



HAL
open science

Développement de méthodes d'analyse de l'ADN par clivage d'une chimère ARN/ADN et par spectrométrie de masse MALDI-TOF

Florence Mauger

► **To cite this version:**

Florence Mauger. Développement de méthodes d'analyse de l'ADN par clivage d'une chimère ARN/ADN et par spectrométrie de masse MALDI-TOF. Génétique. Université Pierre et Marie Curie - Paris VI, 2012. Français. NNT : 2012PAO66252 . tel-00833267

HAL Id: tel-00833267

<https://theses.hal.science/tel-00833267>

Submitted on 12 Jun 2013

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

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

THESE DE DOCTORAT DE L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité Chimie Moléculaire ED406

Présentée par

Florence Mauger

Pour obtenir le grade de Docteur de l'université Pierre et Marie Curie

**Développement de méthodes d'analyse de l'ADN
par clivage
d'une chimère ARN/ADN
et par spectrométrie de masse MALDI-TOF**

Soutenue le 10 Juillet 2012 devant le jury composé de :

Mme. Jeanine Tortajada
M. Codjo Hountondji
M. Christian Rolando
M. Germain Trugnan
M. Jean-Claude Tabet
M. Ivo Gut

Université EVRY Val d'Essonne
Université Pierre et Marie Curie
Université Lille 1
Université Pierre et Marie Curie
Université Pierre et Marie Curie
CNAG Barcelona

Rapporteur
Rapporteur
Examineur
Examineur
Directeur de thèse
co-Directeur de thèse

Remerciements

Ce projet de thèse présenté dans ce mémoire a été réalisé au Laboratoire de Développement et de Production au sein du CEA/DSV/Institut de Génomique/Centre National de Génotypage. Il a été effectué sous la direction du Professeur Jean-Claude Tabet, Professeur émérite de l'Université Pierre et Marie Curie et du Docteur Ivo Gut, Directeur du Centro Nacional de Analisis Genomico à Barcelona et responsable du projet READNA.

Je tiens à remercier le Docteur Mark Lathrop, ancien Directeur du CNG, et les membres de la commission de formation du CEA pour m'avoir permis d'effectuer cette thèse dans le cadre d'une formation professionnelle. Je tiens plus particulièrement à remercier le Professeur Jean-Claude Tabet et le Docteur Ivo Gut d'avoir accepté d'être mes directeurs de thèse, pour le temps qu'ils ont consacré à mes recherches et pour leurs disponibilités.

Je remercie vivement la Professeur Jeanine Tortajada et le Professeur Codjo Hountondji d'avoir accepté d'être rapporteurs ainsi que le Directeur de recherche Christian Rolando et le Professeur Germain Trugnan d'avoir bien voulu être examinateurs du jury de cette thèse.

Je tiens également à remercier le Docteur David Gelfang, le Docteur Keith Bauer et le Docteur Thomas Myers, de la société ROCHE MOLECULAR pour leur collaboration essentielle concernant le développement des ADN polymérase.

Je remercie également Jérémy Semhoun pour l'élaboration du logiciel de calcul des masses, Caroline Horgues, Steven McGinn, Jorg Tost et Alexandre How Kit pour leur collaboration sur le projet de l'analyse de la méthylation de l'ADN et Ekatherina Darii pour son aide pour les expériences par ESI-ITMSⁿ. Je tiens également à remercier, Céline Besse, Christian Daviaud, Céline Lacrouts d'avoir eu la gentillesse et la patience de corriger mon mémoire.

Enfin, je remercie la Communauté européenne pour le financement de cette étude au sein du projet READNA 2008-2012.

