponi	Anti-TNF; Human IgG1	Rheumatoid and psoriatic arthritis, ankylosing spondylitis	2009 (2009)
Canakinumab	Ilaris	Anti-IL1b; Human IgG1	Muckle-Wells syndrome	2009 (2009)
Catumaxomab	Removab	Anti-EPCAM/CD3;Rat/mouse bispecific mAb	Malignant ascites	2009 (NA)
Ustekinumab	Stelara	Anti-IL12/23; Human IgG1	Psoriasis	2009 (2009)
Tocilizumab	RoActemra, Actemra	Anti-IL6R; Humanized IgG1	Rheumatoid arthritis	2009 (2010)
Ofatumumab	Arzerra	Anti-CD20; Human IgG1	Chronic lymphocytic leukemia	2010 (2009)
Denosumab	Prolia	Anti-RANK-L; Human IgG2	Bone Loss	2010 (2010)
Belimumab	Benlysta	Anti-BLyS; Human IgG1	Systemic lupus erythematosus	2011 (2011)
Ipilimumab	Yervoy	Anti-CTLA-4; Human IgG1	Metastatic melanoma	2011 (2011)
Brentuximab vedotin	Adcetris	Anti-CD30; Chimeric IgG1; immunoconjugate	Hodgkin lymphoma	2012 (2011)
Pertuzumab	Perjeta	Anti-HER2; humanized IgG1	Breast Cancer	2013 (2012)
Raxibacumab	(Pending)	Anti-*B. anthrasis* PA; Human IgG1	Anthrax infection	NA (2012)
Trastuzumab emtansine	Kadcyla	Anti-HER2; humanized IgG1; immunoconjugate	Breast cancer	In review (2013)
Vedolizumab	(Pending)	Anti-alpha4beta7 integrin; humanized IgG1	Ulcerative colitis, Crohn disease	In review (NA)

*Country-specific approval; approved under concertation procedure;
#Voluntarily withdrawn from market.

BLyS, B lymphocyte stimulator; **C5**, complement 5; **CD**, cluster of differentiation; **CTLA-4**, cytotoxic T lymphocyte antigen 4; **EGFR**, epidermal growth factor receptor; **EPCAM**, epithelial cell adhesion molecule; **GP**, glycoprotein; **IL**, interleukin; **NA**, not approved; **PA**, protective antigen; **RANK-L**, receptor activator of NFkb ligand; **RSV**, respiratory syncytial virus; **TNF**, tumor necrosis factor; **VEGF**, vascular endothelial growth factor.

Source: Janice M. Reichert, Editor-in-Chief, mAbs.

1.5.2. Human and Humanized monoclonal antibodies

Some of the early attempts at generating fully human mAbs used human hybridomas. However, numerous difficulties, including the instability of the human hybridomas, a preponderance of low affinity IgM mAbs, and the difficulty or ethical problems of finding humans immunized against the target antigen, resulted in this technology being used only rarely [53]. However, new approaches have been developed to interrogate human memory B cells and plasma cells with high efficiency and to isolate several broadly neutralizing antiviral antibodies against highly variable pathogens such as HIV-1 and influenza virus. These antibodies not only provide new tools for prophylaxis and therapy for viral diseases but also identify conserved epitopes that may be used to design new vaccines capable of conferring broader protection [54].

Another approach attempts to chimerized mAbs derived from mice by linking the murine variable region to human constant part. While requiring only relatively simple antibody engineering techniques, chimerized mAbs still may cause HAMA response, because of the fully murine framework and complementarity determining region (CDR) sequences of the antibody variable regions [55].

More sophisticated technologies such as humanization use tools of molecular biology and prokaryotic genetics to re-engineer antibodies to resemble their human counterparts. Humanization technology starts with a murine antibody with the desired characteristics and then changes, from a mouse-like to human-like usage, those amino acids thought to be non-critical to antigen recognition [56]. Typically, this technology requires individual tailoring for each antibody, including extensive molecular modeling and manipulation of the DNA encoding the mAb. Indeed, amino acid changes that would be predicted to have little effect on the antibody could unexpectedly abrogate antibody function. Finally, humanization leaves murine amino acids in the antibody, allowing the possibility of a HAMA response.

1.5.3. Phage or yeast display

In phage display technology, combinatorial libraries of human naive variable regions displayed on the surface of bacteriophage are used to develop antigen-specific antibodies [57]. This technology typically requires successive rounds of in vitro mutagenesis, V gene shuffling and/or panning to produce antibodies with subnanomolar affinity. The synthesis of a complete mAb from phage display technology requires to link the DNA encoding the *in vitro* affinity-matured V regions with DNA encoding constant regions in a suitable expression vector; to express the construct in a tissue culture system, and finally to confirm that the synthetic human mAb does indeed have the desired function.

A number of variations of phage display methods and alternative display technologies have been developed in an effort to overcome limitations of the phage approach. The most widely adopted of these approaches for antibody engineering to date is the yeast display. Two primary advantages distinguish yeast display from technologies based on heterologous expression of antibodies in *Escherichia coli*, including phage display: (1) post-translational processing of secreted proteins in yeast mimics that of mammalian cells and (2) multi-copy display on cell-sized particles enables quantitative library analysis and screening by flow cytometry. In addition, recovered antibodies demonstrated nanomolar affinity for more than a dozen antigens [58].

1.5.4. Human monoclonal antibodies produced in humanized non-human species

An alternative strategy for producing human mAbs is to alter the mouse system cell compartment so that it will produce fully human antibodies, obviating the need for re-engineering of the actual antibodies themselves and leaving intact the powerful natural mechanisms for class switching and affinity maturation. This strategy would require considerable effort up front to genetically engineer the mouse, but the resulting technology would be user friendly, enabling the end-user to produce high-affinity human antibodies against human antigens using only standard hybridoma procedures.

The strategy for the creation of a transgenic mouse expressing human mAbs aimed at the recapitulation of the human humoral immune system in mice. To this end two major genetic manipulations of the mouse genome are required: 1) the inactivation of genes coding for endogenous antibodies, and 2) the stable cloning and introduction of the human immunoglobulin heavy and light chain *loci* [59].

1.6. Production of humanized transgenic mice approaches

Fundamental basic research in mouse embryology and molecular biology led to the development, in the early 1980s, of a set of tools for the manipulation of the mouse genome. Genetically engineered mice were generated by direct microinjection of cloned DNA sequences into the pro-nuclei of single-cell half-day embryos. The microinjected DNA constructs, which are integrated into mouse chromosomes and are propagated through the germline, could include transcriptional regulatory sequences to direct expression to restricted differentiated cell types, including B cell expression of antibody genes [60]. This first report of an expressed immunoglobulin gene in transgenic mice involved a very small transgene. However, it has been demonstrated that a mouse can be genetically engineered by pronuclear injection of much larger transgenes [61].

The evolution of transgenic mouse technologies was first marked by the report of a transgenic mouse containing a predominantly human IgH minilocus, permitting the expression of IgM antibodies with human μ chains, but containing mouse light chains [62]. Analysis of resulting mAbs from these and other mice revealed that rearrangement and hypermutation occurred, indicating that the endogenous cell signaling machinery of mice was compatible with human immunoglobulins [63].

The invention of methods for introducing specific modifications into the mouse germ line [64, 65, 66] fueled the race to generate a mouse that comprised diverse human heavy- and light chain repertoires capable of contributing to a true immune response of high-affinity human mAbs, in the background of disrupted mouse heavy- and κ light chain genes.

In 1994, two articles, one from Lonberg *et al.* and the other from Green *et al.*, reported the generation of mice with four different germline modifications: two targeted disruptions (the endogenous mouse heavy-and κ light chain genes) and two introduced

human transgenes (encoding the heavy chain and κ light chain). Although both articles report the use of homologous recombination in mouse ES cells to engineer similar disruptions of the endogenous mouse *loci*, different technologies were used to construct and deliver the human sequence transgenes. Lonberg *et al.* used pronuclear microinjection to introduce reconstructed minilocus transgenes. In contrast, Green *et al.* used fusion of yeast protoplasts to deliver yeast artificial chromosome (YAC)-based minilocus transgenes. Both constructs underwent V(D)J joining and expressed both IgM and IgD. Furthermore, fully human monoclonal IgM [67] and IgG [2] mAbs recognizing the target antigens against which the mice had been immunized were produced.

Indeed, few years later, Mendez *et al.* generated transgenic mice having nearly complete heavy chain V repertoires and approximately half the κ light chain V repertoire, and compared them with the minilocus mice of Green et al, mentioned above [68]. This study, and a later analysis of the same mouse strains by Green and Jakobovits [69], showed that V-region repertoire size had a profound effect on multiple checkpoints in B cell development, with larger repertoires capable of restoring B cell compartments to near normal levels. Despite the fact that human immunoglobulin transgenic mice express B cell receptors that are essentially hybrids of mouse and human components (e.g., human immunoglobulin, mouse Igα and Igβ), their B cells develop and mature normally. The immunoglobulin transgenes undergo V(D)J joining, random nucleotide (N-region) addition, class switching, and somatic mutation to generate high-affinity mAbs to a variety of different antigens. The process of affinity maturation in these animals even recapitulates the normal pattern of somatic mutation hotspots observed in authentic human secondary repertoire antibodies [70]. Following these reports of transgenic mice comprising either a relatively limited or a complete germline repertoires, a variety of additional mouse strains created appeared in the literature [71]. Most of these mice strain were engineered through pronuclear microinjection and yeast protoplast fusion with ES cell.

2. Aim of the project

My PhD project has been developed in the context of the generation of transgenic mice that produce human antibodies. This procedure requires four genetic manipulations of the genome, in two phases: i) knockout of mouse HC and LC genes and ii) introduction of the human Ig HC and Ig LC *loci* in the mouse genome.

Thus the aims of my project have been:

1) the generation of transgenic vectors in order to transfer the human Ig genes into the mouse genome. To this end, two approaches were followed: i) the main approach consisted of the reconstitution of the almost entire human Ig *loci* coding for the Ig HC and LC, in the form of yeast artificial chromosomes (YACs) and ii) a second and alternative approach aimed to engineer several size-reduced genetic constructs coding for human Ig HC;

2) the evaluation of functionality of the Ig HC constructs, developed in the first part of the project, *in vitro* (cell line), as well as *in vivo* (transgenic mice).

3. Methods and results

3.1. Main approach: human Ig *loci* cloning as YAC

The aim of the main approach is the building of three genetic vectors corresponding to the three integral human Ig *loci* (IgH, Igκ and Igλ), in order to generate, with each of them, transgenic mice. Located on separated chromosomes (IgH on chromosome 14, Igκ on chromosome 2 and Igλ on chromosome 22), each human Ig *loci* cover between 1. 10^6 and 1.8 10^6 base paire (bp) into the human genome [72].

The isolation by the use of yeast artificial chromosome (YAC) of large human genomic regions, covering almost the entire antibody *loci*, has been previously described [73]. YACs are genetically engineered DNA molecule used to clone DNA sequences in yeast *Saccharomyces cerevisiae*. YACs are often used in connection with the mapping and sequencing of genomes. Segments of exogenous genomic DNA, from 100 to 1 000 kbp in length, can be inserted into YAC. The YACs, with their inserted DNA, are then taken up by yeast cells. As the yeast cells grow and divide, they amplify the YAC DNA, which can be isolated and used for DNA mapping and sequencing [74]. YACs were initially used for the Human Genome Project, but due to stability issues, YACs were abandoned for the use of bacterial artificial chromosomes (BAC), an engineered DNA molecule used to clone genomic DNA sequences in bacterial cells (*E. coli*). BACs are often related to DNA sequencing as they were used for the isolation of large genomic regions in order to provide the intermediate substrate for the international genome sequencing effort [75].

BAC inserts are smaller in size compared to YAC (100-200 kbp) and thus the reconstitution of the entire antibody *loci* needed several fusions: i) 4 individual BACs to cover the almost entire human HC *locus*, ii) 2 individual BACs for the human LCκ *locus* and iii) 3 individual BACs for the human LCλ *locus*.

In situ re-assembly of 2 individual BACs in one single piece of genomic DNA has been previously described but with expected limitation in the size of the final product due to the recombination into *E. coli* [76].

However, a recent described cloning technique [3] allows the isolation of large chromosomal regions or genes directly from genomic DNA without prior construction of a genomic library. This technique involves homologous recombination during

spheroplast transformation between the genomic DNA of interest and a TAR vector. Recombination leads to circular Yeast Artificial Chromosome (YAC) formation. In addition circular YACs are described as much more stable as the original linear YACs [77]. In order to build up vectors that should be transferred into transgenic mouse, we decided to use this yeast recombination technology to link our BACs together in one single piece of DNA.

BAC n° (RP11-) 702F21, 72N10, 259B19 and 448N5 were linked together to reassemble the human HC *locus*. BAC n° (RP11-)685C7 and 946A16 were linked together to reconstitute the human LCκ *locus*, and BAC n° (RP11-)22L18, 761L13 and 2079L4 were used for the generation of the human LCλ *locus*. By linking these mentioned BACs together in one single piece of DNA for each human antibody *locus*, we reconstituted *loci* in size range of 660 kbp for HC *locus*, 398 kbp for κ *locus* and 360 kbp for λ *locus*.

3.1.1. Genetic material

Coming from such Bacterial Artificial Chromosome library, three BAC collections covering almost the entire human HC, LCκ and LCλ *loci* were obtained from Invitrogen. As an example, the entire collection of BACs that covers the human HC *locus* is shown in **Figure 34**. The BAC collection covering the human LCκ *loci* and the human LCλ *loci* present the same organization (not shown). Each individual BAC contains insert of 150-200 kbp of human genomic DNA cloned into the vector pBACe3.6.

pBACe3.6 (*ref*: https://bacpac.chori.org/pbace36.htm) contains the chloramphenicol resistance gene for bacterial selection. BAC transformed *E. Coli* bacteria were propagated in Lysogeny Broth (LB) medium containing 30 μg/mL of chloramphenicol. BACs were purified from 100 mL of bacterial culture, using QIAGEN Large-Construct Kit (*ref*: 12462) for alkaline cell lysis, followed by phenol/ chloroform purification and DNA precipitation using sodium acetate/ethanol mixture. From purified BAC DNA, the human genomic insert was separated from the pBACe3.6 vector by *Not*I digestion before performing TAR cloning.

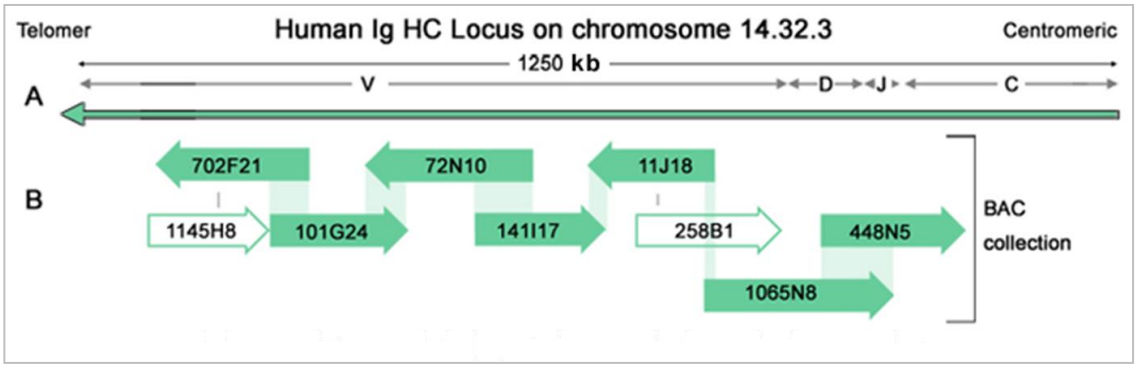

Figure 34: The BAC collection covering human HC *loc*us. Each arrow stands for the genomic insert contained in one BAC (B). Each BAC has its own reference (e.g. 702F21, 72N10). The direction of arrows indicates in which direction the genomic inserts were cloned in the vector pBACe3.6. These 9 BACs cover almost the entire human HC *locus*, which is represented in the upper panel (A). Note that each BAC overlaps with neighboring BACs, therefore green arrows are sufficient to reconstitute human HC *locus*, while white arrows are "optional".

3.1.2. Introduction to TAR cloning

Transforming Associated Recombination (TAR) cloning was first pioneered by Botstein and colleagues (1987) [78] and has since been further developed by Larionov and coworkers (1994) [79]. Its principle is based on homologous recombination occurring at a high rate in the yeast *S. cerevisiae* after the yeast spheroplast have been co-transformed with genomic DNA of interest and a TAR shuttle vector containing two specific target sequences termed "hooks".

The key feature of TAR cloning is the use of a specific TAR shuttle vector (**Figure 35**), which contains the following elements: a) a bacterial origin of replication (*ORI*), b) a selection bacterial marker like the ampicillin resistance gene (Amp^R), c) a polylinker with white/blue selection (*β-Gal*), d) a yeast centromere (*CEN*), e) the gene encoding histidine (*His*) for selection in yeast and f) optionally, the neomycin resistance gene (Neo^R) for selection in mammalian cells. A TAR vector contains also two gene-specific targeting sequences ("hooks"), which are designed to be homologous to the 5' and the 3' extremities of the genomic region to clone. Each targeting sequence is a unique specific sequence of minimum ~60 bp up to 1.5 kbp [80].

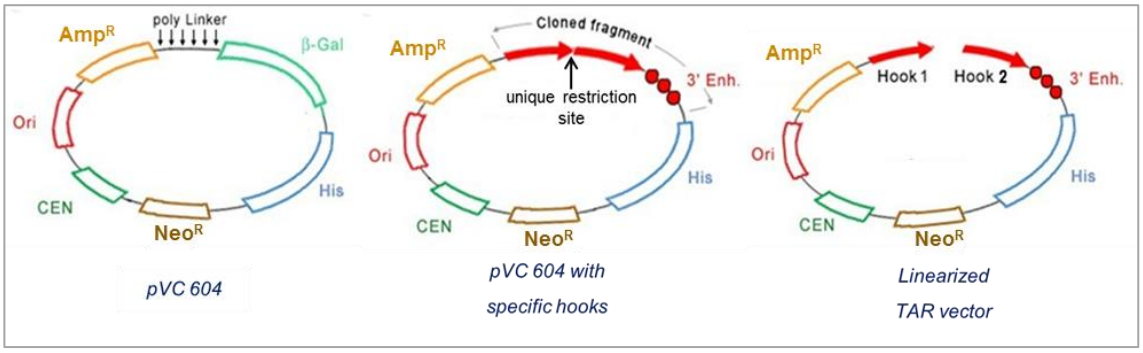

Figure 35: Schematic representation of the pVC604 derived TAR vector. TAR vector is based on pVC604 (*left panel*), which contains all genes described in the paragraph above. For each TAR cloning, two specific hooks are cloned in the polylinker (*middle panel*). In the scheme, the 3' enhancer (3' Enh) is added as an additional element. Before TAR proceeding, the vector is linearized using the unique restriction site placed between both hooks (*right panel*), this makes them able to recombine with their homologous sequences.

Once the TAR shuttle vector and the genomic DNA to clone are transformed into yeast spheroplasts, recombination occurs between the genomic DNA and the linearized TAR vector through the targeting sequences (hooks) and the corresponding homologous DNA sequences along the genomic region. This homologous recombination between the hooks of the TAR vector and the corresponding homologous genomic sequences results in the generation of a stable circular YAC that is able to segregate and be selected in yeast. The yield of gene-positive clones varies from 1% to 5%. The size of the cloned sequence depends on the position of the targeted sequences in the genome and can be, as descripted by Kouprina and Larianov, up to 600 kbp [3]. The basic principle of the TAR cloning procedure is explained in **Figure 36**.

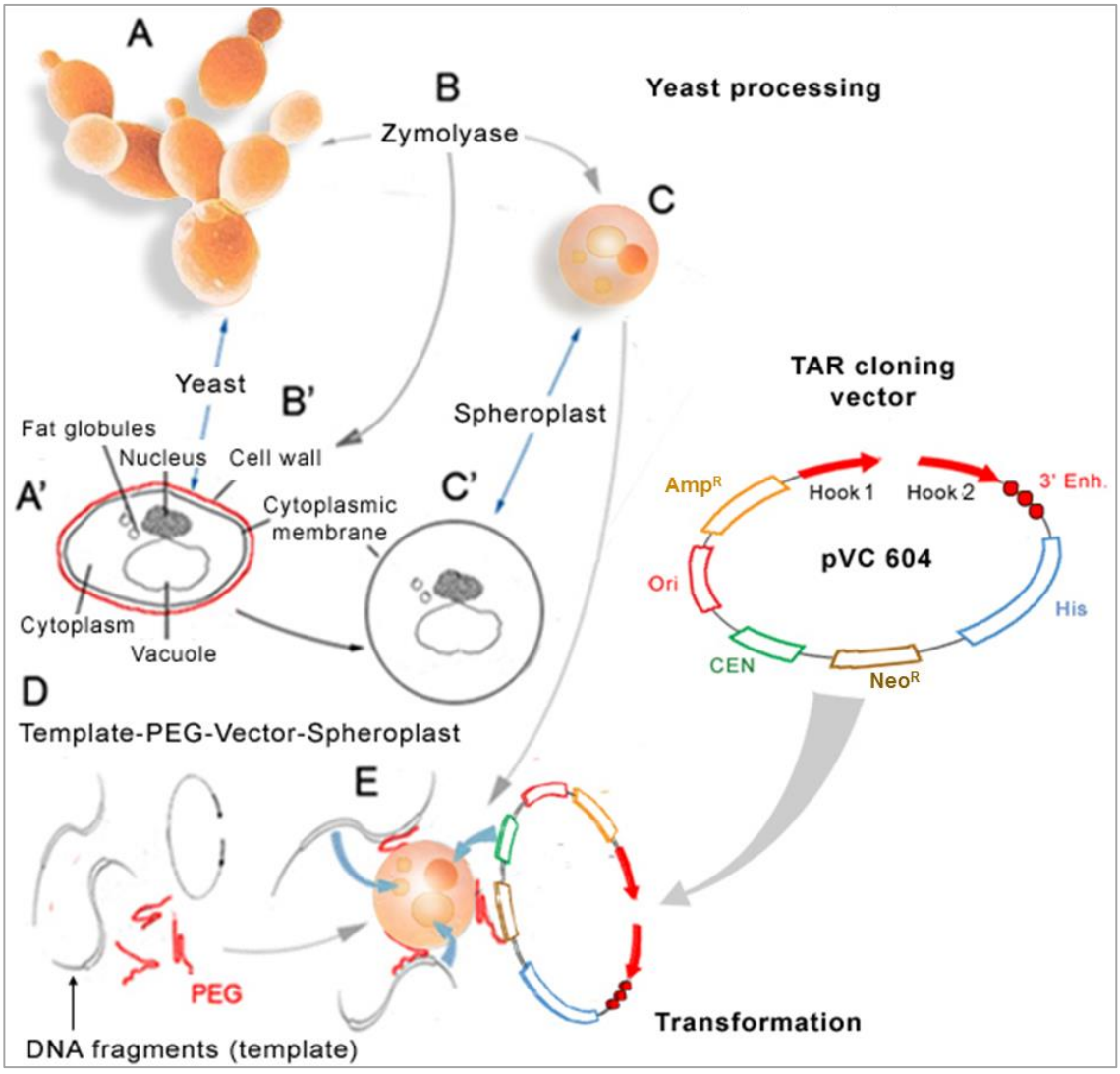

Figure 36: The TAR cloning procedure. Genomic DNA (*D*) is mixed with an excess of linearized TAR cloning vector and is presented to yeast *Saccharomyces cerevisiae* (*A, A'*). Yeast cells are previously treated with zymolyase (*B*), which removes their cell wall. At this point, yeasts appear spherical, and are called spheroplast (*C, C'*). Once mixed with the linearized vector, template DNA, polyethylene glycol (PEG), the spheroplasts can absorb the template DNA and the vector (*E*). Using its normal DNA duplication and repair machinery, the yeasts proceed the recombination between DNA template and vector, leading to YAC formation.

Regarding TAR recombination, **Figure 37** illustrates the molecular mechanisms involved in TAR cloning. Using DNA duplication and DNA repair enzymes; yeast recombines together the genomic DNA and the TAR vector, through homologous DNA sequences between hooks sequences and genomic region of interest. However, TAR

cloning was developed for the selective isolation of one continuous DNA region from complex genome. In order to re-assemble several BACs, that contain antibodies genomic regions, in one single piece, we adapted the previously described TAR method to our purpose. Thus, we implemented the "TAR bridging" method, described in the paragraph below.

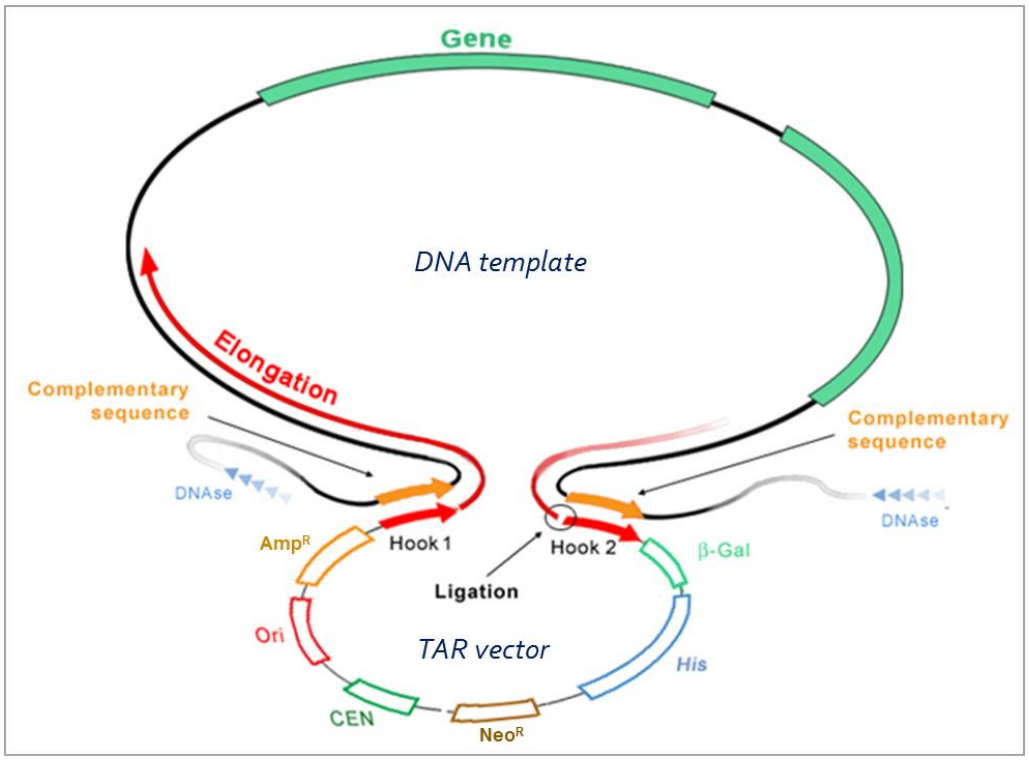

Figure 37: Putative recombination principle in TAR cloning. On this picture, the genomic region to clone is represented with both green segments (Gene). The gene-specific targeting sequences (*red arrows*, Hooks 1 and 2) at both ends of the linear TAR vector are homologous to the sequences (*orange arrows*) on either side of the region to clone. DNA ends are digested by DNAse (*blue*) while DNA polymerase copy DNA from the template (Elongation), using the 5' hook (Hook 1) as a primer. On the other side, the Hook 2-complementary sequence is elongated using the vector as template. Finally, ligase covalently links the junctions (Ligation). Once the process is finished, the artificial chromosome will stay stably in the yeast and segregated at each cell division. The yeast can be selected for the presence of histidine expressed by the gene *His* from the vector and which is absent of the culture medium.

3.1.3. Adaptation of TAR cloning method: TAR bridging strategy

The "TAR bridging" method, developed in our laboratory, enables linking two or more fragments of genomic DNA (BAC inserts) into a single circular YAC; and therefore, re-assembling the three antibody *loci* into three different circular YACs. The basic strategy is the same as for the described TAR cloning, except that an additional specifically designed DNA fragment, termed "bridge", was added to the previously described TAR vector, allowing the "bringing" of two distinct pieces of genomic DNA. The bridge is composed of two linked hook sequences, between which, any sequence may be inserted, such as selectable marker or a useful regulatory element absent from the selected genomic fragments. **Figure 38** illustrates the putative mechanism involved.

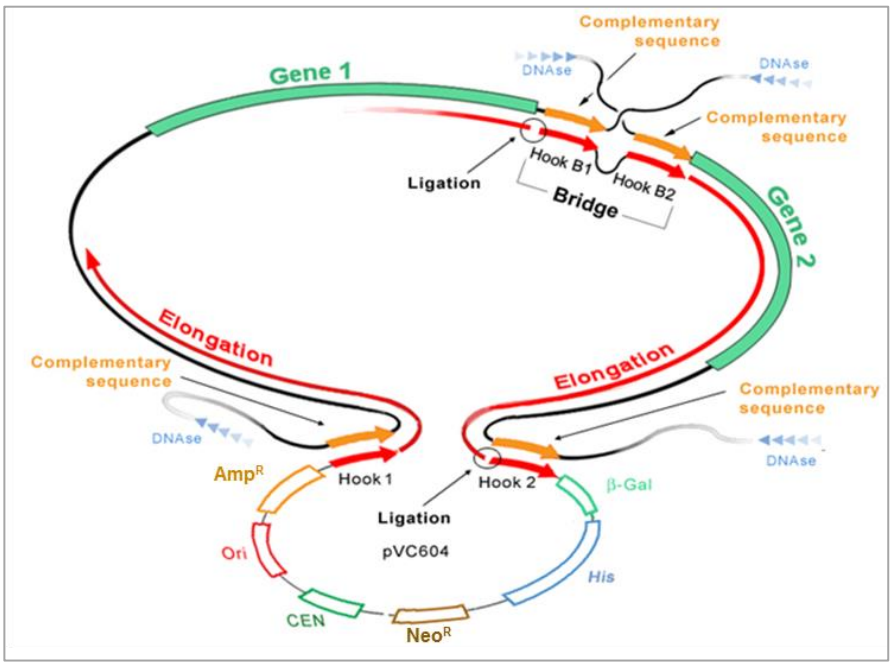

Figure 38: The TAR bridging strategy. On this picture, both independent genomic regions to clone are schematized with green segments (Gene 1 and Gene 2). Originally the TAR vector contains four hooks, Hook 1, Hook 2 and both hooks composing the bridge (*i.e.* Hook B1, Hook B2). After specific digestions between Hook 1 and Hook B1, and between Hook B2 and Hook 2, each hook sequence (*red arrows*) can recombine with its homologous sequence (*orange arrows*). The yeast DNA duplication and DNA repair enzymes undertake recombination and lead to stable YAC formation.

The BAC inserts containing the human genomic antibody gene segments were isolated from the BAC vector by *Not*I digestion. The TAR shuttle vectors were appropriately

digested for each experiment. The generation of the three circular YACs is described in detail below. Regarding the other steps of the procedure, the method of TAR cloning described by Kouprina and Larionov [3] was strictly applied. The yeast strain VL6-48 used in each TAR experiments was provided by V. Larionov. However, in the step of spheroplast transformation, we used 3 ug of the corresponding BAC DNA instead of genomic DNA, along with 1 μg of the specific linearized TAR vector.

3.1.4. Human Ig *loci* cloning

3.1.4.1. Human Ig LC kappa cloning

In order to reconstitute a functional human LCκ *locus*, two BACs (n° 685C7 and n° 946A16) were selected as genomic DNA source for TAR cloning. The almost entire 5' V cluster of the native human LCκ *locus* is the result of a recent inverted duplication which contains V segments that are almost identical to the C proximal one [31, 81]. Therefore, the 5' V set was ignored. By linking together the two selected BACs, the reconstituted human LCκ *locus* will contain 15 out of 21 V segments, the entire J cluster and Cκ, as well as both enhancer elements. The TAR shuttle vector backbone, pVC604 (**Figure 39**) was obtained from V Larionov and modified by the insertion of two home-designed hook sequences previously synthetized by GeneArt® Gene Synthesis, as well as the neomycin resistance gene (*neo*) for selection in mammalian cells. The resulting Kappa TAR vector (**Figure 40**) was transformed in *Escherichia coli* and transformed bacteria were cultivated in LB medium containing 100 μg/mL of ampicillin. The purpose of this bacterial transformation is to use bacteria DNA replication machinery to amplify the plasmid in order to make large quantities of it. Vector DNA was purified from bacterial culture using QIAprep Spin Miniprep Kit (*ref*: Qiagen, 27104).

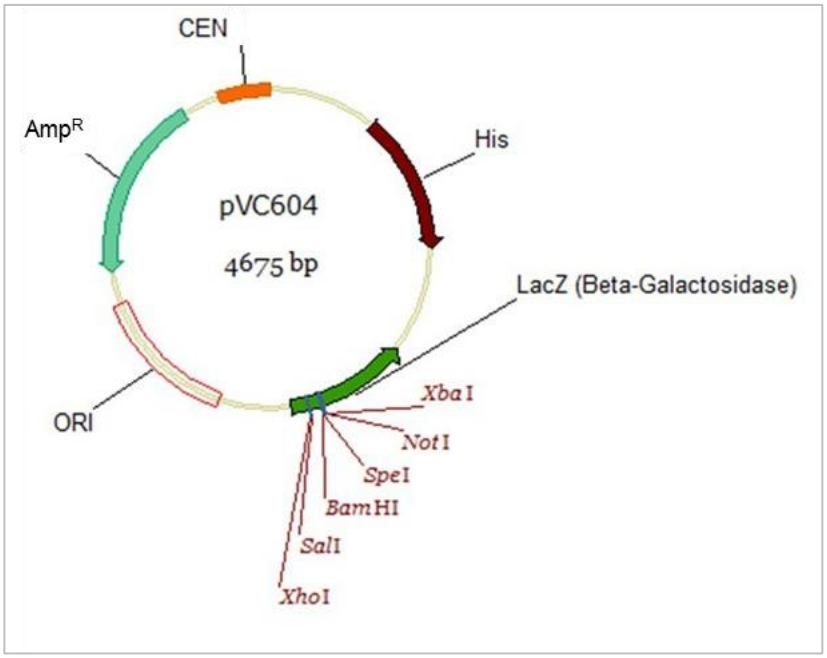

Figure 39: pVC604 as TAR vector backbone, 4 675 bp. *ORI*: bacterial origin of replication; *Amp*^{R}: ampicillin resistance gene as selection bacterial marker; *LacZ*: gene coding for β-galactosidase, containing the polylinker (*Xho*I, *Sal*I, *Bam*HI, *Spe*I, *Not*I and *Xba*I); *CEN*: a yeast centromere; *His*: gene encoding histidine for selection in yeast.

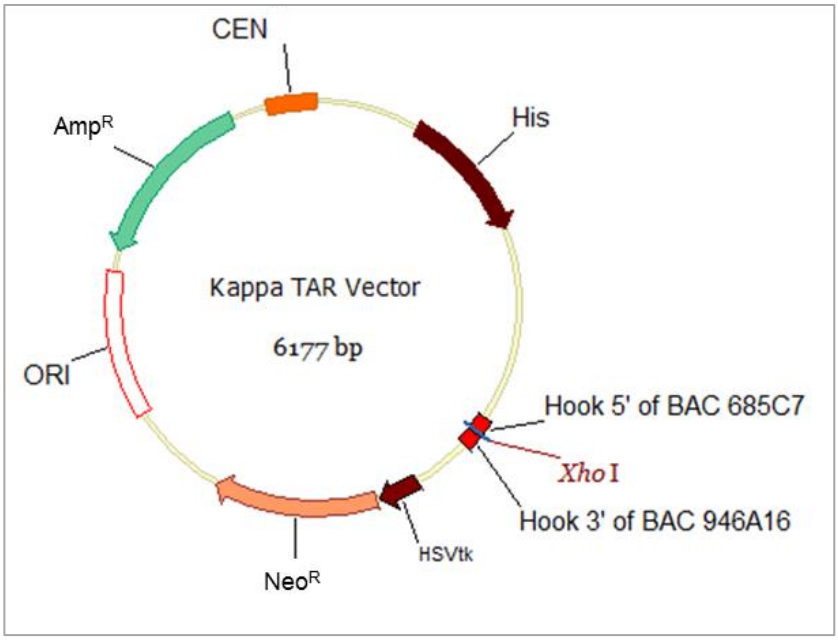

Figure 40: The Kappa TAR vector, 6 177 bp. To the vector pVC604 were added the neomycin resistance gene (*Neo*^{R}) under HSVtk promoter and two hook sequences (Hook 5' of BAC 685C7 and Hook 3' of BAC 946A16) of 75 bp each. *Xho*I is the unique restriction site placed between both hooks.

Regarding genomic template, both BAC inserts were isolate and ligated together before proceeding with TAR (**Figure 41**). In brief, BAC n° 685C7 and BAC n° 946A19 were digested with *Not*I, 2 hours at 37 °C before being ligated together with T4 DNA Ligase (ref.: NEB #M0202S). Ligation products were precipitated using sodium acetate/ethanol mixture and washed with 70 % ethanol before dilution in Tris-EDTA (TE) Buffer. Finally, 3 µg of purified ligation product and 1 µg of *Xho*I linearized Kappa TAR vector were transformed into yeast spheroplasts.

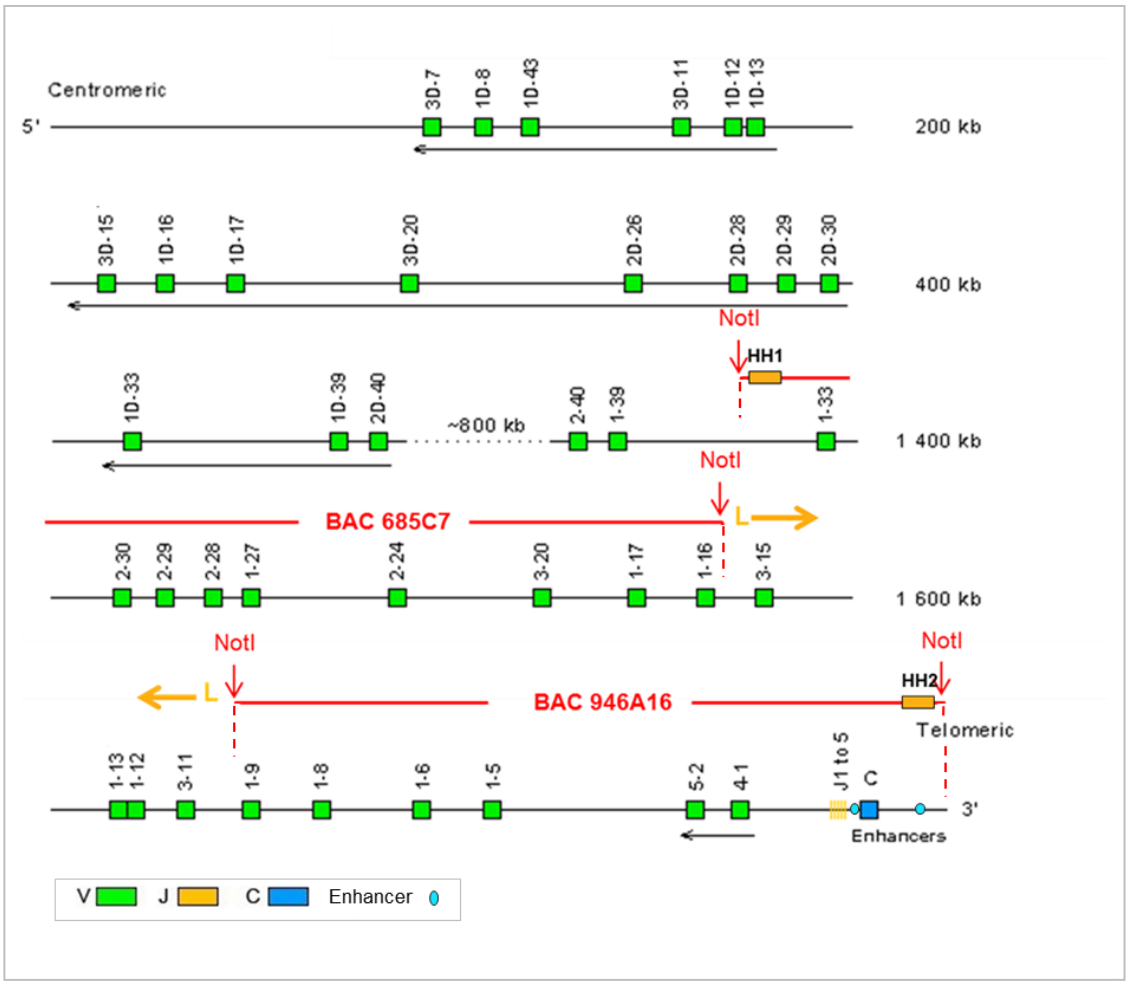

Figure 41: Schematic view of the human Ig LCκ *locus* (1 820 kbp) indicating both BAC inserts used to reconstitute the *locus* (*red lines*). The almost entire 5' V cluster (e.g. 2D.40, 1D.39) is the result of a recent inverted duplication. Selected genomic regions are linked by ligation (L) and include all genes except six V segments which are missing compared to the native Ig LCκ *locus*. The orange boxes show the position of the hook homology: HH1 stands for Hook 5' of BAC 685C7 and HH2 stands for Hook 3' of BAC 946A16. (Picture: Lefranc,M.-P.IMGT®).

Following the protocol of Kouprina and Larionov [3] transformed yeast spheroplasts were plated on solid selection medium exempted from histidine amino acid, for 1 week at 30 °C. Transformation was performed in triplicate and after incubation; around 50 to 100 yeast clone became visible. Clones were picked and transferred into liquid culture for PCR analysis.

A first PCR screening was directly performed on a small cell pellet, as a primary selection. Negative yeast clones were eliminated and 12 clones were selected to be further analyzed. Because PCR analysis directly performed on cell pellet is not optimal, we purified total DNA from the 12 positive yeast clones, using kit Genomic-tip 20 of Qiagen (#10223). Total DNA was used as template for intensive genotyping by PCR. After completed the PCR run, samples were loaded on 1% agarose gel electrophoresis containing 0.01% of RedSafe DNA staining solution (ref: 21141), which allows DNA detection after electrophoresis run. PCR products were then extracted from agarose gel, using QIAquick Gel Extraction Kit (ref: Qiagen, 28704). Purified PCR products were then sequenced and those with more than 99% of the expected homology were considered as positive "hit". The PCR screening of each yeast clone derived total DNA was performed with up to 50 primer pairs, distributed along the expected cloned sequence. The goal was to have maximum positive hits to insure the integrity of the cloned sequence. The clone YAC LC Kappa #8 showed the highest number of PCR positive hits and was therefore selected. The final construct YAC LC Kappa and all positive PCR Hits are represented in **Figure 42**.

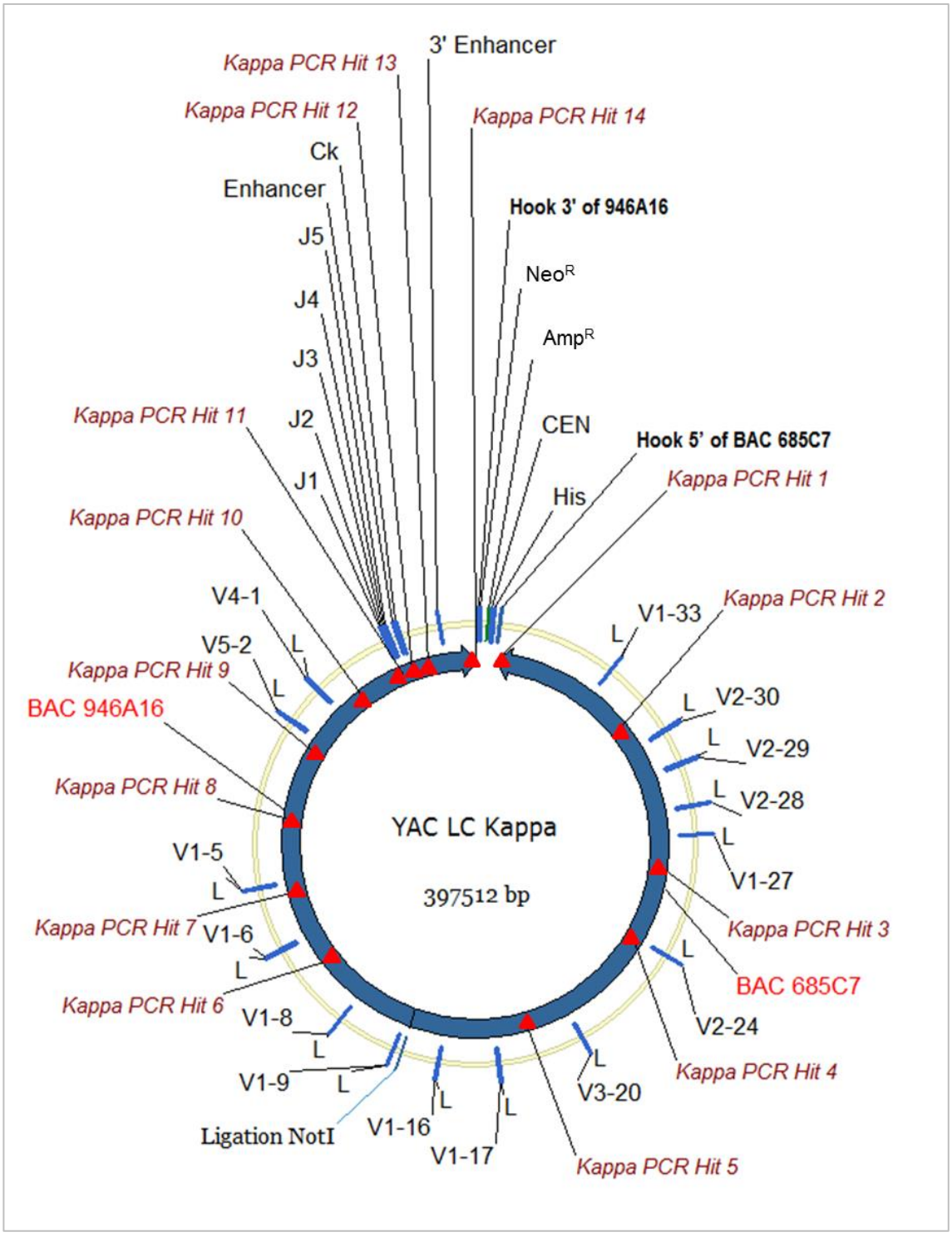

Figure 42: The YAC LC Kappa #8, 397 512 bp. YAC LC Kappa comprises BAC n° 685C7 and 946A16. All V segments with their corresponding leader sequence (L) are given, as well as the five J segments and the constant exon Cκ. Genetic elements as *Neo*^{R}, *Amp*^{R}, *CEN* and *His* are pVC604 original components. All positive PCR Hits are represented by red triangles.

