



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-thesesexercice-contact@univ-lorraine.fr](mailto:ddoc-thesesexercice-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

**UNIVERSITE DE LORRAINE**  
**2015**

---

**FACULTE DE PHARMACIE**

**MEMOIRE**  
**du DIPLOME D'ETUDES SPECIALISEES**  
**de BIOLOGIE MEDICALE**

Soutenu devant le Jury Interrégional  
Le 10 Septembre 2015

Par Alexandra JACQUES  
née le 18 Octobre 1987 à Blois (41)

Conformément aux dispositions de l'arrêté  
du 4 octobre 1988 tient lieu de

**THESE**  
**pour le DIPLOME D'ETAT**  
**de DOCTEUR en PHARMACIE**

---

**Intérêt du séquençage dans les typages HLA  
et la greffe de cellules souches hématopoïétiques**

---

Membres du Jury

Président : Mme le Professeur D. BENSOUSSAN, PU-PH, Faculté de Pharmacie de Nancy

Juges : Mme le Docteur A. AARNINK, PH-U, CHU de Nancy,  
Directrice de thèse  
M. le Professeur L. FRIMAT, PU-PH, CHU de Nancy  
Mme le Docteur C. BONNET, MCU-PH, CHU de Nancy  
Mme le Docteur V. MOALIC-ALLAIN, PH, CHU de Brest

**UNIVERSITÉ DE LORRAINE**  
**FACULTÉ DE PHARMACIE**  
**Année universitaire 2014-2015**

**DOYEN**

Francine PAULUS

**Vice-Doyen**

Béatrice FAIVRE

**Directeur des Etudes**

Virginie PICHON

**Conseil de la Pédagogie**

Président, Brigitte LEININGER-MULLER

**Collège d'Enseignement Pharmaceutique Hospitalier**

Président, Béatrice DEMORE

**Commission Prospective Facultaire**

Président, Christophe GANTZER

Vice-Président, Jean-Louis MERLIN

**Commission de la Recherche**

Président, Raphaël DUVAL

**Responsable de la filière Officine**

Béatrice FAIVRE

**Responsables de la filière Industrie**

Isabelle LARTAUD,

Jean-Bernard REGNOUF de VAINS

**Responsable de la filière Hôpital**

Béatrice DEMORE

**Responsable Pharma Plus ENSIC**

Jean-Bernard REGNOUF de VAINS

**Responsable Pharma Plus ENSAIA**

Raphaël DUVAL

**Responsable de la Communication**

Marie-Paule SAUDER

**Responsable de la Cellule de Formation Continue  
et individuelle**

Béatrice FAIVRE

**Responsable de la Commission d'agrément  
des maîtres de stage**

Béatrice FAIVRE

**Responsables des échanges  
internationaux**

Bertrand RIHN

**Responsable ERASMUS**

Mihayl VARBANOV

**DOYENS HONORAIRES**

Chantal FINANCE

Claude VIGNERON

**PROFESSEURS EMERITES**

Jeffrey ATKINSON

Max HENRY

Gérard SIEST

Claude VIGNERON

**PROFESSEURS HONORAIRES**

Roger BONALY

Pierre DIXNEUF

Marie-Madeleine GALTEAU

Thérèse GIRARD

Michel JACQUE

Pierre LABRUDE

Lucien LALLOZ

Pierre LECTARD

Vincent LOPPINET

Marcel MIRJOLET

Maurice PIERFITTE

Janine SCHWARTZBROD

**MAITRES DE CONFERENCES  
HONORAIRES**

Monique ALBERT

Mariette BEAUD

Gérald CATAU

Jean-Claude CHEVIN

Jocelyne COLLOMB

Bernard DANGIEN

Marie-Claude FUZELLIER

Françoise HINZELIN

Marie-Hélène LIVERTOUX

Bernard MIGNOT

Jean-Louis MONAL

Blandine MOREAU

Louis SCHWARTZBROD

**ASSISTANTS HONORAIRES**

Marie-Catherine BERTHE  
Annie PAVIS

Dominique NOTTER  
Christine PERDICAKIS  
Marie-France POCHON  
Anne ROVEL  
Maria WELLMAN-ROUSSEAU

<b>ENSEIGNANTS</b>	<b>Section CNU*</b>	<b>Discipline d'enseignement</b>
<b>PROFESSEURS DES UNIVERSITES - PRATICIENS HOSPITALIERS</b>		
Danièle BENSOUSSAN-LEJZEROWICZ	82	Thérapie cellulaire
Chantal FINANCE	82	Virologie, Immunologie
Jean-Louis MERLIN	82	Biologie cellulaire
Alain NICOLAS	80	Chimie analytique et Bromatologie
Jean-Michel SIMON	81	Economie de la santé, Législation pharmaceutique
<b>PROFESSEURS DES UNIVERSITES</b>		
Jean-Claude BLOCK	87	Santé publique
Christine CAPDEVILLE-ATKINSON	86	Pharmacologie
Raphaël DUVAL	87	Microbiologie clinique
Béatrice FAIVRE	87	Biologie cellulaire, Hématologie
Luc FERRARI	86	Toxicologie
Pascale FRIANT-MICHEL	85	Mathématiques, Physique
Christophe GANTZER	87	Microbiologie
Frédéric JORAND	87	Eau, Santé, Environnement
Isabelle LARTAUD	86	Pharmacologie
Dominique LAURAIN-MATTAR	86	Pharmacognosie
Brigitte LEININGER-MULLER	87	Biochimie
Pierre LEROY	85	Chimie physique
Philippe MAINCENT	85	Pharmacie galénique
Alain MARSURA	32	Chimie organique
Patrick MENU	86	Physiologie
Jean-Bernard REGNOUF de VAINS	86	Chimie thérapeutique
Bertrand RIHN	87	Biochimie, Biologie moléculaire
<b>MAITRES DE CONFÉRENCES DES UNIVERSITÉS - PRATICIENS HOSPITALIERS</b>		
Béatrice DEMORE	81	Pharmacie clinique
Julien PERRIN	82	Hématologie biologique
Marie SOCHA	81	Pharmacie clinique, thérapeutique et biotechnique
Nathalie THILLY	81	Santé publique
<b>MAITRES DE CONFÉRENCES</b>		
Sandrine BANAS	87	Parasitologie
Xavier BELLANGER	87	Parasitologie, Mycologie médicale
Emmanuelle BENOIT	86	Communication et Santé
Isabelle BERTRAND	87	Microbiologie
Michel BOISBRUN	86	Chimie thérapeutique
François BONNEAUX	86	Chimie thérapeutique
Ariane BOUDIER	85	Chimie Physique
Cédric BOURA	86	Physiologie
Igor CLAROT	85	Chimie analytique
Joël COULON	87	Biochimie
Sébastien DADE	85	Bio-informatique
Dominique DECOLIN	85	Chimie analytique
Roudayna DIAB	85	Pharmacie galénique
Natacha DREUMONT	87	Biochimie générale, Biochimie clinique
Joël DUCOURNEAU	85	Biophysique, Acoustique

<b>ENSEIGNANTS (suite)</b>	<b>Section CNU</b>	<b>Discipline d'enseignement</b>
Florence DUMARCAY	86	Chimie thérapeutique
François DUPUIS	86	Pharmacologie
Adil FAIZ	85	Biophysique, Acoustique
Anthony GANDIN	87	Mycologie, Botanique
Caroline GAUCHER	85/86	Chimie physique, Pharmacologie
Stéphane GIBAUD	86	Pharmacie clinique
Thierry HUMBERT	86	Chimie organique
Olivier JOUBERT	86	Toxicologie, Sécurité sanitaire
Francine KEDZIEREWICZ	85	Pharmacie galénique
Alexandrine LAMBERT	85	Informatique, Biostatistiques
Julie LEONHARD	86	Droit en Santé
Faten MERHI-SOUSSI	87	Hématologie
Christophe MERLIN	87	Microbiologie environnementale
Maxime MOURER	86	Chimie organique
Coumba NDIAYE	86	Epidémiologie et Santé publique
Francine PAULUS	85	Informatique
Caroline PERRIN-SARRADO	86	Pharmacologie
Virginie PICHON	85	Biophysique
Sophie PINEL	85	Informatique en Santé (e-santé)
Anne SAPIN-MINET	85	Pharmacie galénique
Marie-Paule SAUDER	87	Mycologie, Botanique
Rosella SPINA	86	Pharmacognosie
Gabriel TROCKLE	86	Pharmacologie
Mihayl VARBANOV	87	Immuno-Virologie
Marie-Noëlle VAULTIER	87	Mycologie, Botanique
Emilie VELOT	86	Physiologie-Physiopathologie humaines
Mohamed ZAIUO	87	Biochimie et Biologie moléculaire
Colette ZINUTTI	85	Pharmacie galénique

**PROFESSEUR ASSOCIE**

Anne MAHEUT-BOSSER 86 Sémiologie

**PROFESSEUR AGREGE**

Christophe COCHAUD 11 Anglais

\*Disciplines du Conseil National des Universités :

80 : Personnels enseignants et hospitaliers de pharmacie en sciences physico-chimiques et ingénierie appliquée à la santé

81 : Personnels enseignants et hospitaliers de pharmacie en sciences du médicament et des autres produits de santé

82 : Personnels enseignants et hospitaliers de pharmacie en sciences biologiques, fondamentales et cliniques

85 ; Personnels enseignants-chercheurs de pharmacie en sciences physico-chimiques et ingénierie appliquée à la santé

86 : Personnels enseignants-chercheurs de pharmacie en sciences du médicament et des autres produits de santé

87 : Personnels enseignants-chercheurs de pharmacie en sciences biologiques, fondamentales et cliniques

32 : Personnel enseignant-chercheur de sciences en chimie organique, minérale, industrielle

11 : Professeur agrégé de lettres et sciences humaines en langues et littératures anglaises et anglo-saxonnes

# SERMENT DES APOTHICAIRES



**Je jure, en présence des maîtres de la Faculté, des conseillers de l'ordre des pharmaciens et de mes condisciples :**

**D'** honorer ceux qui m'ont instruit dans les préceptes de mon art et de leur témoigner ma reconnaissance en restant fidèle à leur enseignement.

**D'**exercer, dans l'intérêt de la santé publique, ma profession avec conscience et de respecter non seulement la législation en vigueur, mais aussi les règles de l'honneur, de la probité et du désintéressement.

**D'**e ne jamais oublier ma responsabilité et mes devoirs envers le malade et sa dignité humaine ; en aucun cas, je ne consentirai à utiliser mes connaissances et mon état pour corrompre les mœurs et favoriser des actes criminels.

**Que les hommes m'accordent leur estime si je suis fidèle à mes promesses.**

**Que je sois couvert d'opprobre et méprisé de mes confrères si j'y manque.**



« LA FACULTE N'ENTEND DONNER AUCUNE  
APPROBATION, NI IMPROBATION AUX  
OPINIONS EMISES DANS LES THESES, CES  
OPINIONS DOIVENT ETRE CONSIDEREES  
COMME PROPRES A LEUR AUTEUR ».

# **REMERCIEMENTS**

**A Madame le Professeur Danièle BENSOUSSAN,**

PU-PH, Faculté de Pharmacie de Nancy, Unité de thérapie cellulaire et tissus-CHRU de Nancy

Vous me faites l'honneur d'accepter la présidence de cette thèse.

Veillez recevoir l'expression de mon respect et de ma profonde reconnaissance.

**A Monsieur le Professeur Luc FRIMAT,**

PU-PH, Faculté de Médecine de Nancy, Service de Néphrologie – CHRU de Nancy

C'est pour moi un grand honneur de vous compter parmi mes juges et je vous remercie sincèrement pour votre disponibilité. Soyez assuré de ma reconnaissance et de mon profond respect.

**A Madame le Docteur Alice AARNINK,**

PH-U, Faculté de Médecine de Nancy, Laboratoire d'Histocompatibilité – CHRU de Nancy

Je te remercie pour la confiance que tu m'as accordée en me dirigeant dans ce travail, pour ton soutien, pour ton énergie et ton enthousiasme quotidien. Un grand merci également pour tout le temps que tu m'as consacré, pour ta pédagogie et tes conseils qui m'ont beaucoup appris.

Sois assurée de ma sincère reconnaissance.

**A Madame le Docteur Virginie MOALIC-ALLAIN,**

PH, Laboratoire d'Histocompatibilité - CHU de Brest

Vous me faites l'honneur de juger cette thèse et d'apporter vos compétences à la critique de ce travail.

Je vous remercie d'avoir fait le déplacement jusqu'à Nancy pour être présente parmi nous.

Veillez trouver ici l'expression de mon profond respect.

**A Madame le Docteur Céline BONNET,**

MCU-PH, Faculté de Médecine de Nancy, Laboratoire de Génétique - CHRU de Nancy

Vous me faites l'honneur de juger cette thèse. Je vous remercie sincèrement de l'intérêt que vous avez bien voulu porter à mon travail. Veuillez recevoir l'expression de toute ma gratitude.

Je tiens également à remercier :

**Le Docteur Pascale PERRIER,**

Vous m'avez accueillie dans votre service pendant mon internat de biologie et médicale et m'avez fait partager vos connaissances, votre expérience et même votre passion pour l'histocompatibilité. Je vous remercie sincèrement pour la confiance que vous m'avez accordée. Veuillez trouver ici le témoignage de mon plus grand respect.

**Le docteur Anne KENNEL,**

Un grand merci pour ta gentillesse, ta bonne humeur (malgré les journées deux étoiles) et ta disponibilité qui m'a permis de découvrir et apprécier le HLA. C'est un plaisir de travailler à tes côtés. Sois assurée de ma sincère reconnaissance.

**Le docteur Odette AGULLES,**

Merci pour votre soutien pendant mes semestres d'internat et pour votre sympathie. Veuillez recevoir ici l'expression de toute ma gratitude.

**Toute l'équipe du laboratoire d'histocompatibilité,**

Merci pour votre accueil et votre sympathie. Merci pour avoir pris le temps de me former et pour m'avoir donné envie de poursuivre dans cette discipline. Merci en particulier à Maxime pour son aide dans la réalisation de ce travail.

**Merci à tous les professionnels médecins et pharmaciens** qui m'ont accueillie dans leur service et qui m'ont enseigné le savoir-faire et le savoir être du biologiste. A **tous les techniciens et techniciennes** qui ont contribué à ma formation tout au long de ces années d'internat.

**A Papa,**

Tu nous as quittés bien trop tôt. Ton amour, tes encouragements et ta fierté me manquent tous les jours. Tu es pour moi le plus beau modèle de courage et je te promets de faire au mieux pour le suivre. Je te dédie cette thèse.

**A Maman,**

Merci pour ton amour, ton soutien et pour tout le temps que tu m'as donnés, sans lesquels je n'en serais jamais arrivée là aujourd'hui. Sache que je t'en serai toujours reconnaissante.

**A JF,**

Mon frère et mon complice dans tous les moments de la vie, merci pour ton écoute et ton soutien.

**A FX,**

Merci pour ton amour, pour ta patience et ton soutien qui me sont si précieux.

**A l'ensemble de ma famille** que je vois trop peu, vivement les cousinades et les fêtes qui donnent l'occasion de se retrouver !

**A Claudine et Jean-Hugues,** merci pour vos encouragements et pour les bons moments partagés.

**A la famille Hofer : Philippe, Steph et Juju**, pour les super souvenirs de vacances qu'on a partagés.

Aux amis de la fac **Valérie et Pierre, Claire, Benoît, Marine, Henri**...en souvenir des années d'études, et des nombreuses soirées!!

Aux amis **Maeva, Aude et Clément**, qui sont partis bien trop loin de la Lorraine, vous me manquez!

A tous mes co-internes, qui ont rendu encore plus agréables ces années d'internat. Aux plus anciens **Angé, Eliette, Lulu et Lolo, Alain**...mais aussi aux p'tits jeunes!

**A Hélène**, pour les quelques 240 heures de trajets qu'on a partagées et au **chef Léon** qui a nous a supporté dans son bureau pendant 6 mois.

A la Mercy team, **Deckju, Sibi, Cadolu** pour les pauses thés et tartes meringuées de ce dernier semestre (et peut être un jour un pique-nique au château).

**A Delphine et Benoît** pour les parties de squash-(potins).

Aux joueurs et joueuses de baby foot qui se reconnaîtront (je préfère garder leur anonymat...)  
pour avoir partagé cette discipline des pauses midi...

Aux joueurs de billard aussi bien-sûr, même si je n'ai jamais été habilitée!

Enfin j'adresse mes remerciements les plus sincères à tous ceux qui ont de près ou de loin contribué à la réussite de mon parcours, à mon épanouissement personnel et qui me sont chers.

# TABLE DES MATIERES

LISTE DES ABREVIATIONS .....	3
LISTE DES FIGURES .....	6
INTRODUCTION.....	8
GENERALITES.....	9
I.    CMH / Système HLA.....	9
1.    Découverte du CMH.....	9
2.    Définitions .....	10
3.    Gènes et molécules du CMH chez l'Homme .....	13
4.    Fonctions des molécules HLA.....	24
5.    Nomenclature.....	26
6.    Système HLA et transplantation, notion d'allo-réactivité.....	29
II.    La greffe de moelle osseuse.....	31
1.    Quelques définitions :.....	31
2.    La réaction du greffon contre l'hôte (GvH).....	34
3.    Historique de la greffe de CSH.....	39
4.    Choix d'un donneur .....	40
III.   Intérêt du séquençage dans la greffe de moelle.....	44
1.    Typage HLA à l'heure actuelle : sérologie, PCR-SSP et PCR-SSO .....	44
2.    Historique du séquençage .....	48
3.    Séquençage de deuxième génération.....	58
4.    Comparaison entre les différentes technologies proposées .....	70
5.    NGS et informatique.....	74
6.    Séquençage de troisième génération.....	79
OBJECTIFS .....	83
METHODE ET RESULTATS .....	84
I.    Présentation des résultats d'EEQ .....	84
II.   Problèmes d'assignation d'allèle rare.....	85
1.    Cas n°1 : détection d'un allèle rare (4 <sup>ième</sup> digit).....	85
2.    Cas n°2 : détection d'un allèle rare (6 <sup>ième</sup> digit).....	88

III.	Discordance entre techniques de biologie moléculaire .....	91
1.	Cas n°3 : discordance entre PCR-SSP et PCR-SSO .....	91
2.	Cas n°4: discordance entre PCR-SSO, SSP et SSBT .....	93
IV.	Cas n°5 : Discordance sérologie-biologie moléculaire.....	96
	DISCUSSION .....	99
I.	Quelle approche : <i>shot gun</i> ou <i>exon-sequencing</i> ? .....	99
II.	Influence de la méthode de fragmentation dans l'approche <i>shot gun</i> .....	101
III.	Intérêt du séquençage de nouvelle génération dans les typages HLA.....	102
IV.	Limites de la NGS .....	104
V.	Intérêt du séquençage de l'ARN ? .....	106
VI.	Intérêt d'élargir le séquençage à d'autres familles de gènes .....	107
	CONCLUSION .....	109
	BIBLIOGRAPHIE .....	111
	ANNEXES .....	121
	RESUME : .....	1

# LISTE DES ABREVIATIONS

ADN : Acide DésoxyRibonucléique  
ADP : Adénosine DiPhosphate  
APS : Adénosine PhosphoSulfate  
ARN : Acide RiboNucléique  
ASCII : *American Standard Code for Information Interchange*  
ATP : Adénosine TriPhosphate  
CCD : *Charge Coupled Device*  
CMH : Complexe Majeur d'Histocompatibilité  
CMV : *Cytomegalo virus*  
CPA : Cellule Présentatrice d'Antigène  
CSH : Cellule Souche Hématopoïétique  
CSP : Cellule Souche Périphérique  
CWD : *Commun and Well Defined*  
DLI : *Donnor Lymphocyte Infusion*  
DVMO : Donneur Volontaire de Moelle Osseuse  
EBV : *Epstein barr virus*  
EFI : *European Federation for Immunogenetics*  
dNTP : DésoxyNucléotide TriPhosphate  
FRET: *Fluorescence Resonance Energy Transfer*  
G-CSF : *Granulocyte-Colony Stimulating Factor*  
GSA : *Group Specific Amplification*  
GvH : *Graft versus Host*  
GvI : *Graft versus Infection*  
GvL : *Graft versus Leukemia*  
HA : *Human Antigen*  
HVG : *Host versus Graft*  
HLA : *Human Leucocyte Antigen*

HSV : *Herpes simplex virus*  
IFN : Interféron  
IL : Interleukine  
ISP : *Ion Sphere Particule*  
kDa : kiloDalton  
KIR : *Killer cell Ig-like Receptor*  
LT : Lymphocyte T  
(M)QV : *(Mapping) Quality Value*  
Mb : Mégabases  
MID : *Multiplex Identifier*  
NGS : *New Generation Sequencing*  
NK: *Natural killer*  
NMDP: *National Marrow Donor Program*  
OLP, OLS : Organes Lymphoïdes Primaires, Secondaires  
Pb : Paire de bases  
PCR : *Polymerase Chain Reaction*  
PTP : *Pico Titer Plate*  
RDT : *Reverse Dye Terminator*  
RIC : *Reduced Intensity Conditionnement*  
SSBT : *Sanger Sequence Based Typing*  
SMRT : *Single Molecule Real Time sequencing*  
SSO : *Sequence Specific Oligonucleotide probes*  
SSP : *Single Specific Primer*  
TCR : *T Cell Receptor*  
TNF : *Tumor Necrosis Factor*  
USP : Unité de Sang Placentaire  
VZV : *Varicel Zoster virus*  
ZMW: *Zero-Mode Waweguide*



# LISTE DES FIGURES

Figure 1 : Schéma de la carte génétique du système HLA, localisé sur le bras court du chromosome 6 .....	11
Figure 2 : Représentation schématique des gènes de classe Ia du CMH, des ARNm transcrits et des protéines obtenues. ....	15
Figure 3 : Schéma d'apprêtement des protéines endogènes par les molécules HLA de classe I et présentation croisée de protéines exogènes. ....	19
Figure 4 : Représentation schématique des gènes de classe II du CMH, de l'ARNm transcrit et des protéines obtenues. ....	21
Figure 5 : Schéma d'apprêtement des protéines exogènes par une molécule HLA de classe II.....	23
Figure 6 : Nomenclature des allèles HLA. ....	27
Figure 7 : Mécanismes d'allo-réactivité directe et indirecte. ....	30
Figure 8: Schéma du processus physiopathologique de la GvH.....	37
Figure 9 : Graphique représentant la probabilité d'avoir une GvH dans un certain délai selon la compatibilité choisie. ....	41
Figure 10 : Présentation des principes de la PCR-SSP et de la PCR-SSO.....	47
Figure 11 : Structure des dNTP et ddNTP et présentation d'un autoradiogramme. ....	52
Figure 12 : Principe de séquençage selon la méthode Sanger.....	54
Figure 13 : Exemple de polymorphisme cis/trans. ....	56
Figure 14 : Sélection des fragments possédant un adaptateur différent à chaque extrémité, par l'utilisation de billes magnétiques et de l'affinité biotine-streptavidine. ....	60
Figure 15 : Schéma représentatif d'une séquence cible entourée de ses adaptateurs et de code-barres avec l'approche <i>shot gun</i> . ....	61
Figure 16 : Schéma du principe de PCR en émulsion, exemple des billes fixant les adaptateurs B. ....	62
Figure 17 : Principe de la <i>Bridge PCR</i> (PCR en pont). ....	64
Figure 18 : Principe du pyroséquençage.....	66
Figure 19 : Principe de séquençage utilisant la technologie des semi-conducteurs. ....	67
Figure 20 : Principe de séquençage par la technologie Illumina.....	68
Figure 21 : Principe du <i>paired end read</i> .....	69
Figure 22 : Récapitulatif de la comparaison des différentes technologies disponibles. ....	72
Figure 23 : Représentation de la profondeur de lecture.....	72
Figure 24 : Exemple de données par le logiciel Ion Torrent.....	75
Figure 25 : Représentation d'une séquence nucléotidique en format FASTQ. ....	77
Figure 26 : Principe du séquençage de 3 <sup>ième</sup> génération selon la technologie des nanopores.....	80
Figure 27 : Principe du séquençage de 3 <sup>ième</sup> génération selon la technologie des cellules SMRT.....	82
Figure 28 : Validation sur le logiciel HLA Fusion <sup>®</sup> , présentation des informations disponibles pour la validation d'un résultat obtenu par PCR-SSO. ....	86
Figure 29 : Validation des résultats de séquençage SSBT sur le logiciel JSI <sup>®</sup> . ....	87
Figure 30 : Alignement des séquences protéiques et nucléiques des allèles DRB1*13:01:01 et DRB1*13:15 selon le site de l'IMGT <i>database</i> . ....	87
Figure 31 : Profils des billes 39 et 98, consultables sur le logiciel HLA Fusion <sup>®</sup> .....	89
Figure 32 : Alignement des séquences protéiques et nucléiques des allèles A*02: 01:01 et A*02:01:09 selon le site de l'IMGT <i>database</i> . Comparaison au résultat trouvé sur le logiciel JSI <sup>®</sup> par séquençage monoallélique. ....	90
Figure 33 : Photographie du gel d'agarose obtenu en PCR-SSP (Olerup <sup>®</sup> ) pour le typage du locus C, spécificité C*15. ....	91
Figure 34 : Analyse du profil de la bille 94 sur le logiciel HLA Fusion <sup>®</sup> . ....	92

Figure 35 : Alignement des séquences protéiques et nucléiques des allèles C*15: 02:01 et C*15:34 selon le site de l'IMGT <i>database</i> . .....	93
Figure 36 : Photographie du gel d'agarose obtenu en PCR-SSP (Olerup®), partie concernant le typage en faible résolution du locus B.....	94
Figure 37 : Résultat des billes informatives pour l'allèle B*35 :01 :01 (c'est à dire les billes attendues positives pour l'allèle HLA-B*35 :01 :01 et négatives pour l'allèle B*44 :02 :01 :01). .....	95
Figure 38 : Emplacement approximatif et schématique des différentes amorces utilisées en PCR-SSO et en SSBT. ....	95
Figure 39 : Photographie du gel d'agarose obtenu en PCR-SSP (Olerup®).....	97
Figure 40 : <i>Reads</i> issus d'artefacts de PCR, recombinaisons <i>in vitro</i> ou <i>cross-over</i> (exemple du gène <i>DRB1</i> et du gène <i>DRB3</i> ).....	105

# INTRODUCTION

Le complexe majeur d'histocompatibilité (CMH) suscite l'intérêt des recherches par son impressionnante densité en informations génétiques et ses multiples fonctions. Son extrême polymorphisme entretient la curiosité et permet encore la découverte régulière de nouveaux allèles. Seuls quelques principaux gènes de l'histocompatibilité sont aujourd'hui bien étudiés.

Les premières techniques d'exploration basées sur le principe de sérologie étaient limitées à la détection de l'expression antigénique en surface cellulaire. Elles ont connu leur apogée dans les années 1970. Elles se sont ensuite faites dépasser par l'arrivée de la biologie moléculaire qui connut rapidement un grand essor, avec en particulier l'automatisation du séquençage Sanger en 1990. Ce dernier permit alors la découverte de nombreux gènes, ce qui lui conféra un immense succès. La biologie moléculaire occupe aujourd'hui une place majoritaire dans les méthodes de typage utilisées en laboratoire à des fins aussi bien cliniques que de recherche. Ces méthodes tendent toutes à être de plus en plus automatisées, et adaptables aux besoins en termes de débit et de coût.

Parallèlement, les avancées en médecine pour le traitement des hémopathies sont marquées par des essais d'allogreffes de moelle osseuse qui apportent un grand espoir thérapeutique. Ces essais ont d'abord été restreints aux jumeaux monozygotes, vers 1960, puis se sont étendus à la fratrie pour aujourd'hui exister entre deux sujets non apparentés mais qui partagent des compatibilités antigéniques très strictes. Si pour autant les greffes sont le seul moyen d'obtenir une guérison totale, elles sont aussi une cause de mortalité par leur effets secondaires dramatiques. La réaction du greffon contre l'hôte est ainsi devenue la principale crainte et donc la cible des voies d'améliorations recherchées pour arriver à un rapport bénéfice/risque acceptable pour les patients.

Cette thèse a pour objectif de démontrer l'intérêt de l'apport de nouvelles technologies de séquençage dans une discipline aussi complexe que l'histocompatibilité et en particulier dans le contexte de la greffe de cellules souches hématopoïétiques. Les conditions de compatibilités requises sont de plus en plus définies et strictes. Par conséquent les résultats de typages sont soumis à des exigences rigoureuses, auxquelles seules des techniques comme le séquençage répondent aujourd'hui. Nous décrirons dans un premier temps le CMH et ses fonctions, qui expliquent son rôle important dans la greffe de cellules souches hématopoïétiques. Nous verrons ensuite les modalités de ce type de greffe, les conditions de sélection d'un donneur et le mécanisme de la réaction du greffon contre l'hôte. Enfin nous aborderons les techniques de typages utilisées en routine et les techniques de séquençage pour comprendre l'intérêt de ces nouvelles technologies dans l'amélioration de la prise en charge du patient.

# GENERALITES

## I. CMH / Système HLA

### 1. Découverte du CMH

L'existence du complexe majeur d'histocompatibilité (CMH) a d'abord été évoquée chez la souris, par Peter Görer, en 1936. Des réactions d'agglutinations d'érythrocytes de souris par des sérums de lapin préalablement immunisés par des érythrocytes murins ont permis de suggérer l'existence d'allo-antigènes en surface cellulaire. Snell observa par la suite que le rejet du greffon chez la souris était associé à une incompatibilité au niveau des allo-antigènes précédemment définis par Görer. C'est ainsi qu'il décrit pour la première fois le Complexe Majeur d'Histocompatibilité murin, nommé «H2» en l'honneur de l'antigène II découvert par Görer [1].

C'est donc sa capacité d'induire une forte réponse immunitaire allogénique, responsable du rejet de greffe, qui le fait connaître dans un premier temps et qui lui fait mériter son nom. Son rôle majeur dans la réponse immunitaire et dans l'éducation des lymphocytes T par la présentation des antigènes, n'a été découvert que 30 ans après.

Le Complexe Majeur d'Histocompatibilité de l'homme fut découvert en 1952 par Jean Dausset, récompensé du prix Nobel de médecine en 1980. Son nom de système HLA (*Human Leukocyte Antigen*) symbolise le groupe sanguin leucocytaire mis en évidence par l'agglutination de leucocytes par des sérums de sujets immunisés à l'occasion de transfusions sanguines. En effet, après ses études de médecine à Paris, Jean Dausset porta un grand intérêt aux patients polytransfusés et leucopéniques, chez qui il découvrit l'existence d'anticorps anti-leucocytes, d'abord soupçonnés d'auto-anticorps. Il démontra rapidement qu'il s'agissait en fait d'une allo-immunisation générée par les nombreuses transfusions. Ce n'est pourtant que six ans plus tard, en 1958, que le premier antigène du CMH humain MAC (HLA-A2) est décrit. A cette époque, des expériences ont été réalisées à partir du sérum d'un patient qui avait reçu des transfusions d'un unique donneur. Elles permirent de constater que les anticorps développés par ce patient réagissaient avec les leucocytes de la moitié d'une population de donneurs volontaires. Les noms des trois premiers donneurs avec qui le sérum n'a montré aucune réaction sont à l'origine de l'acronyme MAC [2].

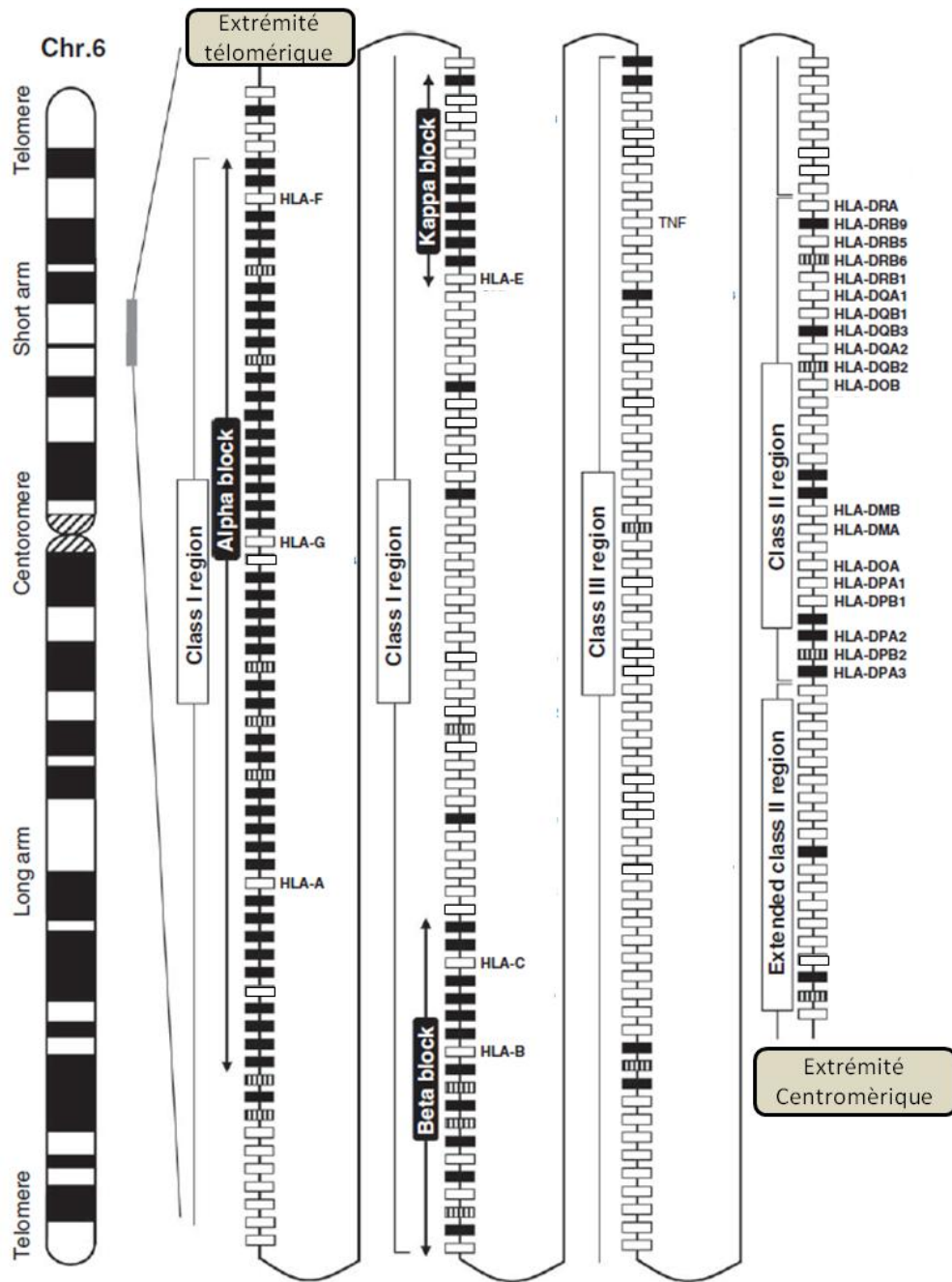
La technique sérologique de lymphocytotoxicité complément-dépendante va ensuite être développée et permettra la découverte de nouveaux antigènes qui révèlent alors progressivement le grand polymorphisme du système HLA. La classe II fut caractérisée plus tardivement, par des réactions lymphocytaires mixtes.

En 1968, Jean Dausset évoque la possibilité d'une relation étroite entre les groupages HLA et la susceptibilité génétique à développer certaines maladies. Le concept de médecine prédictive entraîne alors la nécessité d'établir la carte génétique, en particulier du système HLA, localisé quelques années plus tôt sur le bras court du chromosome 6. Des centaines de laboratoires dans le monde entier sont enthousiasmés par ce projet ce qui permit des progrès rapides en biologie moléculaire et une description de plus en plus précise du CMH, au niveau génétique [2].

## 2. Définitions

Chez l'homme, le complexe majeur d'histocompatibilité représente un segment génomique situé sur le bras court du chromosome 6, occupant environ quatre mégabases. Sa carte génétique est relativement bien détaillée aujourd'hui (**Figure 1**). La première séquence complète, publiée en 1999, est le fruit d'un long travail de collaboration entre quatre principales équipes (*Cancer Research Center, University of Washington, Japanese Science and Technology Corporation, The Sanger Centre*). La carte génétique fut établie à partir des séquences de différents haplotypes et dévoile aujourd'hui 224 gènes. Seulement 60% d'entre eux seraient exprimés. Le CMH est caractérisé par une densité génique élevée, avec en moyenne 35 gènes par mégabase. En comparaison, la densité moyenne dans le génome est de 12 gènes par mégabase. L'étude du CMH de centaines d'individus différents a fait découvrir son extrême polymorphisme, il n'est cependant pas homogène sur l'ensemble de la région (voir paragraphe I.3) [3]. Le terme de **polymorphisme** désigne les différences entre deux êtres d'une même espèce, chez qui un phénotype, ou caractère, est présent sous au moins deux formes (allèles). Ce polymorphisme est dû à l'apparition de mutations génétiques au cours de l'évolution, et peut être expliqué par l'existence de nombreux allèles qui définissent le caractère unique de chaque individu.

Chaque individu possède au maximum deux allèles par locus, un de son père et un de sa mère, qui peuvent donc être différents. Si le même allèle a été transmis par les deux parents l'individu est dit **homozygote**, s'il a reçu deux allèles différents il est dit **hétérozygote**. L'expression des gènes HLA est de caractère **codominant**. Prenons l'exemple d'un individu qui possède au locus HLA-A un allèle HLA-A\*02:01 transmis par son père et un allèle HLA-A\*03:01 transmis par sa mère. Cet individu exprimera à la surface cellulaire les deux molécules HLA correspondant à ces deux allèles. Cet individu est hétérozygote au locus HLA-A.



**Figure 1 : Schéma de la carte génétique du système HLA, localisé sur le bras court du chromosome 6**

D'après "The sequencing MHC consortium" [4].

Elle montre les différentes régions de classe I, classe III et classe II depuis l'extrémité télomérique (en haut) jusqu'au centromère (en bas). Les gènes représentés par des rectangles de couleur blanche, rayée ou noire correspondent respectivement aux gènes exprimés, aux gènes non codant et aux pseudogènes.

Cependant tous les gènes du CMH ne sont pas transcrits avec la même intensité dans tous les types cellulaires. Nous verrons qu'il existe des différences d'expression permettant de parler de gènes classiques, non classiques ou de pseudogènes. Dans ce dernier groupe sont réunis des gènes entiers dont l'expression ne permet pas d'obtenir une molécule fonctionnelle par l'existence d'un codon stop prématuré, mais aussi des fragments de gènes pour lesquels un ou plusieurs exons peuvent manquer et qui sont les vestiges d'un gène ancestral [3][5] [6].

Les gènes du CMH, étant très proches les uns des autres sur le chromosome, sont dans la majorité des cas transmis en «bloc» d'un parent à l'enfant. L'ensemble des gènes du CMH portés par un chromosome forment donc un **haplotype**. Ainsi dans une famille composée d'un père porteur des haplotypes a et b et d'une mère possédant les haplotypes c et d, donneront quatre combinaisons d'haplotypes possibles chez leurs enfants : ac, ad, bc et bd. Cette notion permet de comprendre pourquoi les patients nécessitant une greffe de moelle n'ont pas tous la possibilité de bénéficier d'une greffe intra-familiale, puisqu'il faut que deux enfants possèdent la même combinaison. Cependant, il existe des exceptions à la règle de ségrégation simultanée de deux allèles liés, qui s'explique par des recombinaisons (ou crossing-over) lors de la méiose. Ce point ne sera pas détaillé dans cette thèse.

Ceci conduit à la notion de **déséquilibre de liaison** entre deux allèles proches sur le même chromosome. Il s'agit d'une association préférentielle entre deux ou plusieurs allèles, c'est-à-dire une association rencontrée avec une fréquence plus élevée que la probabilité d'association si ces deux allèles étaient complètement indépendants l'un de l'autre [7]. Par exemple, les allèles HLA-B\*15:01 et HLA-C\*03:03 sont retrouvés sur un même haplotype à une fréquence de 0,035 chez des individus européens. Pourtant la probabilité théorique de rencontrer ces deux allèles sur le même gène est 0,0037 (c'est-à-dire le produit des fréquences isolées de l'allèle HLA-B\*15:01 et HLA-C\*03:03) soit dix fois moins. La prise en compte des déséquilibres de liaison est très utile dans l'aide à la recherche de donneurs non apparentés de moelle osseuse (voir paragraphe «Choix d'un donneur») [8].

L'établissement des haplotypes d'un individu, ou **génotypage**, consiste à déduire d'une étude familiale les allèles présents sur les deux chromosomes homologues [9]. Il est à différencier du **phénotypage**, qui correspond aux situations de typages, par des techniques de sérologie ou de biologie moléculaire, permettant de déterminer les allèles portés par un individu mais sans définir ses haplotypes (typages familiaux non disponibles). Cependant, dans le cadre du HLA, la connaissance des fréquences des haplotypes selon l'origine ethnique peut donner une forte idée de la répartition des allèles sur les deux chromosomes.

Dans le cadre de la transplantation, le génotypage ou le phénotypage du système HLA est restreint à la détermination des allèles de quelques gènes (HLA-A, HLA-B, HLA-C, HLA-DRB1, DRB3, DRB4, DRB5, HLA-DQB1, HLA-DQA1, HLA-DPB1 et HLA-DPA1) dont nous allons faire la description ci-dessous.

### 3. Gènes et molécules du CMH chez l'Homme

Les gènes du CMH sont divisés en trois grandes classes selon leurs caractéristiques structurelles et fonctionnelles :

- ✓ Les gènes de classe I occupent la région télomérique du CMH sur environ deux mégabases (2Mb). On compte 6 gènes transcrits codant pour des molécules HLA (A, B, C, E, G, F).
- ✓ Les gènes de classe II se retrouvent dans la région centromérique du CMH occupant neuf cent kilobases (0.9 Mb), ils comprennent plus de 32 gènes dont au moins 10 gènes HLA transcrits (DRA, DRB1, DRB3, DRB4, DRB5, DPA, DPB1, DQA, DQB1, DMA, DMB, DOA, DOB). (Les gènes DRB3, DRB4 et DRB5 ne sont pas toujours transcrits, cela dépend des gènes DRB1 ; ce point ne sera pas détaillé dans cette thèse).
- ✓ Les gènes de classe III, situés entre les deux classes précédentes, s'étendent sur une zone très dense de plus d'une mégabase (1.1 Mb). Elle est occupée par plus de 39 gènes qui codent principalement pour des molécules ayant un rôle dans les réactions immunitaires, comme les protéines intervenant dans la voie d'activation du complément (C2, C4, Bf), et le *Tumor Necrosis Factor* (TNF $\alpha$  $\beta$ ) [4] [10]. Toutefois, de nombreux gènes n'ayant pas de fonction dans le système immunitaire se trouveraient également dans cette région, comme par exemple les protéines du choc thermique (HSP) ou encore la 21-hydroxylase [10]. Leur polymorphisme est très limité voire inexistant, en contraste avec les gènes de classe I et II.

Les régions de classe I et II du CMH comportent également des gènes qui ne codent pas pour des molécules HLA (comme MICA, MICB, TAP1...) mais aussi des gènes HLA non codants, ou pseudogènes [4] que nous ne détaillerons pas dans cette thèse .

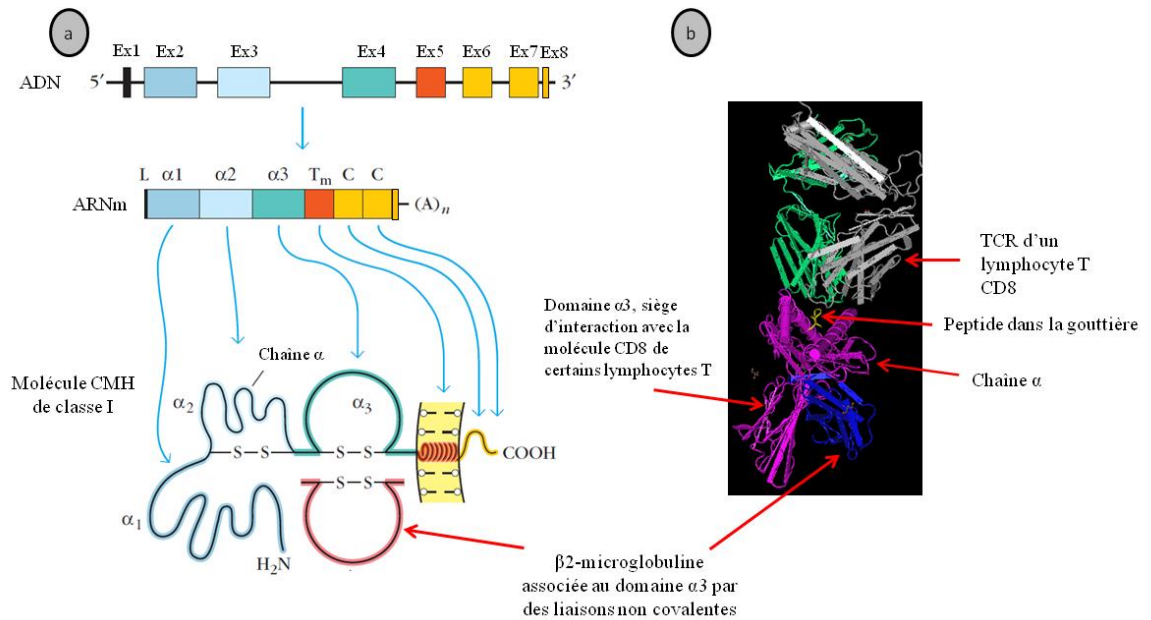
## Structure des gènes et molécules de classe I

Les molécules HLA de classe I sont constituées d'une glycoprotéine transmembranaire, la chaîne lourde alpha ( $\alpha$ ), associée à une chaîne légère non glycosylée, la chaîne  $\beta$ 2-microglobuline, par des liaisons non covalentes (**Figure 2**). On distingue deux sous-classes de molécules de classe I :

- ✓ Les molécules de classe I classiques (Ia) codées par les gènes HLA-A, -B et -C et qui sont d'intérêt majeur en transplantation. Ces gènes sont hautement polymorphiques, et leur expression est pratiquement ubiquitaire (voir paragraphe ci-dessous « expression des molécules de classe I ») [5].
- ✓ Les molécules de classe I non classiques (Ib) codées par les gènes HLA-E, -F, -G situés dans la région de classe I du CMH et le gène HLA-H (aussi connu sous le nom HFE) situé à proximité du CMH [11]. Les trois gènes HLA-E, -F, -G ont un polymorphisme et une expression cellulaire bien plus limités que ceux de classe Ia [10] [12].

Des analyses phylogénétiques laissent penser que la classe Ib proviendrait de la duplication des loci de classe Ia, mais certaines séquences, en particulier celles de régulation transcriptionnelle en 5' ne seraient pas conservées [13]. Cette notion permet de comprendre les risques d'amplification non spécifique et la difficulté de synthétiser des amorces de qualité en biologie moléculaire.

Seule la chaîne lourde  $\alpha$ , d'environ 45 kilodaltons (kDa), est codée par des gènes du CMH et présente un grand polymorphisme. A l'inverse, la chaîne légère de  $\beta$ 2-microglobuline est une chaîne invariante de douze kilodaltons (kDa) codée par un gène situé sur le chromosome 15 [14]. La chaîne  $\alpha$  est composée d'une partie extracellulaire comprenant trois domaines  $\alpha$ 1,  $\alpha$ 2, et  $\alpha$ 3. Les deux premiers constituent la zone la plus polymorphe de la molécule. Leur structure tridimensionnelle, avec deux hélices  $\alpha$  encadrant un plancher de feuillet  $\beta$ , permet de former un sillon destiné à l'accueil des peptides à présenter. On parle aussi de gouttière à peptide ou «*peptide binding groove*». Le peptide présenté, composé de 8 à 10 acides aminés, sera lié par seulement deux résidus à la molécule de classe I (**Figure 2**).



**Figure 2 : Représentation schématique des gènes de classe Ia du CMH, des ARNm transcrits et des protéines obtenues.**

D'après R.A. Goldsby, T.J. Kindt et B.A. Osborne [15].