3.1.4.2. Human Ig LC lambda cloning

In order to reassemble a functional Ig lambda *locus*, 3 BACs (n° 22L18, n° 761L13 and n° 2079L4) have been linked in one single piece of DNA. The three BAC inserts cover genomic regions including 24 functional V genes out of 32, and all Jλ C genes (**Figure 43**). However, the 3' enhancer is not covered by the BAC 2079L4. To ensure proper expression of the human antibody genes, we first inserted a synthetic Kappa light chain 3' enhancer into the TAR vector (**Figure 44**). We designed a first bridge to physically combine BAC n° 22L18 with BAC n° 761L13 and a second bridge to link BAC n° 761L13 and BAC n° 2079I4. Finally, two hooks, homologous to the 5' and in 3' of the three BAC assembly, flanked both bridges in the final Lambda TAR vector. Moreover, appropriates restrictions sites were added in order to linearize the TAR vector and to generate "free" bridges before transformation into yeast spheroplast.

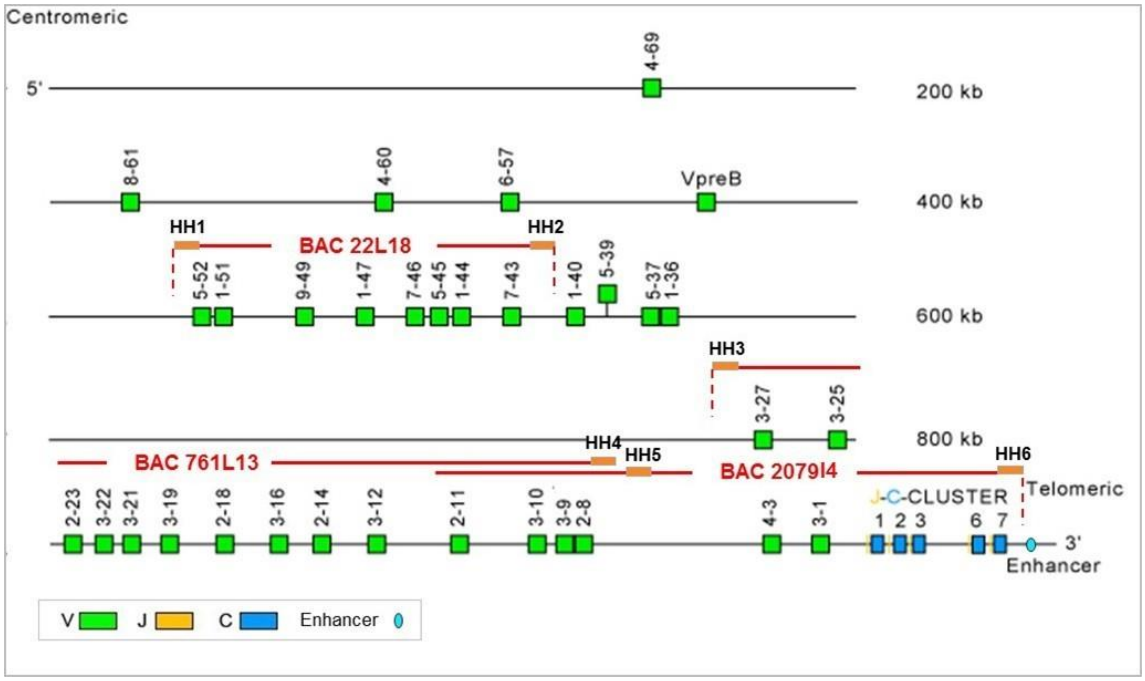

Figure 43: Schematic view of the human Ig LCλ *locus* (1 050 kbp) indicating the three BACs (*red lines*) used to generate the YAC LCλ. The orange boxes show the position of the hook homology (HH). Hooks 1 and 6 are part of the vector; hooks 2 and 3 form the first bridge, while hooks 4 and 5 belong to the second bridge (Picture: Lefranc, M.-P. IMGT®).

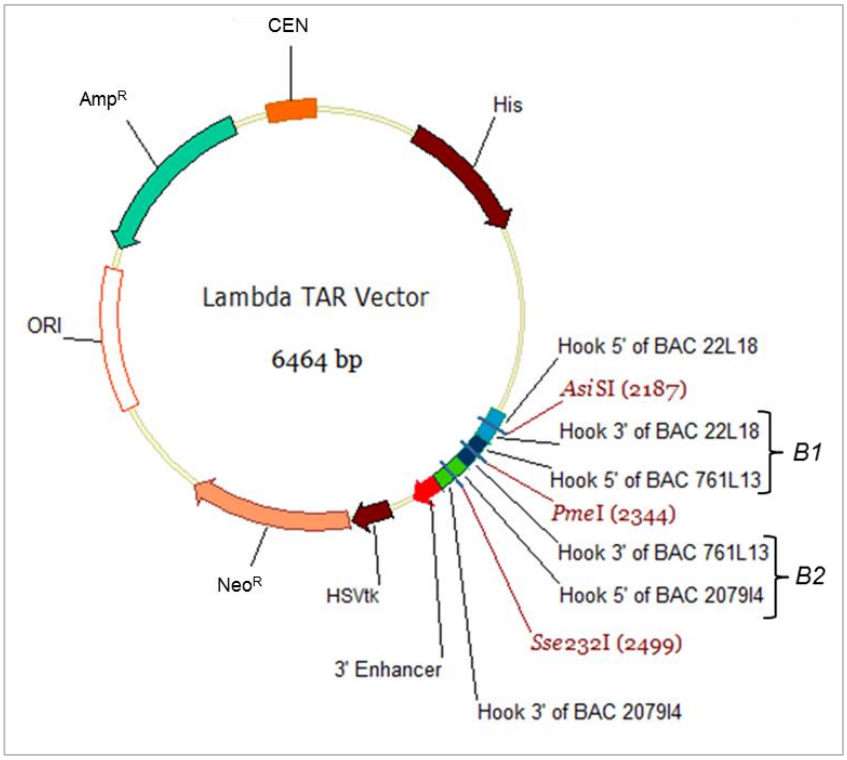

Figure 44: The Lambda TAR vector, 6 464 bp. To the vector pVC604 were added the neomycin resistance gene (*Neo*^{R}) under *HSVtk* promoter, the Kappa LC 3' Enhancer and six hook sequences with appropriate restriction sites. *Asi*SI cuts uniquely between both hooks homologous to BAC 22L18 extremities, *Pme*I cuts uniquely between both hooks designed to recombined with BAC 761L13 extremities and, *Sse*232I cuts uniquely between hooks matching on both side of selected BAC 2079I4 region. Hook 3' of BAC 22L18 and Hook 5' of BAC 761L13 compose the first bridge (B1); and Hook 3' of BAC 761L13 and Hook 5' of BAC 2079I4 compose the second bridge (B2). Each hook sequence accounts for around 75 bp.

BACs n° 22L18, 761L13 and 2079I4 were digested *Not*I, 2 hours at 37 °C. Digestion products were precipitated using sodium acetate/ethanol mixture and washed with 70 % ethanol before being resuspended in Tris-EDTA (TE) Buffer. Lambda TAR vector was digested with *Asi*SI, *Pme*I and *Sse*232I, in order to open the TAR vector and to "free" bridge sequences. For yeast spheroplast transformation, 1 µg of TAR Vector, 3 µg of the three *Not*I digested and purified BAC DNA were mixed together. Transformation was performed in triplicate. As negative control, we also transformed separately, 3 µg of purified BAC DNA template, and 1 µg of Lambda TAR vector. After 1 week of selection at 30 °C, 200 yeast clones in total developed and no yeast colony developed from negative controls.

As for the characterization of the YAC LC Kappa, after several round of PCR analysis, the clone YAC LC Lambda #2 clone showed the highest number of PCR positive hits and was therefore selected. The final construct YAC LC Lambda and all positive PCR Hits are represented in **Figure 45**.

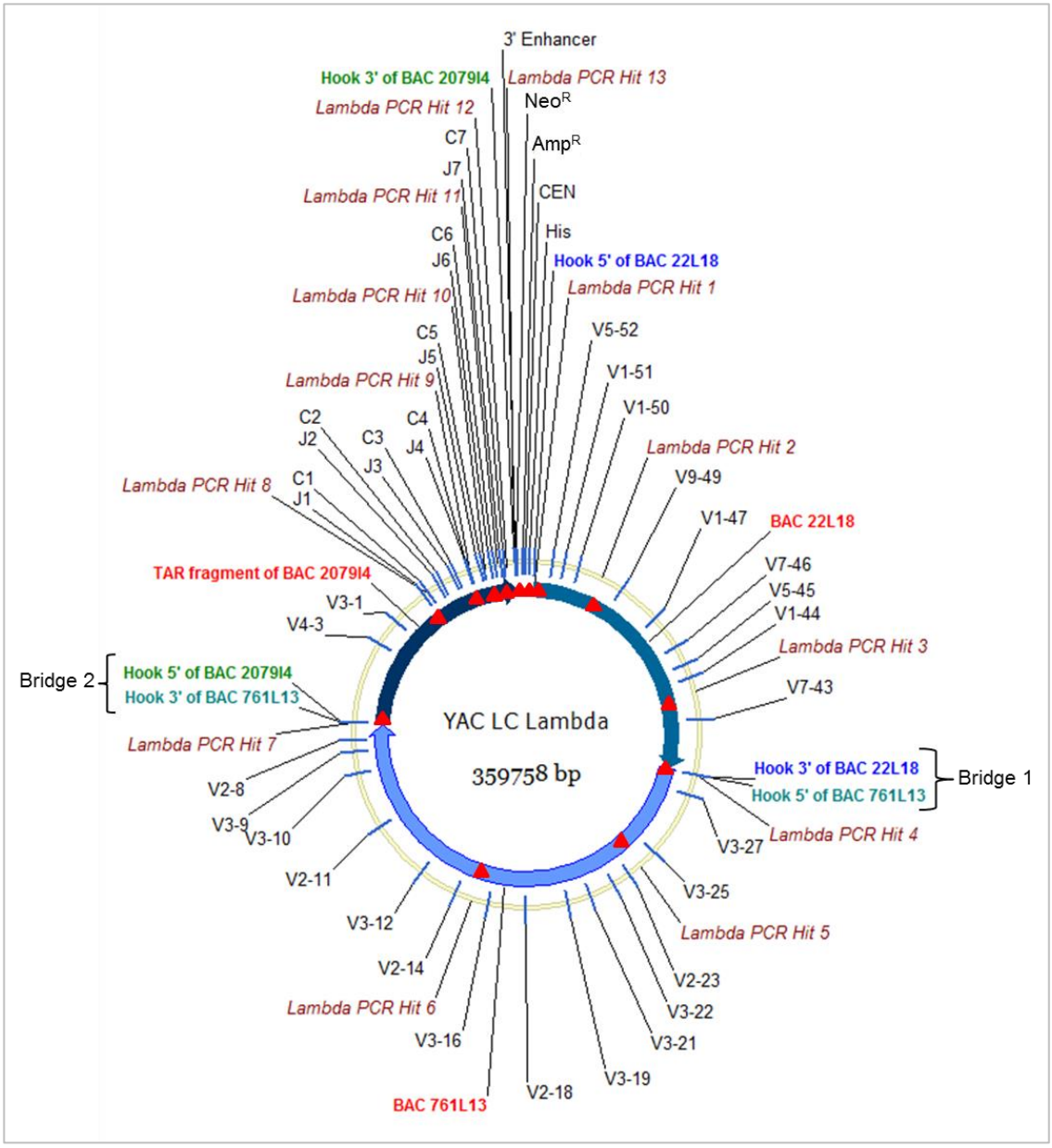

Figure 45: The YAC LC Lambda #8, 359 758 bp. YAC LC Kappa comprises BAC n° 22L18, 761L13 and fragment of BAC n° 2079I4. All V segments are given, and all Jλ-C genes. The V leader sequences were removed from the scheme for a question of readability. Hooks and bridges are also represented. Genetic elements as *Neo*^{R}, *Amp*^{R}, *CEN*, *His* and *3' Enhancer* are pVC604 original components. All positive PCR Hits are represented by red triangles.

3.1.4.3. Human Ig HC cloning

To build up a functional human Heavy Chain *locus*, four different BACs (n° 702F21, n° 72N10, n° 259B19 and n° 448N5), containing the genomic gene segments for a proper expression of human HC, have been linked together in one single piece of DNA. The four selected genomic regions include 31 functional V gene segments out of 41 to 46, 12 D out of 24, the entire J cluster, Cμ, Cδ, Cɣ3 and Cɣ1 gene exons (**Figure 46**). However, as for the Lambda LC reconstitution, we first inserted into the HC TAR vector arm the missing HC 3' Enhancer, in order to ensure proper expression of the human heavy chain genes. We designed bridges and hooks with appropriate restriction sites. This was synthetized by GeneArt® Gene Synthesis and inserted into the final HC TAR vector (**Figure 47**). Between both hooks composing the bridge (Hook 3' of BAC 72N10 and Hook 5' of BAC 259B19) we inserted a selection marker (KanMX4 encoding resistance to kanamycin). Thus after yeast spheroplast transformation, positive clones were selected using medium exempted from histidine (His -) and with 200 μg/mL of Geneticin®.

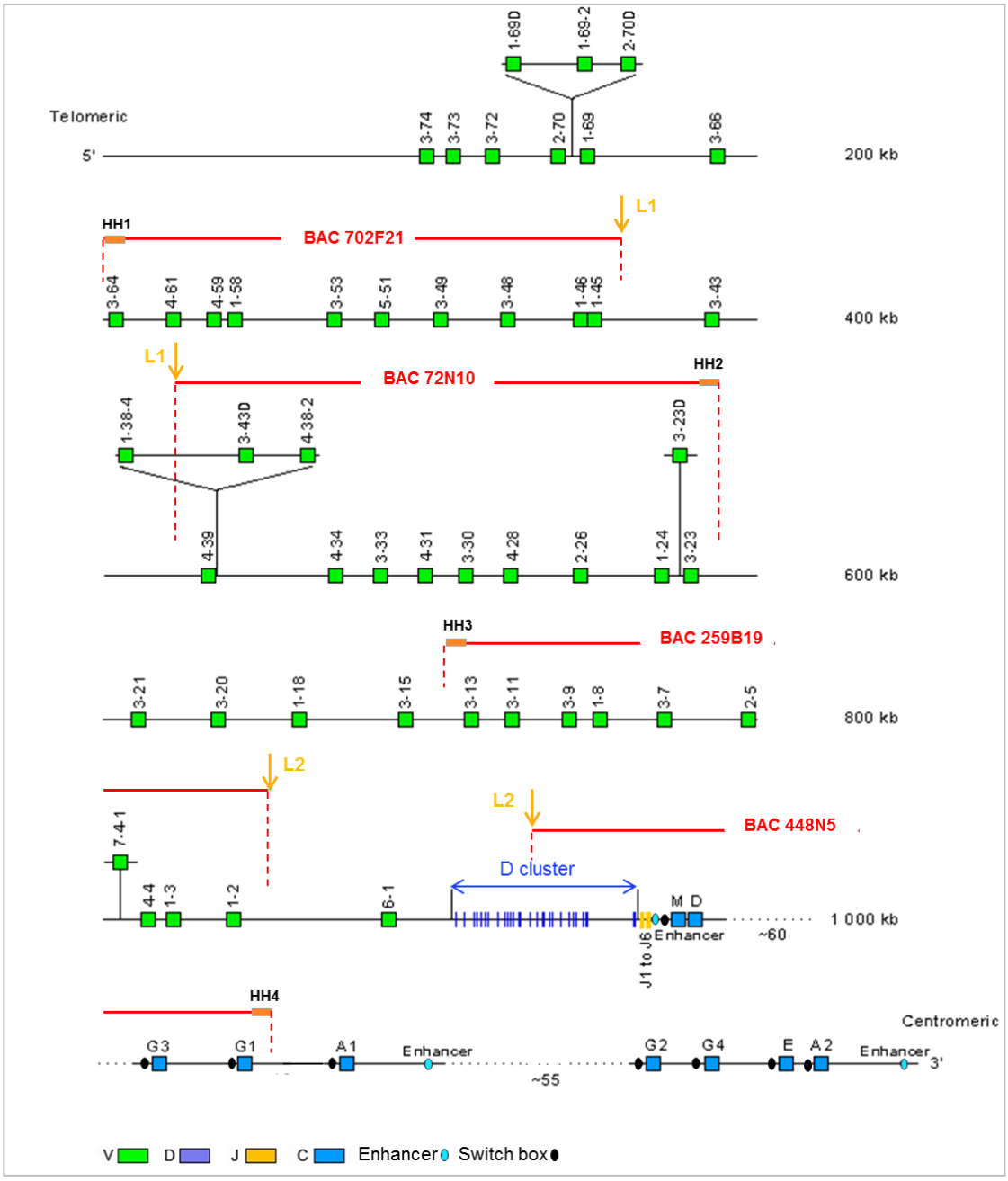

Figure 46: Schematic representation of the human Ig HC *locus* (1 250 kbp) indicating the four BACs (*red lines*) used to reconstitute the *locus* as YAC HC. Constant parts annotated M, D, G3, G1, A1, G2, G4, E and A2 stand for Cμ, Cδ, Cγ3, Cγ1, Cα1, Cγ2, Cγ4, Cε and Cα2. The duplicated V segments are designated by the same numbers as the corresponding genes, with the letter D added. The orange boxes show the position of the hook homology (HH). Hooks 1 and 4 are part of the vector, while hooks 2 and 3 belong to the bridge linking BAC 72N10 and BAC 259B19 together. Between hooks 2 and 3, the gene coding for resistance to kanamycin (*KanMX4*) was included for better selection of transformants. Yellow arrows stand for ligation (L1 = ligation 1, L2 = ligation 2). (Picture: Lefranc, M.-P. IMGT®).

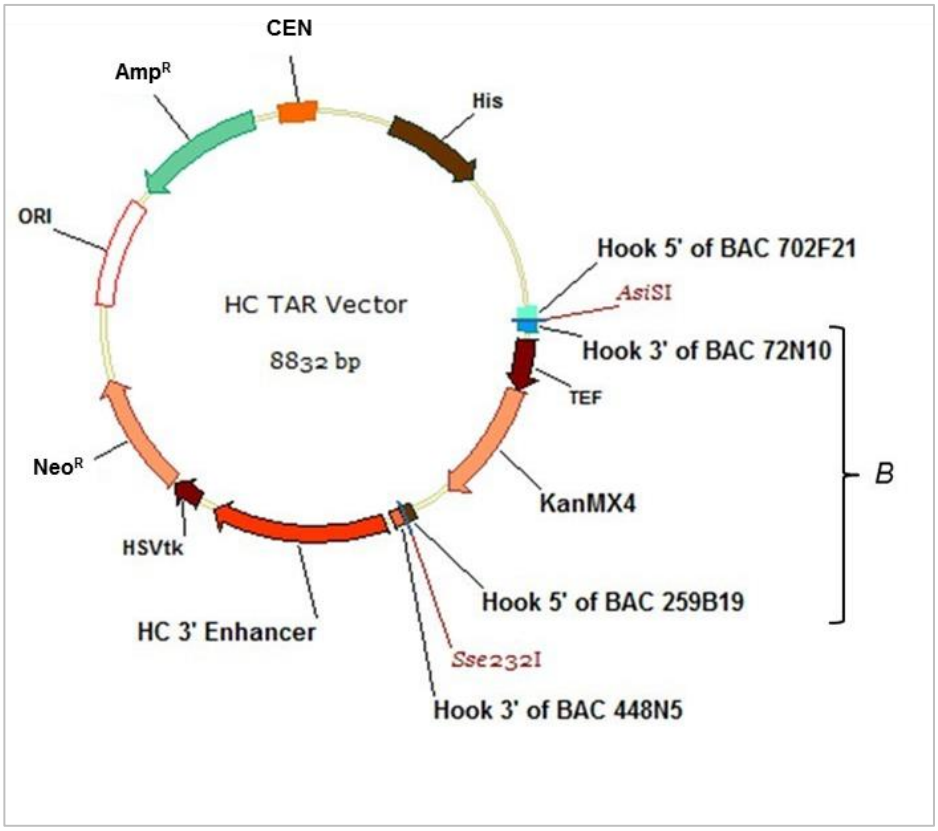

Figure 47: The HC TAR vector, 8 832 bp. To the vector pVC604 were added the neomycin resistance gene (*Neo*^{R}) under HSVtk promoter, the *HC 3' Enhancer*, four hook sequences with appropriate restriction sites and *KanMX4* gene (under TEF promoter) placed between Hook 3' of BAC 72N10 and Hook 5' of BAC 259B19. *Asi*SI cuts uniquely between Hook 5' of BAC 702F21 and Hook 3' of BAC 72N10, while *Sse*232I cuts uniquely between hooks Hook 5' of BAC 259B19 and Hook 3' of BAC 448N5. Hook 3' of BAC 72N10, *KanMX4* gene and Hook 5' of BAC 259B19 compose the bridge (B). Each hook sequence accounts for around 75 bp.

Regarding genomic template, the BAC inserts 702F21 and 72N10 on one hand and BAC inserts 259B19 and 448N5 on the other hand, were isolate and ligated together before proceeding with TAR. In brief, the four BACs were digested *Not*I, 2 hours at 37 °C. After enzyme heat inactivation, 80°C for 10 minutes, BAC n° 702F21 and n° 72N10 were ligated together with T4 DNA Ligase (*ref.*: NEB #M0202S). The same procedure was followed with BAC inserts 259B19 and 448N5. Finally, 3 µg of purified ligation products and 1 µg of *AsiSI* and Sse232I linearized HC TAR vector were transformed into yeast spheroplasts. Transformation was performed in triplicate. Like for the cloning of Lambda *locus*, we used negative controls. After 1 week of double selection (His$^-$ and

200 µg/mL of Geneticin®) at 30 °C, 200 yeast clones in total developed and no yeast colony developed from negative controls. From the 200 clones obtained, 5 showed PCR hits and one of them, YAC HC #2 showed human antibodies sequences derived from each of the four BAC DNA template, as well as bridge elements. Since the YAC HC is too big to be represented as plasmid map, the construct is schematized in **Figure 48**.

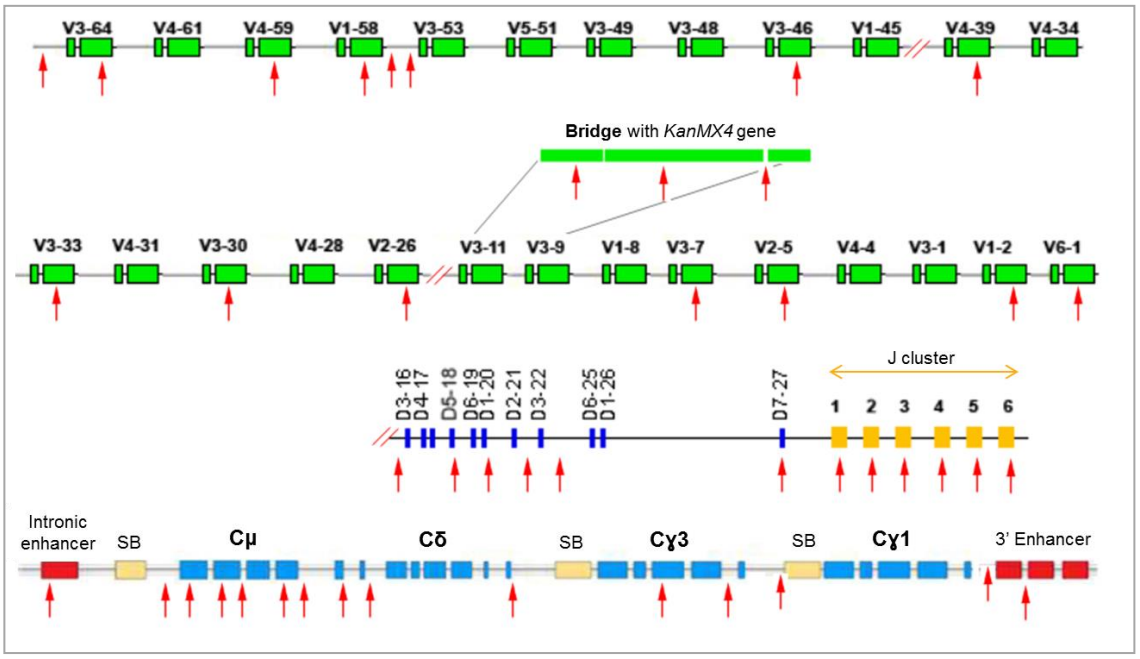

Figure 48: Scheme of YAC HC, 658 686 bp. Green boxes stand for V segments and their corresponding leader sequence; dark blue sticks schematize D segments, while J segments are shown with yellow boxes; blue boxes stand for exons coding for different constant regions. In 5' of Cμ, Cγ3 and Cγ1, beige boxes stand for switch box sequences (SB). Intronic and 3' Enhancers are schematized with red boxes. pVC604 elements are not represented. Each red arrow shows the position of positive PCR hit.

3.1.5. Human Ig YAC collection

TAR cloning allowed us to reconstitute the three human Ig *loci* resulting in the clones schematized in **Figure 49** below.

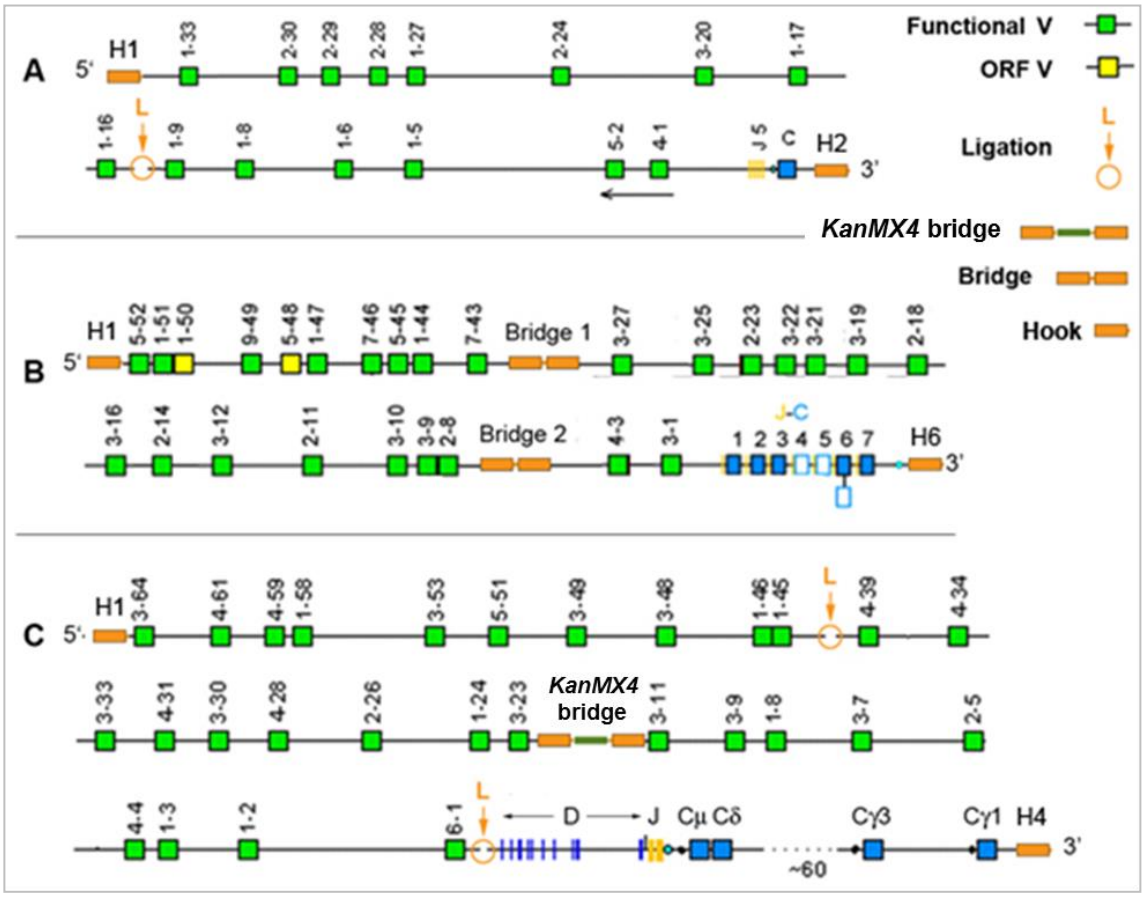

Figure 49: schematic representation of the human Ig clones obtained by TAR cloning.
A; human Ig Kappa clone, 398 kbp.
B; human Ig Lambda clone, 360 kbp.
C; human Ig HC clone, 659 kbp.

3.1.6. YAC purification for transgenesis

Historically, YACs were transferred from yeast into the mouse genome *via* mouse embryonic stem (ES) cells through spheroplast fusion [82]. However the injection of purified YACs directly into mouse oocytes was also previously demonstrated [83] but YAC purification procedure was not completely established. We first, performed fusion between yeast spheroplast containing YAC HC #2 and mouse ES cells, and we obtained two ES clones IgH2.2 and IgH2.8. Both ES cells clone were deeply analyzed by PCR.

Analysis proved the presence of human and synthetized gene segments coming from all BAC inserts and TAR vector elements like the bridge.

However, a recent method demonstrated the isolation of circular YAC in a quantity and a quality suitable for the injection into mouse oocytes [84]. In brief, yeast cells harboring human Ig YAC were cultivated in liquid selection medium, until reaching the optical density (OD) at 600 nm of 1.5. After harvesting the cells, a Zymolyase treatment followed by alkaline lysis, using detergent based buffer leaded to the extraction of total DNA, *i.e.* yeast genomic DNA and YAC molecules. Phenol:chloroform:isoamyl alcohol purification was used to removes proteins and cell debris from such extracted DNA. This was followed by a second extraction consisting in the digestion of the linear yeast genomic DNA by an exonuclease and the isolation of the circular YAC molecules. To this end, extracted DNA samples were treated with exonuclease under optimized conditions. Finally, circular YACs and digest genomic DNA were separated by the use of a Qiagen-tip 500 column (QIAGEN Large-Construct Kit, ref: 12462). Circular YAC molecules retained on the column were eluted and concentrated. An aliquot of purified YAC molecules was visualized on 0.7% agarose gel and compared with an aliquot of total yeast DNA before exonuclease digestion (**Figure 50**).

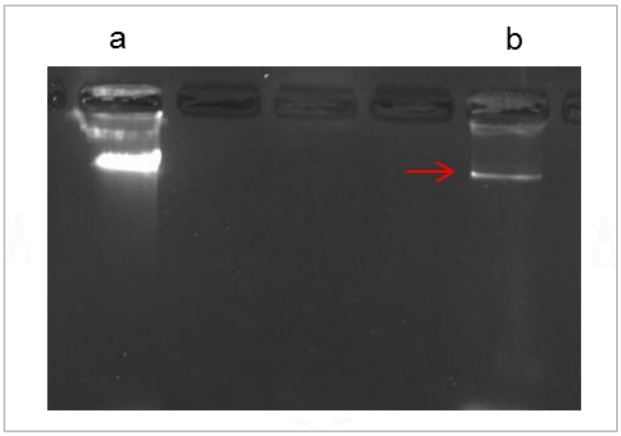

Figure 50: Electrophoresis analysis of circular YAC purification.
a: total DNA, *i.e.* yeast genomic DNA mixed with circular YAC,
b: purified circular YAC DNA (*red arrow*).
Gel: 0,7% agarose, 0.01% of RedSafe DNA staining solution (*ref.*: 21141)
Running conditions: 5 min, 150 V, then 60 min, 60 V.

3.2. Alternative approach: size reduced human Ig HC *loci*

Main project of this thesis

As an alternative to the main approach, for which we attempted to reconstitute the whole three antibodies *loci*, we build up reduced human Ig HC genetic constructs. As for the re-assembling of the human Ig *loci* in YAC (chapter 3.1.), the sources of human antibody genes was the BAC collection (**Figure 34**). Thus, using standard molecular biology techniques, we generated a HC Microlocus Light and a HC Microlocus Classic in form of plasmids as intermediate Ig HC constructs for the final HC Minilocus. The reduced size of such plasmid makes them easier to handle in comparison to the huge YACs generated for the main approach in view also of mouse oocytes injection. However, by adapting the Transformed Associated Recombination (TAR) [3], we also built up a small YAC, that contains a broad range of 13 selected human V segments. Thus, we generated a compromised in size and complexity human HC *locus* called HC Minilocus.

Ig HC *locus* encodes the immunoglobulin heavy chain. This protein results from the random recombination at the genomic DNA level, of three gene segments: V, D and J, with deletion of the DNA located between the used gene segments, to generate a rearranged *VDJ* gene exon. The rearranged *VDJ* gene exon is transcribed with an Ig constant gene (Cμ during B cell development) and translated into an immunoglobulin heavy chain [1]. Therefore, a reduced *locus* encoding human Ig HC should contain a selection of the appropriated genetic elements, whose specific combinations will result in the expression of functional HC proteins.

The reduced HC Minilocus construct was designed following the native organization of human antibody genes. The reduction consisted, first, in the elimination of the non-essential sequences, like pseudo genes or long non coding regions but taking care to keep the essential genetic elements. However, in order to further reduce the size, we also limited the number of V, D and J gene segments. To keep an acceptable diversity, the V gene segments were carefully selected, based on phylogenetic distance analysis. The D segments were selected to be the most representative member per family. The J segments and the C regions have been kept close to their natural state. All well-known antibody regulatory elements have been incorporated into the construct. Some regions,

like D and J, were fully synthesized, reducing this entire cluster from ~60 kbp to 1.25 kbp.

3.2.1. Prerequisite to design reduced human Ig HC *loci*

The goal of this approach was to massively reduce the size of Ig HC *locus* (smaller in size than 50 kb) but retaining functionality compatible with the numerous steps required for B lymphocytes differentiation and maturation. Therefore the genetic organization, including selected genes and crucial control elements, had to be kept as closest as possible to the natural configuration. A proper selection of the essential genes and regulatory elements will ensure the expression of human antibodies in transgenic mice.

However, by progressing in stepwise fashion, several intermediate Ig HC *loci* derived constructs were generated; the HC Microlocus Light and the HC Microlocus Classic. Both were tested for their biological activity in cell line and in transgenic mice.

3.2.2. The design of the HC Minilocus

To build up a reduced *locus* encoding the human Ig HC, we selected and cloned several human V, D and J gene segments. The selection should maintained large antibody diversity and all effector functions. The HC Minilocus building strategy consisted to assemble stepwise the selected DNA fragments in three independent groups before the final assembly:

- Group 1: Assembly of a set of 7 V segments mainly from family 3, a reduced D cluster and the J cluster.
- Group 2: Assembly of a set of 6 V segments from family 1, 2 and 4
- Group 3: Assembly of the selected C regions (Cμ and Cɣ1) including switch boxes and transmembrane regions.

Regulatory elements, like the HC intronic enhancer and the HC 3’ enhancer, have been added during the cloning steps.

Assembling in one vector all the essential gene segments for the production of a functional antibody heavy chain, required many gene amplification steps by PCR.

Specific PCR primer pairs (Mycrosynth) were used with the GoTaq® DNA polymerase from Promega under standard conditions and procedures (www.promega.com). PCR amplifications were performed using standard PCR thermo-cycler. Samples were loaded on 1-1.5% agarose gel electrophoresis containing 0.01% of RedSafe DNA staining solution (ref: 21141), which allows DNA detection after electrophoresis. Gel electrophoresis conditions were 140 volts, during 20 to 30 minutes. PCR products of the expected size were isolated from agarose gel and the DNA extracted using QIAquick Gel Extraction Kit (ref: Qiagen, 28704). PCR products were sequenced before cloning.

3.2.3. Choice of Ig genetic elements

3.2.3.1. Design of the (D) diversity and (J) junction elements

3.2.3.1.1. The D cluster

Native human D cluster spreads over a distance of 53 kbp and is composed of 27 functional D segments distributed in 6 groups containing 4 to 5 D segments each and one unique D7 [29]. The cluster can be largely reduced by designing consensus sequences for each D group. Consensuses were established by keeping for each position the most representative nucleotide. The longest D from each family was retained. This selection strategy could not be strictly respected and became more random in case where two different nucleotides in equal number were found at the same position. **Table 4** summarizes all D sequences for each family and the resulting consensus sequence for D1 to D6. D gene coding segments surrounding sequences were copied from the natural sequence and have a length of 100 to 180 bp, similar to the shortest functional one found in natural HC *locus*. Finally, we took care to keep the somatic recombination signal sequence (RSS) elements flanking the D segments in a correct conformation [85].

Table 4: Consensus sequences for D segments. All native D segments are listed by family in the left column. The corresponding sequences are in the right column. For each family, one D consensus sequence was defined by keeping for each position the most representative nucleotide: in black, the sequence common to all segments within each D family; in blue, the sequence differences between D segments; in red, the selected nucleotide for the consensus sequence.

D1	
D1-1	GGTACAACTGGAACGAC
D1-7	GGTATAACTGGAACTAC
D1-14	GGTATAACCGGAACCAC
D1-20	GGTATAACTGGAACGAC
D1-26	GTATAGTGGGAGCTACTAC
D1 consensus	GGTATAACTGGAACGACTAC
D2	
D2-2	AGGATATTGTAGTAGTACCAGCTGCTATGCC
D2-8	AGGATATTGTACTAATGGTGTATGCTATACC
D2-15	AGGATATTGTAGTGGTGGTAGCTGCTACTCC
D2-21	AGCATATTGTGGTGGTGAT TGCTATTCC
D2 consensus	AGGATATTGTAGTAGTACCAGCTGCTATGCC
D3	
D3-3	GTATTACGATTTTTGGAGTGGTTATTATACC
D3-9	GTATTACGATATTTTGACTGGTTATTATAAC
D3-10	GTATTACTATGGTTCGGGGAGTTATTATAAC
D3-16	GTATTATGATTACGTTTGGGGGAGTTATGCTTATACC
D3-22	GTATTACTATGATAGTAGTGGTTATTACTAC
D3 consensus	GTATTACGATTATTGGAGTGGTTATTATAACTATACC
D4	
D4-4	TGACTACAGTAACTAC
D4-11	TGACTACAGTAACTAC
D4-17	TGACTACGGTGACTAC
D4-23	TGACTACGGTGGTAACTCC
D4 consensus	TGACTACAGTAACTACTCC
D5	
D5-5	GTGGATACAGCTATGGTTAC
D5-12	GTGGATATAGTGGCTACGATTAC
D5-18	GTGGATACAGCTATGGTTAC
D5-24	GTAGAGATGGCTACAATTAC
D5 consensus	GTGGATACAGCTATGGTTAC
D6	
D6-6	GAGTATAGCAGCTCGTCC
D6-13	GGGTATAGCAGCAGCTGGTAC
D6-19	GGGTATAGCAGTGGCTGGTAC
D6-25	GGGTATAGCAGCGGCTAC
D6 consensus	GGGTATAGCAGCGGCTGGTACA
D7	
D7-27	CTAACTGGGGA

3.2.3.1.2. The J cluster

The J cluster spreads over 2 kbp in the natural HC *locus* [86]. By keeping in mind the goal to reduce the artificial *loci*, J5, which is almost identical to J4, was not integrated in the cluster. Sequences between the J segments were kept in their natural configuration but were reduced to ~150 bp. The synthetic J cluster is shorter than 1 kbp. The sequences of J segments are:

J1: 5'GCTGAATACTTCCAGCACTGGGGCCAGGGCACCCTGGTCACCGTCTCCTCAG3'

J2: 5'CTACTGGTACTTCGATCTCTGGGGCCGTGGCACCCTGGTCACTGTCTCCTCAG3'

J3: 5'TGATGCTTTTGATATCTGGGGCCAAGGGACAATGGTCACCGTCTCTTCAG3'

J4: 5'ACTACTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAG3'

J6: 5'ATTACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTC CTCAG3'

3.2.3.1.3. The synthesized D-J construct

The six D consensus, the D7 and the 5 selected J sequences were synthetized as a single DNA fragment (GeneArt® Gene Synthesis). The resulting D-J construct is shown in **Figure 51**. The DJ cluster was inspired by the natural HC *locus*, but due to consensus sequences and shorter intervening sequences, this provides a useful unique genetic content and organization. During functionality tests, this allowed the identification of gene products derived from the construct, and therewith avoiding ambiguity with possible contamination.

This plasmid, containing all required genetic information for the growth of bacteria, was transformed in *Escherichia coli*. The purpose of this bacterial transformation is to introduce the plasmid into bacteria and to use bacteria DNA replication machinery to amplify the plasmid in order to make large quantities of it. After transformation, bacteria were plated on LB-agar medium with 30ug/mL of kanamycin. Only transformed bacteria can grow on this selected medium, therefore visible colonies were picked and cultivated in liquid LB medium with 30ug/mL of kanamycin. Multiple copies of plasmids were purified using QIAprep Spin Miniprep Kit (*ref*: 27104) from 1.5 mL of

culture. Before further cloning step, DNA quality and integrity was controlled by restriction mapping.

The success of the size reduced Ig HC constructs approach depends on the functionality of the D-J construct whose elements are playing a central role in the somatic DNA recombination process. Indeed, all genetic constructs encoding human HC are based on the vector shown in the **Figure 51**.

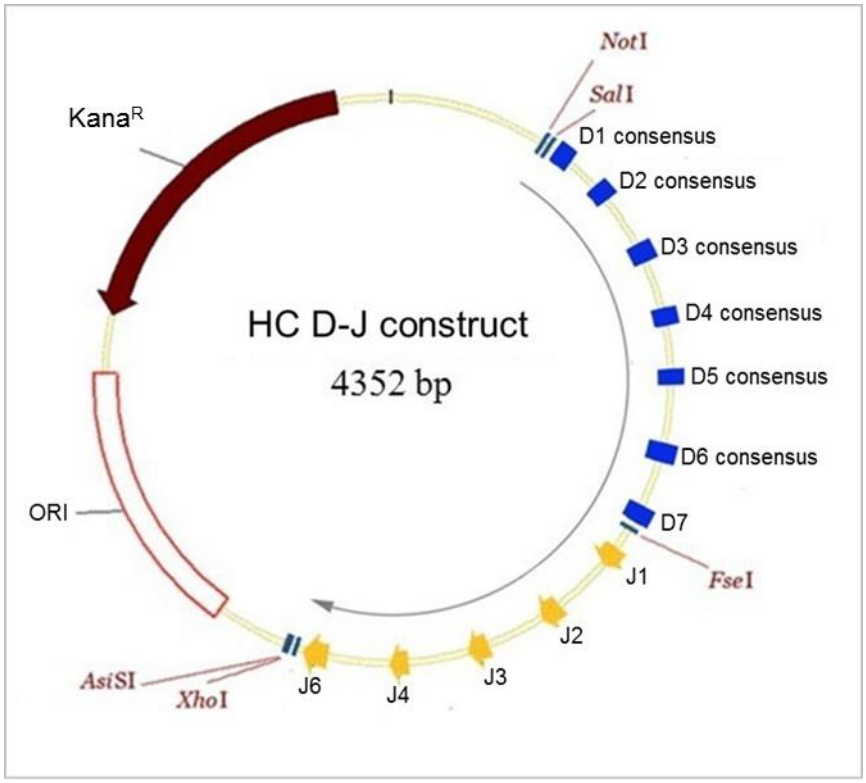

Figure 51: The HC D-J construct; 4,3 kbp. The seven D segments are represented as blue boxes; yellow arrows show the J segments. *ORI*: Origin of Replication; *Kana*^{R}: kanamycin resistance gene as bacterial selectable marker. The grey arrow indicated the coding orientation. Several unique restriction sites are included (i.g. Asi*SI*, Xho*I*, Not*I*, Sal*I*) allowing further cloning steps.

3.2.3.2. The V cluster

3.2.3.2.1. The variable (V) segments selection

Following native human HC *locus* organization, several selected human V segments were cloned in 5' of the D-J cluster. A human V cluster was reconstituted using a set of carefully selected V gene segments, including promoter, leader sequences, V and RSS.

The natural human HC *locus* bears around 45 functional V segments; our human HC BAC collection contains 34 V segments, from which 13 were selected. V segments were amplified by PCR, using specific primers and BAC as DNA template.

Regarding V sequences, there is a different range of similarity between segments. Some of them are very similar: for example, V3-30 and V3-33 are genetically analogous. Both were amplified in the same PCR reaction. For the final selection, only one of them was used (V3-33). Therefore, the selection of V segments relies on genetic distance to be the most representative of the native V cluster (**Figure 52**). Since the V3 family is the most represented in the human genome they account for 6 V gene segments in our selection of 13.

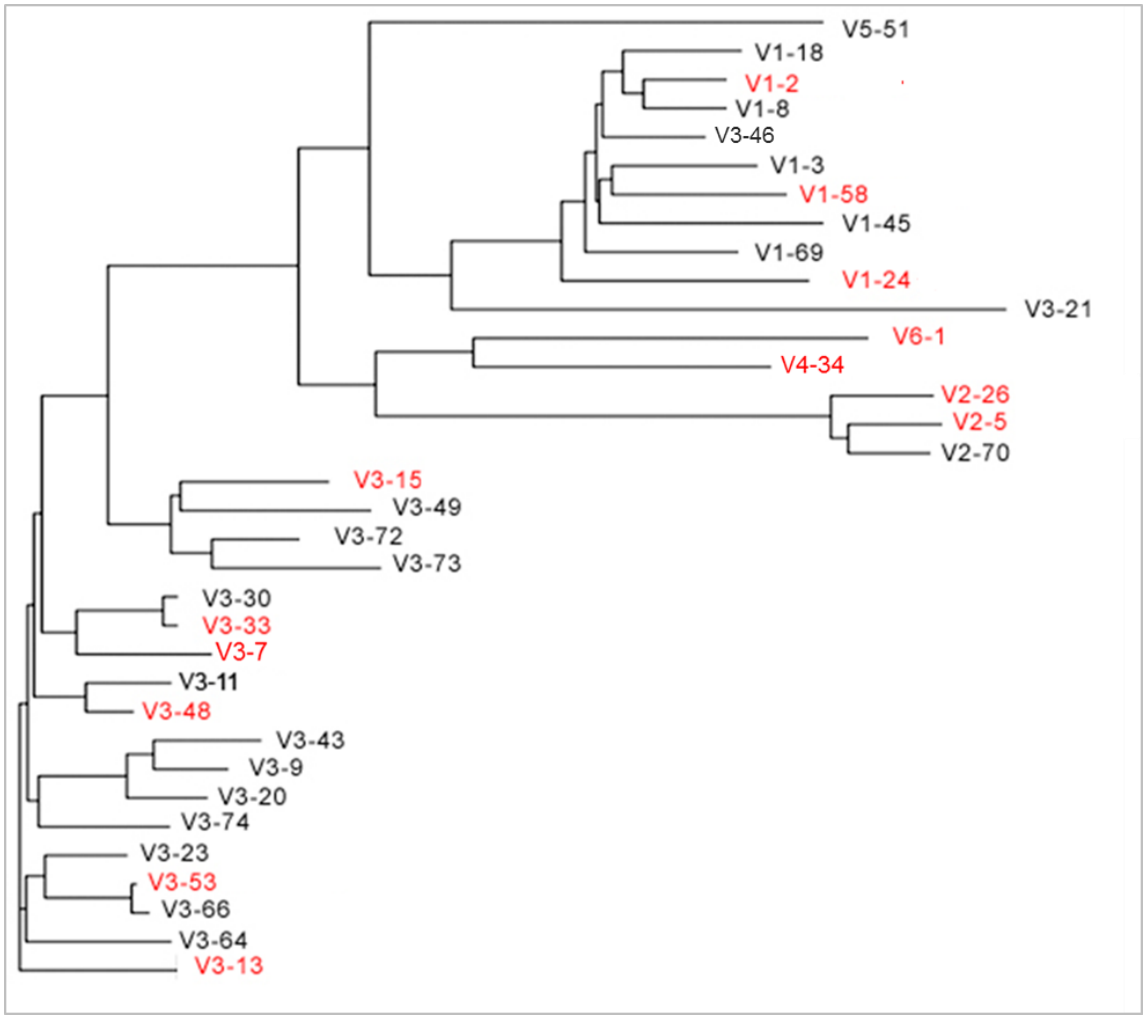

Figure 52: Phylogenetic tree showing genetic distance between human V gene segments. The selected 13 V segments are in red.

3.2.3.2.2. V segments isolation and assembling

The thirteen selected V segments were amplified by PCR, using specific unique primer pairs and the corresponding BAC as DNA template, among BACs n° 702F21, 101G24, 72N10, 141I17, 11J18 and 1065N8. For each of them, two specific primers were designed to amplify fragments of about 1.0 kbp including the promoter, the leader, the V encoding sequence and the 3' RSS. To design the primers, we faced the problem of numerous repetitive sequences present in the human Ig gene *locus*. The forward primer was designed to match approximately at 700 bp 5' of the coding V region and the reverse primer was located 100 bp 3' of the end of the coding V region. The primers were designed with non-homologous tails containing appropriate restrictions sites for the "step by step" cloning. Forward primer contained *Not*I and *Sal*I restriction sites, whereas reverse primer had *Xho*I. Each amplified PCR fragment was visualized by UV illumination after gel electrophoresis (1% agarose) and extracted using QIAquick Gel Extraction Kit (*ref*: 28704). To control their integrity, purified DNA fragments were sequenced before further processing. Sequenced DNA fragments were digested with *Not*I and *Xho*I, while the vector was digested with *Not*I and *Sal*I. *Xho*I and *Sal*I share a cohesive overhang and can be ligated together. After ligation, the resulting site *Sal*I/*Xho*I cannot be cut by either *Sal*I or *Xho*I again. The cloning strategy is schematized in **Figure 53**. After each V segment addition, constructs were transformed in *E. coli* and DNA integrity was controlled by restriction mapping and sequencing.

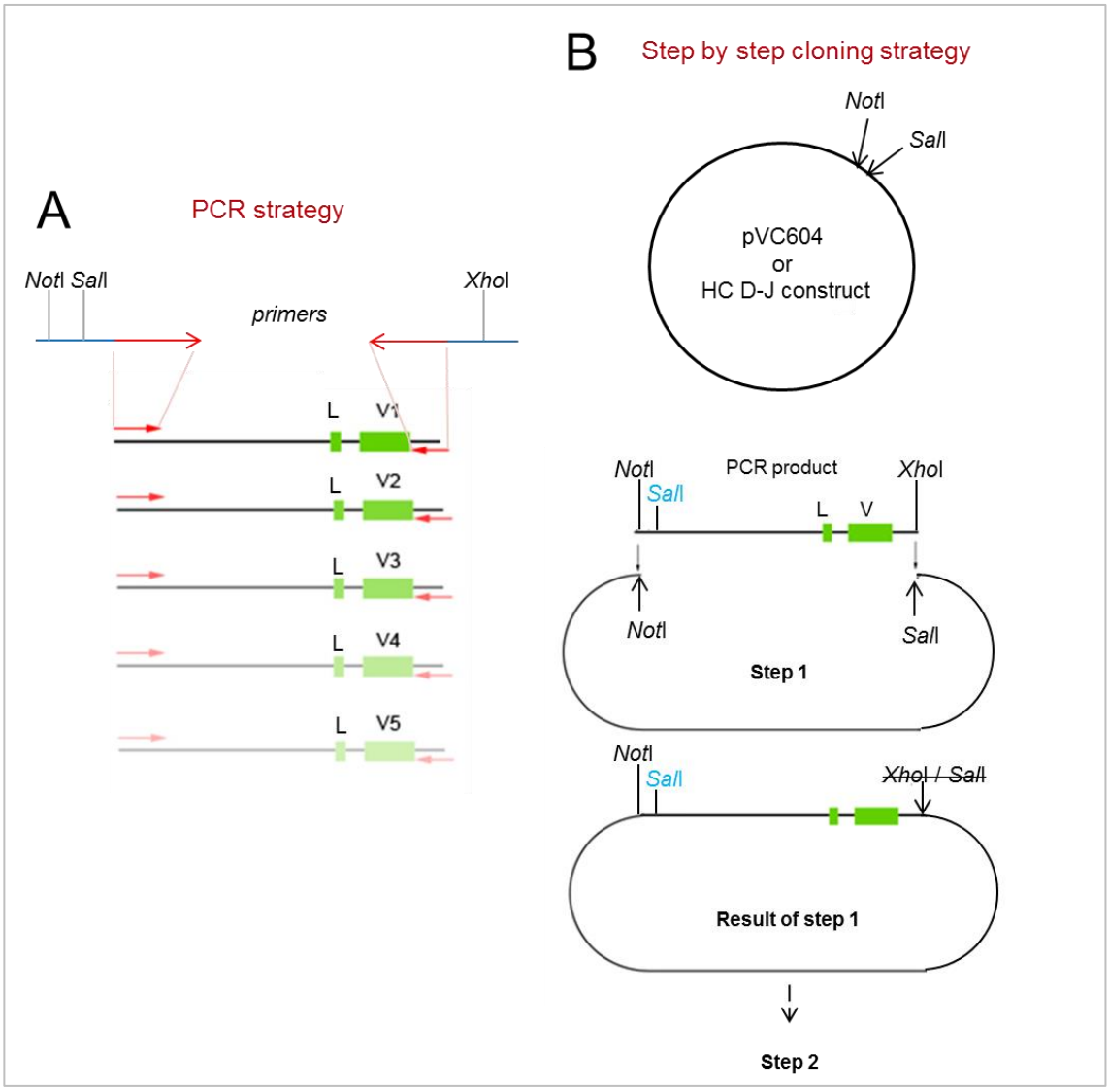

Figure 53: Schematic representation of the step-by-step V segments assembly strategy**.**
A; The PCR primers used (*red arrows*) had a non-homologous tail (*blue line*) containing specific restriction sites (*Not*I and *Sal*I for the forward primer; and *Xho*I for the backward primer). Selected V segments (*green boxes*) and their corresponding leader (L) sequence (*small green boxes*) were amplified by PCR.
B; "step-by-step" cloning strategy. First the vector was digested using *Not*I and *Sal*I, while PCR product was digested with *Not*I and *Xho*I. Vector and insert were ligated *Not*I/*Not*I and *Xho*I/*Sal*I. Each newly ligated PCR product integrated a new *Sal*I restriction site (*in light blue*). Therefore in the next round, the vector can be again digested with *Not*I and *Sal*I.

3.2.3.2.3. The two intermediary constructs

For experimental convenience, we decided to split the V segments selection in two groups and to assemble both groups in two parallel constructs. The assembling strategy is the one described in the previous part. Both constructs are named:

- HC V-D-J construct 1, based on HC D-J construct (**Figure 54**)
- HC V construct 2, based on pVC604 vector (**Figure 55**)

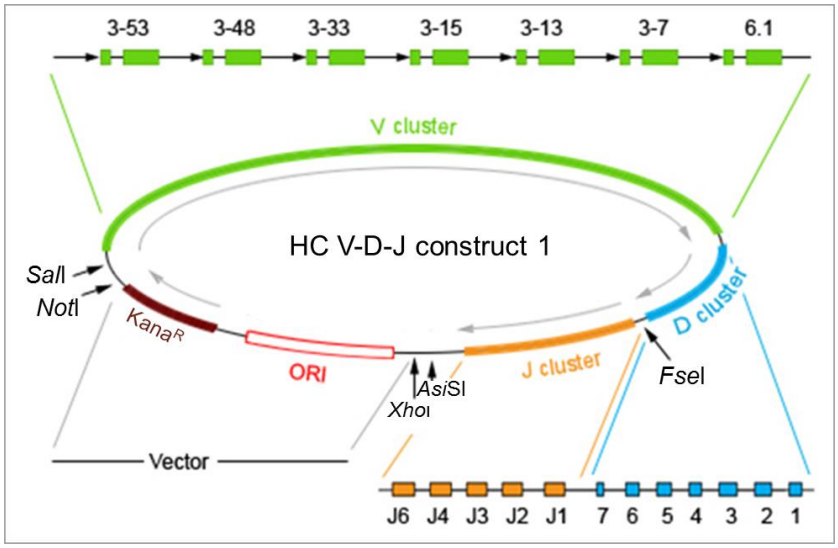

Figure 54: The HC V-D-J Construct 1; 12 kbp. The construct 1 contains the five selected V segments from family 3 and the unique V6-1. Green boxes: V segments and corresponding leader sequence; blue boxes: D cluster; orange boxes: J cluster; *ORI*: Origin of Replication; *Kana*^{R}: gene coding for resistance to kanamycin as selection bacterial marker. *Sal*I, *Not*I, *AsiSI*, *Xho*I and *Fse*I indicate unique restriction site recognized by enzymes of the same name; they serve for further cloning step.

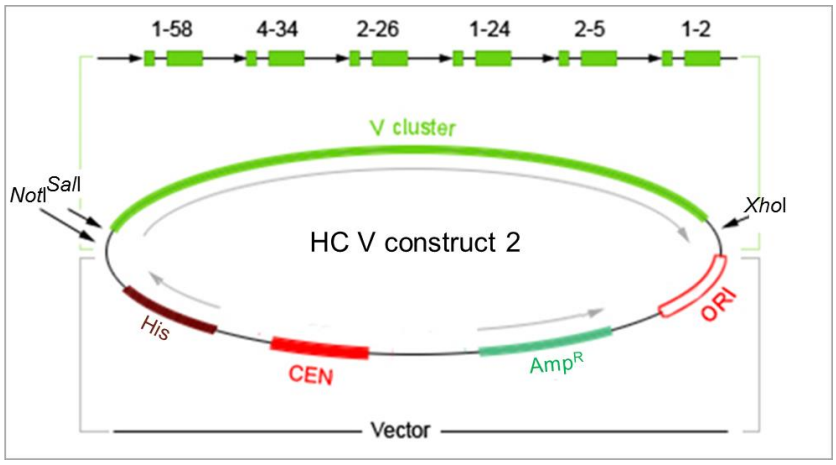

Figure 55: The HC V Construct 2; 11.4 kbp. The construct 2 contains the selected V segments from families 1, 2 and 4. Green boxes: V segments and corresponding leader sequence; *ORI*: Origin of Replication, *Amp*^{R}: gene coding for resistance to ampicillin as selection bacterial marker, *CEN*: a yeast centromer; *His*: the gene encoding histidine for selection in yeast. *Sal*I, *Not*I, *and Xho*I indicate unique restriction site recognized by enzymes of the same name; they serve for further cloning step.

3.2.3.3. Cloning (C) constant regions

IgM is involved in central B cell development and IgG1 is the most representative serum immunoglobulin isotype. Therefore both constant regions, Cμ and Cɣ1 were retained to be cloned in HC Minilocus. Initially, the strategy was designed to clone also Cδ, but reconsidering the size parameter, we decided to focus only on Cμ and Cɣ1. To clone these human constant coding regions we used the TAR bridging approach (see 3.1.3. paragraph). Both isotypes are encoded by a unique BAC (BAC 448N5, **Figure 56**).

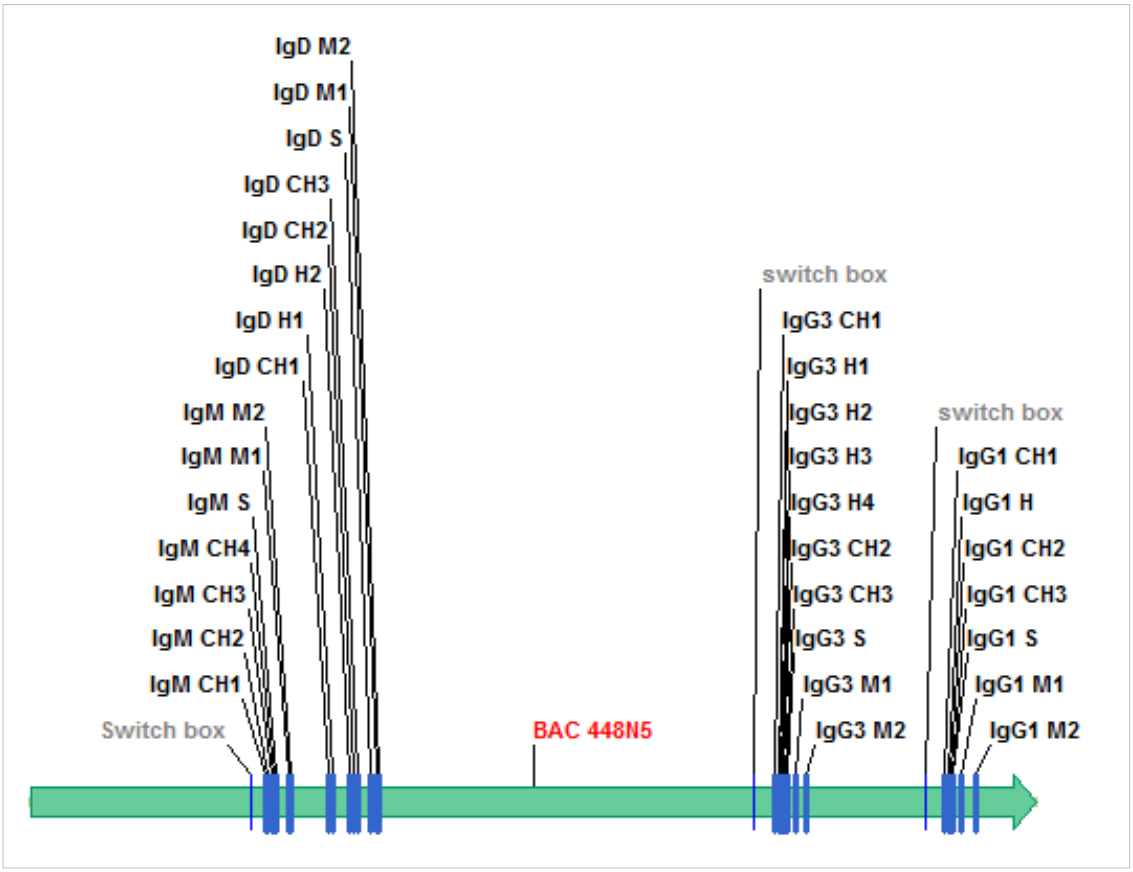

Figure 56: Genomic insert of BAC 448N5, 167 kbp.
The BAC 448N5 contains the gene exons coding for IgM (Cμ), IgD (Cδ), IgG3 (Cɣ3) and IgG1 (Cɣ1). S: coding sequence for secreted Ig, M1, M2: coding sequence for transmembrane regions,
H: coding sequence for hinge region.

We designed three pairs of hooks matching on either sides of Cμ, Cδ and Cɣ1, including switch boxes and transmembrane regions. The hooks were then cloned into the TAR vector (**Figure 57**). In the case of HC Minilocus, only Cμ and Cɣ1 were cloned. Therefore, TAR vector was digested with *Bam*HI and *Pme*I; and the hooks Cμ2, Cδ1, Cδ2 and Cɣ1 functioned as a bridge.

It is important to note that Cɣ1 and Cɣ3 display extended sequence homology, up to 100% in many regions. The two hooks designed on 5' and on 3' of Cɣ1 can therefore recombine on both sides of Cɣ3 with respectively 89% and 94% homology.

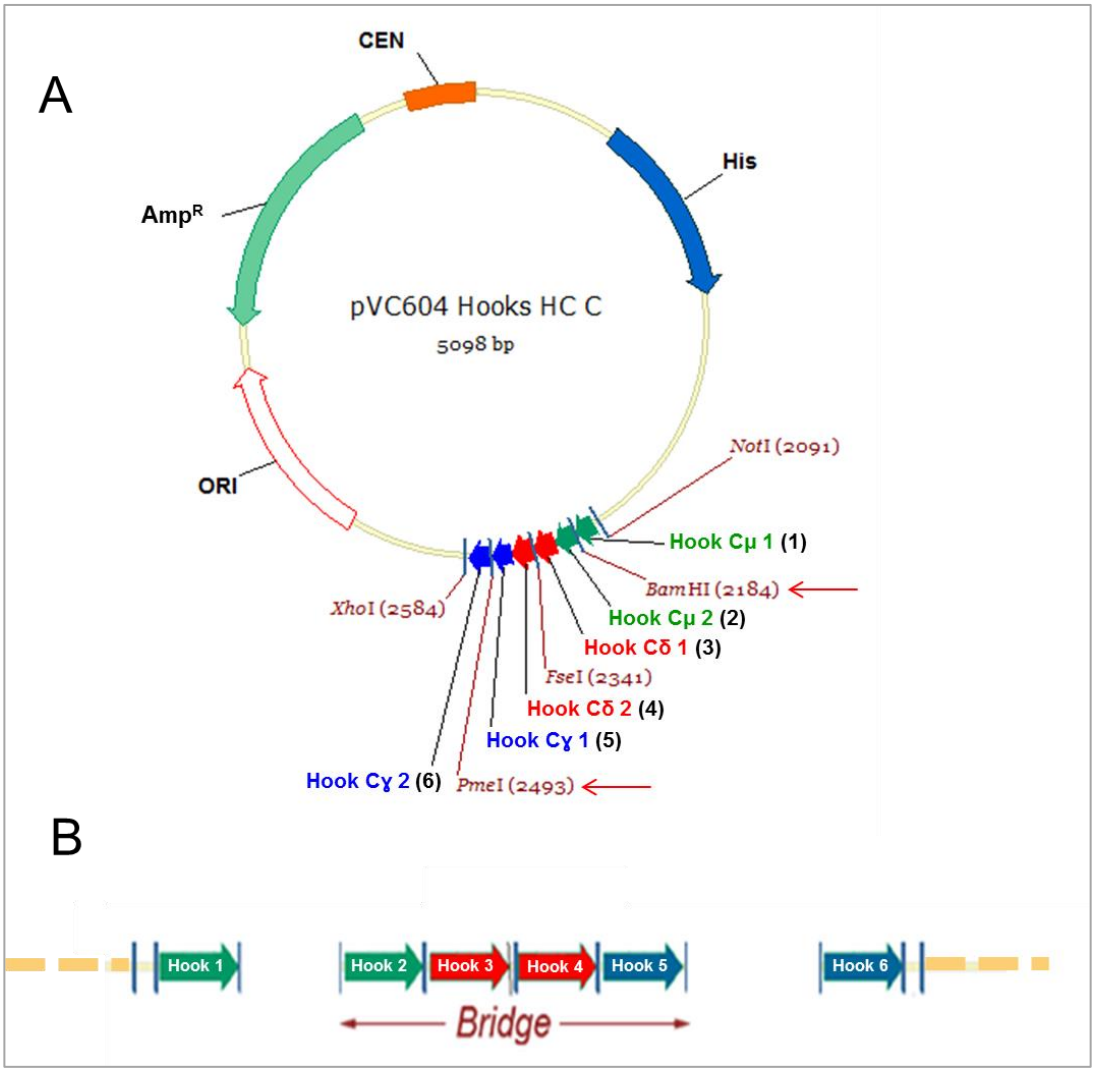

Figure 57: TAR Vector used to clone constant regions for HC Minilocus. **A**: TAR vector based on pVC604 contains *ORI*: Origin of Replication, *amp*: gene coding for resistance to ampicillin as selection bacterial marker, *CEN*: a yeast centromere; *His*: the gene encoding histidine for selection in yeast and specific hook sequences. Hooks Cμ1 (1) and Cμ2 (2) match uniquely on both sides of Cμ. Hooks Cδ1 (3) and Cδ2 (4) match on both sides of Cδ. Hooks Cγ1 (5) Cγ2 (6) match on both sides of genes coding for Cγ1 constant region, including the corresponding 5' switch box sequences. Before TAR cloning, the vector was digested with *Bam*HI and *Pme*I (*red arrows*). **B**: The previously digested vector enabled hooks 1 and 2 to recombine with their homologous sequences on either sides of Cμ and; in the same way; Cγ1 was "fished" by hooks 5 and 6. For HC Minilocus, Cδ is not included; therefore Hooks 2, 3, 4 and 5 worked like a bridge.

Following the protocol of Kouprina and Larionov [3], transformed yeast spheroplasts were plated on solid selection medium without histidine, for 1 week at 30 °C, in order to select transformed yeast cells. Emerging colonies were characterized following the same

strategy as described earlier for YAC HC, LCκ and LCλ (See 3.1.4.). After an extensive PCR screening, the selected clone were found to contain Cμ, Cɣ1 and also Cɣ3. Hook Cɣ1 (5) probably recombines in 5' Cɣ3 instead of 5' Cɣ1 (**Figure 58**). Indeed, this unwanted recombination increased the size of the final HC Minilocus YAC of 27 kbp. However, we decided to continue with this clone; the YAC HC C (**Figure 59**) was used to make the HC Microlocus (See 3.2.5.).

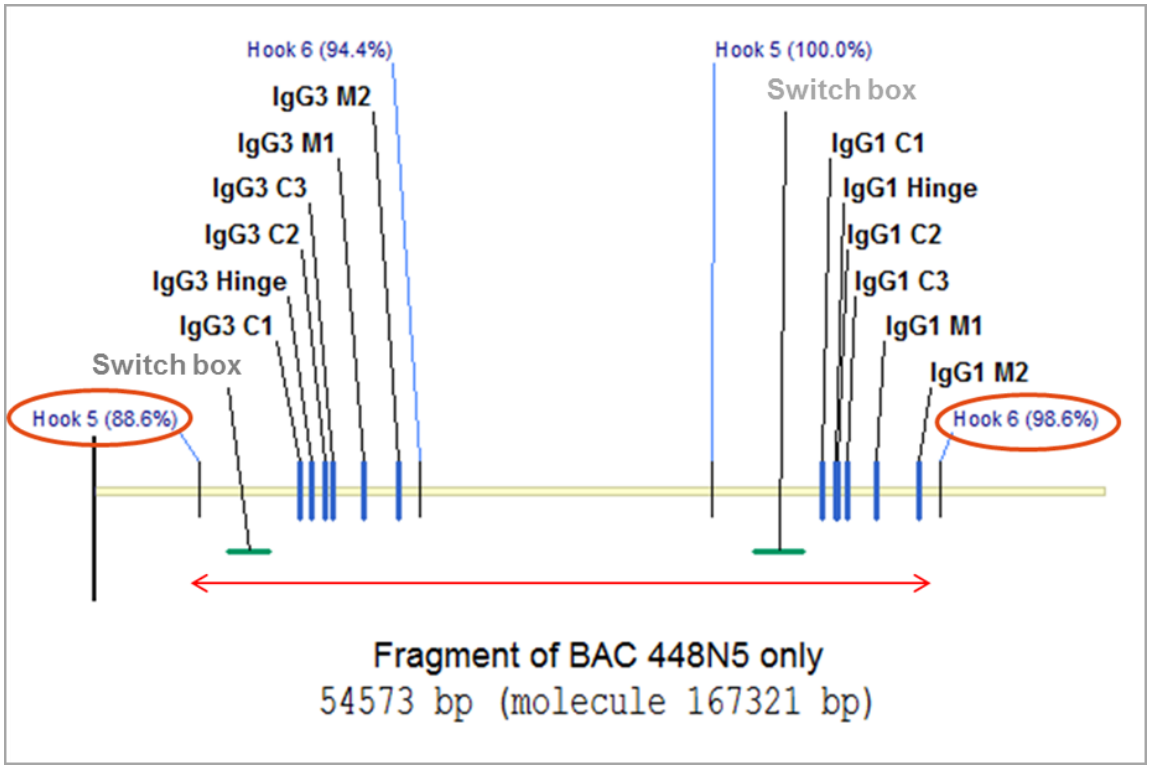

Figure 58: The hypothetical recombination explaining why Cɣ3 and Cɣ1 were cloned together. This picture show the homology of both hooks Cɣ1 (Hook 5) and hook Cɣ2 (Hook 5). The percentage in brackets indicates the homology rate. Considering PCR genotyping results, Hook 5 may have recombined on 5' of Cɣ3 (*circled in red*), instead of 5' of Cɣ1; while Hook 6 recombined, like expected, in 3' if Cɣ1 (*circled in red*). Hook 5 and Hook 6 may have fish Cɣ1 and Cɣ3, and the genomic intervening sequence between both.

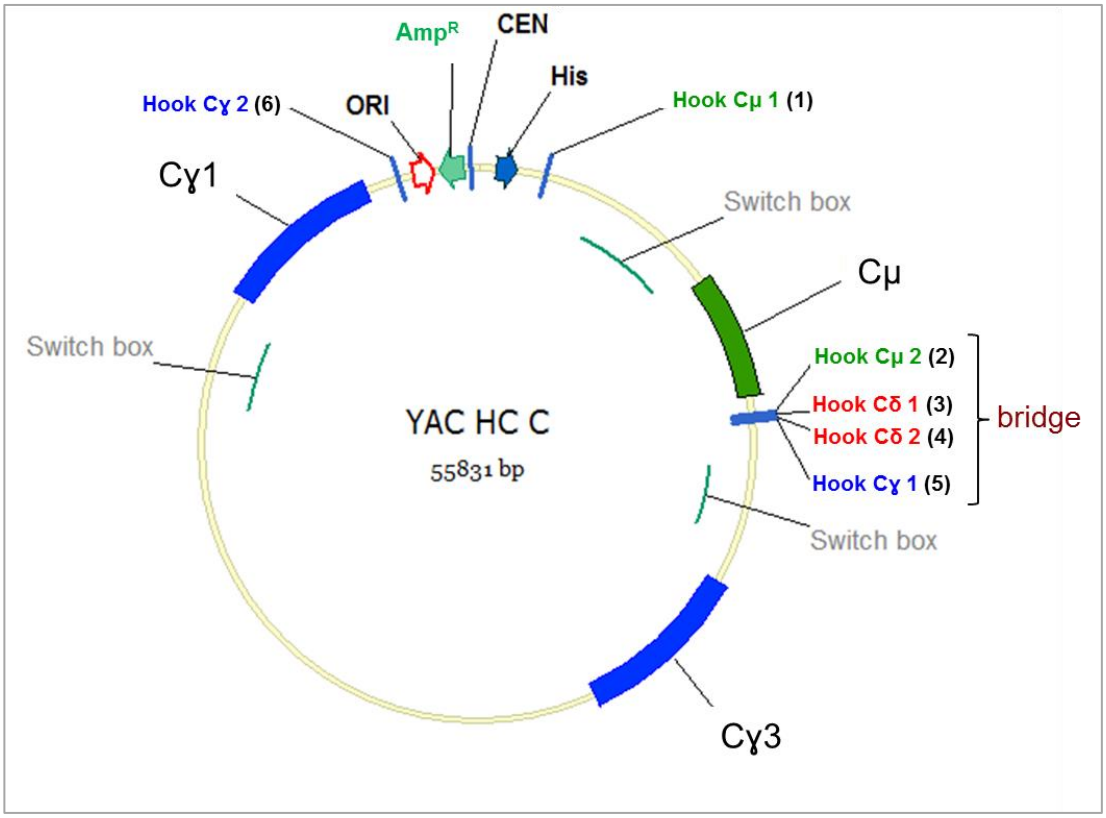

Figure 59: The YAC HC C, 55.8 kbp. The YAC contains Cμ, Cɣ3 and Cɣ1 and the required elements for class switching (switch box). The bridge is composed of four hook sequences: the hook matching on 3' of Cμ, the two hooks initially designed to clone Cδ, and finally, the hook matching on 3' of Cɣ3. *ORI*: Origin of Replication, *amp*: gene coding for resistance to ampicillin as selection bacterial marker, *CEN*: a yeast centromere; *His*: the gene encoding histidine for selection in yeast. Exons coding for each domain are not shown.

3.2.3.4. Regulatory elements

The intronic IgH enhancer and the IgH 3' enhancer are essential for the genetic modulation of IgH *locus* DNA recombination and transcription [87]. In the natural IgH *locus*, the intronic enhancer is located around 1 kbp in 3' of J6 and 1,5 kbp in 5' of the switch box preceding Cμ. From a size of 1.5 kbp, only the central portion of 350 bp, the intronic core enhancer, is sufficient to assume full function [88]. The 3' enhancer is located 25 kbp 3' of Cα2 on the 3' end of the IgH *locus*. The enhancer activity is position-, distance- and orientation-independent. Multiple protein binding sites are present in these regulatory sequences, which can activate or repress transcription, providing numerous possibilities for subtle regulation [89]. For reasons of limited time, we decided to

synthetize the intronic core enhancer and the 3' enhancer elements. The human DNA sequences were obtained from the NCBI data bank. The DNA synthesis was done by an external company (GeneArt® Gene Synthesis) and both enhancer elements were separately cloned into two independent plasmids. Using specific primers we were able to amplify them by PCR and to sub-clone them in reduced human HC constructs keeping them in a similar location as they are in the native *locus*.

3.2.4. Intermediate constructs

The HC V-D-J construct 1 (**Figure 54**) is the genetic core construction planned to generate the more complex HC Minilocus. In order to evaluate the functionality of this core construction, we cloned several version of human Cɣ1 in 3' of the V, D and J elements already assembled. To achieve this goal we engineered two minimized constructs:

- the HC Microlocus Light, which was designed to be the simplest functional human Ig *locus* allowing the production of antibodies. Therefore we inserted into HC V-D-J construct 1, the human Cɣ1 exon lacking the first constant exon CH1 (CH1$^-$). This may allow producing antibodies independently of Ig LC (Heavy Chain only Abs) [90].

- the HC Microlocus Classic, which is based on the HC Microlocus Light approach. The Cɣ1 CH1 exon was added to obtain a complete Cɣ1 version, which should lead to classical human LC dependent antibodies.

The "micro" prefix is because these two constructs contain only 7 V segments, among the 13 selected for the final construct HC Minilocus. The suffixes "light" and "classic" respectively refer to Cɣ1 CH1$^-$ and to Cɣ1.

3.2.4.1. HC Microlocus Light

To the segments V, D and J already assembled in HC V-D-J construct 1, the following elements were added:

- the HC intronic core enhancer,
- the Cɣ1 CH1$^-$ exon (*i.e.* Cɣ1 hinge, CH2, CH3, S, TM1 and TM2 exons),
- the HC 3' Enhancer, and
- the gene Neo^R coding for resistance to neomycin (under promoter HSVtk), as selection marker in mammalian cells.

As described earlier, both enhancer elements were amplified by PCR from synthetic plasmids. Cɣ1 CH1$^-$ was amplified by PCR from the BAC 448N5. For each PCR amplification, we used specific primers including tails, which contain appropriate restrictions sites for cloning directly in the HC V-D-J construct 1. The original vector backbone of HC V-D-J construct 1 was changed to that of pVC604. The resulting construct, HC Microlocus Light is shown in **Figure 60**. This construct was fully sequenced.

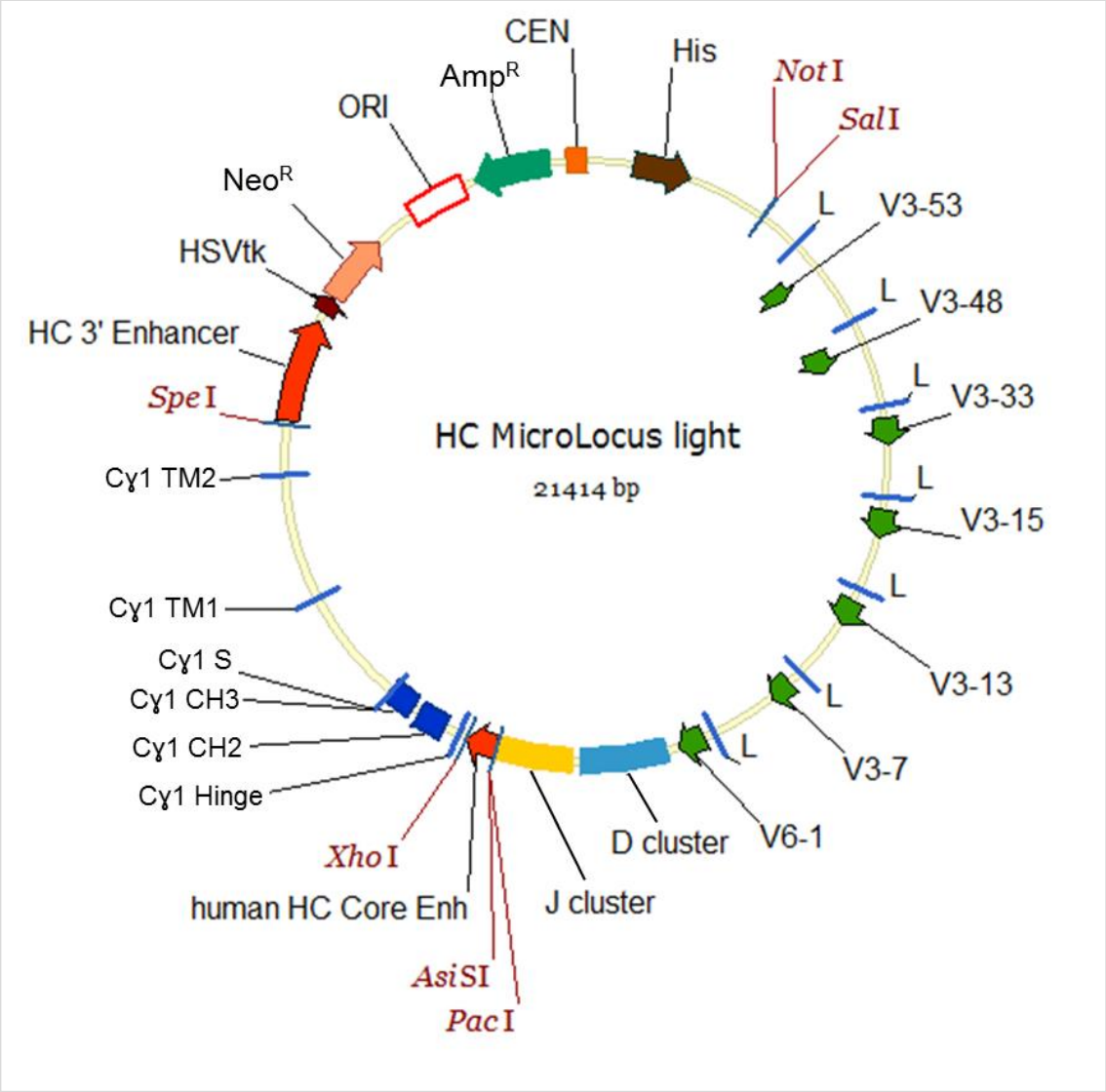

Figure 60: The HC Microlocus Light, 21,4 kbp. The construct contains seven V segments (*green arrows*) and their corresponding leader sequences (L), the D (*light blue box*) and J (*yellow box*) clusters, the human intronic core enhancer (HC Core Enhancer), the Cγ1 CH1$^-$ constant exon composed of the Cγ1 hinge, CH2, CH3, S, TM1 and TM2 exons (*dark blue boxes)*, and finally the 3' enhancer (*HC 3' Enhancer*). All sequences are human. *Not*I, *Sal*I, *Pac*I, *Asi*SI, *Xho*I *and Spe*I indicate unique restriction site recognized by enzymes of the same name. *Neo$^{R:}$* neomycin resistance gene, for selection in mammalian cells (whose expression is regulated by HSVtk promoter); *ORI*: Origin of Replication, *Amp^R:* gene coding for resistance to ampicillin as selection bacterial marker, *CEN*: a yeast centromere; *His*: the gene encoding histidine for selection in yeast.

3.2.4.2. HC Microlocus Classic

In order to have the full Cɣ1 version, Cɣ1 CH1 exon was amplified by PCR from BAC 448N5, with specific primers containing the appropriate restrictions sites. Forward primer contained *Xho*I restriction site and the reverse primer the *Sal*I restriction site. After digestion with both enzymes, Cɣ1 CH1 was ligated in 5' of Cɣ1 $CH1^-$ after *Xho*I digestion of HC Microlocus Light. *Xho*I and *Sal*I can be ligated together by dint of cohesive ends. The resulting HC Microlocus Classic is shown in **Figure 61**. This construct was fully sequenced.

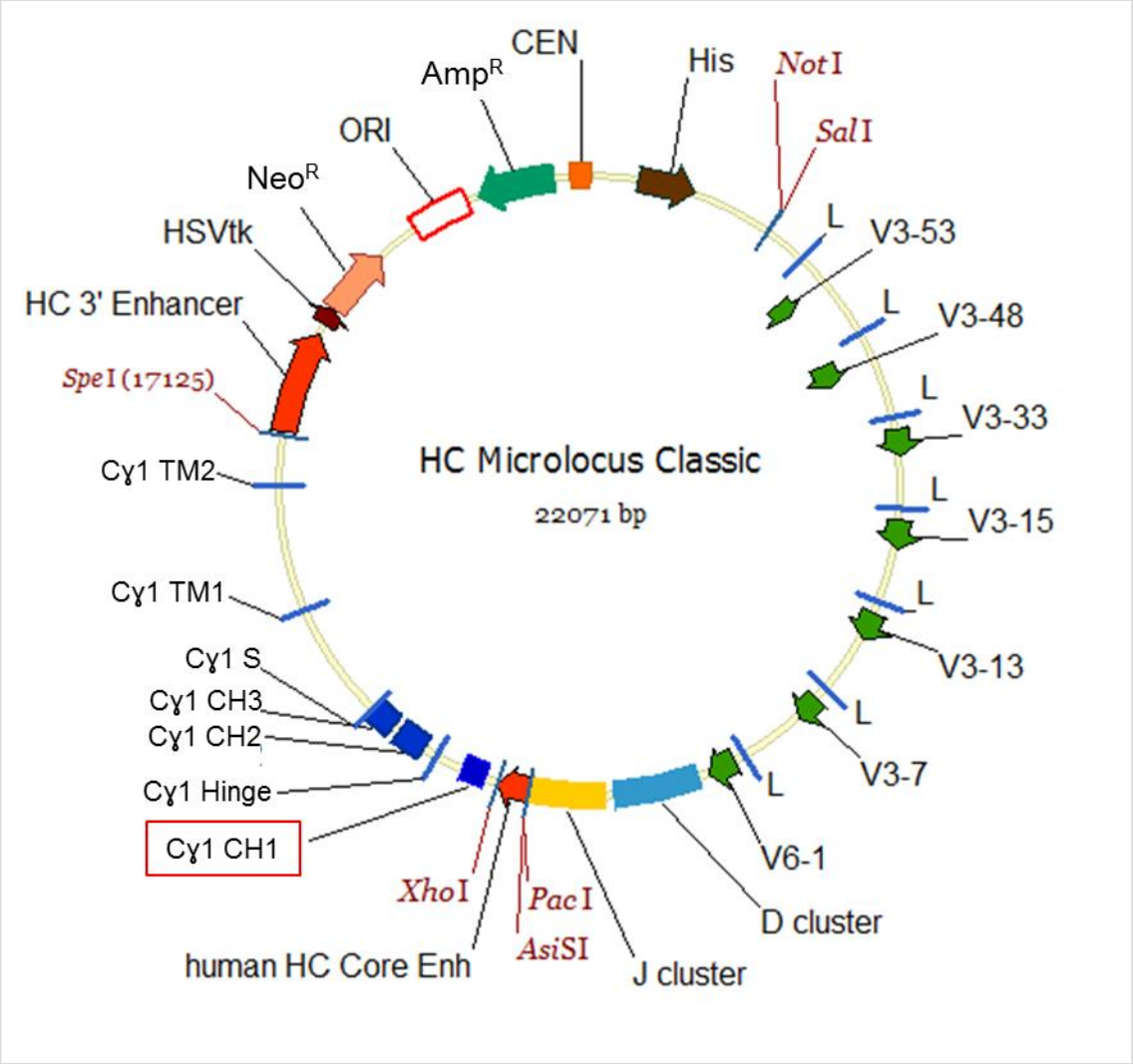

Figure 61: The HC Microlocus Classic, 22 kbp. This second intermediate construct is derived from the HC Microlocus Light by addition of the Cγ1 CH1 exon (*red box*). All other elements are strictly the same as those composing the HC Microlocus Light. The construct contains seven V segments (*green arrows*) and their corresponding leader sequences (*L*), the D (*light blue box*) and J (*yellow box*) clusters, the human intronic core enhancer (*HC Core Enhancer*), the Cγ1 constant exon composed of the Cγ1 CH1, hinge, CH2, CH3, S, TM1 and TM2 exons (*dark blue boxes)*, and finally the 3' enhancer (*HC 3' Enhancer*). All sequences are human. *Not*I, *Sal*I, *Pac*I, *Asi*SI, *Xho*I *and Spe*I indicate unique restriction site recognized by enzymes of the same name. *Neo$^{R:}$* neomycin resistance gene, for selection in mammalian cells (whose expression is regulated by HSVtk promoter); *ORI*: Origin of Replication, *Amp^R:* gene coding for resistance to ampicillin as selection bacterial marker, *CEN*: a yeast centromere; *His*: the gene encoding histidine for selection in yeast.

3.2.4.3. HC Microlocus Light shRNA

Before having the possibility to use murine HC KO mice, we had the opportunity to inject one of our construct into the oocytes of wild type (WT) mice. In order to promote the expression of HC construct in WT background, we planned to down regulate endogenous mouse IgM expression using RNAi. RNA interference (RNAi) is a natural process through which expression of a targeted gene can be down-regulated with high specificity and selectivity [91]. Down regulation of genes involve short hairpin RNA (shRNA) specific repression. The mechanism of shRNA mediated gene repression is based on the sequence-specific degradation of the host mRNA through the cytoplasmic delivery of double stranded RNA (dsRNA) identical to the target sequence [92]. Degradation of target gene expression is achieved through an enzymatic pathway involving the endogenous RNA induced silencing complex (RISC). One strand of the shRNA duplex (the antisense strand) is loaded into the RISC with the assistance of Argonaute (Ago) proteins and double-stranded RNA-binding proteins. The other strand of shRNA duplex (sense strand) is degraded. The RISC then localizes the antisense strand to the complementary mRNA molecule, which is subsequently cleaved by Ago and further degraded by other endogenous nucleases. Thus, mRNA loses its ability to be translated, leading to a down regulation of the product encoded by the initially targeted gene.

Expression of shRNA in WT mice was accomplished by cloning, into the HC Microlocus light, a DNA sequence coding for the specific shRNA. The expression of this shRNA was regulated by an independent promoter. DNA sequences encoding shRNA were determined to be homologous to the mRNA resulting from the transcription of mouse HC Cμ gene product. We designed a DNA sequence coding for four shRNA targeting mouse Cμ transcript. The shRNA complementary targets are four 25 bp regions taking place along the selected mRNA, targeting e mouse Cμ CH1, CH2, CH3 and CH4 (**Figure 62**). The transcription of the shRNA part is regulated by the promoter 7SK, a RNA polymerase III promoter, routinely used to express shRNA molecules. Once transcribed in the nucleus by polymerase III, homologous sense and antisense RNA pieces will combine to form pre-shRNA. The resulting pre-shRNA is exported from the nucleus and is then processed to become shRNA duplex, whose

antisense strand directs RISC to targeted mRNA (**Figure 63**). These processing are supported by endogenous RNAses III. The whole DNA sequence, including promoter and sequences encoding the four shRNA, was synthetized by an external company (GeneArt® Gene Synthesis) and cloned in the HC Microlocus Light construct, in 5' of the V cluster. The resulting construct, HC Microlocus Light shRNA (**Figure 64**), is the third intermediate construct, specifically designed for the injection in WT mice oocytes.