- Représentation schématique d'une molécule d'ADN correspondant à un gène HLA-A (A noter que le gène HLA-B et les gènes de la classe Ib ne comportent que 7 exons), de son transcrit et de la molécule obtenue.
- Structure tridimensionnelle de la molécule HLA de classe Ia, et interaction avec le récepteur (*T cell Receptor*, TCR) d'un lymphocyte T CD8. (PDB ID : 3H9S)

## Expression des molécules de classe I

Les molécules de classe I classiques sont retrouvées à la surface de pratiquement toutes les cellules nucléées, cependant leur niveau d'expression est variable. Ainsi les lymphocytes (B plus que T) et les cellules dendritiques sont les cellules qui expriment le plus les molécules de classe I classiques. Inversement, les cellules musculaires squelettiques et cardiaques, et les cellules endocrines de certains organes (thyroïde, pancréas, muqueuse gastrique par exemple) expriment très faiblement les molécules de classe I classiques [16]. Néanmoins quelques cellules nucléées en semblent exemptes comme par exemple les cellules germinales, l'endothélium cornéen et certaines cellules du système nerveux central [10].

Les mécanismes de régulation qui contrôlent cette expression dans les différents tissus et cellules ne sont pas encore tous élucidés. Ils mettent en jeu diverses cytokines comme les interférons (IFN  $\alpha$ ,  $\beta$  et  $\gamma$ ) et le facteur nécrosant des tumeurs (TNF) qui augmentent l'expression des molécules de classe I. Ils agissent par l'intermédiaire de promoteurs situés en amont des gènes de classe I [17].

L'expression des molécules de classe I classiques à la surface des cellules peut aussi être modulée négativement lorsqu'elles sont infectées par certains virus, comme par exemple le cytomégalovirus (CMV) [18].

Il n'existe pas de généralités concernant l'expression des molécules de classe Ib. Commençons par la molécule **HLA-E**, découverte en 1988. Elle possède un polymorphisme très faible, et une répartition quasi ubiquitaire sur les cellules somatiques, mais est exprimée en bien moindre quantité que les molécules de classe Ia. Bien que neuf allèles différents soient connus, deux seulement semblent être récurrents [12]. Cette molécule, qui reste énigmatique, est décrite pour son rôle de régulation des cellules *Natural Killer* (NK), aussi bien inhibiteur qu'activateur, en particulier dans des situations de stress ou d'infection [11]. Les molécules HLA-E présentent des peptides d'origines variées qui peuvent être d'origine microbienne ou bien provenir de molécules du soi. En effet, elles présentent les peptides *leader* (codé par l'exon 1) des molécules HLA de classe Ia, et reflètent ainsi le niveau de transcription de ces dernières. Cette action est destinée à renseigner les cellules NK qui surveillent la qualité d'expression des molécules HLA de classe I classique [11]. Récemment, la molécule HLA-E fait l'objet de diverses études sur l'influence de son polymorphisme dans la survenue de la réaction du greffon contre l'hôte (GvH, «*Graft versus host disease*») après une greffe de cellules souches hématopoïétiques (CSH) [12].[19]. Certaines études ont conclu que le phénotype HLA-E\*01:03/E\*01:03 du receveur serait associé à un moindre risque de développer une GvH après une greffe de moelle osseuse et globalement à une survie plus longue. Il a aussi été retenu que l'homozygotie du donneur pour HLA-E\*01:01 (comparé aux phénotypes HLA-E\*01:01/E\*01:03 et HLA-E\*01:03/E\*01:03) serait favorable au développement d'infections bactériennes en lien direct avec une mortalité accrue après la greffe. Une capacité de présentation des peptides moins efficace par les cellules du donneur serait à l'origine de cette tendance aux infections [12].

Il est aussi intéressant de s'arrêter sur la molécule **HLA-G**, découverte en 1990. Son polymorphisme est comparable à celui du gène HLA-C. Longtemps considérée comme en «voie d'extinction», HLA-G a finalement suscité de l'intérêt lors de la découverte de ses fonctions en particulier pendant la gestation [16]. Les cellules trophoblastiques n'exprimeraient d'ailleurs que des molécules de classe Ib. Dans des conditions physiologiques, on ne retrouve les molécules HLA-G que dans les cellules trophoblastiques, les cellules épithéliales du thymus et la cornée. Son expression transcriptionnelle est très largement représentée dans les tissus adultes et embryonnaires, en revanche la protéine n'est retrouvée qu'en surface d'un nombre restreint de cellules, notamment au niveau placentaire. En effet, le gène HLA-G est soumis à une importante régulation post-transcriptionnelle qui donne lieu à sept isoformes différentes, quatre membranaires et trois solubles [12]. Une hypothèse émise par l'équipe de Geraghty suggère que ces isoformes solubles pourraient déclencher l'apoptose des cellules T alloréactives maternelles, qui attaqueraient les peptides d'origine paternelle lors de la grossesse [16]. En effet, la grossesse peut être comparée à une greffe semi-allogénique, puisque le fœtus exprime les antigènes maternels mais aussi paternels, ces derniers étant étrangers pour la mère. On comprend que le phénomène de tolérance immunitaire a toute son importance pour le bon déroulement des échanges foeto-maternels, et le maintien en vie du fœtus qui doit échapper aux cellules Natural Killer de sa mère. Le rôle de la molécule HLA-G dans la greffe de tissus a également été démontré avec une corrélation positive entre son expression et la diminution du risque de rejet [12].

Nous ne détaillerons pas la molécule **HLA-F**, qui demeure encore très peu étudiée. Ces molécules seraient transcrites dans la plupart des cellules et exprimées au niveau intracellulaire. Mais récemment, l'équipe de Geraghty montre, grâce à de nouveaux anticorps monoclonaux anti-HLA-F, que contrairement à la plupart des lignées cellulaires, toutes les populations de lymphocytes (B, T et NK), les monocytes et les cellules du trophoblaste extravilleux expriment HLA-F à leur surface [20].

Quant à la molécule **HLA-H**, autrement appelée HFE, elle semble être une exception à l'implication des molécules HLA dans le système immunitaire. Son rôle est plutôt important dans le métabolisme du fer [12] [11].

Au total, les molécules de classe Ib, et plus précisément HLA-E et HLA-G, pourraient jouer un rôle important dans la réaction d'immunotolérance en transplantation d'organe et de CSH [21]. Leurs rôles peu connus continuent d'attirer la curiosité, et les nouvelles technologies de séquençage promettent de faciliter des découvertes à leurs sujets.

Il est important de noter que la surface des globules rouges ne possède aucune molécule HLA chez l'Homme, contrairement à d'autres espèces, notamment la souris, ce qui explique la découverte de Görer. Les thrombocytes expriment les molécules HLA de classe I et présentent de ce fait des risques d'allo-immunisation anti-HLA de classe I lors des transfusions plaquettaires. Actuellement, les précautions transfusionnelles consistent à utiliser des culots globulaires déleucocytés lors de transfusions érythrocytaires pour limiter les risques d'allo-immunisation anti-HLA.

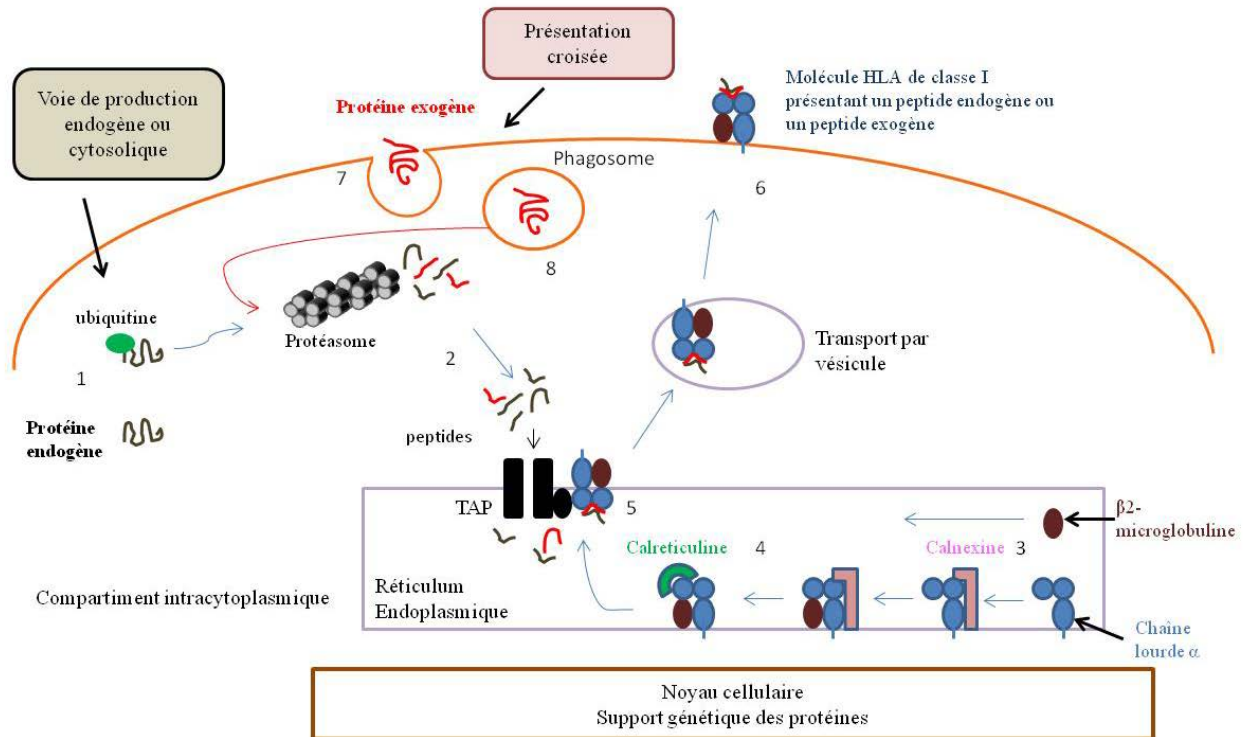
## Nature des peptides présentés

La détection des protéines par les lymphocytes T ne peut avoir lieu qu'après leur transformation en peptide, formant un complexe ternaire avec la molécule HLA et le récepteur du lymphocyte T (TCR, *T cell receptor*) (**Figure 2**). La cavité formée par les domaines  $\alpha 1$  et  $\alpha 2$ , décrite un peu plus tôt, permet de loger un peptide de 8 à 10 acides aminés issu de la dégradation intracytoplasmique de protéines endogènes. Dans le cas des molécules de classe I, cette loge est fermée de façon à retenir les extrémités libres du peptide, que constituent les groupements amine NH<sub>2</sub> et carboxylique COOH. Nous verrons cependant que cette règle n'est pas exclusive. Il existe en effet des protéines d'origine exogène présentées par les molécules de classe I [14].

En l'absence d'infection par un pathogène, les sillons des molécules de classe I de la cellule sont occupés par des peptides dérivés de protéines du soi (*self*) [22]. En revanche, lorsque la cellule héberge un pathogène, les peptides présentés en surface dérivent de l'agent infectieux (protéines du non soi), c'est ce qui va permettre de déclencher une réponse du système immunitaire entraînant ainsi l'élimination des cellules infectées.

Avant d'être à l'état de peptides « apprêtés », les protéines doivent entrer dans la voie de production endogène ou cytosolique des peptides, dont les principaux acteurs sont les protéasomes, des complexes de protéinases à haute activité catalytique. La chaîne lourde  $\alpha$  et la chaîne légère  $\beta 2$ -microglobuline, après leur traduction, sont toutes les deux équipées d'une séquence signal guidant leur entrée dans le réticulum endoplasmique (RE). Tout un processus est mis en jeu pour faciliter cet assemblage, que nous ne détaillerons pas davantage ici. Le schéma de la **figure 3** représente la voie d'assemblage des molécules de classe I et l'apprêtement du peptide [22].

Dans les cellules présentatrices d'antigènes (CPAs) dites professionnelles, comme les cellules dendritiques myéloïdes, les peptides présentés par les molécules de classe I peuvent être d'origine exogène, on parle alors de **présentation croisée**. En d'autres termes, elle est mise en jeu lorsque des molécules HLA de classe I présentent des protéines qui ne sont pas synthétisées par la cellule elle-même. Ce phénomène est également schématisé dans la figure 3 [23].



**Figure 3 : Schéma d'apprêtement des protéines endogènes par les molécules HLA de classe I et présentation croisée de protéines exogènes.**

#### Voie endogène

- 1 : Protéine d'origine endogène, du « soi » ou non soi (virale...) dirigée vers le protéasome par ubiquitinylation post-transcriptionnelle.
- 2 : Dégradation en peptides par le protéasome et transport du peptide, du cytosol à la lumière du RE par une protéine de transport ATP (adénosine triphosphate)-dépendante : TAP.
- 3 : Parallèlement, dans le RE, intervention de la calnexine pour permettre l'association des chaînes α et β2- microglobuline.
- 4 : Une fois le dimère chaîne lourde-chaîne légère formé, une autre protéine chaperonne intervient, la calreticuline. Elle assure la formation d'un complexe stable.
- 5 : Liaison du complexe à la TAP. A cette étape, le chargement du peptide commence, il stabilise le complexe final pendant le transit jusqu'à la surface cellulaire, pour aboutir à sa présentation en surface cellulaire : 6.

#### Présentation croisée

- 7 : Protéine exogène internalisée dans la cellule par phagocytose.
- 8 : Le transfert dans le cytosol est effectué par un mécanisme encore mal connu. Une fois dans le cytosol, l'antigène est hydrolysé par le protéasome en oligopeptides qui sont pris en charge par la protéine de transport TAP et chargés sur les molécules du CMH de classe I au niveau du réticulum endoplasmique.

## Structure des gènes et molécules de classe II

Les molécules HLA de classe II sont des hétérodimères constitués de deux glycoprotéines  $\alpha$  et  $\beta$ , transmembranaires, associées entre elles par des liaisons non covalentes (**Figure 4**). Leur structure est proche de celle des molécules de classe I, en revanche chacune des chaînes glycoprotéiques  $\alpha$  et  $\beta$  sont codées par les gènes du CMH de classe II. Ces derniers sont séparés en cinq sous-régions DR, DQ, DP, DO et DM. Chacune d'elles, contient des gènes A et B, codant respectivement pour les chaînes  $\alpha$  et  $\beta$  des molécules DR, DQ, DP, DO et DM.

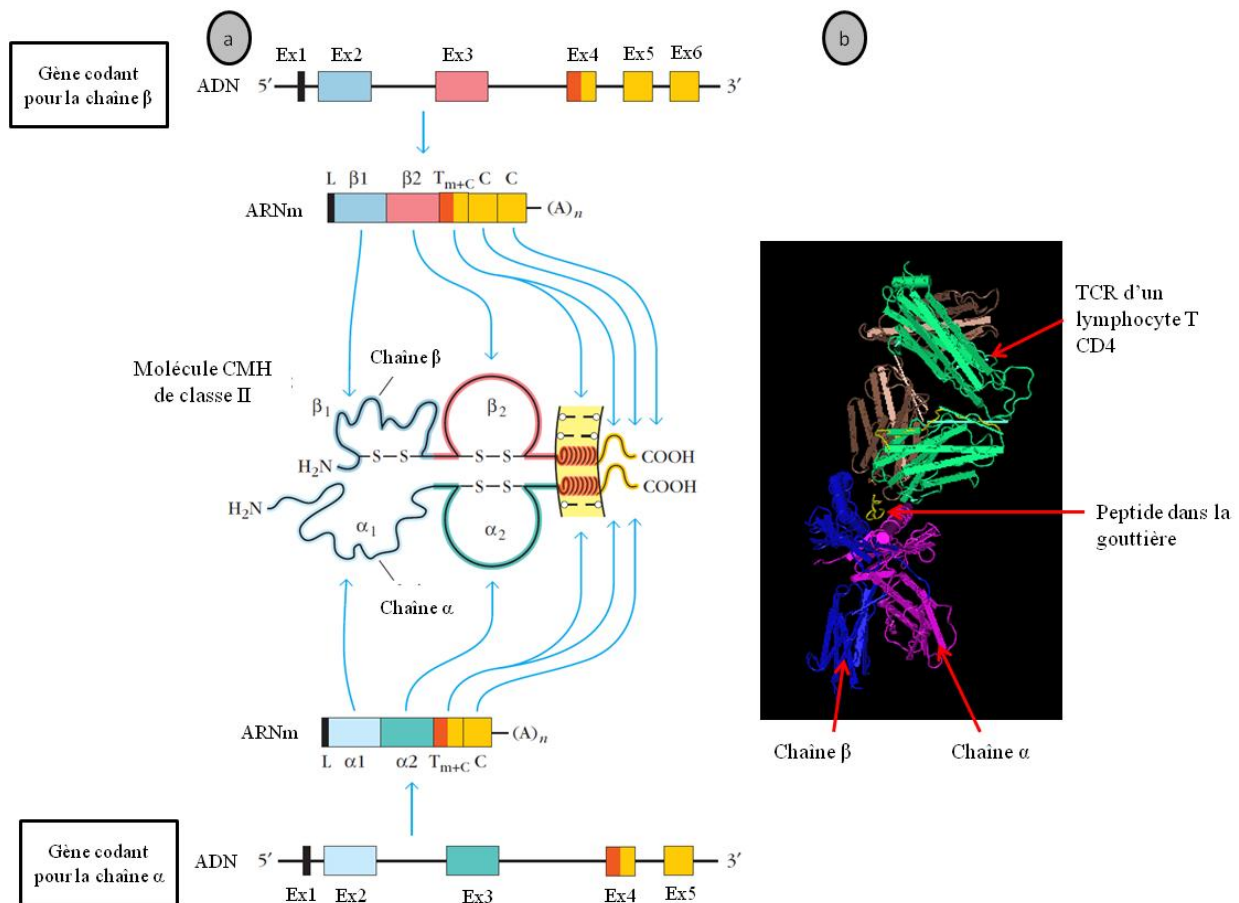
Les gènes A (DRA, DQA, DPA, DOA et DMA) comportent 5 exons, leur polymorphisme est restreint notamment pour le gène DRA qui ne présente que 7 allèles décrits dans l'espèce humaine. Ceci contraste avec la majorité des gènes B de classe II qui sont hautement polymorphes, notamment DRB1. A noter que celui-ci comporte également 5 exons contrairement aux autres gènes B qui en comptent 6.

Les sous-régions ne sont pas comparables en termes de gènes et de polymorphisme. Ainsi on dénombre au moins 10 gènes au niveau de la famille DR : un unique gène DRA pratiquement invariant, qui est à l'origine de la chaîne DR $\alpha$ , et neuf gènes DRB. Parmi eux, quatre sont polymorphes et codent pour une chaîne DR $\beta$  (DRB1, DRB3, DRB4 et DRB5), les cinq autres sont des pseudogènes (DRB2, DRB6, DRB7, DRB8 et DRB9). La molécule membranaire mature HLA-DR est formée de l'association d'une chaîne DR $\alpha$  et d'une chaîne DR $\beta$ , elle doit finalement son grand polymorphisme aux seuls gènes DRB [24].

La sous-région DQ peut être décrite plus simplement, puisqu'elle ne présente que deux gènes HLA-DQA (A1 et A2) et trois gènes HLA-DQB (B1, B2 et B3). Seuls les gènes DQA1 et DQB1 sont fonctionnels. Une description similaire de la région DP peut être apportée, avec cette fois trois gènes connus pour HLA-DPA (A1, A2 et A3) et deux gènes décrits pour HLA-DPB (B1 et B2). De la même façon, seuls les gènes DPA1 et DPB1 sont fonctionnels. HLA-DPA2, -DPA3 et -DPB2 sont des pseudogènes [25].

Quant aux sous régions DO et DM, elles comportent chacune deux gènes (DOA, DOB et DMA, DMB) dont l'expression résulte en une molécule hétérodimérique de classe II non classique impliquée dans le mécanisme d'apprêtement de l'antigène (**Figure 5**) [22].

Au total, six gènes sont principalement impliqués et étudiés en histocompatibilité lors des greffes : trois de classe I (HLA-A, B, C) et trois de classe II (HLA-DRB1, DQB1, et DPB1). Les gènes DRB3, DRB4, DRB5 sont étudiés dans le contexte de la greffe de cellules souches hématopoïétiques mais seulement occasionnellement en transplantation d'organe. Ils ne sont étudiés (comme pour les gènes DQA1 et DPA1) que lorsqu'une immunisation a été mise en évidence chez le receveur.



**Figure 4 : Représentation schématique des gènes de classe II du CMH, de l'ARNm transcrit et des protéines obtenues.**

D'après R.A. Goldsby, T.J. Kindt et B.A. Osborne [15].

- Représentation de deux molécules d'ADN correspondant à des gènes HLA de classe II (codant pour les chaînes  $\alpha$  et  $\beta$ ), de leurs transcrits et de la molécule obtenue.
- Structure tridimensionnelle de la molécule HLA de classe II, et interaction avec le TCR d'un lymphocyte T CD4. (PDB ID : 4MAY)

Les deux chaînes s'associent entre elles pour former une molécule composée d'une partie extracellulaire comprenant deux domaines de chaque chaîne ( $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$  et  $\beta_2$ ). Les extrémités  $\alpha_1$  et  $\beta_1$  constituent le sillon servant à accueillir les peptides. Cette cavité est délimitée par des parois formées de deux hélices  $\alpha$  et par un plancher de feuillets  $\beta$ . Tout comme pour les molécules de classe I, les domaines proximaux ( $\alpha_2$  et  $\beta_2$  pour les molécules de classe II et  $\alpha_3$  pour la chaîne  $\alpha$  des molécules de classe I) ont une structure proche de la superfamille des immunoglobulines [7] [22].

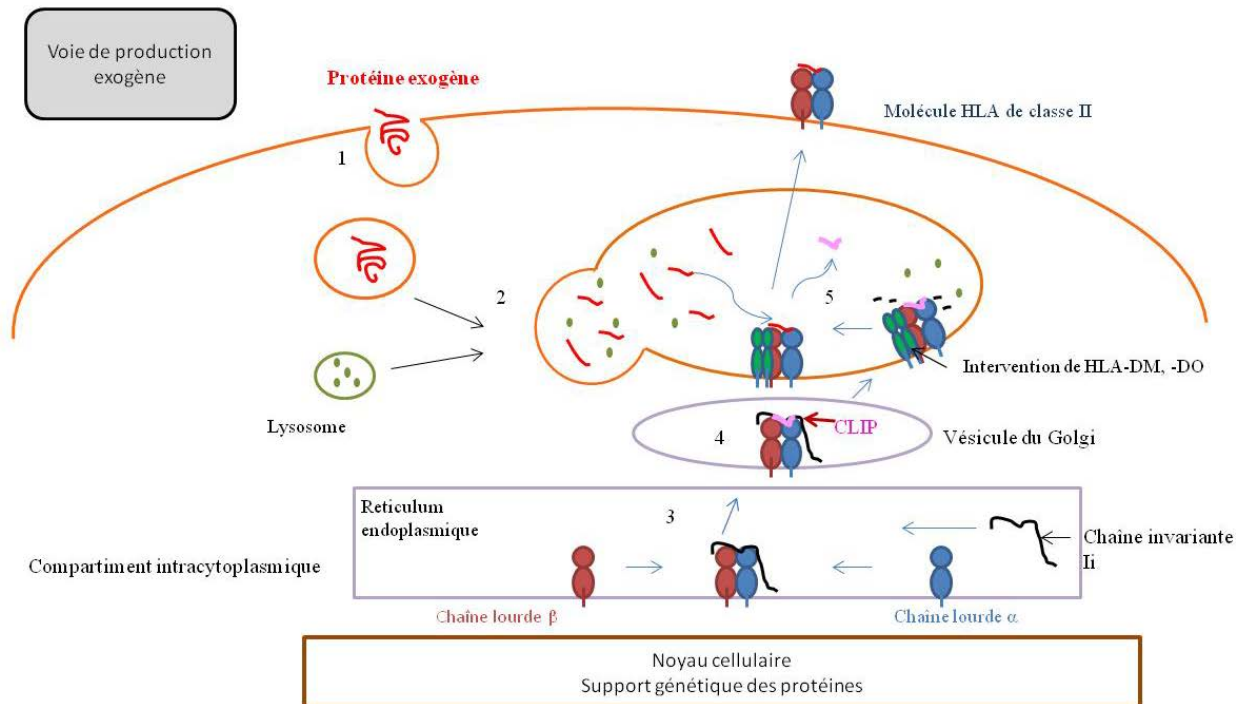
## Expression des molécules de classe II

L'expression cellulaire des molécules de classe II est limitée aux cellules présentatrices d'antigènes (CPAs) du système immunitaire que sont les lymphocytes B, les monocytes / macrophages, les cellules dendritiques, les cellules de Langerhans épidermiques et les cellules de Kupffer du foie. Elles sont chargées de la présentation des peptides aux lymphocytes T CD4+ [26].

Leur expression membranaire est régulée par l'interféron gamma (IFN  $\gamma$ ), qui peut induire une exacerbation de l'expression en surface cellulaire mais aussi générer des molécules de classe II à la surface de nombreuses cellules qui deviennent alors des cellules présentatrices d'antigènes accessoires [24]. Leur expression est aussi sous le contrôle d'autres effecteurs comme les interleukines 4 et 10 (IL-4 et IL-10) ou encore le TNF $\alpha$ . La régulation moléculaire de leur expression fait l'objet d'importantes recherches, afin de comprendre le fonctionnement complexe des promoteurs et des facteurs de transcription qui permettent d'établir un équilibre entre les réactions vis-à-vis de peptides étrangers, et l'autoréactivité contre les peptides du soi [26].

## Nature des peptides présentés

La source des peptides présentés par les molécules de classe II provient de protéines dites exogènes, qui pénètrent dans la cellule après phagocytose par les CPAs, pour subir ensuite une dégradation lysosomale. Il s'agit donc de protéines qui ne sont pas synthétisées dans la cellule. Les peptides logés dans la niche des molécules de classe II peuvent être plus grands que ceux de la classe I, allant de 10 à 24 résidus d'acides aminés, les extrémités de la gouttière étant ouvertes. La voie d'assemblage des molécules HLA de classe II et le mécanisme d'apprêtement des peptides par la voie exogène sont schématisés dans la **figure 5**. A noter que les molécules de classe II qui n'ont pas fixé de peptide après assemblage sont rapidement dégradées par la voie endosomale, et présentées à leur tour par leurs semblables. Ceci explique pourquoi les peptides dérivés de molécules de classe II sont bien représentés dans le répertoire des peptides du soi [22].



**Figure 5 : Schéma d'apprêtement des protéines exogènes par une molécule HLA de classe II**

- 1 : Pénétration de la protéine exogène dans le compartiment cytoplasmique par phagocytose.
- 2 : Fusion du phagosome avec des lysosomes et puis avec des vésicules provenant de l'appareil de Golgi, dans lesquelles s'effectue la maturation des molécules de classe II.
- 3 : Parallèlement, dans le RE, a lieu l'assemblage des chaînes  $\alpha$  et  $\beta$ . Cette opération est assistée par une troisième molécule, la chaîne invariante Ii. Le complexe ternaire formé est maintenant apte à migrer vers l'appareil de Golgi.
- 4 : La chaîne invariante Ii exerce alors un rôle important, empêchant le détournement du complexe vers la voie de dégradation endosomale. Elle assure également le blocage de la gouttière à peptide par une partie de la chaîne, nommée CLIP (*Class II Associated invariant chain peptide*), qui occupe la cavité pour empêcher le chargement d'un peptide endogène.
- 5 : Lorsqu'arrive l'opportunité de chargement d'un peptide exogène, Ii est alors clivée, autorisant le déplacement du complexe vers un compartiment endosomal. Seule la partie CLIP reste en place, elle sera dissociée au moment de la fixation du peptide. C'est à cette même étape que les molécules chaperonnes HLA-DM semblent intervenir, en facilitant cette dissociation. Le rôle de HLA-DO reste encore énigmatique, mais il semblerait similaire à celui de HLA-DM, une aide au chargement des peptides. Une fois le peptide chargé, un changement conformationnel stabilise l'hétérodimère  $\alpha/\beta$ , pour un meilleur transport à la surface cellulaire.

## 4. Fonctions des molécules HLA

Les molécules HLA (classiques) sont réparties en différentes classes selon des critères de structure et d'expression, mais aussi de fonction. Si les classes I et II ne partagent pas complètement les mêmes spectres d'expression, elles ont en commun le rôle de présentation des peptides aux lymphocytes, T CD8<sup>+</sup> pour les molécules de classe I et T CD4<sup>+</sup> pour les molécules de classe II.

Contrairement aux lymphocytes B, qui reconnaissent des antigènes natifs, les lymphocytes T ne peuvent détecter que les antigènes présentés par les molécules HLA de classe I ou de classe II, sous une forme dégradée en peptides. C'est pourquoi on parle de **restriction par le HLA** dans le dernier cas.

Les molécules HLA ont donc un lien étroit avec les lymphocytes T, avec qui elles interagissent grâce aux récepteurs de surface, appelés TCR. Plus précisément, il s'agit du TCR $\alpha\beta$ , qui possède une structure proche des immunoglobulines. Nous ne développerons pas dans cette thèse l'autre population de lymphocytes T qui possèdent des TCR $\gamma\delta$ , que l'on retrouve essentiellement dans les muqueuses. Ces derniers ont la particularité de reconnaître des antigènes viraux ou bactériens indépendamment des molécules HLA. Les lymphocytes T porteurs d'un TCR $\alpha\beta$  et exprimant le corécepteur CD8 reconnaissent les peptides enchassés dans les molécules HLA de classe I, tandis que ceux exprimant le corécepteur CD4 répondent aux antigènes présentés par les molécules de classe II. Les corécepteurs se lient au domaine non polymorphique des molécules HLA, c'est-à-dire aux domaines proximaux extracellulaires ( $\alpha 3$  pour la classe I et  $\alpha 2/\beta 2$  pour la classe II) de la molécule HLA [27]. L'interaction entre le TCR, le peptide apprêté et la molécule HLA génère un signal intracellulaire qui permettra une réponse spécifique du lymphocyte T [28].

Le rôle de présentation est essentiel dans deux grandes étapes de la vie des lymphocytes T, à savoir leur éducation à la tolérance du soi dans le thymus et leur recrutement et stimulation dans la réponse immunitaire adaptative face à des peptides étrangers [29]. Les précurseurs des lymphocytes T sont acheminés de la moelle osseuse vers le thymus. C'est dans cet organe lymphoïde primaire que se joue leur sélection, les autorisant ou non à poursuivre leur vie en dehors du thymus, pour rejoindre les organes lymphoïdes secondaires (OLS). Ils subissent dans le thymus deux sélections :

- ✓ **La sélection positive**, qui veille à promouvoir une réponse immunitaire efficace contre les antigènes étrangers en sélectionnant des lymphocytes T équipés de TCR capables de fonctionner efficacement avec les molécules HLA de l'individu. Autrement dit les lymphocytes T sélectionnés dans le thymus doivent être capables de reconnaître des peptides étrangers au soi, enchassés dans les molécules HLA de l'individu. Les lymphocytes T capables de reconnaître les molécules HLA de classe I *via* leur TCR vont devenir des lymphocytes T CD8<sup>+</sup> alors que ceux capables de

reconnaître les molécules HLA de classe II vont devenir des lymphocytes T CD4+ [30].

- ✓ **La sélection négative**, qui assure le phénomène de tolérance au soi en détectant et détruisant les lymphocytes T autoréactifs, qui reconnaissent les peptides du soi présentés par les molécules HLA.

Après leur maturation thymique, les lymphocytes T naïfs et quiescents migrent dans les OLS. Pendant cette phase, leur rencontre avec un peptide à l'origine de leur stimulation se fait par l'intermédiaire des molécules de classe I et II des CPAs professionnelles. Il s'agit des cellules dendritiques, qui migrent dans les OLS après acquisition d'un antigène étranger. Les CPA présentent les peptides aux lymphocytes T CD8+ et CD4+ grâce, respectivement, à leurs molécules de classe I et de classe II. Elles mettent également en jeu le phénomène de présentation croisée, précédemment détaillé, permettant ainsi la présentation de peptides d'origine exogène par les molécules HLA de classe I, aux lymphocytes TCD8+. L'éducation vis-à-vis des antigènes tumoraux passe aussi par ce mécanisme, après l'endocytose de débris de cellules tumorales par les CPAs. Les lymphocytes T passent ensuite dans le sang périphérique, à la recherche des antigènes étrangers à éliminer. Leur stimulation déclenche l'activité cytolytique des lymphocytes T CD8+ et l'activité de synthèse de cytokines des lymphocytes T CD4+ auxiliaires. Ces cytokines entretiennent la réponse immunitaire à la fois cellulaire, en activant les lymphocytes T, mais aussi humorale en activant les lymphocytes B. Il existe différentes sous-populations de lymphocytes T CD4+ qui ne seront pas détaillées dans cette thèse. Il s'en suit une expansion clonale des lymphocytes T cytotoxiques, cellules effectrices, mais aussi une différenciation en lymphocytes T mémoires. Si ce même motif étranger est de nouveau rencontré par un lymphocyte T qui n'est plus « naïf », sa cytotoxicité sera directement déclenchée par les molécules de classe I des cellules qui présentent l'antigène [23]. Ainsi, en périphérie, a lieu une sélection des lymphocytes capables de reconnaître les antigènes étrangers.

Les molécules HLA de classe I ont pour autre mission la surveillance de l'intégrité cellulaire. En effet, comme nous l'avons déjà évoqué un peu plus tôt, les molécules HLA de classe I sont reconnues par les récepteurs de surface des cellules *Natural Killer*, au niveau de leurs domaines non polymorphiques. Ces récepteurs jouent un rôle fondamental de régulation de l'activité cytolytique et productrice de cytokines des cellules NK. Trois grandes familles de récepteurs ont été décrites à leur surface, la plus connue appartient à la super famille des immunoglobulines, nommée *Killer cell Ig-like Receptor* (KIR). La liaison d'une molécule HLA de classe I au récepteur KIR entraîne la génération d'un signal qui peut être inhibiteur ou activateur, selon le type de protéine KIR [16]. Ainsi, une cellule qui a perdu partiellement ou totalement l'expression de ses molécules HLA de classe I sera détruite lors de sa rencontre avec une cellule NK.

Enfin, les molécules HLA de classe II peuvent servir de récepteurs spécifiques, pour des antigènes qui ne nécessitent aucune dégradation préalable pour activer une reconnaissance et activation des lymphocytes T. On parle de super antigènes. Ils se fixent sur la région non polymorphique  $\alpha 2/\beta 2$  des molécules HLA de classe II.

De façon indirecte, les molécules HLA interviennent dans la susceptibilité à développer certaines pathologies. Du fait de leur important polymorphisme, il existe des différences d'interactions entre les molécules HLA et d'autres molécules, provenant aussi bien de l'organisme que de substances médicamenteuses. Elles conduisent à des effets délétères, ou parfois bénéfiques. De ceci découle une autre application médicale de l'étude du HLA en laquelle croyait beaucoup Jean Dausset, la médecine prédictive, utilisée aujourd'hui couramment [2]. Elle permet l'aide au diagnostic de certaines maladies qui sont en lien avec des allèles HLA, certains pouvant protéger ou au contraire prédisposer à une pathologie ou réaction par des mécanismes divers et variés qui ne seront pas développés ici (exemple : HLA A\*29:02 et Birdshot [31], hypersensibilité à l'abacavir et HLA-B\*57 :01...) [7].

## 5. Nomenclature

La nomenclature HLA doit respecter des critères précis qui permettent son homogénéité entre les résultats au niveau international (**Figure 6**). Elle est tenue à jour par le comité international « *WHO nomenclature committee* » [32].

La nomenclature est adaptée à la technique utilisée pour réaliser le typage :

- ✓ **En sérologie**, on définit des spécificités. Une spécificité regroupe plusieurs allèles qui possèdent le même profil de réactivité sérologique. Il est courant d'utiliser le terme « broad » pour définir une spécificité large que l'on peut subdiviser en « split ». Ces derniers ont été découverts dans un second temps avec la mise en évidence d'anticorps anti-HLA plus spécifiques.

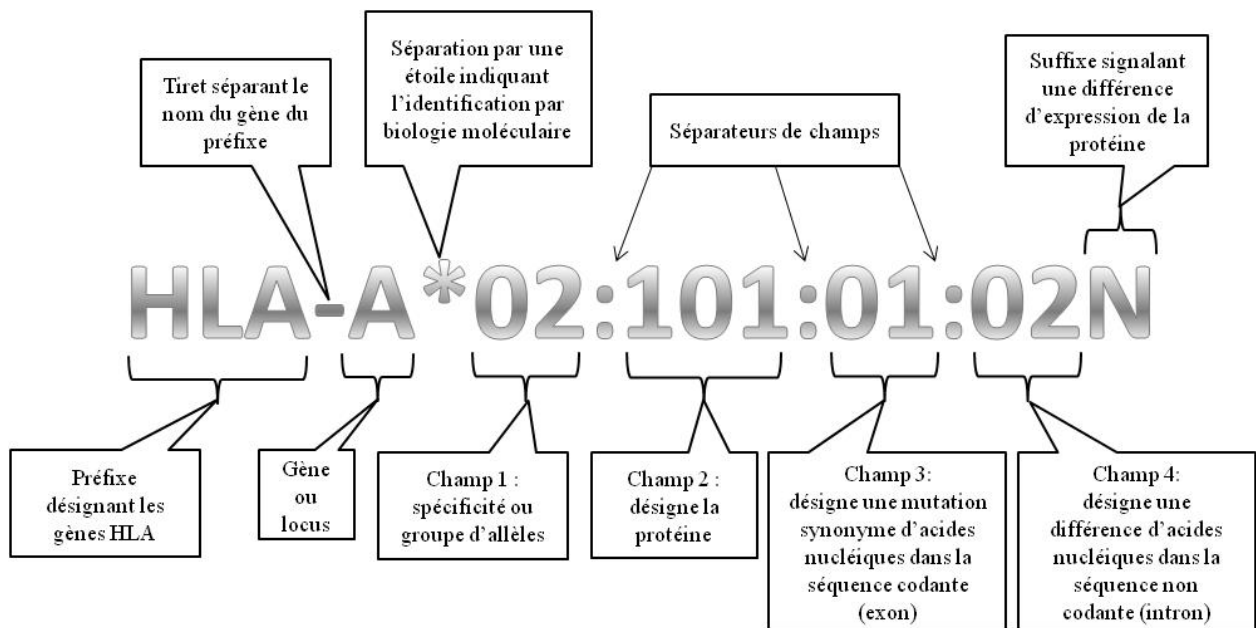
Pour les allèles HLA-C définis en sérologie, la lettre **w** (pour workshop) a été ajoutée afin d'éviter les confusions avec les facteurs du complément. Ex : HLA-Cw1

HLA-A1

- ✓ Pour les allèles définis par **biologie moléculaire**, une étoile suit systématiquement l'identification du locus. On distingue trois types de résolution de typage :
  - Le typage générique, dit de « **basse résolution** », on parle aussi de résolution 2 digits :

HLA-A\*01

- Le typage allélique, dit de « **haute résolution** », on parle de résolution 4 digits. Avec certaines techniques de biologie moléculaire, on peut obtenir une haute résolution de 6 digits, voire même 8 digits [33]. La **figure 6** apporte des précisions sur ces derniers digits.



**Figure 6 : Nomenclature des allèles HLA.**

D'après <http://hla.alleles.org/>.

Le suffixe « N » désigne un allèle nul, c'est-à-dire codant pour une protéine qui n'est pas exprimée à la surface cellulaire. « L » désigne un allèle codant pour une protéine dont l'expression à la surface cellulaire est significativement réduite. « S » correspond aux protéines sécrétées sous forme soluble et enfin « Q » signifie que l'expression demeure indéterminée.

- Dans de nombreux cas, les techniques de biologie moléculaire actuelles comme les techniques de PCR-SSP et PCR-SSO, qui vont être détaillées dans le chapitre III. 1, ne permettent pas d'assigner un unique allèle. Très souvent, un groupe d'allèles dont la séquence est proche, et parmi lesquels on retrouve un allèle majoritairement fréquent, est déterminé. Ce groupe d'allèles correspond à une liste d'ambiguïtés. Elles sont regroupées sous une combinaison de lettres appelée code NMDP (*National Marrow Donor Program*). Le résultat de typage obtenu est dit de « **résolution intermédiaire** » ou « **moyenne résolution** ».

HLA-A\*68:HUBJ

Par exemple le code NMDP « HUBJ » regroupe la liste d'allèles suivants : HLA-A \*68:01/68:06/68:07/68:08/68:11N/68:17/68:21/68:24/68:29/68:32/68:33/68:37/68:38/68:46/68:47.

Lorsqu'un seul antigène ou allèle est identifié à un locus en sérologie ou biologie moléculaire, on parle de blanc, noté (-) pour le deuxième antigène ou allèle. Le phénotype ne peut être présenté avec les antigènes exprimés deux fois que lorsque l'homozygotie est prouvée par l'étude familiale ou lorsque le typage allélique identifie clairement une hétérozygotie avec la présence de deux allèles distincts de même spécificité [9].

Les standards EFI (*European Federation for Immunogenetics*) en vigueur (version 6.2) imposent que les typages HLA réalisés dans le cadre des greffes de CSH avec donneurs non apparentés soient des typages dits de haute résolution. Les seules ambiguïtés tolérées sont celles correspondant à un sous-groupe d'allèles dont le polymorphisme est situé en dehors du/des exons 2 et/ou 3 selon la classe, et qui ne sont donc pas reconnus par les TCR des lymphocytes T. Les allèles nuls appartenant aux ambiguïtés restantes sont tenus d'être recherchés. Actuellement au laboratoire, on combine plusieurs techniques de typage de biologie moléculaire (PCR-SSO et PCR-SSP) afin d'éliminer un grand nombre d'ambiguïtés mais parfois, ces deux techniques ne suffisent pas pour être en conformité avec la réglementation EFI en vigueur [9].

Le typage haute résolution est à différencier d'un **typage allélique** dans le sens où le typage allélique nécessite de définir l'allèle tel qu'il est défini dans la nomenclature en cours (en 2, 4, 6 ou 8 digits selon les allèles comme par exemple HLA-A \* 01: 01: 01: 01, A \* 01: 07). Cela imposerait au laboratoire d'obtenir la séquence entière du gène sans aucune ambiguïté même dans les régions introniques, 5'UTR et 3'UTR).

Les allèles appartenant au même **groupe G** sont les allèles qui partagent une séquence nucléotidique identique au niveau des exons codant pour les domaines d'ancrage du peptide, autrement dits les domaines les plus polymorphiques des molécules HLA (exon 2 et 3 pour la classe I, exon 2 pour la classe II). Ces groupes d'allèles sont désignés par un G majuscule qui suit le 6<sup>ième</sup> digit de la désignation de l'allèle qui porte le plus petit numéro du groupe. Ce groupe est à différencier du **groupe P**, qui rassemble les allèles dont l'expression protéique résulte en la même séquence d'acides aminés au niveau des domaines précédemment cités (exon 2 et 3 pour la classe I, exon 2 pour la classe 2). Un P majuscule suit cette fois le 4<sup>ième</sup> digit du premier allèle du groupe. Deux codons d'acides nucléiques différents peuvent en effet désigner le même acide aminé, dû à la dégénérescence du code génétique [25].

Il existe un catalogue, publié depuis 2007 et mis à jour annuellement, des **Common and Well Documented** allèles (CWD). Les « *common* » allèles correspondent à des allèles que l'on retrouve dans de nombreuses populations et dont la fréquence dans une population de référence (de 1500 individus minimum) est supérieure à 0,001. Les allèles répondant aux critères de la catégorie « *well documented* » ont été observés beaucoup plus rarement que ceux de la première catégorie. Ainsi ils doivent avoir été identifiés par la méthode de séquençage de type Sanger SBT (*Sequencing Based Typing*), ou bien cinq fois chez des individus non apparentés, ou bien au moins trois fois dans un même haplotype. Ce terme de

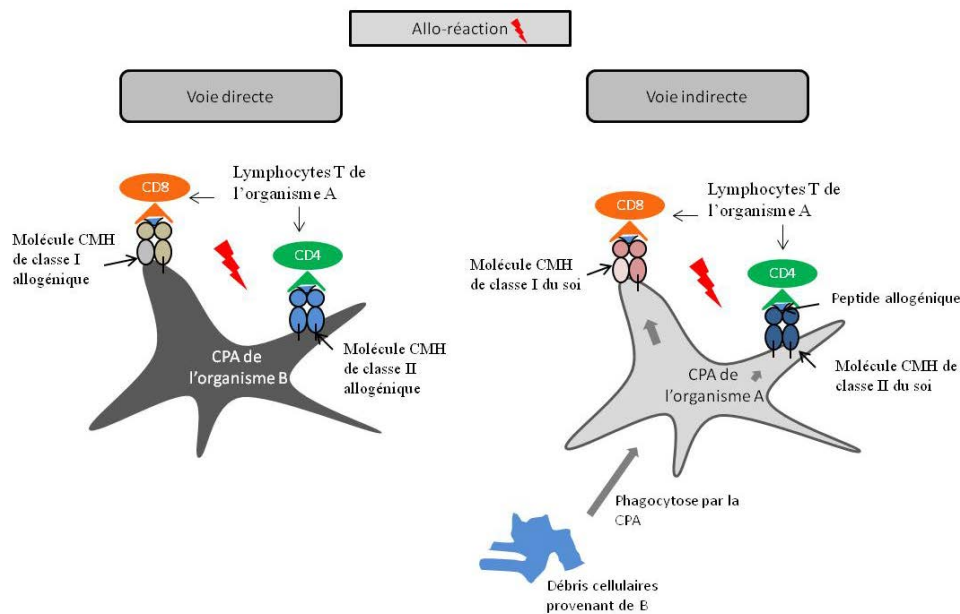
« *well documented* » a été choisi pour éviter des confusions avec la définition des allèles rares qui diffère selon les registres. Les **allèles rares**, d'après la définition du NMDP, répondent à une fréquence maximale qui est fonction de leur classe (par exemple une fréquence < 1/50 000 pour la classe I) [32] [34] [35].

## 6. Système HLA et transplantation, notion d'allo-réactivité

L'importance des molécules HLA en transplantation s'explique par leur impressionnant polymorphisme, avec pas moins de 13023 allèles HLA recensés selon l'IMGT (*ImMunoGeneTics*) database en Août 2015. Parmi eux, on compte actuellement 3107 allèles décrits pour HLA-A, 3887 pour HLA-B et 1726 pour HLA-DRB1 [32]. En conséquence, les molécules HLA occupent le premier rôle dans le phénomène d'allo-réactivité (**Figure 7**). On peut définir **l'allo-réactivité** comme la capacité des lymphocytes T à être stimulés par des molécules provenant d'un organisme étranger au leur. Cependant ce terme convient pour deux organismes qui sont de la même espèce, sinon on parle de xéno-réactivité. Les molécules mises en jeu, appelées allo-antigènes, sont dans 90% des cas des molécules du CMH, d'où leur nom **d'antigènes majeurs** [36]. On verra qu'ils ne sont pas les seuls acteurs de l'allo-réactivité, il existe en effet d'autres antigènes impliqués dans ces conflits immunitaires, appelés **antigènes mineurs**. Brièvement, il s'agit de protéines qui possèdent un polymorphisme et qui ne sont pas issues de la transcription d'un gène appartenant au CMH. Cette variété génétique peut être délétère lors de la greffe de cellules souches hématopoïétiques. Le chapitre concernant la greffe de CSH apporte un paragraphe plus détaillé sur les antigènes mineurs.

Lors d'une greffe, les lymphocytes T du receveur vont rencontrer les CPA du donneur, et reconnaître à leur surface les molécules HLA de classe I et II qu'elles expriment. Cette rencontre active directement la fonction de défense des lymphocytes T, qui détruisent le greffon par leur cytotoxicité. Ce phénomène **d'allo-réactivité directe** entraîne un rejet s'il n'est pas limité par un conditionnement suffisamment myélo-ablatif (dans le cas de la greffe de moelle) et par les immunosuppresseurs administrés en prévention. Cette notion de réactivité directe peut surprendre aux vues des conditions strictes de compatibilité définies dans le cas de la greffe de moelle osseuse, pour le choix du donneur. En effet, elles doivent limiter considérablement les conflits antigéniques et donc les différences de structures entre les molécules HLA du receveur et du donneur, de façon à leurrer les lymphocytes T. Ces derniers ne devraient donc pas réagir vis à vis de « motifs connus », puisqu'ils ont été sélectionnés pendant leur éducation thymique. Cependant l'élimination des clones auto-réactifs au cours de la sélection négative ne prévient pas la réaction face à des molécules HLA allogéniques. Il faut donc admettre qu'il existe une réaction des lymphocytes T vis-à-vis des complexes peptides-CMH allogéniques, même si ces derniers ont une forte analogie avec les complexes peptides-CMH du soi.

Suite à cette allo-réaction directe relativement précoce, les débris engendrés par la destruction des cellules du donneur, sont endocytés par les CPA du receveur, qui entraînent l'apprêtement par leurs molécules HLA, de peptides issus du donneur. Ces motifs antigéniques étrangers présentés aux lymphocytes T du receveur stimulent également la triade de la réponse classique immunitaire décrite précédemment. On parle ici **d'allo-réactivité indirecte**. Ce qui permet aux antigènes mineurs d'être responsables d'une réaction immunitaire [36].



**Figure 7 : Mécanismes d'allo-réactivité directe et indirecte.**

Dans le contexte de la greffe de cellules souches hématopoïétiques, l'organisme A représente le donneur, et l'organisme B le receveur.

Au total, l'allo-réactivité directe consiste en la réaction de lymphocytes T d'un organisme A (donneur ou receveur) vis-à-vis de molécules HLA allogéniques, en surface des CPAs (receveur ou donneur respectivement) provenant d'un organisme B. En revanche, l'allo-réactivité indirecte correspond à la stimulation de lymphocytes T d'un organisme A (donneur ou receveur) vis-à-vis de molécules HLA présentant un peptide issu de B, en surface des CPAs provenant de l'organisme A. Cette définition plus générale permet de mieux comprendre le problème rencontré dans la greffe de cellules souches hématopoïétiques où les mécanismes sont inversés. Alors qu'en transplantation d'organe, le principal produit de l'allo-réactivité directe et indirecte à maîtriser est le rejet, la transplantation de cellules souches fait redouter le phénomène inverse, appelé réaction du greffon contre l'hôte (GvH). Dans ce dernier cas, le greffon contient des cellules immunocompétentes en quantité importante. Inversement, le receveur ne possède qu'une faible quantité de cellules immunocompétentes puisqu'elles ont été détruites par le conditionnement. Par conséquent on craint davantage la GvH. Sa physiopathologie sera détaillée dans le chapitre suivant qui traite plus précisément des complications de la greffe de moelle.

Finalement, le CMH apparaît comme une région très riche en gènes, dont une grande partie n'est pas exprimée. Bien qu'il reste désigné avec le terme «histocompatibilité», de nombreux gènes ne codent pas pour des allo-antigènes, et interviennent dans des processus indépendants du système immunitaire. L'extraordinaire polymorphisme génétique de ce complexe garantit sans aucun doute l'adaptation et la survie de l'espèce, notamment en assurant des défenses immunitaires efficaces et adaptées aux agents infectieux, qui évoluent également de leur côté. Il est aussi la cause de l'allo-immunisation fœto-maternelle et des réactions immunitaires allogéniques provoquées par la transfusion plaquettaire, la transplantation et la greffe de cellules souches hématopoïétiques en cas d'incompatibilité HLA.

## II. La greffe de moelle osseuse

### 1. Quelques définitions :

Chez de nombreux patients atteints d'hémopathies malignes, **l'allogreffe de cellules souches hématopoïétiques (CSH)** constitue le seul traitement curatif. Elle est réalisée par administration intra-veineuse des CSH. Son objectif est de remplacer totalement la moelle osseuse du receveur, qui produit un clone malin, par les CSH saines d'un donneur. On parle parfois de transplantation mais le terme est moins adapté puisqu'il désigne l'implantation d'un organe entier avec rétablissement des connexions vasculaires [37].

Les cellules souches possèdent un marqueur d'immaturité en surface membranaire, le CD34+, qui est utilisé pour la numération des CSH dans les prélèvements effectués chez le donneur et en apprécier la richesse. La greffe implique au préalable la destruction complète ou quasi complète des cellules hématopoïétiques du receveur, et par conséquent celle des cellules tumorales. On appelle cette étape obligatoire le **conditionnement**, il permettra également la vacuité médullaire ainsi que l'immunosuppression de l'hôte pour éviter le rejet de la greffe. On distingue deux types de conditionnement :

- ✓ **Myéloablatif** : le patient est mis en aplasie totale par une chimiothérapie qui peut être accompagnée d'une irradiation corporelle totale. Ce traitement est fortement risqué pour le patient qui peut subir des effets secondaires immédiats ou tardifs pouvant mettre en jeu le pronostic vital. Ce type de conditionnement, beaucoup moins utilisé maintenant, ne laissait pas accès à la greffe chez des personnes âgées ou fragiles.
- ✓ **Non myéloablatif** (ou *Reduced Intensity Conditionnement*, RIC) : le patient subit une chimiothérapie et/ou radiothérapie moins intense que dans le cas précédent et par conséquent les effets secondaires sont mieux tolérés. Le principal intérêt de ce conditionnement non myélo-ablatif est plutôt l'immunodépression du patient qui doit éviter le rejet de la greffe. C'est l'effet GvL (*Graft versus Leukemia*, défini ci-dessous)

de la greffe qui devra détruire des cellules tumorales non éliminées par le conditionnement.

L'intensité du conditionnement dépend du risque de rechute du patient dans l'hémopathie mais également du donneur sélectionné, c'est-à-dire du degré d'histocompatibilité de la greffe [38].

Dans le cas des CSH, des conditions importantes de compatibilité sont requises pour limiter les différentes réactions immunologiques pouvant avoir des conséquences dramatiques. La première réaction redoutée dans toute greffe (organe, tissulaire et cellulaire) est le **rejet du greffon** (*Host-versus-Graft disease* ou HvG). Elle est causée par la réaction des lymphocytes T de l'hôte qui n'ont pas été totalement détruits par le conditionnement, et qui s'activent contre les antigènes HLA étrangers du greffon. Ce processus physiopathologique a été décrit précédemment avec l'allo-réactivité directe et indirecte. La deuxième réaction n'existe pas dans les autres types de greffe (organe et tissulaire), il s'agit de la **réaction du greffon contre l'hôte** (*Graft-versus-Host disease* ou GvH). Elle est liée au mécanisme inverse de la première réaction, c'est-à-dire à l'activation des lymphocytes T du greffon, immunocompétents, qui détectent les alloantigènes HLA du receveur. Il peut en résulter des dommages sur de nombreux organes du receveur selon le degré de gravité de la GvH, pouvant aller jusqu'au décès du patient. Cette notion sera abordée plus précisément dans la suite de ce chapitre. Cette incompatibilité antigénique comporte cependant un avantage, en effet les lymphocytes T et les cellules NK apportées par le greffon reconnaissent également les cellules tumorales résiduelles et entraîneront leur destruction. On parle alors de réaction du **greffon contre la leucémie** (*Graft versus Leukemia* ou GvL). On comprend donc que la difficulté dans toute greffe est de trouver un niveau d'incompatibilité minimum qui permette une GvL suffisante mais qui n'entraîne pas une réaction GvH trop sévère.

Les indications des greffes de moelle osseuse ne sont pas limitées aux leucémies, elles se sont étendues dans les années 1980 à des pathologies telles que d'autres hémopathies, malignes (syndromes myélodysplasiques), ou non comme les thalassémies, les drépanocytoses et les déficits immunitaires congénitaux [39].

Il existe plusieurs modalités de greffes :

- ✓ La greffe **autologue** consiste à greffer au patient ses propres cellules souches, elles sont par conséquent HLA identiques. Un prélèvement de moelle osseuse ou de cellules souches périphériques est réalisé chez le patient en phase de rémission complète, puis les cellules sont congelées jusqu'à leur réinjection. Dans ce cas il n'existe aucun risque de GvH, en revanche le risque de rechute leucémique est élevé dans ce contexte. Les principales indications sont le lymphome non hodgkinien et le myélome multiple [40].
- ✓ La greffe **syngénique** est une greffe réalisée entre deux jumeaux monozygotes, donc HLA identiques. Ce type de greffe est rarement réalisé, il permet éventuellement le remplacement des greffes autologues (greffon potentiellement contaminé par des cellules tumorales), chez des patients en aplasie après une très forte dose de

chimiothérapie, dans des cas de cancers métastasés par exemple. Les résultats obtenus seraient équivalents à ceux des greffes autologues, bien que l'objectif escompté soit de limiter le risque de rechute néoplasique [41]. Dans ce contexte, aucun effet GvL n'est recherché.

- ✓ La greffe **allogénique**, quant à elle, est réalisée avec un donneur. Il peut être apparenté (mais pas un jumeau monozygote) et le plus souvent dans ce cas, il sera HLA **géo-identique**. Les différences portent sur des gènes en dehors du système HLA. Il s'agira donc d'un membre de la fratrie du receveur. Le donneur peut être non apparenté et dans ce cas on parle de greffe HLA **phéno-identique**. Il s'agira d'un donneur inscrit sur les registres des Donneurs Volontaires de Moelle Osseuse (DVMO). Ce point est détaillé dans le paragraphe « Choix d'un donneur ».

Après l'injection du greffon, les cellules souches hématopoïétiques du donneur rejoignent la moelle osseuse de l'hôte. Si elles prennent possession des lieux, elles permettront une nouvelle hématopoïèse, apportant des cellules immunitaires saines et immunocompétentes. Les cellules résiduelles du receveur, lymphocytes T et Natural killer, s'opposent à cette installation et tentent de rejeter le greffon, ce qui explique la nécessité de les éliminer au maximum avec le conditionnement pré-greffe et de les anergiser avec un traitement immunosuppresseur. Les cellules immunocompétentes du greffon et en particulier les cellules NK, vont fortement participer à la neutralisation des cellules résiduelles et favoriser ainsi la prise de greffe [42].

On estime qu'une quantité de 4 à 8 x 10<sup>6</sup> cellules CD34+/kilogramme (poids du receveur) est nécessaire à la réalisation d'une allogreffe [43]. Dans le cas des greffes allogéniques, les cellules souches hématopoïétiques ont trois origines possibles :

- ✓ La **moelle osseuse**, prélevée le plus souvent par une ponction au niveau des crêtes iliaques, nécessite une anesthésie générale du donneur. Pour obtenir une quantité de CSH suffisante, il faut prélever en moyenne 1400 mL de moelle chez un adulte, avec une limite à 20 mL/kg du poids du donneur et 10 mL/kg pour les enfants [44].
- ✓ Les **cellules souches périphériques (CSP)**, pour lesquelles le prélèvement réalisé par cytophérèse, nécessitent au préalable plusieurs injections au donneur de facteur de croissance granulopoïétique (G-CSF), dans les cinq jours précédents, pour faire «sortir» les CSH de la moelle osseuse et les recueillir dans le sang circulant. Ainsi on obtient un greffon qui peut contenir jusqu'à quatre fois plus de cellules souches CD34+ et dix fois plus de lymphocytes T qu'un greffon médullaire [45].
- ✓ Les **cellules souches de sang de cordon** (Unité de Sang Placentaire ou **USP**) constituent une troisième source, disponible dans des banques, après prélèvement à la naissance du sang de cordon ombilical et congélation [43]. Le nombre de cellules souches est moins élevé que dans un greffon médullaire mais elles ont une très grande capacité de régénération et une immaturité immunologique qui limite le risque de GvH aiguë. Les conditions d'histocompatibilité recherchées sont moins strictes que pour les

sources précédentes, cette notion sera développée dans le paragraphe « Choix d'un donneur ». En raison du faible nombre de cellules souches apportées, les USP sont plus souvent utilisées chez les enfants [38].

Si l'apport de lymphocytes T par le greffon est nécessaire pour obtenir un effet GvL et éviter tout risque de rechute, un excès de lymphocytes T est délétère car l'effet GvH prendra le dessus. C'est l'un des faits constatés avec le développement des greffes à partir de cellules souches périphériques.

Il est possible de réaliser des administrations de lymphocytes T du donneur (*Donnor Lymphocyte Infusion* ou DLI) à distance de la greffe (>21 jours). Ceci permet d'obtenir l'effet GvL suffisant en évitant une GvH, à condition d'injecter une dose limite maximale de lymphocytes T [46].

Quelle que soit l'origine des cellules souches, elles sont administrées au receveur par transfusion lente dans un cathéter central, peu de temps après leur prélèvement (moins de 24h), excepté dans le cas des greffes autologues, qui nécessitent une congélation des cellules puisque l'administration est faite après un délai plus long.

## 2. La réaction du greffon contre l'hôte (GvH)

Comme nous l'avons vu précédemment, la GvH est une complication majeure de la greffe de cellules souches très difficile à contrôler. Elle a été décrite chez l'homme par Billingham dès 1966 [47]. La mortalité post-greffe qu'elle entraîne peut aller de 10% à 50% [46].

Environ 40% des patients avec donneur apparenté HLA identique (cette notion sera détaillée dans le paragraphe « Choix d'un donneur ») et 50% à 70% de patients avec donneur non apparenté HLA identique (10/10) subissent une GvH aiguë. Des données similaires sont retrouvées pour les GvH chroniques. Elles constituent la première cause de mortalité après la rechute leucémique, principalement à cause des complications infectieuses ou de la dégradation progressive des organes [48] [47].

On distingue la GvH aiguë et la GvH chronique selon le délai de survenue, mais celui-ci n'étant pas toujours corrélé, la clinique ou l'histologie est parfois plus discriminante. La GvH aiguë survient dans les 100 premiers jours suivant la greffe. Sa sévérité est classée en grades de I à IV selon les critères de Glucksberg et de A à D selon l'index de l'*International Bone Marrow Transplant Registry* [49]. Les principaux symptômes sont un rash érythémateux maculo-papulaire, des symptômes gastro-intestinaux à type de diarrhées, mucites ou encore une atteinte hépatique. La classification de Glucksberg tient compte de la sévérité de l'atteinte de ces trois organes (peau, système digestif et foie). Les GvH de grade III à IV sont souvent la cause du décès, directe ou indirecte, du patient.

On parle de GvH chronique lorsqu'elle survient dans un délai supérieur à 100 jours post-greffe, il peut s'étendre à plus de 400 jours. On peut la définir comme une pathologie auto-immune qui atteint les mêmes organes que la forme aiguë mais qui peut s'étendre au-delà, jusqu'à donner une atteinte multi-viscérale. Les causes les plus fréquentes de mortalité sont liées aux infections bactériennes, fongiques ou virales qui se développent avec l'immunodépression. Il existe des formes plus ou moins sévères de GvH chronique dont on distingue trois types, selon l'existence ou non d'une GvH aiguë au préalable, et selon sa guérison [47].

Avec le développement des conditionnements d'intensité réduite et l'administration de plus en plus fréquente des DLI, une classification est en réflexion par un groupe de travail au *National Institute of Health* qui distinguerait :

- Les GvH aiguës classiques
- Les GvH aiguës d'apparition tardive (survenant plus de 100 jours après la greffe ou la DLI) mais similaire à la précédente sur le plan clinique
- Les GvH chroniques classiques
- Les syndrômes de chevauchement, qui regroupent les caractéristiques cliniques à la fois des GvH aiguës et chroniques.

Même dans les meilleures conditions de compatibilité, une GvH peut être constatée. L'apparition de GvH dans les greffes de moelle entre individus non apparentés 10/10 ou même entre deux individus apparentés géno-identiques a montré que de telles précautions ne suffisent pas. Les réactions du greffon contre l'hôte décrites dans ces situations, ont permis de penser qu'il existe d'autres antigènes impliqués dans l'allo-réactivité que les molécules du CMH. Les capacités de stimulation allogénique du CMH ne proviennent pas nécessairement de la molécule HLA elle-même, dont les motifs polymorphiques sont rendus relativement inaccessibles par le peptide occupant la cavité, mais seraient plutôt la conséquence du polymorphisme du peptide lui-même [46]. Ces peptides sont issus de la dégradation de protéines qu'on appelle **antigènes mineurs**. Comme nous l'avons vu, ils correspondent à des protéines intracellulaires qui possèdent un fort polymorphisme et qui ne sont pas codées par les gènes du CMH. Cependant leur présentation est restreinte par celui-ci. Les peptides qui dérivent de la dégradation de l'antigène mineur, sont apprêtés par des molécules HLA, et peuvent être à l'origine de la reconnaissance d'un motif étranger par les lymphocytes T, et des dégâts qui en découlent. Ces protéines sont codées par des allèles situés sur différents chromosomes, autosomiques et gonosomiques. On ne peut échapper à cette différence d'antigènes mineurs que dans le cas des greffes entre jumeaux monozygotes.

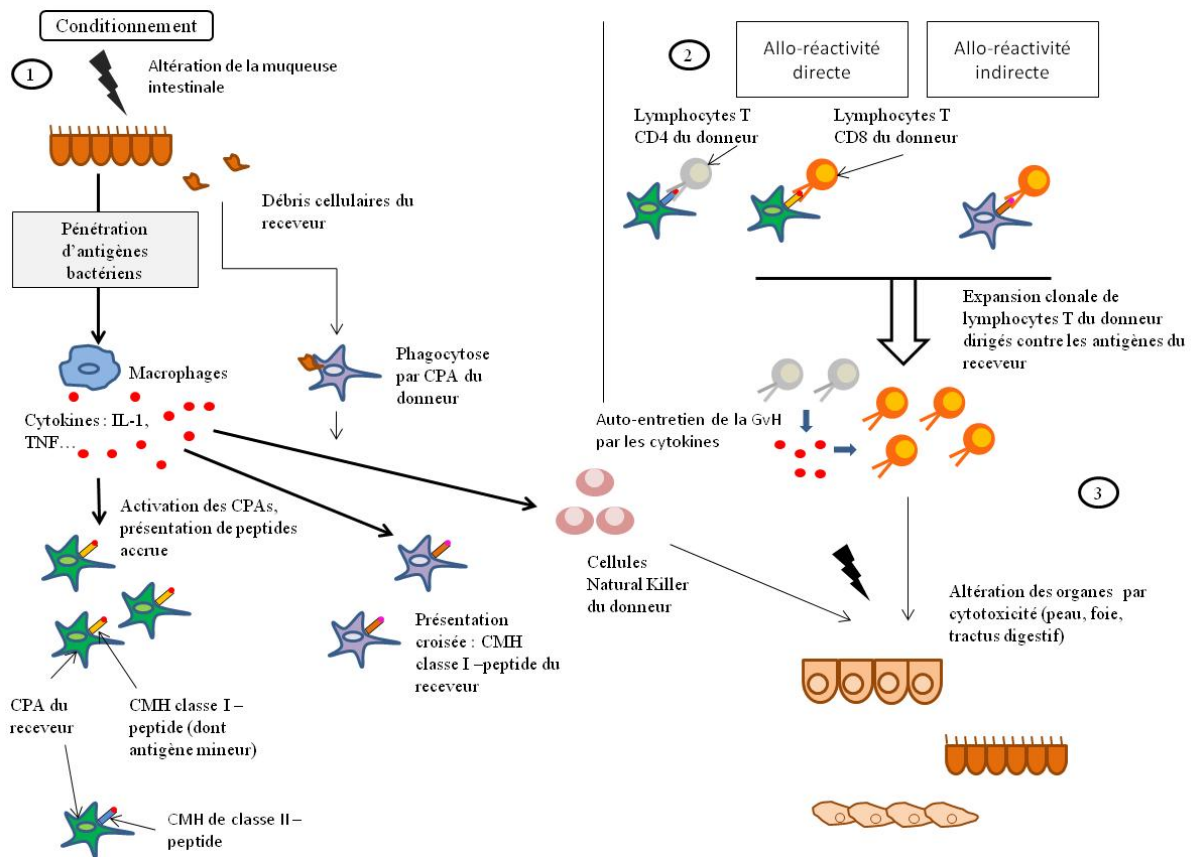
Le meilleur exemple d'antigène mineur est celui de l'antigène H-Y (Human Y chromosome antigen), codé par un gène situé sur le chromosome Y, il n'est donc retrouvé que chez les hommes. Son expression est ubiquitaire dans l'organisme. Il permet de comprendre l'importance de ne pas greffer un receveur masculin avec un donneur de sexe féminin, le greffon risquant fortement de réagir avec cet antigène inconnu. Il existe également d'autres antigènes tels que HA1, HA2 et HA5 (Human Antigen 1, 2 et 5), qui sont codés par des

autosomes. Ces derniers ont une expression restreinte aux cellules hématopoïétiques, HA1 n'est retrouvé que sur les cellules de la lignée lymphocytaire B [46][42].

D'autres systèmes comme celui des récepteurs KIR présentent un polymorphisme compté parmi celui des antigènes mineurs [50]. D'après certaines études, l'incompatibilité des systèmes KIR entre donneur et receveur serait favorable à une diminution du risque de GvH, et à un effet GvL très satisfaisant, mais ces conclusions semblent limitées aux leucémies de type myéloïde. Il semblerait intéressant également d'étudier les incompatibilités KIR-ligand, qui pourraient être associées à une augmentation de la mortalité post-greffe. Les cellules Natural Killer interagissent par l'intermédiaire de leur récepteur KIR avec les molécules HLA de classe I, si cette dernière est identifiée comme appartenant au soi, elle inhibe la cytotoxicité de la cellule NK. En revanche si la molécule HLA de classe I n'est pas reconnue, ou bien est absente (comme c'est le cas sur des cellules tumorales en majorité), la destruction cellulaire est immédiate. Contrairement aux lymphocytes T, les cellules NK agissent directement, par libération de perforines et granzymes, sans passer par une phase d'activation et d'expansion. Les cellules Natural Killer ont donc un rôle plus important dans l'effet GvL, tandis que les lymphocytes T interviennent majoritairement dans la GvH [42].

D'autres facteurs interviennent dans les risques de survenue de GvH, et notamment la variabilité d'expression des gènes contrôlant la synthèse des cytokines. En effet la réponse face à un allo-antigène est variable d'un individu à l'autre et entraîne une production plus ou moins élevée de cytokines, donc un risque génétique différent de rejet et de GvH qu'il pourrait être nécessaire de considérer lors d'une greffe [51] [52]. Les gènes codant pour les nombreuses cytokines sont répartis sur divers chromosomes, par exemple le gène de l'interféron gamma (IFN $\gamma$ ) est localisé sur le chromosome 12, celui de l'interleukine 1 (IL1) est situé sur le chromosome 2. Le polymorphisme serait plutôt retrouvé au niveau des régions 5' et 3' régulatrices d'expression, entraînant des modifications de structure au niveau des sites de liaison des promoteurs. La différence d'expression et de production de cytokines aurait une influence importante dans l'intensité de «l'orage cytokinique» qui constitue la première phase dans la survenue des GvH. Il permet d'expliquer en partie que les GvH existent dans des proportions variées (de 30 à 80%) dans les cas de greffes géno-identiques avec donneur intrafamilial [42] [53].

Le mécanisme physiopathologique de la GvH est complexe et reste mal connu, en particulier celui des GvH chroniques. Dans la forme aiguë, trois phases sont décrites pour tenter d'expliquer les phénomènes imbriqués dans cette réaction (**Figure 8**).



**Figure 8: Schéma du processus physiopathologique de la GvH.**

1 : Dans un premier temps la fragilisation des différents tissus et les apoptoses cellulaires, qui résultent du conditionnement ou des diverses chimiothérapies précédant la greffe, déclenchent la synthèse de cytokines pro-inflammatoires de type  $TNF\ \alpha$ , interleukine 1 (IL1). On parle «d'orage cytokinique» [42]. Il est entretenu par l'endommagement de la muqueuse intestinale qui laisse pénétrer plus facilement les bactéries et contribue à la stimulation du système immunitaire. Les CPA résiduelles de l'hôte et celles du donneur sont alors fortement stimulées.

2 : L'allo-réactivité entre en jeu, lorsque les lymphocytes T du donneur rencontrent les molécules HLA des tissus sains du receveur présentant des peptides étrangers au donneur issus des molécules HLA ou des antigènes mineurs du receveur.

3 : La troisième phase, marquée par une dégradation cellulaire affectant de nombreux organes, commence avec l'action des lymphocytes T cytotoxiques. Elle est amplifiée et auto-entretenu par l'action des cytokines pro-inflammatoires qui permettent, en outre, le recrutement des cellules Natural Killer et des macrophages [54] [55].

Ces allo-réactions sont à redouter lors de toute transfusion de leucocytes à un patient immunodéprimé. Les lymphocytes T du donneur immunocompétent risquent d'entraîner une réaction cytotoxique vis-à-vis du receveur affaibli. Depuis Avril 1998, tous les produits sanguins labiles sont déleucocytés par filtration afin de limiter les risques de transmission de germe intra-leucocytaire mais aussi pour diminuer le risque d'apparition d'allo-immunisation HLA. De plus, dans certaines indications bien définies, en particulier les patients immunodéprimés ayant subi une greffe de CSH, les produits sanguins labiles subissent une irradiation, permettant de limiter le risque de GvH en empêchant la réplication des lymphocytes T résiduels éventuels [56].

Dans la GvH chronique, on suppose plutôt une mise en jeu de mécanismes auto-immuns par l'action de lymphocytes T auto-réactifs du donneur. Ils ne seraient plus sélectionnés efficacement par le thymus, endommagé par le conditionnement. Des auto-anticorps dirigés contre les antigènes du receveur seraient aussi responsables de ce processus auto-immun. La GvH aiguë, l'âge du patient, l'intensité du conditionnement, sont autant de facteurs qui favorisent le développement de GvH chronique [47].

La physiologie de l'effet GvL est similaire à celui de la GvH, mais cette fois les cellules ciblées sont les cellules tumorales du receveur. La présentation d'antigènes directement par ces dernières aux lymphocytes T du donneur déclencherait une expansion clonale spécifique anti-tumorale [42].

Après la greffe, la reconstruction d'un système immunitaire complètement fonctionnel est longue, pouvant aller jusqu'à des mois ou des années. Ce délai dépend à la fois des conditions de compatibilité, mais encore de bien d'autres paramètres comme les statuts viraux du couple donneur-receveur. Le répertoire des lymphocytes T post greffe est constitué en majorité de lymphocytes matures du donneur, effecteurs et mémoires, il y a en revanche un faible nombre de lymphocytes T naïfs. Ils ont subi leur maturation thymique chez le donneur. Les cellules mémoires, selon le statut du donneur, ont de grandes chances d'être compétentes pour détecter les antigènes des principaux virus comme le *Cytomegalo virus* (CMV), l'*Epstein barr virus* (EBV), le *Varicel Zoster virus* (VZV) et l'*Herpes simplex virus* (HSV), permettant la lutte contre les infections (ou *Graft versus Infection*) chez le receveur immunodéprimé. Au cours de la prise de greffe, les précurseurs T synthétisés à partir des cellules souches du donneur subiront une éducation et une maturation dans le thymus de leur hôte. Ils accéderont aux mêmes sélections positive et négative, mais à partir cette fois de peptides de leur « nouveau soi ». Chez les patients âgés, le thymus demeure souvent endommagé et incomplètement fonctionnel; le conditionnement y contribue fortement. La reconstitution immunitaire sera plus lente [42]. Ceci explique en partie la plus grande proportion de GvH chronique dans cette tranche d'âge, découlant d'une mauvaise sélection négative des lymphocytes T, qui tendent à être libérés dans la circulation même s'ils sont autoréactifs. La formation d'un nouveau répertoire peut attendre plusieurs années post-greffe. Un conditionnement de forte intensité favorise davantage une GvH par la détérioration thymique qu'il engendre.

Bien que le typage des antigènes majeurs d'histocompatibilité reste une première étape primordiale, le développement des techniques de séquençage conduit vers l'étude de

nouveaux polymorphismes encore peu explorés. Ils semblent pourtant avoir leur importance dans les mécanismes délétères et bénéfiques entourant la greffe de cellules souches hématopoïétiques. Le développement de ces techniques montre encore une fois un intérêt dans la prise en charge et un plus grand espoir dans la survie en post-greffe à long terme [57].

### 3. Historique de la greffe de CSH

Les premiers essais de greffe de moelle allogénique ont été réalisés sur des souris leucémiques qu'on soumettait à de puissantes irradiations. Dès ces premières expériences, la réaction du greffon contre l'hôte ne se fit pas attendre, mais l'idée d'une réaction du greffon contre la leucémie est également évoquée avec la mise en jeu de réactions immunitaires allogéniques. Des essais sur les chiens ont permis une exploration plus ample des paramètres à contrôler avant la greffe [58].

Les premières greffes humaines de moelle osseuse ont été réalisées dans les années 1960 entre des jumeaux monozygotes après une irradiation mettant le receveur en aplasie sévère [37]. C'est une dizaine d'années plus tard que les greffes entre membres d'une même fratrie, non jumeaux mais HLA-identiques, commencent à se développer. Les typages HLA sont alors réalisés uniquement par sérologie et les difficultés d'étude de compatibilité limitent les greffes aux donneurs apparentés. Le développement des médicaments immunosuppresseurs permettra ensuite l'amélioration des possibilités de greffe. Les indications se sont également élargies à des leucémies en première rémission et non plus en dernière intention après plusieurs rémissions.

En 1971, l'histoire d'un petit garçon atteint du syndrome de Wiskott-Aldrich, Anthony Nolan, va permettre une avancée importante. Ses seules chances de survie sont alors la greffe de moelle osseuse, pourtant les espoirs sont pauvres lorsque sa famille apprend l'absence de donneur compatible dans la fratrie. A cette époque, aucune greffe de moelle avec donneur non apparenté n'avait atteint son but et il n'existait aucune aide pour rechercher un potentiel donneur compatible. Il faut attendre 1973 pour voir aboutir une 1<sup>ière</sup> greffe HLA identique avec donneur non apparenté [59]. A l'époque, les compatibilités étaient étudiées sur les loci HLA-A, B et DRB1 par les techniques de sérologie et de culture lymphocytaire mixte [60]. Cet événement fit naître un nouvel espoir pour la maman d'Anthony, qui fût à l'origine, en 1974, de la mise en place du premier registre de donneurs volontaires de moelle osseuse en Angleterre : le registre « Anthony Nolan ». Les fonds récoltés par les campagnes grâce à l'énergie des familles ont permis l'ouverture d'un premier laboratoire de typage qui alimentait cette base de données. Malheureusement, Anthony Nolan est décédé à l'âge de huit ans sans pouvoir bénéficier d'une greffe.

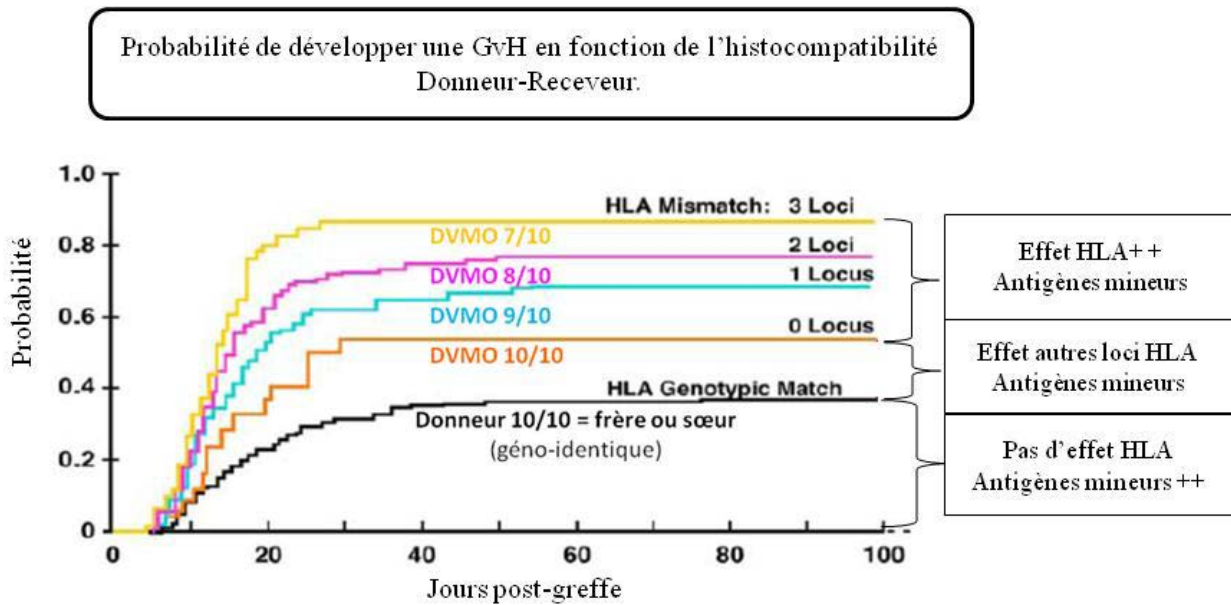
En 1986-1987, l'association France Greffe de Moelle permet la création du registre français de donneurs volontaires de moelle osseuse pendant qu'au même moment est ouvert le registre américain NMDP (*National Marrow Donnor Program*) [61].

Progressivement, l'extension des registres, parallèlement au développement de la biologie moléculaire dans les techniques de typage HLA, permet d'augmenter la qualité des greffes avec donneurs non apparentés.

#### 4. Choix d'un donneur

Chez un patient nécessitant une greffe de cellules souches hématopoïétiques, la recherche d'un donneur compatible commence par l'étude de la fratrie. En effet, comme on l'a vu précédemment, le patient receveur possède une chance sur quatre d'avoir un frère ou une sœur qui soit HLA géno-identique. C'est le cas d'environ un tiers des patients en France [39].

S'il ne possède aucune fratrie, ou si aucune personne de la fratrie n'est compatible, la deuxième solution est la recherche d'un donneur non apparenté inscrit sur les registres internationaux de **Donneur Volontaire de Moelle Osseuse (DVMO)**. Il s'agit d'un registre international qui regroupe, comme son nom l'indique, des donneurs volontaires, anonymes et non rémunérés, qui peuvent être choisis à tout moment (du début de leur inscription jusqu'à leur retrait du registre à l'âge de 60 ans révolu) pour un don de CSH, à condition que leur état de santé le permette. Le patient possède en moyenne une chance sur un million de trouver un donneur compatible, mais cela dépend de son typage HLA. En France, un donneur est dit phéno-identique lorsqu'il existe une compatibilité («match») entre les deux allèles de chacun des cinq loci: HLA-A, -B, -C, -DRB1 et -DQB1, c'est pourquoi dans ce cas on parle de donneur 10/10 (sur les allèles évalués) (**Figure 9**). Dans le cas où la greffe est le dernier recours, une greffe avec un donneur 9/10 peut être envisagée, c'est-à-dire où l'un des dix allèles est incompatible, on dit alors qu'il y a un «mismatch». Cependant, le gold standard international actuel pour les donneurs non apparentés est une compatibilité 8/8 allélique entre les loci HLA-A, -B, -C, et -DRB1 [50]. Dans certains cas, lorsque le patient possède un typage HLA relativement «fréquent» et que plusieurs donneurs 10/10 s'offrent à lui, d'autres critères entrent alors en jeu pour sélectionner le donneur qui aura le plus faible risque de provoquer une GvH. Un grand nombre de ces critères sont en cours d'étude, car ils sont encore loin d'être identiques d'un centre greffeur à l'autre. L'idéal serait d'établir des algorithmes décisionnels au niveau national voire international, pour faciliter la stratégie de recherche du donneur ce que tentent actuellement de réaliser des groupes de travail, notamment en France [62].



**Figure 9 : Graphique représentant la probabilité d'avoir une GvH dans un certain délai selon la compatibilité choisie.**

D'après Hansen et al. [63].

Dans le cas de donneurs non apparentés, compatibles 10/10, on peut supposer l'intervention des autres loci HLA non typés (voire HLA-E, -G,...).

Globalement, les lymphocytes T du donneur sont exposés à rencontrer quatre types d'antigènes sur les cellules du receveur, qui sont susceptibles de les stimuler. Les trois premiers types d'antigènes responsables d'une GvH et/ou d'une GvL communs aux greffes phéno et géno-identiques sont **les antigènes mineurs d'histocompatibilité, les protéines non polymorphiques de l'organisme qui seraient exprimées de façon aberrante** par une cellule tumorale, et enfin des **antigènes spécifiques aux cellules tumorales**. A cela s'ajoutent les **antigènes majeurs d'histocompatibilité incompatibles** dans le cas des greffes phéno-identiques (représentés par DPB1 dans 80% des cas) [42].

Parmi les critères de choix, on considère parfois un locus supplémentaire. Ainsi on réalise une extension du typage sur le locus DPB1 pour obtenir une éventuelle compatibilité 12/12. Aujourd'hui de nombreuses études tendent à évaluer l'importance des compatibilités de chaque locus, et dans le cas d'une incompatibilité, estimer l'importance de son caractère antigénique (deux digits) ou allélique (quatre digits) [50].

Toutes les études statistiques menées *a posteriori* aboutissent à la même conclusion concernant les greffes avec donneurs non apparentés présentant des incompatibilités 6/8 puis 7/8 contre 8/8. «Elles sont associées à une augmentation significative de la survenue de GvH de haut grade (IV), de GvH chronique, de mortalité liée à la greffe et de mortalité globale» [64]. Une incompatibilité au niveau d'un des loci HLA-A, -B, -C, ou -DRB1 augmentent considérablement le risque de GvH, il ne faut donc pas déroger à cette règle [65]. Toutefois, dans le cas où un sacrifice doit être fait sur la compatibilité d'un des loci A, B, C ou DRB1, il paraît préférable de prioriser le locus -C puis A. En revanche, l'incompatibilité ne doit pas

porter sur le locus DRB1 car elle engendrerait un plus grand risque de GvH. Une incompatibilité sur le locus B semble associée à une augmentation significative de la sévérité des GvH par rapport à celle du locus C [64] [66]. Le problème du fort déséquilibre de liaison entre les loci B et C se pose également : un mismatch B a de fort risque d'être accompagné d'un mismatch C. Une incompatibilité sur le locus DQB1 semble avoir une incidence mineure sur la survie post-greffe en l'absence d'incompatibilité aux quatre loci A, B, C et DRB1, en revanche si elle s'ajoute à l'une d'elles, on observe une augmentation de la mortalité [50] [67].

Les **incompatibilités HLA-DPB1** semblent moins évidentes à évaluer, les études menées à partir de couples 10/10 et compatibles ou non sur le locus DPB1 (11/12, 10/12) montrent qu'elles peuvent être séparées en deux groupes selon le «*T cell epitope*» (séquence protéique reconnue par le TCR des cellules T), celles qui apparaissent comme tolérées (permissives) et celles qui sont à l'origine d'une augmentation de GvH (non permissives). L'appartenance à un groupe est liée à un risque plus ou moins fort d'immunogénicité. Ce risque peut être calculé à partir des typages HLA-DPB1 du couple donneur-receveur grâce à un outil mis à disposition sur le site <http://www.ebi.ac.uk/cgi-bin/ipd/imgt/hla/dpb.cgi>. Ainsi, il faut privilégier les donneurs n'ayant aucun mismatch DPB1 ou ayant un mismatch permissif [64] [65].

Les comparaisons portant sur le niveau d'incompatibilité, **allélique contre antigénique**, n'aboutissent pas à des résultats homogènes. Certaines tendent à conclure que seule la nature des mismatches portant sur le locus C montre une différence d'impact, un mismatch antigénique (plutôt qu'allélique) serait fortement délétère [50]. D'autres démontrent que les incompatibilités antigéniques sur le locus B engendreraient des risques de GvH plus sévères (grade IV vs II) [64].

Une étude rapporte que la présence d'un mismatch aurait plus d'impact sur la survie globale dans le cadre du traitement d'une hémopathie à un stade encore peu évolué, que dans les hémopathies à un stade avancé (après une deuxième rémission) [50] [60][59][58].

Dans le cas de l'utilisation d'unités de sang placentaire, les critères sont moins stricts, et s'arrêtent à la compatibilité 6/6. On s'intéresse en premier lieu aux loci HLA-A, -B au niveau antigénique, puis -DRB1 au niveau allélique [68]. Elles sont utilisées en dernier recours chez les adultes [39] [69]. Toutefois, il semble que l'absence de mismatch sur le locus C soit associée à un moindre risque de mortalité post-greffe, c'est pourquoi certains praticiens demandent le typage de ce locus [70].

La **compatibilité ABO** n'est pas un critère aussi important qu'en greffe d'organe, puisque le receveur adoptera le groupe sanguin du donneur, les cellules souches hématopoïétiques greffées étant à l'origine de la synthèse érythrocytaire. L'incompatibilité ABO dans une greffe avec donneur non apparenté ne semble pas influencer la survenue de GvH, ni les rechutes post-greffe [64]. Cependant, il y a des dispositions à prendre lors d'une telle incompatibilité, le pré-traitement du greffon doit inclure une déplasmatisation et/ou une desérythrocytation.

La **source des cellules souches** influence également le risque de développer une GvH aiguë : le plus fort risque est atteint avec des cellules souches périphériques tandis que les unités de sang placentaire semblent présenter le risque le plus faible en raison de la plus grande immaturité cellulaire. La moelle se situe à un niveau de risque intermédiaire [37].

Plus **l'âge du donneur** (et du receveur) sont avancés, plus le risque de GvH augmente. On évite de sélectionner un donneur entre 40 et 45 ans si plusieurs autres donneurs plus jeunes sont disponibles [50].

La **différence de sexe** entre donneur et receveur est un facteur de risque supplémentaire de GvH si un donneur féminin est choisi pour un receveur masculin. Comme nous l'avons vu précédemment, la présence de l'antigène H-Y chez le receveur est un facteur de risque accru de GvH. Un receveur masculin doit absolument faire rechercher un donneur masculin, en revanche pour un receveur féminin, les donneurs des deux sexes sont possibles [65].

Les **donneuses nullipares** sont choisies préférentiellement aux multipares qui ont pu s'immuniser contre des antigènes d'histocompatibilité paternels au cours de leurs grossesses.

Le **statut CMV** est également pris en considération lors de la sélection d'un donneur, le risque de réactivation ou de primo-infection à cytomégalovirus chez un sujet immunodéprimé pouvant être létal. Un receveur CMV négatif doit être greffé de préférence avec un donneur CMV négatif pour ne pas risquer d'infection. Dans le cas d'un receveur CMV positif, il doit recevoir un greffon du même statut (CMV positif), qui apporte ainsi des cellules immunocompétentes vis-à-vis du virus présent dans l'organisme greffé et évite sa réactivation [71].

Le lancement d'une recherche de DVMO sur le registre, appelée «interrogation de fichier», présente un coût important pouvant aller jusqu'à plus de 2000€ par patient. Elle s'effectue *via* un logiciel appelé Syrenad, et permet l'interrogation du registre « France greffe de moelle » mais aussi celle d'autres registres internationaux représentant plus de 80 % de la totalité des DVMO inscrits dans le monde. Des logiciels statistiques gratuits comme «Easy match» ou «Haplostat» permettent d'estimer le nombre de donneurs 10/10 ou 9/10 potentiels disponibles sur les registres internationaux pour un patient de typage donné, avant de lancer une interrogation. Dans les jours qui suivent le déclenchement d'une interrogation, les différents registres nationaux et internationaux proposent les donneurs dont le typage se rapproche au maximum de celui du patient. La difficulté vient du fait que les typages de DVMO à leur inscription diffèrent d'un registre à l'autre. Par exemple, en France, les donneurs sont typés en haute résolution sur les quatre loci d'intérêt majeur HLA-A, -B, -C et -DRB1, on ne type pas DQB1 contrairement à l'Allemagne par exemple. Sur d'autres registres, seuls les loci A, B et DRB1 sont typés et parfois en résolution deux digits. Ce manque d'homogénéité conduit à la nécessité de demander des typages complémentaires avant de pouvoir recruter un donneur. Ces typages complémentaires augmentent les délais pré greffe de plusieurs semaines et le coût d'une recherche. Avec le développement de nouvelles technologies comme le séquençage à haut débit permettant des typages quasi alléliques, l'inscription et la recherche des donneurs devraient être considérablement simplifiées.

Actuellement, environ un tiers des patients n'ont aucun donneur compatible 10/10, apparenté ou non.

On recense 22.9 millions de DVMO inscrits sur l'ensemble des registres, et 606 554 USP enregistrés sur l'ensemble des banques internationales. On comptait 221 460 DVMO inscrits sur le registre français au 31/12/2013, et 31 230 USP (données les plus récentes) [40]. En juin 2015, un courrier de l'agence de biomédecine nous annonce que l'on atteint les **240 000 DVMO** et nous rappelle l'objectif de l'année 2015, qui vise l'inscription de plus de 20 000 nouveaux donneurs.

### III. Intérêt du séquençage dans la greffe de moelle

#### 1. Typage HLA à l'heure actuelle : sérologie, PCR-SSP et PCR-SSO

Les techniques de typage dont dispose actuellement la très grande majorité des laboratoires d'histocompatibilité sont des techniques qui utilisent les principes d'une part de sérologie, avec le typage par **lymphocytotoxicité**, et d'autre part de biologie moléculaire avec deux techniques faisant appel à la polymérisation en chaîne *Polymerase Chain reaction* (PCR), *Single Specific Primer* (SSP) et *Sequence Specific Oligonucleotide probes* (SSO).

#### Lymphocytotoxicité

Le principe de **lymphocytotoxicité** est basé sur la détection des antigènes HLA présents sur les lymphocytes (et monocytes) du sang circulant. Cette technique de sérologie nécessite un prélèvement sanguin sur tube ACD (Acide acétique, Citrate de sodium, Dextrose), qui assure une bonne conservation des cellules, puis un isolement de la couche de cellules mononucléées après centrifugation en gradient de densité Ficoll-Isopaque. Le principe est donc de mettre en contact les cellules cibles (lymphocytes T ou lymphocytes B selon qu'on réalise un typage HLA de classe I ou classe II respectivement) avec des anticorps anti-HLA de spécificité connue, cytotoxiques en présence de complément de lapin, et préalablement déposés dans la microplaque. Si les anticorps se fixent sur les antigènes de la surface cellulaire, le complément est spécifiquement activé. Il en résulte une lésion membranaire irréversible entraînant la lyse cellulaire, elle est visualisée par addition du colorant AO/EB (Acridine Orange/Ethidium Bromide). La proportion de cellules mortes (noyau coloré en orange par l'EB) et de cellules vivantes (colorées en vert par l'AO) est évaluée au microscope inversé à fluorescence. Cette technique possède l'avantage de lever les ambiguïtés d'allèles

nuls obtenus avec une technique de biologie moléculaire pour les loci HLA-A, -B, DRB1 et DQB1. Les typages HLA-C, DPB1, DQA1 et DPA1 ne peuvent pas être réalisés par lymphocytotoxicité. Cette technique est peu coûteuse mais ne permet que la réalisation de typage de faible résolution et est de moins en moins utilisée.

### PCR-SSP

De manière générale, la biologie moléculaire est réalisée à partir d'un échantillon sanguin prélevé sur tube EDTA à partir duquel on extrait l'ADN.

La technique de **PCR SSP**, développée en 1992, utilise une méthode d'amplification dont les couples d'amorces sont complémentaires et spécifiques d'allèles (kit de typage allélique) ou groupe d'allèles (kit de typage générique) selon la résolution du kit utilisé. Si l'amplification a lieu, elle témoigne de la présence de l'allèle ou groupe d'allèles ciblé par les amorces. Pour cela, les produits d'amplification subissent ensuite une migration dépendante de leur poids moléculaire sur des gels d'agarose ou *via* une électrophorèse capillaire. Dans le premier cas, les bandes sont visualisées sur le gel sous rayons ultra-violet grâce à une coloration de l'ADN par un agent intercalant (**Figure 10**). La combinaison de l'ensemble des puits positifs (où apparaissent une migration de bande de taille spécifique) permet d'assigner un allèle, avec des ambiguïtés alléliques de fréquences rares associées le plus souvent [72] [73].

Cette technique est la plus utilisée actuellement en histocompatibilité dans le contexte urgent des typages en faible résolution des donneurs d'organes cadavériques. En revanche son application au typage en haute résolution des DVMO reste inadéquate, car inadaptée à la réalisation de séries. Elle permet le typage d'un seul DVMO (typage des loci HLA-A, -B, -C et -DRB quasi allélique) en 48 heures [74]. Son interprétation biologique est très longue, entre une heure et une heure trente par patient dans le cas d'une haute résolution. Elle possède cependant l'avantage d'éviter les phénomènes de *drop out*, problème dont on reparlera à la fin de ce chapitre. En quelques mots, le *drop out* correspond à l'amplification préférentielle d'un allèle au détriment du second. Il est redouté en histocompatibilité puisqu'il peut amener à l'assignation d'un résultat homozygote, chez un individu qui est en fait hétérozygote. Une technique utilisant un grand nombre d'amorces présente ainsi moins de risques d'amplifications préférentielles. En revanche la PCR SSP ne permet pas la définition de nouveaux allèles et n'aboutit pas (ou très rarement) à l'assignation d'allèles en 4 digits, et laisse fréquemment des ambiguïtés à lever. Pour cette raison, elle est utilisée comme complément et parallèlement à la PCR SSO afin d'éliminer un maximum d'ambiguïtés.

## PCR-SSO

La **PCR SSO** a vu une grande évolution depuis ses débuts et est désormais couplée à la technologie de cytométrie en flux Luminex® (**Figure 10**). Cette technique est une technique de *reverse SSO* en phase liquide.