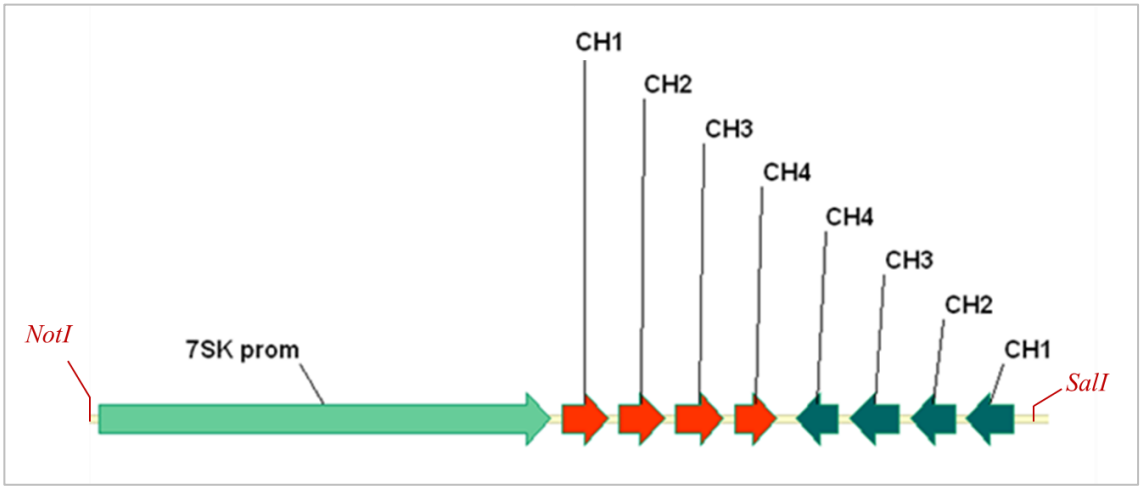

Figure 62: The p7SK shRNA targeting mouse IgM transcripts, 500 bp. Expression of double stranded RNA is regulated by 7 SK promoter (7SK prom). Red arrows code for the sense strand of shRNA duplex, while dark green arrows encode the antisense strand. Each resulting shRNA duplex targets one of the four constant exons encoded in mouse IgM transcript (*CH1*, *CH2*, *CH3* and *CH4*). *Not*I and *Sal*I indicate unique restriction sites recognized by enzymes of the same name; they served to clone this fragment in HC Microlocus Light.

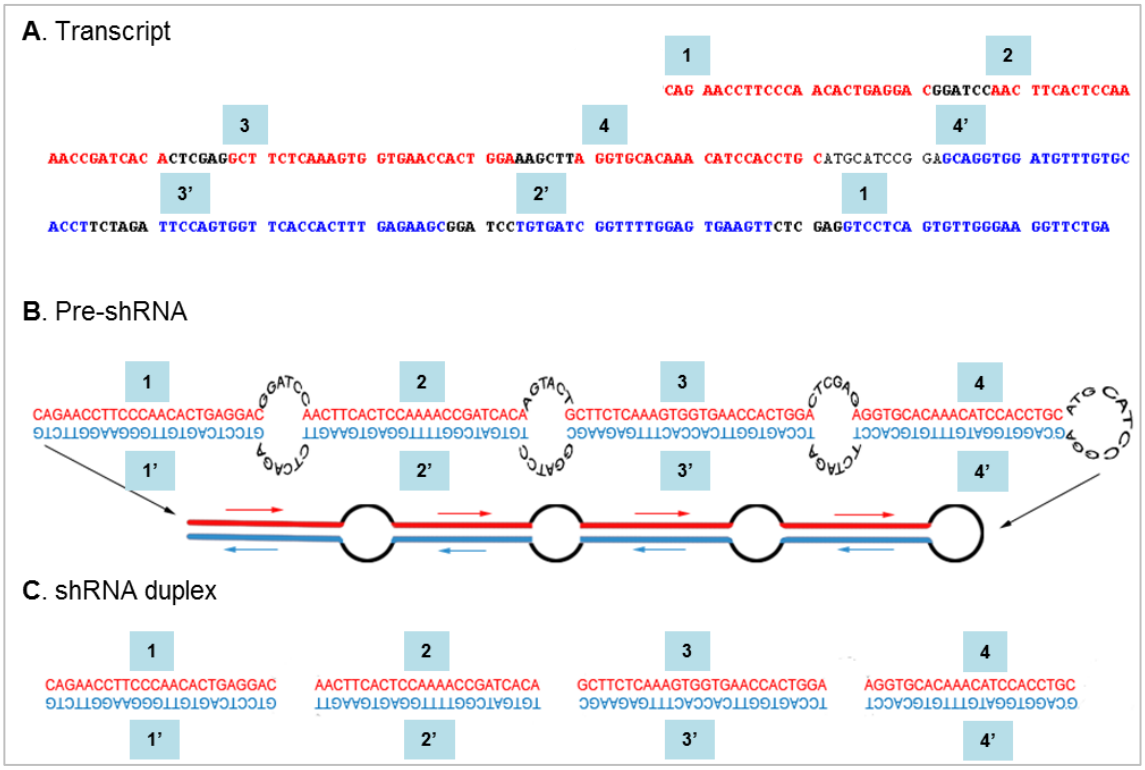

Figure 63: The shRNA processing.

A: primary transcript of the DNA construct coding for shRNA.

B: Pre-shRNA, consisting in association of sense and antisense RNA segments.

C: shRNA duplex.

On these three panels, sense strands are stamped with numbers 1, 2, 3 and 4; while antisense strands are stamped with numbers 1', 2', 3', and 4'.

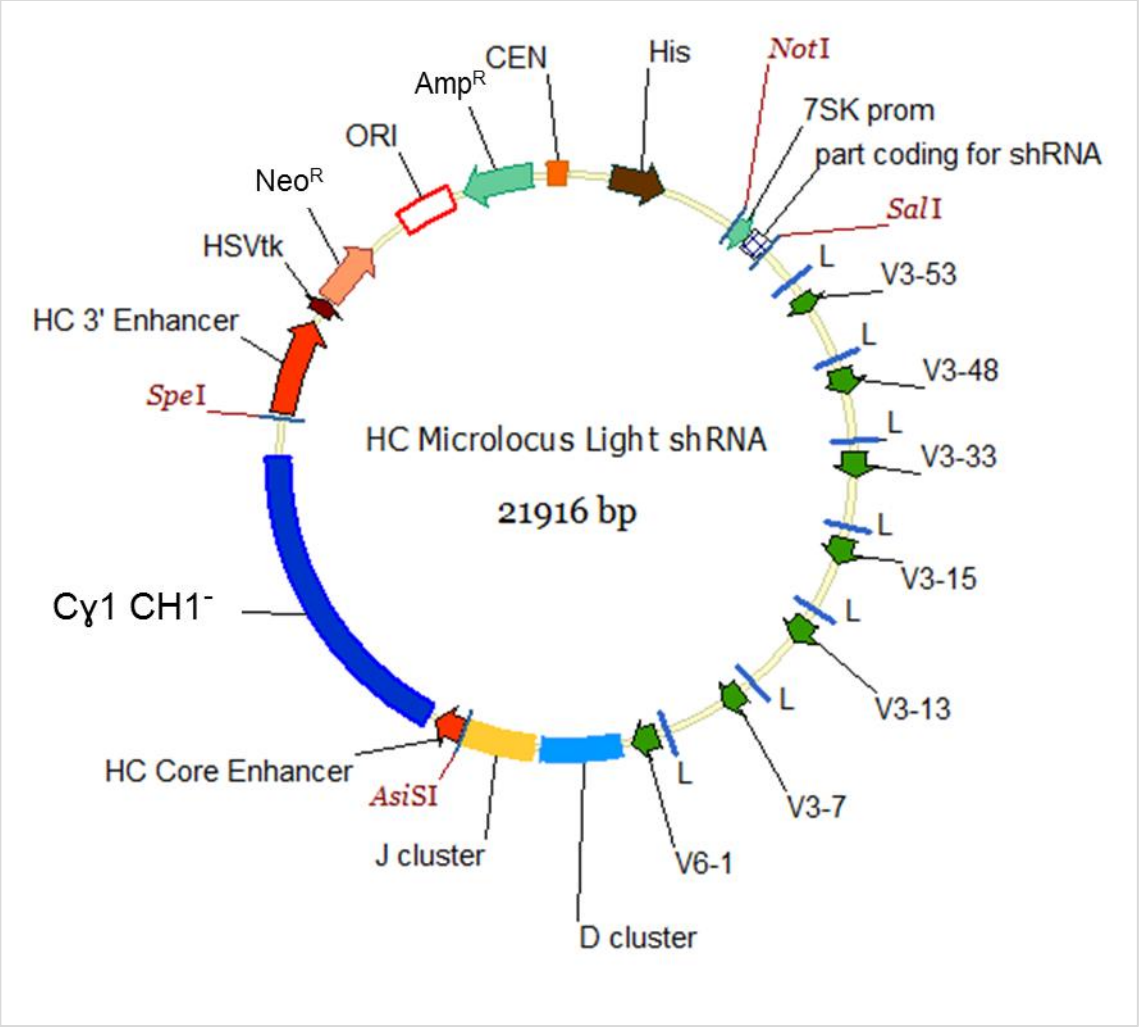

Figure 64: HC Microlocus Light shRNA, 22 kbp. This third intermediate construct is derived from the HC Microlocus light by addition of an independently regulated region coding for four shRNA (part coding for shRNA) targeting mouse IgM transcript. The promoter 7SK (7SK prom) regulates shRNA transcription. All other elements are strictly the same as those composing the HC Microlocus light: seven V segments (*green arrows*) and their corresponding leader sequences (*L*), the D (*light blue box*) and J (*yellow box*) clusters, the human intronic core enhancer (*HC Core Enhancer*), the constant part Cγ1 CH1⁻ (*dark blue box*) composed of Cγ1 Hinge, CH2, CH3 and coding parts for secreted and transmembrane Ig versions (to simplify the scheme, these exons are not shown), and finally the 3' enhancer (*HC 3' Enhancer*). All sequences are human. *NotI*, *SalI*, *AsiSI* and *SpeI* indicate unique restriction site recognized by enzymes of the same name. *Neo^R:* neomycin resistance gene, for selection in mammalian cells (whose expression is regulated by HSVtk promoter); *ORI*: Origin of Replication, *Amp^R:* gene coding for resistance to ampicillin as selection bacterial marker, *CEN*: a yeast centromere; *His*: the gene encoding histidine for selection in yeast.

3.2.5. HC Minilocus

The HC Minilocus YAC contains the 13 selected V segments, the synthetized D and J clusters and the genes coding for full version of Cμ, Cɣ3 and Cɣ1, including switch boxes and transmembrane parts. All these elements were carefully assembled. First, selected V, D and J segments were assembled in a unique construct. In a second step, the constant parts Cμ, Cɣ3 and Cɣ1 were combined with all V, D and J elements.

3.2.5.1. Assembling V, D and J regions

HC V-D-J construct 1 (**Figure 54**) contains the five selected V segments from family 3, the unique V6-1, and the synthetic D and J clusters. The HC V construct 2 (**Figure 55**) contains six V segments from families 1, 2 and 4. The V-D-J region of HC Minilocus is based on the combination of all elements previously mentioned. HC V-D-J construct 1 and HC V construct 2 have specific restriction sites, allowing insertion of the V cluster contained in construct 2 into construct 1. First, the V cluster of HC V construct 2 was isolated using specific restrictions enzymes. After restriction enzyme digestion, DNA fragments were separated by gel electrophoresis and the fragment of interest, *i.e.* the fragment containing the V cluster, was purified from gel, using QIAquick Gel Extraction Kit (*ref*: 28704). Purified DNA fragment was then ligated in previously linearized HC V-D-J construct 1. The combination procedure is described with the schema presented in **Figure 65**. The resulting cassette, final HC V-D-J construct, is schematized in **Figure 66**.

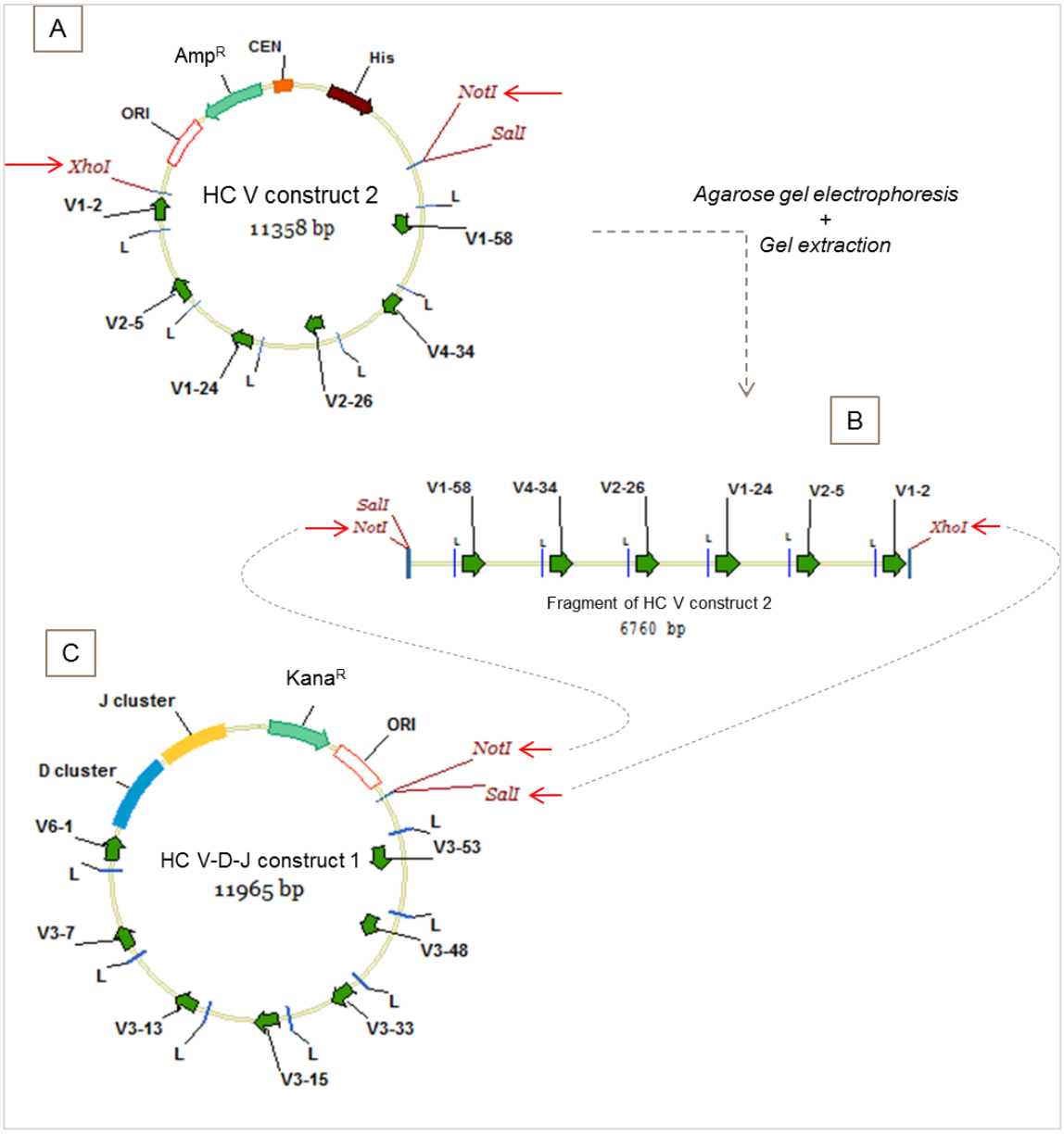

Figure 65: strategy used to combine all selected V, D and J elements into a single construct.

A: The V cluster contained in HC V construct 2 was isolated by digestion with enzymes *Not*I and *Xho*I. Both enzymes recognize restriction site with the same name (*red arrows*) which are unique in HC V construct 2. After digestion, two fragments were obtained.

B: After fragments separation using gel electrophoresis, the fragment containing the 6 V segments (6 760 bp) was extracted from the gel.

C: The resulting DNA fragment was ligated in HC V-D-J construct 1, which was beforehand digested with *Not*I and *Sal*I (*lower panel*). Because of cohesive ends, *Sal*I and *Xho*I can be ligated together.

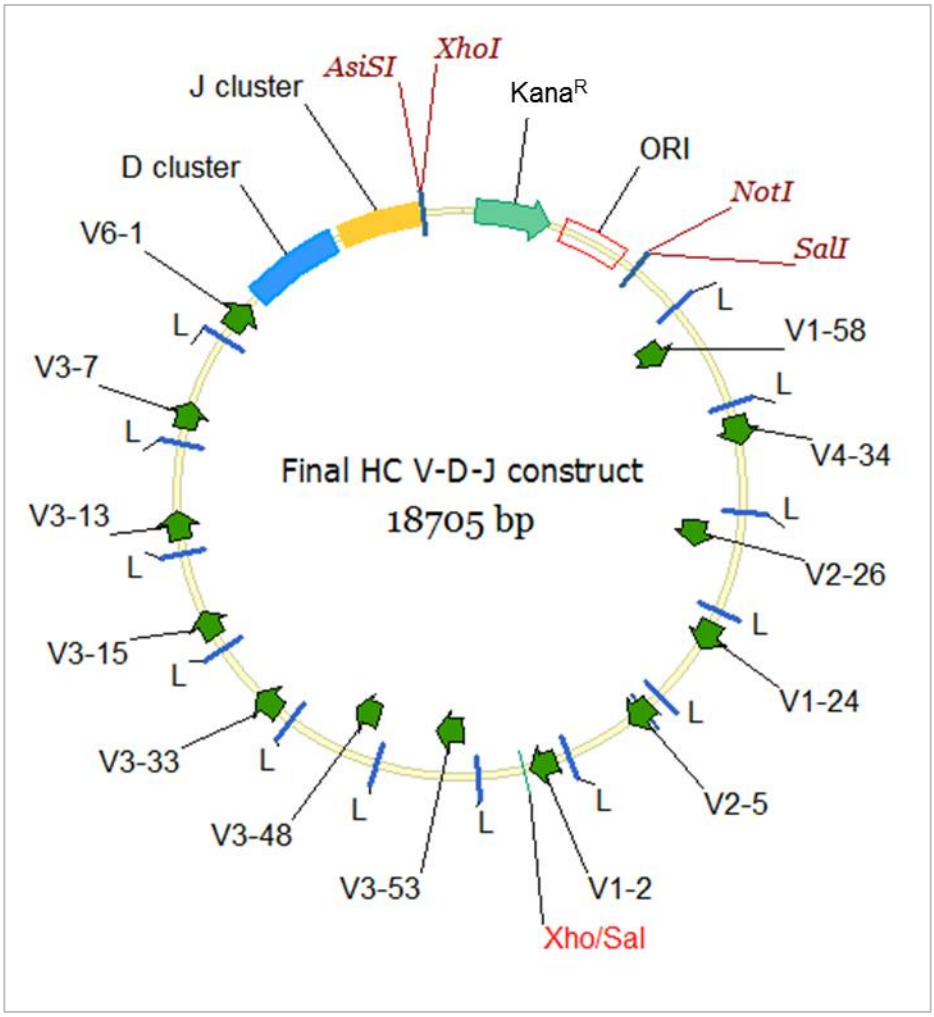

Figure 66: Final HC V-D-J construct, 18.7 kbp. This construct contains the 13 selected V segments: the 7 V segments originally cloned in HC V-D-J construct 1 (in 3' of Xho/Sal) and the 6 V segments originally assembled in HC V construct 2 (in 5' of Xho/Sal). *Green boxes*: V segments and corresponding leader sequence (*L*); *blue box*: D cluster; *yellow box*: J cluster; *ORI*: Origin of Replication; *Kana*^{R}: gene coding for resistance to kanamycin as selection bacterial marker. *Asi*SI, *Xho*I, *Not*I and *Sal*I indicate unique restriction site recognized by enzymes of the same name; they serve for further cloning step. Xho/Sal shows the site of ligation between *Xho*I and *Sal*I cohesive ends.

3.2.5.2. Assembling V and C parts by TAR

The constant parts Cμ, Cγ3 and Cγ1 were already cloned in YAC HC C (see 3.2.3.3. and **Figure 59**). Initially, we wanted to retrofit this YAC as BAC in bacteria. This would have allowed simple combination of constant parts with final HC V-D-J construct.

However, we did not succeed with this approach because of YAC DNA breakage and uncontrolled recombination.

An alternative strategy consisting of combining by homologous recombination in yeast the final HC V-D-J construct (see below) with the Ig constant regions cloned in YAC was performed. To achieve this combination, specific hooks were designed and cloned in the final HC V-D-J construct. One hook is homologous to 5' of Cμ and the other one is homologous to 3' of Cγ1 of the cloned YAC HC C.

Thus, the final HC V-D-J construct (**Figure 66**) was used as TAR vector. Several elements have been added:

- the HC 3' Enhancer;
- the HC intronic core enhancer;
- the gene *KanMX4* coding for resistance to geneticin (under TEF promoter), as selection marker in yeast cells;
- the gene Neo^R coding for resistance to neomycin (under promoter HSVtk), as selection marker in mammalian cells;
- the gene *mCherry* coding for the fluorophore mCherry (a red fluorescent protein) [93], under CMV promoter. After mammalian cell transformation, mCherry can be excited at 590 nm, and transformed cells will emit a red light at 610 nm; and,
- two hooks sequences: Hook HC C 1 and Hook HC C 2. These both hook sequences allow homologous recombination of final HC V-D-J construct into YAC HC C.

All these elements were amplified from synthetic plasmids, using specific primers; and they were assembled using standard molecular biology methods.

The resulting TAR vector, final HC V-D-J construct _HOOKS, was linearized using unique restriction site between both hook sequences. Following the TAR procedure described earlier (see 3.1.2.), yeasts containing YAC HC C were treated with zymoliase to become spheroplasts, which are permeable to DNA. Mixed with the linearized vector final HC V-D-J construct _HOOKS and PEG, the spheroplasts were able to absorb the vector. The yeast then proceeded to recombination between the vector final HC V-D-J construct _HOOKS and the YAC HC C. These recombinations, schematized in **Figure 67,** led to YAC HC Minilocus formation (**Figure 68**).

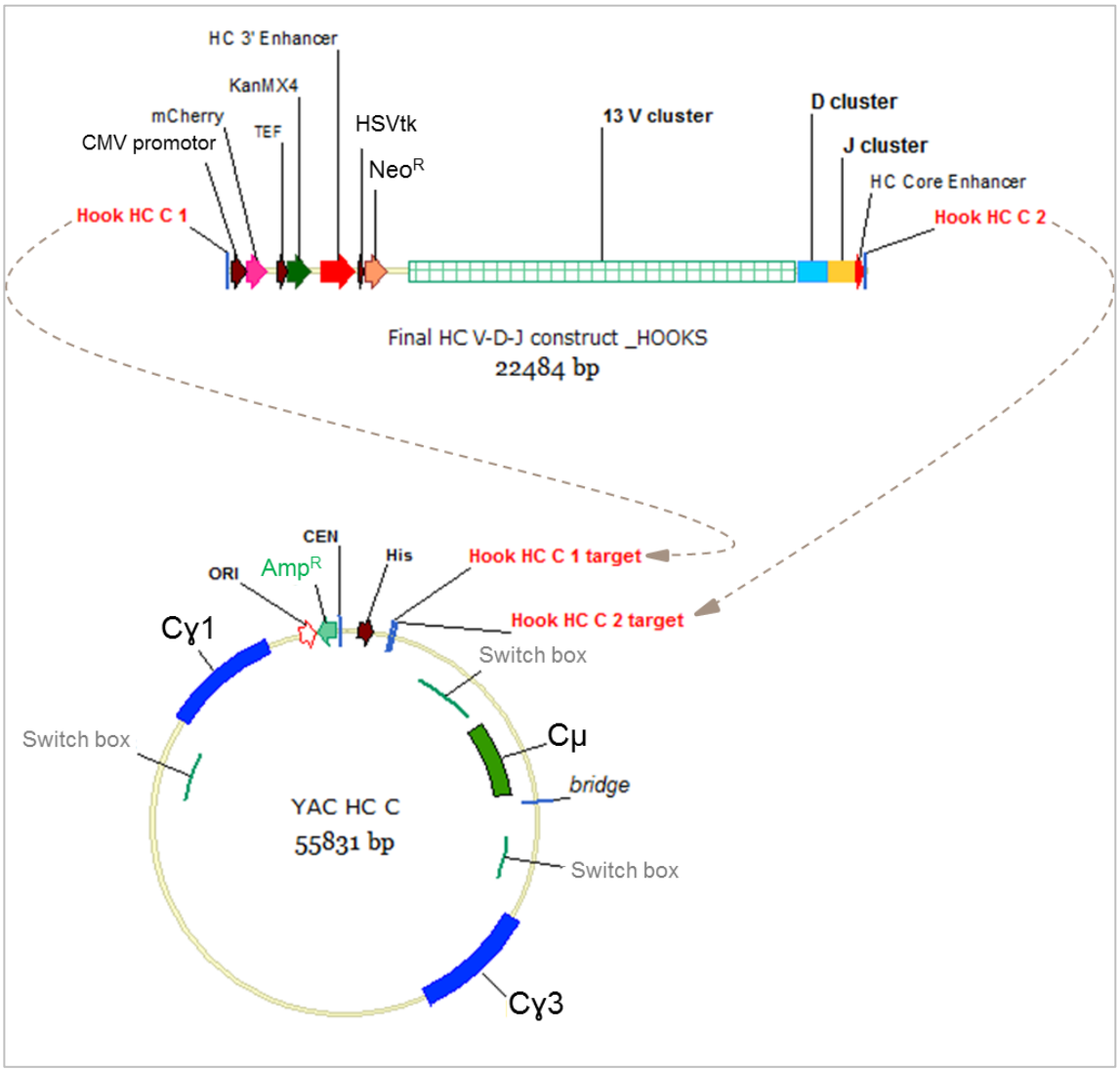

Figure 67: HC Minilocus assembly *via* TAR cloning. Final HC V-D-J construct _HOOKS contains all selected V (*green grid*), D (*light blue*) and J (*yellow*) segments; both enhancer elements (HC 3' Enhancer and HC Core Enhancer); *mCherry* gene coding for red fluorescent protein, whose expression in mammalian cells is regulated by CMV promoter; the *KanMX4* resistance gene as selection marker in yeast cells and the *Neo*^{R} resistance gene as selection mammalian marker. YAC HC C contains Cμ, Cγ3 and Cγ1 and the required elements for class switching (*switch box*); and the gene encoding histidine for selection in yeast cells (*His*).
The TAR vector final HC V-D-J construct _HOOKS was linearized by digestion with *Sse231*I (not shown), which cut uniquely between the two hook sequences (*upper panel*). Yeasts containing YAC HC C were transformed with final HC V-D-J construct _HOOKS. Therefore, Hook HC C 1 and Hook HC C 2 recombined with their homologous sequences in the YAC HC C (Hook HC 1 target and Hook HC C 2 target).

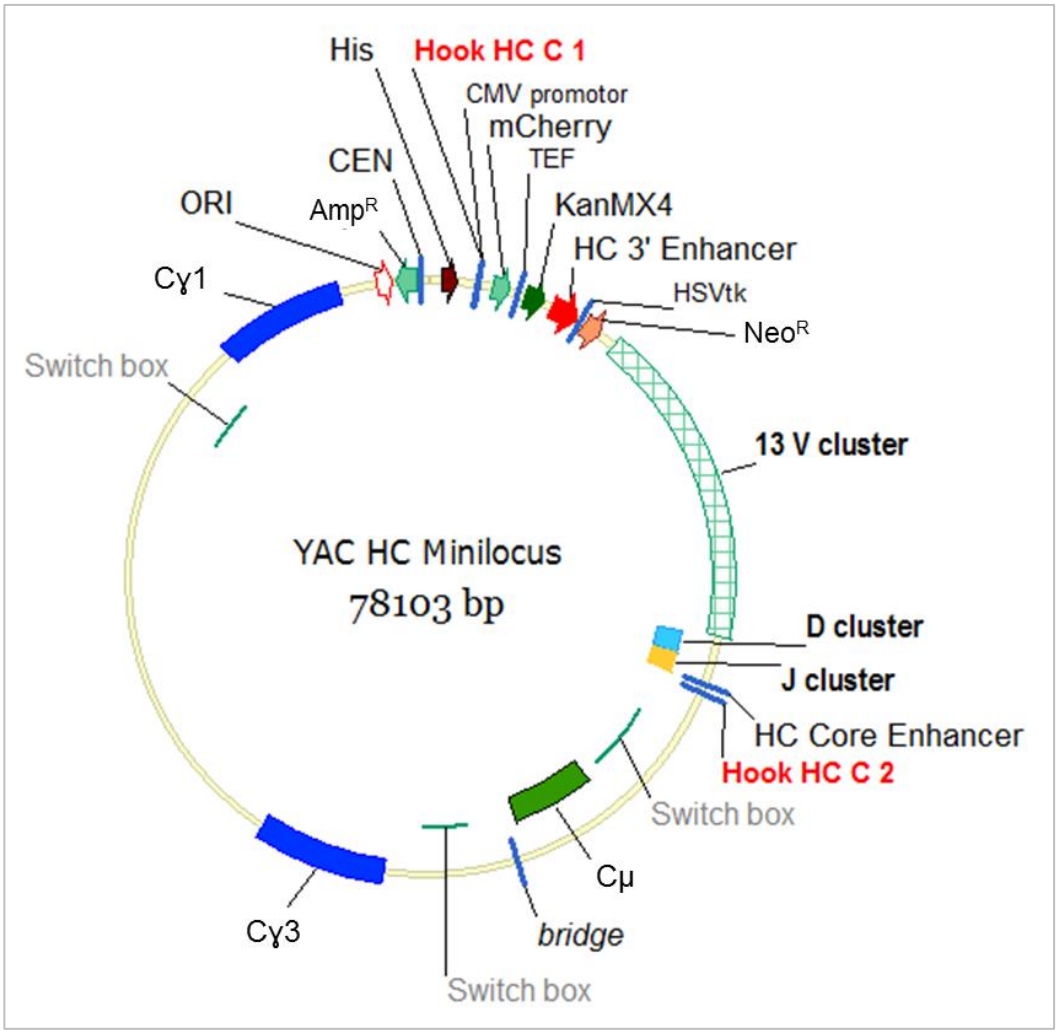

Figure 68: The HC Minilocus, 78 kbp. This construct contains the 13 selected V (*green grid*), the D (*light blue*) and J (*yellow*) clusters; both enhancer elements (*HC Core Enhancer* and *HC 3' Enhancer*); Cμ, Cγ3 and Cγ1 and the required elements for class switching (*switch box*); *His* gene encoding histidine for selection in yeast cells, *mCherry* gene coding for red fluorescent protein, whose expression in mammalian cells is regulated by *CMV promoter*; the *KanMX4* resistance gene as selection marker in yeast cells (whose expression is regulated by *TEF* promoter) and the *Neo*^{R} resistance gene as selection mammalian marker (whose expression is regulated by *HSVtk* promoter).

After TAR cloning, positive clones have grown in selection medium without histidine (selection marker *His* borne by YAC HC C) and with geneticin (selection marker *KanMX4* brought by Final HC V-D-J construct _HOOKS). Selected clones were intensively genotyped (**Figure 69**).

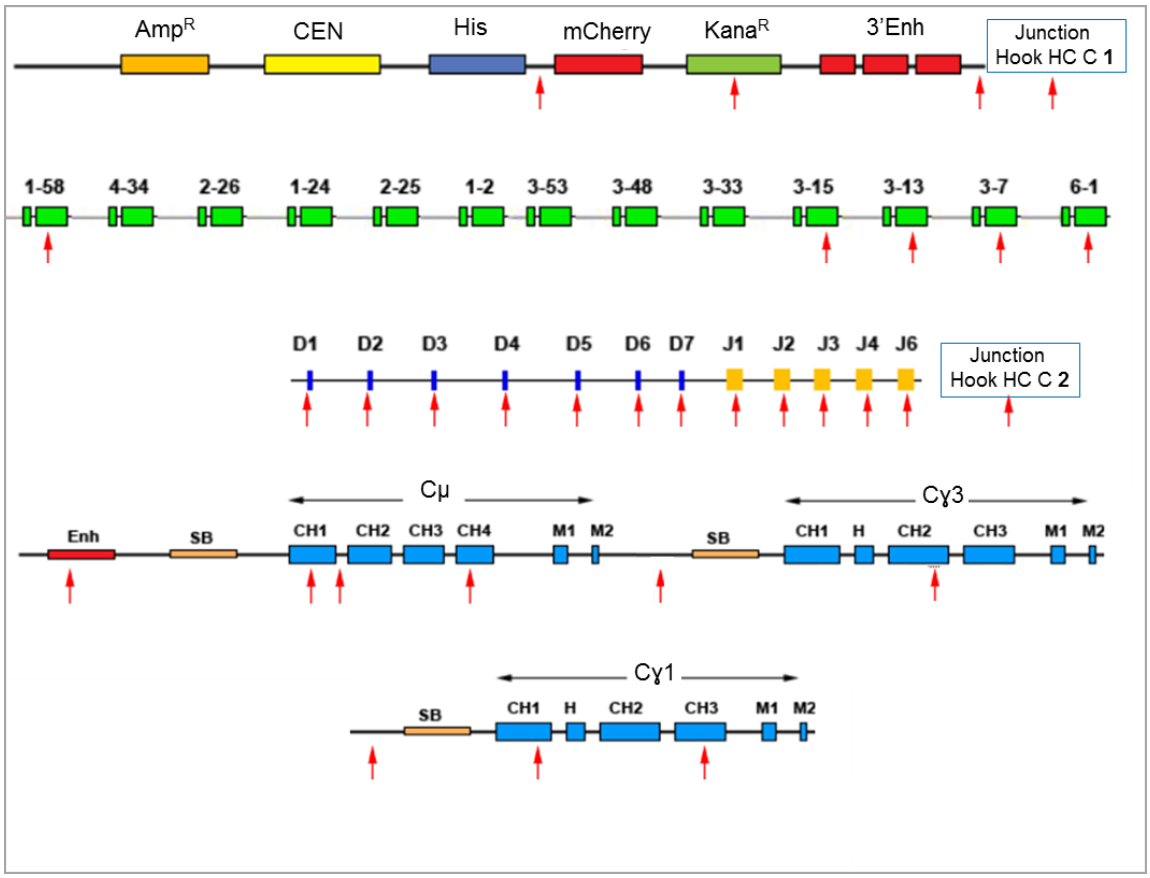

Figure 69: Genotyping of the HC Minilocus clone #H2. Green boxes stand for V segments and its corresponding leader sequence; dark blue sticks schematize D segments, while J segments are shown with yellow boxes; blue boxes stand for exons coding for different constant domains. In 5' of Cμ, Cɣ3 and Cɣ1, beige boxes stand for switch box sequences (SB). Intronic and 3' Enhancers are schematized with red boxes. Each red arrow shows the position of positive PCR hit.

3.2.6. Construct collection

The HC Minilocus and the three intermediate constructs, which are summarized in **Figure 70**, were analyzed for their functionality.

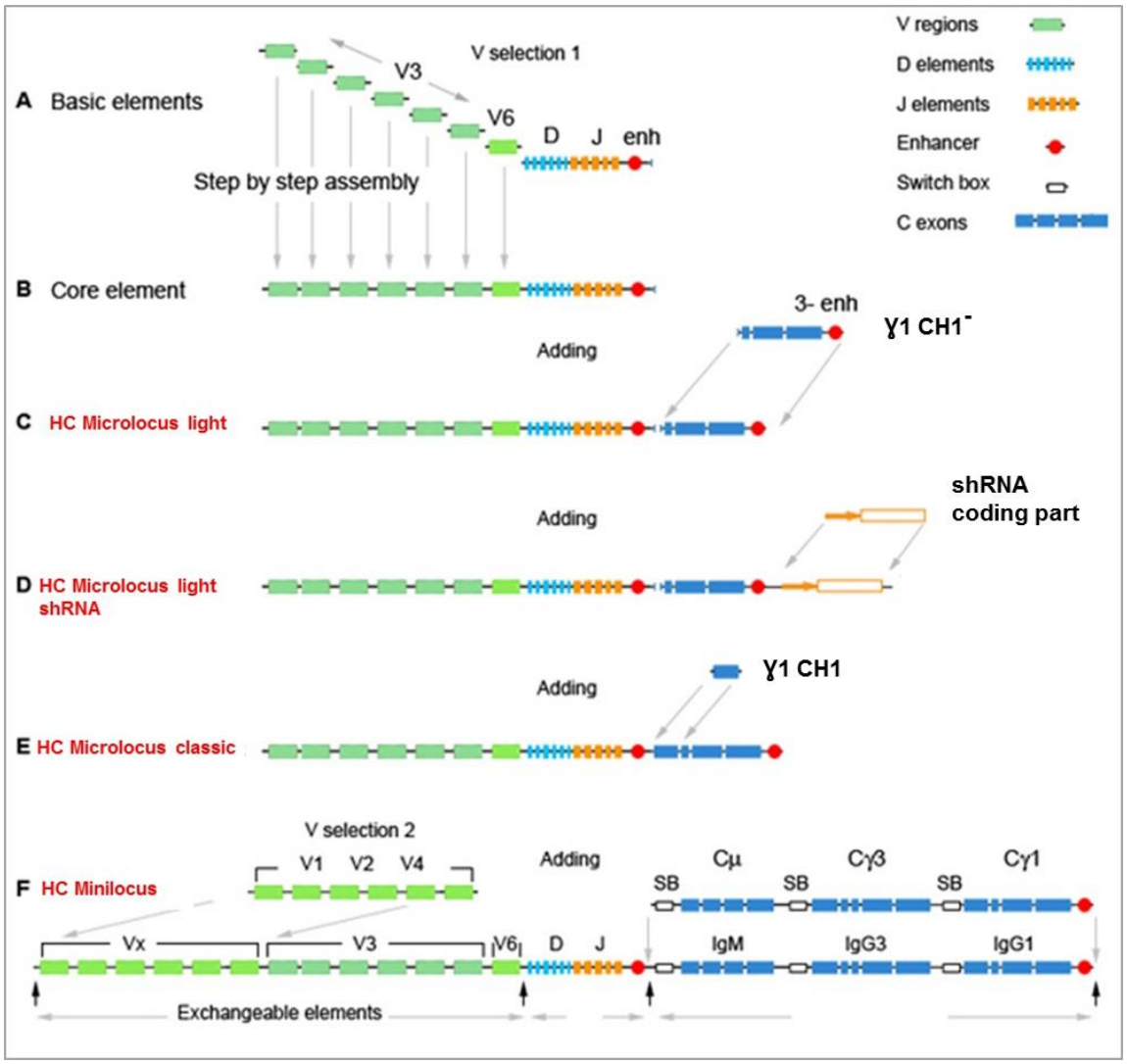

Figure 70: Summary of all human HC reduced constructs. The panels *A* and *B* represent the basic elements V, D, J and the intronic core enhancer (enh). This assembly constituted the basis of human HC constructs (*C*, *D*, *E* and *F*). Indeed, different elements were stepwise added to the core construction (*B*):

- addition of the Cγ1 CH1$^-$ and the 3'enhancer (3-enh) led to HC Microlocus light (*C*);
- addition of part coding for shRNA into the HC Microlocus light (*C*) led to HC Microlocus light shRNA (*D*);
- addition of Cγ1 CH1 exon to HC Microlocus light (*C*) led to HC Microlocus classic (*E*);
- addition of further V segments, and constant parts Cμ, Cγ3 and Cγ1 to the core construction (*B*) led to HC Minilocus (*F*).

3.3. Testing reduced Ig *loci* in the 300-19 cell line

In order to demonstrate the biologic functionality of the human HC reduced constructs described in the previous chapter, we transfected them into a mouse lymphoid cell line. The aims of 300-19 mouse cell line transfection experiments were to evaluate the following functional aspects:

- rearrangement D to J and V to DJ,
- recombinatory diversity: the gene segment usage,
- junctional diversity: N nucleotides addition,
- transcription,
- mRNA splicing,
- alternative splicing: secreted and transmembrane forms,
- expression: antibody chain production, and
- transport of the Ig HC to the cell surface.

3.3.1. 300-19 cell line

The Abelson murine leukemia virus (A-MuLV) immortalized 300-19 cell line was derived from the bone marrow of an adult NIH/Swiss outbred mouse. The original 300-19 clonal isolates carried DJ_H rearrangement on both alleles [5], which corresponds to the molecular genotype characteristic of pro-B cells. In 1985, Reth and colleagues demonstrated that such lines, when propagated in culture, can progress from pro-B cell to the mature B cell stage. Approximately 20-30% of the V_H to DJ_H joining in the 300-19 line are productive, leading to μ chain expression, and a few of them undergo Vκ assembly [6]. Thus, 300-19 cell line provides a tool to assess the functionality of human HC reduced constructs, from VDJ rearrangement to protein expression. 300-19 cells were maintained at 37°C (5% CO_2), in RPMI complete medium [94], whose formulation is given in **Table 5.**

Table 5: Formulation of RPMI complete medium.

RPMI 1640, GlutaMAX	GIBCO	61870	500mL
Fetal calf serum (FCS)	GIBCO	26010-074	25-50mL
Penicillin-Streptomycin-Glutamine (PSG 100X)	GIBCO	10378-016	5mL
2-Mercaptoethanol (50 mM)	GIBCO	31350-010	500µL
Non-Essential Amino Acids Solution (NEAA 100X)	GIBCO	11140	5mL
Sodium Pyruvate (100 mM)	GIBCO	11360	5mL

3.3.2. Transfection of 300-19 cells with reduced HC constructs

Four human HC reduced constructs (HC Microlocus Light, HC Microlocus Light shRNA, HC Microlocus Classic and HC Minilocus) were transfected into 300-19 cells by electroporation. Electroporation was performed using Nucleofector® device (Amaxa; Cat. No. AAD-1001) and the Cell line Nucleofector® Kit R (Amaxa; Cat. No. VCA-1003). The "General Protocol for Nucleofection of Suspension Cell Lines" was followed. For 300-19 cell type, the appropriate Nucleofector® program is U-015.

As the vector arm of the four HC reduced constructs contains the gene Neo^R under the control of HSVtk promoter, transfected 300-19 cells grew under G418 selection. G418, Geneticin® (*ref*: Gibco, 10131035) was used at 1 mg/mL in complete RPMI medium. An additional marker gene, *mCherry* [93], was added to the HC Minilocus. Thus, 300-19 cells transfected with the HC Minilocus emitted red light at 610 nm after excitation at 590 nm (**Figure 71**).

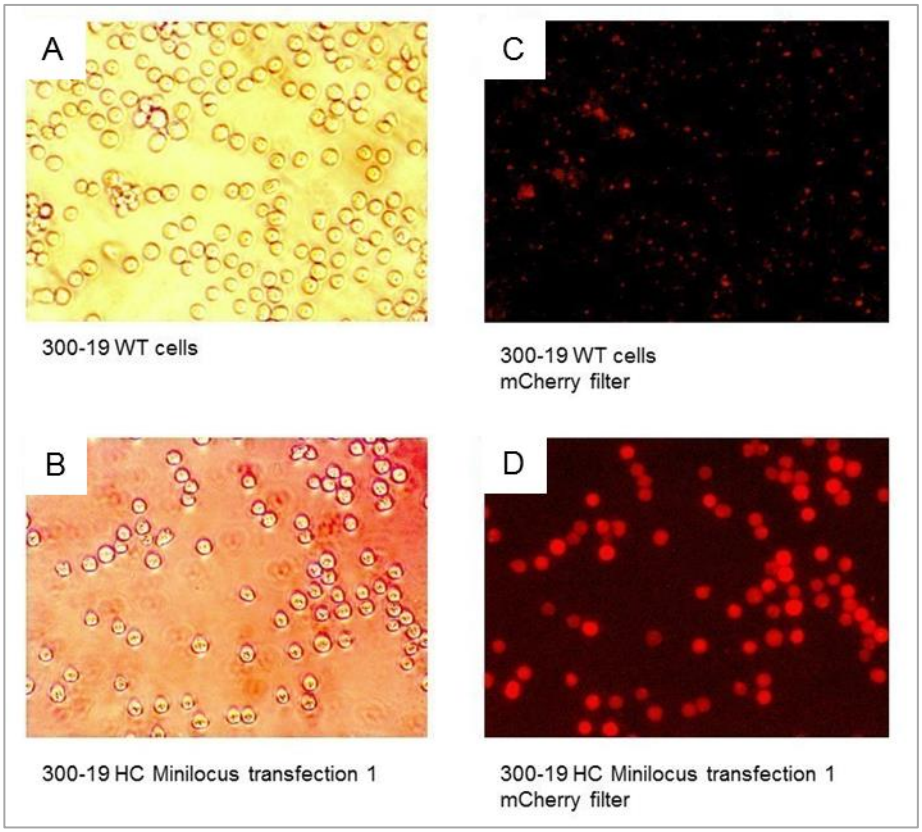

Figure 71: 300-19 HC Minilocus cells express mCherry fluorescent protein. Pictures A and B show the cells without excitation, while pictures C and D are the same images after 590 nm wavelength excitation. The 300-19 cells transfected with HC Minilocus (*panel D*) emit a red light (610 nm) when compared with 300-19 wild type cells (*panel C*). These pictures were taken two weeks after transfection (under G418 selection).

3.3.3. Human HC expression analysis by FACS

When a pre-B cell generates *via* somatic recombination a functional *VDJ* gene, *i.e.* in open reading frame (ORF), the rearranged *VDJ* gene, including the Ig constant region exons, is transcribed into mRNA which is translated into an immunoglobulin heavy chain polypeptide. Translation takes place in cytoplasm and after protein maturation in different intracellular compartments (endoplasmic reticulum and Golgi), HC proteins are exported to the cell surface, assuming that the HC is associated with a LC or a surrogate LC [95]. Surface expression of HC in pre-B cells is significantly lower than in B cells despite comparable amounts of intracellular HC in the two cell types. Non-stoichiometric amounts of the pre-BCR components, mainly λ5, account for its low

surface expression despite high HC intracellular levels [96]. Such statement leads us to consider surface as well as intracellular human ɣ1 HC expression.