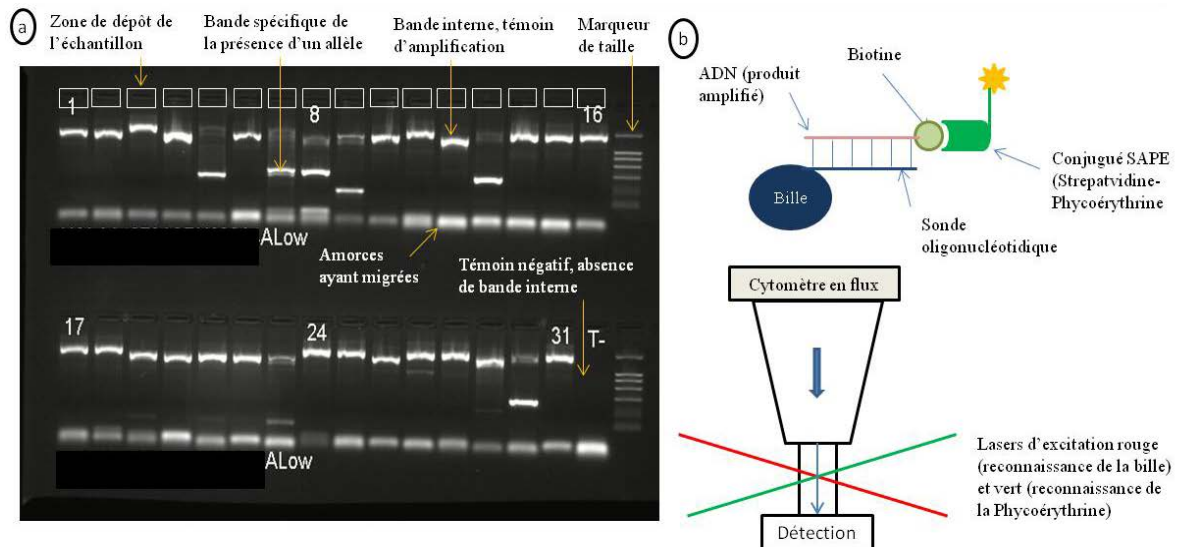
Dans une première étape, une amplification spécifique de locus est réalisée, en utilisant des amorces biotinylées qui s'hybrident dans les régions non polymorphiques du gène d'intérêt. Les produits d'amplification marqués par la biotine sont ensuite mis en incubation en présence d'un mélange de billes (microsphères en polystyrène, de l'ordre du micromètre de diamètre), chacune couplée à un court fragment nucléotidique unique, appelé sonde. Ces sondes vont alors s'hybrider, ou non, selon leur complémentarité à la séquence de l'allèle présent. Après cette étape, la streptavidine couplée à une molécule fluorescente (la phycoérythrine) est ajoutée au milieu et se lie aux éventuels hybrides biotinylés précédemment formés. Le mélange de billes passe ensuite dans le cytomètre en flux. Cet instrument est capable d'exciter des molécules fluorescentes par différents lasers et de détecter la fluorescence émise. De ce fait, il peut identifier individuellement plus d'une centaine de billes différant par leur contenu en fluorescence rouge et infra-rouge (1<sup>ier</sup> laser). L'hybridation des oligonucléotides avec les amplicons est révélée par la phycoérythrine (2<sup>ème</sup> laser). Un profil de billes positives (hybridées) et négatives (non hybridées) est finalement obtenu, et comparé à une base de données intégrée au logiciel de l'automate, proposant l'assignation des allèles de chaque locus [73].

Brièvement, il existe d'autres techniques de PCR-SSO, où les produits d'amplification sont immobilisés sur des membranes de nitrocellulose et mis en présence de sondes d'oligonucléotides pour permettre une éventuelle hybridation. La technique directe n'est actuellement plus utilisée car elle nécessitait une membrane par type de sonde. Chaque sonde s'hybridait à une température propre ce qui compliquait la manipulation. Les radio-isotopes utilisés initialement pour la révélation, furent ensuite remplacés par des réactions de chimiluminescence, moins toxiques. Aujourd'hui l'uniformisation des températures d'hybridation a considérablement simplifiée la manipulation, qui reste pourtant encore laborieuse.

Cette technique permet le typage d'un DVMO (typage des loci HLA-A, -B, -C et –DRB1) en résolution intermédiaire en 24 heures. Contrairement à la PCR-SSP, elle est adaptée à la réalisation de série de 20 patients (4 loci) simultanément. Son interprétation biologique est relativement rapide, environ trois heures pour 20 patients. De plus, elle permet la détection d'allèles rares avec une bonne spécificité, cependant elle n'est pas utilisable pour la définition de nouveaux allèles. Les ambiguïtés à lever sont fréquentes et amènent à l'utilisation parallèle de la PCR-SSP.

Une des difficultés des méthodes de typage réside dans la détermination des positions cis/trans d'un polymorphisme détecté à un locus donné, c'est-à-dire déterminer sur lequel des deux allèles du locus il se trouve. Dans le cas où cette réponse n'est pas apportée, il en résulte

un plus grand nombre d'ambiguïtés, parfois génériques. Ce problème de *phasing* est illustré dans le paragraphe III. 3 « technique Sanger ». Il est résolu dans le cas de la PCR-SSP, qui couvre de plus longs fragments grâce à ses paires d'amorces spécifiques d'allèles, contrairement aux sondes oligonucléotidiques de la SSO qui sont trop courtes pour faire le lien entre les polymorphismes. Leur élongation a permis de réduire en partie ce problème [67]. On verra que cette difficulté se retrouve dans le cas du séquençage bi-allélique de type Sanger, ce qui confère l'avantage au séquençage mono-allélique, cette notion sera expliquée dans le chapitre portant sur cette technique [75].



**Figure 10 : Présentation des principes de la PCR-SSP et de la PCR-SSO**

- a) Exemple de photo d'un gel d'agarose obtenu pour l'interprétation d'un typage basse résolution du locus A, en PCR-SSP.
- b) Schéma du principe de la technologie Luminex® associée à la PCR-SSO.

Au total, ces deux approches ne permettent plus de répondre à la définition des typages « haute résolution » demandés par les standards EFI (EFI D1.320) lors du typage HLA des donneurs et receveurs de CSH. Le résultat fourni par les kits alléliques de PCR-SSP et/ou de PCR-SSO laisse un grand nombre d'alternatives alléliques et des allèles nuls au locus C qui ne peuvent être éliminés par la technique sérologique (celle-ci ne typant pas les antigènes HLA-C). Les résolutions des typages HLA par PCR-SSO et PCR-SSP restent limitées par le nombre de sondes (SSO) ou d'amorces (SSP) utilisables. La seule stratégie adéquate pour satisfaire à la définition de la haute résolution est le séquençage des gènes HLA.

## 2. Historique du séquençage

### De l'ADN à la protéine

La découverte de la molécule sur laquelle repose la vie et l'évolution a passionné de nombreux scientifiques depuis les siècles derniers. La structure hélicoïdale de l'ADN double brins, composée de ses quatre bases (Adénine, Guanine, Cytosine et Thymidine), a été déterminée en 1953 par James Watson et Francis Crick alors âgés respectivement de 24 ans et 36 ans [76]. Leurs travaux ont été confirmés par ceux de Briton Maurice et Rosalind Franklin avec la cristallographie aux rayons X et par la suite avec des études physico-chimiques ainsi que par la microscopie électronique [77] [78].

Les quatre constituants majeurs de la molécule d'ADN sont les nucléotides (désoxy «Base» Mono Phosphate : dAMP, dTMP, dGMP et dCMP). Ils sont composés d'une molécule de désoxyribose (pentose) à laquelle sont greffés une base azotée, et un acide phosphorique. Il existe deux bases puriques (Adénine A, Guanine G) et deux bases pyrimidiques (Cytosine C, Thymine T). Les liaisons entre nucléotides sont assurées par des liaisons phosphodiesters, entre le phosphate d'un nucléotide et le pentose du nucléotide précédent. Il en résulte une extrémité libre terminant par un groupement phosphate, appelée 5'P, et une autre terminant par une molécule de pentose appelée 3'OH. Le sens conventionnel de lecture d'une molécule d'ADN est de 5' en 3'. Quand Erwin Chargaff et son équipe découvrirent qu'une molécule d'ADN contenait une quantité égale d'Adénine et de Thymine d'une part et de Guanine et Cytosine d'autre part, ceci conduisit à suggérer leur appariement deux à deux [77]. Une molécule est habituellement constituée de deux brins de nucléotides, disposés en configuration hélicoïdale. Chaque paire de bases forme des liaisons hydrogènes qui assurent en partie la stabilité de l'hélice. Les paires sont toujours formées d'une Adénine associée à une Thymine, et d'une Cytosine reliée à une Guanine, de ce fait les brins sont dits **complémentaires**. Le terme **antiparallèles** qui les caractérise également signifie que lors de la lecture d'une molécule d'ADN, un brin est dans le sens 5'→3' (brin sens) et l'autre dans le sens 3'→5' (brin anti-sens ou antiparallèle).

A la fin des années 1970, deux équipes indépendantes développent des techniques de séquençage : l'anglais Frederick Sanger utilise une technique par synthèse enzymatique et «terminaison de chaîne», on parle alors de séquençage de type Sanger dit de première génération, pendant qu'aux Etats-Unis Walter Gilbert utilise la dégradation chimique [79]. Le séquençage a pour but de déterminer l'enchaînement des nucléotides qui constituent une molécule d'ADN ou d'ARN. En 1980, Frederick Sanger et Walter Gilbert sont colauréats de la moitié du prix Nobel de chimie (l'autre moitié a été remise à Paul Berg) « pour leur contribution à la détermination des séquences d'acides nucléiques » [80]. Il est alors possible de lire des séquences de 200 à 400 paires de bases [79].

Par la suite, l'établissement d'un lien entre la carte génétique d'un gène et la structure primaire de la protéine correspondante a permis la découverte du code génétique. C'est

Francis Crick qui propose une théorie sur la synthèse biphasique des protéines en 1958 : une première phase nucléaire correspondant à la **transcription** de l'information contenue dans l'ADN en une molécule servant de messenger intermédiaire ARNm (Acide ribonucléique messenger), baptisée ainsi par Jacob et Monod en 1961 [81] [77]. La deuxième phase, se déroulerait dans le cytoplasme où migre l'ARNm, pour subir la **traduction** de sa séquence en une chaîne d'acides aminés selon le code génétique. Cependant, on sait désormais qu'il existe quelques étapes supplémentaires pour obtenir une protéine fonctionnelle à partir de l'ARNm. D'ailleurs, toutes les molécules d'ARN produites ne sont pas à l'origine d'une protéine, il existe des ARN de transfert (ARNt), ribosomiaux (ARNr) ou encore *small nuclear* (ARNsn) qui participent uniquement au processus de synthèse.

Le code génétique est dit « *comma-less triplet code* » c'est-à-dire que l'unité codante, le codon, est représentée par trois nucléotides et « sans ponctuation » pour les séparer. Cette unité correspond à un acide aminé particulier de la structure primaire d'une protéine. Ce code génétique a été démontré comme étant universel c'est-à-dire identique chez toutes les espèces vivant sur terre (bactéries, virus, végétaux et animaux...). Une exception existe cependant dans le génome mitochondrial où certains codons ne codent pas pour le même acide aminé que dans le noyau cellulaire eucaryote [77].

Une découverte surprenante est venue ajouter quelques difficultés lors de l'étude des premiers ADN humains séquencés, et des structures primaires des protéines correspondantes. L'analyse des acides aminés de l'albumine permettait d'établir un enchaînement des codons traduits qui était vraisemblablement interrompu par de longues séries de codons non traduits en acides aminés. Cette constatation a amené Sharp, Broker et Roberts à en apprendre davantage sur la transcription et les phénomènes d'excision-épissage chez les organismes plus évolués. Plus précisément, les segments nucléotidiques d'un gène sont une succession d'**exons** et d'**introns**. Les deux sont transcrits en un pré-ARNm, en revanche les introns ne seront pas traduits. La molécule d'ARNm, obtenue après maturation du pré-ARNm est donc une copie des exons d'un gène mis bout à bout. La maturation consiste en une **excision** visant à enlever les introns, suivie d'un **épissage** permettant la réunion des fragments nucléotidiques restants [82].

Finalement, « seulement un à trois pour cent des séquences nucléotidiques du génome humain représentent l'information génétique codant la totalité des protéines » [83].

Les différentes étapes qui permettent d'aboutir à une protéine à partir d'un brin d'ADN sont soumises à une régulation. Contrairement à ce qui fut pensé initialement, les introns ont un rôle capital. Ils constituent notamment des sites de fixation pour les facteurs de transcription, indispensables à la régulation de l'expression du gène. Selon le taux de protéines synthétisées, des signaux transmis par la fixation de molécules à différents endroits sur l'intron engendreront l'arrêt ou au contraire favoriseront la transcription. Il existe aussi des mécanismes de régulation post-transcriptionnels, affectant l'épissage par exemple.

## Le séquençage, mode d'emploi

L'ARN diffère de l'ADN en différents points à la fois structuraux et fonctionnels : le pentose est cette fois du ribose, et la base azotée Thymine est remplacée par une base appelée Uracile qui s'apparie également avec l'Adénine. Ces appariements existent, dans le cas où la molécule se replie sur elle-même, mais la molécule d'ARN ne possède qu'un seul brin. Ces caractéristiques qui la différencient de l'ADN, font de l'ARN une molécule moins stable, dont la synthèse est également plus propice aux erreurs par les enzymes ARN-polymérase. Ceci permet de comprendre pourquoi le séquençage d'une molécule d'ARN nécessite sa conversion en ADN complémentaire (ADNc) par une transcriptase inverse. En effet, si on travaillait sur l'ARN, les enzymes chargées de sa polymérisation ne seraient pas suffisamment performantes pour corriger et limiter les risques d'erreurs, contrairement à certaines ADN-polymérases.

Dans l'organisme humain, chaque type cellulaire a ses propres fonctions et ses propres besoins en protéines selon sa spécialisation. Cependant, toutes les cellules nucléées de l'organisme renferment le même ADN, prêt à servir de support pour l'expression des protéines qui lui sont nécessaires. Ainsi, dans une cellule spécialisée, les séquences codantes ne seront pas toutes transcrites. Les gènes exprimés dans une cellule musculaire différeront de ceux exprimés dans une cellule pancréatique par exemple. Cette notion est importante pour comprendre que le séquençage de l'ARN d'une cellule ne conduira pas automatiquement au même résultat que le séquençage de son ADN. On comprend aisément que la découverte du séquençage a montré de nombreux intérêts, notamment dans la compréhension des fonctions d'un grand nombre de molécules et cellules de l'organisme, en étudiant les différences d'ARN produits d'une cellule à l'autre. La recherche de mutations, prédisposant ou affirmant le diagnostic d'une pathologie fait également l'objet d'une grande utilisation du séquençage.

**Dans le cas du typage HLA**, l'intérêt du séquençage consiste à définir les allèles des gènes codant les molécules HLA portées par un individu. C'est la technique de typage qui permet d'obtenir la meilleure résolution, en comparaison des autres techniques existantes (sérologie, *PCR-SSP*, *PCR-SSO*). Depuis son invention par Sanger, elle a subi une grande et rapide évolution, laissant apparaître une seconde et même une troisième génération.

Il a fallu attendre l'invention des premières techniques d'amplification pour bénéficier d'une quantité suffisante de matériel génétique pour permettre, par exemple, la détermination des cartes génétiques. La première méthode fût développée par Paul Berg et son équipe au début des années 1970. Elle est basée sur l'utilisation de vecteurs pour introduire un segment d'ADN d'intérêt dans un génome viral ou un plasmide bactérien par exemple, et créer ainsi une molécule d'ADN recombinante. Une fois cette dernière introduite dans une cellule bactérienne, le segment pourra être amplifié [77]. Un peu plus tard en 1983, un autre biochimiste américain, Kary Mullis, met au point le principe de réaction de polymérisation en chaîne (PCR) qui connaîtra un développement et une utilisation quasi omniprésente dans le monde de la biologie moléculaire [77]. Brièvement, elle consiste en la répétition cyclique de trois étapes marquées par des températures réactionnelles différentes (dénaturation des

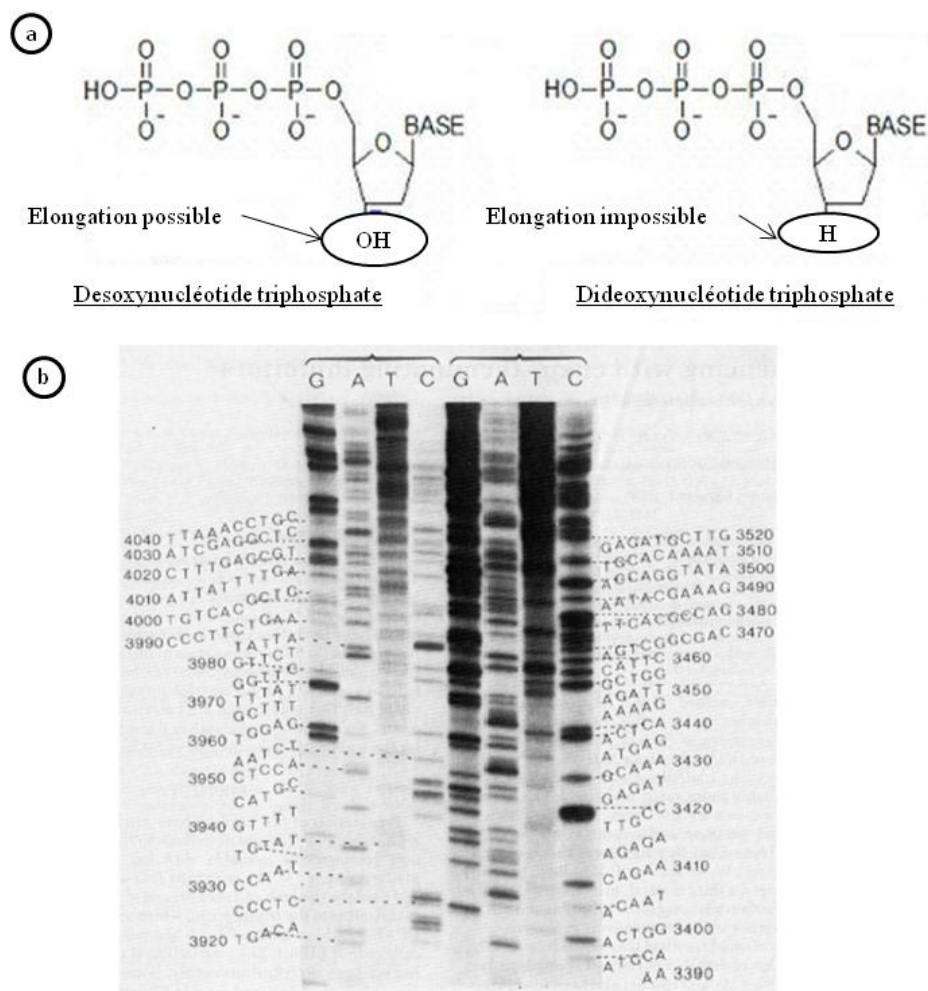
doubles brins, hybridation d'amorces et polymérisation). Ce principe donnera d'ailleurs le nom de thermocycleurs au matériel développé pour son automatisation. En plus du fragment d'ADN double brin à amplifier, il faut apporter au milieu réactionnel des désoxynucléotides triphosphates et deux amorces qui encadrent la séquence cible, et qui sont chacune complémentaires d'un des deux brins. Elles permettront l'action ciblée de l'ADN-polymérase, enzyme indispensable à la synthèse des nouveaux brins. Chaque brin néosynthétisé servira de matrice pour les réactions suivantes, ce qui multiplie les quantités de fragments de façon exponentielle. Une trentaine de cycles permet l'obtention d'une quantité d'ADN adéquate pour l'analyse de séquences. Initialement, les changements de température successifs inactivaient l'enzyme et nécessitaient son ajout à chaque cycle. La découverte en 1988 d'une ADN-polymérase thermorésistante, la **Taq** polymérase extraite d'une bactérie vivant dans les sources d'eaux chaudes (*Thermus aquaticus*), a permis l'automatisation de cette technique et un gain de temps considérable [83].

## Technique Sanger

### Principe de la méthode

La réalisation de la méthode Sanger nécessite à ses débuts l'utilisation de quatre milieux réactionnels différents, un pour chaque base, dont le volume est de l'ordre du microlitre ( $\mu\text{L}$ ) [84]. Dans chaque tube il y a l'ADN matrice à séquencer, une ADN polymérase I, une amorce complémentaire au début du segment à séquencer et un mélange de désoxynucléotides triphosphates (dNTP) et de didéoxynucléotides (ddNTP). Les ddNTP sont aussi appelés « terminateurs de chaîne », ils diffèrent des dNTP par leur extrémité 3' où la fonction hydroxyle est remplacée par un simple hydrogène (**Figure 11a**). Cette différence empêche la formation de la liaison phosphodiester permettant l'incorporation du nucléotide suivant. Au début de la technique, les ddNTP sont greffés d'un marqueur radioactif (Soufre 35, Phosphore 32...). La structure des ddNTP entraîne donc l'arrêt de la prolongation de la chaîne lorsqu'ils sont incorporés et permet la détection des fragments obtenus grâce au marquage. Une fois le mélange constitué, la Taq polymérase commence son travail de polymérisation et sélectionne aléatoirement des dNTP et ddNTP dans le milieu, qui sont en compétition. Le mélange en proportion 9/10 (dNTP/ddNTP) et l'affinité supérieure de la Taq pour les ddNTP sont tels que les ddNTP seront statistiquement intégrés à toutes les positions possibles [84]. Après de multiples réactions de polymérisation, on obtient alors dans chaque tube tous les fragments possibles de différentes tailles finissant par un ddNTP marqué. Une migration de ces fragments est réalisée sur gel, ceux de petites tailles migrent rapidement tandis que les plus grands migrent lentement. Ils sont ensuite révélés par leur marquage radioactif et l'analyse de la séquence est faite manuellement [85]. Le résultat obtenu est présenté sous la forme d'un autoradiogramme (**Figure 11b**). A l'époque, il fallait compter une journée entière pour le séquençage de 500 à 1000 nucléotides [86].

Une autre technologie a été mise au point par Maxam et Gilbert. Elle était basée sur le principe de dégradation chimique de l'ADN. Après marquage radioactif de l'extrémité 5' des ADN par du phosphore 32 ( $^{32}\text{P}$ ), la molécule était dénaturée pour obtenir de l'ADN simple brin. Puis différents traitements chimiques étaient utilisés pour réaliser des coupures sélectives selon les bases, avec un milieu réactionnel par agent [84]. L'acide formique était utilisé par exemple pour la coupure après des bases puriques et l'hydrazine ciblait les bases pyrimidiques. La séquence était établie en retrouvant l'ordre des coupures, révélée par la migration selon la taille des fragments obtenus, sur gel d'électrophorèse. Elle n'a pas eu autant de succès du fait de son efficacité modérée et de la toxicité des produits utilisés, c'est donc la méthode Sanger qui a servi de modèle par la suite pour le développement des techniques de séquençage. Des améliorations ont permis un gain de temps et une analyse simplifiée des résultats. Des marqueurs non radioactifs ont ensuite été utilisés [87].



**Figure 11 : Structure des dNTP et ddNTP et présentation d'un autoradiogramme.**

D'après F. Sanger et al. [84].

a) Différence de structure entre désoxynucléotide (dNTP) et didéoxynucléotide (ddNTP).

b) Autoradiogramme datant de 1977, obtenu sur gel de polyacrylamide. Sur les côtés, l'interprétation minutieuse de la séquence.

Ainsi en 1987, Applied Biosystems sort un premier séquenceur AB370, il utilise des didéoxynucléotides marqués par des fluorophores différents ce qui permet la réalisation des quatre réactions dans le même mélange (**Figure 12**). Ainsi la Thymine correspond à la couleur rouge, l'Adénine au vert, la Cytosine au bleu et la Guanine au noir sur les électrophorogrammes obtenus avec le séquenceur Applied Biosystem®. Chaque milieu réactionnel, destiné au séquençage d'une molécule d'ADN, est contenu dans un puits d'une microplaque, facilitant le séquençage de plusieurs échantillons simultanément. L'étape de migration sur gel a été remplacée par une migration en électrophorèse capillaire, autorisant ainsi une analyse informatisée avec le développement en parallèle de logiciels. Les fragments obtenus dans un puits migreront dans un capillaire. Ainsi c'est le nombre de capillaires qui conditionne le nombre de réactions de séquençage. Pour donner une idée du progrès, à ses débuts l'automate peut séquencer 96 paires de bases (pb) en une étape, et 500 Kb par jour. Aujourd'hui il atteint les 2.88 Mb par jour soit 5 à 6 fois plus, avec une longueur de lecture par échantillon proche des 900 pb [88].

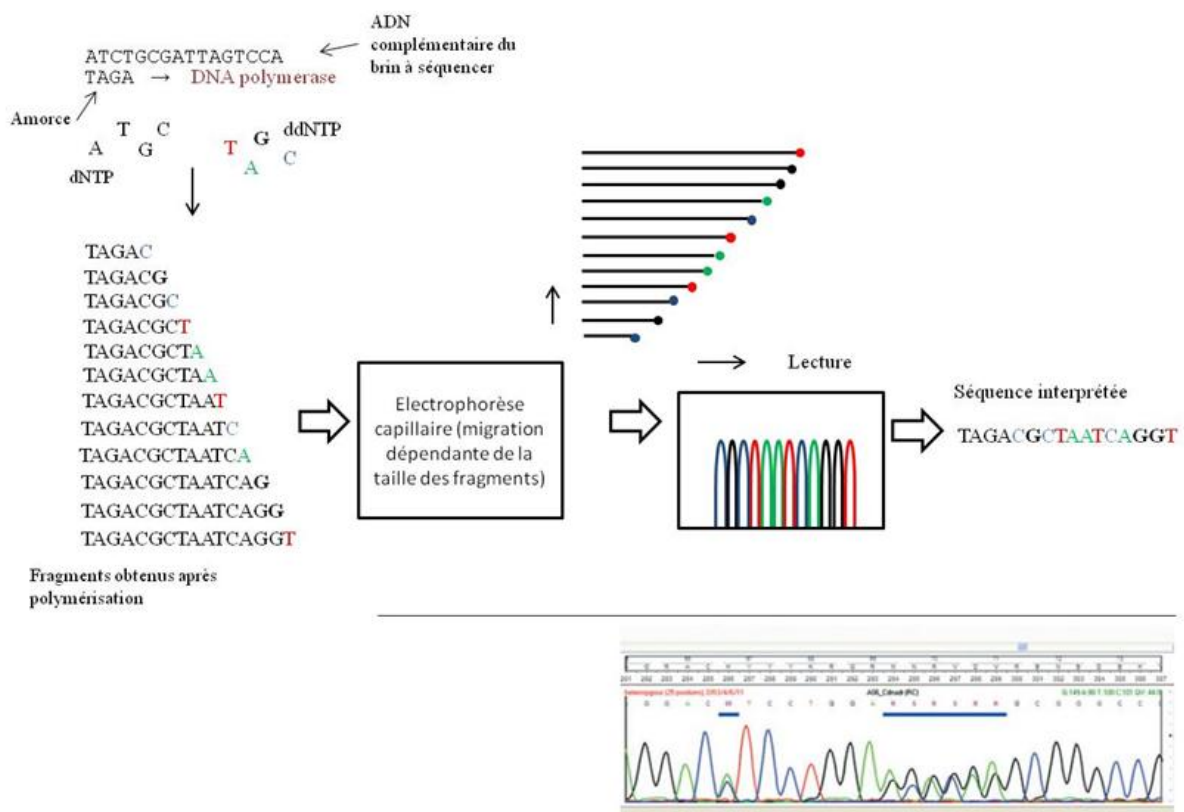
Aujourd'hui, la technologie des Big Dye Terminator ® (BDT) permet la détection des fragments dans le capillaire par la technique FRET (Transfert d'énergie par résonance). Elle permet d'améliorer la sensibilité de la méthode par augmentation du signal produit. Les ddNTP sont alors marqués par deux fluorochromes qui possèdent des longueurs d'onde d'excitation différentes, l'un jouant le rôle du donneur (fluorescéine) et l'autre d'accepteur (dichloroRhodamine). Ce dernier possède un spectre d'émission qui diffère pour chaque type de ddNTP (ddATP, ddTTP, ddGTP et ddCTP). Un laser à argon va exciter spécifiquement la fluorescéine, qui libèrera son énergie par l'émission d'une fluorescence correspondant au spectre d'excitation de la dichloroRhodamine. Le fluorochrome accepteur excité à son tour va émettre une fluorescence spécifique du ddNTP sur lequel il se trouve. Le signal est capté par une caméra *Charge Couple Device* (CCD). Le principe d'un détecteur dit CCD est basé sur la conversion d'un signal lumineux en un signal électrique.

### **Application à l'histocompatibilité**

La *Sanger Sequencing-Based Typing methode* (SSBT) est devenue une méthode de choix et de référence pour le typage HLA, permettant une haute résolution. L'étape de séparation des fragments par électrophorèse capillaire pour un exon d'environ 300 paires de bases prend environ 30 minutes [74]. L'analyse de 200 séquences (d'environ 750 pb) est possible en 30 heures sans intervention humaine, grâce à l'automatisation. La taille des capillaires utilisés dépendra de la taille même des fragments analysés, de 22 centimètres pour les petits capillaires jusqu'à 80 centimètres pour l'analyse de plus grande séquences.

Avec les automates de séquençage Sanger actuels, la collection et l'extraction automatique des données brutes issues de la lecture par la caméra permettent leur analyse par un logiciel qui doit tenir compte des différences de mobilité phorétique des fluorochromes et

des différences d'intensité de fluorescence de chaque fluorochrome accepteur. Un autre logiciel compare les séquences obtenues à une base de données (IMGT *data base*). L'exploitation des données requiert une analyse de la part de l'utilisateur qui vérifie les seuils choisis pour l'identification correcte d'une base. Le logiciel doit être capable d'identifier les hétérozygoties et attribue un code correspondant aux deux bases identifiées (par exemple, M correspond à A et C). Il est important qu'une analyse visuelle de l'électrophorégramme soit pratiquée pour détecter d'éventuelles erreurs dans les séquences, selon la taille et la pureté des pics obtenus. Le développement du séquençage occasionne des découvertes plus fréquentes de nouveaux allèles. Les logiciels d'interprétation doivent donc tenir compte de cette évolutivité constante de la base de données regroupant les allèles HLA et doivent pouvoir subir des mises à jour régulières, au moins annuelles selon les standards EFI [74].



**Figure 12 : Principe de séquençage selon la méthode Sanger.**

D'après P.P.J. Dunn [74].

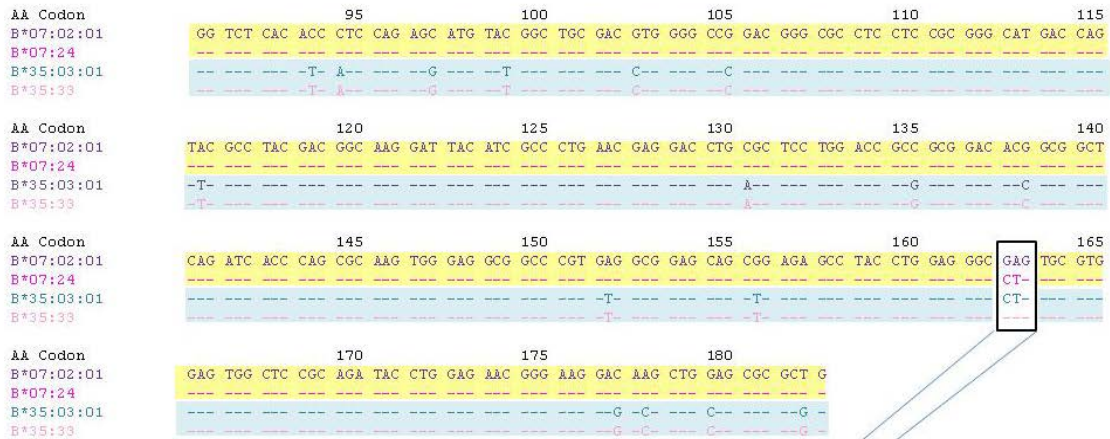
Le brin d'ADN complémentaire de celui à séquencer est mis en présence des dNTP, ddNTP et d'ADN polymérase. De nombreux fragments de taille variable sont synthétisés aléatoirement, ils sont ensuite soumis à une électrophorèse capillaire. Leur migration dépendante de la taille associée à une détection spécifique de nucléotide permet de reconstituer la séquence.

En bas à droite, exemple d'un électrophorégramme obtenu par SSBT.

Aujourd'hui, plusieurs fournisseurs proposent des kits de typage HLA adaptés aux automates de séquençage, associés à un logiciel d'aide à l'interprétation [74]. La différence majeure qui distingue les kits est la possibilité d'effectuer ou non du séquençage **mono-allélique**. Il s'agit dans ce cas de séquencer les deux allèles d'un même locus séparément et d'obtenir un résultat assignant indépendamment ces allèles. L'autre possibilité, dite de séquençage **bi-allélique**, consiste à séquencer en même temps les deux allèles d'un même gène, les résultats d'assignation dépendront alors l'un de l'autre en raison des polymorphismes de phase rencontrés (ou polymorphisme cis/trans), laissant éventuellement des ambiguïtés non résolues. Ce problème de *phasing*, évoqué précédemment, est expliqué avec un exemple concret dans la **figure 13**.

La technique du séquençage mono-allélique est basée sur une première amplification spécifique de locus, suivie d'amplifications spécifiques de groupe d'allèles ou *Group-specific amplification* (GSA) qui se déroulent séparément dans les puits d'une microplaque. Parfois plusieurs spécificités différentes appartiennent au même groupe, et sont ainsi amplifiées dans le même puits. Le contenu de ces derniers est soumis à une migration électrophorétique sur gel d'agarose, qui permet la détermination des puits où a eu lieu une amplification. Dans le cas d'un échantillon de typage totalement inconnu, il peut être soumis à toutes les amplifications par les amorces spécifiques de groupe. Si elles conduisent à l'obtention de produits d'amplification dans deux puits différents, ceci permettra le séquençage individuel des allèles. Si les deux allèles se trouvent en revanche dans le même groupe d'amplification, il n'y a pas d'autre choix que de le traiter en mode bi-allélique. Il est possible également de réaliser un typage préalable de l'échantillon en faible résolution pour cibler le ou les kits d'amorces de GSA à utiliser. Cette solution permet aussi de repérer les éventuels problèmes d'amplification préférentielle (*drop-out*), qui peuvent exister si les deux allèles sont dans le même groupe, en confirmant la nature hétérozygote d'un échantillon. Les produits amplifiés seront ensuite purifiés et mis en présence d'amorces qui serviront au séquençage. Ces amorces, de par leur spécificité, déterminent à la fois l'exon et le sens du brin séquencés. Ainsi, pour la classe I, les exons 2 et 3 sont séquencés systématiquement. Pour la classe II, seul l'exon 2 est systématique. L'exon 4 de la classe I, qui code pour la partie  $\alpha 3$  du domaine extracellulaire, et l'exon 3 des gènes DQB1 peuvent également être séquencés, ils permettent d'éliminer quelques allèles nuls distingués par ces exons [89].

Il apparaît clairement qu'il est plus difficile d'obtenir une assignation unique de deux allèles dans le cas du séquençage bi-allélique. La méthode mono-allélique présente un avantage indéniable, elle est néanmoins deux fois plus coûteuse. Seuls les kits Protrans® propose cette technique. L'obtention de séquences de qualité, avec un nombre restreint d'ambiguïtés, permet une interprétation bien plus rapide et un gain de temps considérable. Les kits de séquençage bi-allélique sont actuellement en fin de commercialisation en raison du nombre d'ambiguïtés non résolues.



2 paires de séquences similaires,  
différenciables uniquement par deux  
acides nucléiques polymorphiques situés  
à la même position pour les deux paires.

B\*070201 TGGAGGGC**G**AGTGCCTGGAG  
B\*0724 TGGAGGGC**CT**GTGCCTGGAG  
B\*3503 TGGAGGGC**CT**GTGCCTGGAG  
B\*3533 TGGAGGGC**G**AGTGCCTGGAG

**Figure 13 : Exemple de polymorphisme cis/trans.**

D'après S.D. Adams et al. [90].

D'après l'alignement des séquences des 4 allèles ci-dessus, on retrouve un polymorphisme au niveau des deux mêmes positions, qui donnent deux combinaisons possibles, chacune caractéristique d'une paire d'allèles. Si GA est attribué à l'allèle 1 (séquence jaune) et donc CT à l'allèle 2 (séquence bleue), le résultat sera alors B\*07:02, B\*35:03. Dans le cas contraire, on obtient un résultat tout à fait différent : B\*07:24, B\*35:33. L'amplification séparée des deux allèles permet d'éviter ces problèmes, à l'origine d'ambiguïtés génériques, mais aussi alléliques.

Le séquençage par technique Sanger présente à l'heure actuelle l'intérêt de révéler la totalité des polymorphismes présents sur les exons les plus polymorphes, contrairement aux méthodes vues précédemment, comme la PCR SSO qui n'assurent pas la couverture complète de ces exons avec l'ensemble des sondes fixées sur les billes. Le séquençage apporte ainsi des résultats plus précis et plus discriminants. Cette technique possède toutefois quelques limites. Les hétérozygoties qui ne peuvent pas être séparées par les kits d'amorces spécifiques de groupe restent plus compliquées à résoudre. De plus, les allèles qui diffèrent sur des nucléotides situés dans les exons non séquencés ne seront pas identifiés mais les standards EFI ne demandent pas un tel niveau de résolution à l'heure actuelle. Le résultat devant être rendu en groupe G, aucune incertitude ne doit persister concernant les exons 2 et/ou 3. Cependant les allèles nuls compris dans un groupe G doivent être éliminés, ce qui reste parfois problématique lorsque la mutation est située en dehors des exons précédemment cités et que le typage par lymphocytotoxicité ne peut pas être effectué ou n'est pas informatif. La *Next Generation Sequencing* (NGS) apporte une potentielle solution puisque le séquençage de la totalité du gène, de son extrémité 5' à 3' devient réalisable.

Le projet de séquençage du génome humain en 2001 a fortement stimulé le développement d'outils informatiques pour l'analyse des données, mais également le développement de nouveaux instruments pour diminuer le temps de main d'œuvre et le coût de l'analyse. On voit alors apparaître la *Next Generation Sequencing* qui regroupe en fait plusieurs générations : le séquençage de deuxième génération *Amplified Single Molecule Sequencing*, puis celui de troisième génération qu'on peut qualifier de *Single Molecule Sequencing* car elle permettrait de s'affranchir de l'étape de PCR. Le point commun à toutes ces technologies est finalement retrouvé dans la terminologie française de séquençage haut débit.

Aujourd'hui trois technologies de 2<sup>ème</sup> génération sont disponibles sur le marché, proposant chacune leurs avantages en termes de temps d'analyse, technologie, équipement, et performances [88][74].

### 3. Séquençage de deuxième génération

La 2<sup>ème</sup> génération permet le séquençage massif grâce à différentes innovations techniques. La possibilité de travailler avec des volumes réactionnels de l'ordre du micro/picolitre, le développement de supports solides miniatures (puces) ou encore la mise au point d'outils informatiques de plus en plus performants ont rendu possible la réalisation de millions de réactions de séquençage en même temps ainsi que l'analyse des données générées. On peut donc en un minimum de temps séquencer de très nombreuses bases, que ce soit le génome entier d'un individu, ou bien un seul gène mais pour de nombreux individus simultanément [91].

Dans le domaine de l'histocompatibilité, l'intérêt majeur de la NGS est d'obtenir des résultats de typage de haute résolution pour de nombreux échantillons et en une seule expérience. La découverte de nouveaux allèles est aussi un point fort de cette technologie, qui respecte et facilite la conformité aux Standards EFI [91].

Les trois technologies disponibles à l'heure actuelle sur le marché sont proposées par trois fournisseurs différents : Roche®, Illumina® et Ion Torrent® (Life technologie®). Il existe pour chaque fournisseur une gamme d'automates disponibles selon les besoins. Certains sont adaptés à une utilisation à grande échelle, d'autres sont des modèles plus petits permettant la mise à disposition dans les laboratoires d'analyses de biologie médicale de plus petite ampleur. Les différents automates disponibles sont représentés dans le chapitre suivant (III. 4).

Chacune des technologies possède des avantages et inconvénients que nous verrons par la suite, mais les grandes étapes qui se succèdent pour aboutir à un séquençage restent identiques :

- Extraction de l'ADN
- Préparation d'une librairie
- Amplification clonale de la librairie par « PCR en émulsion » (Roche ou Life technologie) ou « bridge PCR » (Illumina)
- Séquençage
- Analyse informatique des données

Dans les protocoles proposés pour le typage par NGS, il existe deux approches possibles selon la cible à séquencer :

- ✓ ***L'amplicon-based exon sequencing***: il s'agit de pré-amplifier la séquence d'intérêt avant la préparation de la librairie par une PCR spécifique d'exon. Ceci convient pour des séquences de faible taille (<1000 pb) qui ne nécessitent pas d'être fragmentées avant le séquençage.
- ✓ Le ***shot gun*** consiste à travailler sur un gène entier ou bien un grand fragment d'ADN issu d'une étape de *long range PCR*, celui-ci sera alors fragmenté en morceaux de taille aléatoire, qui seront sélectionnés et « ajustés » grâce à une exonucléase de façon

à obtenir des extrémités franches permettant la fixation d'un adaptateur, on parle d'*end repair*. La méthode de fragmentation peut être mécanique (nébulisation, ultrasons...) ou bien enzymatique. La fragmentation enzymatique est plus rapide mais plus coûteuse [89]. Dans le kit NGS disponible à l'heure actuelle pour la réalisation de typage HLA, c'est cette dernière méthode qui est proposée.

### Extraction de l'ADN

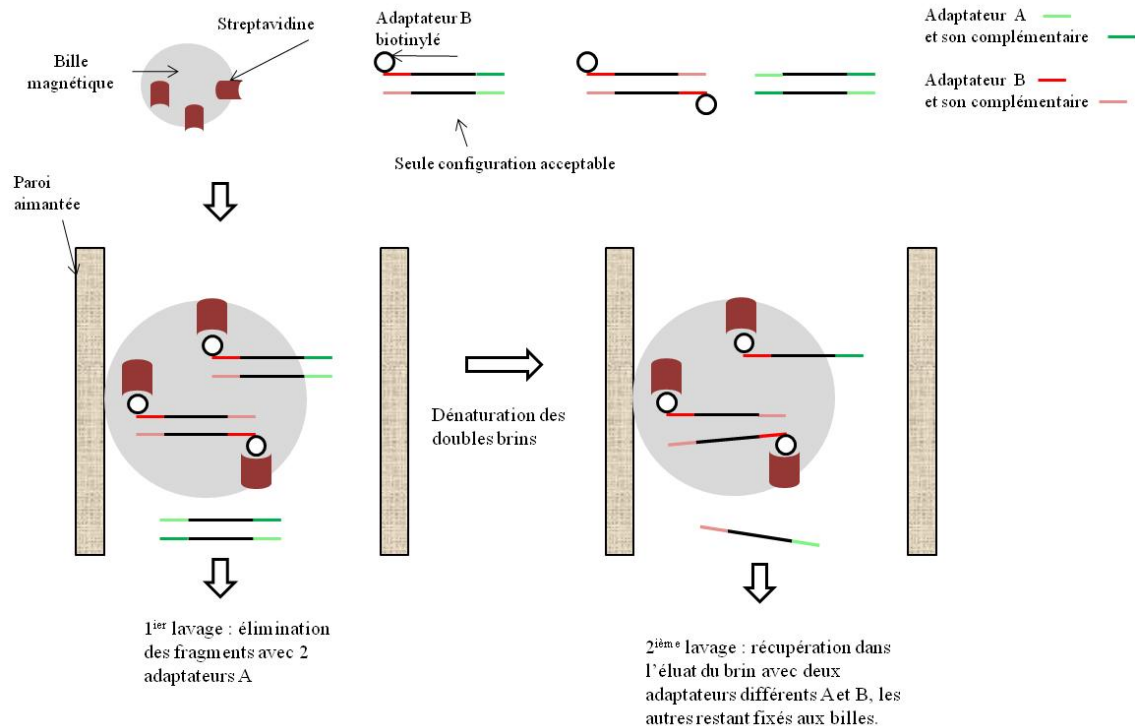
La méthode d'extraction utilisée influence la qualité de l'ADN obtenu. Dans le cas de l'approche par *long range PCR*, il apparaît important d'obtenir un ADN dont l'intégrité est bien conservée, autrement dit un ADN le moins fragmenté possible. Cette étape aura un impact direct sur la qualité des résultats de typage. Une étude a été menée pour comparer différentes méthodes d'extraction, classées selon la technique (billes magnétiques, extraction sur colonne et enfin par précipitation). Une analyse de l'ADN obtenu a été effectuée, en particulier avec une électrophorèse sur gel en champ pulsé permettant l'observation du poids moléculaire des molécules d'ADN extraites. Cette analyse a permis de montrer des différences significatives entre les résultats de typage obtenus (à partir d'une *long range PCR*) avec chaque technique, et que les meilleurs résultats sont constatés avec l'extraction par précipitation, puis par les billes magnétiques. L'utilisation de l'extraction sur colonne semble la moins adaptée [92]. En revanche, les résultats de typage obtenus dans une autre étude, basée cette fois sur une approche d'*exon-sequencing*, ne semblent pas varier selon les différentes méthodes d'extraction [93].

### Préparation de la librairie:

On appelle librairie l'ensemble des fragments d'ADN à séquencer, préparés avec des adaptateurs fixés à leurs extrémités, et éventuellement des identifiants. Cette préparation à partir d'échantillons d'ADN double brin est nécessaire pour qu'ils puissent subir les étapes d'amplification clonale puis de séquençage mais également pour les rendre identifiables en cas de multiplexage. Un adaptateur est une séquence nucléotidique de quelques paires de bases, il est fixé par une ligase au fragment d'ADN double brin. Il servira à la fois à la fixation des fragments sur un support et d'amorces universelles pendant les étapes précédemment citées. Le terme « universel » est utilisé car il suffira seulement de deux modèles d'amorce pour amplifier et séquencer un grand nombre de fragments différents, sans se préoccuper de leur séquence.

On parle de *paired-end library* lorsqu'on utilise un adaptateur à chaque extrémité. Il est possible aussi pour des gros fragments d'ADN (> deux kilobases) de les circulariser en utilisant un seul adaptateur qui joint les deux extrémités, la réaction sera alors générée dans les deux sens de l'ADN, à partir de l'adaptateur et sera donc ainsi plus rapide, on parle dans ce cas de *mate pair library* [89]. Ce type de librairie ne sera pas détaillé car il ne concerne pas

le typage HLA. Dans le cas de la *paired-end library*, un adaptateur A est placé à une extrémité de l'ADN double brin et le B à la seconde. Dans le cas des technologies de Roche® et d'Illumina®, les fragments qui ont deux adaptateurs identiques à leurs extrémités sont éliminés, pour le bon déroulement de l'étape de PCR (**Figure 14**). On comprendra pourquoi dans le paragraphe dédié à l'amplification clonale. Chez Ion Torrent®, cette étape est réalisée après la PCR en émulsion. Le principe de *paired end library* permet l'application dite du *paired end read*, qui est propre au fournisseur Illumina® (détaillé dans un paragraphe ci-dessous).

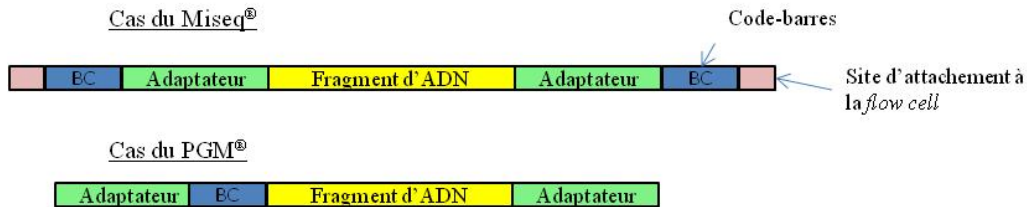


**Figure 14 : Sélection des fragments possédant un adaptateur différent à chaque extrémité, par l'utilisation de billes magnétiques et de l'affinité biotine-streptavidine.**

Pour ce faire l'adaptateur B est biotinylé. Les divers fragments formés sont mis en contact avec des billes magnétiques sur lesquelles sont greffées des molécules de streptavidine, qui possèdent une haute affinité pour la biotine. Les fragments bornés uniquement par deux adaptateurs A ne pourront donc pas être récupérés sur ces billes et seront éliminés par lavage. Ensuite, l'ADN qui est resté fixé sur les billes subit une dénaturation et seules les molécules d'ADN (devenues simple brin), bornées par des adaptateurs différents, auront la faculté de se décrocher pour être récupérées. En revanche, les deux brins des molécules d'ADN possédant deux adaptateurs B restent piégés par les billes.

Un des grands avantages de la NGS est le multiplexage. Il est en effet devenu possible de séquencer l'ADN de plusieurs individus en un même *run*. Un *run* correspond à la réalisation d'un cycle complet (ou processus) de séquençage par l'automate. C'est pendant la préparation de la librairie qu'il faut ajouter un «code-barres» aux fragments d'ADN, en plus de l'adaptateur (**Figure 15**). Ces courtes séquences de quelques nucléotides sont attribuées à chaque échantillon et permettront de les identifier lors de l'analyse informatique. On parle de

MID (*Multiplex Identifier*), de *tag*, de *barcode*, ou d'*index* en fonction des fournisseurs. Lors de la préparation d'une librairie en approche *exon-sequencing*, l'utilisation de *primers de fusion* permet la ligation des adaptateurs simultanément aux codes-barres. Ces *primers de fusion* sont des amorces spécifiques d'exon, auxquelles sont greffés un adaptateur et un identifiant, permettant la ligation en même temps que l'amplification. Dans l'approche *shot gun*, avec l'Ion Torrent<sup>®</sup>, le *barcode* est ajouté pendant l'étape d'*end-repair* en même temps que les adaptateurs, le code-barres étant contenu dans l'un des adaptateurs [94]. Dans le cas du Miseq, leur ligation peut être faite séparément de celle des adaptateurs.



**Figure 15 : Schéma représentatif d'une séquence cible entourée de ses adaptateurs et de code-barres avec l'approche *shot gun*.**

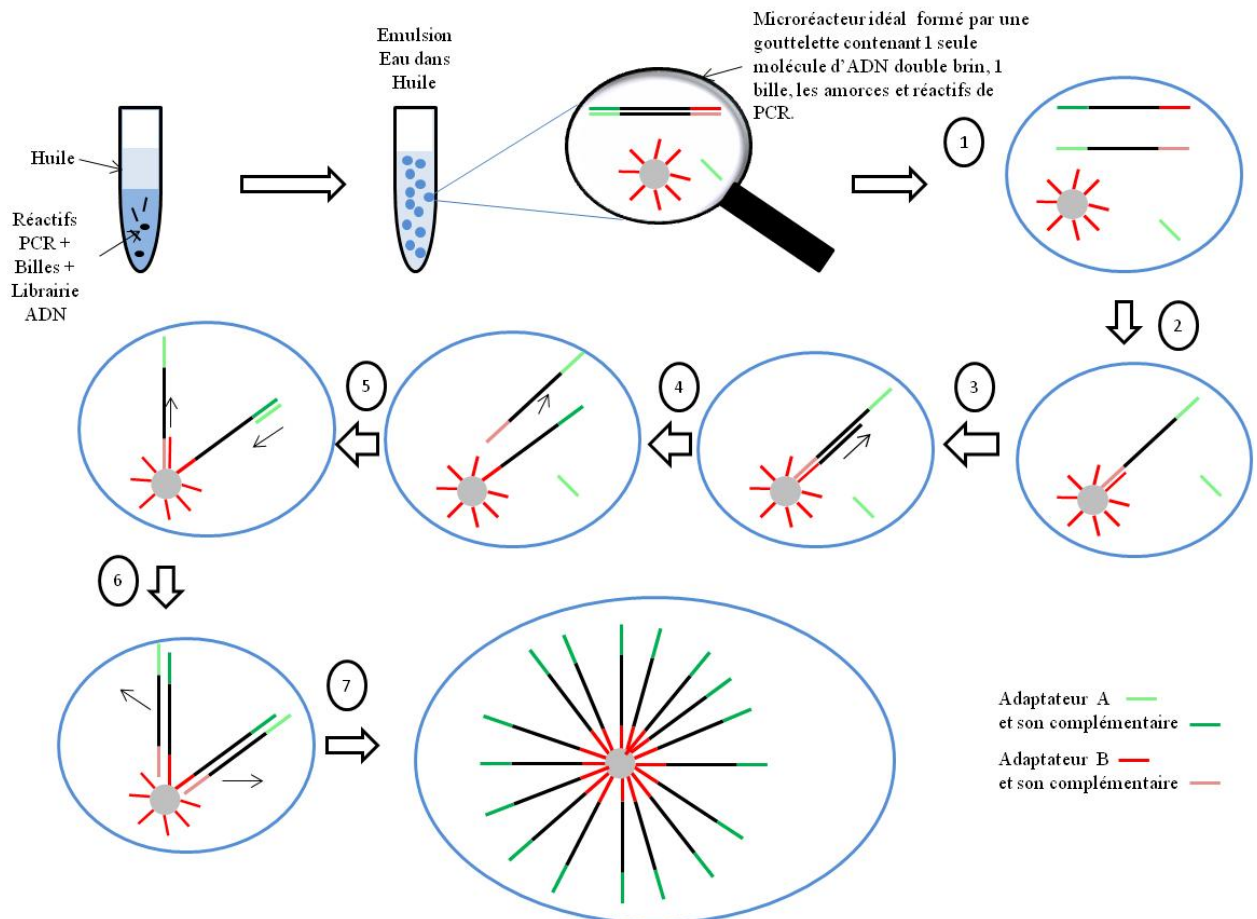
### Etape de PCR ou amplification clonale de la librairie:

On a vu qu'il existait deux types de PCR qui diffèrent principalement par le support utilisé:

- ✓ La **PCR en émulsion** proposée par Life technologie<sup>®</sup> et Roche<sup>®</sup>:

Les fragments d'ADN simple brin sont liés à des microbilles grâce aux adaptateurs, puis un mélange « eau dans huile » crée une émulsion dans laquelle chaque goutte d'eau constitue un micro-réacteur contenant une bille et des réactifs d'amplification (**Figure 16**). Ainsi des millions de réactions de PCR ont lieu en même temps dans un volume réduit permettant d'obtenir tous les amplicons correspondants [89]. Le résultat obtenu est appelé «polonies». Les billes sont appelées IPS pour Ion Sphere Particule chez Ion Torrent. Le volume réactionnel constitué par la gouttelette d'eau est de l'ordre du nanolitre (nL). Cette infime dimension impose la formation de gouttelettes de taille et de forme homogène pour le bon déroulement de la PCR, notamment sa précision et sa reproductibilité. Brièvement, elle peut être assurée grâce à l'arrivée simultanée dans un petit orifice des deux phases non miscibles, qui circulent initialement dans des canaux séparés. La phase aqueuse contient tous les réactifs de la PCR, les fragments d'ADN et les billes [95].

Ainsi, ce type de PCR ne peut pas être effectué dans un thermocycleur classique, chaque fournisseur propose un instrument qui servira à cette étape : il s'agit du Ion touch<sup>®</sup> pour Ion Torrent<sup>®</sup>. Ce dernier est un système microfluidique qui assure en premier lieu la génération de gouttelettes homogènes. Une fois les micro-réacteurs formés, ils traversent un circuit en capillaires, établi de façon à leur faire parcourir différentes zones de température fixe, pour permettre les cycles de dénaturation et d'hybridation [95]. Cette étape dure six à huit heures et comporte en moyenne 50 cycles.



**Figure 16 : Schéma du principe de PCR en émulsion, exemple des billes fixant les adaptateurs B.**

1 : Début de la PCR par dénaturation des fragments de la bibliothèque. Cette étape n'est effectuée que pour la technologie Ion Torrent® qui n'a pas réalisé la sélection des fragments simple brin avec deux adaptateurs différents.

2 : Hybridation de l'adaptateur B du fragment d'ADN sur l'adaptateur complémentaire attaché à la bille.

3 : Synthèse du brin complémentaire à partir de l'adaptateur B servant d'amorce à la polymérase.

4 : Dénaturation du brin matrice tandis que la molécule d'ADN simple brin néo-synthétisée reste fixée sur la bille (avec à son extrémité un adaptateur A).

5 : Un nouveau cycle commence pour le fragment matrice qui se fixe sur un autre site de la bille. Simultanément, l'amorce complémentaire de l'adaptateur A présente dans le milieu réactionnel, se fixe à l'extrémité distale du brin néo-synthétisé.

6 : Chaque brin sert donc de modèle à la formation de nouveaux fragments, attachés à la bille par leur extrémité complémentaire à l'adaptateur B.

7 : Après plusieurs cycles de PCR, la bille est ainsi recouverte de plusieurs millions de molécules d'ADN simple brin identiques.

On comprend ici l'importance d'avoir des fragments avec les deux adaptateurs différents à chaque extrémité.

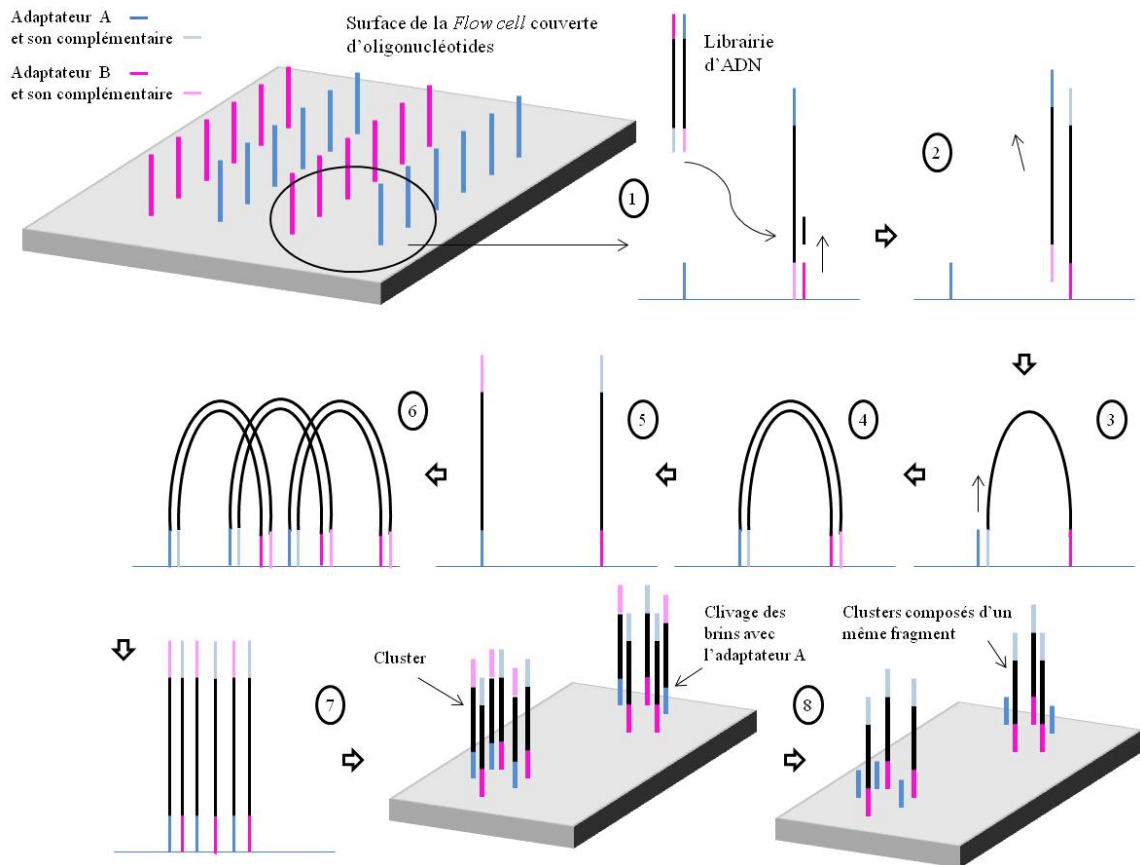
La PCR est bidirectionnelle si deux types de billes fixant chacune un adaptateur sont utilisées, les brins sont alors amplifiés dans leur deux sens. On obtient une PCR unidirectionnelle si les billes utilisées ne fixent qu'un type d'adaptateur.

A la fin de l'analyse, les billes doivent être triées et sélectionnées de façon à ne garder que celles qui sont dites clonales, c'est-à-dire qui n'ont amplifié qu'un seul et même fragment d'ADN. L'utilisation d'une faible concentration d'échantillon pour la préparation de l'émulsion est primordiale, afin d'obtenir une majorité de billes clonales et éviter l'amplification de plusieurs fragments sur une même bille (billes polyclonales). Cette concentration doit tout de même être suffisante pour ne pas laisser trop de billes n'ayant fixé aucun fragment, ou billes négatives. En réalité, les billes polyclonales ne pourront pas être éliminées avant le séquençage, mais les données issues des séquences qu'elles ont générées seront éliminées pendant l'analyse.

Des modules d'enrichissement comme l'Ion touch ES (Ion Torrent®) et le Robotic Enrichement Module (Roche®) permettent ainsi la sélection de billes positives (clonales et polyclonales). Ils utilisent le même principe de capture sur bille que celui précédemment cité, avec la streptavidine et la biotine.

✓ La **Bridge PCR** ou *solid phase amplification* proposée par Illumina® :

Cette fois le support servant de thermocycleur est une surface plane appelée *flow cell*. Elle est revêtue d'oligonucléotides monocaténares de deux types qui sont complémentaires aux séquences des adaptateurs ligaturés pendant la phase de préparation de la librairie. Ils servent aussi d'amorces pour la PCR. Ils sont clivables chimiquement, par des agents différents, pour permettre un détachement de la *flow cell*, sélectif de certains brins. Le processus est beaucoup plus facile à comprendre en image (**Figure 17**).



**Figure 17 : Principe de la *Bridge PCR* (PCR en pont).**

- 1 : La librairie est dénaturée chimiquement et les différents fragments d'ADN simple brin obtenus sont hybridés à une amorce liée à la surface de la cellule, par l'intermédiaire de leur adaptateur.
- 2 : Des réactifs d'amplification sont ajoutés et un cycle d'élongation permet la formation d'un brin complémentaire de celui à cloner qui sera fixé au support.
- 3 : Une variation thermique permettra à ces brins de se replier pour s'hybrider à l'amorce voisine. Une extension des brins peut commencer, ils forment alors des «ponts» d'où le nom donné à cette technique.
- 4 à 7 : Grâce à des changements cycliques de température, communs à une PCR classique, on obtient à la fin de cette étape des millions de «clusters» à la surface de la cellule. Ils correspondent chacun à un groupe de produits d'amplification issu d'un même fragment.
- 8 : Les brins formés à partir d'un même adaptateur sont détachés du support grâce à une fonction « diol » d'un des deux types d'oligomères qui subit une dégradation chimique. Ainsi chaque fragment est représenté dans un cluster dans une seule direction (sens ou antisens) [96].

L'espace suffisant de chaque type de fragment sur la *flow cell* est permis grâce à cette astuce, seuls les brins ayant le même sens sont gardés dans un même cluster. Les risques de polyclonalité des clusters sont fortement réduits par une forte dilution de la librairie, de l'ordre du picomole.

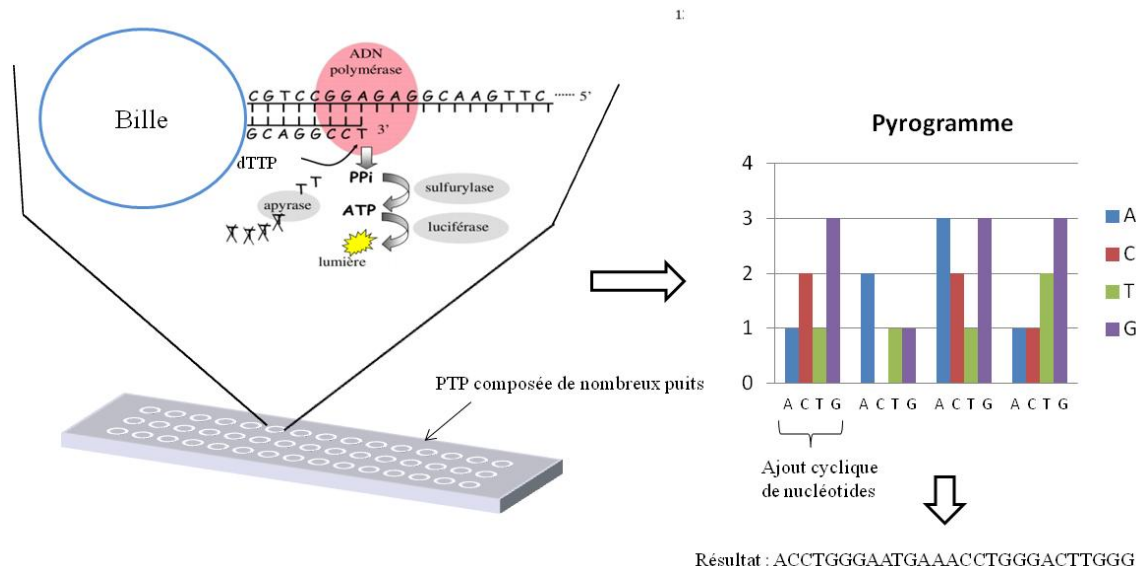
Cette technique d'amplification en pont possède donc l'avantage d'un gain de temps par rapport aux technologies qui utilisent la PCR en émulsion, puisqu'elle ne nécessite pas d'étape d'enrichissement. De plus, les produits d'amplification obtenus sont directement prêts au séquençage sur la *flow cell* permettant l'automatisation maximale du processus. La *flow cell* est seulement déplacée dans le séquenceur [89]. Cette étape de formation des clusters prend environ quatre heures et nécessite 35 cycles.

### Etape de séquençage :

C'est à cette étape que les différents fournisseurs se distinguent le plus.

#### ✓ Roche®, exemple du Genome Sequencer (GS), Junior 454

La technique utilisée est ici le pyroséquençage (**Figure 18**). Après l'étape de PCR en émulsion, le mélange de billes est déposé sur une plaque PTP (Pico Titer Plate®) contenant des millions de puits dont les diamètres ne permettent d'accueillir qu'une seule bille. Ces puits constituent des microréacteurs où vont s'effectuer les différents cycles du pyroséquençage. Le brin d'ADN fixé sur la bille sert de matrice et l'amorce A (dans le cas de billes B) est utilisée pour initier cette polymérisation.



**Figure 18 : Principe du pyroséquençage.**

D'après M. Falque [97].

A chaque étape d'un cycle, 1 seul des 4 nucléotides (dNTPs) est ajouté au milieu. S'il est attendu dans la séquence, il est incorporé par la polymérase, un signal est alors généré. Sinon il est éliminé par l'étape de lavage qui succède puis le nucléotide suivant est ajouté et ainsi de suite. Cet enchaînement d'étapes est défini par le terme de *flow*. Le lavage des dNTP non incorporés est assuré par une enzyme : l'apyrase. Si le même nucléotide est incorporé plusieurs fois en une seule étape, le signal est proportionnel au nombre de nucléotides incorporés car la réaction utilisée est stœchiométrique. A chaque incorporation d'un dNTP, un pyrophosphate est libéré en quantité équimolaire. Ce dernier est utilisé par une ATP sulfurylase qui convertit l'Adénosine 5'phosphosulfate (APS) en ATP. Celui-ci permet l'action d'une enzyme, la luciférase, qui produit alors une molécule d'oxyluciférine, source de lumière. Ainsi, le signal correspond à une émission de lumière mesurée par un capteur CDD (*Charge Coupled Device*). Les résultats sont présentés sous forme d'un pyrogramme [89].

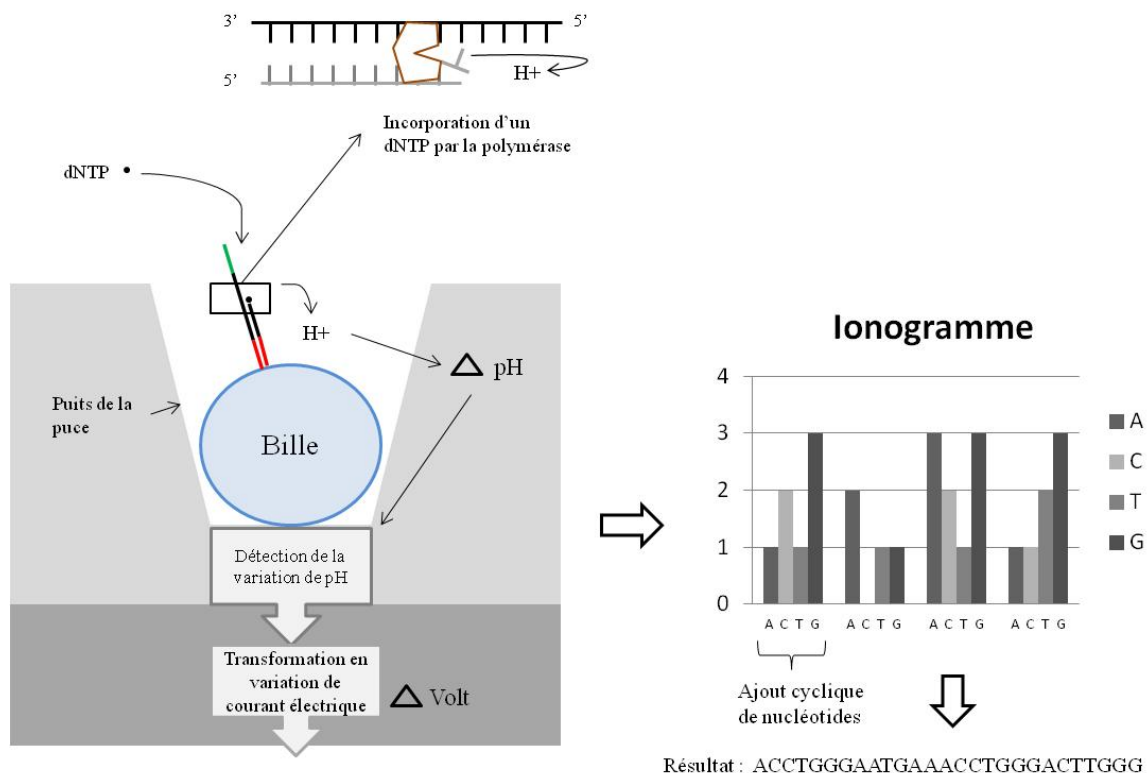
Pour éviter toute interférence, la désoxyadénosine triphosphate (dATP) utilisée pour l'élongation du brin d'ADN est ici remplacée par de la désoxyadénosine alpha-thio-triphosphate (dATPaS), qui est aussi bien reconnue par l'ADN polymérase, mais qui ne peut pas être utilisée par la luciférase.

On comprend qu'avec cette méthode, la distinction de séquence homopolymère du type GGGG ou GGGGG nécessite une sensibilité importante du signal. Pourtant il devient difficile de distinguer un enchaînement de plusieurs nucléotides identiques, car l'émission de lumière n'est pas assez discriminante, c'est le point faible des deux premières technologies évoquées. Les homopolymères entraînent des scores de basse qualité et constitue la grande faiblesse de ce séquenceur, ceci se traduit par un taux de délétion/insertion plus élevé qu'avec d'autres technologies. Il semble qu'un homopolymère de plus de six paires de base engendre un grand risque d'erreur [88] [94].

✓ **Life technology®**, exemple du **Personal Genome Machine (PGM)**, **Ion torrent®**

La technologie utilisée est celle des semi-conducteurs [98]. Les billes utilisées pour la PCR en émulsion sont appelées ici *Ion Sphere Particle (ISP)*. Après l'étape d'enrichissement qui permet de ne récupérer que les billes positives, ces dernières sont «chargées» dans une puce. Il en existe de différentes capacités en fonction du résultat souhaité et de la quantité d'échantillons à tester. La puce, ou *chip*, est constituée de millions de puits dont les diamètres ne permettent d'accueillir qu'une seule bille, elle est l'équivalent de la plaque PTP de Roche®. Selon le même principe que la technologie Roche®, les dNTP sont ajoutés de façon séquentielle. Mais cette fois le signal qui suit l'incorporation du nucléotide est engendré par la libération d'un proton ( $H^+$ ) qui va modifier le pH (potentiel en Hydrogène) dans la chambre réactionnelle (**Figure 19**). Cette variation immédiate du pH est détectée par la cellule de mesure puis transformée en un courant exprimé en Volt. Les résultats sont présentés sous la forme d'un ionogramme. Le nombre de protons libérés est identique au nombre de dNTP incorporés et par conséquent le signal est proportionnel [99]. De ce fait, les homopolymères sont de potentielles sources de problèmes avec cette technologie également.

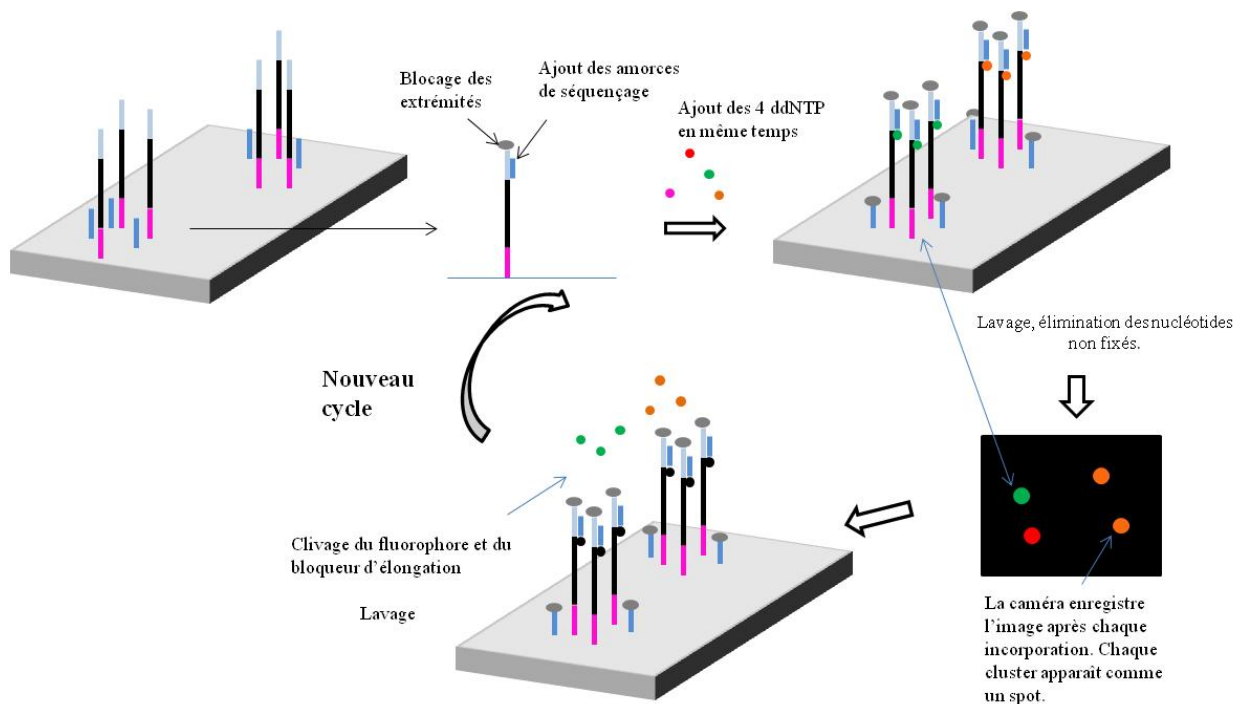
L'originalité de cette technique réside dans le fait qu'elle n'utilise pas de caméra pour la détection du signal ni de dNTP avec un marquage fluorescent. Son mode de détection est basé sur la pHmètrie et constitue un avantage au niveau du coût [100].



**Figure 19 : Principe de séquençage utilisant la technologie des semi-conducteurs.**

## ✓ Illumina® Solexa, exemple du Miseq system

La technologie utilisée est appelée *Reverse Dye Terminator* (RDT) car elle consiste en l'utilisation de didésoxynucléotides (ddNTP) qui possèdent chacun un marqueur fluorescent différent (*reversible terminators*) et un bloqueur d'élongation. Ils diffèrent de ceux utilisés dans la méthode Sanger car le blocage à l'extrémité 3' peut être supprimé par un agent chimique, il est donc réversible. Le séquençage est réalisé dans la *flow cell* où se trouvent tous les clusters (**Figure 20**). D'autres types de bloqueurs sont utilisés pour bloquer les extrémités libres des brins fixés sur la *flow cell* (ovales gris sur le schéma).

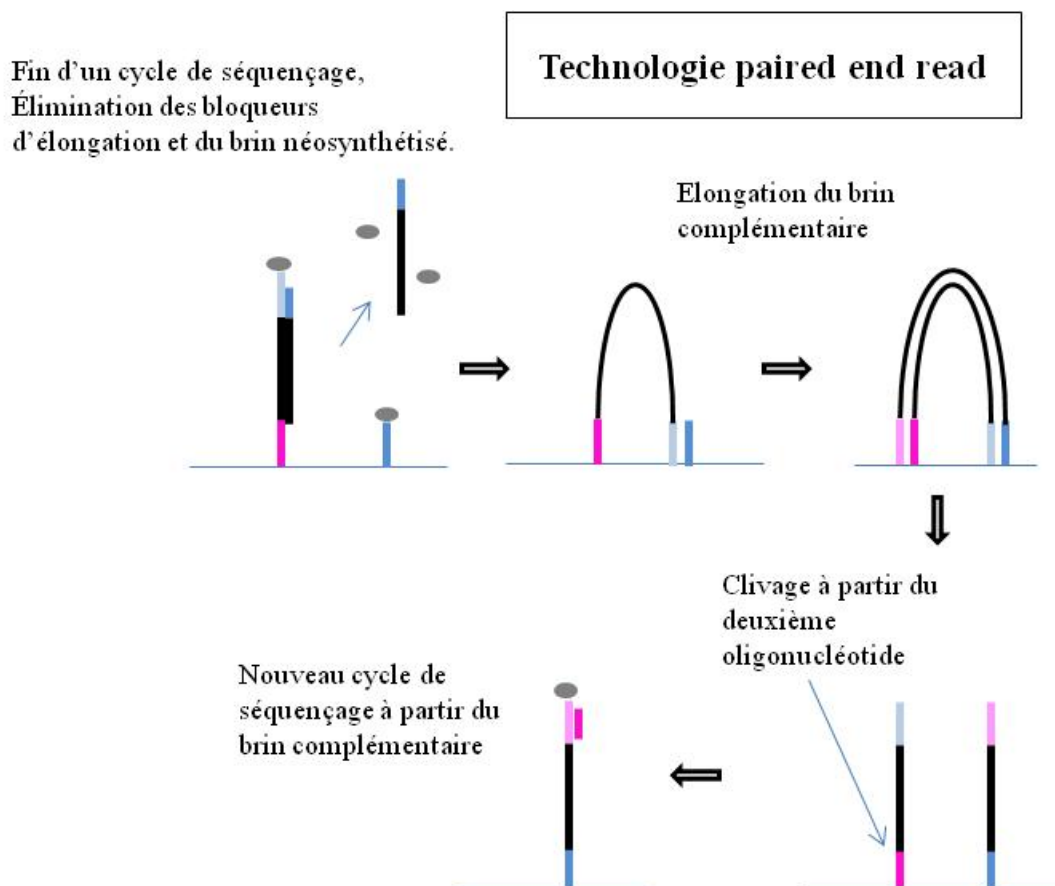


**Figure 20 : Principe de séquençage par la technologie Illumina.**

Cette fois les quatre nucléotides sont incorporés simultanément. Ils sont synthétisés de façon à stopper l'élongation donc un seul ddNTP est incorporé à chaque cycle de polymérisation. Les ddNTP non liés sont éliminés par une étape de lavage. Ensuite une image est enregistrée par une caméra CCD, permettant de connaître le nucléotide incorporé au niveau de chaque cluster grâce à la fluorescence des marqueurs. Enfin, le marqueur fluorescent est clivé ainsi que le terminateur qui empêchait l'élongation. Un nouveau cycle d'élongation peut donc avoir lieu.

L'impossibilité d'incorporer plusieurs dNTP en un même cycle rend cette technologie plus précise que les deux précédentes notamment dans la détermination des séquences répétées ou homopolymères. Il y a autant de cycles de séquençage que de bases dans un *read* final.

Cette technologie possède un avantage supplémentaire, grâce à la méthode du *paired end read*. Ceci permet de doubler le nombre de *reads*, et donc la profondeur de lecture, sans diminuer les capacités de séquençage. Pour cela, lorsque la fin de la séquence d'un fragment est atteinte, celui-ci subit un « retournement » comme indiqué dans la **figure 21**. Une deuxième session de séquençage, débute alors dans un même *run*, à partir de l'extrémité opposée à la précédente. Un même cluster est donc lu deux fois en mode *paired-end read*, permettant la lecture des brins dans les deux sens. En lecture simple il est lu une fois, dans un seul sens du brin [96]. Dans le cas des autres technologies (Ion torrent ou 454), les brins peuvent également être lus dans les deux sens, mais cela requiert l'utilisation de deux fois plus de billes et diminue ainsi la diversité des reads séquencés. En d'autres termes, les billes utilisées pour séquencer le brin dans l'autre sens ne seront pas utilisées pour séquencer un brin d'ADN différent et diminue ainsi les « capacités » d'un *run*.



**Figure 21 : Principe du *paired end read*.**

## 4. Comparaison entre les différentes technologies proposées

Les technologies de séquençage de deuxième génération offrent différentes approches qui peuvent être comparées sur de nombreux points, chacune présentant ses avantages et inconvénients. La qualité des résultats obtenus pourra se traduire par différents paramètres que l'on va définir. Parmi les inconvénients, la nécessité d'intervention humaine est à considérer car elle constitue une perte de temps dans chaque étape menant au résultat. Elle augmente également les risques d'erreurs. Le tarif de l'automate, du matériel associé et des réactifs est fortement variable d'un fournisseur à l'autre et doit être étudié soigneusement dans le choix d'une technologie, en fonction des besoins du laboratoire. Dans cette thèse seront comparés les trois automates appropriés à une utilisation en laboratoire de biologie médicale, à savoir le Genome Sequencer Junior (GS) de Roche<sup>®</sup>, le Personal Genome Machine (PGM) de Life technology<sup>®</sup> et le Miseq system d'Illumina<sup>®</sup>. La **figure 22** récapitule les principaux critères de comparaison.

Globalement, les technologies les plus coûteuses sont celles du Miseq et du GS et la moins onéreuse est celle du PGM [101]. Les différences de prix, pour Roche<sup>®</sup> et Illumina<sup>®</sup>, s'expliquent en partie par la nécessité d'une caméra pour la détection de fluorescence et de nucléotides marqués, plus coûteux. Le matériel de détection prend également plus de place dans ce cas. L'originalité d'Ion Torrent est qu'il est le seul à proposer une méthode non basée sur la détection d'une lumière, mais sur des variations de pH. Ceci impose des contrôles de la qualité de l'eau et nécessite l'installation d'une station de purification (utilisation d'eau milli-Q) ainsi qu'une maintenance technique lourde et longue (plusieurs heures) pour la préparation de l'automate avant analyse.

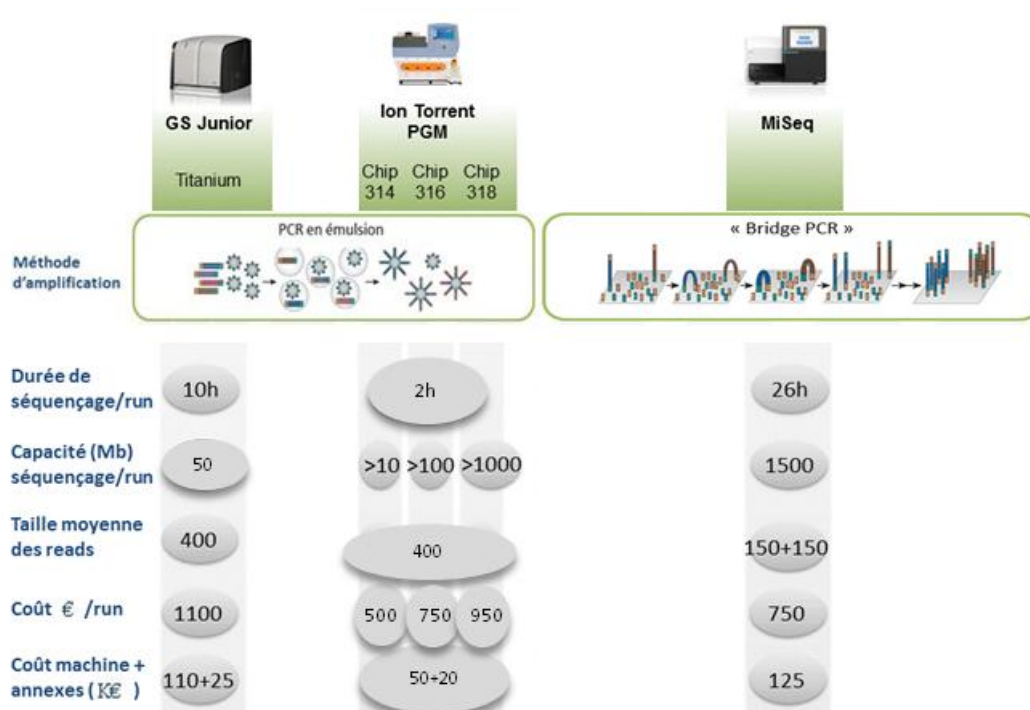
Débits et capacités s'expriment en nombre de paires de bases séquencées pendant un *run*. Une capacité importante offre la possibilité de séquencer un grand nombre d'échantillons en un run, ou bien, un nombre restreint d'échantillons, mais avec une profondeur de lecture plus importante et donc un résultat de séquence plus fiable. La **profondeur de lecture** correspond au nombre de lectures couvrant une position donnée de la séquence. En d'autres termes elle correspond au nombre de fois où une base est séquencée (**Figure 23**). Elle est exprimée en « fois » ou « x ». On comprend pourquoi l'augmentation de la profondeur est synonyme de résultat plus fiable et que selon l'application, une profondeur variant de 20X à 1000X sera exigée. Par exemple, dans le cas où il s'agit de retrouver des mutations très rares et ponctuelles à l'origine d'une pathologie (mutations somatiques), il faudra une profondeur élevée (1000X) pour assurer une bonne sensibilité. Dans le cas du typage HLA, où un polymorphisme est recherché, il sera donc théoriquement représenté pour moitié à une position donnée chez un patient à l'état hétérozygote. Une profondeur plus faible (de l'ordre de 50X par exemple) suffit à s'assurer du polymorphisme. Ainsi l'automate qui présente la plus petite capacité est le GS et le plus performant est le Miseq. Le PGM présente des

capacités intermédiaires et possède l'avantage d'avoir des capacités adaptables selon l'activité de l'utilisateur car trois puces sont proposées (Chip 314, 316 et 318) [101]. Pour donner une idée plus précise des capacités, d'après le fournisseur, la puce 314 de l'Ion Torrent permet de typer en un *run* les loci HLA-A, -B, -C, -DRB1, DPB1 de six patients avec une couverture moyenne (définie dans le paragraphe suivant) supérieure à 100X, soit une couverture tout à fait satisfaisante. La puce 316 permet d'augmenter le nombre de patients à 24 par *run*. La puce 318 permettrait de typer 48 patients en un *run* en admettant que de nouveaux *barcodes* soient disponibles, en attendant elle permet de typer 11 loci de 24 patients [94]. Le Miseq permet le typage des loci HLA-A, -B, -C et -DRB1 de 96 échantillons en un seul *run*. [102].

Une **lecture ou read** correspond à un fragment nucléotidique dont la séquence est déterminée par l'automate. La taille des *reads* doit être la plus grande possible pour éviter les problèmes de reconstitution de séquence après l'analyse, assurer une bonne couverture, et donc simplifier le travail. Si les *reads* sont trop courts, il est plus difficile pour le logiciel de rétablir leur enchaînement complet car leur chevauchement est plus faible, et le risque de génération d'ambiguïtés augmente [94]. Les longueurs proposées n'offrent pas toutes la possibilité de séquencer un exon entier en un seul *read*, ce qui est pourtant avantageux dans le cas de l'approche « *amplicon-based exon sequencing* » dont nous parlerons dans la discussion. De plus la qualité de lecture est médiocre en fin de *read*, il faut donc prévoir une longueur de *read* supérieure à celle de l'exon afin d'obtenir un résultat de séquençage satisfaisant (>300 pb environ) [89].

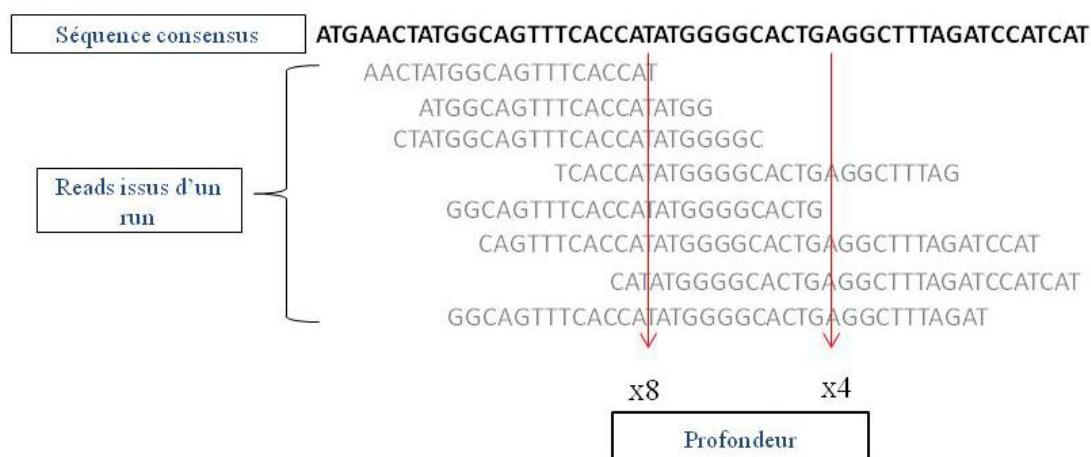
La notion de **couverture** correspond au pourcentage des régions d'intérêt séquencées, c'est-à-dire au pourcentage de bases couvertes par une profondeur définie. Ce paramètre représente la **sensibilité**. La possibilité de couvrir une base par de nombreux *reads* permet à la NGS d'offrir une sensibilité bien meilleure par rapport à une méthode Sanger.

Un lien entre la longueur moyenne des *reads* (L), le nombre de *reads* (N) et la profondeur (P) peut être établi par la formule suivante :  $P = \frac{N * L}{G}$  avec G la longueur totale du/des gène(s) à séquencer en un *run*. Cette équation simple permet de comprendre l'équilibre qui existe entre la qualité des séquences obtenues et l'exploitation optimale des capacités du séquenceur. En effet, la qualité des séquences obtenue est maîtrisée en grande partie par une profondeur de lecture suffisamment élevée. Si on augmente la variable G, qui reflète le nombre d'échantillons ou la taille des gènes de l'échantillon à séquencer, on diminue P, on perd donc en qualité. Il faut donc trouver le bon équilibre entre les deux variables dans le but d'obtenir un maximum de séquences de bonne qualité en un seul *run*. La capacité de l'automate doit être choisie en fonction de son mode et de sa fréquence d'utilisation.



**Figure 22 : Récapitulatif de la comparaison des différentes technologies disponibles.**

D'après <http://www.biorigami.com>



**Figure 23 : Représentation de la profondeur de lecture.**

La longueur des *reads* obtenus avec le GS et l'Ion Torrent sont les plus intéressantes, ils peuvent aller jusqu'à 400 paires de bases alors que le Miseq n'arrive qu'à 150 [103]. Une amélioration récente permettrait toutefois d'atteindre les 300 pb [94]. En revanche, il est le seul à disposer du *paired-end read*, permettant de multiplier par deux le nombre de lectures. En effet, comme expliqué précédemment, une seconde session de séquençage est débutée à partir de la deuxième extrémité de chaque fragment. On obtient ainsi des lectures d'une qualité satisfaisante à partir des deux extrémités de la séquence. Dans les autres cas, on observe des taux d'erreurs plus importants en fin de *read*. L'intérêt majeur du *paired-end-read* est de gagner en profondeur de lecture sans diminuer la capacité totale de séquençage (le nombre d'échantillons) d'un run.

L'obtention de plus grandes capacités demande également des temps d'exécution par *run* plus importants. Il faut compter moins de 10 heures par *run* pour le GS, alors qu'on atteint 26 heures pour le Miseq. Le PGM réalise un *run* en deux à trois heures selon la puce utilisée.

Cependant, il faut aussi prendre en compte le temps nécessaire pour la préparation des échantillons, qui est propre à chaque technologie. Par exemple, pour le typage de huit échantillons, avec l'Ion torrent, il faut compter 13 heures de préparation de librairie [101]. La préparation des librairies est une étape indispensable et critique au bon déroulement du séquençage. Elle est la plus chronophage de tout le processus et les moyens déployés pour la faciliter sont recherchés. Comme nous l'avons évoqué précédemment, la technologie de PCR en pont d'Illumina, réalisée sur le même support que l'étape de séquençage (*flow cell*), permet une automatisation complète. Elle nécessite peu d'intervention manuelle entre ces deux étapes, la *flow cell* est simplement déplacée vers le séquenceur.

Chez Ion Torrent, la préparation de la librairie est de plus en plus automatisable avec la sortie du Ion Chef qui rassemble plusieurs étapes. Après la préparation de la librairie et le contrôle de sa concentration, les échantillons sont chargés dans l'Ion Chef qui réalise la PCR en émulsion, le contrôle de sa qualité et l'enrichissement. Enfin la puce qui sert de support à l'étape de séquençage est automatiquement chargée des billes enrichies. Il possède un système de lecture des codes-barres qui permet l'identification des échantillons et des réactifs. Toute cette étape ne nécessite que 15 minutes d'intervention technique (au lieu de plusieurs heures) et permet d'aboutir directement à l'étape de séquençage. Cependant, cet automate n'est rentable que si les deux puces qu'il prépare peuvent être chargées sur le PGM, c'est-à-dire si les capacités (en terme de longueur de *reads*) ne nécessitent pas plus de 500 *flows* par puce car le maximum de *flows* réalisables dans un run est 1000 (2 x 500). Le séquençage des gènes HLA nécessite 850 *flows*, il n'est donc pas possible actuellement d'utiliser deux puces sur un même *run* dans ce contexte.

Quelle que soit la technologie choisie, PCR en émulsion ou en pont, la librairie doit être impérativement quantifiée avant son amplification clonale. Comme nous l'avons précisé précédemment, la concentration doit être favorable à la formation de microréacteurs efficaces. Plus précisément, il s'agit d'obtenir des gouttelettes contenant une bille avec un seul fragment à amplifier, et l'équivalent pour les clusters. Un ajustement de la concentration pourra être nécessaire voire indispensable dans le dernier cas. Les instruments Quant-It PicoGreen®

dsDNA assay et Qubit® fluorometer de Life Technologies® peuvent être utilisés pour leur rapidité [94]. Ils permettent d'éviter les interférences avec les molécules de type acides nucléiques, ARN et protéines grâce à l'utilisation d'un fluorochrome spécifique intercalant. Ce dernier possède une fluorescence très faible qui s'intensifie considérablement lors de sa liaison à la molécule d'ADN double brin. L'absorbance mesurée à la longueur d'onde spécifique du fluorochrome est ainsi proportionnelle à la concentration d'ADN double brin du milieu, sans être modifiée par la présence d'autres composants. Il existe un fluorophore spécifique pour chaque type de molécule, de façon à permettre, selon les besoins, les mesures des concentrations en protéines, ARN ou ADN d'un même milieu. Le Qubit® est destiné à un nombre réduit d'échantillons puisque le passage est individuel, alors que le Quant-It® est adapté à la lecture de microplaque de 96 puits donc à de nombreuses lectures simultanées.

Afin de s'assurer que la librairie en cours de préparation permette d'aboutir à un résultat de séquençage acceptable, plusieurs étapes de vérification apprécient la qualité de la manipulation. Elles évitent de prolonger inutilement une préparation fastidieuse, qui entraînerait un gaspillage de réactifs. Deux étapes de purification et de sélection de taille des fragments peuvent être nécessaires selon l'approche. Dans tous les cas, la vérification de la taille des fragments après ligation des adaptateurs est nécessaire, ainsi que la mesure de la concentration de la librairie après PCR et enrichissement, cette dernière pouvant être ajustée.