After three weeks of G418 selection (2 passages per week), HC Microlocus Light, HC Microlocus Light shRNA, HC Microlocus Classic and HC Minilocus transfected 300-19 cells were analyzed by FACS. Expression of the human IgG was evaluated on cell surface as well as intracellular. For HC Minilocus transfected 300-19 cells, the expression of the human IgM was first assessed. The staining protocol is described below.

3.3.3.1. Method

For each cell transfection, 5. 10^5 cells were collected and washed two times with FACS buffer (PBS, 0.5% FCS, 2mM EDTA). Before intracellular staining, cells were fixed and permeabilized using Cytofix/Cytoperm™ from BD (ref: BD 554715). HC Microlocus light, HC Microlocus light shRNA and HC Microlocus classic transformed cells were stained with anti-human IgG Fc antibody, coupled with the fluorochrome PE (ref: eB 12-4998-82). HC Minilocus classic 300-19 cells were stained with anti-human IgM antibody, coupled with fluorochrom APC (ref: BD, 551062). As negative control, wild type 300-19 cells were stained following the same protocol. Samples were incubated 30 minutes at 4°C in dark. After washing, stained cell samples were acquired with FACS equipment LSR Fortessa. Analyses were performed with FlowJo software. Human Ig expression level was assessed on live cells based on their size (FSC parameter) and granularity (SSC parameter).

3.3.3.2. Results

After few months in culture, transformed 300-19 cells expressed intracellular human ɣ1 HC, thus suggesting that all required functions to the step of polypeptide synthesis are assumed by the genetic constructs. As expected, the surface expression was more difficult to demonstrate. Indeed, like the conventional μHC involved in pre-BCR, the human ɣ1 HC may need associated partners like the mouse surrogate light chain and the mouse heterodimer Igα/Igβ, to reach the cell surface. Such molecules may be selective or poorly expressed in 300-19 cells.

By FACS analysis of the HC Minilocus transfected cells, we faced several problems due to the strong expression of the mCherry fluorophore, which interfered with all fluorochroms used for the analysis. Therefore the FACS results for HC Minilocus transformed cells are ambiguous and they are not presented here.

Intracellular antibody chains detection

Three months after transfection, 300-19 cells were tested by FACS analysis, and intracellular human ɣ1 HC expression was detected (**Figure 72**). Between 4 and 8% of cells, transfected with different constructs, expressed a high level of human ɣ1 HC.

Surface expression

Three and an half months after the transfection of the 300-19 cells with the HC Microlocus Light, HC Microlocus Light shRNA and HC Microlocus Classic, few cells started to show surface expression of human ɣ1 HC (**Figure 73**).

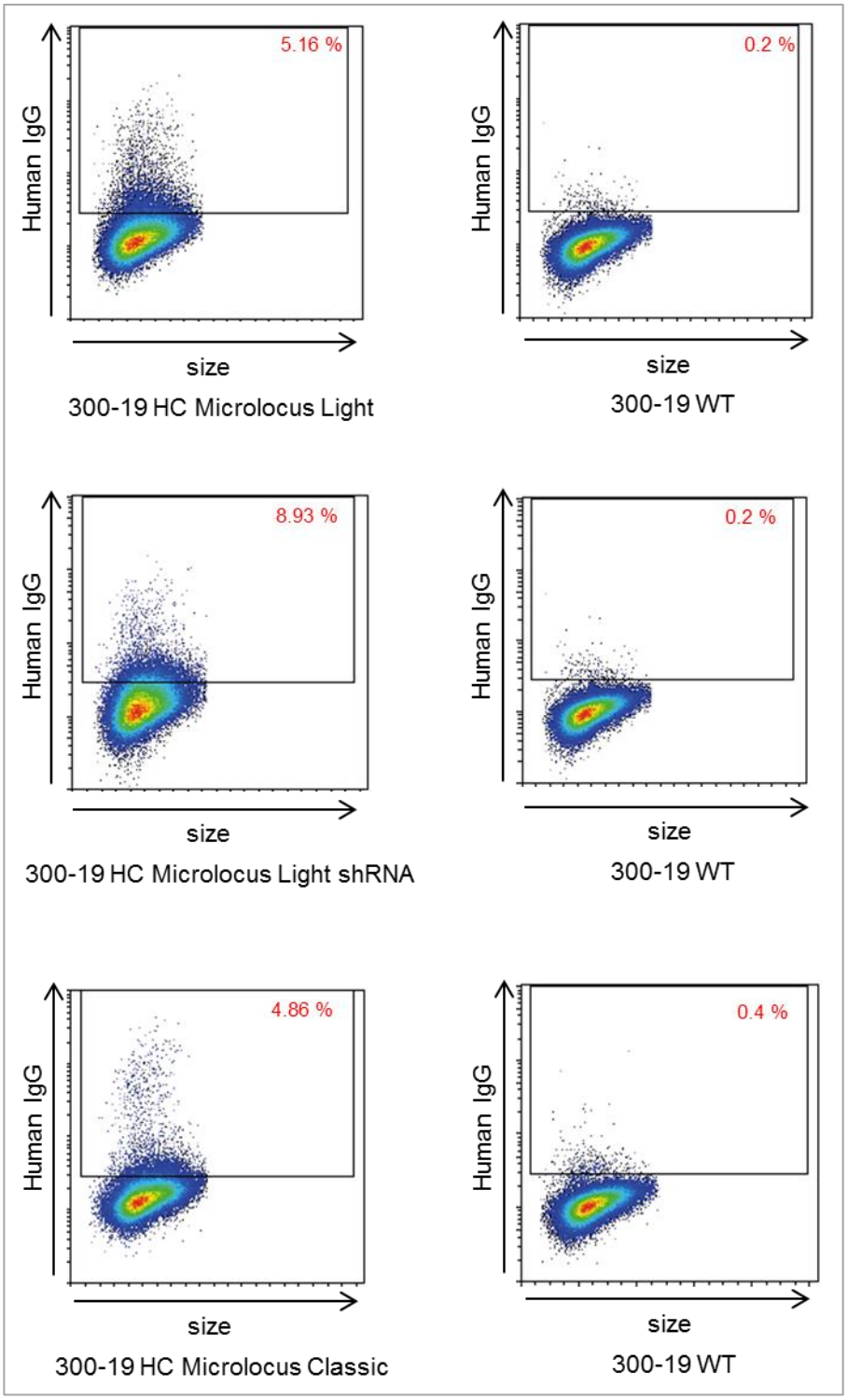

Figure 72: Intracellular detection of human ɣ1 HC, three months in culture after transfection. Transformed cells (*left panel*) are compared to wild type cells (*right panel*). Human ɣ1 HC positive cells were assessed on total cells. The percentage of positive cells is in red color, in the top right-hand corner of the dot plot.

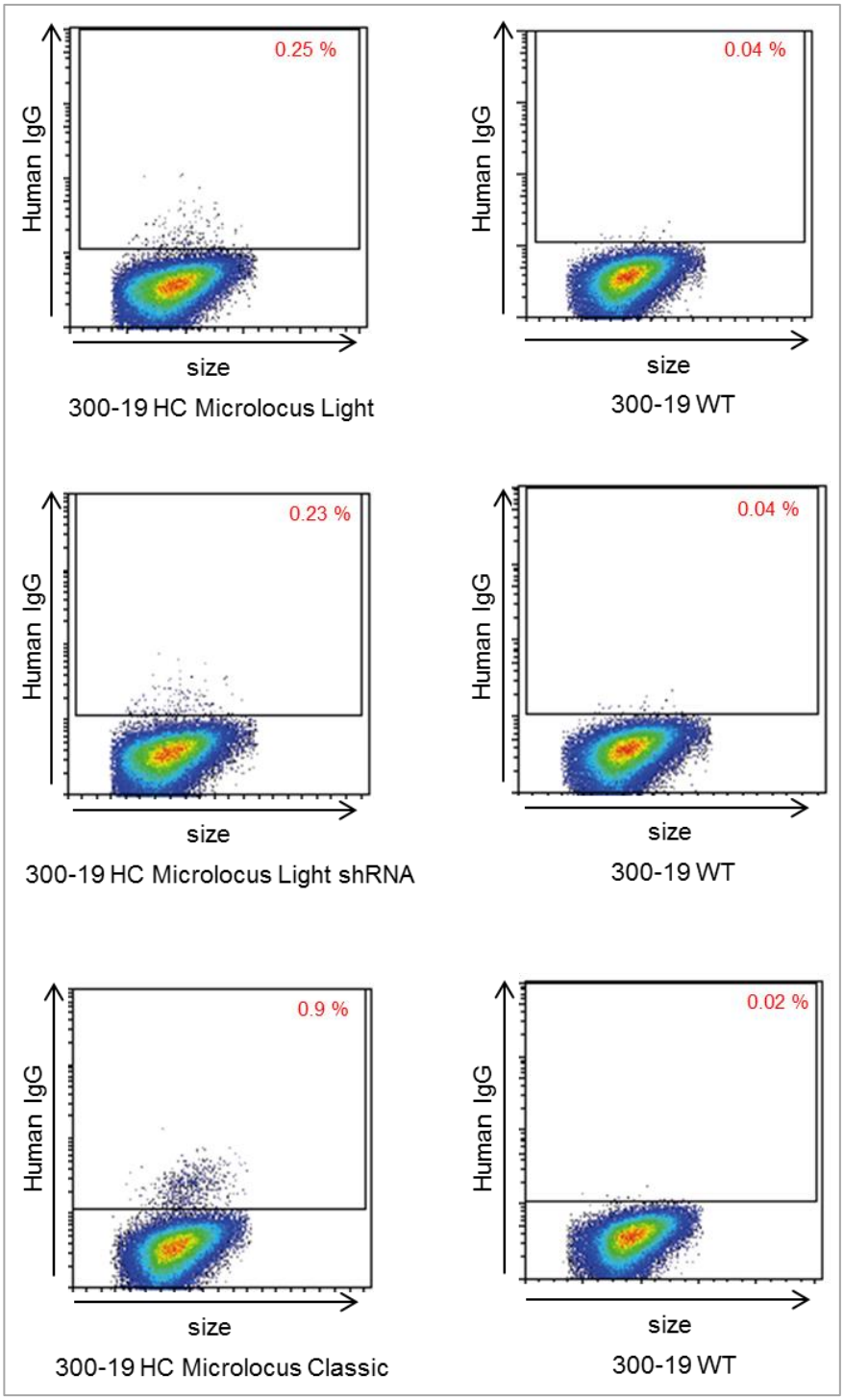

Figure 73: Surface detection of human γ1 HC, three and an half months in culture after transfection. A small ratio of cells, transfected with different constructs, expressed human γ1 HC on cell surface (*left panel*) when compared with wild type cells (*right panel*). Alive cells gating was based on their size and granularity (not shown); human γ1 HC expression was assessed on alive cells. The percentage of positive cells is in red color, in the top right-hand corner of the dot plot.

The low frequency of intracellular and surface human ɣ1 HC positive cells could be related to the fact than 300-19 cells are pro-B cells, and only few of them are able to differentiate to later stage of B cell development [6]. Nonetheless these results are a proof of the functionality of the designed constructs HC Microlocus Light, HC Microlocus Light shRNA and HC Microlocus Classic.

In a second time, we concentrated our efforts to enrich the few surface positive population by panning or sorting strategies allowing an increase of human ɣ1 HC expressing cell ratio and therefore, a more accurate characterization of these cells.

3.3.4. Positive cell enrichment

In order to characterize deeper human ɣ1 HC expressing 300-19 cells, at the VDJ rearrangement as well as at the protein expression level, human heavy chain expressing cells were enriched by *3.3.4.1.)* cell panning using an anti-human IgG antibody and *3.3.4.2.*) human ɣ1 HC positive cell sorting. The cell panning was first tested with HC Microlocus Classic transformed cells, while positive cell enrichment by cell sorting was performed on HC Microlocus Classic and HC Microlocus Light transformed cells.

3.3.4.1. Human ɣ1 HC positive cell panning

As observed in the last experiment, almost 1% of 300-19 cells transfected with the HC Microlocus Classic construct express the human ɣ1 heavy chain on surface (**Figure 73**). The human ɣ1 HC surface positive population was enriched by panning. OmniTray MaxiSorp (Nunc®) plates were coated with 5 µg/mL of goat anti-human IgG (ɣ) antibody, mouse absorbed (Invitrogen H10500), and incubated overnight at 4°C. After washing, 300-19 HC Microlocus Classic cells, showing initially 0.9% of cell expressing ɣ1 HC on surface, were transferred on the plate and incubated 2 hours at room temperature. This method allows the capture of positive cells expressing surface ɣ1 HC. After washing, most of negative cells were removed. The remaining positive cells were then incubated with rich RPMI medium containing 1mg/mL of Geneticin®. After one week incubation, cell population obtained by panning was analyzed by FACS. Human

IgG positive cell population reached 15% (**Figure 74**). This experiment shows that we were able to enrich human ɣ1 HC expressing cells by cell panning with an anti-human IgG antibody.

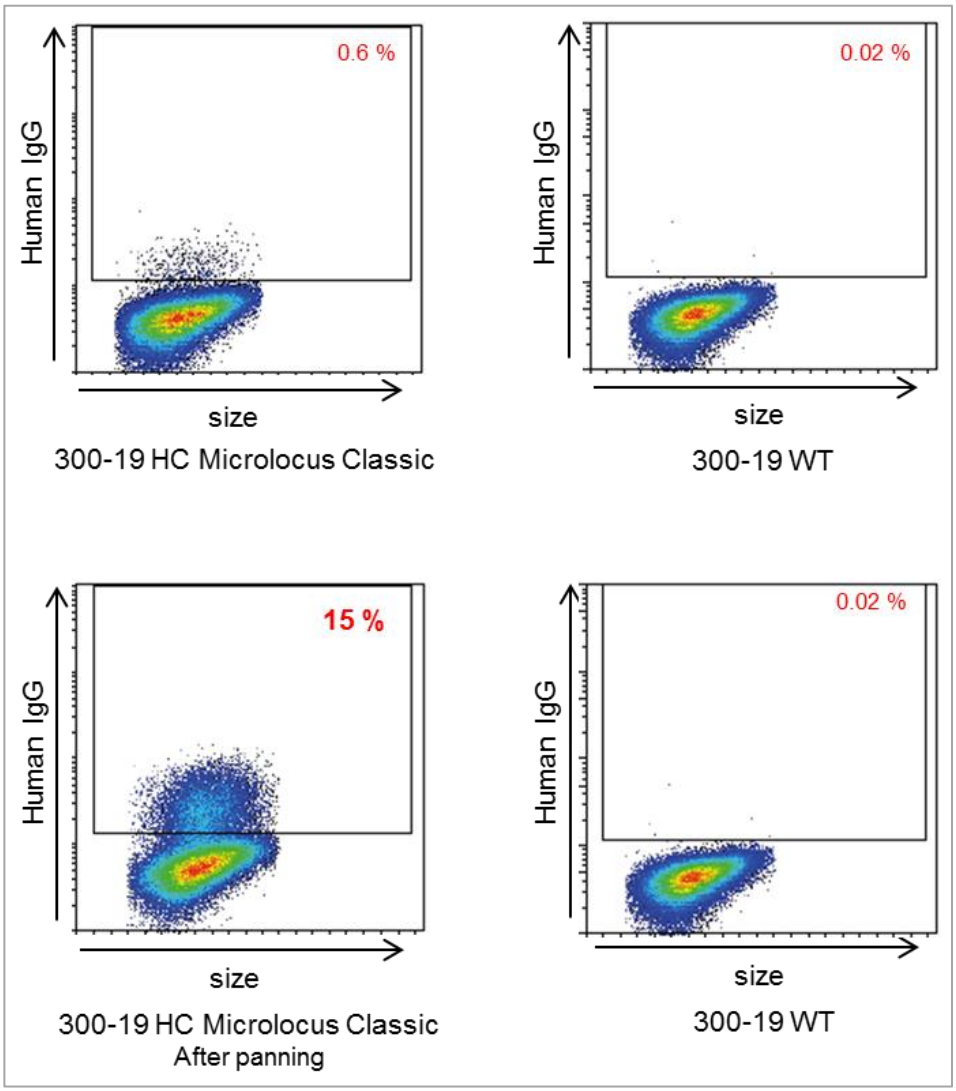

Figure 74: Increased human ɣ1 HC expressing cell ratio after panning. Human ɣ1 HC expressing cell shows a remarkable increase after panning (*lower panel*) when compared with expression level in original cell culture (*upper panel*). Alive cells gating was based on their size and granularity; and human ɣ1 HC positive cells were assessed on live cells (not shown). Human IgG gating was defined according negative control, *i.e.* WT 300-19 cells (*right panel*). The percentage of positive cells is in red color, in the top right-hand corner of dot plot.

3.3.4.2. Human ɣ1 HC positive cell sorting by flow cytometry

The purpose of cell sorting was to retrieve the cell population of interest, *i.e.* human ɣ1 HC expressing cells, from a heterogeneous population for further study. If a cell can be specifically identified by its physical characteristics, and this is the case, it can be separated using a flow sorter.

Sample preparation prior to sorting is important: in fact, successful sorting depends almost entirely on the state of the input sample. A prerequisite for flow cytometry is that cells should be in a monodispersed suspension. Human ɣ1 HC expressing cells from HC Microlocus Light and HC Microlocus Classic cell cultures were sorted. For both cell cultures, 8.10^6 of cells were washed two times with sterile FACS buffer (PBS, 0.5% FCS, 2mM EDTA). Both cell samples were stained with an optimal quantity of anti-human IgG Fc antibody (*ref.*: eB 12-4998-82). As negative control, wild type 300-19 cells were stained following the same protocol. Samples were incubated 30 minutes at 4°C in the dark. After washing, cell samples were acquired and sorted with FACS Aria III. Human IgG expression level was assessed on live cells based on their size (FSC parameter) and granularity (SSC parameter). Negative control was used to define the human IgG gate (not shown). Human IgG positive cells were sorted as pool in a tube containing complete RPMI medium, 20% FC, 1 mg/mL Geneticin® (*ref*: Gibco, 10131035). Electronically gating is shown in **Figure 75**. For HC Microlocus Light transformed cells, around 60 000 human IgG positive cells were sorted, while for HC Microlocus Classic, around 40 000 positive cells were sorted. Sorted cells were incubated at 37°C (5% CO_2). One week later, sorted cells were analyzed for human IgG expression by FACS. Cell samples were acquired with LSR Fortessa. Analyses were performed with FlowJo software. Expression levels were assessed on live cells based on their size (FSC parameter) and granularity (SSC parameter). The human ɣ1 HC positive cell population showed a great increase; indeed HC Microlocus Light and HC Microlocus Classic 300-19 cell culture reached respectively 83% and 53% of positive cells (**Figure 76**). This experiment shows that using cell sorting by flow cytometry, we were able to enrich human ɣ1 HC expressing cells.

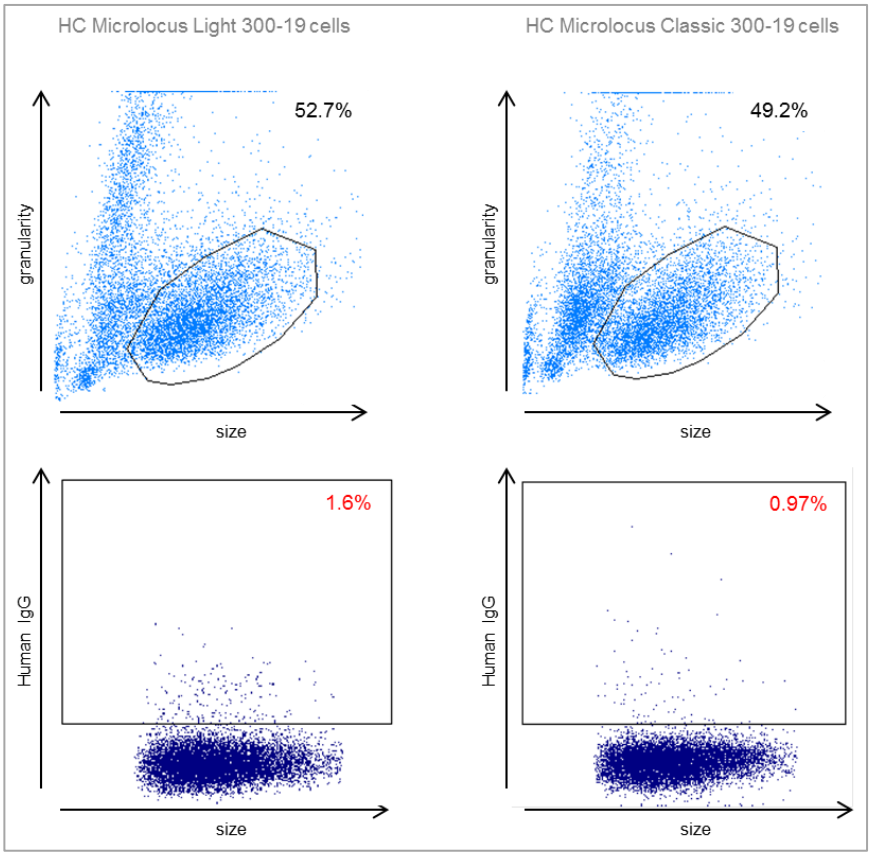

Figure 75: Electronically gating for human ɣ1 HC positive cell sorting. For HC Microlocus Light (*left panel*) and HC Microlocus Classic (*right panel*), live cells gating was based on their size and granularity (*upper panel*). Human ɣ1 HC positive cells were assessed on live cells (*lower panel*). Human IgG gating was defined according negative control *i.e.* non-transfected 300-19 cells (not shown). The percentage of positive cells is in red color, in the top right-hand corner of the dot plot.

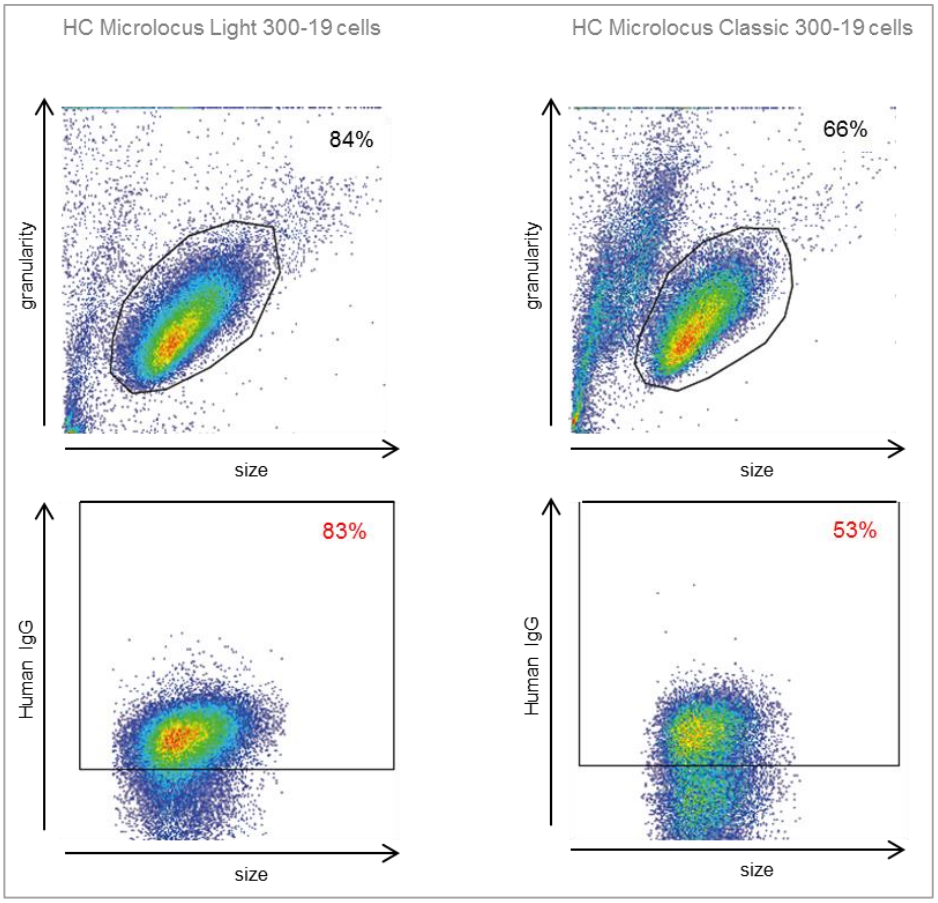

Figure 76: Human ɣ1 HC positive cell population after sorting. For HC Microlocus Light (*left panel*) and HC Microlocus Classic (*right panel*), live cells gating is based on their size and granularity (*upper panel*). Human ɣ1 HC positive cells were assessed on live cells (*lower panel*). Human IgG gating was defined according negative control (not shown). The percentage of positive cells is in red color, in the top right-hand corner of the dot plot.

3.3.5. 300-19 ɣ1 HC phenotype characterization

In order to better understand how HC Microlocus Light and HC Microlocus Classic constructs are processed in 300-19 cell line, we analyzed the expression of different surface markers involved in BCR signaling and in B cell development.

HC Microlocus Light and HC Microlocus Classic lead respectively to LC independent and LC dependent antibodies, *i.e.* to be expressed at the cell surface, ɣ1 HC $CH1^-$ encoded by HC Microlocus Light does not require LC (or SLC) pairing, while ɣ1 HC encoded by HC Microlocus Classic need to associate with surrogate LC or LC.

3.3.5.1. Method

After enrichment of the 300-19 cell population expressing human ɣ1 HC, it was attempted to characterize them at cell surface markers. Expression of BCR co-receptors Igα and Igβ was assessed by FACS. The original 300-19 clonal isolates have been described as blocked at the pro-B stage [5] therefore surrogate LC elements (VpreB, λ5) expression were also analyzed. In parallel, it has been described that, when propagated in culture, 300-19 cells can progress from pro-B cell to the mature B cell stage [6], therefore mouse IgM, mouse Igκ and mouse Igλ expression were also evaluated. We used WT mouse splenocytes as positive control. Lymphoid mononuclear cells were isolated by smashing the spleen obtained from another experiment. Analyzed HC Microlocus Classic and HC Microlocus Light cell culture showed respectively 19% and 25% of cell expressing human ɣ1 HC. The staining protocol used is the same as described earlier (see 3.3.3.1.).

Antibodies used for the different staining combinations were:

- PE anti-human IgG Fc antibody, *ref.*: eB 12-4998-82,
- APC anti-mouse CD79a (Igα) antibody, *ref.*: biolegend 133105,
- FITC anti-mouse CD79b (Igβ) antibody, *ref.*: AbD serotec MCA2209T,
- Per CP Cy5 anti-mouse Ig κ antibody, *ref.*: BD 560668,
- PE anti-mouse Ig λ antibody, *ref.*: BL 407308,
- V450 anti-mouse IgM antibody, *ref.*: BD 560575,
- PE anti-mouse CD179a (VpreB) antibody, *ref.*: biolegend 143603, and
- PE anti-mouse CD179b (λ5) antibody, *ref.*: biolegend 349803.

For each staining combination, appropriate compensation controls were used. Samples were acquired with LSR Fortessa. Analyses were performed with FlowJo software. Different expression levels were assessed on alive cells based on their size (FSC parameter) and granularity (SSC parameter).

3.3.5.2. Results

Mouse IgM, mouse Igλ and mouse Igκ expression

No mouse IgM and no mouse Igλ were detected (data not shown) on the cell surface of 300-19 WT and transfected cells. However, mouse Igκ was slightly but significantly detected in HC Microlocus Classic transfected 300-19 cells (**Figure 77**). HC Microlocus Light 300-19 cells also show a small population of cells expressing mouse Igκ, but the percentage is too low to be considered as significant.

Expression of mouse Igκ in transfected 300-19 cells is related to the fact that, when propagated in culture, 300-19 cells can partially undergo differentiation and therefore express LCκ. However to be detected at the cell surface, a LC in necessary associated with a HC. Since HC Microlocus Classic cells do not show mouse IgM expression, mouse LCκ is certainly associated with the human γ1 HC. This could suppose that HC Microlocus Classic expression drive 300-19 cell differentiation.

Surrogate Light Chain (SLC) elements and BCR co-receptors Igα and Igβ

Despite high background in 300-19 WT cells, mouse λ5 is expressed on HC Microlocus Classic 300-19 cells. Althought HC Microlocus Light should allow producing CH1 missing antibodies, few transfected cells were also detected as positive for λ5 expression. VpreB was detected in HC Microlocus Classic as well as HC Microlocus Light transfected cells. Mouse Igβ is expressed on 300-19 WT cell and on transfected cells; however transfected cells show a greater percentage of positive cells, which implies that human γ1 HC expression somehow drive Igβ expression. Mouse Igα detection is not so evident. These results suggest that the constructs (HC Microlocus Light and HC Microlocus Classic) are able to drive the surface expression of some molecules crucial for pre-BCR expression and signaling. All these data are summarized in **Figure 78**.

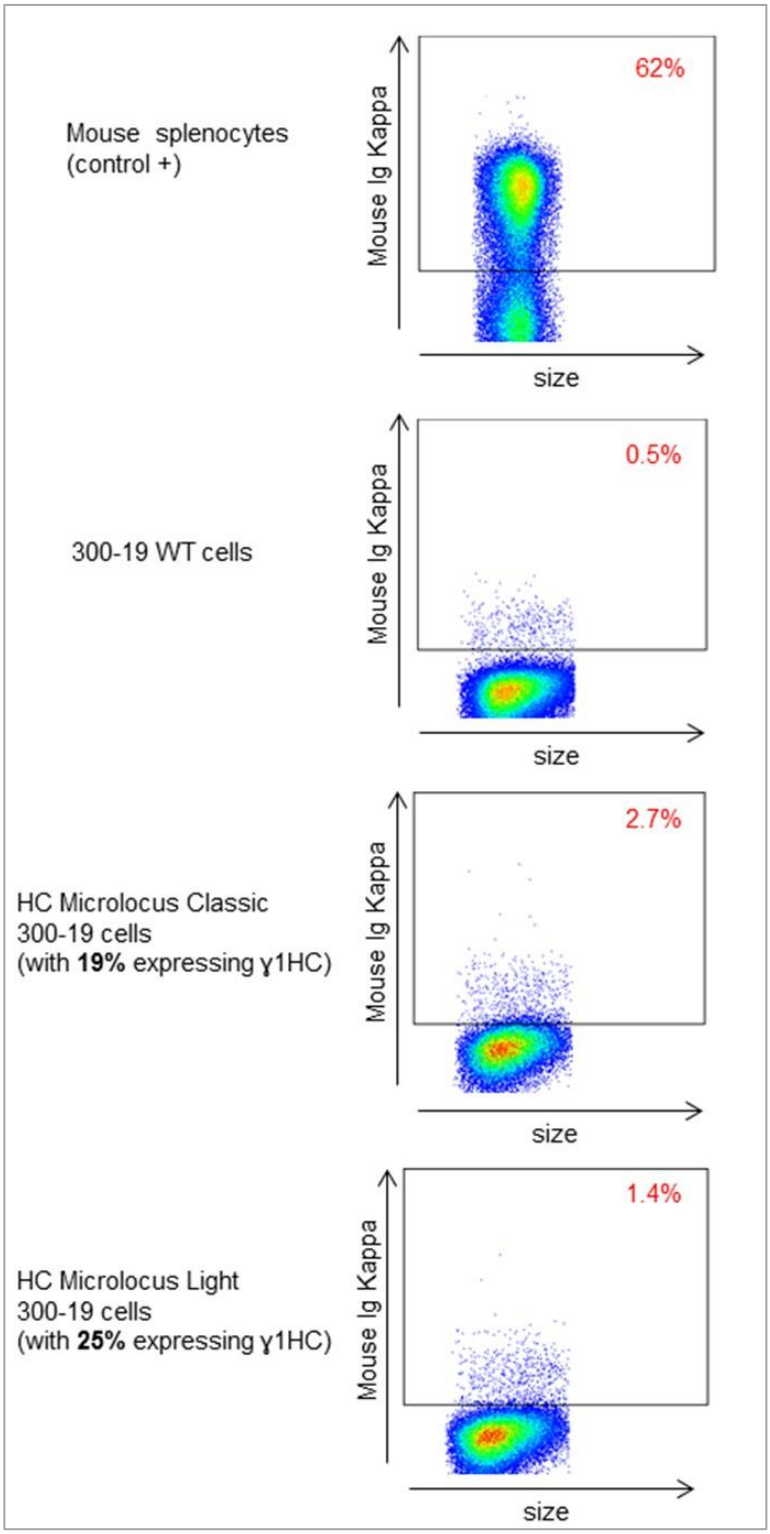

Figure 77: Mouse Igκ expression assessed on alive cells. Mouse splenocytes were used as positive control to define mouse Igκ gate (*upper panel*), which was applied to 300-19 WT cells (*2nd panel*), HC Microlocus Classic 300-19 cells (*3rd panel*), and HC Microlocus Light 300-19 cells (*lower panel*). The percentage of positive cells is reported, in the top right-hand corner of the dot plot.

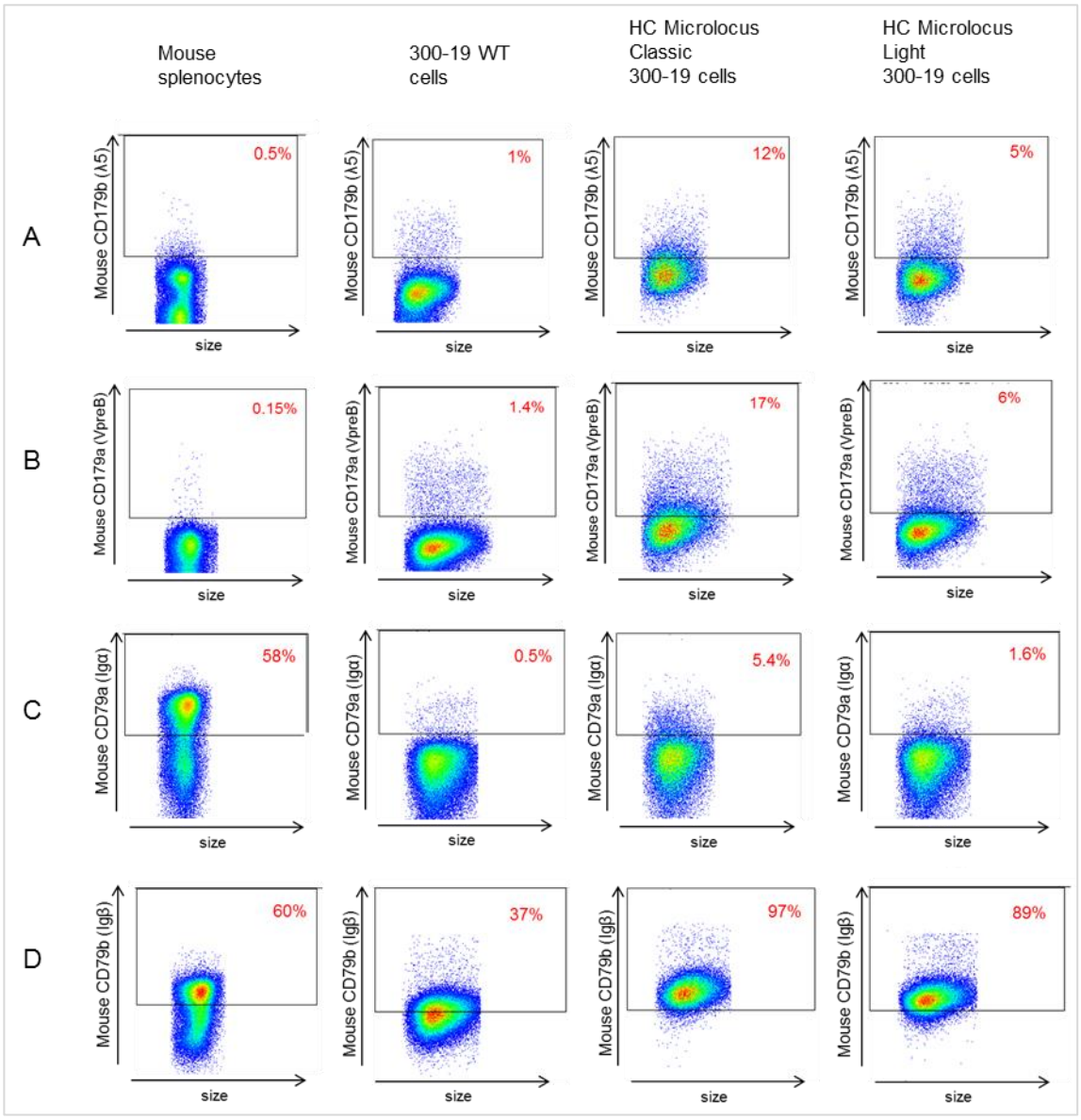

Figure 78: Evaluation of mouse λ5 (*A*), VpreB (*B*), Igα (*C*) and Igβ (*D*) expression on cell surface. The expression rates were assessed on alive cells. Mouse splenocytes were used as negative control to define λ5 and VpreB gates and as positive control for Igα and Igβ (*1st column*). Each gate was then applied to 300-19 WT cells (*2nd column*), HC Microlocus Classic 300-19 cells (*3rd column*), and HC Microlocus Light 300-19 cells (*4th column*). The percentage of positive cells is reported, in the top right-hand corner of the dot plots.

3.3.6. Human Ig secretion *in vitro* measured by ELISA assay

We have shown that HC Microlocus Light and HC Microlocus Classic transfected cells express intracellular and surface human HC. Considering the fact that 300-19 cell line can sequentially undergo all of the immunoglobulin gene rearrangement and expression events associated with the progression to the B cell stage [6], we asked whether 300-19 human HC transfected cells could secrete human IgG in the supernatants. This assay could also demonstrate that HC Microlocus Light and HC Microlocus Classic can produce both, membrane and secreted forms. Alternative splicing is a key function to encode these both splice products.
In parallel, we also analyzed the cell supernatant for secreted mouse IgM and mouse IgG.

3.3.6.1. Method

After enrichment in human HC positive cells (by panning or sorting), cells were grown at confluence. The culture media was then harvested and the presence of human Ig has been assessed by ELISA (Enzyme-Linked Immunosorbent Assay).

ELISA plates (96 wells ELISA plates Greiner medium binding, *ref*: 655001) were coated with 70 µl/well of unlabeled capture antibody solution (final concentration: 10 µg/mL in PBS) and incubated overnight at 4 °C. As capture antibody, we used a goat anti-human IgG antibody purchased from Southern Biotech (*ref*: 2040-01). Plates were washed three times with PBS supplemented with 0.05% of Tween 20. Once capture antibodies have been properly adsorbed onto the ELISA plate, the next critical step in creating a reliable immunoassay is the blocking of the plate. Each well were blocked with 200 µl of blocking solution (PBS supplemented with 10% FCS) and incubate 2 hours at room temperature. After removing blocking solution, 50 µl of cell supernatants were incubated over night at 4°C. We used serial dilutions (0 to 100 ng/mL) of human IgG (*ref.*: Sigma, I4506-10MG), diluted in sterile RPMI medium in order to keep the same conditions as for the samples to test. After cell culture supernatant incubation, the plates were three times washed with PBS supplemented with 0.05% of Tween 20. Wash solution was carefully removed. To detect any human Ig previously captured, wells were incubated for 2 hours at room temperature, with 50 µl of detection antibody diluted at

2 µg/mL in blocking buffer. As detection antibody, we used a goat anti-human IgG antibody conjugated with alkaline phosphatase purchased from Southern Biotech (*ref*: 2040-04). After three washing with PBS supplemented with 0.05% of Tween 20, 100 µl of alkaline phosphatase fluorescent substrate (AttoPhos®, *ref*.: S1001, 0.2 mM diluted in PBS supplemented with 0.05% of Tween 20) were added in each well.

In parallel, we evaluated mouse IgG and mouse IgM levels in cell supernatant by using a goat anti-mouse IgG antibody (*ref*: 1030-01) and a goat anti-mouse IgM antibody (*ref*: 1021-01) as capture antibodies. Detection antibody used were the same, except that they are conjugated with alkaline phosphatase (*ref*: 1030-04, 1021-04).

3.3.6.2. Results

ELISA assays performed with HC Microlocus Light and HC Microlocus Classic cell supernatant for human IgG measurement, showed a significant signal when compare to WT 300-19 cell supernatant. We were able to evaluate the human ɣ1 HC concentration in both transfected cell lines, by referring to standard curve (**Figure 79**).

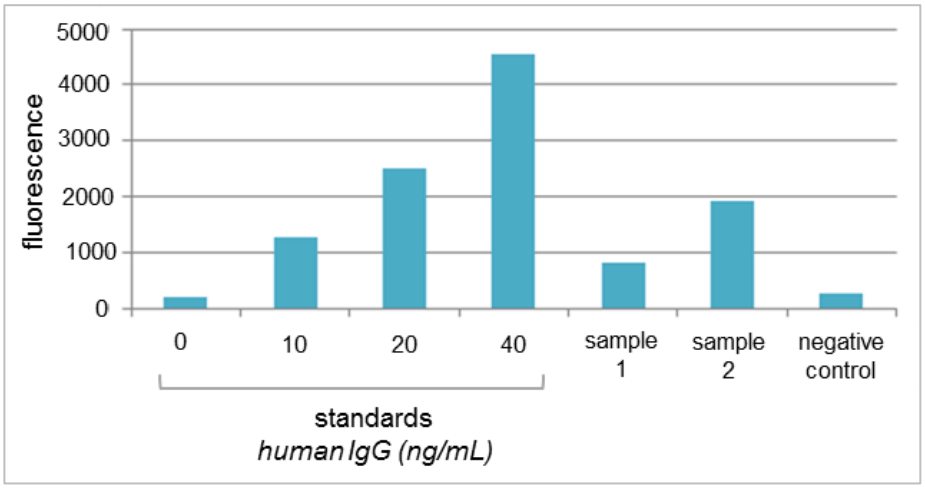

Figure 79: Assessment of human IgG concentration by ELISA. The standards correspond to RPMI medium supplemented with different concentration of human IgG (*i.e.* 10, 20 and 40 ng/mL). Negative control corresponds to RPMI medium. Sample 1 corresponds to HC Microlocus Classic 300-19 cell supernatant, while sample 2 corresponds to HC Microlocus Light 300-19 cell supernatant. All measurements were performed in the same assay.

Using standard curve we can detect human IgG in transformed cell supernatant. Furthermore, the assay performed with WT 300-19 cell supernatant, only shows a weak background, similar to the signal background observed with the negative control. Thus, we were able to estimate the concentration of human ɣ1 HC in cell supernatant; HC Microlocus Light and HC Microlocus Classic transfected cell supernatant contain respectively, around 16 ng/mL and 7 ng/mL of human ɣ1 HC. In parallel to this experiment, we analyzed mouse IgM and IgG concentrations in cell supernatant, but neither were detected.

Because the detection of human ɣ1 HC expression is the absolute evidence that human HC constructs are functional, we concentrated our efforts to confirm these latter results by using another Ig detection method, the ELISPOT assay (3.3.7.), but also by trying to visualize these proteins *via* Western Blot assay (3.3.8.).

3.3.7. Human Ig secretion *in vitro*; ELISPOT assay

HC Microlocus Light and HC Microlocus Classic transfected cells secrete human ɣ1 HC. To verify these results, we decided to perform an Enzyme-Linked ImmunoSpot (ELISPOT) assay. This method is a very sensitive immunoassay which measures the frequency of immunoglobulin secreting cells. Cells are cultured on a surface coated with a specific capture antibody. Secreted proteins are captured by the specific antibodies on the surface. After an appropriate incubation time, cells are removed and the secreted Ig detected using a similar procedure to ELISA. By using a substrate with a precipitating rather than a soluble product, the end result is seen as visible spots on the surface where each spot corresponds to an immunoglobulin-secreting cell. In order to confirm the last observation indicating that HC Microlocus Light and HC Microlocus Classic 300-19 cells secrete human ɣ1 HC, we have performed an ELISPOT targeting human IgG secreting cells. This experiment was performed on cells expressing human ɣ1 HC at the cell surface. Thus, for HC Microlocus Light and HC Microlocus Classic cell culture, around 300 000 human ɣ1 HC expressing 300-19 cells were previously sorted following the protocol described in 3.3.4.2.

3.3.7.1. Method

We used 96 well plates (*ref*: MSIPS4510), which were coated with 70 µL per well of capture antibody at 10 µg/ml in coating buffer (see below). The capture antibody used was a goat anti-human IgG antibody, purchased from Southern Biotech (*ref.*: 2040-01). The plate was incubated 2 hours at room temperature. After washing with PBS, wells were blocked with 200 µL of blocking solution (PBS supplemented with 1% BSA) and incubate 30 minutes at 37°C and then pre-incubate with 100 µL per well of complete RPMI medium (see **Table 5**, page 116).

Human ɣ1 HC expressing 300-19 cells were titrated in a 1 to 2 serial dilution, (starting from 10^5 cells/well) in 100 µl of complete medium. Cells were incubated 24 and 48 hours, at 37°C, 5% CO_2. The number of Ig cell secreting should be the same in both assays; however the amount of secreted Ig should be greater after 48 hours cell incubation, when compared with 24 hours cell incubation. As negative control, two wells were incubated with sterile complete RPMI medium.

After incubation, cells were removed and the wells were washed with PBS supplemented with 0.25% Tween 20. In order to reveal secreted human ɣ1 HC, each well were incubated with 70 µL of 0.25 µg/mL detection antibody diluted in blocking buffer. As detection antibody we used a goat anti-human IgG antibody biotinylated, purchased from Southern Biotech (*ref.*: 2043-08). The plate was incubated for 2 hours at room temperature.

After washing, we added to each well 70 µL of 2.5 µg/mL streptavidin (SA) coupled with Horseradish peroxidase (HRP), diluted in blocking buffer. Streptavidin, which have high affinity for biotin, will bind all detection antibodies specifically retained on the well surface. The plate was incubated 1 hour at room temperature and then intensively washed. Finally, 70 µL per well of developing solution (see below) was added. The developing solution contains 3-Amino-9-ethylcarbazole, which is a peroxidase substrate. This substrate produces an insoluble end product that is red in color and can be observed visually. The plate was incubated in the dark at room temperature until it was possible to individuate red-orange spots (5-10 minutes).