## 5. NGS et informatique

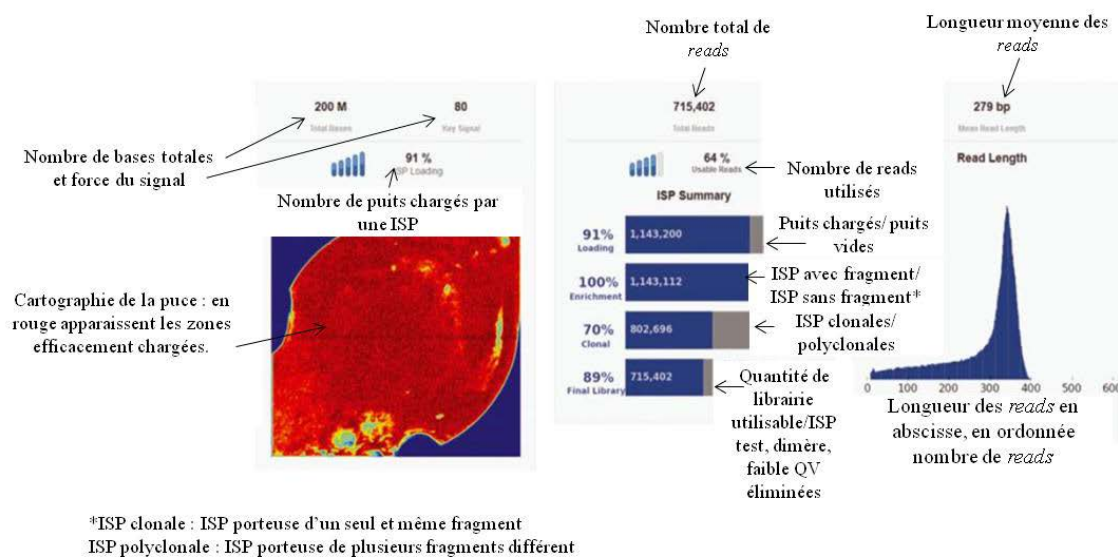
Les technologies de séquençage de nouvelle génération fournissent énormément de données, jusqu'à plusieurs Giga paires de bases (Gpb) sont produites par jour pour un séquenceur. Ces données sont impossibles à analyser sans logiciel à tel point que les facteurs limitant sont actuellement les logiciels d'analyses des données mais également leur moyen de stockage [104].

Les logiciels proposent un résultat, à l'aide d'algorithmes bien établis, et renseignent également par différentes façons sur la qualité du résultat obtenu. Dans le cadre du typage HLA, le **ratio allélique** (*allelic imbalance*) est un paramètre très utile. Il est calculé par le rapport des profondeurs moyennes des deux allèles. Ainsi il correspond au nombre de lectures provenant de chacun des allèles pour chaque locus. Si ce ratio est déséquilibré (éloigné de un) il peut signaler un éventuel phénomène de *drop out*, ou bien attiré l'attention sur un résultat homozygote (dans ce cas il doit évidemment être le plus déséquilibré possible). Les *drop out* sont souvent dûs à une mauvaise conception d'amorces génomiques de PCR mais peuvent être également induits par la présence de mutations qui empêchent l'hybridation d'une amorce. Ces mutations ponctuelles peuvent être rencontrées plus fréquemment dans le contexte d'hémopathies malignes, ce qui impose une vigilance accrue [101] [89] [105].

Il n'existe pas de consensus établi à l'heure actuelle pour la définition des valeurs des paramètres qualité acceptables pour une validation de typage HLA. Cependant, l'expérience de différentes études rapporte qu'une longueur de lecture moyenne d'environ 200 paires de

bases au minimum apporterait une qualité satisfaisante d'analyse. Des critères fixés à une profondeur moyenne de 100X et une profondeur minimum de 20X permettraient d'assurer des résultats robustes [94]. Un ratio allélique placé à 20-80% pour des échantillons hétérozygotes serait souhaitable, en d'autres termes, aucun des deux allèles ne doit être représenté par plus de 80% des lectures [94]. Cependant, il semble plus prudent que ces critères soient testés et établis après plusieurs essais lors de la mise en place de la technique dans un laboratoire.

Quelques exemples de données apportées par le logiciel du PGM sont représentés dans la **figure 24**. A ces données s'ajoutent, par exemple un paramètre appelé **uniformité de couverture**, qui est indiquée par le pourcentage de bases qui sont couvertes par au moins 20% de la profondeur moyenne. Le nombre de *reads* alignés (sur la séquence de référence) par rapport au nombre de *reads* générés est également un renseignement intéressant.



**Figure 24 : Exemple de données par le logiciel Ion Torrent.**

D'après C. Gabriel et al. [94].

Dans des études comparatives basées sur des séquences identiques analysées parallèlement, on observe de faibles variations entre les profondeurs de lecture des deux automates GS et PGM pour un même endroit de la séquence. Ceci permet de penser que la profondeur obtenue dépend principalement de la séquence elle-même, des types de nucléotides et de leur enchaînement plutôt que d'autres facteurs comme la préparation de la librairie [101].

Des études, réalisées sur le GS, ont également montré qu'en moyenne, 65-70% des *reads* étaient exempts d'erreurs, et que dans les autres moins de trois nucléotides étaient faux. La répétition de l'analyse conduisait dans plus d'un tiers des cas aux mêmes erreurs, ce qui laisse encore penser que certaines combinaisons de nucléotides favorisent les erreurs [94]. Finalement, seulement 50 à 80% des données brutes issues d'un *run* participent à l'établissement de la séquence consensus [106]. La plupart des séquences rejetées proviennent de billes polyclonales qui ont donné lieu au séquençage de plusieurs brins d'ADN différents dans le même puits [89].

On a vu que le multiplexage est possible avec les technologies de nouvelle génération, il est en effet très intéressant pour gagner en rapidité et réaliser des séries de typage. Les identifiants sont des oligonucléotides de quelques paires de bases (une dizaine), ils doivent être synthétisés de façon à distinguer les échantillons entre eux. Ils doivent aussi pouvoir se lier aux adaptateurs et ne pas entraîner de risque d'inhibition d'amplification et de séquençage. Chaque identifiant doit donc être validé et vérifié sur plusieurs échantillons connus ce qui les rend difficiles à élaborer et onéreux [94]. De ce fait, ils peuvent être un facteur limitant des capacités de séquençage, il existe un nombre limité de *barcode* disponibles pour chaque fournisseur (24 pour GenDx par exemple qui est le seul fournisseur à proposer un kit de typage HLA par NGS sur le marché). Les barres-codes ne sont pas à l'abri d'erreurs de séquençage, et sont étudiés de façon à ce qu'une ou plusieurs erreurs de séquençage dans ceux-ci soient bien détectées comme des erreurs afin de ne pas attribuer le *read* à un autre individu (présentant un barre-code proche du barre-code séquencé). La détection de ces erreurs entraîne automatiquement un rejet du *read* par le logiciel, qui ne sera attribué à aucun échantillon. D'après des données estimées avec le Miseq, ces erreurs peuvent atteindre 10 à 20% des *reads* [94].

Les différentes technologies proposées ont chacune leurs faiblesses. Comme nous l'avons déjà évoqué précédemment, les homopolymères sont un grand point faible de l'automate de Roche®, et également de l'Ion Torrent qui montrent des taux de délétions et insertions bien plus élevés que le Miseq. En effet, rappelons que chez ce dernier, l'ajout des nucléotides ne peut se faire qu'un par un, grâce au bloqueur d'élongation, ce qui limite ce type d'erreurs. Elles compliquent la tâche aux logiciels qui sont chargés d'aligner les séquences. Si les délétions/insertions ne sont pas une faiblesse du Miseq, en revanche, il présente des erreurs de substitution [107]. Les substitutions sont dues à des erreurs survenant pendant la PCR ou bien à un mauvais clivage du marquage fluorescent, propre à cette technologie. Globalement, le Miseq possède actuellement des données brutes plus fiables que le GS et le PGM.

Il existe plusieurs stratégies pour calculer des scores de qualité en séquençage, la plus connue étant le programme Phred, qui permet un calcul logarithmique lié à une probabilité d'erreur selon la formule  $Q = -10 \log_{10} P$  avec P la probabilité d'erreur d'identification d'une base. Cette probabilité est estimée par le séquenceur, en fonction de l'intensité du signal perçu et de la position de la base dans le *read* (score médiocre en fin de *read*, présence d'un homopolymère avant la base). Ce programme a été conçu initialement pour permettre des comparaisons entre séquenceurs. Un score de qualité Q est attribué à chaque base nucléique issue d'un électrophorégramme de séquençage automatique. Plus concrètement, un score qualité Phred Q10 correspond à une chance sur dix d'avoir une identification incorrecte d'une base, autrement dit une précision d'assignation d'une base de 90%. Un score Q20 correspond respectivement à une chance sur cent, ou une précision de 99% et ainsi de suite...[101] [108]. Au total, ils permettent de filtrer et d'éliminer les séquences incorrectes. Dans le cas des automates de NGS, les logiciels utilisent ces scores de qualité basés sur l'estimation Phred ou sur un autre type de calcul (qui ne sera pas détaillé) pour attribuer une *Quality Value* ou QV aux bases.

Le séquenceur effectue une série de corrections lors de l'acquisition de ces données (correction du bruit de fond, évaluation du degré de confiance de l'assignation), avant de les convertir en format FASTQ, c'est-à-dire en séquences nucléotidiques auxquelles sont associés des scores de qualité [88]. Le fichier texte FASTQ est une forme de stockage des séquences, il comprend pour chaque *read* quatre lignes (**Figure 25**).

```
@SEQ_ID
GATTGGGGTTCAAAGCAGTATCG
+
!''*(((((***+))%%%)++)(%%).1
```

**Figure 25 : Représentation d'une séquence nucléotidique en format FASTQ.**

La première ligne constitue l'identification du *read*, précédée d'un symbole @. La seconde ligne est la séquence de nucléotides en elle-même. Ces deux premières lignes constituent ensemble un fichier texte FASTA, à la seule différence du « @ » remplacé par un « > ». La troisième ligne est un signe « + », il n'est présent aujourd'hui que pour des raisons de compatibilité entre logiciels. La quatrième ligne est un enchaînement de score de qualité correspondant chacun aux nucléotides de la deuxième ligne. Ces scores sont convertis en caractères ASCII (*American Standard Code for Information Interchange*) qui est une norme de codage informatique. Dans le cas du séquençage Sanger et Ion Torrent<sup>®</sup>, les scores qualités sont basés sur le calcul Phred, les automates de chez Illumina<sup>®</sup> utilisent un autre type de calcul. Selon la version de la norme ASCII utilisée, les conversions ne sont pas forcément comparables [94] [108].

L'obtention de résultats sous forme d'une séquence consensus, c'est-à-dire une unique séquence établie à partir d'une combinaison de toutes les lectures obtenues, nécessite plusieurs étapes et logiciels. Ces derniers reconnaissent également les codes-barres et attribuent la séquence à l'échantillon correspondant. Le logiciel du séquenceur s'occupe de l'acquisition des données brutes et détermine les enchaînements de nucléotides, on parle de **base calling**. La reconstitution de la séquence analysée à partir des *reads* s'appelle le **mapping**, elle est réalisée par un second logiciel. Il peut être réalisé par alignement des lectures obtenues par rapport à une séquence de référence connue, ou bien, si aucun génome de référence n'est disponible, une nouvelle séquence peut être formée par assemblage *de novo*. Dans ce dernier cas, le chevauchement des *reads* permet de les aligner, il faut donc prévoir un chevauchement suffisant et des longueurs de *reads* assez importantes [94]. Plusieurs difficultés peuvent être rencontrées durant la reconstitution des séquences. Plus les lectures sont courtes, plus elles peuvent correspondre à plusieurs parties d'une référence, ou à plusieurs allèles dans le cas du typage HLA. L'objectif du séquençage étant principalement la recherche des polymorphismes, il faut que le logiciel accepte un minimum de variations lors de la reconstitution (mutations, insertions ou délétions). La difficulté est alors de considérer cette variation comme une réalité ou comme une erreur de la technologie. Cette réponse sera apportée par la multitude des lectures couvrant une même base, autrement dit sa profondeur.

La recherche de séquences de référence similaires à la séquence consensus obtenue est réalisée grâce à différents programmes d'alignement du type BLAT. L'alignement des *reads* est réalisé par rapport à des séquences regroupées dans la base de données de l'IMGT-HLA *data base*, selon des algorithmes permettant d'attribuer le meilleur emplacement au *read* [104]. Le logiciel calcule un score de *mapping* qui lui permet d'attribuer les *reads* à un emplacement sur une séquence. Ce calcul ne sera pas détaillé ici. La qualité générale de l'alignement est appréciée par la *Mapping Quality Value* (MQV). Elle correspond à une probabilité d'avoir un bon alignement qui prend en compte les scores de *mapping* et la longueur des *reads* alignés. Si aucun alignement d'un *read* à une référence n'est trouvé, il est comparé aux artefacts possibles qui sont regroupés dans une table. Les artefacts référencés sont des séquences qui peuvent être générées si l'amorce ne s'est pas hybridée à l'endroit attendu, par exemple avec les pseudogènes. En cas de suspicion d'erreur de ce type, certains logiciels, comme le Conexio ATF software, classent les *reads* erronés dans une « corbeille ». L'utilisateur a toujours la possibilité de déplacer ces données pour qu'elles soient prises en compte pour l'assignation [106] [109].

Dans le cas du typage HLA en *long Range PCR*, le chevauchement des *reads* est nécessaire pour l'assemblage final car les séquences des gènes entiers sont encore loin d'être toutes référencées dans la base de données. Pour l'approche *exon-sequencing*, l'assemblage par rapport à une référence est beaucoup plus simple pour des exons bien documentés et qui nécessitent peu de lectures pour être couverts en longueur [94].

Cette informatisation a pour avantage de permettre la réanalyse des données brutes issues d'un *run*. En effet, dans le cas où les *reads* n'ont pas une qualité satisfaisante, il est possible de les soumettre à des conditions plus strictes de *trimming*, c'est-à-dire de « couper » plus facilement les extrémités. La réanalyse des nouvelles données peut aboutir à un résultat qui aura perdu en couverture mais qui sera plus fiable.

Au total, une seule société commerciale propose une solution clé en main depuis la préparation des échantillons (avec l'option *long Range PCR* qui semble la plus plébiscitée) jusqu'à l'analyse de résultats via un logiciel dédié, il s'agit de **GenDx**<sup>®</sup> (pour les automates Miseq et Ion Torrent). A l'heure actuelle, en France, aucune autre méthode fournisseur n'est disponible avant Octobre 2015.

Finalement, Roche<sup>®</sup> propose un automate qui permet des résultats rapides, avec des longueurs de lectures appréciables mais les capacités de l'automate de paillasse sont limitées par rapport à ses concurrents avec un coût élevé par *run*. La résolution des homopolymères reste le problème majeur de cette technologie qui fut la première à être utilisée dans le cadre du typage HLA, mais qui annonce l'arrêt du marché dans le développement de kit pour le typage HLA. Cet arrêt de marché est prévu en 2016, laissant la voie aux deux autres séquenceurs. Les homopolymères sont aussi un problème pour l'Ion Torrent, qui représente pourtant la technologie la plus récente. Cette technologie présente un coût plus attrayant que son principal concurrent le Miseq, grâce à l'éviction de nucléotides au marquage fluorescent et des détecteurs associés. Sa rapidité de séquençage pour un débit et des longueurs de lectures tout à fait correctes en font un concurrent respectable, malgré le temps de préparation

de bibliothèques encore long. Le Miseq possède l'avantage indéniable d'être complètement automatisé à partir de l'étape d'amplification clonale de la bibliothèque, limitant les interventions techniques, et ainsi la perte de temps et le risque d'erreurs. La chimie utilisée a largement fait ses preuves dans l'histoire du séquençage et permet d'éviter les problèmes rencontrés chez les automates précédents. Cependant les erreurs les plus rencontrées seraient des incorporations d'un mauvais nucléotide, en particulier après une Guanine [89]. Il propose également la possibilité de faire du *paired-end read*, permettant une amélioration de la robustesse des résultats.

## 6. Séquençage de troisième génération

La troisième génération de séquençage est actuellement en cours de développement et d'amélioration. Elle est déjà utilisée mais est encore réservée aux activités de recherche.

Cette génération est qualifiée de *Single Molecule Sequencing* puisqu'elle permet d'obtenir un résultat de séquence à partir d'une seule molécule. Nous pouvons ainsi espérer s'affranchir de toute étape d'amplification et donc de la préparation d'une bibliothèque. Il s'agit d'une avancée prometteuse car ces étapes d'amplification sont les causes majeures d'erreurs de séquençage. Un autre avantage est également d'autoriser des lectures de plus grande taille et en un temps très court.

Il faut nuancer les explications concernant l'absence de nécessité d'amplification, ceci dépend entièrement de l'application utilisée. Dans le cas du typage HLA, il serait nécessaire de pré amplifier le gène à typer pour sélectionner la séquence d'intérêt et y ajouter des codes-barres nécessaires au multiplexage d'échantillons. Mais dans tous les cas, cette amplification (classique) resterait bien plus simple et rapide que la préparation d'une bibliothèque de deuxième génération.

### Technologie des nanopores

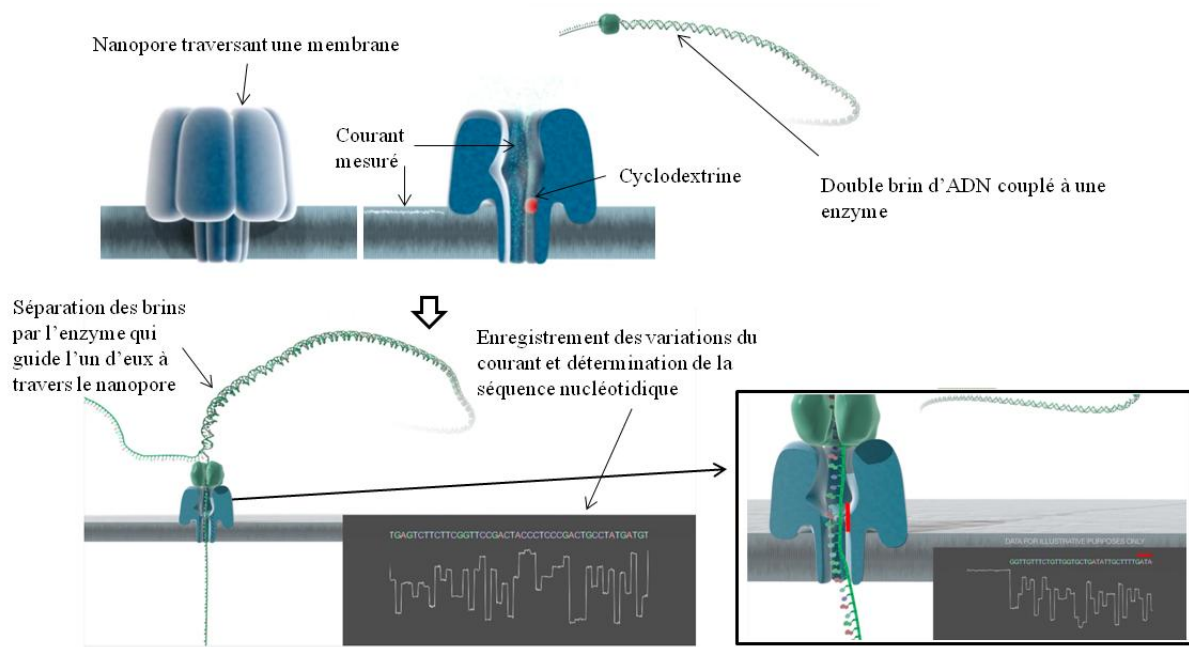
L'une des deux technologies de 3<sup>ème</sup> génération actuellement connue est proposée par Oxford nanopore technologies<sup>®</sup>. C'est une approche basée sur l'utilisation de pores de très faible diamètre, de l'ordre du nanomètre (nm), intégrés dans une membrane qui sépare deux compartiments remplis d'électrolytes, à travers lesquels passe un courant ionique mesuré en continu. Si le courant est ralenti par l'obstruction partielle du pore, ceci se traduit par une diminution du courant mesuré. Deux méthodes sont développées ensuite : l'*exonuclease sequencing* où chaque base est clivée individuellement par une enzyme et passe par le nanopore pour être identifiée ou bien le *strand sequencing* : une molécule d'ADN simple brin entière est menée par une enzyme à travers le nanopore (**Figure 26**). L'identité des bases est obtenue lors de leur passage une par une dans le pore au niveau d'une cyclodextrine en

mesurant la variation du courant ionique qui diffère selon la base qui bloque le pore. Ce principe a subi plusieurs améliorations qui tendent à gagner en précision et en longueur des lectures envisageables.

Ces nanopores peuvent être biologiques, comme par exemple la toxine  $\alpha$ -hémolysine du Staphylocoque doré, stable et de diamètre intérieur idéal pour qu'une molécule d'ADN simple brin puisse être "conduite" à travers le pore, nucléotide par nucléotide selon l'ordre de la séquence. Des supports solides peuvent être aussi utilisés à base de silicium ou de graphène. Ceci s'applique à des ADN simple brin ou des ARN dont la longueur peut atteindre plusieurs kilobases ou bien à de petites molécules (nucléosides).

Un séquençage de faible coût est alors permis du fait de l'absence de réactifs fluorescents, de système optique, et d'étape d'amplification. La purification et l'extraction nécessaires sont réalisables avec des kits existant sur le marché, à partir de 20 millilitres de sang. Le système de traitement des données est relativement simple [74][110][111].

Plusieurs difficultés ont été rencontrées au cours du développement de cette technologie, principalement concernant la vitesse de mobilité du brin d'ADN à travers le pore qu'il fallait réduire.



**Figure 26 : Principe du séquençage de 3<sup>ème</sup> génération selon la technologie des nanopores.**

D'après <https://nanoporetech.com> [112].

## Technologie des Zero-mode Waweguides

L'autre technologie est proposée par la société Pacific biosciences<sup>®</sup> avec l'automate PacBio RS. Cette technologie repose sur l'utilisation de cellules SMRT (*Single molecule real time*) qui contiennent plusieurs milliers de puits (**Figure 27**) appelés Zero-mode waweguides (ZMW). Dans chaque puits, se déroule une réaction de séquençage en temps réel grâce à une ADN polymérase, attachée au fond du puits. Chacun des puits contient un seul brin d'ADN et est alimenté en nucléotides, marqués chacun par un fluorophore différent. La polymérase peut alors les incorporer successivement et selon l'ordre de séquence. L'incorporation du nucléotide permet l'excitation du fluorophore correspondant car les ZMW sont exposés par-dessous à une lumière qui permet d'exciter le fluorophore. Ce faisceau d'excitation dont la longueur d'onde est trop importante pour traverser efficacement le nanopore, illumine uniquement la polymérase. Le faisceau d'excitation émis est donc fortement atténué ce qui crée un volume de détection très faible (de l'ordre du zeptolitre soit  $10^{-21}$  L) dans le fond du puits d'où l'appellation de « zero-mode waveguide ». Une fois l'incorporation du nucléotide achevée, la libération du fluorophore dans le puits met fin au signal lumineux (car il ne reçoit plus le faisceau d'excitation qui « illumine » uniquement la polymérase). Une caméra de détection est placée au dessus de la cellule SMRT pour capter la lumière émise à chaque incorporation. Ainsi un ZMW agit comme un microscope ultra puissant et la lumière émise à chaque incorporation est détectée par la caméra. Comme toutes les réactions se déroulent simultanément dans les puits de la cellule, le signal provenant des ZMW est ainsi considérablement amplifié. Le bruit de fond est quasi inexistant grâce au faible volume de détection atteint [113] [114]. La difficulté reste de synchroniser le séquençage de tous les brins simultanément.

L'avantage des plus grandes lectures proposées par ces nouvelles techniques est d'envisager une séparation facile des haplotypes, qui serait d'un grand intérêt en histocompatibilité.

Des tests de typage HLA ont déjà été réalisés sur le PacBio RS, avec multiplexage de huit à vingt échantillons sur les gènes de classe I. Les résultats sont très encourageants, et aboutissent à une réussite proche de 100% [115].

Ceci permet d'envisager également une première application de la 3<sup>ième</sup> génération, comme un moyen d'obtenir des lectures de grands fragments avec une faible précision. Ces derniers serviraient de « modèle » pour l'assemblage et la combinaison des nombreux petits fragments obtenus avec une grande précision par les technologies de 2<sup>ième</sup> génération, voire par la suite avec la 3<sup>ième</sup> génération elle-même [110].

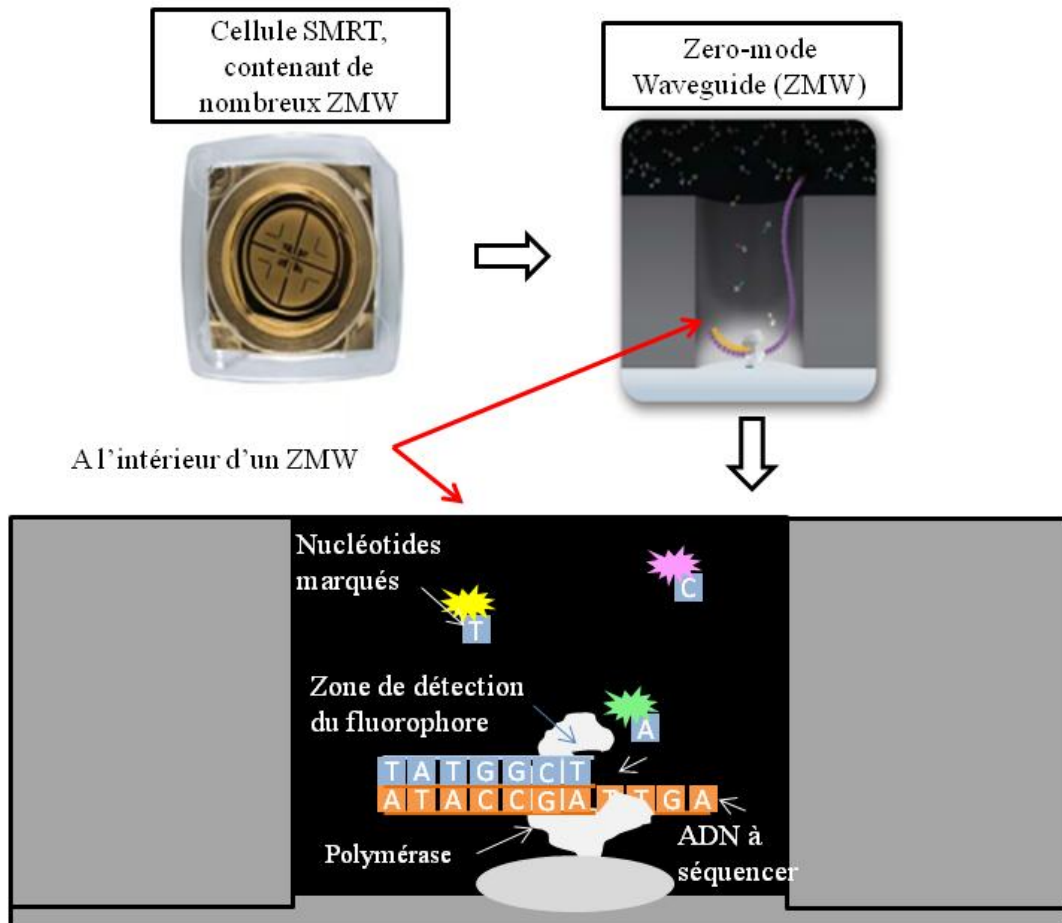


Figure 27 : Principe du séquençage de 3<sup>ème</sup> génération selon la technologie des cellules SMRT.

D'après [www.pacificbiosciences.com](http://www.pacificbiosciences.com) [113].

## OBJECTIFS

L'objectif de cette thèse est à la fois d'évaluer la mise en place du séquençage Sanger au laboratoire d'histocompatibilité du CHRU de Nancy, et de réfléchir à la mise en place d'une technique de séquençage de nouvelle génération pour le typage HLA des couples donneur/receveur de CSH et le typage des nouveaux DVMO.

La mise en place de la technique du séquençage Sanger mono-allélique a été demandée en Janvier 2013 par le laboratoire d'histocompatibilité du CHRU de Nancy. Cette demande fait suite à la modification des exigences des standards de l'EFI (version 6.0) concernant la définition du typage de haute résolution, demandé dans le cas des greffes de CSH avec donneurs non apparentés (Section D1.4 et I4.3). Il est maintenant obligatoire de lever toutes les ambiguïtés relatives au polymorphisme des exons 2 (et 3 pour la classe I), c'est-à-dire de rendre un résultat de typage en groupe G. Il est également demandé d'éliminer la possibilité d'allèles nuls faisant partie de ce groupe G. L'utilisation de cette technologie est devenu un besoin réel car les kits de PCR-SSP dits « alléliques » (même « combinés » aux kits de PCR-SSO) ne permettent plus de respecter la conformité aux standards EFI lors du typage des donneurs et receveurs de CSH. Les techniques de PCR-SSP et de PCR-SSO laissent un grand nombre d'alternatives alléliques et des allèles nuls qui ne peuvent pas toujours être éliminés (pas de typage en sérologie pour le locus C).

La mise en place du séquençage Sanger mono-allélique au laboratoire d'histocompatibilité du CHRU de Nancy n'a été effective qu'à partir de Février 2015, en raison des contraintes administratives. Afin de montrer que cette action a permis d'être en conformité avec les exigences EFI, nous réalisons une étude d'intérêt, par la comparaison des résultats de contrôles d'Evaluation Externe de la Qualité (EEQ) effectués avant et après l'utilisation du séquençage Sanger.

D'autres technologies de séquençage, dites de nouvelle génération, font leur apparition dans le domaine de l'histocompatibilité. Les améliorations recherchées concernent principalement le cadre de la greffe de cellules souches hématopoïétiques, et portent sur deux paramètres principaux : la résolution et le débit. En effet, l'agence de biomédecine souhaite maintenant que les DVMO soient inscrits avec un typage allélique, en remplacement des codes NMDP actuels, pour être compétitifs vis-à-vis des autres registres. Il existe déjà une saturation des inscriptions des DVMO en raison du manque de moyens techniques et de personnel, le séquençage Sanger n'apportera pas de solution à ce problème car il n'est pas adapté au typage de séries. Ainsi, devant cette augmentation des exigences, il est important de réfléchir à la mise en place d'une nouvelle technique adaptée à la réalisation de séries de typages de haute résolution, et aux autres perspectives qu'elle apporte. Le rôle du biologiste est de choisir la technologie qui lui semble la plus appropriée et la plus fiable, en fonction des moyens dont il dispose. Ainsi nous discuterons des points qui semblent importants à prendre en considération lors de la mise en place de la NGS.

# METHODE ET RESULTATS

Une première partie présente l'analyse des EEQ, avec la comparaison des résultats obtenus avant et après l'utilisation du séquençage Sanger afin d'évaluer l'apport de sa mise en place au sein du laboratoire.

La seconde partie concerne une vingtaine de dossiers qui ont présenté des difficultés lors de la réalisation du typage au laboratoire du CHRU de Nancy sur l'année 2014 - début 2015. Parmi ces dossiers, cinq ont été choisis pour illustrer la majorité des problèmes pouvant être rencontrés avec les typages de patients par les techniques actuellement utilisées. Nous pouvons les séparer en différentes catégories : la première concerne la capacité à détecter les allèles rares, la seconde concerne les discordances entre techniques de biologie moléculaire et enfin les discordances de typage entre la sérologie et la biologie moléculaire. Parmi eux, deux dossiers permettent de montrer les limites du séquençage selon la méthode Sanger et l'intérêt du séquençage de nouvelle génération.

## I. Présentation des résultats d'EEQ

Dans le tableau ci-dessous sont consignés les pourcentages de résultats au niveau haute résolution, demandé par les organismes d'évaluation externe de la qualité. Les résultats de l'année 2012 représentent douze échantillons d'EEQ qui ont été rendus par une technique de PCR-SSO, associée si nécessaire à une technique de PCR-SSP. L'année 2015 représente les résultats de typage de douze échantillons, testés à l'aveugle par séquençage Sanger. Ces échantillons, dont le typage attendu est connu, comprennent des EEQ des années précédentes et des DVMO typés au 6<sup>ième</sup> digit dans les laboratoires HLA extérieurs accrédités EFI. Le typage de ces échantillons porte sur les 5 loci HLA-A, -B, -C, -DRB1 et -DQB1.

	RSSO (2012)	SSBT (2015)
	Haute résolution (HR)	
Loc	% de résultats au niveau HR	
A	48%	92%
B	42%	96%
C	25%	96%
DRB1	62,5%	96%
DQB1	75%	100%

Nous pouvons constater une nette amélioration apportée par la technologie de séquençage dans la résolution des ambiguïtés et la qualité des résultats rendus. La SSBT permet donc de mieux répondre aux exigences de qualité, et d'être en conformité avec les standards EFI.

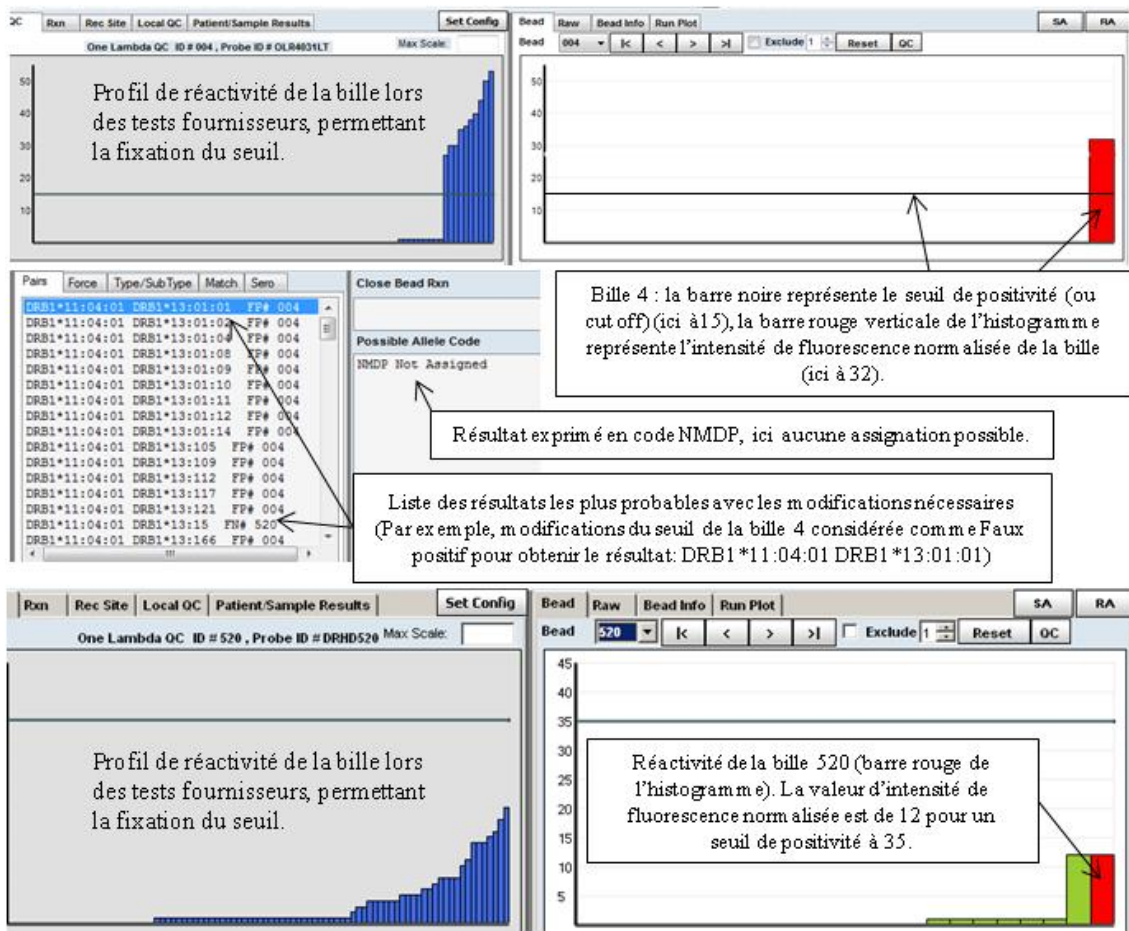
## II. Problèmes d'assignation d'allèle rare

### 1. Cas n°1 : détection d'un allèle rare (4<sup>ième</sup> digit)

Le typage est réalisé en PCR SSO dans le but d'inscrire un nouveau DVMO sur le registre. Le premier essai aboutit à une impossibilité d'assignation, toutefois deux résultats sont présentés comme étant les plus probables en admettant des modifications de seuil pour certaines billes (**Figure 28**) :

- ✓ Le résultat obtenu serait DRB1\*11:04:01 DRB1\*13 :ACSSR avec ACSSR = 01/28/105/109/112/117/121/166, en considérant la bille 4 comme étant faussement positive c'est-à-dire en modifiant le seuil de positivité de 15 à 33.
- ✓ Le résultat obtenu serait DRB1\*11:04:01, DRB1\*13:15, en considérant la bille 520 comme faussement négative c'est-à-dire en modifiant le seuil de positivité de 35 à 12.

Ce typage ne peut être validé selon les critères établis au laboratoire car il nécessite, dans un cas comme dans l'autre, la modification d'un seuil de plus de 50%. Les mêmes résultats sont obtenus lors du contrôle de ce test.

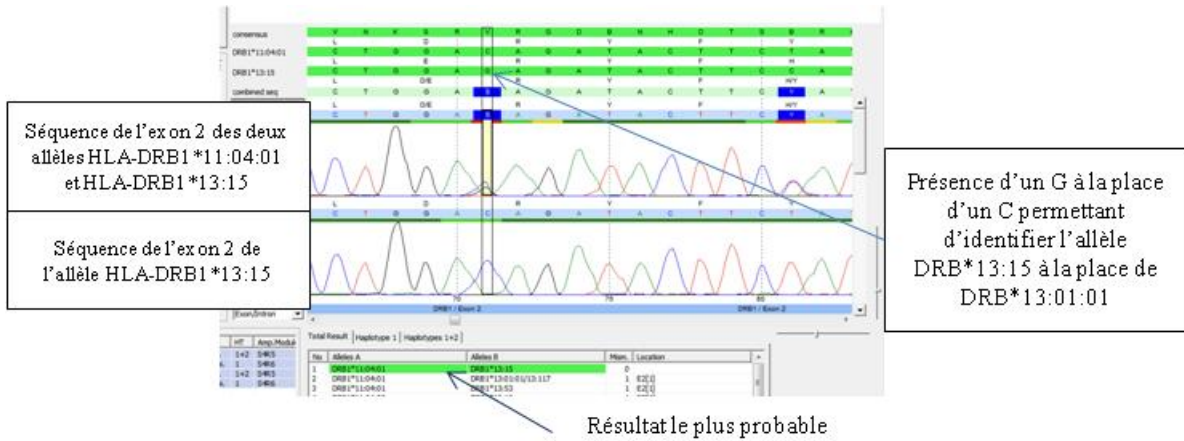


**Figure 28 : Validation sur le logiciel HLA Fusion<sup>®</sup>, présentation des informations disponibles pour la validation d'un résultat obtenu par PCR-SSO.**

Les deux histogrammes en haut à droite et en bas à droite présentent les réactivités des billes 4 et 520 respectivement.

Le typage par PCR-SSP allélique conduit à des ambiguïtés alléliques avec deux combinaisons comportant chacune un allèle rare (dont l'une avec DRB1\*13:15). Il ne permet alors aucun rendu de résultat de typage.

Un typage HLA-DRB1 en SSBT (exon 2 *forward*) est alors réalisé. Il permet l'assignation d'un résultat DRB1\*11:04:01, DRB1\*13:15 sans aucun problème, qui révèle un allèle rare (l'allèle DRB1\*13:15) (**Figure 29**).

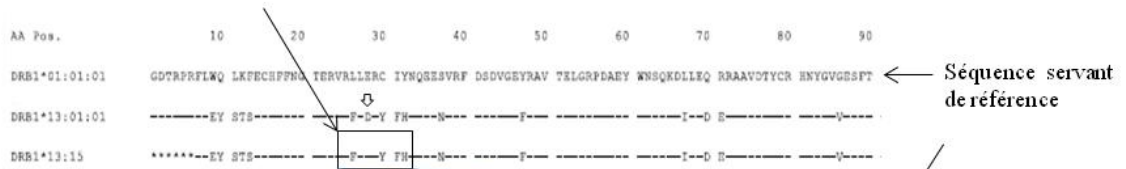


**Figure 29 : Validation des résultats de séquençage SSBT sur le logiciel JSI®.**

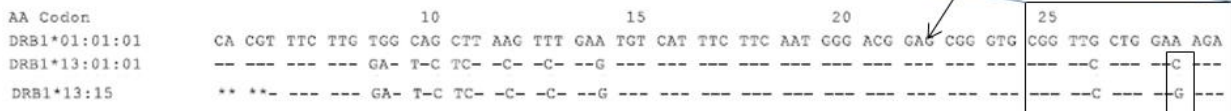
En revenant sur les résultats obtenus en PCR-SSO, on constate que la bille 4 est bien attendue positive en présence d'un allèle DRB1\*13:15 (et négative pour les allèles DRB1\*13:01:01 et DRB1\*11:04:01) : elle n'est donc pas faussement positive. En revanche, la bille 520 est bien faussement négative, ce qui explique l'absence d'assignation de résultat. Les zones d'hybridation respectives des deux billes sont représentées dans la **figure 30**.

Alignement des séquences protéiques

Zone d'hybridation des sondes de la bille 4 (25-F---YFH32) et de la bille 520 (24--F---29), différenciant l'allèles rare de l'allèle le plus fréquent.



Alignement des séquences nucléiques



Mutation d'un codon GAC (DRB1\*13:01:01) en GAG (DRB1\*13:15) entrainant le remplacement d'une Asparagine en Glutamine.

**Figure 30 : Alignement des séquences protéiques et nucléiques des allèles DRB1\*13:01:01 et DRB1\*13:15 selon le site de l'IMGT database.**

Les tirets représentent les analogies entre séquences, les étoiles représentent les zones encore inconnues dans les séquences. La mutation encadrée dans le schéma correspond à celle retrouvée dans la figure 29 par la technique de séquençage.

D'après les alignements réalisés sur le site de l'IMGT *database*, seule la mutation évoquée précédemment permet de différencier les allèles DRB1\*13 :01 et DRB1\*13 :15 sur l'exon 2. On peut donc penser que les deux billes (4 et 520), spécifiques de DRB1\*13:15, doivent s'hybrider sur le même site de l'allèle. On peut émettre l'hypothèse qu'une concurrence entre les deux billes pourrait expliquer la faible réaction de la bille 520, cette dernière possédant une zone d'hybridation moins grande que la bille 4 sur la séquence.

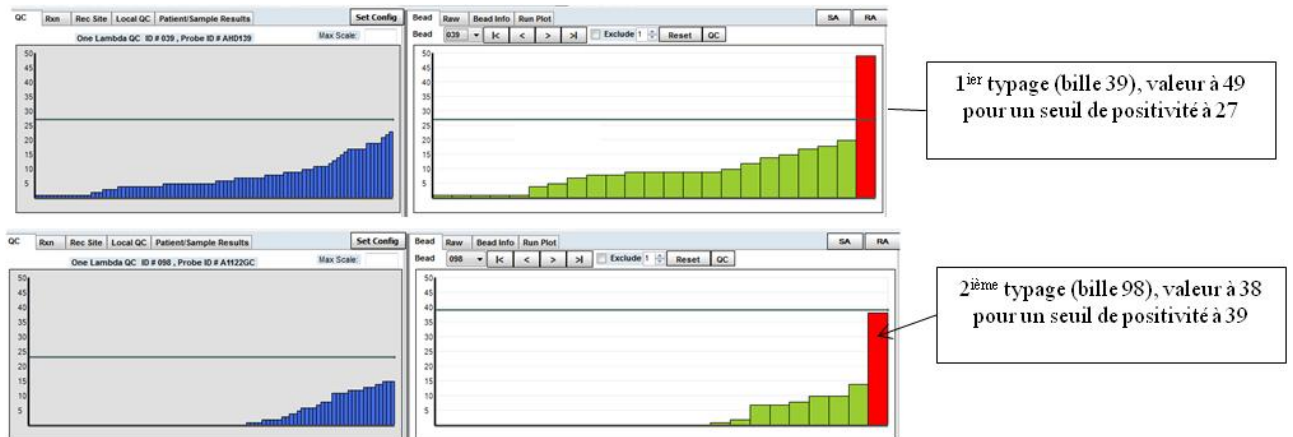
Ce dossier montre l'importance d'utiliser des techniques performantes, pour éviter des résultats ambigus. Le contexte d'une greffe de CSH nécessite un résultat allélique sur le locus DRB1, pour lequel il faut respecter obligatoirement des compatibilités strictes. Dans ce cas, l'impossibilité d'assigner un résultat sur le locus DRB1 entraînait un retard dans la prise en charge du patient. Pour obtenir un résultat en SSBT, l'externalisation du typage était indispensable avant la mise en place du séquençage Sanger au laboratoire.

## 2. Cas n°2 : détection d'un allèle rare (6<sup>ième</sup> digit)

Dans ce second cas d'inscription de DVMO sur le registre, le typage réalisé en PCR SSO permet, comme dans le cas précédent, d'obtenir deux propositions de résultats:

- ✓ A\*02:01:01, A\*02:01:09, correspondant à un typage avec un allèle rare (A\*02 :01 :09). Il n'y a pas dans cette proposition de changement de seuil de positivité de l'une des billes : ce typage répond aux exigences techniques du laboratoire cependant il doit être confirmé car chaque allèle rare doit être confirmé par une seconde technique.
- ✓ A\*02:ABVAT, A\*02:ABVAT est une autre possibilité que rend le logiciel d'interprétation de PCR-SSO. Mais cette proposition oblige à considérer la bille 39 comme faussement positive, c'est-à-dire qu'il faut modifier le seuil de positivité de cette bille de 27 à 49 (modification du seuil de plus de 50%) (**Figure 31**). Le code NMDP ABVAT correspond à la liste d'ambiguïtés ci-dessous :  
02:01/02:01L/02:01Q/02:09/02:43N/02:66/02:75/02:83N/02:89/02:97/02:132/02:134/02:140/02:153/02:241/02:252/02:256/02:266/02:291/02:294/02:305N/02:327/02:329/02:340/02:341/02:343/02:347/02:348/02:349/02:350N/02:351/02:353/02:354/02:356N/02:357/02:360/02:361/02:363/02:365/02:366N/02:368/02:371/02:372/02:374/02:375/02:377/02:378/02:380/02:381/02:383/02:384/02:389/02:390/02:392/02:394/02:395N/02:397/02:399/02:401/02:407/02:410/02:411/02:414/02:418/02:422/02:423/02:424/02:434/02:435/02:439N/02:440Q/02:441/02:443/02:444/02:445/02:446/02:455/02:456/02:458/02:460/02:461/02:469. Au vu de ce résultat, le résultat le plus probable est A\*02 :01 homozygote si l'on accepte de modifier le seuil de positivité de la bille 39 à plus de 50%.

Au total, la seule bille différenciant A\*02:01:01 de A\*02:01:09 est la bille 39, attendue négative pour le plus fréquent des allèles soit A\*02:01:01. Elle apparaît ici clairement positive avec une valeur à 49 pour un seuil de positivité à 27 (**Figure 31**). L'allèle rare A\*02:01:09 semble donc présent.