All reagents used in the ELISPOT method are described below:

Coating buffer:
- solution A: 8.4 g $NaHCO_3$ in 100 ml H_2O,
- solution B: 10.6 g Na_2CO_3 in 100 ml H_2O,
- mix 11.3 mL of solution A with 4.5 mL solution B, and complete up to 250 mL with H_2O.

Developing solution:
- add 1 tablet AEC (Sigma, *ref*: A6926) in 1.6 mL dimethylformammide, dissolve by vortex,
- add 200 µL of AEC solution to 9 mL sodium acetate buffer (see composition below),
- filter the solution with 0.45 nm filter,
- add 12 µL of H_2O_2 30%,
- prepare the solution immediately before use.

Sodium acetate buffer:
- solution A: 11.5 mL glacial acetic acid in 1 L H_2O (final concentration: 0.2M),
- solution B: 27.2 g sodium acetate hydrate in 1 L H_2O (final concentration: 0.2M),
- mix 14.8 mL of solution A to 35.2 mL of solution B, and complete up to 100 mL with H_2O.

3.3.7.2. Results

After 24 or 48 hours incubation, cell secreting human ɣ1 HC were detected. Both plates were photographed using CTL ImmunoSpot® reader (**Figure 80** and **81**). Each orange colored dot represents a human ɣ1 HC secreting cell. The results obtained with ELISPOT assay are fully coherent, indeed the assay performed with 48 hours cell incubation show stronger orange coloration intensity when compared with 24 hours incubation. Furthermore, no orange coloration developed on wells incubated with WT 300-19 cells. However, dots are small and weakly colored; this may indicate a low secretion rate. These results confirm that 300-19 cells transfected with HC Microlocus Light and HC Microlocus Classic, detected as positive for human ɣ1 HC expression on

surface, can also secrete human ɣ1 HC. These results, which are consistent with those obtained by ELISA assay, demonstrates that HC reduced constructs are fully functional and can be well processed by the mouse machinery.

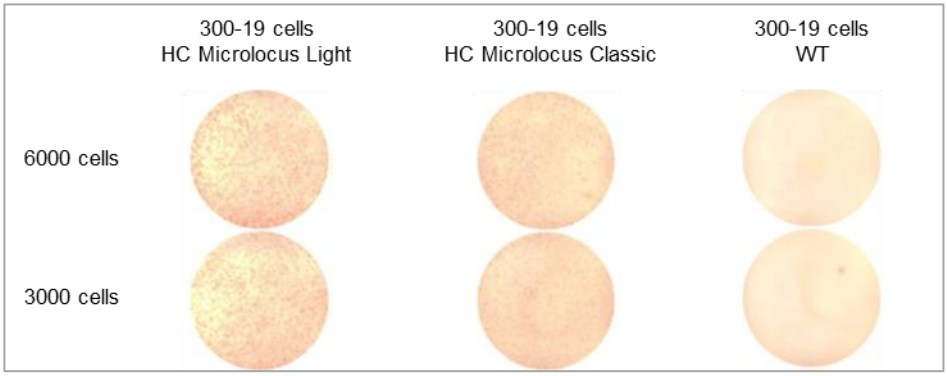

Figure 80: ELISPOT assay, human ɣ1 HC secreting cells detected after 24 hours incubation.

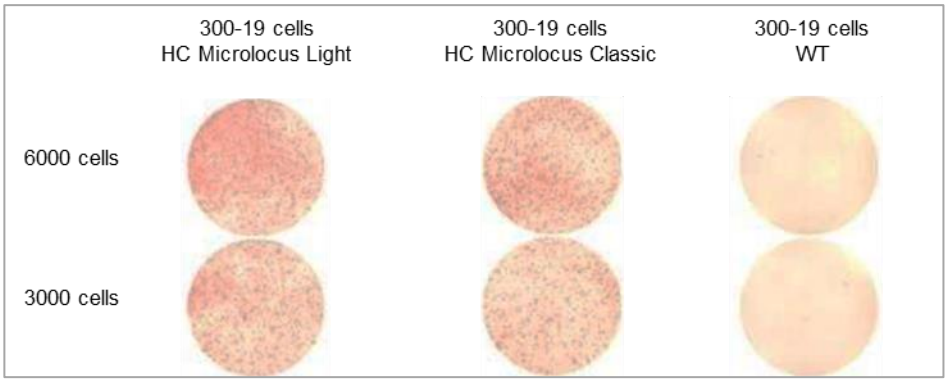

Figure 81: ELISPOT assay, human ɣ1 HC secreting cells detected after 48 hours incubation.

3.3.8. Analysis of secreted human antibodies

In order to better characterize the secreted human Ig proteins, we performed Western Blot with cell supernatants. This analytical method is used to detect specific proteins and to evaluate their size. HC Microlocus Light encodes ɣ1 HC lacking CH1 domain (ɣ1 HC $CH1^-$). Such immunoglobulins are defined as HC only antibodies because they are expressed independently of LC [90]. The mass of one ɣ1 HC $CH1^-$ chain is estimated at around 40 kDa. HC Microlocus Classic encodes the classical ɣ1 HC, which represents a

50 kDa protein. Furthermore ɣ1 HC CH1$^-$ and ɣ1 HC may be associated with surrogate LC components, VpreB (~16 kDa) and λ5 (~22kDa). **Figure 82** shows the proteins that may be produced in 300-19 cells transfected with HC Microlocus Light, and their corresponding masses. **Figure 83** proposes the expected forms of Ig secreted by 300-19 HC Microlocus Classic.

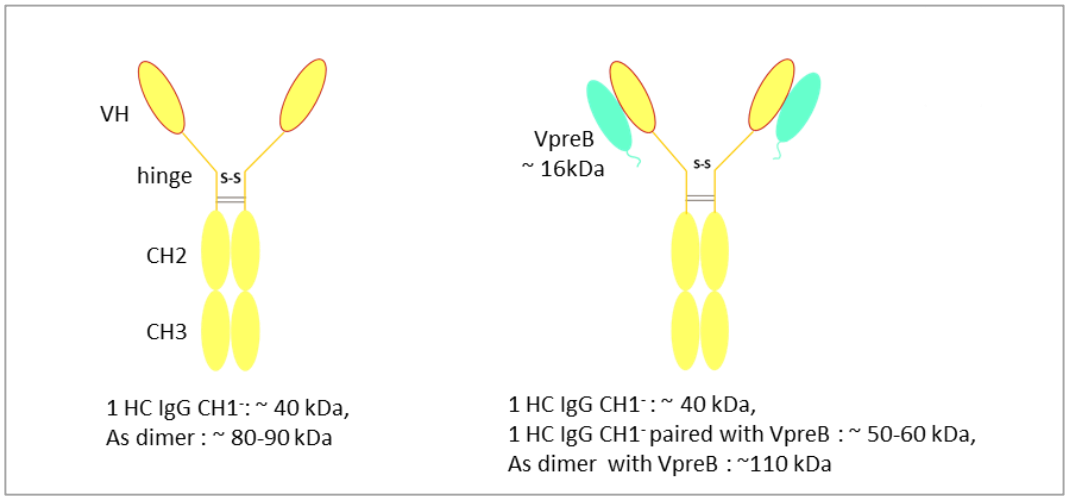

Figure 82: Schematic representation of Ig produced by 300-19 HC Microlocus Light cells.

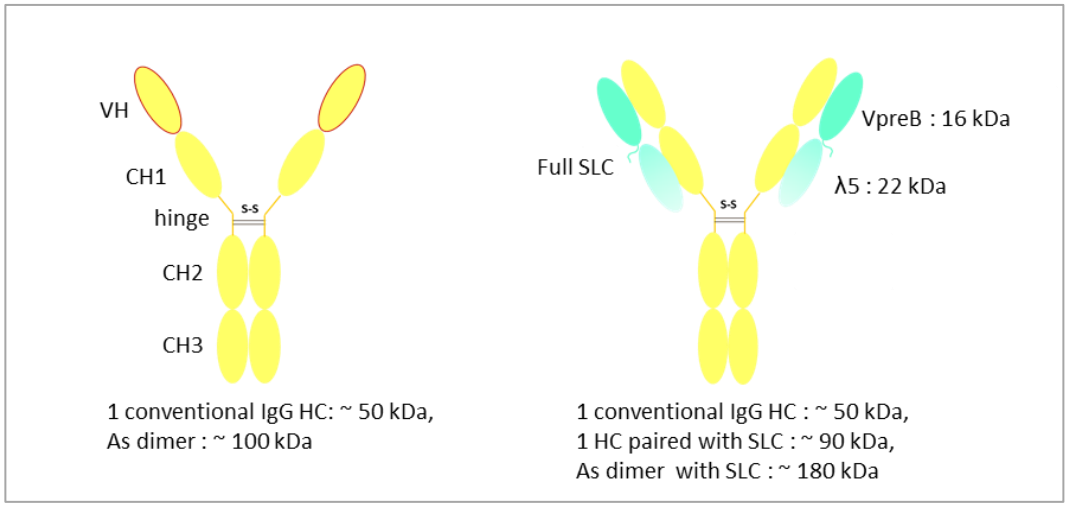

Figure 83: schematic representation of Ig produced by 300-19 HC Microlocus Classic cells.

3.3.8.1. Method

Since the human ɣ1 HC concentration in cell supernatant is low (HC Microlocus Light cell supernatant: ~ 16 ng/mL; and HC Microlocus Classic cell supernatant: ~ 7 ng/mL)

we performed a concentration step of the ɣ1 HC proteins using protein A column (MabSelect SuRe LX from GE Healthcare, *ref.*: 17-5474-01): from 20 mL cell culture supernatant, ɣ1 HC proteins were concentrated to 20 µL. As control, 300-19 WT cell supernatant and sterile RPMI medium supplemented with 10 or 20% of FCS were treated in the same manner.

For each samples, proteins were analyzed under reduced and non-reduced conditions. Each sample was divided in two: one for reduced, the other one for native proteins (10 µL each). To each of them was added 3.3 µL of loading buffer (LDS 4X Invitrogen, *ref.*: NP0007). For reduced samples, 1.4 µL of reducing agent (NuPAGE® Sample Reducing Agent (10X), Invitrogen, *ref.*: NP0009) was added to each sample. The reducing agent acts to disrupt any disulfide bonds between antibody chains. All samples were heated 10 minutes at 80 °C.

In a first step, each protein mixture was separated according to molecular size. Thus, all samples have been run on SDS-PAGE (Gel 12wells, Invitrogen, *ref*: NP0322). To be able to evaluate protein size, we also used 5 µL of pre-stained protein ladder (see blue plus2, Invitrogen, r*ef*: LC5925). The running conditions were 35 minutes at 180 volts. After electrophoresis, proteins were transferred onto 0.45µm PVDF membrane using a trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad, ref: #170-3940). The transfer conditions were 30 minutes at 25 V. After washing and blocking, the membrane was incubated with 1 µg/mL of anti-human IgG antibody conjugated to alkaline phosphatase (*ref*: Rockland, 709-1516) for 2 hours at room temperature. After washing, the membrane was incubated in developing solution composed of one substrate tablet (SIGMA *FAST*™, *ref*: B5655), an insoluble substrate for the detection of alkaline phosphatase, diluted in 10 mL of water. The membrane was incubated until blue bands appeared. The developing was stopped by rinsing the membrane with water. After being correctly dried, the membrane was scanned.

3.3.8.2. Results

Several aliquots of HC Microlocus Light and HC Microlocus Classic 300-19 cell supernatant have been analyzed following the method previously described.

The Western Blot membranes corresponding to native and reduced proteins are respectively shown in **Figure 84** and **Figure 85.** These results confirmed the presence of

human ɣ1 HC in HC Microlocus Light and HC Microlocus Classic transfected cell supernatant. Furthermore, these results indicate a size difference between both proteins. This difference could correspond to ɣ1 CH1 domain, which is not encoded in HC Microlocus Light construct. These data are consistent with what we expected.

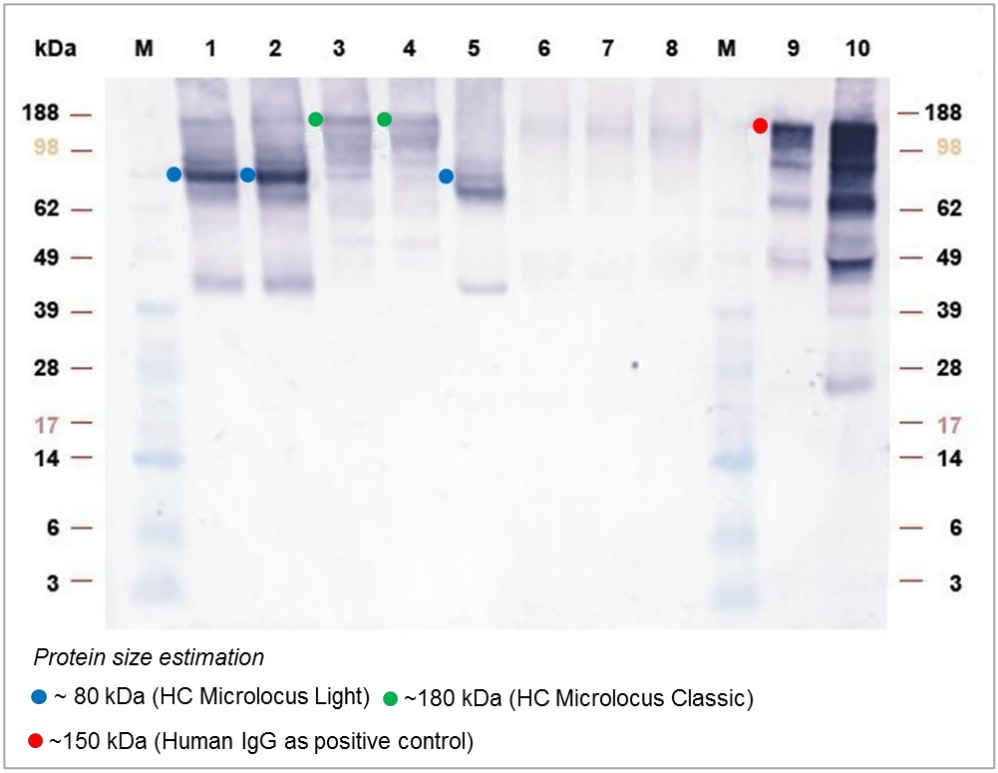

Figure 84: Western Blot after electrophoresis in absence of reducing agent.
1. 300-19 HC Microlocus Light, aliquot 1; **2.** 300-19 HC Microlocus Light, aliquot 2; **3.** 300-19 HC Microlocus Classic, aliquot 1; **4.** 300-19 HC Microlocus Classic, aliquot 2; **5.** 300-19 HC Microlocus Light, aliquot 3; **6.** 300-19 WT; **7.** Medium 10% FCS (GIBCO); **8.** Medium 20% FCS; **9.** Human IgG, 0.1 µg; **10.** Human IgG, 1.0 µg

Human IgG was detected in all HC construct transfected cell supernatants (samples **1** to **5**). As expected, no signal was observed in 300-19 WT cell supernatant (sample **6**), neither in negative controls (samples **7** and **8**). Samples corresponding to HC Microlocus Light (samples **1**, **2** and **5**) show a different pattern when compared with samples corresponding to HC Microlocus Classic (samples **3** and **4**). The bands of samples **1** and **2** correlate approximately to an 80 kDa protein, which can be interpreted as ɣ1 HC $CH1^-$ dimer. A similar band is also present in sample **5** but the protein partially degraded during the process. Regarding both samples corresponding to HC Microlocus Classic, we are not able to conclude because the signal is too weak.

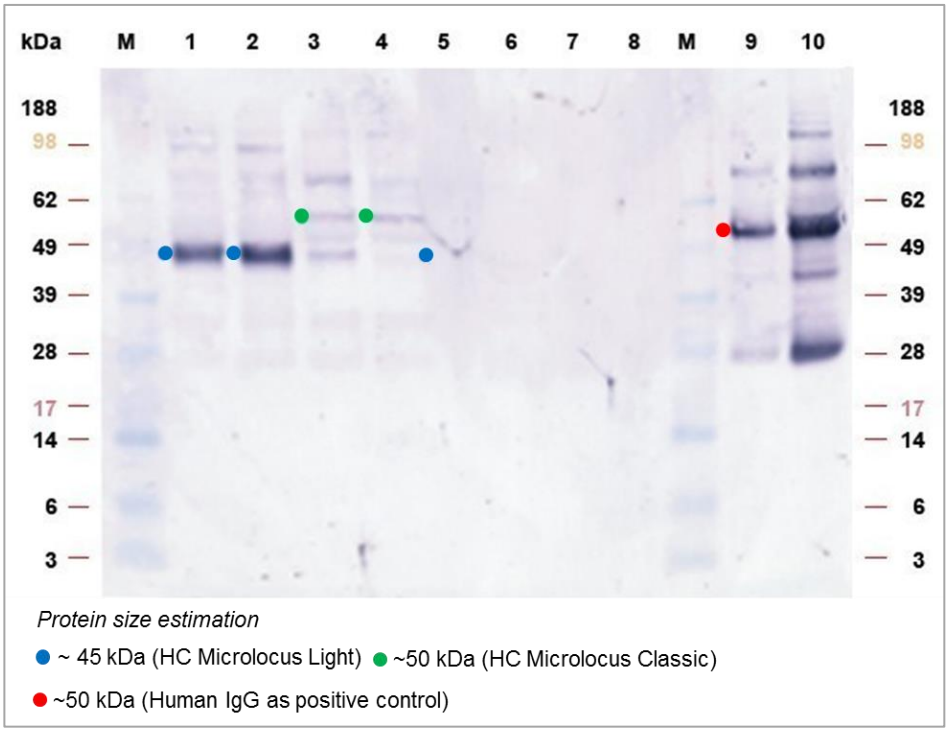

Figure 85: Western Blot after electrophoresis in presence of reducing agent.
1. 300-19 HC Microlocus Light, aliquot 1; **2.** 300-19 HC Microlocus Light, aliquot 2; **3.** 300-19 HC Microlocus Classic, aliquot 1; **4.** 300-19 HC Microlocus Classic, aliquot 2; **5.** 300-19 HC Microlocus Light, aliquot 3; **6.** 300-19 WT; **7.** Medium 10% FCS (GIBCO); **8.** Medium 20% FCS; **9.** Human IgG, 0.1 µg; **10.** Human IgG, 1.0 µg

Human IgG was detected in samples 300-19 HC Microlocus Light cell supernatant (except sample **5** which is not readable due to a gel default), as well as in samples 300-19 HC Microlocus Classic cell supernatant. As expected, no signal was observed in 300-19 WT cell supernatant (sample **6**), neither in negative controls (samples **7** and **8**). According to the size (~45 kDa), the bands in samples **1** and **2**, corresponding to HC Microlocus Light cell supernatant, are interpreted to be ɣ1 HC lacking the CH1 domain. HC Microlocus Classic cell supernatants (samples **3** and **4)** show a common band at around 50 kDa, this could correspond to a conventional ɣ1 HC.

3.3.9. DJ rearrangement analysis

The somatic DNA rearrangement of antibody genes is the basic process to generate antibodies. When transfected 300-19 cells started to express cytoplasmic human ɣ1 HC and on cell surface, we decided to investigate the study of DJ and VDJ rearrangements. Although we have not been able to clearly demonstrate the expression of human Ig from HC Minilocus transfected 300-19 cells, we analyzed these cells as well. The goal of DJ and VDJ rearrangement analysis was to verify that the constructs are well processed, and that different segments V, D and J were randomly recombined, leading to antibody diversity (recombinatory diversity). The analysis was performed at the DNA level.

3.3.9.1. Method

After FACS analyzes showing human ɣ1 HC expression in transfected 300-19 cell culture (see 3.3.3.), we pelleted the cells from theses culture and stored them at -20°C. In order to analyze rearrangements in these cells, they were lysate using a cell lysis buffer containing SDS, and the DNA was then purified with phenol chloroform followed by precipitation using sodium acetate/ethanol mixture. The method include a DNA quality control, which consists of looking for the presence of mouse Cε and mouse Cα genes, naturally present in 300-19 cells, by specific PCR performed on freshly purified DNA:

PCR control mouse IgE: 1306x1307 (469 bp):

- primer 1306 (forward): 5’ GTCAGCTGGCTAATGGACGATC 3’, which is homologous to mouse Cε CH2, and
- primer 1307 (reverse): 5’ GTTATGGTGGTGCTTAGTGTAC 3’, which is homologous to Cε CH3.

PCR control mouse Cα: 1308x1309 (601 bp):

- primer 1308 (forward): 5’ GTTCAGATGCCAGCCTCACATG 3’, which is homologous to mouse Cα CH2, and
- primer 1309 (reverse): 5’ CAGCTCCTCATTTCCATGCAGC 3’, which is homologous to Cα CH3.

After PCR cycling run, samples were loaded on 1% agarose gel electrophoresis containing 0.01% of RedSafe DNA staining solution (*ref.*: 21141), which allows DNA detection after electrophoresis (**Figure 86**).

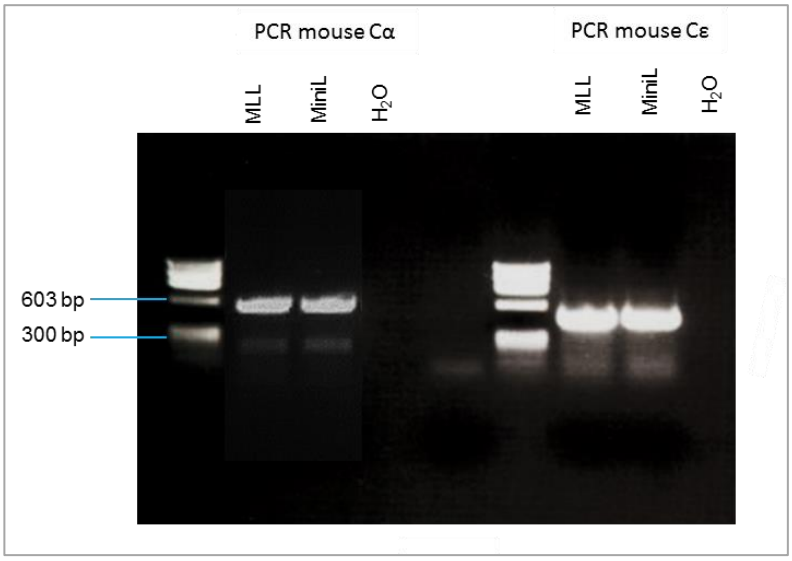

Figure 86: Gel picture of DNA quality control PCR. Gel: 1% agarose, running conditions: 25 minutes, 140 Volts. Size marker: ΦX174 DNA-HaeIII digest (*ref.*: NEM #N3026L). *MLL*: HC Microlocus light, *MiniL*: HC Minilocus.

PCR product were then extracted from agarose gel, using QIAquick Gel Extraction Kit (ref: Qiagen, 28704), and then sequenced. After sequencing it was confirmed that amplification products correspond to mouse Cα and Cε. All DNA sequences were analyzed with the program SeqMan. The gel picture in **Figure 86** only shows PCR products from cells transfected with HC Microlocus light and HC Minilocus, but cells transformed with other constructs were similarly analyzed. These PCR control certified the quality of DNA extracted from cell pellet was suitable for DJ rearrangements analyzes.

To analyze DJ rearrangements, specific primers were designed:

- primer 1429 (5' GGTATAACTGGAACGACTAC 3'), which is homologous to D1 sequence, and
- primer 1393 (5' GCTTAATTAAATTCTTACCT 3') which is homologous to the J6 3' end sequence.

The D-J region which is fully synthesized, show a unique genetic content in our constructs coding for human HC and therefore, any ambiguous results, such as

contamination, can be excluded. **Figure 87** shows the strategy to detect any DJ rearrangement coming from HC constructs.

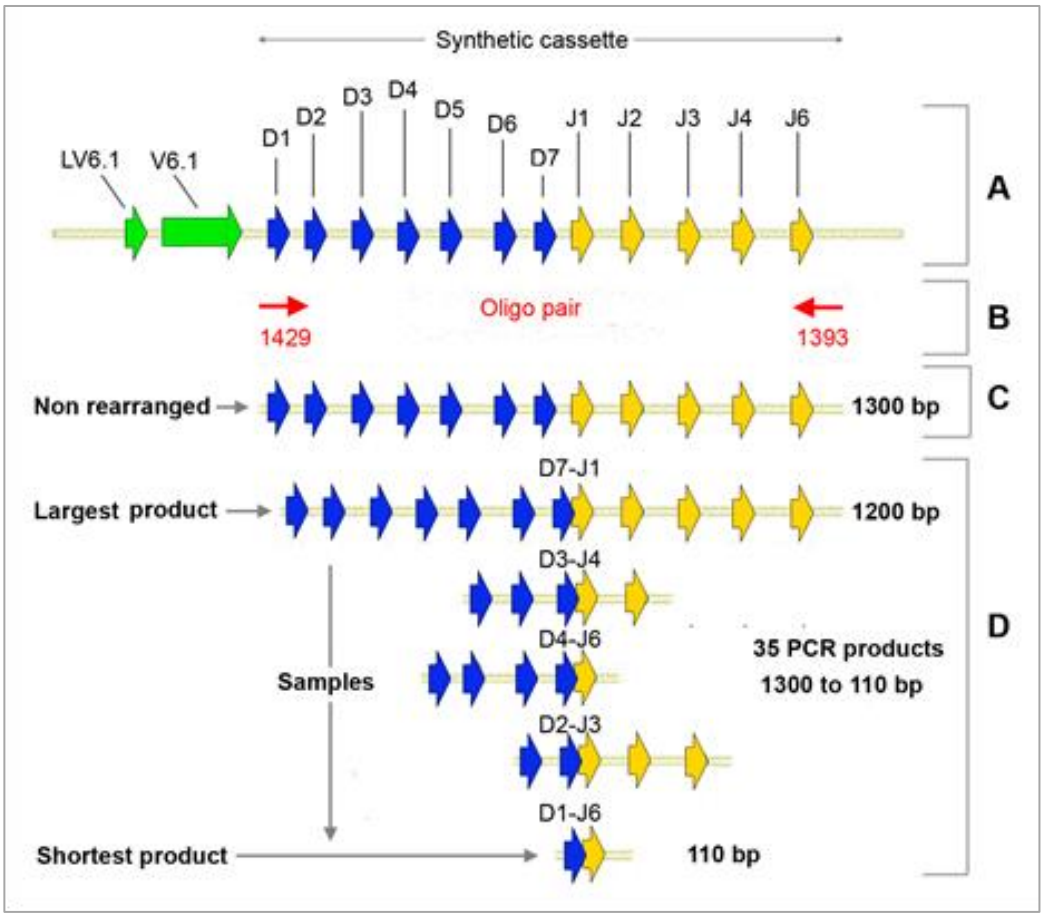

Figure 87: Strategy for DJ rearrangements analysis in 300-19 cells transfected with HC constructs. **A;** the D-J synthesized cassette, **B;** location of the primers used for PCR, **C**; Size of non-rearranged PCR products (WT), **D;** examples of expected PCR products.

As shown in **Figure 87**, the size of expected PCR products can vary from 110 bp to 1200 bp, thus PCR amplification was performed in duplicate, one with an elongation time of 30 seconds and the other one with an elongation time of 2 minutes. When the PCR run completed, samples were loaded on 1,5% agarose gel electrophoresis containing 0.01% of RedSafe DNA staining solution (*ref*.: 21141), which allows PCR product visualization. Detectable PCR product were extracted from agarose gel, using QIAquick Gel Extraction Kit (*ref*: Qiagen, 28704), and then sequenced. Weaker PCR product or smear were also purified from agarose gel and sub-cloned using TOPO® TA Cloning® Kit from Invitrogen. This sub-cloning step allows sequencing of weakly amplified PCR product.

3.3.9.2. Results

Considering that DJ rearrangements were analyzed on DNA from cell pool and that all cells do not have the same rearrangement, it was expected to get a large panel of PCR products with different sizes. Obviously, the most represented rearrangements were preferentially amplified and well detectable after gel electrophoresis. Regarding the rearrangements less frequently represented, they were only visible as smear, sometimes very weak, but by sub-cloning, we were able to identify the sequences. An example of PCR product is shown in **Figure 88**. This corresponds to PCR products obtained after amplification with primers 1429 and 1393 (30 seconds elongation); the DNA template was DNA purified from HC Microlocus Classic cell pellet.

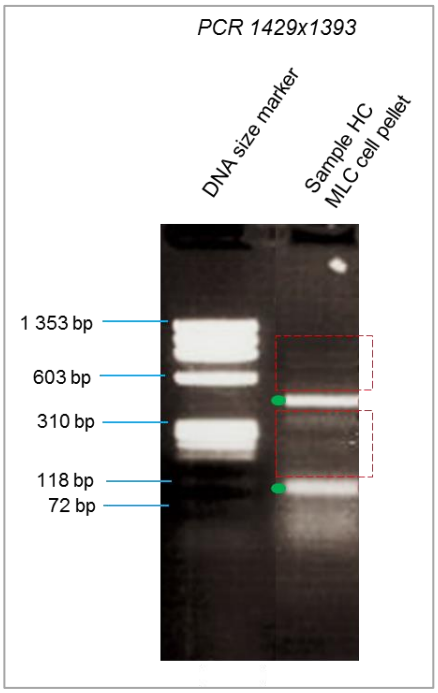

Figure 88: DJ rearrangements amplification (PCR with primers 1429 and 1393). The two bands marked with a green dot were extracted from the gel and sequenced. The smears framed in red were extracted from the gel and sub-cloned using TOPO® TA Cloning® Kit before being sequenced. Size marker: ΦX174 DNA-HaeIII digest, *ref.*: NEM #N3026L. (MLC: HC Microlocus Classic).

In this experiment the bright bands at around 400 bp and 100 bp were directly sequenced with both primers used for amplification (*i.e.* 1429 and 1393). The bands located above and below the 400 bp band are really weak, suggesting that they are amplicons from less represented DJ rearrangements. These PCR products are too less concentrated to be

directly sequenced, therefore, they were sub-cloned using TOPO® TA Cloning® Kit from Invitrogen. This sub-cloning step allows sequencing of weakly amplified PCR products. All DNA sequences were analyzed with the program SeqMan.

Despite the fact that we often amplified nonspecific products from mouse genomic DNA, this method allowed identifying numerous and diverse DJ rearrangements from HC Microlocus Light (**Table 6**), HC Microlocus Classic (**Table 7**), HC Microlocus Light shRNA (**Table 8**), and HC Minilocus (**Table 9**) transfected 300-19 cells. In these tables, the sequences of the different DJ rearrangements are presented. For each, the identity of the involved D and J segments is given. At the junction between D and J, the N nucleotides as well as the cropped nucleotides are shown.

Table 6: Rearrangements DJ from HC Microlocus Light. For each rearrangement, the D and J segment identities are given in the two first columns. The sequence corresponding to the involved D segment is in yellow, and the sequence corresponding to the J segment is in blue. The cropped nucleotides in 3' of D and on 5' of J segments are schematized with asterisk (*). The N nucleotides are shown in red. The 3' sequence (*AGGT*) corresponds to the splice donor site for splicing to Cγ1 Hinge.

D	J	DNA sequence (5' to 3')
D3	J1	GTATTACGATTATTGGAGTGGTTATTATAACTA******TGAATACTTCCAGCACTGGGGCCAGGGCACCCTGGTCACCGTCTCCTCAGGT
D2	J4	AGGATATTGTAGTAGTACCAGCTG*********CTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D1	J4	GGTATAACTGGAACGACTAC*****TTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D5	J4	GTGGATACAGCTATGG******TTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D2	J6	AGGATATTGTAGTAGTACCAGCTGCTATG****ACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTCCTCAGGT
D2	J6	AGGATATTGTAGTAGTACCAGCTGCTATG********ACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTCCTCAGGT
D2	J6	AGGATATTGTAGTAGTACCAGCTGCTATG**ATTACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTCCTCAGGT
D2	J4	AGGATATTGTAGTAGTACCAGCTGCT*****T****CTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D1	J4	GGTATAACTGGAACGACTACT***ACTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D3	J4	GTATTACGATTATTGGAGTGGTTATTATAACTATA**G**********CTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D2	J3	AGGATATTGTAGTAGTACCAGCTGCTATG********TTTTGATATCTGGGGCCAAGGGACGATGGTCACCGTCTCTTCAGGT
D1	J6	GGTATAACTGGAACGACTACG**TACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTCCTCAGGT
D3	J4	GTATTACGATTATTGGAGTGGTTATTATAACTATA**GT*****TTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D3	J6	GTATTACGATTATTGGAGTGGTTATTATAACTA*********CTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTCCTCAGGT
D4	J1	TGACTACAGTAACTACTC*A*******ACTTCCAGCACTGGGGCCAGGGCACCCTGGTCACCGTCTCCTCAGGT
D7	J2	CTAACTGGGGAG*****GGTACTTCGATCTCTGGGGCCGTGGCACCCTGGTCACTGTCTCCTCAGGT

Table 7: Rearrangements DJ from HC Microlocus Classic. For each rearrangement, the D and J segment identities are given in the two first columns. The sequence corresponding to the involved D segment is in yellow, and the sequence corresponding to the J segment is in blue. The cropped nucleotides in 3' of D and on 5' of J segments are schematized with asterisk (*). The N nucleotides are shown in red. The 3' sequence (*AGGT*) corresponds to the splice donor site for splicing to Cγ1 CH1.

D	J	DNA sequence (5' to 3')
D2	J4	AGGATATTGTAGTAGTACCAGCTGCTATG**********TTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D2	J6	AGGATATTGTAGTAGTACCAGCTGCTATGC********ACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTC ACCGTCTCCTCAGGT
D2	J6	AGGATATTGTAGTAGTACCAGCTGCTATG**********ACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCA CCGTCTCCTCAGGT
D2	J6	AGGATATTGTAGTAGTACCAGCTGCTATGC*ATTACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGT CACCGTCTCCTCAGGT
D1	J6	GGTATAACTGGAACGAC*******TACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTCCTC AGGT
D1	J6	GGTATAACTGGAACGACT******************ACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTCCTCAGGT
D2	J4	AGGATATTGTAGTAGTACCAGCTGCT***************TTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D1	J4	GGTATAACTGGAACGACTACT*******ACTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D1	J4	GGTATAACTGGAACGACTACC************GACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D1	J4	GGTATAACTGGAACGACTAC**************TACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D1	J4	GGTATAACTGGAACGACTACGGGA************GACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D1	J3	GGTATAACTGGAACGACTACGGG********GCTTTTGATATCTGGGGCCAAGGGACAATGGTCACCGTCTCTTCAGGT
D2	J6	AGGATATTGTAGTAGTACCAGCTGCTA********GTTCATTACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCAC GGTCACCGTCTCCTCAGGT
D1	J1	GGTATAACTGGAACGACTAC*******GAATACTTCCAGCACTGGGGCCAGGGCACCCTGGTCACCGTCTCCTCAGGT
D5	J2	GTGGATACAGCTATGGTTACCCCT**********GTACTTCGATCTCTGGGGCCGTGGCACCCTGGTCACTGTCTCCTCAGGT
D5	J1	GTGGATACAGCTATGGTTACG*******GAATACTTCCAGCACTGGGGCCAGGGCACCCTGGTCACCGTCTCCTCAGGT

Table 8: Rearrangements DJ from HC Microlocus Light shRNA. For each rearrangement, the D and J segment identities are given in the two first columns. The sequence corresponding to the involved D segment is in yellow, and the sequence corresponding to the J segment is in blue. The cropped nucleotides in 3' of D and on 5' of J segments are schematized with asterisk (*). The N nucleotides are shown in red. The 3' sequence (*AGGT*) corresponds to the splice donor site for splicing to Cγ1 Hinge.

D	J	DNA sequence (5' to 3')
D1	J4	GGTATAACTGGAACGACTACT*******ACTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D1	J4	GGTATAACTGGAACGACTACGTAGACTACTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D1	J3	GGTATAACTGGAACGACTAC********GCTTTTGATATCTGGGGCCAAGGGACAATGGTCACCGTCTCTTCAGGT
D1	J3	GGTATAACTGGAACGACTACGCATGATGCTTTTGATATCTGGGGCCAAGGGACAATGGTCACCGTCTCTTCAGGT
D3	J1	GTATTACGATTATTGGAGTGGTTATTATAACTATACCG*********ATACTTCCAGCACTGGGGCCAGGGCACCCTGGTCACCGT CTCCTCAGGT

Table 9: Rearrangements DJ from HC Minilocus. For each rearrangement, the D and J segment identities are given in the two first columns. The sequence corresponding to the involved D segment is in yellow, and the sequence corresponding to the J segment is in blue. The cropped nucleotides in 3' of D and on 5' of J segments are schematized with asterisk (*). The N nucleotides are shown in red. The 3' sequence (*AGGT*) corresponds to the splice donor site for splicing to Cμ CH1.

D	J	DNA sequence (5' to 3')
D2	J6	AGGATATTGTAGTAGTACCAGCTGCTATG*****ACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTC ACCGTCTCCTCAGGT
D1	J6	GGTATAACTGGAACGACTACGCCTTATTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCGT CTCCTCAGGT
D2	J6	AGGATATTGTAGTAGTACCAGCTGCT*******ACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCA CCGTCTCCTCAGGT
D2	J6	AGGATATTGTAGTAGTACCAGCTGCTATG**ACT***ACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACG GTCACCGTCTCCTCAGGT
D2	J6	AGGATATTGTAGTAGTACCAGCTGCTAT***AACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCAC CGTCTCCTCAGGT
D2	J4	AGGATATTGTAGTAGTACCAGCTGCTATG**T****CTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
D1	J6	GGTATAACTGGAACGACTACGCCTTAT*****TACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCG TCTCCTCAGGT

The DJ rearrangements analyzed so far show appreciable recombinatory and junctional diversities like expected. Indeed, for all HC constructs, D and J segment usage is varied and we observed different types of junction. As an example, regarding DJ rearrangements arising from HC Minilocus, four out of seven involve D2 and J6 segments, but the junctions are completely different. The same phenomenon is observed in all rearrangements arising from the three other constructs. This diversity will be explained in part 3.3.11. These results encouraged us to continue with VDJ rearrangement analysis.

3.3.10. VDJ rearrangements analysis

3.3.10.1. Analysis at DNA level

Following DJ rearrangements analysis, another strategy was established to identify VDJ rearrangements in HC Microlocus Light, HC Microlocus Classic, HC Microlocus Light shRNA and HC Minilocus transfected 300-19 cells. We followed the same DNA extraction procedure as for DJ rearrangements analysis.

Because all constructs contain a large number of V gene segments from family 3, it was decided to focus first on V3DJ rearrangement. As forward primer, 1430 (5' GTGCAGCTGGTGGAGTCTGG 3'), which is specific to all V3, V1-58 and V1-2, was designed. The primer matches at the beginning of the V coding region. Primer 1393, matching on the 3' end of J6 was used as reverse primer. The strategy to detect any VDJ rearrangement coming from HC constructs is shown in **Figure 89**.

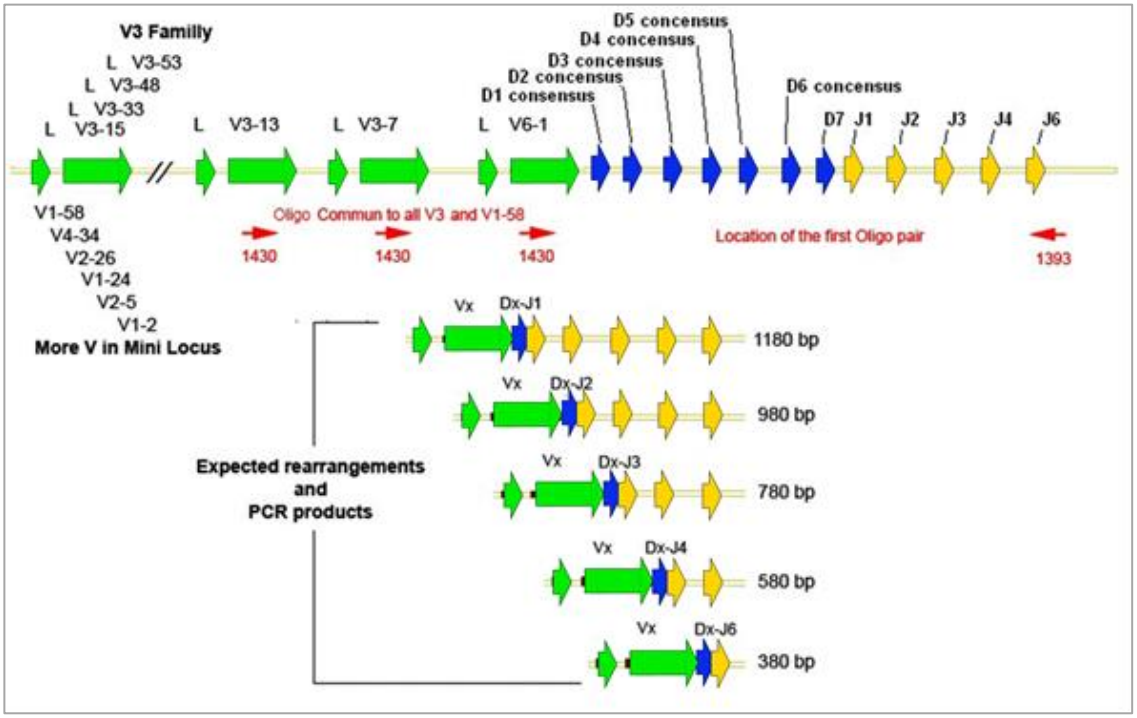

Figure 89: Strategy for V3DJ rearrangements analysis in 300-19 cells transformed with HC constructs. On the top of the schema, the V-D-J cassette and the matching sites of primers are shown. The upper schema shows the size of the expected rearrangements.

Following the same strategy, other forward primers were designed for other V segments:

- V1 family, primer 1461: 5' CTCAGTGAAGGTCTCCTGCAAG 3'
- V2 family, primer 1462: 5' AGACCCTCACGCTGACCTGCAC 3'
- V4 family, primer 1464: 5' CTATGGTGGGTCCTTCAGTGGTT 3'
- V6 family, primer 1465: 5' CATCGAGAGGCCTTGAGTGGCTG 3'

Like for DJ rearrangements, the PCR amplifications were performed with different elongation time, *i.e.* 30 seconds and 2 minutes. When the PCR run completed, samples were loaded on 1,5% agarose gel electrophoresis containing 0.01% of RedSafe DNA staining solution (*ref.*: 21141). Detectable PCR product were extracted from agarose gel, using QIAquick Gel Extraction Kit (*ref*: Qiagen, 28704), and then sequenced. Weaker PCR product or smear were also purified from agarose gel and sub-cloned using TOPO® TA Cloning® Kit from Invitrogen. This sub-cloning step allows sequencing of weakly amplified PCR products. All DNA sequences were analyzed with the program SeqMan.

3.3.10.2. Results

Despite the fact that we often amplified nonspecific products from mouse DNA, this method allowed identifying seven diverse VDJ rearrangements from HC Microlocus Light, HC Microlocus Classic and HC Minilocus transfected 300-19 cells. Theses VDJ sequences are given in **Table 10**. For each, the identity of the involved V, D and J segments is reported. At the junction between V and D, and between D and J, the N nucleotides as well as the cropped nucleotides are shown.

Table 10: VDJ rearrangements characterized in transformed 300-19 cells, at the DNA level. The name of the construct corresponding to the analyzed rearrangements is given in the top of each rearrangement clusters. For each rearrangement, the V, D and J segment identities are given in the three first columns. The sequence corresponding to the involved V segment is in green, the one that corresponds to D segment is in yellow, and the sequence corresponding to the J segment is in blue. The eventual cropped nucleotides in 3' of V, in 3' and 5' of D and in 5' of J segments are schematized with asterisk (*). The N nucleotides are shown in red. The 3' sequence (*AGGT*) corresponds to the splice donor site for splicing to Cγ1 Hinge for HC Microlocus light; to Cγ1 CH1 for HC Microlocus classic; and to Cμ CH1 for HC Minilocus.

V	D	J	DNA sequence (5' to 3')
HC Microlocus light			
V3-7	no D	J1	AACAGCCTGAGAGCCGAGGACACGGCCGTGTATTACTGTGCGAGAG**CTGAATACTTCCAGCACTGGGGCCA GGGCACCCTGGTCACCGTCTCCTCAGGT
V3-7	no D	J6	GCCAAGAACTCACTGTATCTGCAAATGAACAGCCTG**********GGGCAAAGGGACCACGGTCACCGTCTCCTCAGGT
V3-7	D1	J6	AGCCGAGGACACGGCCGTGTATTACTGTGCGAGAG********CTGGAACGAC*****ACTACTACTACTACTACA TGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTCCTCAGGT
HC Microlocus classic			
V3-7	D1	J6	AATGAACAGCCTGAGAGCCGAGGACACGGCCGTGTATTACTGTGCGAGAGATC*********GGAAC******TGG** TACTACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTCCTCAGGT
HC Minilocus			
V1-58	D1	J6	CTGAGATCCGAGGACACGGCCGTGTATTACTGTGC******CC*****AACTGGAACGACTACGCCTTATTACTACT ACTACTACTACATGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTCCTCAGGT
V1-58	no D	J4	ACAAGCACAGCCTACATGGAGCTGAGCAGCCTGAGATCCGAGGACACGGCCGTGTATTACTGTGCGGCAGACT ****CTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGT
V1-58	D1	J3	ACAGCCTACATGGAGCTGAGCAGCCTGAGATCCGAGGACACGGCCGTGTATTACTGTGCGGC***GGGTATAA CTGGAACGACT***GATGCTTTTGATATCTGGGGCCAAGGGACAATGGTCACCGTCTCTTCAGGT

These results demonstrate diversity in gene usage since different V, D and J segments are involved. Remarkably the diversity observed at the junction is quite high. For example, if we consider the third rearrangement identified from HC Microlocus Light and the unique rearrangement arising from HC Microlocus Classic, both involve V3-7, D1 and J6, but the junctions are completely different; for one, D1 segment is dramatically cropped and N nucleotides were added, for the other, no N nucleotide were added and D1 segment extremities are less cropped. In some case, like the two first rearrangements arising from HC Microlocus Light, the D segment has been so cropped that it disappeared. The rearrangement diversity will be better characterized in the part 3.3.11.