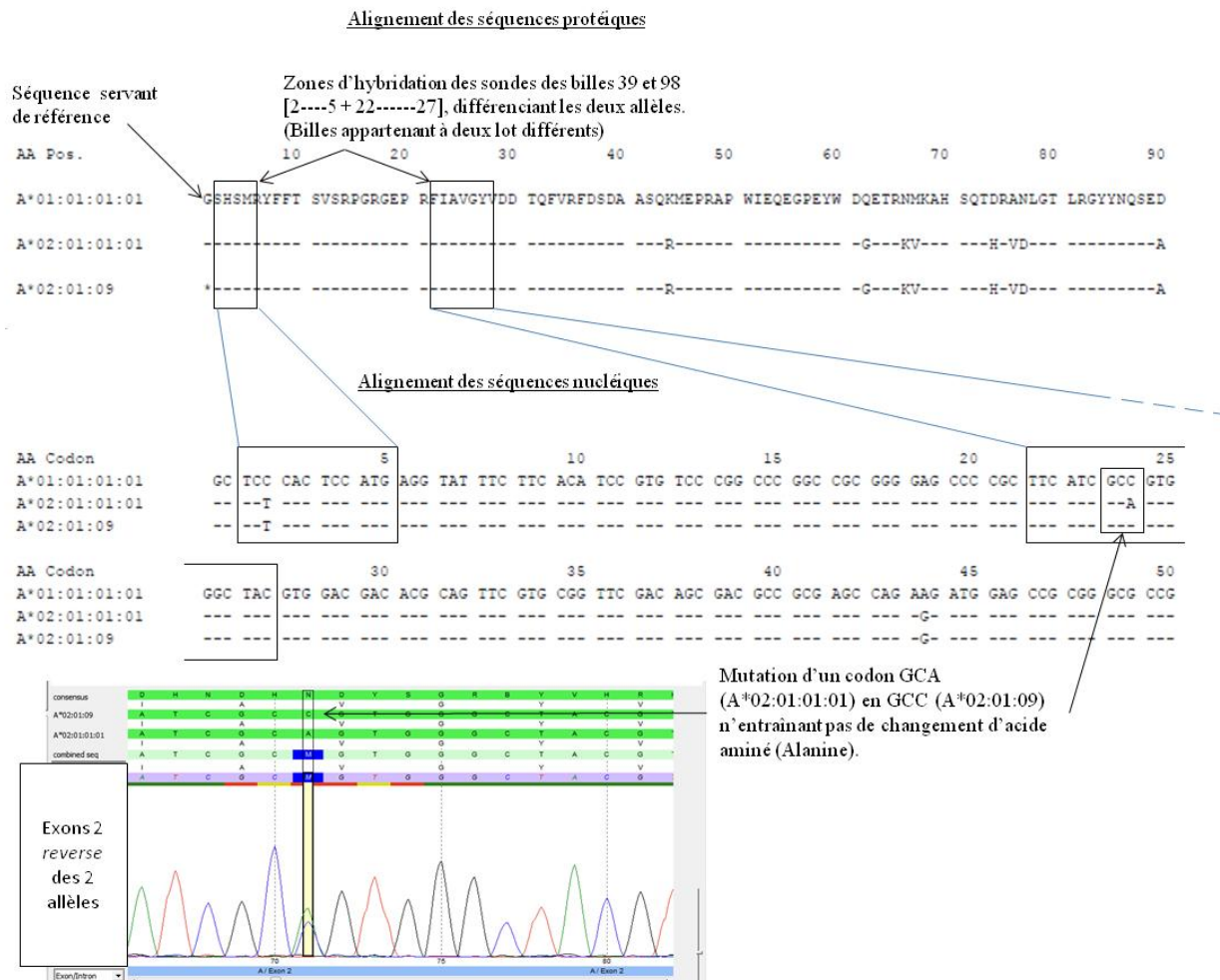


**Figure 31 : Profils des billes 39 et 98, consultables sur le logiciel HLA Fusion®**

La confirmation de cet allèle rare lors du deuxième typage par PCR-SSO n'a pas été possible. Réalisé avec un lot de billes différent, le résultat : A\*02:01:01, A\*02:01:09 nécessitait la diminution du seuil de positivité de la bille 98 de 39 à 38, permettant de la considérer comme positive (**Figure 31**). Cependant, la modification d'un seuil n'est pas autorisée pour l'assignation d'un allèle rare.

La PCR-SSP n'est pas utilisable puisque les kits disponibles ne détectent pas l'allèle A\*02:01:09.

Le typage de contrôle est donc effectué en séquençage SSBT et permet la confirmation du résultat A\*02:01:01, A\*02:01:09 sans aucun problème, par la présence d'une seule mutation dans l'exon 2 (**Figure 32**).



**Figure 32 : Alignement des séquences protéiques et nucléiques des allèles A\*02: 01:01 et A\*02:01:09 selon le site de l'IMGT database. Comparaison au résultat trouvé sur le logiciel JSI® par séquençage monoallélique.**

Ces deux cas illustrent bien la difficulté d'assignation d'allèles rares, que ce soit de l'ordre du 4<sup>ème</sup> ou 6<sup>ème</sup> digit, par les techniques actuelles comme la PCR-SSO, pourtant réputée sensible pour la détection d'allèle rare, et la PCR-SSP. Le séquençage a permis dans les deux cas, sans difficulté, la validation des allèles rares.

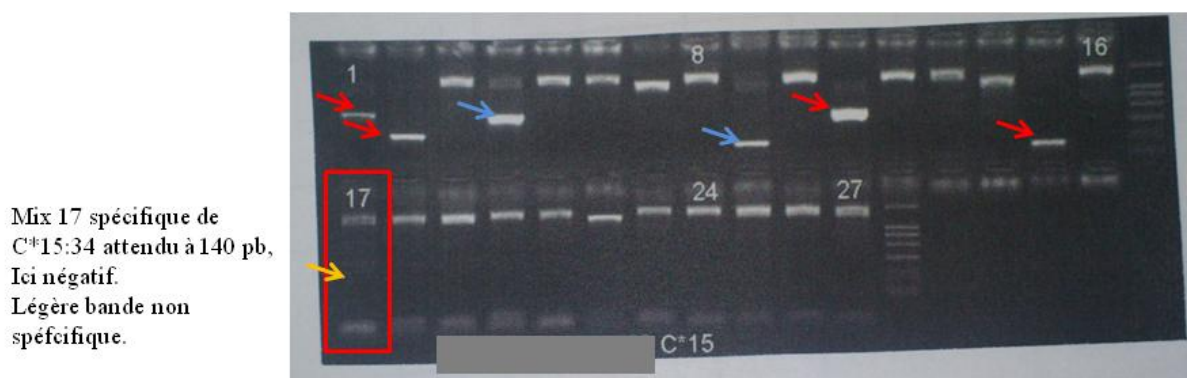
Dans ce second dossier, l'impact de l'ambiguïté de typage est sans conséquence clinique (même protéine), cependant le retard dans la validation des résultats aurait été le même, avec la nécessité d'externalisation.

### III. Discordance entre techniques de biologie moléculaire

#### 1. Cas n°3 : discordance entre PCR-SSP et PCR-SSO

Dans ce dossier, un typage est réalisé chez un receveur de rein, un premier typage fait par PCR-SSO donne le résultat brut C\*02:10 et C\* 15:34, sans signalement d'éventuelles billes faussement positives et/ou négatives.

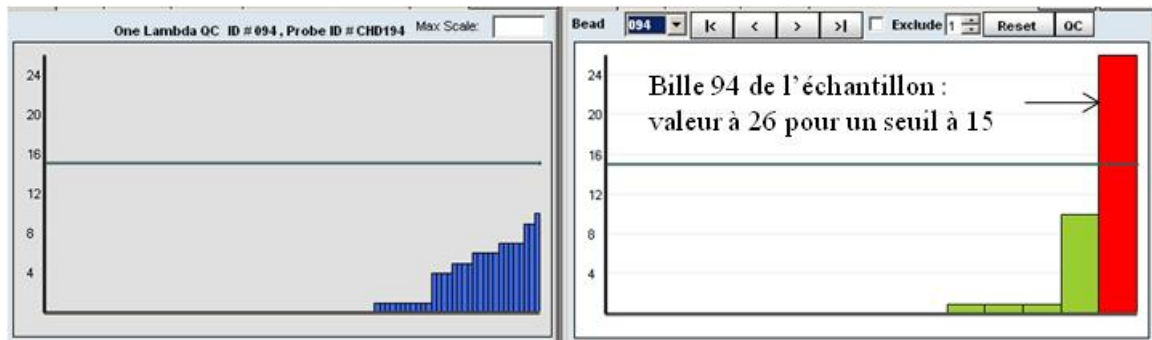
L'allèle C\*02:10 a déjà été retrouvé plusieurs fois au laboratoire avec notamment l'allèle HLA-B\*15:03 présent également chez ce patient. En revanche l'allèle C\*15:34 est un allèle rare jamais trouvé au laboratoire. La confirmation de sa présence est donc recherchée par un typage en PCR-SSP. Un typage de l'allèle C\*15 est donc entrepris avec un kit de PCR-SSP haute résolution (Olerup SSP<sup>®</sup>) afin de vérifier le résultat du premier typage. Le résultat obtenu est présenté dans la **figure 33**, il donne un typage discordant de celui trouvé en PCR-SSO puisqu'il révèle un allèle C\* 15:02:01, attendu d'après les déséquilibres de liaison avec l'allèle HLA-B\*51:01 présent chez ce patient (ce résultat comporte une liste d'ambiguïtés ne comprenant pas C\* 15:34).



**Figure 33 : Photographie du gel d'agarose obtenu en PCR-SSP (Olerup®) pour le typage du locus C, spécificité C\*15.**

Les flèches rouges signalent les bandes spécifiques de l'allèle C\* 15:02:01, les bleues correspondent à l'allèle associé C\*02 et la flèche jaune montre l'absence de bande spécifique au niveau du mix 17, témoignant de l'absence de l'allèle C\* 15:34.

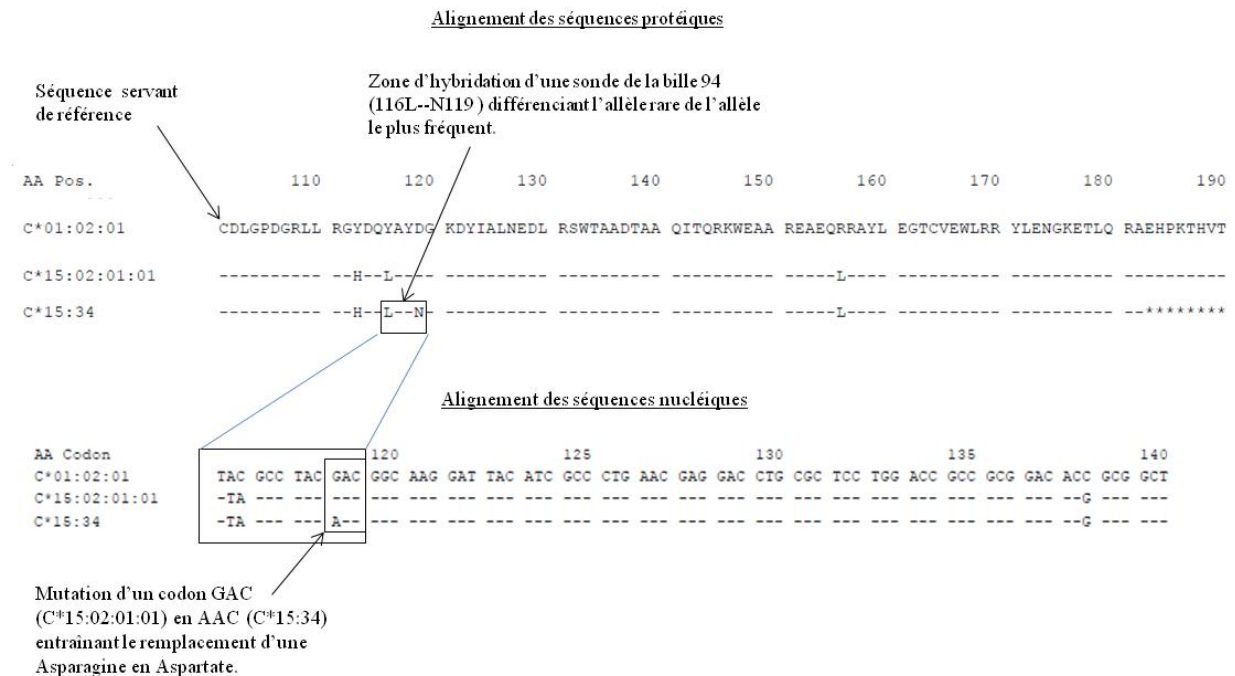
L'allèle C\*15:34 est différencié de C\*15:02:01:01 par la bille 94 (négative pour ce dernier), elle est ici positive dans la technique de PCR-SSO (**Figure 34**).



**Figure 34 : Analyse du profil de la bille 94 sur le logiciel HLA Fusion®.**

Le typage réalisé ensuite en SSBT confirme l'absence de l'allèle C\* 15:34 et permet l'assignation du résultat C\*02:10, C\*15:02:01G sans difficulté.

Dans ce cas, la bille 94 en PCR-SSO, apparaît comme étant faussement positive. La comparaison des alignements de séquence (**Figure 35**) montre la zone de fixation différenciant les deux allèles (C\*15 :02 :01G et C\*15 :34), cependant cette bille possède plusieurs sondes servant à la discrimination d'autres allèles comme l'indique les informations du fournisseur : [112-HD-L-117 + 160---E163] / 144QG-L-148 / 116(-/S)--H119 / 116L--N119 / 106--LL109 / 108-LV110 / 111---K-115 / 112--NRF116 / 113-D-YP--119 / 178---HA182 / 180--AK183 / 101----S105 / 103L-T---108 / 125----N--131 / 128-N---132 / 160-K-T-164 / 161D-E--165 / 108-L-C--113 / 109L-H--D114 / 156L-A-R--T163 / 156--AZ159 / 120---H---127 / [127-----131 + 154--W-A-159]. Cette multitude de sites de fixation pourrait être à l'origine d'une hybridation non spécifique, expliquant le nombre de réactions faussement positives répertoriées pour cette bille dans les relevés statistiques du logiciel Fusion® (0,63% de faux positifs pour la bille 94 du lot concerné).



**Figure 35 : Alignement des séquences protéiques et nucléiques des allèles C\*15: 02:01 et C\*15:34 selon le site de l'IMGT database.**

Ainsi des discordances entre techniques de biologie moléculaire peuvent exister, cette discordance rapportée témoigne ici d'un manque de spécificité de la PCR-SSO.

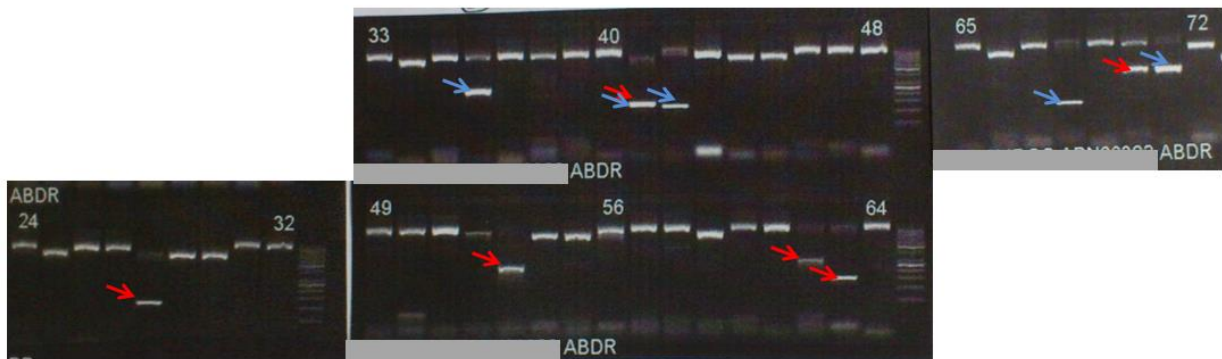
## 2. Cas n°4: discordance entre PCR-SSO, SSP et SSBT

Ce dossier au départ consistait en une recherche de l'allèle HLA-B\*57 :01 dans le cadre de l'administration de l'abacavir chez un patient infecté par le VIH. En effet, avant toute administration d'abacavir chez un patient, il faut vérifier l'absence de l'allèle HLA-B\*57 :01. Son interaction avec le médicament peut être responsable d'une réaction d'hypersensibilité au médicament, potentiellement mortelle [116].

Dans ce contexte, le typage est réalisé en PCR-SSO, mais il aboutit à une impossibilité totale d'assignation de résultat, en raison d'un trop grand nombre de billes faussement négatives.

Une deuxième tentative de typage en PCR-SSO est réalisée sur le même prélèvement. De nouveau, aucun résultat n'est assignable mais un résultat est suggéré : B\*44:02:01:01 B\*44:43:02 en considérant une bille comme faussement positive. Ce résultat n'est pas assignable selon les critères du laboratoire (présence d'un allèle rare associé à une modification trop importante du seuil de positivité d'une bille).

Pour éclaircir la situation, une PCR-SSP générique est utilisée (**Figure 36**).



**Figure 36 : Photographie du gel d'agarose obtenu en PCR-SSP (Olerup®), partie concernant le typage en faible résolution du locus B.**

Les flèches rouges indiquent les bandes spécifiques de l'allèle B\*44, les bleues indiquent les bandes spécifiques de l'allèle B\*35. Il faut souligner que la bande d'amplification au niveau du mix 41 est caractéristique et indispensable pour l'assignation de l'allèle B\*44, cependant ce mix peut être, ou non, pris en compte pour l'assignation du deuxième allèle, il est donc responsable d'une ambiguïté générique. Ainsi, deux combinaisons sont proposées : B\*35 B\*44 ou B\*18, B\*44, cette dernière étant une combinaison avec un allèle rare (B\*18 :29) qui la rend peu probable.

Finalement la PCR-SSP apporte un résultat discordant avec celui de la PCR-SSO, avec deux groupes d'allèles HLA-B distincts et une ambiguïté générique : B\*18, B\*44 ou B\*35, B\*44.

Un troisième typage est réalisé devant ce cas problématique, par SSBT. Une amplification des exons 2 et 3 (dont l'efficacité est vérifiée par migration sur un gel) est réalisée, puis selon les recommandations du fournisseur, le séquençage des exons 2 et 3 est effectué en sens *reverse*.

Le séquençage de l'amplicon de l'exon 2 de l'un des deux allèles à partir de l'amorce *reverse* a échoué. D'après le logiciel JSI (Sequence Pilot™), le résultat obtenu apporte alors une nouvelle ambiguïté générique et une discordance supplémentaire: B\*15:20, B\*44:02 ou B\*35:01:01:01, B\*44:02:01:01. Nous décidons alors de séquencer l'amplicon de l'exon 2 à partir de l'amorce *forward*, qui se révèle efficace, et on obtient cette fois un résultat sans ambiguïté : B\*35:01:01G, B\*44:02:01G.

D'après l'analyse de la PCR-SSO, l'étude *a posteriori* du profil des billes attendues positives pour l'allèle B\*35:01:01:01 permet de constater que les sondes B\*35 devant s'hybrider à l'exon 2 sont négatives, et que seules les sondes B\*35 s'hybridant à l'exon 3 sont positives (**Figure 37**). Ceci conduit à penser qu'il existe une mutation (ou délétion/insertion...) de cet allèle au niveau du site de reconnaissance de l'une des deux amorces utilisées pour amplifier l'exon 2, empêchant son amplification. Cependant, la bille


témoin d'amplification de l'exon 2 (sonde liant un site non polymorphique) montre un signal faible (supérieur au signal seuil défini pour la validation) ce qui laisse penser que l'amplification a eu lieu, mais avec une très faible efficacité.


En séquençage Sanger, l'amplification de cet exon 2 a fonctionné, en revanche c'est l'amorce *reverse* qui n'a pas pu se fixer. Ceci permet d'avoir une idée plus précise du lieu de la mutation, au niveau de l'intron 2 d'après le raisonnement expliqué dans la **figure 38**.

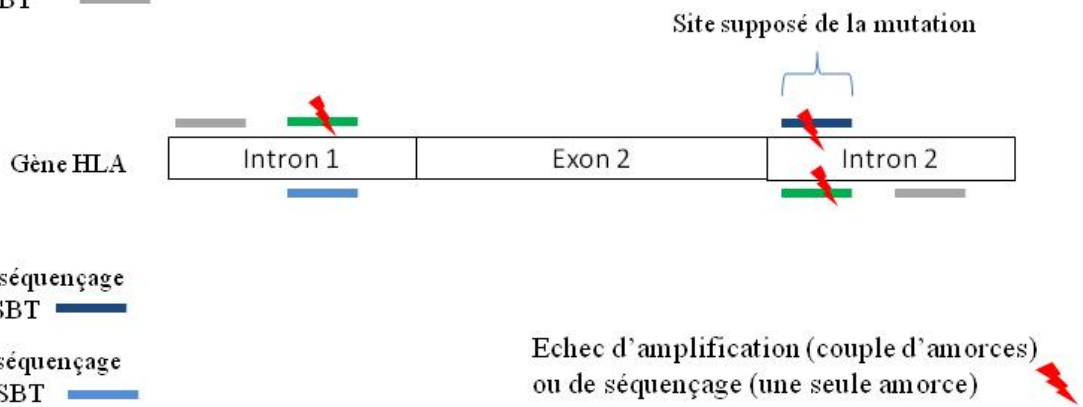
Bille	62	73	87	508	514	518	529	556	588	590	594
Résultat pour l'échantillon	+	-	+	-	-	-	-	+	-	+	-
Exon cible	3	2	3	2	2	2	2	3	2	3	2

**Figure 37 : Résultat des billes informatives pour l'allèle B\*35 :01 :01 (c'est à dire les billes attendues positives pour l'allèle HLA-B\*35 :01 :01 et négatives pour l'allèle B\*44 :02 :01 :01).**

Les sondes B\*35 qui s'hybrident normalement sur l'exon 2 sont négatives, seules les sondes B\*35 s'hybridant sur l'exon 3 sont positives.

Amorces utilisées pour l'amplification de l'exon 2 en PCR SSO 

Amorces utilisées pour l'amplification de l'exon 2 en SBT 



**Figure 38 : Emplacement approximatif et schématique des différentes amorces utilisées en PCR-SSO et en SSBT.**

L'étude de la séquence de l'exon 2 obtenue à partir de l'amorce *forward*, nous permet de déterminer une partie de la séquence de l'intron 2. L'alignement de la séquence de l'intron 2 de notre allèle par rapport à la séquence de l'allèle B\*35:01:01 de la *database* IMGT dévoile une mutation d'une Cytosine en Adénine en position 505 sur l'intron 2. Il s'agit donc d'un nouvel allèle non référencé dans l'IMGT *database*. La mutation se situe effectivement sur les zones de fixation des amorces d'amplification de la PCR-SSO et des amorces de séquençage *reverse*.

Le séquençage de nouvelle génération en *long range PCR* aurait apporté une réponse directe, sans poser de problèmes analytiques dus à cette mutation. Ce dossier montre l'importance de disposer de plusieurs techniques de typage, ou du moins l'importance d'associer plusieurs couples d'amorces qui se fixeront sur différents sites du gène et limiteront le risque de manquer une amplification.

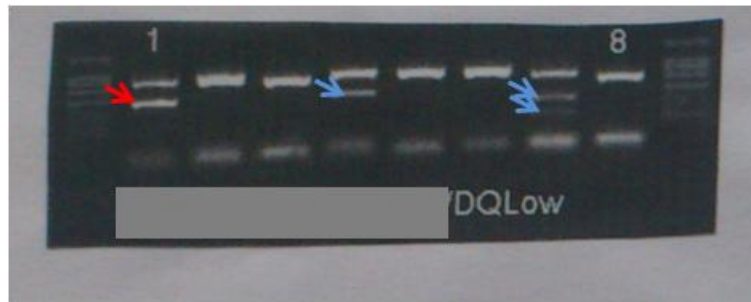
#### IV. Cas n°5 : Discordance sérologie-biologie moléculaire

Dans ce dossier de greffe d'organe, un premier typage a été effectué en sérologie permettant l'assignation en classe II du résultat suivant : DR4, DR14; DQ7.

Dans le cas des dossiers de greffes (organes ou CSH), le typage doit être confirmé par une deuxième technique sur un nouveau prélèvement. Ainsi le deuxième typage, réalisé par PCR-SSO Luminex<sup>®</sup> sur un deuxième prélèvement, révèle alors les résultats les plus probables suivants : DRB1\*04 :01, DRB1\*14 :01; DQB1\*03 :01, DQB1\*05 :03. Pour une meilleure compréhension du dossier, il faut connaître les équivalences de nomenclature entre la biologie moléculaire et la sérologie, à savoir DRB1\*04 :01 équivaut à DR4, DRB1\*14 :01 équivaut à DR14, DQB1\*03 :01 équivaut à DQ7 et DQB1\* 05 :03 donne une image sérologique correspondant à DQ5. Un allèle DQB1\*05 :03 (DQ5) est donc trouvé en SSO alors qu'aucune réaction DQ5 n'est détectée en sérologie.

Suite à cette discordance, la technique de sérologie a été répétée sur le deuxième prélèvement utilisé pour la PCR-SSO. Les résultats obtenus sont identiques aux premiers résultats de sérologie à savoir en classe II : DR4, DR14; DQ7.

Le typage a été dans un premier temps contrôlé en PCR-SSP générique (kit DQ Low, Olerup SSP<sup>®</sup>) (**Figure 39**), où l'on retrouve la preuve de la présence des deux groupes d'allèles DQB1\*03 et DQB1\*05.



**Figure 39 : Photographie du gel d'agarose obtenu en PCR-SSP (Olerup®)**

La flèche rouge désigne la bande spécifique témoignant de la présence d'un allèle du groupe DQB1\*05, les flèches bleues désignent DQB1\*03. La présence de DQB1\*05 ne laisse ici pas de doute possible.

Le contrôle en séquençage Sanger monoallélique (SSBT) est réalisé sur un troisième prélèvement. Les exons 2 et 3 du gène DQB1 ont été séquencés dans les deux sens (*reverse* et *forward*). Le résultat confirme celui trouvé précédemment en biologie moléculaire: DQB1\*03:01:01G, DQB1\*05:03:01G.

Les discordances entre une méthode de sérologie et de biologie moléculaire s'expliquent principalement, après exclusion des erreurs techniques, par la présence d'un allèle nul ou d'expression faible, non mis en évidence en sérologie. Le caractère nul ou faible de l'allèle peut être dû à l'existence d'une mutation, délétion ou insertion entraînant un codon stop, un décalage du cadre de lecture, une modification de la fixation des facteurs de transcription ou encore des problèmes pendant la phase d'épissage. Elles peuvent conduire à la diminution d'expression de la synthèse d'une protéine, à la synthèse d'une protéine aberrante, ou bien à l'absence de protéine, qui ne sera donc pas retrouvée en surface cellulaire par la sérologie. La localisation variable de la mutation explique que les techniques de biologie moléculaire peuvent parfois aboutir à un résultat non aberrant, dans le cas où la mutation est située en dehors des exons 2 et 3 principalement étudiés.

Dans ce cas, un séquençage du gène entier (du 5'UTR au 3'UTR) à partir d'une *long range PCR* en approche *shot gun* permettrait d'identifier la mutation responsable. Il pourrait s'agir d'un nouvel allèle.

Cette *long range PCR* n'est pas envisageable avec le seul kit disponible sur le marché (GenDx®), en effet, l'amplification du gène DQB1 se fait uniquement à partir de l'exon 2 jusqu'à l'exon 4. En revanche, pour Octobre 2015, il est prévu la mise sur le marché d'amorces qui permettront d'amplifier le gène DQB1 en entier, à partir du 5'UTR. A ce moment, nous pourrions envisager le séquençage de cet allèle. Ce cas n'est donc actuellement pas résolu.

Ce type de problème est rencontré assez fréquemment en histocompatibilité. Nous pouvons citer l'exemple de l'allèle HLA-A\*02 :01 :01 :02L qui est caractérisé par une mutation singulière (T →C) présente dans le promoteur du gène HLA-A (*enhancer B*), responsable d'une diminution de son expression. La fixation des facteurs de transcription peut en effet être perturbée par l'existence d'une unique mutation. Le typage de cet allèle en sérologie montre une faible réactivité, tandis que le séquençage mono-allélique des exons 2, 3 et 4 conduit à assigner l'allèle HLA-A\*02 :01 :01 :01. Dans ce cas, seul le séquençage entier du gène permet de résoudre ce typage [117].

Au total, les dossiers sélectionnés ici montrent la fiabilité et l'intérêt du séquençage lorsque deux autres techniques amènent à des résultats discordants ou insuffisants (cas des allèles rares). La mise en place du séquençage a permis un gain de temps en évitant l'externalisation des demandes de typages qui étaient nécessaires dans ces cas.

Les derniers dossiers montrent également les limites de cette technologie et l'intérêt de mettre en place la NGS, avec une approche *long range PCR*. Elle permettrait de mettre en relation l'expression de la molécule avec les polymorphismes des régions 5'UTR et introniques.

## DISCUSSION

Les dossiers étudiés précédemment, représentatifs de nombreux cas, mettent bien en évidence les faiblesses et les limites des techniques de typage HLA traditionnelles (PCR-SSO et PCR-SSP). Elles présentent de nombreux pièges pouvant amener à un résultat erroné, qui pourrait avoir des conséquences gravissimes pour le patient. Ces pièges sont évités par la mise en place de critères de validation stricts, mais aussi par la combinaison de plusieurs techniques différentes de typage. Le séquençage Sanger apporte une aide considérable pour limiter ces erreurs, même s'il n'est pas complètement épargné par ce type de problèmes.

En pratique, le séquençage de nouvelle génération peut être comparé à la technique de *PCR-SSO*, puisqu'il possède l'avantage de permettre la réalisation de plusieurs typages dans une seule série (plus d'une vingtaine d'échantillons typés sur les 6 loci en parallèle). Le séquençage de type Sanger remplacerait plutôt la PCR-SSP car seul le typage d'un patient sur les 5 loci peut être réalisé simultanément. Cependant, le temps de manipulation reste très long dans le cas de la NGS, 2 à 3 jours selon la technologie et avec plusieurs étapes d'interventions manuelles dans le cas du Ion torrent<sup>®</sup>, elle est surtout avantageuse dans le contexte des typages en grande série, comme ceux des DVMO. La résolution obtenue dépend de l'utilisation ou non du principe de *long range PCR* et peut être potentiellement très élevée avec ce dernier. En effet, différentes approches sont possibles, plus ou moins appropriées selon l'application considérée.

Nous allons aborder les différentes questions qui se posent à l'heure où la nouvelle génération de séquençage commence à se développer pour le typage HLA. Quelle approche semble la plus appropriée ? La NGS apporte-elle réellement un intérêt ? Et quelles sont les perspectives envisageables pour optimiser cette technologie, quelles en sont ses limites ?

### I. Quelle approche : *shot gun* ou *exon-sequencing* ?

Comme nous l'avons vu précédemment, il existe deux approches en séquençage qui peuvent concerner le typage HLA.

L'approche *exon-sequencing* consiste à amplifier les exons d'intérêt (2 ± 3 et 4) séparément et rejoint de ce fait en partie la méthode Sanger. Elle demande un temps de préparation moins long que la deuxième approche. En effet, elle ne nécessite pas d'étape de fragmentation (et donc une étape en moins de sélection de taille des fragments). De plus, la possibilité d'utiliser des *primers de fusion*, qui sont des amorces spécifiques d'exon, auxquelles sont préalablement liés des adaptateurs et des identifiants, permet d'obtenir une librairie en une seule étape, dès la fin de la PCR initiale [75]. Ainsi l'approche *exon-sequencing* présente un réel avantage de simplicité et de rapidité. Néanmoins, la multiplicité des amplifications nécessaires dans cette approche lui confère un inconvénient important, à

savoir la quantité d'amorces à synthétiser et valider. Cependant, l'avantage que l'on peut en tirer est le moindre risque de *drop out*. En effet, il paraît peu probable qu'un allèle ne réponde à aucune amplification en utilisant plusieurs couples d'amorces ciblant des régions différentes. En revanche, l'amplification séparée des exons permet de penser que les problèmes de *phasing* ne seront pas résolus avec cette stratégie. L'approche *exon-sequencing* promet plutôt une résolution équivalente à celles des techniques habituelles [93], il semble donc dommage de mettre en place des technologies aussi complexe pour ne pas gagner en résolution. Finalement cette possibilité semble présenter plus d'inconvénients que d'avantages.

L'approche *shot gun* consiste à amplifier dans une première étape le gène d'intérêt entier, du promoteur à l'extrémité 3'UTR, par l'utilisation d'une *long range PCR*. Une fragmentation est ensuite indispensable, elle coupera de façon aléatoire les ADN doubles brins. Après la réparation des extrémités, la préparation de la librairie (fixation des identifiants et adaptateurs) pourra seulement commencer. Cette approche est plus longue que la précédente mais permet d'obtenir beaucoup d'informations sur chaque gène. La *long range PCR* semble être une approche de choix pour pallier à ces difficultés de *phasing*. Ainsi, elle laisse place à beaucoup moins d'ambiguïtés cis/trans [118] car elle permet l'obtention d'amplicons de plus de 10 kilobases et donc d'amplifier la plupart des gènes HLA en entier. Dans le cadre du typage HLA, le kit de la firme GenDx permet le séquençage des gènes HLA-A, B, C, DRB3, DRB4, DRB5 et DQA1 entiers, de l'extrémité 5'UTR jusqu'à l'extrémité 3' UTR. En revanche, les gènes DRB1 et DQB1 ne sont pas amplifiés en entier, mais de l'exon 2 à l'exon 4 inclus, en raison des difficultés à synthétiser des amorces spécifiques des locus DRB1 et DQB1 dans les régions 5'UTR et 3'UTR permettant d'amplifier tous les allèles DRB1 ou DQB1 décrits. Cependant, il existe des amorces décrites dans la littérature permettant l'amplification du 5'UTR au 3'UTR des gènes DRB1 et DQB1 [119]. Les kits utilisant ces amorces devraient être disponibles en Octobre 2015. Quant au gène DPB1, sa taille importante nécessite de « diviser » en deux l'amplification (de l'extrémité 5'UTR jusqu'à l'exon 2, et de l'exon 2 à l'extrémité 3'UTR) [101]. Il est à noter que le locus DRB1 peut également être amplifié à partir d'amorces spécifiques de groupe afin d'éviter les amplifications non spécifiques des gènes DRB3, DRB4, DRB5 [101].

Le séquençage de tous les exons, et même des introns permet en théorie d'atteindre une résolution maximale, au 8<sup>ième</sup> digit. L'intérêt d'une telle résolution sera discuté dans le paragraphe « Intérêt du séquençage ». En outre, cette stratégie permet la découverte de nouveaux allèles à plus grande échelle. La *long range PCR* possède toutefois l'inconvénient d'augmenter les risques de *drop out*, en raison du faible nombre d'amorces utilisées.

## II. Influence de la méthode de fragmentation dans l'approche *shot gun*

La préparation de la librairie est une étape cruciale dont les différentes étapes possèdent une influence sur la qualité des résultats. Si la méthode d'extraction a déjà été discutée dans le chapitre précédent, montrant toute son importance dans la pratique d'une *long range PCR*, la fragmentation peut soulever les mêmes questions. Elle serait même l'une des étapes les plus influentes sur les biais de couverture qui correspondent à des différences de profondeur. La profondeur est en lien étroit avec la sensibilité et la spécificité de détection des allèles [120].

Il existe trois principaux outils, mécaniques ou chimiques, utilisés pour fragmenter des brins d'ADN. D'un point de vue mécanique, la **sonication** utilise les vibrations créées par des ultrasons dans un milieu liquide pour casser les molécules d'ADN de façon aléatoire. La **nébulisation** utilise les forces créées par de l'air ou de l'azote comprimé qui entraîne la rupture des brins d'ADN. La méthode chimique fait intervenir des mélanges d'**enzymes**, qui « grignotent » l'un des deux brins de la molécule d'ADN, le brin opposé est ensuite coupé à son tour.

Dans tous les cas, les extrémités obtenues peuvent ne pas être franches, l'un des deux brins peut être plus court que l'autre, la réparation des extrémités par une enzyme à propriété polymérase, dite *end repair* ou *nick repair*, est nécessaire. Dans les trois situations, les fragments générés sont de taille variable, elle dépend en partie du temps d'incubation. De trop petits fragments entraînent une perte de profondeur de lecture, ils sont favorisés par une incubation longue. Il est donc important d'effectuer une sélection de la taille des fragments avant de les soumettre au séquençage, pour obtenir la taille la plus homogène possible. Cette sélection peut être effectuée en électrophorèse en gel d'agarose ou par l'utilisation de billes magnétiques [120]. Globalement, d'après une étude comparative utilisant la technologie Roche®, menée avec les trois méthodes de fragmentation en parallèle, les résultats de séquençage ne montrent pas de différence significative en terme de qualité. Le choix de la fragmentation peut être porté en priorité sur des critères organisationnels qui conviennent le mieux au laboratoire [121].

Cette étape de fragmentation est suivie de l'indexage des fragments pour leur identification ultérieure. Pour gagner en rapidité, une autre stratégie de fragmentation, utilisant l'approche enzymatique, est proposée pour permettre la cassure des brins, leur réparation et leur indexation en une seule étape. Il s'agit de la **tagmentation**, ou méthode des *transposases*.

D'après une étude comparative utilisant la technologie du Miseq, les méthodes de tagmentation ont montré une profondeur de lecture inférieure à la moyenne notamment dans les zones riches en bases GC en comparaison à une technique classique de sonication [120]. Ces zones correspondent principalement aux régions codant les niches à peptide, et donc aux exons dont la séquence est primordiale pour le résultat de typage (exons 2 et 3 classe I, exon 2 classe II). On comprend pourquoi dans ce cas des biais importants de couverture entraînent

facilement des erreurs de résultat de typage ou des impossibilités d'assignation. Ceci pourrait s'expliquer par le fait que l'intervention des transposases ne serait pas complètement aléatoire ce qui entraînerait la génération de fragments de tailles très différentes selon la séquence nucléotidique de l'allèle et une élimination préférentielle de certains amplicons pendant l'étape de sélection des fragments de tailles convenables. De la même manière, plus l'étape d'amplification clonale comporte de cycles de PCR, plus elle favorise l'amplification préférentielle des fragments sélectionnés. Ceci conduit parfois à des différences de représentation des deux allèles, selon les fragments qui ont été le mieux amplifiés. Si la profondeur de couverture est complètement déséquilibrée entre les deux allèles, l'assignation par le logiciel peut ne pas prendre en compte les fragments minoritaires qui seront classés comme des bruits de fond ou des artefacts. Cela dépend du seuil défini pour la balance allélique, qui sert de critère pour prendre en compte ou non l'assignation de deux allèles. Une profondeur de lecture minimale ne suffit donc pas comme critère, elle doit être bien répartie entre les allèles [120].

D'autres motifs que les bases GC sont corrélés à une perte de profondeur de lecture. C'est le cas des séquences répétées de quelques dizaines d'acides nucléiques, microsatellites ou *Short Tandem repeats*, que l'on trouve par exemple au niveau de la jonction exon 2- intron 2 du gène DRB1 [122].

### III. Intérêt du séquençage de nouvelle génération dans les typages HLA

Alors que le séquençage selon la méthode Sanger nécessite beaucoup de pré-requis sur la séquence à analyser, la NGS permet le séquençage de fragments peu connus grâce à l'utilisation du *shot gun* et d'amorces universelles [123]. Bien que le taux d'erreur d'une technique de NGS soit plus élevé qu'avec la méthode Sanger, la génération de nombreuses lectures (*reads*) qui couvrent une même position permet d'assigner un polymorphisme, même inhabituel, avec une certaine confiance, dont l'importance augmente avec la profondeur de lecture. Ceci confère un avantage majeur à la NGS [102].

Le typage de type Sanger SBT bi-allélique peut aboutir dans de nombreux cas à des résultats multiples qui proposent différentes combinaisons d'allèles. Une étude comparative a permis l'estimation de la réduction des ambiguïtés par la NGS. Ces ambiguïtés proviennent de typages initialement obtenus avec une méthode Sanger SBT bi-allélique. Ainsi pour le locus HLA-B, on obtient 92,7% de réduction, 85,7% pour HLA-A, 100% avec HLA-DQB1, 68,2% pour DPB1, et 46,1% seulement pour HLA-DRB1. L'amélioration semble moins évidente pour ce dernier locus, car son séquençage en méthode Sanger est réalisé à partir d'amplification spécifique de groupe qui revient à une méthode mono-allélique, déjà plus performante que la méthode bi-allélique. De plus le manque d'informations sur l'exon 3 dans la base de données ne permet pas de réduire suffisamment le nombre d'ambiguïtés [106]. Ces problèmes de résultats ont pour origine trois situations majeures :

La première situation concerne les ambiguïtés qui ne peuvent être résolues que par **l'amplification séparée des deux allèles ou bien par l'amplification du gène entier**, pour ne pas engendrer de difficultés de *phasing*. Le séquençage Sanger mono-allélique a déjà permis de réduire considérablement ces ambiguïtés [90]. La NGS avec l'approche utilisant la *long range PCR* semble apporter une solution encore plus prometteuse.

La deuxième situation, relativement fréquente, relève des **polymorphismes situés en dehors des régions classiquement étudiées (exons 2 et 3)**. Ce sont des ambiguïtés qui persistent car les méthodes actuelles ont atteint leurs limites. Elles peuvent donc passer à côté de sites polymorphiques nouveaux, non ciblés par les sondes ou les amorces disponibles dans les kits de typage. Le séquençage de nouvelle génération permet la détermination complète des séquences des gènes [90].

La troisième situation concerne **les allèles dont la séquence n'est pas entièrement connue** dans les régions analysées (par exemple l'allèle A\*01 :01 :02 n'est connu qu'au niveau des exons 2 et 3, A\*01 :01 :01 est connu des exons 1 à 8) [90]. Ils feront partie du même groupe G ce qui n'entraîne finalement pas de problème en terme de résultat actuellement, mais les exigences pourraient changer avec de nouveaux objectifs. La découverte de nouvelles séquences va augmenter avec le déploiement des techniques de NGS. Les bases de données devraient considérablement s'enrichir et ce type d'ambiguïtés diminuer.

Le séquençage des gènes entiers donne accès à une super haute résolution ou résolution allélique, allant jusqu'au 6<sup>ième</sup>, voire 8<sup>ième</sup> digit. L'intérêt de chercher une telle résolution réside principalement dans le domaine de la recherche actuellement. Avec cette approche basée sur le concept du *Whole Gene sequencing*, la possibilité de découvrir de nouvelles séquences avec une importance fonctionnelle est un avantage à côté duquel on pourrait passer avec les méthodes conventionnelles de séquençage partiel [102].

On ne connaît pas encore s'il existe une pertinence clinique à rechercher un résultat de typage au 8<sup>ième</sup> digit. L'intérêt des compatibilités entre donneur et receveur ne sont pas encore étudiées à un tel niveau, bien que les connaissances fonctionnelles des molécules HLA puissent faire penser qu'il n'y ait pas d'intérêt majeur. Toutefois, le séquençage de la région 5'UTR et intronique des gènes HLA pourrait renseigner sur leur niveau d'expression. Un nouvel allèle HLA faiblement ou nullement exprimé n'est actuellement suspecté qu'avec la sérologie, qui n'est plus pratiquée dans tous les laboratoires. Si un allèle de faible expression n'entraîne actuellement pas de différence de stratégie dans une greffe, il n'en est pas de même pour les allèles nuls.

L'incrémentation de la base de données de l'*IMGT database* paraît un premier objectif du séquençage massif, essentiel pour pouvoir utiliser l'application de la haute résolution à bon escient. L'autre objectif est de favoriser et valoriser la sélection de donneurs sur le registre Français, il faut savoir que 80% des donneurs prélevés pour les patients nationaux proviennent de registres extérieurs présentant des typages HLA haute résolution.

Pour les typages de routine en laboratoire, comme le typage des nouveaux DVMO, nécessaire à leur inscription sur le registre, le séquençage haut débit semble parfaitement adapté, puisqu'il s'agit de typage en série. L'approche *exon-sequencing* peut paraître suffisante à condition qu'elle ne laisse pas trop d'ambiguïtés à résoudre. Cependant, la perspective d'inscrire les DVMO avec un typage complet et en haute résolution, pour un coût pas plus élevé que celui des typages actuels, est fortement intéressante. La recherche d'un donneur compatible serait extrêmement facilitée par la présentation des typages en résolution allélique sans ambiguïté et permettrait un gain de temps bénéfique pour les patients. En effet actuellement, les typages des DVMO sont encore incomplets sur les registres, et présentent encore de nombreuses ambiguïtés (voire un niveau de faible résolution) qui nécessitent des demandes de typages complémentaires. Elles constituent une perte de temps considérable.

#### IV. Limites de la NGS

Ces nouvelles technologies peuvent encore réserver quelques pièges. Les problèmes les plus rencontrés proviennent de la conception des amorces qui sont un vrai enjeu pour la réussite du séquençage. Elles doivent être hautement spécifiques, ce qui est complexe dans un système comme le CMH où les gènes conservent une grande homologie entre eux et un grand polymorphisme chacun. Parfois des co-amplifications de pseudogènes ou des recombinaisons *in vitro* peuvent constituer des séquences parasites et entraîner une perte de profondeur. Le développement des amorces doit donc tenir compte de ces difficultés pour ne passer à côté d'aucun allèle.

Ainsi, il peut exister des recombinaisons de gènes *in vitro*, également appelées *cross over*, principalement entre les gènes DRB1 et les autres gènes DRB (DRB3, DRB4, DR5). Elles engendrent des artefacts d'amplification (**Figure 40**). Ceci survient plutôt en fin de PCR et est favorisé par une augmentation du nombre de cycles. En d'autres termes il s'agit d'une recombinaison des deux gènes, simulée par un mauvais appariement de l'amorce [75].

Certains logiciels sont capables de faire la différence entre les artefacts de recombinaisons et des allèles proches. Ainsi ces *reads* ne créent pas de bruit de fond, contrairement à la méthode Sanger, mais résultent en une quantité variable de données inutilisées pour l'assignation.

DRB1*03:01:01 <i>Forward</i>	<pre> TTTCTTGGAGTACTCTACGTCTGAGTGTCAATTTCTTCAATGGGACGGAGCGGGTGCGGTA CCTGGACAGATACTTCCATAACCAGGAGGAGAAACGTGCGCTTCGACAGCGACGTGGGGG AGTTCCGGGCGGTGACGGAGCTGGGGCGGCCTGATGCCGAGTACTGGAACAGCCAGAA GGACCTCCTGGAGCAGAAGCGGGGCCGGTGGACAATACTACTGCAGACACAATACTACGG GTTGTG </pre>
DRB1*03:42 <i>Forward</i>	<pre> TTTCTTGGAGCTGCGTAA/GTCTGAGTGTCAATTTCTTCAATGGGACGGAGCGGGTGCGGTA CCTGGACAGATACTTCCATAACCAGGAGGAG/AACGTGCGCTTCGACAGCGACGTGGGG GAGTTCCGGGCGGTGACGGAGCTGGGGCGGCCTGATGCCGAGTACTGGAACAGCCAGA AGGACCTCCTGGAGCAGAAGCGGGGCCGGTGGACAATACTACTGCAGACACAATACTACGG GGTTGTG </pre>
DRB3*01:01:02 <i>Forward</i>	<pre> TTTCTTGGAGCTGCGTAAAGTCTGAGTGTCAATTTCTTCAATGGGACGGAGCGGGTGCGGTA CCTGGACAGATACTTCCATAACCAGGAGGAGTTCTGCGCTTCGACAGCGACGTGGGGG AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGTCGCCGAGTCTGGAACAGCCAGAA GGACCTCCTGGAGCAGAAGCGGGGCCGGTGGACAATACTACTGCAGACACAATACTACGG GTTGGT </pre>
DRB3*01:14 <i>Forward</i>	<pre> TTTCTTGGAGTACTCTAC/GTCTGAGTGTCAATTTCTTCAATGGGACGGAGCGGGTGCGGTA CCTGGACAGATACTTCCATAACCAGGAGGAG TTCTGCGCTTCGACAGCGACGTGGGGG AGTACCGGGCGGTGACGGAGCTGGGGCGGCCTGTCGCCGAGTCTGGAACAGCCAGAA GGACCTCCTGGAGCAGAAGCGGGGCCGGTGGACAATACTACTGCAGACACAATACTACGG GTTGGT </pre>

**Figure 40 : Reads issus d'artefacts de PCR, recombinaisons *in vitro* ou *cross-over* (exemple du gène *DRB1* et du gène *DRB3*).**

D'après H. Erlich [75].

Les *reads* hybrides issus d'une recombinaison peuvent correspondre à la séquence d'un autre allèle répertorié dans la base de données de l'IMGT. La séquence en vert représente celle du gène *DRB1*, celle en rouge correspond au gène *DRB3*, et celle en bleu représente la séquence commune aux deux gènes (zone où a eu lieu le *cross-over*). Les séquences hybrides obtenues correspondent aux allèles *DRB1\*03 :42* et *DRB3\*01 :14*.

Le séquençage a également des limites dans la découverte des nouveaux allèles, ou même dans la détection d'allèles connus, qui auraient subi une mutation ponctuelle chez un individu (contexte d'hémopathie par exemple). Nous l'avons vu, un *drop out* peut avoir lieu, notamment avec la *long range PCR*. Dans ce cas un résultat peut apparaître comme homozygote alors qu'il s'agit d'une amplification manquée d'un des deux allèles chez un hétérozygote [105]. Les conséquences d'une telle erreur de typage seraient dramatiques. Ce problème peut être limité par l'utilisation de techniques complémentaires, basées sur des amorces ou des sondes qui s'hybrident sur un nombre de sites le plus varié possible.

La mise en place d'une technique de séquençage de nouvelle génération en routine est une démarche fastidieuse. Elle demande une validation ou vérification de méthode complexe car de nombreux points sont à aborder. La formation du personnel est également compliquée. Il faut également prévoir des dispositions importantes au niveau du stockage des données informatiques, qui est devenu le principal facteur limitant d'une telle technologie. A chaque *run* sont accumulées des quantités impressionnantes de données. Il faut avoir la capacité de les archiver, et seules des solutions onéreuses sont proposées par les fournisseurs.

## V. Intérêt du séquençage de l'ARN ?

Cette approche consiste à séquencer les molécules d'ARN messager extraites d'un tissu. Nous avons évoqué cette possibilité dans le chapitre dédié à l'histoire du séquençage, avec la nécessité d'effectuer une rétrotranscription pour convertir l'ARN en ADN complémentaire (ADNc). Il subira ensuite les mêmes étapes de fragmentation et ligation d'adaptateurs pour obtenir une librairie prête au séquençage. Les molécules d'ARN d'un type cellulaire ne reflètent pas totalement la séquence génétique de l'ADN qu'il renferme puisque tous les gènes ne sont pas nécessairement transcrits et le niveau d'expression des gènes est variable selon la spécialisation cellulaire. Ainsi, il faudrait travailler sur les lymphocytes B pour obtenir les typages des molécules de classe II. De plus, seuls les exons d'un gène peuvent être déduits du séquençage de l'ADNc.

Cependant, la résolution obtenue en RNAseq reste trop faible pour l'utiliser dans le cadre de la greffe de moelle osseuse. Si des résultats tout à fait satisfaisants sont obtenus en 2 digit, la résolution 4 digit reste à améliorer, alors qu'on parle déjà de 8 digit en travaillant sur de l'ADN [124]. La nécessité d'obtenir des séquences introniques est un argument pour encourager le développement de la NGS, il serait donc dommage de se contenter de l'ADNc, excepté dans le cas de typage en faible résolution à haut débit.

Travailler sur la molécule d'ARN reste néanmoins une idée très intéressante pour les futures générations de séquençage (3<sup>ième</sup>) en cours de développement qui permettraient de séquencer directement une molécule d'ARNm. Ainsi avec une séquence relativement courte d'ARN (1 à 1,5 kB), on pourrait réaliser le typage des exons en une seule fois sans le fragmenter et ainsi résoudre totalement les ambiguïtés de phase.

## VI. Intérêt d'élargir le séquençage à d'autres familles de gènes

Les techniques de séquençage global permettront un gain de temps mais aussi un gain financier pour les typages nécessaires en routine. Il sera donc envisageable de s'intéresser, dans l'avenir, à d'autres gènes, appartenant au CMH ou non, comme les gènes HLA-E, G, MICA, MICB ou encore ceux des cytokines pro-inflammatoires. Agrandir le périmètre de recherche peut amener à découvrir de potentielles relations entre ces gènes et la survenue d'une réaction du greffon contre l'hôte. L'étude des gènes entiers, avec la séquence de fixation des facteurs de transcription, permettrait de déduire le niveau d'expression des gènes. Ceci a toute son importance dans le cas des cytokines, qui nous l'avons vu, forment l'orage cytokinique qui génère et entretient la GvH. Une adaptation du traitement préventif immunosuppresseur selon le profil génétique du patient sera peut être envisageable. Plusieurs études ont déjà été menées sur différents paramètres, qui nécessitent des techniques de séquençage pour être explorés à grande échelle. Par exemple, il a été constaté que la présence de l'antigène HLA-DRB1\*15 chez des patients greffés semblait être associée à un moindre risque de rechute. Cette observation est expliquée par la transmission haplotypique de HLA-DRB1 et du gène du TNF, ce dernier étant localisé à proximité dans la région HLA de classe III. Il existe ainsi un déséquilibre de liaison entre HLA-DRB1\*15 et un polymorphisme particulier du gène du TNF. L'impact clinique de ce polymorphisme, situé dans le promoteur du gène, serait une diminution du risque de GvH aiguës. Ceci suggère que d'autres variations spécifiques d'haplotypes pourraient exister, leur connaissance entraînerait une prise en charge différente pour la transplantation [125]. Ceci nous amène également à une autre question qui est l'importance de la connaissance des haplotypes du couple donneur/receveur lors d'une greffe. En effet, en dehors des allèles des molécules HLA situés sur un même haplotype, des SNPs spécifiques d'un haplotype pourraient avoir un rôle important dans le mécanisme de GvH. Ainsi, rechercher une compatibilité supplémentaire, sur le plan des haplotypes, permettrait de limiter le nombre de SNPs entre deux patients et réduirait potentiellement le risque de GvH. La troisième génération de séquençage semble plus favorable à la résolution des haplotypes par les longueurs de fragments qu'il semble possible de séquencer [126].

Dans le cas des greffes d'organe, l'étude du polymorphisme des cytokines pourrait également se révéler très utile pour l'adaptation des doses de traitement immunosuppresseur. Des techniques non invasives sont actuellement en cours d'évaluation pour la détermination du risque de rejet aiguë suite à une greffe d'organe. Elles sont basées sur l'étude de l'expression des gènes, avec l'utilisation d'une PCR quantitative sur les ARNm d'un échantillon sanguin. La technique kSORT (Kidney Solid Organ Response Test) étudie 17 gènes. Ceci permettrait de limiter les biopsies réalisées dans ce contexte. Dans l'avenir, l'étude de l'expression des gènes pourrait peut être se réaliser par le séquençage de leur extrémité 5'UTR, et permettre l'évaluation *a priori*, du risque de rejet aiguë [127].

La découverte d'une importance clinique de nouveaux gènes ou de leur expression va de paire avec une augmentation de la complexité du choix d'un donneur non apparenté. Des

nouvelles priorités seront peut être établies prochainement, ces perspectives restent à découvrir...

# CONCLUSION

Ce travail avait pour objectif d'aider à la réflexion pour la mise en place de nouvelles techniques de typage HLA, en particulier dans le cadre de l'allogreffe de CSH, compte tenu de l'apparition des nouvelles technologies, regroupées sous le sigle NGS pour *Next generation sequencing* ou séquençage de nouvelle génération.

Dans ce travail, nous avons d'abord rappelé l'importance du système HLA en transplantation et souligné les particularités de l'allogreffe de CSH, particulièrement en matière de choix du donneur et de réaction du greffon contre l'hôte.

Nous avons dressé un état des lieux des techniques actuellement utilisées au laboratoire d'histocompatibilité du CHRU de Nancy, leur niveau de résolution et leurs limites respectives : lymphocytotoxicité (LCT), PCR-SSO, PCR-SSP, séquençage mono-allélique selon la méthode de Sanger.

Puis, nous avons exposé les nouvelles technologies NGS avec les trois principales techniques de séquençage de 2<sup>ème</sup> génération actuellement disponibles (GS Junior 454 de Roche<sup>®</sup>, PGM Ion Torrent de Life technology<sup>®</sup> et Miseq d'Illumina<sup>®</sup>) et leurs qualités respectives de résultats. Nous avons également montré les perspectives nouvelles avec le développement en cours des techniques de séquençage de 3<sup>ème</sup> génération.

La mise en place des techniques de séquençage est de nos jours incontournable dans un laboratoire d'histocompatibilité typant des donneurs de moelle et affilié à un centre d'allogreffe de CSH, de par les exigences de l'Agence de la Biomédecine pour le typage des nouveaux donneurs de moelle d'une part, de par les standards EFI traitant de l'allogreffe de CSH avec donneur non apparenté d'autre part. C'est pourquoi, nous avons souhaité faire le point sur l'apport de la technique de séquençage mono-allélique selon la méthode de Sanger depuis son introduction au laboratoire (2014).

Nous avons montré que le niveau de typage haute résolution atteint avec cette technique allait de 92 à 100% selon les loci A, B, C, DRB1 et DQB1 (contre 25 à 75% précédemment, sans l'utilisation de cette technique).

A l'aide de cinq cas sélectionnés en raison d'un problème d'interprétation de typage en utilisant la LCT, la PCR-SSO et la PCR-SSP, nous avons montré que le séquençage mono-allélique selon la méthode de Sanger permettait, sans difficulté :

- soit d'assigner un allèle rare dans deux cas (DRB1\*13:15, A\*02:01:09)
- soit d'assigner un résultat dans un cas (C\*02:10, C\*15:02:01G)
- soit, dans un autre cas, de confirmer l'existence d'une mutation dans l'intron 2, (suspectée lors du typage par PCR-SSO) ne permettant pas la fixation d'une amorce de PCR-SSO et de SSBT *reverse* et expliquant les discordances de typage selon les techniques (il s'agit donc d'un nouvel allèle à soumettre à l'*IMGT database*)
- le 5<sup>ème</sup> cas fait envisager un allèle DQB1 nul ou un nouvel allèle DQB1 mais, en l'absence de possibilité actuelle de séquençage global du gène DQB1 - du 5'UTR au 3'UTR – ce cas n'est pas actuellement résolu.

La discussion a porté sur les techniques de séquençage de nouvelle génération avec leurs apports, leurs pièges et limites, leurs contraintes (en particulier informatiques) et les nouvelles approches qui conviendraient mieux pour le typage HLA (*long range PCR*). Enfin, la discussion a été ouverte sur l'intérêt de typer d'autres gènes que les gènes HLA dans le cadre de la prévention du risque de rejet ou du risque de GVH dans le cadre de l'allogreffe de CSH.

**En conclusion**, la nécessité de mettre en place le séquençage NGS dans le laboratoire d'histocompatibilité du CHRU de Nancy n'est plus à discuter : contraintes institutionnelles (ABM), standards EFI, harmonisation des pratiques, typages à haut débit de grandes séries, plus faible coût. Cette mise en place n'est pas sans difficulté. Certaines étapes présentent des risques d'erreur importants (comme la préparation de la librairie) et nécessitent la plus complète automatisation. Les logiciels d'analyse des données et le stockage des multiples données générées doivent être pris en charge par des bio-informaticiens. Enfin, comme pour toute technique utilisée dans un laboratoire de biologie médicale, des critères rigoureux de qualité doivent être définis lors de la période de validation de la technique par le laboratoire en tenant compte des standards EFI qui se mettent en place pour le séquençage NGS et qui se préciseront au cours des mois à venir.

## BIBLIOGRAPHIE

- [1] J. Klein, “George Snell’s first foray into the unexplored territory of the major histocompatibility complex” *Genetics*, vol. 159, no. 2, pp. 435–439, 2001.
- [2] E. D. Carosella, “From MAC to HLA” *Hum. Immunol.*, vol. 70, no. 9, pp. 661–662, Sep. 2009.
- [3] R. A. Gibbs, J. W. Belmont, P. Hardenbol, T. D. Willis, F. Yu, H. Yang, L.-Y. Ch’ang, W. Huang, B. Liu, Y. Shen, and others “The international HapMap project,” *Nature*, vol. 426, no. 6968, pp. 789–796, 2003.
- [4] The sequencing MHC consortium, “Complete sequence and gene map of a human major histocompatibility complex” *Nature*, vol. 401, 1999.
- [5] J. Proll, M. Danzer, S. Stabentheiner, N. Niklas, C. Hackl, K. Hofer, S. Atzmuller, P. Hufnagl, C. Gully, H. Hauser, O. Krieger, and C. Gabriel, “Sequence Capture and Next Generation Resequencing of the MHC Region Highlights Potential Transplantation Determinants in HLA Identical Haematopoietic Stem Cell Transplantation” *DNA Res.*, vol. 18, no. 4, pp. 201–210, Aug. 2011.
- [6] S. Takuno, T. Nishio, Y. Satta, and H. Innan, “Preservation of a Pseudogene by Gene Conversion and Diversifying Selection” *Genetics*, vol. 180, no. 1, pp. 517–531, Aug. 2008.
- [7] J. Dausset and M. Pla, *HLA complexe majeur d’histocompatibilité de l’Homme*, Flammarion. 1993.
- [8] “Statistics < IMGT/HLA < IPD < EMBL-EBI.” .
- [9] European Federation for Immunogenetics, “Standards for histocompatibility and immunogenetics testing” 2013.
- [10] L. Pichon, T. Giffon, B. Chauvel, J.-Y. Le Gall, and V. David, “La région HLA de classe I du CMH: une des régions les plus complexes du génome humain?” 1996.
- [11] V. M. Braud, D. S. Allan, and A. J. McMichael, “Functions of nonclassical MHC and non-MHC-encoded class I molecules” *Curr. Opin. Immunol.*, vol. 11, no. 1, pp. 100–108, 1999.
- [12] M. A. Pabón, C. E. Navarro, J. C. Osorio, N. Gómez, J. P. Moreno, A. F. Donado, H. C. Pérez, and E. Lozano, “Impact of Human Leukocyte Antigen Molecules E, F, and G on the Outcome of Transplantation” *Transplant. Proc.*, vol. 46, no. 9, pp. 2957–2965, Nov. 2014.
- [13] A. L. Hughes, “Origin and evolution of HLA class I pseudogenes” *Mol. Biol. Evol.*, vol. 12, no. 2, pp. 247–258, 1995.

- [14] A. Maffei, K. Papadopoulos, and P. E. Harris “MHC class I antigen processing pathways,” *Hum. Immunol.*, vol. 54, no. 2, pp. 91–103, 1997.
- [15] R. A. Goldsby, T. J. Kindt, and B. A. Osborne, *Immunologie : Le cours de Janis Kuby*, Dunod. p.182, 2000.
- [16] P. Le Bouteiller and F. Lenfant, “Gène HLA-G: le plus classique des non classiques” 1997.
- [17] P. J. van den Elsen, T. M. Holling, H. F. Kuipers, and N. van der Stoep, “Transcriptional regulation of antigen presentation” *Curr. Opin. Immunol.*, vol. 16, no. 1, pp. 67–75, Feb. 2004.
- [18] A.-M. Fillet and B. Sénéchal, “Infection à cytomégalovirus” *John Libbey Eurotext*, vol. 6, no. 1, pp. 46–65, 2000.
- [19] E. Hosseini, A. P. Schwarzer, and M. Ghasemzadeh, “Do human leukocyte antigen E polymorphisms influence graft-versus-leukemia after allogeneic hematopoietic stem cell transplantation?” *Exp. Hematol.*, Nov. 2014.
- [20] N. Lee, A. Ishitani, and D. E. Geraghty, “HLA-F is a surface marker on activated lymphocytes” vol. 40, no. 8, pp. 2308–2318, 2010.
- [21] M. A. Pabón, C. E. Navarro, J. C. Osorio, N. Gómez, J. P. Moreno, A. F. Donado, H. C. Pérez, and E. Lozano, “Impact of Human Leukocyte Antigen Molecules E, F, and G on the Outcome of Transplantation” *Transplant. Proc.*, vol. 46, no. 9, pp. 2957–2965, Nov. 2014.
- [22] A. Maffei and P. E. Harris, “Peptides bound to major histocompatibility complex molecules” *Peptides*, vol. 19, no. 1, pp. 179–198, 1998.
- [23] K. L. Rock and L. Shen, “Cross-presentation: underlying mechanisms and role in immune surveillance” *Immunol. Rev.*, vol. 207, no. 1, pp. 166–183, 2005.
- [24] J.-F. Eliaou, “Immunité adaptative : structures reconnues (CMH et antigènes)” 2007.
- [25] “HLA Nomenclature @ hla.alleles.org” [Online]. Available: <http://hla.alleles.org/genes/index.html>. [Accessed: 22-Feb-2015].
- [26] J. P.-Y. Ting and J. Trowsdale, “Genetic control of MHC class II expression” *Cell*, vol. 109, no. 2, pp. S21–S33, 2002.
- [27] D. Russell, “A binding site for the TCR co-receptor CD8 and the alpha3 domain of HLA-A2” *Nature*, vol. 345, 1990.
- [28] M. Labalette, “Le complexe majeur d’histocompatibilité humain (HLA)” .

- [29] L. Klein, B. Kyewski, P. M. Allen, and K. A. Hogquist, “Positive and negative selection of the T cell repertoire: what thymocytes see (and don’t see)” *Nat. Rev. Immunol.*, vol. 14, no. 6, pp. 377–391, May 2014.
- [30] T. K. Starr, S. C. Jameson, and K. A. Hogquist, “Positive and negative selection of T cells” *Annu. Rev. Immunol.*, vol. 21, no. 1, pp. 139–176, Apr. 2003.
- [31] A. P. Brézin, D. Monnet, J. H. M. Cohen, and R. D. Levinson, “HLA-A29 and Birdshot Chorioretinopathy” *Ocul. Immunol. Inflamm.*, vol. 19, no. 6, pp. 397–400, Dec. 2011.
- [32] “HLA Nomenclature @ hla.alleles.org” [Online]. Available: <http://hla.alleles.org/nomenclature/stats.html>. [Accessed: 01-May-2015].
- [33] T. Shiina, S. Suzuki, Y. Ozaki, H. Taira, E. Kikkawa, A. Shigenari, A. Oka, T. Umemura, S. Joshita, O. Takahashi, Y. Hayashi, M. Paumen, Y. Katsuyama, S. Mitsunaga, M. Ota, J. K. Kulski, and H. Inoko, “Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers: Super high-resolution DNA typing of HLA loci” *Tissue Antigens*, vol. 80, no. 4, pp. 305–316, Oct. 2012.
- [34] “ImmPort: Immunology Database and Analysis Portal” [Online]. Available: <https://immport.niaid.nih.gov/immportWeb/home/home.do?loginType=full>. [Accessed: 15-May-2015].
- [35] S. J. Mack, P. Cano, J. A. Hollenbach, J. He, C. K. Hurley, D. Middleton, M. E. Moraes, S. E. Pereira, J. H. Kempenich, E. F. Reed, M. Setterholm, A. G. Smith, M. G. Tilanus, M. Torres, M. D. Varney, C. E. M. Voorter, G. F. Fischer, K. Fleischhauer, D. Goodridge, W. Klitz, A.-M. Little, M. Maiers, S. G. E. Marsh, C. R. Müller, H. Noreen, E. H. Rozemuller, A. Sanchez-Mazas, D. Senitzer, E. Trachtenberg, and M. Fernandez-Vina, “Common and well-documented HLA alleles: 2012 update to the CWD catalogue: CWD 2.0.0 catalogue” *Tissue Antigens*, vol. 81, no. 4, pp. 194–203, Apr. 2013.
- [36] M. D. Carvalho, C. Baron, G. Blanco, M. Labalette, H. Moins, G. Sterkers, and P. Tiberghien, “Mécanismes de l’alloréactivité, des rejets de greffe et de la réaction du greffon contre l’hôte” .
- [37] P. Martin, “La greffe de cellules souches hématopoïétiques” no. 20, pp. 16–28, 2009.
- [38] C. Pautas, Y. Hicheri, and K. Debbache, “Place de l’allogreffe de cellules souches hématopoïétiques dans le traitement des hémopathies malignes de l’adulte” *RFL*, no. 395, 2007.
- [39] J.-H. Dalle, “L’allogreffe de cellules souches hématopoïétiques en 2012 : Pour qui? Comment? Dans quelles conditions?” *Elsevier Masson*, 2013.
- [40] “Agence de la biomédecine - rapport médical et scientifique” [Online]. Available: <http://www.agence-biomedecine.fr/annexes/bilan2010/donnees/cellules/01-greffe/synthese.htm#t2>. [Accessed: 17-Jan-2015].

- [41] J. D. Rizzo, S. Williams, J. T. Wu, A. L. Pecora, H. M. Lazarus, B. Bolwell, K. K. Fields, R. P. Gale, G. Elfenbein, M. M. Horowitz, and K. H. Antman, “Syngeneic hematopoietic stem cell transplantation for women with metastatic breast cancer” *Bone Marrow Transplant.*, vol. 32, no. 2, pp. 151–155, Jul. 2003.
- [42] A. J. Barrett, K. Rezvani, S. Solomon, A. M. Dickinson, X. N. Wang, G. Stark, H. Cullup, M. Jarvis, P. G. Middleton, and N. Chao, “New developments in allotransplant immunology” *ASH Educ. Program Book*, vol. 2003, no. 1, pp. 350–371, 2003.
- [43] O. Hequet, “Les différentes modalités de prélèvement des cellules souches hématopoïétiques” *Transfus. Clin. Biol.*, vol. 18, no. 2, pp. 230–234, Apr. 2011.
- [44] C. Leberre, “Procédure prélèvement de moelle osseuse au bloc opératoire, CHU de Renne” 2008.
- [45] M. Michallet, “Allogreffes de cellules souches hématopoïétiques dans le traitement des hémopathies malignes” vol. 88, no. 9, pp. 908–26, 2011.
- [46] P. Tiberghien, J.-Y. Cahn, and P. Hervé, “Modulation de la réactivité allogénique après la greffe de cellules souches hématopoïétiques” 1997.
- [47] P. Martin, “La réaction du greffon contre l’hôte” no. 20, pp. 29–34, 2009.
- [48] J. Koreth and J. H. Antin, “Current and future approaches for control of graft-versus-host disease” 2008.
- [49] J.-Y. Cahn, J. P. Klein, S. J. Lee, N. Milpied, D. Blaise, J. H. Antin, V. Leblond, N. Ifrah, J.-P. Jouet, F. Loberiza, and others, “Prospective evaluation of two acute graft-versus-host (GVHD) grading systems: A joint Societe Francaise de Greffe de Moelle et Therapie Cellulaire (SFGM-TC), Dana Farber Cancer Institute (DFCI) and International Bone Marrow Transplant Registry (IBMTR) prospective study” *Blood*, 2005.
- [50] B. E. Shaw, R. Arguello, C. A. Garcia-Sepulveda, and J. A. Madrigal, “The impact of HLA genotyping on survival following unrelated donor haematopoietic stem cell transplantation” *Br. J. Haematol.*, vol. 150, no. 3, pp. 251–258, Jun. 2010.
- [51] A. M. Dickinson and D. Charron, “Non-HLA immunogenetics in hematopoietic stem cell transplantation” *Curr. Opin. Immunol.*, vol. 17, no. 5, pp. 517–525, Oct. 2005.
- [52] A. Dickinson and P. Middleton, “Beyond the HLA typing age: Genetic polymorphisms predicting transplant outcome” *Blood Rev.*, vol. 19, no. 6, pp. 333–340, Nov. 2005.
- [53] A. M. Dickinson, “Risk assessment in haematopoietic stem cell transplantation: Pre-transplant patient and donor factors: non-HLA genetics” *Best Pract. Res. Clin. Haematol.*, vol. 20, no. 2, pp. 189–207, Jun. 2007.
- [54] M.-G. Roncarolo and M. Battaglia, “Regulatory T-cell immunotherapy for tolerance to self antigens and alloantigens in humans” *Nat Rev Immunol*, vol. 7, no. 8, pp. 585–598, Aug. 2007.

- [55] W. D. Shlomchik, “Graft-versus-host disease” *Nat Rev Immunol*, vol. 7, no. 5, pp. 340–352, May 2007.
- [56] AFSSAPS, “Transfusion de globules rouges homologues : produits, indications, alternatives” 2002.
- [57] M. S. S. Kotschote and others, “Next-generation Sequencing” 2011.
- [58] E. D. Thomas and R. B. Epstein, “Bone marrow transplantation in acute leukemia” *Cancer Res.*, vol. 25, no. 9 Part 1, pp. 1521–1524, 1965.
- [59] K. Hughes-Jones, S. Selwyn, and P. G. Riches, “Who pioneered the use of alternative donors (and stem cells from peripheral blood) in bone marrow transplantation?” *Arch. Dis. Child.*, vol. 66, no. 9, p. 1102, 1991.
- [60] G. Fillet, J. Bury, and M. Baudrihaye, “Allogreffes et autogreffes de moelle osseuse, principes et indications” *Rev. Médicale Liège*, vol. XLII, no. 10, pp. 503–509, 1987.
- [61] “Registre France Greffe de Moelle - Agence de la biomédecine” [Online]. Available: <http://www.agence-biomedecine.fr/Registre-France-Greffe-de-Moelle>. [Accessed: 17-Jan-2015].
- [62] D. Fürst, C. Müller, V. Vucinic, D. Bunjes, W. Herr, M. Gramatzki, R. Schwerdtfeger, R. Arnold, H. Einsele, G. Wulf, and others, “High-resolution HLA matching in hematopoietic stem cell transplantation: a retrospective collaborative analysis” *Blood*, vol. 122, no. 18, pp. 3220–3229, 2013.
- [63] J. A. Hansen, E. W. Petersdorf, M.-T. Lin, S. Wang, J. W. Chien, B. Storer, and P. J. Martin, “Genetics of allogeneic hematopoietic cell transplantation. Role of HLA matching, functional variation in immune response genes” *Immunol. Res.*, vol. 41, no. 1, pp. 56–78, May 2008.
- [64] J. Pidala, S. J. Lee, K. W. Ahn, S. Spellman, H.-L. Wang, M. Aljurf, M. Askar, J. Dehn, M. F. Viña, A. Gratwohl, and others, “Nonpermissive HLA-DPB1 mismatch increases mortality after myeloablative unrelated allogeneic hematopoietic cell transplantation” *Blood*, vol. 124, no. 16, p. 2596, 2014.
- [65] E. W. Petersdorf, M. Malkki, K. Hsu, P. Bardy, A. Cesbron, A. Dickinson, V. Dubois, K. Fleischhauer, T. Kawase, A. Madrigal, Y. Morishima, B. Shaw, S. Spellman, E. Spierings, M. Stern, J.-M. Tiercy, A. Velardi, T. Gooley, and on behalf of the International Histocompatibility Working Group in Hematopoietic Cell Transplantation, “16th IHIW: International Histocompatibility Working Group in Hematopoietic Cell Transplantation” *Int. J. Immunogenet*, pp. 2-10, Dec. 2012.
- [66] Y. Chalandon, J. M. Tiercy, U. Schanz, T. Gungor, R. Seger, J. Halter, C. Helg, B. Chapuis, A. Gratwohl, A. Tichelli, and others, “Impact of high-resolution matching in allogeneic unrelated donor stem cell transplantation in Switzerland” *Bone Marrow Transplant.*, vol. 37, no. 10, pp. 909–916, 2006.

- [67] M. Beksac, Ed., *Bone marrow and stem cell transplantation*. Totowa, N.J: Humana Press, 2007.
- [68] P. B. Dahi, D. M. Ponce, S. Devlin, K. L. Evans, M. Lubin, A. M. Gonzales, C. Byam, M. Sideroff, D. Wells, S. Giralt, N. A. Kernan, A. Scaradavou, and J. N. Barker, “Donor–recipient allele-level HLA matching of unrelated cord blood units reveals high degrees of mismatch and alters graft selection” *Bone Marrow Transplant.*, vol. 49, no. 9, pp. 1184–1186, Sep. 2014.
- [69] L. Gragert, M. Eapen, E. Williams, J. Freeman, S. Spellman, R. Baitty, R. Hartzman, J. D. Rizzo, M. Horowitz, D. Confer, and M. Maiers, “HLA Match Likelihoods for Hematopoietic Stem-Cell Grafts in the U.S. Registry” *N. Engl. J. Med.*, vol. 371, no. 4, pp. 339–348, Jul. 2014.
- [70] M. Eapen, J. P. Klein, G. F. Sanz, S. Spellman, A. Ruggeri, C. Anasetti, M. Brown, R. E. Champlin, J. Garcia-Lopez, G. Hattersely, G. Koegler, M. J. Laughlin, G. Michel, S. K. Nabhan, F. O. Smith, M. M. Horowitz, E. Gluckman, and V. Rocha, “Effect of donor–recipient HLA matching at HLA A, B, C, and DRB1 on outcomes after umbilical-cord blood transplantation for leukaemia and myelodysplastic syndrome: a retrospective analysis” *Lancet Oncol.*, vol. 12, no. 13, pp. 1214–1221, Dec. 2011.
- [71] E. Jaskula, J. Bochenska, E. Koewin, A. Tarnowska, and A. Lange, “CMV Serostatus of Donor-Recipient Pairs Influences the Risk of CMV Infection/Reactivation in HSCT Patients” *Bone Marrow Res.*, vol. 2012, pp. 1–8, 2012.
- [72] Olerup SSP, “Olerup\_HLA\_Instructions\_for\_Use\_IF.” 2012.
- [73] V. Moalic, “Comment est réalisé un typage HLA ?” *Réanimation*, vol. 17, no. 4, pp. 407–411, Jun. 2008.
- [74] P. P. J. Dunn, “Human leucocyte antigen typing: techniques and technology, a critical appraisal: HLA typing technology, a critical appraisal” *Int. J. Immunogenet.*, vol. 38, no. 6, pp. 463–473, Dec. 2011.
- [75] H. Erlich, “HLA DNA typing: past, present, and future” *Tissue Antigens*, vol. 80, no. 1, pp. 1–11, Jul. 2012.
- [76] J. Watson and F. Crick, “Molecular structure of nucleic acids” *Nature*, no. 4356, 1953.
- [77] P. Portin, “The birth and development of the DNA theory of inheritance: sixty years since the discovery of the structure of DNA” *J. Genet.*, vol. 93, no. 1, pp. 293–302, 2014.
- [78] A. Klug, “From macromolecules to biological assemblies (nobel lecture)” *Angew. Chem. Int. Ed. Engl.*, vol. 22, no. 8, pp. 565–582, 1983.
- [79] W. Gilbert, “DNA sequencing and gene structure, Nobel lecture” *Bioscience Reports*, pp. 353–375, décembre-1980.

- [80] F. Sanger, “Determination of Nucleotide Sequences in DNA” *Biosci. Rep.*, vol. 24, no. 4–5, pp. 237–253, Aug. 2004.
- [81] P. Berg, “Dissections and reconstructions of genes and chromosomes” *Biosci. Rep.*, vol. 1, no. 4, pp. 269–287, 1981.
- [82] W. Gilbert, “DNA sequencing and gene structure” *Biosci. Rep.*, vol. I, pp. 353–375, 1981.
- [83] J. Etienne and E. Clauser, *Biochimie génétique, biologie moléculaire.*, Masson. 2004.
- [84] F. Sanger, S. Nicklen, and A. R. Coulson, “DNA sequencing with chain-terminating inhibitors” *Proc. Natl. Acad. Sci.*, vol. 74, no. 12, pp. 5463–5467, 1977.
- [85] C. Philippe, “Apport du Séquençage Nouvelle génération en Génétique Médicale” 2014.
- [86] F. Sanger, “Determination of Nucleotide Sequences in DNA” *Biosci. Rep.*, vol. 24, no. 4–5, pp. 237–253, Aug. 2004.
- [87] V. Ecochard, “Techniques et stratégies en biologie moléculaire” 2011.
- [88] L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu, and M. Law, “Comparison of Next-Generation Sequencing Systems” *J. Biomed. Biotechnol.*, vol. 2012, pp. 1–11, 2012.
- [89] C. Gabriel, S. Stabentheiner, M. Danzer, and J. Pröll, “What Next? The Next Transit from Biology to Diagnostics: Next Generation Sequencing for Immunogenetics” *Transfus. Med. Hemotherapy*, vol. 38, no. 5, pp. 308–317, 2011.
- [90] S. D. Adams, K. C. Barracchini, D. Chen, F. Robbins, L. Wang, P. Larsen, R. Luhm, and D. F. Stroncek, “Ambiguous allele combinations in HLA Class I and Class II sequence-based typing: when precise nucleotide sequencing leads to imprecise allele identification” *J. Transl. Med.*, vol. 2, no. 1, p. 30, 2004.
- [91] A. G. Smith, C.-W. Pyo, W. Nelson, E. Gow, R. Wang, S. Shen, M. Sprague, S. E. Pereira, D. E. Geraghty, and J. A. Hansen, “Next generation sequencing to determine HLA class II genotypes in a cohort of hematopoietic cell transplant patients and donors” *Hum. Immunol.*, vol. 75, no. 10, pp. 1040–1046, Oct. 2014.
- [92] F. Malentacchi, C. M. Ciniselli, M. Pazzagli, P. Verderio, L. Barraud, C. C. Hartmann, S. Pizzamiglio, S. Weisbuch, R. Wyrich, and S. Gelmini, “Influence of pre-analytical procedures on genomic DNA integrity in blood samples: The SPIDIA experience” *Clin. Chim. Acta*, vol. 440, pp. 205–210, Feb. 2015.

- [93] C. L. Holcomb, B. Höglund, M. W. Anderson, L. A. Blake, I. Böhme, M. Egholm, D. Ferriola, C. Gabriel, S. E. Gelber, D. Goodridge, S. Hawbecker, R. Klein, M. Ladner, C. Lind, D. Monos, M. J. Pando, J. Pröll, D. C. Sayer, G. Schmitz-Agheguian, B. B. Simen, B. Thiele, E. A. Trachtenberg, D. B. Tyan, R. Wassmuth, S. White, and H. A. Erlich, “A multi-site study using high-resolution HLA genotyping by next generation sequencing” *Tissue Antigens*, vol. 77, no. 3, pp. 206–217, Mar. 2011.
- [94] C. Gabriel, D. Fürst, I. Faé, S. Wenda, C. Zollikofer, J. Mytilineos, and G. F. Fischer, “HLA typing by next-generation sequencing - getting closer to reality: HLA typing by NGS” *Tissue Antigens*, vol. 83, no. 2, pp. 65–75, Feb. 2014.
- [95] Z. Zhu, G. Jenkins, W. Zhang, M. Zhang, Z. Guan, and C. J. Yang, “Single-molecule emulsion PCR in microfluidic droplets” *Anal. Bioanal. Chem.*, vol. 403, no. 8, pp. 2127–2143, Jun. 2012.
- [96] M. S. Son and R. K. Taylor, “Preparing DNA Libraries for Multiplexed Paired-End Deep Sequencing for Illumina GA Sequencers” in *Current Protocols in Microbiology*, R. Coico, T. Kowalik, J. Quarles, B. Stevenson, and R. Taylor, Eds. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2011.
- [97] M. Falque, “Evolution des méthodes de génotypage” 2011.
- [98] J. M. Rothberg, W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth, and J. Bustillo, “An integrated semiconductor device enabling non-optical genome sequencing” *Nature*, vol. 475, no. 7356, pp. 348–352, Jul. 2011.
- [99] B. Merriman, I. T. R&D Team, and J. M. Rothberg, “Progress in Ion Torrent semiconductor chip based sequencing: Nanoanalysis” *ELECTROPHORESIS*, vol. 33, no. 23, pp. 3397–3417, Dec. 2012.
- [100] W. J. Ansorge, “Next-generation DNA sequencing techniques” *New Biotechnol.*, vol. 25, no. 4, pp. 195–203, Apr. 2009.
- [101] J. K., S. Suzuki, Y. Ozaki, S. Mitsunaga, H. Inoko, and T. Shii, “In Phase HLA Genotyping by Next Generation Sequencing — A Comparison Between Two Massively Parallel Sequencing Bench-Top Systems, the Roche GS Junior and Ion Torrent PGM” in *HLA and Associated Important Diseases*, Y. Xi, Ed. InTech, 2014, pp. 141–181.
- [102] P. K. Ehrenberg, A. Geretz, K. M. Baldwin, R. Apps, V. R. Polonis, M. L. Robb, J. H. Kim, N. L. Michael, and R. Thomas, “High-throughput multiplex HLA genotyping by next-generation sequencing using multi-locus individual tagging” *BMC Genomics*, vol. 15, no. 1, p. 864, 2014.

- [103] B. Grumbt, S. H. Eck, T. Hinrichsen, and K. Hirv, “Diagnostic Applications of Next Generation Sequencing in Immunogenetics and Molecular Oncology” *Transfus. Med. Hemotherapy*, vol. 40, no. 3, pp. 196–206, 2013.
- [104] H. Li and N. Homer, “A survey of sequence alignment algorithms for next-generation sequencing” *Brief. Bioinform.*, vol. 11, no. 5, pp. 473–483, Sep. 2010.
- [105] C. Cheng, Z. M. Kashi, R. Martin, G. Woodruff, D. Dinauer, and T. Agostini, “HLA-C locus allelic dropout in Sanger sequence-based typing due to intronic single nucleotide polymorphism” *Hum. Immunol.*, vol. 75, no. 12, pp. 1239–1243, Dec. 2014.
- [106] M. Danzer, N. Niklas, S. Stabentheiner, K. Hofer, J. Pröll, C. Stückler, E. Raml, H. Polin, and C. Gabriel, “Rapid, scalable and highly automated HLA genotyping using next-generation sequencing: a transition from research to diagnostics” *BMC Genomics*, vol. 14, no. 1, p. 221, 2013.
- [107] A. Magi, M. Benelli, A. Gozzini, F. Girolami, F. Torricelli, and M. L. Brandi, “Bioinformatics for Next Generation Sequencing Data” *Genes*, vol. 1, no. 2, pp. 294–307, Sep. 2010.
- [108] D. S. Horner, G. Pavesi, T. Castrignano, P. D. De Meo, S. Liuni, M. Sammeth, E. Picardi, and G. Pesole, “Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing” *Brief. Bioinform.*, vol. 11, no. 2, pp. 181–197, Mar. 2010.
- [109] V. Lange, I. Böhme, J. Hofmann, K. Lang, J. Sauter, B. Schöne, P. Paul, V. Albrecht, J. M. Andreas, D. M. Baier, and others, “Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing” *BMC Genomics*, vol. 15, no. 1, p. 63, 2014.
- [110] D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, and J. A. Schloss, “The potential and challenges of nanopore sequencing” *Nat. Biotechnol.*, vol. 26, no. 10, pp. 1146–1153, Oct. 2008.
- [111] “Automatic optimisation of system performance - Technology - Oxford Nanopore Technologies” [Online]. Available: <https://nanoporetech.com/technology/automatic-optimisation-of-system-performance>. [Accessed: 22-Jan-2015].
- [112] “Oxford Nanopore Technologies” [Online]. Available: <https://nanoporetech.com/>. [Accessed: 28-Jun-2015].
- [113] “Pacific Biosciences: Overview [www.pacificbiosciences.com](http://www.pacificbiosciences.com)” [Online]. Available: <http://www.pacificbiosciences.com/products/>. [Accessed: 22-Jan-2015].
- [114] P. Coupland, T. Chandra, M. Quail, W. Reik, and H. Swerdlow, “Direct sequencing of small genomes on the Pacific Biosciences RS without library preparation” *BioTechniques*, vol. 53, no. 6, Dec. 2012.

- [115] N. Mayor, J. Robinson, A. McWhinnie, and S. Ranade, “Genomic DNA sequences of HLA class I alleles generated using multiplexed barcodes and SMRT DNA sequencing technology” Anthony Nolan, Pacific Bioscience.
- [116] C. C. Bell, L. Faulkner, K. Martinsson, J. Farrell, A. Alfirevic, J. Tugwood, M. Pirmohamed, D. J. Naisbitt, and B. K. Park, “T-Cells from HLA-B\*57:01+ Human Subjects Are Activated with Abacavir through Two Independent Pathways and Induce Cell Death by Multiple Mechanisms” *Chem. Res. Toxicol.*, vol. 26, no. 5, pp. 759–766, May 2013.
- [117] P. Perrier, A. Dormoy, C. Andre-Botte, and N. Froelich, “HLA-A\*02010102L: a laborious assignment” *Tissue Antigens*, vol. 68, no. 5, pp. 442–445, Nov. 2006.
- [118] K. Hosomichi, T. A. Jinam, S. Mitsunaga, H. Nakaoka, and I. Inoue, “Phase-defined complete sequencing of the HLA genes by next-generation sequencing” *BMC Genomics*, vol. 14, no. 1, p. 355, 2013.
- [119] Y. Ozaki, S. Suzuki, K. Kashiwase, A. Shigenari, Y. Okudaira, S. Ito, A. Masuya, F. Azuma, T. Yabe, S. Morishima, S. Mitsunaga, M. Satake, M. Ota, Y. Morishima, J. K. Kulski, K. Saito, H. Inoko, and T. Shiina, “Cost-efficient multiplex PCR for routine genotyping of up to nine classical HLA loci in a single analytical run of multiple samples by next generation sequencing” *BMC Genomics*, vol. 16, no. 1, Dec. 2015.
- [120] J. H. Lan, Y. Yin, E. F. Reed, K. Moua, K. Thomas, and Q. Zhang, “Impact of three Illumina library construction methods on GC bias and HLA genotype calling” *Hum. Immunol.*, vol. 76, no. 2–3, pp. 166–175, Mar. 2015.
- [121] E. Knierim, B. Lucke, J. M. Schwarz, M. Schuelke, and D. Seelow, “Systematic Comparison of Three Methods for Fragmentation of Long-Range PCR Products for Next Generation Sequencing” *PLoS ONE*, vol. 6, no. 11, p. e28240, Nov. 2011.
- [122] D. Aird, M. G. Ross, W.-S. Chen, M. Danielsson, T. Fennell, C. Russ, D. B. Jaffe, C. Nusbaum, and A. Gnirke, “Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries” *Genome Biol.*, vol. 12, no. 2, p. R18, 2011.
- [123] S. Behjati and P. S. Tarpey, “What is next generation sequencing?” *Arch. Dis. Child. - Educ. Pract.*, vol. 98, no. 6, pp. 236–238, Nov. 2013.
- [124] S. Boegel, M. Lower, M. Schafer, T. Bukur, J. De Graaf, V. Boisguérin, O. Tureci, M. Diken, J. C. Castle, and U. Sahin “HLA typing from RNA-Seq sequence reads” *Genome Med* vol. 4, no. 12, p. 102, 2013.
- [125] E. W. Petersdorf, “Genetics of graft-versus-host disease: The major histocompatibility complex” *Blood Rev.*, vol. 27, no. 1, pp. 1–12, Jan. 2013.
- [126] E. W. Petersdorf, “The major histocompatibility complex: a model for understanding graft-versus-host disease” *Blood*, vol. 122, no. 11, pp. 1863–1872, Sep. 2013.
- [127] S. Roedder, T. Sigdel, N. Salomonis, S. Hsieh, and M. M. Sarwal “The kSORT Assay to Detect Renal Transplant Patients at High Risk for Acute rejection : results of the Multicenter AART Study” *PLOS Med.*, vol. 11, pp. 1–8, 2014.

## ANNEXES

Récapitulatif des principales caractéristiques et différences entre les molécules de classe I et II.

<b>Principales différences entre classe I et classe II</b>		
<b>Molécule HLA</b>	<b>Classe I</b>	<b>Classe II</b>
<b>Loci d'intérêt en transplantation</b>	A, B, C	DRB1, DQB1, DPB1
<b>Taille des gènes</b>	Environ 3 kb	De 4 à 10 kb
<b>Nb d'exons</b>	7 ou 8	5 ou 6
<b>Polymorphisme</b>	Exons 2 et 3	Exon 2
<b>Expression cellulaire</b>	Quasiment toutes les cellules nucléées (+ plaquettes) (- os, cartilage, neurones...)	Restreinte : lymphocytes B+++
<b>Peptides présentés</b>	Endogènes (synthétisé par la cellule elle-même) (viraux + soi)	Exogènes
<b>Lymphocytes T</b>	CD8+	CD4+

**Tableau récapitulatif des sources de cellules souches hématopoïétiques**

Comparaison/résumé des différentes sources de cellules souches hématopoïétiques			
Origine des cellules souches	Moelle osseuse MO	Cellules Souches Périphériques (CSP)	Unités de Sang Placentaire (USP)
Quantité de CSH (Qualité)	Grande quantité de CSH pour un faible volume de moelle	Grande quantité de CSH	Quantité de CSH faible et limité (non repreneable). (2 USP pour un adulte)
Quantité de lymphocytes T (potentiel GvH)	Quantité correcte de lymphocytes T pour un faible volume de moelle	Grande quantité de lymphocytes T (risque de GvH ↑/GvL)	Quantité lymphocyte T faible (potentiel GvH diminué mais également GvL)
Modalité prélèvement	Prélèvement délicat, contrainte élevé pour le donneur, bonne tolérance	Prélèvement relativement simple, mais injection de G-CSF nécessaire : bonne tolérance mais < MO	Prélèvement simple, aucune contrainte pour autrui
Compatibilités	10/10 voire 9/10	10/10 voire 9/10	6/6, moins de contraintes
Post-greffe	Reconstitution avec un délai intermédiaire au deux autres sources	Reconstitution hématopoïétique plus rapide qu'avec un greffon médullaire	Risque de non prise de greffe, reconstitution plus lente qu'avec MO ou CSP Moins de cellules immunocompétentes : infections post greffe+
DLI	Possible	Possible	Un seul recueil possible, pas de DLI
Disponibilité	Risque de changement d'avis, mauvaise compliance, indisponibilité	Risque de changement d'avis, mauvaise compliance, indisponibilité	Disponibilité rapide des cellules souches
Risques infectieux	Risque infectieux par le greffon +	Risque infectieux par le greffon +	Sécurité ++ (moins de risques infectieux par le greffon)



## DEMANDE D'IMPRIMATUR

Date de soutenance : 10 Septembre 2015

**DIPLOME D'ETAT DE DOCTEUR  
EN PHARMACIE**

présenté par : Alexandra JACQUES

Sujet : Intérêt du séquençage dans les typages HLA et la greffe de cellules souches hématopoïétiquesJury :Président (Co-directeur) : Pr BENSOUSSAN Danièle, PU-PH,  
Université de Lorraine, CHU de NancyDirecteur : Dr AARNINK Alice, PH-U, Pharmacien,  
Université de Lorraine, CHU de NancyJuges : Pr FRIMAT Luc, PU-PH, Médecin, Université de  
Lorraine, CHU de NancyDr BONNET Céline, MCU-PH, Médecin-  
biologiste CHU de NancyDr MOALIC-ALLAIN Virginie, PH, Pharmacien-  
biologiste, CHU de Brest

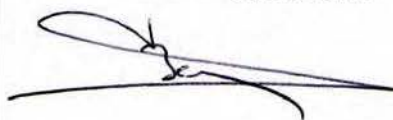
Vu,

Nancy, le 21 Juillet 2015

Le Président du Jury

Directeur de Thèse

Mme D. BENSOUSSAN - M. M. A. AARNINK



Vu et approuvé,

Nancy, le 29.07.2015

Doyen de la Faculté de Pharmacie  
de l'Université de Lorraine,  
Béatrice FAIVRE  
Vice-Doyen

Francine PAULUS

Vu,

Nancy, le 28 AOUT 2015

Le Président de l'Université de Lorraine,



Pierre MUTZENHARDT

N° d'enregistrement :

7048

N° d'identification :

**TITRE**

**Intérêt du séquençage dans les typages HLA et la greffe de cellules souches hématopoïétiques**

**Thèse soutenue le 10 Septembre 2015  
Par Alexandra JACQUES**

**RESUME :**

La greffe de cellules souches hématopoïétiques est parfois le seul espoir thérapeutique dans le cas d'hémopathies malignes ou de déficit immunitaires congénitaux. Différentes sources sont utilisées : la moelle osseuse, les cellules souches du sang périphérique ou encore les unités de sang placentaire. Chacune possède ses avantages et inconvénients, mais dans tous les cas la greffe présente des effets secondaires redoutables, principalement la réaction du greffon contre l'hôte, responsable d'une mortalité post- greffe élevée, après les rechutes des hémopathies.

Les conditions optimales sont encore imparfaitement connues, qu'il s'agisse des compatibilités des molécules du Complexe Majeur d'Histocompatibilité ou bien d'autres facteurs génétiques qui pourraient avoir un rôle important.

Les techniques de typage HLA actuellement utilisées n'arrivent plus à satisfaire les exigences de plus en plus strictes des standards de l'EFI. Le Séquençage Sanger est devenu une technique indispensable pour permettre au laboratoire d'être en conformité avec ces exigences. Parallèlement, l'arrivée du séquençage de nouvelle génération, apporte des solutions intéressantes, en termes de résolution de typage et de débit.

Le premier objectif de cette thèse est de montrer l'amélioration apportée par le séquençage Sanger, dans la qualité des résultats rendus par le laboratoire du CHRU de Nancy. Nous constaterons également que le séquençage Sanger possède ses propres limites, repoussées par les technologies de 2<sup>ième</sup> et 3<sup>ième</sup> génération arrivant sur le marché. Nous discuterons ainsi de l'intérêt de la mise en place de ces nouvelles générations, et nous comparerons les différentes technologies disponibles sur le marché. Ces avancées nous permettront d'explorer plus aisément les facteurs influençant la réussite d'une greffe de moelle. Elles faciliteront également l'incrémentation des registres de Donneurs Volontaires de Moelle Osseuse avec des typages plus complets, contribuant à la valorisation des donneurs inscrits sur le registre français.

**MOTS CLES :**

Histocompatibilité - HLA – Séquençage – NGS– Typage – Greffe - Cellules souches - Moelle osseuse

<b>Directeur de thèse</b>	<b>Intitulé du laboratoire</b>	<b>Nature</b>	
<b>Dr Alice AARNINK, PH-U</b>  <b>Dr Danièle BENSOUSSAN,</b> <b>PU-PH, co-directeur</b>	<b>Faculté de Pharmacie de</b> <b>Nancy,</b> <b>Université de Lorraine</b>  <b>Laboratoire</b> <b>d'Histocompatibilité,</b> <b>CHU de Nancy</b>	<b>Expérimentale</b>	<input type="checkbox"/>
		<b>Bibliographique</b>	<input checked="" type="checkbox"/>
		<b>Thème</b>	<b>5</b>

**Thèmes**

**1 – Sciences fondamentales**

**3 – Médicament**

**⑤ - Biologie**

**2 – Hygiène/Environnement**

**4 – Alimentation – Nutrition**

**6 – Pratique professionnelle**