3.3.10.3. Analysis at RNA level

We also applied the single cell PCR to evaluate VDJ recombination. This method relies on the prior sorting (one cell per well) of human Cɣ1 HC expressing cells followed by cell lysis, reverse transcription (RT) and VDJ analysis PCR on cDNA. Forward primers bind in the 5' end of V segments, reverse primers bind in Cɣ1 CH1 or in Cɣ1 CH2, depending on the construct. The exact primer sequences are given below:

- primer RS6153 (forward): AAGGTGTCCAGTGTSAGGTGCAG, which is homologous to all leader sequences of V3 segments; S nucleotide: C or G;
- primer RS6154 (forward): GTCCTGTCCCAGGTGCAGCTGCAG, which is homologous to all leader sequences of V4 and V6 segments;
- primer RS6157 (reverse): GGAAGGTGTGCACGCCGCTGGTC, which is homologous to IgG1 CH1 (HC Microlocus Classic); and
- primer 1538 (reverse): CTCACGTCCACCACCACGCA, which is homologous to IgG1 CH2 (HC Microlocus Light).

This method was only applied to HC Microlocus Light and HC Microlocus Classic transfected cells, because these are the only showing an appreciable human ɣ1 HC surface expression rate, suitable for cell sorting. HC Microlocus Light and HC Microlocus Classic 300-19 cells were stained with an optimal quantity of anti-human IgG Fc antibody (*ref.*: eB 12-4998-82). As negative control, wild type 300-19 cells were stained following the same protocol. Samples were incubated 30 minutes at 4°C. HC positive samples were sorted with FACS equipment Aria III. Human IgG expression level was assessed on alive cells based on their size (FSC parameter) and granularity (SSC parameter). Human ɣ1 HC positive cells were sorted at 1 cell per well. After cell lysis and reverse transcription, VDJ PCR was performed on the resulting cDNA.

When PCR program completed, samples were loaded on 1% agarose gel electrophoresis containing 0.01% of RedSafe DNA staining solution (ref: 21141), which allows DNA detection after gel electrophoresis. Gel electrophoresis conditions were 140 volts, during 20 minutes. Detectable PCR products were extracted from agarose gel, using QIAquick

Gel Extraction Kit (ref: Qiagen, 28704), and sequenced. All DNA sequences were analyzed with the program SeqMan.

3.3.10.4. Results

In contrast to the analysis performed on DNA from cell pool, the single cell PCR can only amplify one rearrangement because one single cell transcripts one unique rearrangement (or two, when both alleles are rearranged). This reflects in the visualization of PCR amplifications after gel electrophoresis which is much clearer. An example of gel PCR is shown in **Figure 90**.

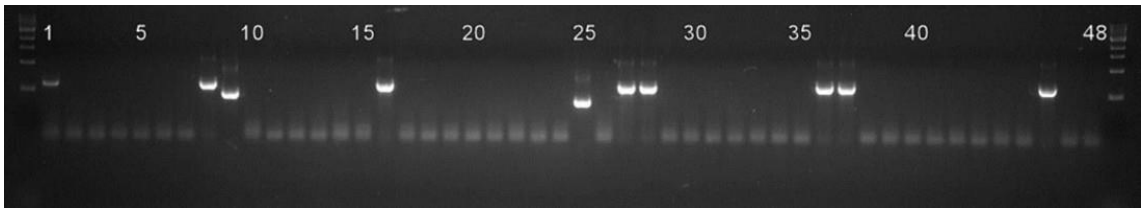

Figure 90: VDJ rearrangements amplification, at the RNA level.
Gel: 1,5% agarose, running conditions: 25 minutes, 140 Volts.
Size marker: GeneRuler 100 bp DNA Ladder (*ref.*: Thermo scientific # SM0242)

The 300-19 cell line is blocked at the pro-B cell stage, and even if these cells can progress through further differentiation stage when propagated in culture, progenitor B cells do not express Ig as much as mature B cells. This implies that in 300-19 cells, VDJ rearrangements transcripts are present in low copy numbers. This may explain why the efficiency of this method is low. However, we were able to identify three different VDJ rearrangements, which are presented in **Table 11**. They all come from HC Microlocus Classic transfected 300-19 cells.

Table 11: VDJ rearrangements characterized in 300-19 cells transformed with HC Microlocus classic, at the RNA level. For each rearrangement, the V, D and J segment identities are given in the three first columns. The sequence corresponding to the involved V segment is in green, the one that corresponds to D segment is in yellow, and the sequence corresponding to the J segment is in blue. The eventual cropped nucleotides in 3' of V, in 3' and 5' of D and in 5' of J segments are schematized with asterisk (*). The N nucleotides are shown in red. For each, the J segment is sliced on Cɣ1 CH1 (*in italic dark blue*)

V	D	J	cDNA sequence (5' to 3')
V3-7	D1	J4	GTGTATTACTGTGCGAGAGA****TAACTGGAACGACTAC*****TTTGACTACTGGGGCCAGGGAACCCT GGTCACCGTCTCCTCAG*CCTCCACCAAGGGCCCATCGGTCTTCCCCC*
V3-7	D1	J3	GTGTATTACTGTGCGAGAGA**TATAACTGGAACGACTAC***TGCTTTTGATATCTGGGGCCAAGGGAC AATGGTCACCGTCTCTTCAG*CCTCCACCAAGGGCCCATCGGTCTTCCCCC*
V6-1	D7	J3	GGCTGTGTATTACTGTGCAAGAGAT***ACTGGG***GATGCTTTTGATATCTGGGGCCAAGGGACAATG GTCACCGTCTCTTCAG*CCTCCACCAAGGGCCCATCGGTCTTCCCCC*

These three rearrangements show diversity in gene usage; indeed different V, D and J segments are involved. The diversity observed at the junction is detectable as well. The rearrangement diversity will be better characterized in the part 3.3.11. Furthermore, in these three cases, the J segment is correctly spliced to the Cɣ1 CH1 constant exon, indicating that mRNA splicing is correctly performed.

3.3.10.5. Unusual rearrangements

By analyzing VDJ rearrangements at the DNA level (3.3.10.1.), we identified two unexpected rearrangements in contradiction with the classical model. We found one V segment rearranged to one D segment and further 3' another D rearranged to one J segment, on the same allele (**Table 12**). Both sequences come from HC Minilocus 300-19 cells.

Table 12: Unconventional rearrangement found in 300-19 cells transformed with HC Minilocus, showing two examples of VD and DJ rearrangements on the same allele. For both sequences, the V, D and J segment identities are given in the three first columns. The sequence corresponding to the involved V segment is in green, the one that corresponds to D segments are in yellow, and the sequence corresponding to the J segment is in blue. The intervening sequence between both D segments is in black. The eventual cropped nucleotides in 3' of V, in 3' and 5' of D and in 5' of J segments are schematized with asterisk (*).The 3' sequence (*AGGT*) corresponds to the splice donor site for splicing to Cμ CH1.

V	D--D	J	cDNA sequence (5' to 3')
V4-34	D1---D2	J6	GTGACCGCCGCGGACACGGCTGTGTATTACTGTGCGAGAG*****TAACTGGAACGACTACC ACAGTGAGAAAAACTGTGACAAAAACCAACTAGCCAGAGACAGCAAGAGGGGACTCAGT GACTCCCGCGGGGACAGGGGTTTTTGTGGGGGCTCGTGTCACAGTGAGGATATTGTAG TAGTACCAGCTGCTATG******ACTACTACTACTACTACATGGACGTCTGGGGCAAAGGGAC CACGGTCACCGTCTCCTCAGGT
V1-58	D1---D2	J4	CCTGAGATCCGAGGACACGGCCGTGTATTACTGTGCGGC***CCG*****AACTGGAACGACT ACCACAGTGAGAAAAACTGTGACAAAAACCAACTAGCCAGAGACAGCAAGAGGGGACTC AGTGACTCCCGCGGGGACAGGGGTTTTTGTGGGGGCTCGTGTCACAGTGAGGATATTG TAGTAGTACCAGCTGCTATG**TCTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTC CTCAGGT

3.3.11. Diversity

In total, 44 DJ and 10 VDJ rearrangements were analyzed so far. Considering these data, the recombinatory and the junctional diversities were evaluated.

3.3.11.1. Recombinatory diversity/ Gene usage

Considering DJ and VDJ rearrangements data, HC Microlocus Light, HC Microlocus Light shRNA, HC Microlocus Classic and HC Minilocus are fully functional to generate diverse Ig HC gene recombinations. The usage frequency of D and J segment was evaluated (number or rearrangement involving a specific segment/total number of rearrangement*100), and is presented as a histogram in **Figure 91**.

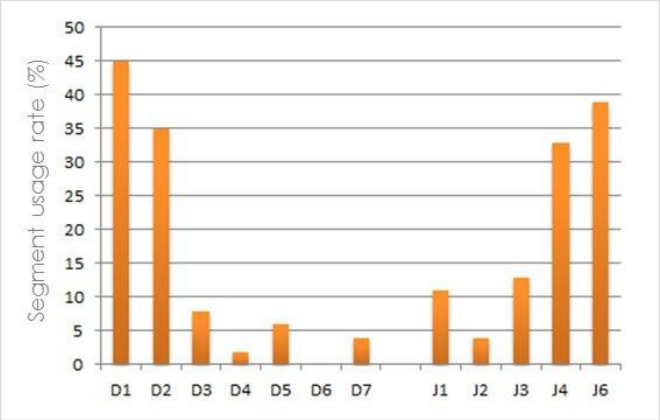

Figure 91: Segment usage frequency (%). Usage rates are calculated considering 44 DJ and 10 VDJ rearrangements.

The most 5' D segments and the most 3' J segments show greater usage rate than the others. This is because almost all rearrangements were identified at the DNA level and considering the method used, PCR product on rearrangements using the most 5' D and the most 3' J segments, have the smallest size and are therefore much easier to amplify and therefore to detect (see **Figure 87**). The V segments involved in VDJ rearrangement analyzed so far are V1-58, V4-34, V6-1 and V3-7. The analysis will be extended to highlight the use of other V segments.

3.3.11.2. Junctional diversity

During gene segments recombination, the junction between V and D, as well as the junction between D and J are imprecise. The segment extremities can be cropped by exonuclease and N nucleotides can be randomly added by terminal deoxynucleotidyl transferase. Considering the considering 44 DJ and 10 VDJ rearrangements identified so far, all types of junction were observed:

- no N nucleotide addition
- 1 to 7 N nucleotides addition
- 3' V segment cropped (between 1 and 38 bp)
- 3' D segment cropped (between 1 and 7 bp)
- 5' D segment cropped (between 1 and 10 bp)
- 5' J segment cropped (between 1 and 32bp)
- Segments extremities cropped and no N addition
- Segments extremities cropped and N addition

As a conclusion, the recombinatory and the junctional diversities arising from HC Microlocus Light, HC Microlocus Classic, HC Microlocus Light shRNA and HC Minilocus, are quite evident. This is the evidence that HC constructs can generate diversity like expected.

3.3.12. Alternative splicing

Since we observed that cells transformed with HC Microlocus Light and HC Microlocus Classic express transmembrane and secreted form of human ɣ1 HC, a strategy was established to test the alternative splicing process at the RNA level. Alternative splicing of the pre-messenger RNA of the heavy chain can yield either a membrane heavy chain, or a secreted heavy chain, which retain the same VDJ rearrangement and the same constant region allowing the production of secreted or membrane bound antibodies.

3.3.12.1. Method

The sequence encoding the 20 C-terminal residues of secreted form (S) is derived from DNA contiguous with the last constant domain exon ɣ1 CH3, whereas in the mRNA encoding transmembrane Ig, the last constant domain exon derives from two exons, TM1 and TM2. A set of primers was designed to allow detection, on one hand, of transcripts leading to secreted protein expression, on the other hand, of transcripts leading to transmembrane protein expression:

- primer 1 (forward): 5’ GGACGTGAGCCACGAAGACCCT 3’, which is homologous to ɣ1CH2,
- primer 2 (reverse): 5’ ATGCTGGGTGCCTGGGAAG 3’, which is homologous to the non-coding sequence between ɣ1S and polyA site, and
- primer 3 (reverse): 5’ TGGCACTGTAGCACACGCTTAAC 3’, which is homologous to ɣ1TM1 exon.

PCR with primers 1 and 2 may therefore detect transcripts for the secreted form of ɣ1; while PCR with primers 1 and 3 may highlight transcripts for the transmembrane form. The matching site of each primer is shown in **Figures 92.**

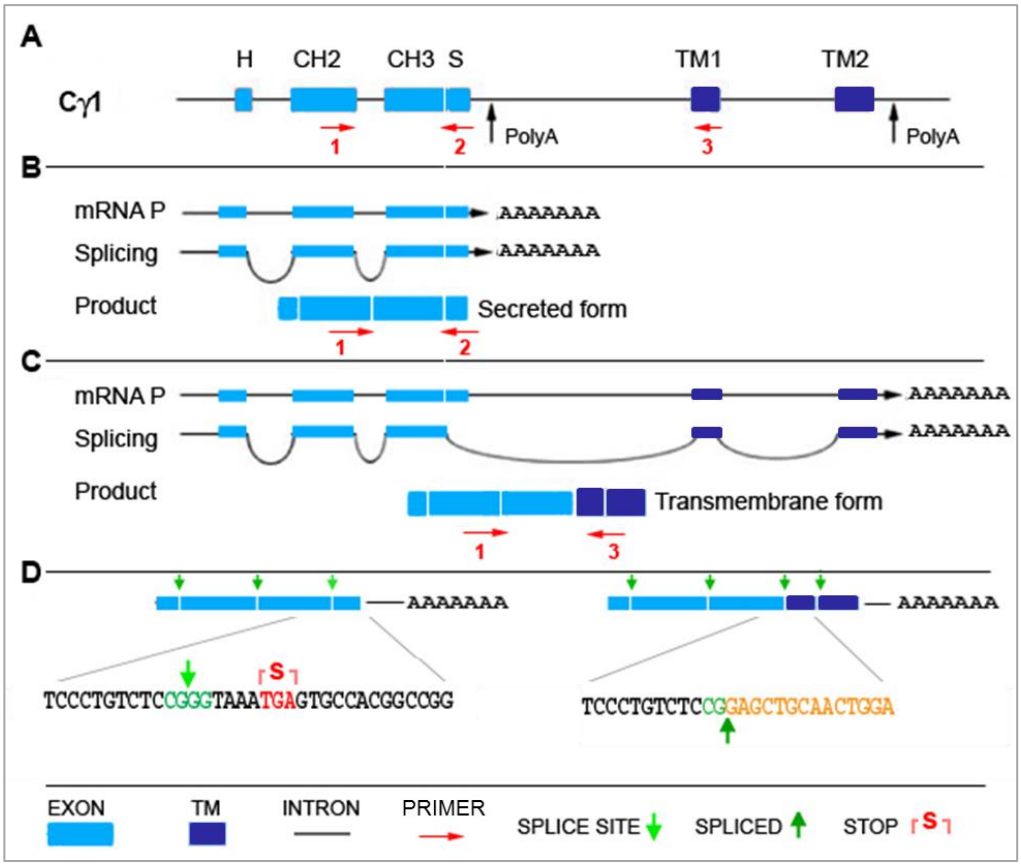

Figure 92: Strategy to analyze alternative splicing.
A: genomic organization of Cɣ1 exon and the two potential polyadenylation sites (PolyA). (H: hinge, S: "secreted" exon, TM1 and TM2: "transmembrane" exons)
B: transcription for the secreted form of ɣ1: the primary transcript is cleaved and polyadenylated at the first polyA site, eliminating the TM exons and giving rise to the secreted form of the heavy chain.
C: transcription for the transmembrane form of ɣ1: the splicing between a splice site located between the Cɣ1 CH3 exon and the Cɣ1 S sequence, and a second splice site at the 5′ end of the Cɣ1 TM exons, results in removal of the Cɣ1 S sequence and joining of the TM exons to the Cɣ1 CH3 exon. This transcript is polyadenylated (AAA) at the second polyA site.
D: The left panel corresponds to the secreted form transcripts with the sequence of ɣ1 CH3 followed by ɣ1 S coding region and the stop codon; the splice site remains intact. The right panel corresponds to the transmembrane form transcript with ɣ1 CH3 spliced on TM1 (yellow sequence).

This analysis requires high quality complementary DNA (cDNA) synthesis. From HC Microlocus Light and HC Microlocus Classic 300-19 cell pellets, mRNA was extracted using RNeasy Kit from QIAGEN®. After reverse transcription of the mRNA using

ThermScript™ RT-PCR System from Invitrogen™, we obtained cDNA, suitable for PCR analyzes. PCR amplifications with primer pairs 1 and 2, and 1 and 3, were performed with 1 minute elongation time. When the PCR run was completed, samples were loaded on 1% agarose gel electrophoresis containing 0.01% of RedSafe DNA staining solution (*ref.*: 21141). PCR products showing the right size, *i.e.* 550 bp for the secreted form transcript and 670 bp for the transmembrane form transcript, were extracted from agarose gel, using QIAquick Gel Extraction Kit (*ref*: Qiagen, 28704), and then sequenced. All DNA sequences were analyzed with the program SeqMan.

3.3.12.2. Results

After PCR product sequencing and sequence alignment, we were able to identify both transcript types; one encoding the transmembrane form (**Figure 93**) and the other one encoding the secreted form (**Figure 94**). In the case of transmembrane form transcript, ɣ1 CH3 exon is correctly spliced to ɣ1 TM1 exon. In parallel, ɣ1 S exon remains contiguous with ɣ1 CH3 exon in the case of secreted form transcript. This experiment demonstrated that alternative splicing works as expected in 300-19 cells transformed with HC Microlocus Light and HC Microlocus Classic.

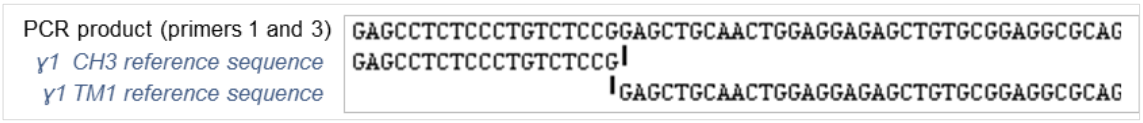

Figure 93: Sequence alignment. The sequence of the PCR amplicon is on top. ɣ1 CH3 and ɣ1 TM1 reference sequences come from IMGT database.

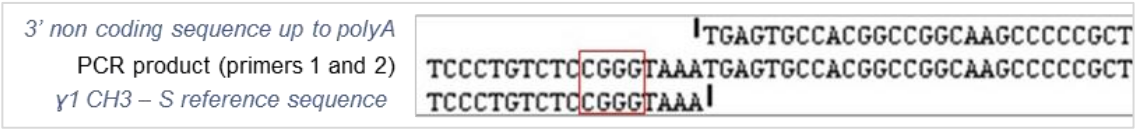

Figure 94: Sequence alignment. The sequence of the PCR amplicon is in the middle. ɣ1 CH3-S reference sequence and the non-coding sequence between ɣ1 CH3-S 3' up to polyA site come from IMGT database. The splice site used for splicing to TM1 remains intact (*red box*).

3.4. Transgenesis in mice

In order to generate transgenic mice that produce human antibodies we build up yeast artificial chromosomes that cover the almost entire human HC, LCκ and LCλ *loci*; and, reduced *loci* in form of plasmids, encoding the human HC. The generation of both types of genetic constructs was described in chapter 3.1. and chapter 3.2. respectively. We tested first our reduced human *loci* in lymphoid mouse 300-19 cell line. Such experiments, described in detail in the previous chapter (3.3), demonstrated that the HC Microlocus Light, the HC Microlocus Light shRNA and the HC Microlocus Classic were correctly processed by the murine recombinatory-, transcriptional- and translational machineries, leading to human HC expression on 300-19 cell surface and secretion into the cell media. Thus, we decided to inject the HC Microlocus Light shRNA into WT mice oocytes and, the HC Microlocus Light as well as the HC Microlocus Classic into the HC knock-out (KO) mouse oocytes.

Standard micro-injection of our three vectors in both, WT and HC KO mice derived oocytes, were performed at the Friedrich Miescher Institute (Basel) and by Novartis, according to the procedure described by Nagy and colleagues [97]. Twenty days after birth offspring mice were genotyped to identify transgenic founders.

Mouse biopsies were treated with DirectPCR Lysis Reagent from ViagenBiotech (*ref.*: 402-E); according to the protocol delivered with the lysis reagent, PCR reactions were directly performed on 0.5 uL of lysate with specific primers. PCR products were loaded on agarose gel, and those showing the correct size were extracted from the gel, using QIAquick Gel Extraction Kit (ref: Qiagen, 28704), and then sequenced. Positive genotyped mice were analyzed by FACS, to demonstrate the presence of human IgG, in the blood. FACS analyses were performed according to the method described in part 3.3.3.1. In order to identify human IgG expression in mouse B cells, anti-mouse CD19 antibody coupled with fluorochrome PE (*ref.*: BL 115508) for B cell detection and anti-human IgG Fc antibody coupled with fluorochrome PE (*ref.*: eB 12-4998-82) for transgene expression evaluation were used.

In this section, the generation of an HC Microlocus Light shRNA transgenic mice line and the phenotyping of the HC Microlocus Light and the HC Microlocus Classic into HC KO mouse will be presented.

3.4.1. Generation of HC KO mice

In order to produce human antibodies in transgenic mice, the endogenous antibody *loci* have to be silenced. Therefore, an Ig HC knock-out mouse line was generated by deleting the mouse heavy chain J region according to the previously described Ig gene knock-out strategy [98]. The absence of J segment in the mouse HC *locus* prevents HC DNA rearrangement. It results in the blocking of the B cell differentiation in the stage of early pro B cells, which depends of surface HC expression to progress to the next step. Therefore these mice do not produce any antibody. However these mice maintain functional trans-acting factors for antibody rearrangements and expression and therefore provide the proper genetic background for introducing the germ-line human HC *loci*.

3.4.2. Injections in HC KO background

3.4.2.1. HC Microlocus Light transgenesis

The HC Microlocus Light vector was prepared for oocyte micro-injection as followed. *E. Coli* bacteria containing the construct were propagated in Lysogeny Broth (LB) medium containing 100 μg/mL of ampicillin. Plasmids were purified from 100 mL of bacterial culture, using QIAGEN Large-Construct Kit (*ref*: 12462) for alkaline cell lysis, followed by phenol/chloroform purification and DNA precipitation using sodium acetate/ethanol mixture. DNA was diluted at 1μg/mL, in sterile environment, in injection buffer composed of Tris 5 mM and EDTA 0.1 mM.

Micro-injection of circular HC Microlocus Light plasmid into HC KO mouse oocytes were performed by Thomas Hennek according to standard procedure [97].

After micro-injections of HC Microlocus Light construct into 400 oocytes derived from HC KO mice, 112 mice were born. 17 mice were transgenic since they show human HC specific PCR products. FACS analyses on blood cells were performed regularly during 10 months. Unfortunately, no CD19+ B cells were found in these 17 transgenic mice.

3.4.2.2. HC Microlocus Classic transgenesis

Like for HC Microlocus Light, large quantity of DNA was purified and sterile diluted in injections buffer. After micro-injections of 200 oocytes derived from HC KO mice, 60 mice were born and 5 were transgenic as shown by human HC specific PCR products. Unfortunately, FACS analyses on blood cells from these transgenic animals did not show any CD19+ cell compartment reconstitution.

3.4.3. Mice HC Microlocus light shRNA in WT background

Plasmid DNA of HC Microlocus light shRNA was prepared for WT mouse oocyte injection as described in 3.4.2.1. Micro-injection of WT mouse oocytes [97] were performed by Jean-Francois Spetz at the Friedrich Miescher Institute of Novartis in Basel.

3.4.3.1. Genotyping

Around 300 B6 mice oocytes were injected and 218 were implanted into pseudo-pregnant B6 foster mother mice. A total of 17 mice were born (identified as numbers #1 to #17). PCR analysis revealed that 5 mice have the transgene integrated in their genome: #5, 7, 13, 16 and 17. This indicates good transgenesis efficiency. Whole blood cells of the five positive genotyped mice were analyzed by FACS to evaluate human IgG expression. An example of genotyping PCR in shown in **Figure 95**, the PCR was performed on 0.5uL of biopsy lysate, using primers 1252 and 1253, which match specifically in V3-15. The expected PCR product size is 221 bp. PCR products from mice #5, 7, 13, 16 and 17 correspond perfectly to the human V3-15 sequence. All other signals correspond to unspecific amplification of mouse genes.

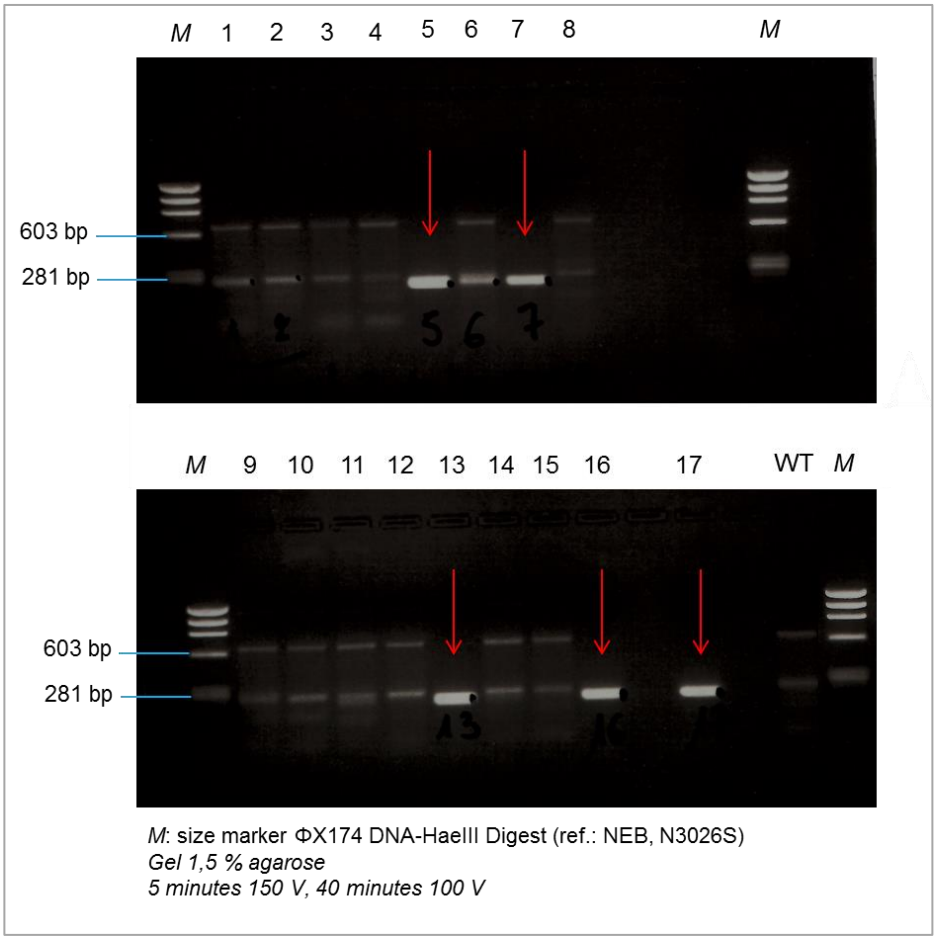

Figure 95: Example of PCR genotyping, mice HC Microlocus light shRNA #1 to #17. PCR products from mice #5, 7, 13, 16 and 17 (red arrows) show the correct size (~221 bp).

3.4.3.2. Detection of human ɣ1 HC by FACS

Genotyped positive mice were tested for human ɣ1 HC expression by FACS. Blood cells were analyzed for surface as well as intracellular human ɣ1 HC expression. FACS assays were performed following the method described in part 3.3.3.1. Only mouse HC Microlocus light shRNA #5 1 year old showed intracellular (**Figure 96**) and surface (**Figure 97**) human ɣ1 HC positive cells. The results are presented below.

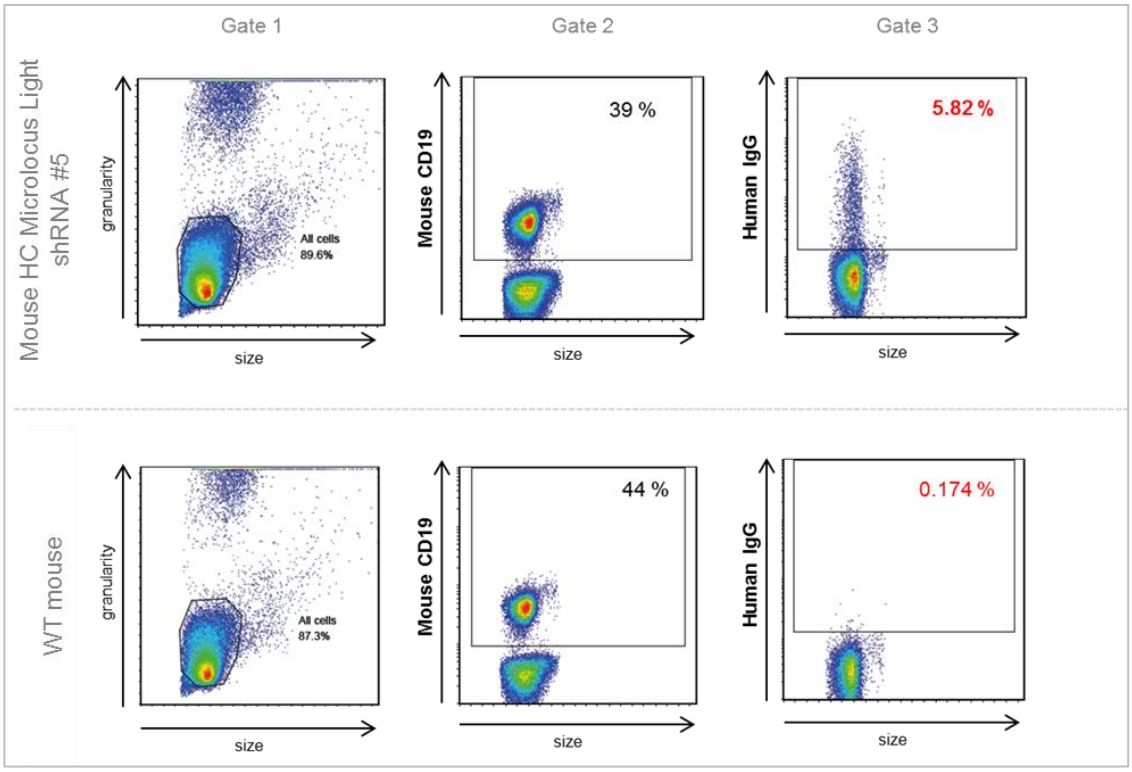

Figure 96: Intracellular detection of human ɣ1 HC, in mouse HC Microlocus Light shRNA #5 blood cells (*upper panel*), compared to wild type mouse blood cells (*lower panel*). Alive cells gating is based on their size and granularity (*gate 1*). Mouse CD19 positive cells were enumerated on alive cells (*gate 2*); finally, human ɣ1 HC expressing cells were evaluated on mouse CD19 positive cell pool (*gate 3*). The percentage of intracellular human ɣ1 HC expressing cells is in red color, in the top right-hand corner of the dot plot.

The transgenic mouse HC Microlocus Light shRNA #5 has shown more than 5% of mouse CD19 positive cells expressing intracellular human ɣ1 HC. This indicates that the construct can be processed and expressed intracellularly *in vivo*.

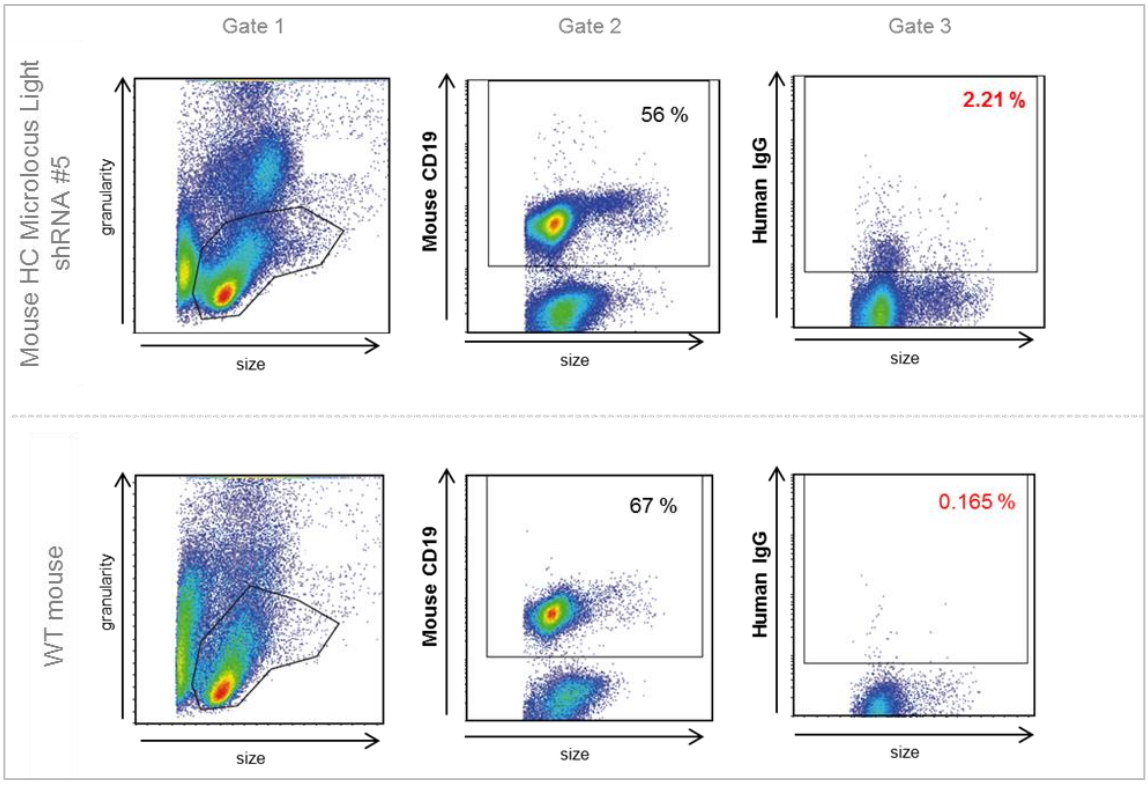

Figure 97: Detection of surface human ɣ1 HC, in mouse HC Microlocus Light shRNA #5 blood cells (*upper panel*), compared to wild type mouse blood cells (*lower panel*). Alive cells gating strategy is based on their size and granularity (*gate 1*). Mouse CD19 positive cells were assessed on alive cells (*gate 2*); finally, human ɣ1 HC expressing cells were evaluated on mouse CD19 positive cell pool (*gate 3*). The percentage of surface human ɣ1 HC expressing cells is in red color, in the top right-hand corner of dot plot.

The transgenic mouse HC Microlocus Light shRNA #5 has shown more than 2% of mouse CD19 positive cells expressing surface human ɣ1 HC. By comparing the intracellular expression percentage (~5.5%, **Figure 96**), there are less cells expressing human ɣ1 HC on cell surface. This observation could indicate difficulties for some human ɣ1 HC to be exported to the cell surface.

3.4.3.3. shRNA effect

Initially, we inserted, in 5' of the V cluster of HC Microlocus Light, a DNA piece coding for four pairs of shRNA targeting mouse IgM transcript in order to down-regulate its expression with the aim of promoting transgenic expression. The transcription of the shRNA part is independently regulated by the promoter 7SK, a RNA polymerase III promoter, routinely used to express shRNA molecules. In this assay we aimed to evaluate the efficacy of this specific repression.

To evaluate shRNA effect, blood cells of three HC Microlocus Light shRNA mice were assessed for mouse IgM expression:

- HC Microlocus Light shRNA #5 mouse, which is transgenic and express human ɣ1 HC;
- HC Microlocus Light shRNA #7 mouse, which is transgenic and does not express human ɣ1 HC; and,
- WT mouse

Cells were stained with anti-mouse IgM antibody, coupled with fluorochrome FITC (*ref.*: BD 560575) and analyzed by flow cytometry.

Cell samples were acquired with FACS equipment LSR Fortessa. Analyses were performed with FlowJo software. Mouse IgM expression level was assessed on alive cells based on their size (FSC parameter) and granularity (SSC parameter). In order to evaluate mouse IgM expression, FITC intensity curves as a function of cell number were overlaid (**Figure 98**).

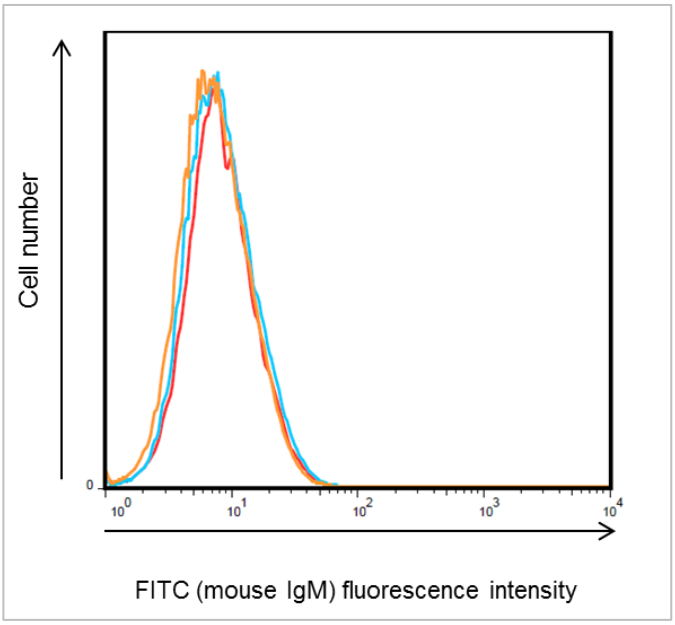

Figure 98: Mouse IgM expression rate comparison – estimation of shRNA effect. FITC intensity as a function of cell number curves superposition. *Orange*: mouse #7; *blue*: mouse #5; *red*: WT mouse.

Any shift would have been representative of mouse μ expression's down-regulation but this is not the case. The analysis performed shows that the shRNA approach does not have any effect on mouse IgM expression.

3.4.3.4. VDJ rearrangement

Since we detected intracellular and surface expression of human HC in mouse HC Microlocus Light shRNA #5, we analyzed the VDJ rearrangements.

The VDJ rearrangement analysis, following the method described in 3.3.10.3. part, is still ongoing. The only rearrangement found so far is presented in **Table 13**.

Table 13: VDJ rearrangement characterized in mouse HC Microlocus Light shRNA #5, at the RNA level. The V, D and J segment identities are given in the three first columns. The sequence corresponding to the involved V segment is in green, the one that corresponds to D segment is in yellow, and the sequence corresponding to the J segment is in blue. The cropped nucleotides in 3' of V, in 3' and 5' of D and in 5' of J segments are schematized with asterisk (*). The randomly added N nucleotides are shown in red. The J segment is sliced on Cɣ1 Hinge (*in italic dark blue*).

V	D	J	cDNA sequence (5' to 3')
V3-7	D1	J4	AGGACACGGCCGTGTATTACTGTGCGAGA**AGGGGGGG******ACTGGA*******CCGG****CTTTGATT ACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAG*AGCCCAAATCTTGTGACAAAACTCACAC* *ATGCCCACCGT*

This unique rearrangement shows imprecise junction between V and D, as well as between D and J; indeed extremities segments are more or less cropped, and N nucleotides were added. This sequence analysis shows a correct mRNA splicing; effectively J4 segment is correctly spliced on Cɣ1 Hinge coding region.

3.4.3.5. Breeding of Mouse HC Microlocus Light shRNA #5 and offspring mice analyses

Since mouse HC Microlocus Light shRNA #5 shows human ɣ1 HC expression, we started a breeding program between this mouse and wild type mouse. With this breeding we expected to have the half of offspring mice bearing the transgene. Transgene expression depends on transgene integration in the mouse genome. Transgene integration fashion in offspring mice will stay the same as in mouse HC Microlocus Light shRNA #5. Therefore, all transgenic offspring should express human ɣ1 HC in the same way as mouse HC Microlocus Light shRNA #5.

Eighty offspring mice were born and are identified with numbers #18 to #87. They were genotyped and analyzed by FACS, following the method described in 3.3.3.1., in order to evaluate human ɣ1 HC expression.

Like for the first generation, offspring mice were genotyped. Biopsies were treated with DirectPCR Lysis Reagent from ViagenBiotech; and genotyping PCR were directly performed with 0.5 uL of lysate.

Like expected, half of progeny had the transgene integrated in genome: 42 mice out of 80 were genotyped as positive. Positive genotyped mice were then evaluated for ɣ1 HC expression. Blood cells were therefore stained with an anti-mouse CD19 antibody coupled with fluorochrome PE (*ref.*: BL 115508) for B cell detection and an anti-human IgG Fc antibody coupled with fluorochrome PE (*ref.*: eB 12-4998-82) for transgene expression evaluation.

This analysis revealed expression of human ɣ1 HC in all 42 mice whole blood cells. Indeed around 5% of blood cells express cytoplasmic human ɣ1 HC and between 0.3 and 1.5 % of blood cells express human ɣ1 HC on surface. These results are fully consistent with what we expected.

3.4.3.6. Serum analysis of human ɣ1 HC and mouse IgM

We evaluated the serum levels of human ɣ1 HC and mouse IgM by ELISA. We followed the method described in 3.3.6.1. As capture antibodies we used a goat anti-human IgG antibody (*ref.*: 2040-01), and a goat anti-mouse IgM antibody (*ref.*: 1021-01). Detection antibodies are the same, except that they are conjugated to alkaline phosphatase (*ref.*: 2040-04, 1021-04). All these antibodies were purchased from Southern Biotech.

Mouse HC Microlocus Light shRNA #5 as well as positive genotyped offspring expressing human ɣ1 HC, have around 30 µg/mL of secreted human ɣ1 HC in serum. Regarding mouse IgM level, there is no significant difference between mice expressing the transgene and WT mouse; they both show a serum level of around 7 mg/mL.

3.4.3.7. Human γ1 HC expression in progenitor B cell subsets

HC Microlocus light shRNA is identical to the HC Microlocus Light concerning the antibody gene part, since both encode human γ1 HC $CH1^{-}$. B cell development occurs in bone marrow through different stages. One of the first critical point in B cell development is the expression of rearranged HC gene. In association with surrogate LC and co-receptors Igα/β, HC polypeptide is presented to cell surface as pre-BCR.

Signaling *via* the pre-BCR provides feedback about the functionality of the recombined HC gene, so only pre-B cells that express a signaling-competent receptor can further mature. Disruption in assembly, expression, and therefore in signaling capability, results in developmental arrest.

Considering the data shown in previous section (3.4.2.), we know that HC Microlocus Light in mouse HC KO background does not lead to B cell reconstitution; this may indicate that the human HC *locus* is not able to replace the disrupted endogenous HC *locus* necessary for B cell development. In contrast, HC Microlocus light shRNA in mouse WT background shows human γ1 HC expression in mouse peripheral blood cells. The hypothesis is that γ1 HC $CH1^-$ does not participate to B cell development and is only expressed in parallel to endogenous mouse immunoglobulin expression. To verify this, we evaluated the expression of the transgene in different bone marrow B cell subsets, namely pro-, pre-, immature and mature B cells, of a mouse HC Microlocus shRNA. In pro-B cell subset, no mouse μ HC is expressed; in pre-B cells mouse μ HC is expressed as pre-BCR; immature and mature B cells express mouse IgM. We attempted to see if the expression of the transgene follows the same evolution.

These analyzes were performed on lymphocytes isolated from bone marrow from mouse HC Microlocus shRNA #20. This mouse, offspring of mouse HC Microlocus shRNA #5, was genotyped positive and showed intracellular and surface human ɣ1 HC expression. Moreover we could detect serum human γ1 HC (40 μg/mL).

Method and results

The purpose of this analysis was to evaluate the expression of human ɣ1 HC in different bone marrow (BM) B cell subsets, *i.e.* in pro-, pre-, immature- and mature B cells. Mouse HC Microlocus Light shRNA #20 was sacrificed at one year of age. Lymphoid mononuclear cells were isolated by flushing out the BM with a syringe in complete medium. Each progenitor B cell subset expresses particular cell surface markers (**Table 14**), thus by staining bone marrow cells with appropriate antibodies, we were able to discriminate the different subsets and evaluate γ1 HC $CH1^-$ expression in each.

Table 14: Murine cell surface markers used to discriminate different B cell subsets in bone marrow

MOUSE B cell subsets in bone marrow					
Early Lymphoid Progenitor (ELP)	Early Pro B cell	Pro B cell	Pre B cell	immature B cell	mature B cell
B220 - CD43+	B220+ CD43 low CD19-	B220+ CD43+ CD19+ cμ-	B220+ CD43 low CD19+ cμ+	B220+ CD19+ IgM+	B220 high CD19 high IgM+

Following the method described in 3.3.3.1., BM cells of HC Microlocus light shRNA #20 and WT mice have been labelled with FITC anti-mouse CD43 (*ref*.: eBioscience, 11-0431), APC anti-human IgG (*ref*.: Jackson Immunoresearch, 109-606-098), PerCP anti-mouse B220 (*ref*.: BD, 553093), V450 anti-mouse IgM antibody (*ref*.: BD, 560575), and PE-Cy7 anti-mouse CD19 (*ref*.: BD, 561739). Each antibody was used in its optimal concentration. After staining, cells were acquired in the LSR Fortessa.

Pro-B cell subset was defined as $B220^{low}$, $CD19^{low}$, IgM^{-} and $CD43^{+}$. The frequency of pro-B cell is increased in transgenic mouse compared with WT mouse. This could indicate a potential alteration in the early step of B cell development in the transgenic mouse; analysis of several mice is needed in order to confirm this observation. Cell surface expression of human ɣ1 HC was evaluated in this cell subset. The electronically gating used is presented in **Figure 99**.

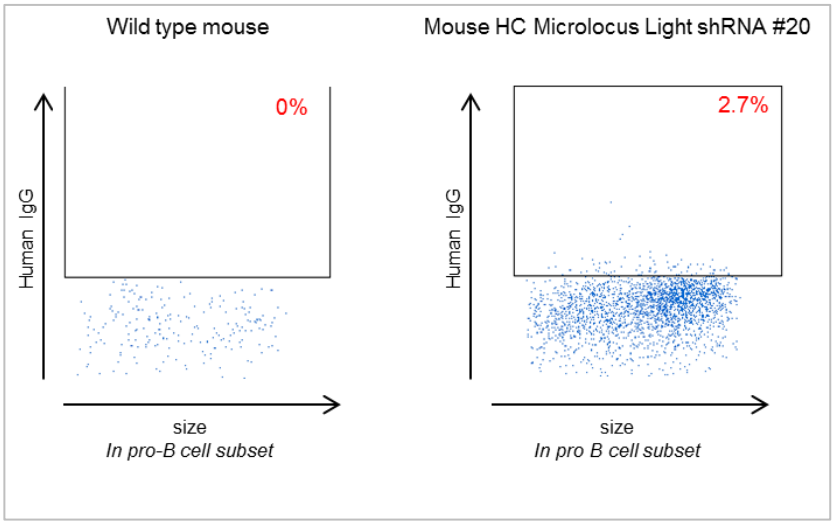

Figure 99: Human ɣ1 HC expression in BM pro-B cell subset. The percentage of positive cells is indicated in the top right-hand corner of the dot plot. Analyses on WT mouse cells (*left panel*) serve as negative controls.

Theoretically, at the pro-B cell stage no µ HC is expressed, but a few percentage (2.7%) of these cells already express human ɣ1 HC on cell surface. This result may indicate that, in few cells, the transgene encoding human HC is recombined and express slightly earlier than the endogenous one.

In a second step, pre-B cells were defined as $B220^{low}$, $CD19^{low}$, IgM^{-} and $CD43^{-}$. The frequency of pre-B cell is comparable between transgenic and WT mice. Cell surface expression of human ɣ1 HC was evaluated in this cell subset. The electronically gating used is presented in **Figure 100**.

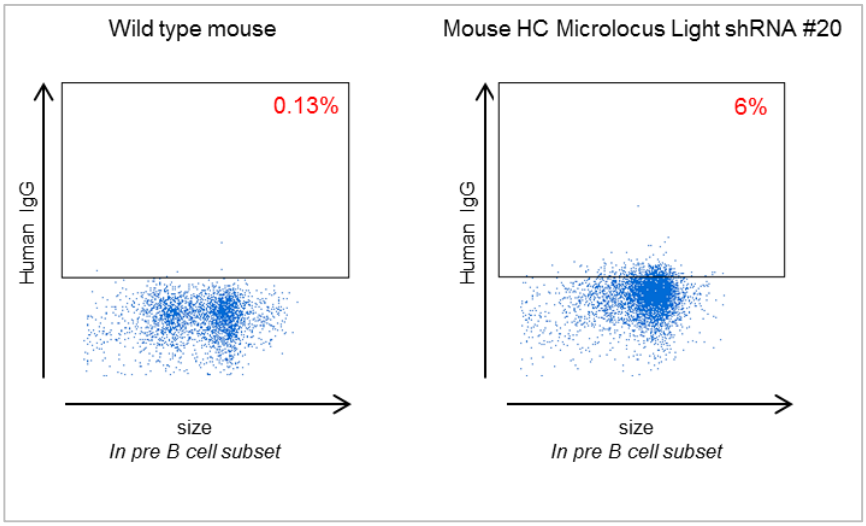

Figure 100: Human ɣ1 HC expression in BM pre-B cell subset. The percentage of positive cells is indicated in the top right-hand corner of the dot plot. WT mouse cells (*left panel*) were used as negative controls.

At the pre-B cell stage, all cells express μ HC as pre-BCR, and 6% of them also express human ɣ1 HC. To progress to the next developmental stage, *i.e.* immature B cell, pre-BCR is controlled for functionality. It is supposed that this control is only supported by endogenous μ HC. If this is the case, the 6% of cells co-expressing μ HC and ɣ1 HC will also progress to immature B cell stage (assuming that their endogenous μ HC involved in pre-BCR is functional).

Thus, immature B cell subset was defined as $B220^{low}$, $CD19^{low}$, IgM^{+} and $CD43^{-}$. The frequency of immature B cell is comparable between transgenic and WT mice. Cell surface expression of human ɣ1 HC was evaluated in this cell subset. The electronically gating used is presented below in **Figure 101**.

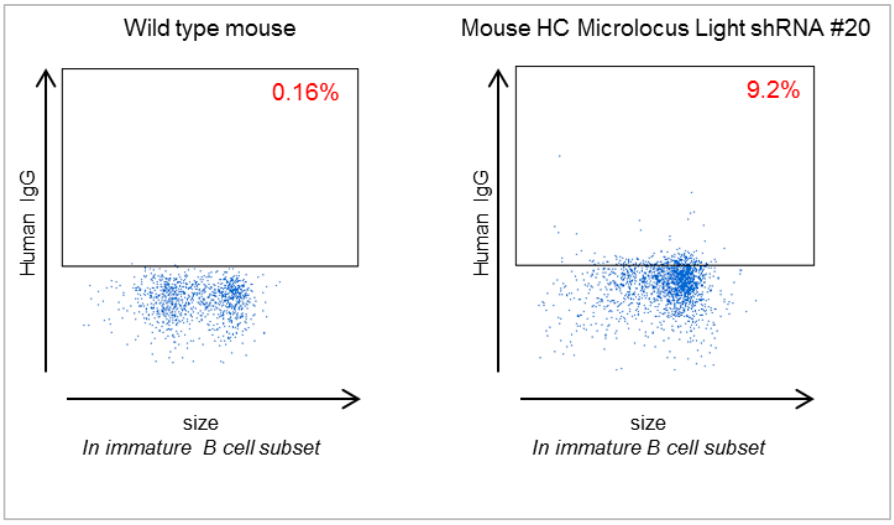

Figure 101: Human ɣ1 HC expression in BM immature B cell subset. The percentage of positive cells is indicated in the top right-hand corner of the dot plot. WT mouse cells (*left panel*) were used as negative controls.

Immature B cells express IgM, and 9% of these cells also express human ɣ1 HC on cell surface. This result indicates that the previous 6% of pre-B cells co-expressing mouse μ HC and human ɣ1 HC were able to progress to the next developmental stage.

Finally, the mature cell subset was defined as $B220^{high}$, $CD19^{high}$ and IgM^{+}. The frequency of immature B cell was equivalent in transgenic and in WT mice. Cell surface expression of human ɣ1 HC was evaluated in this cell subset. The electronically gating used is presented below in **Figure 102**.

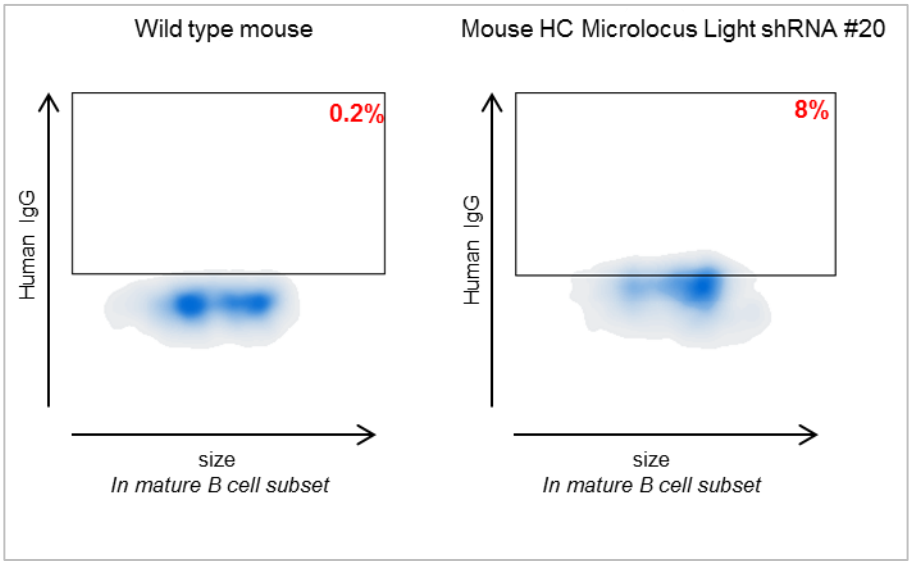

Figure 102: Human ɣ1 HC expression in BM mature B cell subset. The percentage of positive cells is indicated in the top right-hand corner of the dot plot. WT mouse cells (*left panel*) were used as negative controls.

In mature B cell population, 8% of cells express mouse µ and human ɣ1 HC. This observation indicates that endogenous IgM and transgenic human ɣ1 HC are co-expressed during B cell development.

This experiment suggests that human ɣ1 HC expression follow the same evolution as endogenous mouse Cµ expression in B cell development. Indeed, the human ɣ1 HC expression increases during the B cell development process: pro B cell subset display the lowest percentage of human IgG positive cells while the highest frequency is detectable in immature B cell population.

The most important observations are that endogenous mouse IgM and transgenic human ɣ1 HC seem to be co-expressed at the surface of mature B cells, and that no cell population expressing only human ɣ1 HC were identified. This suggests that the endogenous Cµ expression is necessary for B cell development, because pre-BCR composed of human ɣ1 HC may not trigger adequate signaling and therefore the B cell differentiation is interrupted. These results will be confronted to those obtained in HC KO mice (3.4.2.) and further discussed in the conclusion.

Discussion

With the goal of generating a mouse producing human antibodies, the human Ig genes were cloned following two approaches. As main approach, the almost entire human Ig *loci* coding for the Ig HC and LC were cloned as circular yeast artificial chromosomes (YACs). A second alternative approach aimed to engineer a reduced HC *locus*, the HC Minilocus. Through this approach, three HC Microlocus versions emerged as useful side products.

Regarding the main approach, the very large size of the human Ig *loci* (1 250 kbp for HC *locus*, 1 820 kbp for κ LC *locus* and 1 050 kbp for λ LC *locus*) led us to use circular YACs as they are the unique DNA molecules with such a high size cloning capacity. Thus, we generated three YACs (one for each human Ig *loci*) by developing a new approach derived from the Transformation Associated Recombination (TAR), which is a method that enables to clone by homologous recombination any gene or *locus* directly from genomic DNA template. The disadvantages of using YACs include DNA breakage, complex purification procedure and the risk of unwanted genetic recombination in yeast cells. Moreover, the injection of YACs into mice oocytes, as well as the fusion between mouse ES cells and yeast containing YACs, in order to generate transgenic mice are not well established methods. They can lead to undesired or complex transgene integration. Therefore, as a backup of the risky YAC approach, an alternative strategy was designed. This intended to generate a largely reduced human Ig HC *locus* by using standard molecular biology technologies and well established cloning methods.

- Genetic constructs: reduced human Ig HC *loci*

Meticulously selected antibody gene segments and genetic control elements were stepwise assembled into reduced Ig *loci* resulting in up to 50 times size reduction compared with the natural genetic counterpart. The representative gene segments were selected to maintain an acceptable diversity and to contain all required genetic control elements for optimal processing and expression of human HC in mouse B cells. Four constructs were created. The HC Minilocus contains 13 V segments, a synthetic D-J region and the constant regions Cμ, Cɣ3 and Cɣ1. The three other constructs were

derived from intermediate steps and are based on a set of 7 V segments and the same synthetic D-J region. One has the Cɣ1 exon (HC Microlocus Classic). The two others contain Cɣ1 lacking the CH1 exon (HC Microlocus Light) and consequently, encode Heavy Chain only antibodies. One of the two last constructs also contains an independently regulated DNA sequence coding for shRNA targeting mouse Cμ transcripts (HC Microlocus Light shRNA). Notably, all constructs were designed to be easily modified using standard molecular biology methods, allowing all kind of combinations.

- *In vitro* functionality analysis

Before starting the transgenesis by oocyte injection, the functionality of the four reduced human HC constructs was assessed by transfecting the 300-19 cell line, a mouse pro-B cell line able to process endogenous or transgenic Ig genes from rearrangement to Ig protein expression [6]. Intermediate constructs HC Microlocus Light, HC Microlocus Light shRNA and HC Microlocus Classic functionality was deeply investigated and almost all steps of antibody generation were analyzed. These include DJ and VDJ rearrangements, recombinatory diversity, junctional diversity, transcription, mRNA maturation, alternative splicing, protein surface expression and secretion. HC Minilocus transformed cell FACS analysis was initially not possible due to the too strong expression of mCherry fluorescent protein, used as cell transfection efficiency readout. After removing the *mCherry* gene from this construct, transfected cells could be re-analyzed using similar approaches.

Transgene expressing cells were further characterized by evaluating by FACS mouse VpreB, λ5, Igβ, Igα, LCκ, LCλ and IgM surface expression. No mouse LCλ and no mouse IgM were detected. However, HC Microlocus Classic transformed cells expressed human ɣ1 HC, which seems to be associated with mouse surrogate light chain components (VpreB, λ5). In addition, these cells start weakly to express mouse LCκ, showing that human HC can associate with mouse LC and be expressed at the cell surface. On the other hand, human ɣ1 HC expressing cells showed a much higher percentage of Igβ positive cells when compared with WT 300-19 cells. This suggests

that in these cells, the expression of the constructs drives the export of cell surface molecules crucial for pre-BCR assembly and signaling.

The secreted human ɣ1 HC were characterized by Western Blot. The first results showed the expected size difference between HC secreted by HC Microlocus Light (~40-45 kDa / chain) and by HC Microlocus Classic (~50 kDa / chain). After human IgG1 purification, the corresponding proteins could be visualized on SDS-PAGE. According to the size, they could be purified from the gel and characterized by mass spectrometry.

Knowing that the recombinatory and junctional diversities are significant, HC Microlocus Light transformed cells expressing human ɣ1 HC constitute a source for libraries of primary antibodies having undergone no negative selection. This means that these antibodies have a recognition potential of a wide range of antigens. The same type of library might be obtained from HC Microlocus Classic cells after transfecting them with a construct encoding the human LC.

- *In vivo* functionality analysis

The results obtained in the 300-19 cell line encourage us to proceed to transgenesis by injecting the HC Microlocus Light and HC Microlocus Classic in HC knock out mouse oocytes and the HC Microlocus Light shRNA in wild type mouse oocytes. Human HC constructions injected in wild type mice led to human γ1 HC expression, whereas injections in HC KO mice did not show any B cell population reconstitution and consequently no human HC γ1 expression.

Transgenesis in mouse HC KO background

The HC Microlocus light construct encodes γ1 HC lacking CH1 domain, which is inspired by HC only antibodies discovered in *camelidae*. In 1989, it has been found that chimeric proteins composed of a molecule of interest (for example cell surface marker like CD4), in association with Ig constant part lacking CH1 domain can be produced in mouse myeloma cells [99]. More recent studies have demonstrated the perfect processing and expression of rearranged dromedary HC constructs in mouse myeloma cell [100]. Indeed, the specific HC only antibodies were secreted as disulfide linked homodimer

and displayed on the mouse cell surface, independently of LC. This proves the correct processing of such a construct in mature B cell, but its correct participation in B cell development remains unproven. The HC Microlocus Light was injected in HC KO mice oocytes. Such mice show a block in B cell differentiation at the pro-B cell stage and consequently, they lack mature CD19 positive cells in blood and in secondary lymphoid organs. Therefore, we expected that the injection of the construct that encodes HC would allow B cell compartment reconstitution through HC expression. HC Microlocus Light transgenic mice were generated, but we did not observe mouse CD19 positive cell development. The HC Microlocus Classic, which encodes classical human γ1 HC was also injected in HC KO mouse oocytes. In this case, we obtained only few HC Microlocus Classic transgenic mice and none of them showed mature B cell development. Several assumptions could explain these results. First, we cannot exclude the lack of transgene expression, *i.e.* we do not know if pro-B cells express cytoplasmic human Ig chains. Secondly, the absence of mature B cell does not exclude the possibility that human γ1 HC can participate to the early stage of B cell development, but without reaching the mature B cell stage. This could be verified by analyzing progenitor B cell compartments of these mice. Most likely is that the constructs lead to human γ1 HC expression but do not participate to B cell differentiation.

Transgenesis in mouse WT background

The HC Microlocus Light shRNA was injected in WT mouse oocytes and one out of five transgenic mice has shown 2% of blood B cell expressing surface γ1 HC $CH1^-$. We started a breeding program between this mouse and wild type mouse, leading to a colony of 42 transgenic mice in which the expression of the transgene was confirmed. Considering that HC Microlocus Light in HC KO background did not lead to transgene expression and the fact that HC Microlocus light shRNA (which is exactly identical to the Microlocus Light concerning the antibody gene part) is expressed in mouse wild type background; this led us to the hypothesis that γ1 HC $CH1^-$ can only be expressed, in parallel to normal endogenous mouse Ig expression. To better characterize transgene processing and expression, we investigated the human γ1 HC expression in progenitor B cell compartment in bone marrow of one of these mice. For that purpose, a transgenic

mouse HC Microlocus light shRNA was sacrificed. This mouse was positively genotyped and FACS analysis on blood sample showed that 1.7% of blood B cells were expressing human γ1 HC at the surface. Progenitor B cell analysis has shown that the human ɣ1 HC expression rate seems to increase with maturity of cells: the lowest rate is in the pro-B cell population (2.7%), while the highest is in the immature B cell population (9%). Finally, 8% of mature B cells co-expressed human γ1 HC and mouse IgM at the cell surface. It will thus be important to test whether *in vitro* and *in vivo* triggering of the ɣ1 HC receptor can lead to activation of the mature B cells.

Conclusion and perspectives

We demonstrated that the design of the reduced human antibody HC *loci* is perfectly functional in all biological aspects when analyzed in a transfected pro B cell line. The VDJ rearrangements including addition of random nucleotides, gene segment usage, translational and post-translational processing up to the mature protein expression behaved like the natural counterparts. Nevertheless, whereas the main goal has been reached, some issues of this work remain to be clarified.

- Unexpected rearrangements

VDJ rearrangement analysis in transformed 300-19 cells allowed us to make an unexpected new observation. The classical Ig HC rearranging model argues for a sequential two steps recombination process, were in the first step one D segment rearranges to one J segment and in a second step a V joins the previously rearranged DJ [101]. In contrast to this model, we detected one V segment rearranged to one D segment and further 3', another D segment rearranged to one J segment, on the same allele. This finding, which was only possible because of the small size of the construct, could contradict the sequential rearranging model. Does this finding reflect the reality or is it a particularity of our artificial gene cluster? On one hand, VD-DJ analysis in natural genetic environment, and specially on the same allele, was almost impossible to perform at the time the sequential rearranging model was described. Since the native mouse D cluster covers more than 70 kbp, the two events located on the same chromosome, may be separated by a very large portion of genomic DNA. On the other hand, we cannot exclude that our finding, based on a largely reduced genetic organization, is an artifact. This can only be partially answered by analyzing *in vivo* endogenous mouse recombination events and looking for the presence of VD and DJ rearrangements in the same cell. This can be performed by targeted PCR analysis but this would only show if VD events are present but not if the two events are located on the same chromosome. To address this latest question, cloning by TAR the region of interest covering the two involved segment recombinations would be the most appropriate solution.

- B cell differentiation issues

As we initiated transgenesis with some of our constructs, we noticed that the HC Microlocus Light and HC Microlocus Classic introduced into endogenous HC KO mice did not allow reconstituting the missing B cell compartment as expected. From this observation we suspected that the transgenic products, human γ1 HC $CH1^-$ and human γ1 HC respectively, do not support B cell hematopoiesis. One of the early stages of B cell differentiation is the production of a functional HC in the form of a pre B cell receptor (pre-BCR) on the surface of pre B lymphocytes. If this is achieved, the pre-BCR is able to initiate intracellular signaling, which drives the cell into the next differentiation stage. Only recently, the mode of pre-BCR function was described [7]. In the natural situation, the pre-BCR is present on the cell surface as a μHC homodimer associated with the surrogate light chains VpreB and λ5. Also Igα and Igβ signaling molecules are included in the complex. The binding of λ5 to the first constant region domain of μHC *via* a conserved N linked glycan mediates the initiation of signal transduction. This conserved N-linked glycosylation is specific to IgM isotype. As a consequence, since HC Microlocus Light and HC Microlocus Classic only encode ɣ1 HC, λ5 has no partner to associate properly and therefore pre-BCR signaling cannot be initiated, resulting in a B cell differentiation blockage at the pro B cell stage. Unfortunately we could not predict this issue since this mode of pre-BCR function was not published at the time we engaged transgenesis in mice. In order to experimentally test the human γ1 (with or without CH1 domain) signaling failure, we could, in HC Microlocus Classic, exchange the γ1 CH1 with μ CH1 exon and inject this adapted construct in HC KO mouse oocytes. A positive result would lead to normal B cell development and confirm the requirement on μ CH1 participation in pre-BCR signaling.
The most promising approach, close to the natural situation, to produce normal human antibodies in transgenic mice, is the HC Minilocus which meets all the today known required criteria to give the best results. HC Minilocus was injected in HC KO mouse oocytes but no transgenic animals were obtained so far.

- Human γ1 HC CH1$^-$ as bystander in WT mice

The HC Microlocus Light shRNA is identical to the HC Microlocus Light discussed above, except an additional independent gene coding for mouse IgM shRNA. The purpose of shRNA usage was to reduce the endogenous IgM production and therefore promote human Ig expression. We generated HC Microlocus Light shRNA transgenic mice with WT background and detected human γ1 HC CH1$^-$ on B cell surface of these animals. This was unexpected considering the results obtained in HC Microlocus Light transgenic mice with HC KO background.

We noticed that the shRNA does not seem to have fulfilled its role since the endogenous IgM expression and distribution are comparable to non-transgenic WT mice. We observed that the human Ig was always present on cells having also mouse Ig on the surface. Since we interpreted the results in KO mice as γ1 CH1$^-$ signaling failure, the presence of human antibodies in these WT background mice was a surprise. This could only be explained if the human Ig behaves like "blind passengers" and that the endogenous mouse Ig takes over the B cell differentiation. This is a unique situation were the human antibodies are transported by the B cells without being subject to the B cell differentiation check points since they are not able to signal. As a consequence, the human γ1 HC, co-expressed with mouse IgM on WT mouse B cells, was not subject to selection and human γ1 HC molecules will constitute a virgin antibody repertoire which could lead to the generation of human heavy chain only antibody libraries with a significant innovative opening for new therapeutically biologics discovery. They can hopefully be considered as the most potent antibodies to interact with human molecules, especially against cross-species conserved epitopes, in the context of new therapeutic molecule discovery.

The "blind passenger" theory discussed above attempts to explain the γ1 HC CH1$^-$ signaling failure during B cell hematopoiesis. However, the functionality of the human γ1 HC CH1$^-$ bystander involved in BCR on peripheral mature B cells should be addressed in the HC Microlocus Light shRNA transgenic mice through B cell activation *via* specific human BCR stimulation.

Bibliography

1. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575-581 (1983).

2. Lonberg, N. *et al.* Antigen-specific human antibodies from mice comprising four distinct genetic modifications. *Nature* **368**, 856-859 (1994).

3. Kouprina, N. & Larionov, V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. *Nature protocols* **3**, 371-377 (2008).

4. Hamers-Casterman, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446-448 (1993).

5. Alt, F., Rosenberg, N., Lewis, S., Thomas, E. & Baltimore, D. Organization and reorganization of immunoglobulin genes in A-MULV-transformed cells: rearrangement of heavy but not light chain genes. *Cell* **27**, 381-390 (1981).

6. Reth, M.G., Ammirati, P., Jackson, S. & Alt, F.W. Regulated progression of a cultured pre-B-cell line to the B-cell stage. *Nature* **317**, 353-355 (1985).

7. Ubelhart, R. *et al.* N-linked glycosylation selectively regulates autonomous precursor BCR function. *Nature immunology* **11**, 759-765 (2010).

8. Edelman, G.M., Benacerraf, B., Ovary, Z. & Poulik, M.D. Structural differences among antibodies of different specificities. *Proceedings of the National Academy of Sciences of the United States of America* **47**, 1751-1758 (1961).

9. Snell, G.D. Methods for the study of histocompatibility genes. *Journal of genetics* **49**, 87-108 (1948).

10. Kohler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495-497 (1975).

11. Hozumi, N. & Tonegawa, S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proceedings of the National Academy of Sciences of the United States of America* **73**, 3628-3632 (1976).

12. Capecchi, M.R. Altering the genome by homologous recombination. *Science* **244**, 1288-1292 (1989).

13. Spangrude, G.J., Heimfeld, S. & Weissman, I.L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58-62 (1988).

14. Kondo, M., Weissman, I.L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661-672 (1997).

15. Schenten, D. & Medzhitov, R. The control of adaptive immune responses by the innate immune system. *Advances in immunology* **109**, 87-124 (2011).

16. Hardy, R.R. & Hayakawa, K. B cell development pathways. *Annual review of immunology* **19**, 595-621 (2001).

17. Victora, G.D. & Nussenzweig, M.C. Germinal centers. *Annual review of immunology* **30**, 429-457 (2012).

18. Pieper, K., Grimbacher, B. & Eibel, H. B-cell biology and development. *The Journal of allergy and clinical immunology* **131**, 959-971 (2013).

19. Germain, R.N. T-cell development and the CD4-CD8 lineage decision. *Nature reviews. Immunology* **2**, 309-322 (2002).

20. Hansen, T.H. & Bouvier, M. MHC class I antigen presentation: learning from viral evasion strategies. *Nature reviews. Immunology* **9**, 503-513 (2009).

21. Cresswell, P. Assembly, transport, and function of MHC class II molecules. *Annual review of immunology* **12**, 259-293 (1994).

22. Griesser, H. & Mak, T.W. The T-cell receptor--structure, function, and clinical application. *Hematologic pathology* **8**, 1-23 (1994).

23. Maki, R. *et al.* The role of DNA rearrangement and alternative RNA processing in the expression of immunoglobulin delta genes. *Cell* **24**, 353-365 (1981).

24. Schamel, W.W. & Reth, M. Monomeric and oligomeric complexes of the B cell antigen receptor. *Immunity* **13**, 5-14 (2000).

25. Nguyen, V.K., Hamers, R., Wyns, L. & Muyldermans, S. Camel heavy-chain antibodies: diverse germline V(H)H and specific mechanisms enlarge the antigen-binding repertoire. *The EMBO journal* **19**, 921-930 (2000).

26. Vu, K.B., Ghahroudi, M.A., Wyns, L. & Muyldermans, S. Comparison of llama VH sequences from conventional and heavy chain antibodies. *Molecular immunology* **34**, 1121-1131 (1997).

27. Su, C., Nguyen, V.K. & Nei, M. Adaptive evolution of variable region genes encoding an unusual type of immunoglobulin in camelids. *Molecular biology and evolution* **19**, 205-215 (2002).

28. Janssens, R. *et al.* Generation of heavy-chain-only antibodies in mice. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15130-15135 (2006).

29. Ichihara, Y., Matsuoka, H. & Kurosawa, Y. Organization of human immunoglobulin heavy chain diversity gene loci. *The EMBO journal* **7**, 4141-4150 (1988).

30. Pallares, N., Lefebvre, S., Contet, V., Matsuda, F. & Lefranc, M.P. The human immunoglobulin heavy variable genes. *Experimental and clinical immunogenetics* **16**, 36-60 (1999).

31. Schable, K.F. & Zachau, H.G. The variable genes of the human immunoglobulin kappa locus. *Biological chemistry Hoppe-Seyler* **374**, 1001-1022 (1993).

32. Matthyssens, G., Hozumi, N. & Tonegawa, S. Somatic generation of antibody diversity. *Annales d'immunologie* **127**, 439-448 (1976).

33. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. *Nature* **286**, 676-683 (1980).

34. Dudley, D.D., Chaudhuri, J., Bassing, C.H. & Alt, F.W. Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Advances in immunology* **86**, 43-112 (2005).

35. Desiderio, S.V. *et al.* Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. *Nature* **311**, 752-755 (1984).

36. Bentolila, L.A. *et al.* Extensive junctional diversity in Ig light chain genes from early B cell progenitors of mu MT mice. *J Immunol* **162**, 2123-2128 (1999).

37. Benedict, C.L., Gilfillan, S., Thai, T.H. & Kearney, J.F. Terminal deoxynucleotidyl transferase and repertoire development. *Immunological reviews* **175**, 150-157 (2000).

38. Rabbitts, T.H., Forster, A., Dunnick, W. & Bentley, D.L. The role of gene deletion in the immunoglobulin heavy chain switch. *Nature* **283**, 351-356 (1980).

39. Leff, S.E., Rosenfeld, M.G. & Evans, R.M. Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Annual review of biochemistry* **55**, 1091-1117 (1986).

40. Herzog, S., Reth, M. & Jumaa, H. Regulation of B-cell proliferation and differentiation by pre-B-cell receptor signalling. *Nature reviews. Immunology* **9**, 195-205 (2009).

41. Kitamura, D., Roes, J., Kuhn, R. & Rajewsky, K. A B cell-deficient mouse by targeted disruption of the membrane exon of the immunoglobulin mu chain gene. *Nature* **350**, 423-426 (1991).

42. Elkabetz, Y., Argon, Y. & Bar-Nun, S. Cysteines in CH1 underlie retention of unassembled Ig heavy chains. *The Journal of biological chemistry* **280**, 14402-14412 (2005).

43. Geier, J.K. & Schlissel, M.S. Pre-BCR signals and the control of Ig gene rearrangements. *Seminars in immunology* **18**, 31-39 (2006).

44. Engel, H., Rolink, A. & Weiss, S. B cells are programmed to activate kappa and lambda for rearrangement at consecutive developmental stages. *European journal of immunology* **29**, 2167-2176 (1999).

45. Takeda, S. *et al.* Deletion of the immunoglobulin kappa chain intron enhancer abolishes kappa chain gene rearrangement in cis but not lambda chain gene rearrangement in trans. *The EMBO journal* **12**, 2329-2336 (1993).

46. Gorman, J.R. & Alt, F.W. Regulation of immunoglobulin light chain isotype expression. *Advances in immunology* **69**, 113-181 (1998).

47. Pelanda, R. & Torres, R.M. Central B-cell tolerance: where selection begins. *Cold Spring Harbor perspectives in biology* **4**, a007146 (2012).

48. Tiegs, S.L., Russell, D.M. & Nemazee, D. Receptor editing in self-reactive bone marrow B cells. *The Journal of experimental medicine* **177**, 1009-1020 (1993).

49. Goodnow, C.C., Adelstein, S. & Basten, A. The need for central and peripheral tolerance in the B cell repertoire. *Science* **248**, 1373-1379 (1990).

50. Green, L.L. Antibody engineering via genetic engineering of the mouse: XenoMouse strains are a vehicle for the facile generation of therapeutic human monoclonal antibodies. *Journal of immunological methods* **231**, 11-23 (1999).

51. Zuckier, L.S., Rodriguez, L.D. & Scharff, M.D. Immunologic and pharmacologic concepts of monoclonal antibodies. *Seminars in nuclear medicine* **19**, 166-186 (1989).

52. Choy, E.H., Schantz, A., Pitzalis, C., Kingsley, G.H. & Panayi, G.S. The pharmacokinetics and human anti-mouse antibody response in rheumatoid arthritis patients treated with a chimeric anti-CD4 monoclonal antibody. *British journal of rheumatology* **37**, 801-802 (1998).

53. Winter, G. & Milstein, C. Man-made antibodies. *Nature* **349**, 293-299 (1991).

54. Corti, D. & Lanzavecchia, A. Broadly neutralizing antiviral antibodies. *Annual review of immunology* **31**, 705-742 (2013).

55. Elliott, M.J. *et al.* Repeated therapy with monoclonal antibody to tumour necrosis factor alpha (cA2) in patients with rheumatoid arthritis. *Lancet* **344**, 1125-1127 (1994).

56. Riechmann, L., Clark, M., Waldmann, H. & Winter, G. Reshaping human antibodies for therapy. *Nature* **332**, 323-327 (1988).

57. Burton, D.R. & Barbas, C.F., 3rd Human antibodies to HIV-1 by recombinant DNA methods. *Chemical immunology* **56**, 112-126 (1993).

58. Boder, E.T., Raeeszadeh-Sarmazdeh, M. & Price, J.V. Engineering antibodies by yeast display. *Archives of biochemistry and biophysics* **526**, 99-106 (2012).

59. Jakobovits, A., Amado, R.G., Yang, X., Roskos, L. & Schwab, G. From XenoMouse technology to panitumumab, the first fully human antibody product from transgenic mice. *Nature biotechnology* **25**, 1134-1143 (2007).

60. Brinster, R.L. *et al.* Expression of a microinjected immunoglobulin gene in the spleen of transgenic mice. *Nature* **306**, 332-336 (1983).

61. Schedl, A. *et al.* A method for the generation of YAC transgenic mice by pronuclear microinjection. *Nucleic acids research* **21**, 4783-4787 (1993).

62. Bruggemann, M. *et al.* A repertoire of monoclonal antibodies with human heavy chains from transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 6709-6713 (1989).

63. Taylor, L.D. *et al.* A transgenic mouse that expresses a diversity of human sequence heavy and light chain immunoglobulins. *Nucleic acids research* **20**, 6287-6295 (1992).

64. Gordon, J.W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A. & Ruddle, F.H. Genetic transformation of mouse embryos by microinjection of purified DNA. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 7380-7384 (1980).

65. Gordon, J.W. & Ruddle, F.H. Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* **214**, 1244-1246 (1981).

66. Mansour, S.L., Thomas, K.R. & Capecchi, M.R. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348-352 (1988).

67. Green, L.L. *et al.* Antigen-specific human monoclonal antibodies from mice engineered with human Ig heavy and light chain YACs. *Nature genetics* **7**, 13-21 (1994).

68. Mendez, M.J. *et al.* Functional transplant of megabase human immunoglobulin loci recapitulates human antibody response in mice. *Nature genetics* **15**, 146-156 (1997).

69. Green, L.L. & Jakobovits, A. Regulation of B cell development by variable gene complexity in mice reconstituted with human immunoglobulin yeast artificial chromosomes. *The Journal of experimental medicine* **188**, 483-495 (1998).

70. Harding, F.A. & Lonberg, N. Class switching in human immunoglobulin transgenic mice. *Annals of the New York Academy of Sciences* **764**, 536-546 (1995).

71. Lonberg, N. Human antibodies from transgenic animals. *Nature biotechnology* **23**, 1117-1125 (2005).

72. Taussig, M.J. Molecular genetics of immunoglobulins. *Immunology. Supplement* **1**, 7-15 (1988).

73. Mendez, M.J. *et al.* Analysis of the structural integrity of YACs comprising human immunoglobulin genes in yeast and in embryonic stem cells. *Genomics* **26**, 294-307 (1995).

74. Burke, D.T., Carle, G.F. & Olson, M.V. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806-812 (1987).

75. Osoegawa, K. *et al.* A bacterial artificial chromosome library for sequencing the complete human genome. *Genome research* **11**, 483-496 (2001).

76. Zhang, X.M. & Huang, J.D. Combination of overlapping bacterial artificial chromosomes by a two-step recombinogenic engineering method. *Nucleic acids research* **31**, e81 (2003).

77. Kouprina, N., Noskov, V.N., Koriabine, M., Leem, S.H. & Larionov, V. Exploring transformation-associated recombination cloning for selective isolation of genomic regions. *Methods Mol Biol* **255**, 69-89 (2004).

78. Ma, H., Kunes, S., Schatz, P.J. & Botstein, D. Plasmid construction by homologous recombination in yeast. *Gene* **58**, 201-216 (1987).

79. Larionov, V. *et al.* Transformation-associated recombination between diverged and homologous DNA repeats is induced by strand breaks. *Yeast* **10**, 93-104 (1994).

80. Noskov, V.N. *et al.* Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning. *Nucleic acids research* **29**, E32 (2001).

81. Huber, C., Huber, E., Lautner-Rieske, A., Schable, K.F. & Zachau, H.G. The human immunoglobulin kappa locus. Characterization of the partially duplicated L regions. *European journal of immunology* **23**, 2860-2867 (1993).

82. Jakobovits, A. *et al.* Germ-line transmission and expression of a human-derived yeast artificial chromosome. *Nature* **362**, 255-258 (1993).

83. Schedl, A., Grimes, B. & Montoliu, L. YAC transfer by microinjection. *Methods Mol Biol* **54**, 293-306 (1996).

84. Noskov, V.N. *et al.* Isolation of circular yeast artificial chromosomes for synthetic biology and functional genomics studies. *Nature protocols* **6**, 89-96 (2011).

85. Siebenlist, U., Ravetch, J.V., Korsmeyer, S., Waldmann, T. & Leder, P. Human immunoglobulin D segments encoded in tandem multigenic families. *Nature* **294**, 631-635 (1981).

86. Ruiz, M., Pallares, N., Contet, V., Barbi, V. & Lefranc, M.P. The human immunoglobulin heavy diversity (IGHD) and joining (IGHJ) segments. *Experimental and clinical immunogenetics* **16**, 173-184 (1999).

87. Vincent-Fabert, C. *et al.* Genomic deletion of the whole IgH 3' regulatory region (hs3a, hs1,2, hs3b, and hs4) dramatically affects class switch recombination and Ig secretion to all isotypes. *Blood* **116**, 1895-1898 (2010).

88. Sakai, E., Bottaro, A. & Alt, F.W. The Ig heavy chain intronic enhancer core region is necessary and sufficient to promote efficient class switch recombination. *International immunology* **11**, 1709-1713 (1999).

89. Peterson, C.L., Orth, K. & Calame, K.L. Binding in vitro of multiple cellular proteins to immunoglobulin heavy-chain enhancer DNA. *Molecular and cellular biology* **6**, 4168-4178 (1986).

90. Zou, X. *et al.* Heavy chain-only antibodies are spontaneously produced in light chain-deficient mice. *The Journal of experimental medicine* **204**, 3271-3283 (2007).

91. Rao, D.D., Vorhies, J.S., Senzer, N. & Nemunaitis, J. siRNA vs. shRNA: similarities and differences. *Advanced drug delivery reviews* **61**, 746-759 (2009).

92. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811 (1998).

93. Shaner, N.C., Steinbach, P.A. & Tsien, R.Y. A guide to choosing fluorescent proteins. *Nature methods* **2**, 905-909 (2005).

94. Schlissel, M.S., Corcoran, L.M. & Baltimore, D. Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. *The Journal of experimental medicine* **173**, 711-720 (1991).

95. Fang, T., Smith, B.P. & Roman, C.A. Conventional and surrogate light chains differentially regulate Ig mu and Dmu heavy chain maturation and surface expression. *J Immunol* **167**, 3846-3857 (2001).

96. Cohen, S., Haimovich, J. & Hollander, N. Distinct processing of the pre-B cell receptor and the B cell receptor. *Molecular immunology* **54**, 115-121 (2013).

97. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. Microinjection setup. *CSH protocols* **2006** (2006).

98. Jakobovits, A. *et al.* Analysis of homozygous mutant chimeric mice: deletion of the immunoglobulin heavy-chain joining region blocks B-cell development and antibody production. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 2551-2555 (1993).

99. Traunecker, A., Schneider, J., Kiefer, H. & Karjalainen, K. Highly efficient neutralization of HIV with recombinant CD4-immunoglobulin molecules. *Nature* **339**, 68-70 (1989).

100. Nguyen, V.K. *et al.* Heavy-chain only antibodies derived from dromedary are secreted and displayed by mouse B cells. *Immunology* **109**, 93-101 (2003).

101. Alt, F.W. *et al.* Ordered rearrangement of immunoglobulin heavy chain variable region segments. *The EMBO journal* **3**, 1209-1219 (1984).

UNIVERSITÉ DE STRASBOURG

NOVARTIS

Aurore Villemin

Generation of transgenic vectors encoding human immunoglobulins, functionality assays and transgenesis in mice

Résumé

Dans le but de générer une souris transgénique produisant des anticorps humains, les trois *loci* humains (HC, LCκ et LCλ) ont été reconstitués sous forme de YAC circulaires. Puis, pour simplifier la manipulation des gènes codant pour la chaîne lourde, quatre vecteurs plasmidiques qui résument l'information génétique contenue dans le *locus* humain de la chaîne lourde ont été réalisés. La construction HC Minilocus (78 kbp) est donc composée des 13 segments V sélectionnés, de la région D-J synthétisée et des gènes codant pour les parties constantes de type μ, γ3 et γ1. Trois autres constructions (~22 kbp) ont été dérivées des étapes intermédiaires de clonage. Elles sont basées sur 7 segments V et la région D-J synthétisée. Le HC Microlocus Classic (22 kb) a été obtenu après clonage du gène codant pour la partie constante γ1 en 3' des éléments V, D et J précédemment cités; cette construction code pour des chaînes lourdes d'IgG1 (γ1 HC). De la même façon, le HC Microlocus Light (21.5 kb) contient le gène codant pour γ1 CH1$^-$, cette construction code donc pour des chaînes lourdes IgG sans domaine CH1 (γ1 HC CH1$^-$). Les chaînes lourdes de ce type sont exprimées sans être associées à des chaînes légères, à l'image de ce qui est observé chez les camélidés (Heavy Chain only antibodies). Enfin le HC Microlocus Light shRNA (22 kbp) est une construction basée sur le Microlocus Light à laquelle a été ajoutée une séquence codant pour quatre shRNA (small hairpin RNA) visant à réprimer l'expression d'IgM murin. Ce locus est destiné à être injecté dans des souris natives (wilde type, WT). La fonctionnalité de ces quatre vecteurs a été évaluée par transfection dans la lignée cellulaire 300-19, des lymphocytes pro B d'origine murine pouvant progresser du stade pro B au stade de cellule B mature, lorsqu'elles sont maintenues en culture. Pour les quatre constructions, plusieurs réarrangements DJ et VDJ ont été identifiés, montrant une grande diversité combinatoire et jonctionelle. Par cytométrie en flux (FACS), des chaînes lourdes humaines ont été identifiées dans le cytoplasme et à la surface des cellules transfectées avec les trois constructions intermédiaires. Les cellules exprimant des chaînes lourdes en surface ont été enrichies par FACS. Par la suite, en utilisant des méthodes immuno-enzymatiques (ELISA et ELISPOT), il a été prouvé que les chaines lourdes d'anticorps humains étaient secrétées dans le milieu extracellulaire.
La transgénèse avec les constructions HC Microlocus Light et HC Microlocus Light shRNA s'est révélée efficace puisque nous avons obtenu plusieurs animaux transgéniques. Aucune cellule B n'a été détectée dans les souris HC Microlocus Light (background HC KO); alors que cette même construction associée à un shRNA (HC Microlocus Light shRNA) dans une souris transgénique au système immunitaire humoral intact, montre l'expression de chaînes lourdes d'anticorps humains dans le cytoplasme et à la surface des cellules B. La chaine lourde humaine est exprimée en parallèle des immunoglobulines endogènes à la surface des cellules pre-B, ainsi que dans les cellules B immatures et matures.

Mots-clés : Anticorps thérapeutiques, souris transgénique humanisée, *loci* d'anticorps, souris HC KO, transgènes humains, cellules 300-19, réarrangement VDJ, diversité combinatoire, diversité jonctionnelle, expression d'immunoglobuline.

Résumé en anglais

In order to generate a transgenic mouse producing human antibodies, three human *loci* (HC, LCκ et LCλ) were reconstituted in the form of circular YAC. Then, to simplify the manipulation of genes encoding the heavy chain, four vectors summarizing the genetic information contained in the human heavy chain *locus* were designed and cloned. The HC Minilocus (78 kbp) is composed of 13 V segments, a synthetic DJ cluster and the genes encoding the constant parts Cμ, Cγ3 and Cγ1. Three other constructions (~22 kbp) were derived from the intermediate cloning steps. They are based on 7 V segments and a DJ region synthesized. The HC Microlocus Classic (22 kbp) was obtained after cloning of the gene encoding the constant part γ1 in 3' of the V , D and J elements mentioned above; this construct encodes the heavy chain of IgG1 (γ1 HC). The HC Microlocus Light (21.5 kb) contains the gene encoding γ1 CH1$^-$, therefore, this construction encodes IgG1 HC without CH1 domain (γ1HC CH1$^-$). Such HC are expressed without being associated with light chain (Heavy Chain only antibodies). Finally, the HC Microlocus Light shRNA (22 kb) is based on the HC Microlocus Light to which was added a sequence encoding four shRNA (small hairpin RNA) to repress the murine IgM expression. This *locus* is designed to be injected into wild type mouse. The functionality of the four reduced human HC constructs was assessed in mouse cells. The 300-19 cell line was used because is a pro-B cell line able to rearrange endogenous or transgenic Ig genes and to express the rearranged genes as Ig proteins. 300-19 cells were transduced with the four reduced human HC constructs and the expression of the transgenic Ig genes was investigated in detail. Indeed, DJ and VDJ rearrangement, recombinatory diversity, junctional diversity, transcription, mRNA maturation, alternative splicing, Ig surface expression and secretion were assessed. The data demonstrated that these constructs undergo editing and processing in murine cells and give rise to a large diversity of human heavy chains. The HC Microlocus Light was injected in HC knock out mouse oocytes and the HC Microlocus Light shRNA (which is exactly identical to the Microlocus Light concerning the antibody gene part) in wild type mouse oocytes. Injection in wild type mice led to human γ1 HC expression, whereas injections in HC KO mice did not show any B cell population reconstitution and consequently no human HC γ1 expression. Transgenic human γ1 HC and endogenous mouse IgM are co-expressed at the surface of progenitor-, immature- and mature B cells.

Keywords: therapeutic antibodies, humanized transgenic mice, antibody *loci*, HC KO mice, human transgenes, 300-19 cells, VDJ rearrangement, recombinatory diversity, junctional diversity, immunoglobulin expression.