Sommaire

| | |
|--|-----------|
| Remerciements | 2 |
| Sommaire | 3 |
| Abréviation | 11 |
| Introduction générale | 12 |
| 1. Les modifications de l'ADN | 13 |
| 1.1. L'ADN..... | 13 |
| 1.2. Les modifications de l'ADN..... | 15 |
| 1.2.1. Les modifications de la séquence..... | 15 |
| 1.2.2. Les modifications épigénétiques..... | 16 |
| 2. Les méthodes d'analyse de l'ADN | 17 |
| 2.1. Le séquençage de <i>première génération</i> | 18 |
| 2.1.1. Le séquençage de Sanger..... | 18 |
| 2.1.2. Le séquençage de Maxam et Gilbert..... | 20 |
| 2.2. Le séquençage de <i>seconde génération</i> | 20 |
| 2.2.1. Le séquençage sur phase solide..... | 20 |
| 2.2.2. Le séquençage par hybridation sur puce..... | 21 |
| 2.2.3. Le séquençage cyclique sur puce..... | 22 |
| 2.2.4. Le séquençage par microélectrophorèse..... | 23 |
| 2.2.5. Le séquençage par nanopore..... | 23 |
| 2.3. Le séquençage de <i>nouvelle génération</i> <i>NGS</i> | 24 |
| 3. Les méthodes d'analyse de l'ADN par spectrométrie de masse | 25 |
| 3.1. La spectrométrie de masse ESI-IT et MALDI-TOF..... | 25 |
| 3.1.1. La spectrométrie de masse ESI-IT..... | 25 |
| 3.1.2. La spectrométrie de masse MALDI-TOF..... | 31 |
| 3.2. Les méthodes d'analyse de l'ADN..... | 42 |
| 3.2.1. Les méthodes directes..... | 43 |
| 3.2.2. Les méthodes indirectes..... | 45 |

| | |
|--|-----------|
| 4. Développement de méthodes d'analyse de l'ADN par clivage d'une chimère ARN/ADN et par spectrométrie de masse MALDI-TOF | 47 |
| 4.1. Le principe | 47 |
| 4.1.1. La chimère ARN/ADN | 48 |
| 4.1.2. Le clivage de la chimère ARN/ADN | 49 |
| 4.1.3. L'analyse par MALDI-TOF MS | 50 |
| 4.2. Les méthodes | 51 |
| 4.2.1. Les méthodes actuelles | 51 |
| 4.2.2. Le développement de nouvelles méthodes | 53 |
| Chapitre 1. Analyse de la chimère ARN/ADN par spectrométrie de masse en tandem ... | 55 |
| 1. Introduction | 56 |
| 2. Matériels et Méthodes | 58 |
| 2.1. Matériels | 58 |
| 2.2. Méthodes | 60 |
| 2.2.1. Le clivage | 60 |
| 2.2.2. Le dessalage | 60 |
| 2.2.3. L'analyse par spectrométrie de masse MALDI-TOF/TOF | 60 |
| 2.2.4. L'analyse spectrométrie de masse ESI-IT | 61 |
| 2.2.5. Les ions fragments | 61 |
| 3. Résultats | 62 |
| 3.1. La chimère ARN/ADN | 62 |
| 3.1.1. La formation de la chimère ARN/ADN | 62 |
| 3.1.2. La nomenclature des ions fragments | 62 |
| 3.1.3. Le seuil de détection des chimères ARN/ADN par MALDI-TOF/TOF | 63 |
| 3.2. L'analyse des chimères ARN/ADN de 4-mer par MALDI-TOF/TOF | 65 |
| 3.2.1. Les fragmentations en mode CID des chimères avec une base B ₄ (A) | 65 |
| 3.2.2. Les fragmentations en mode CID des chimères avec une base B ₄ (C) | 69 |
| 3.2.3. Les fragmentations en mode CID des chimères avec une base B ₄ (G) | 72 |
| 3.2.4. Les fragmentations en mode CID des chimères avec une base B ₄ (U) | 74 |
| 3.2.5. Le bilan des fragmentations en mode CID | 76 |

| | |
|---|------------|
| 3.3. Les fragmentations en phase gazeuse de la chimère GCTA | 79 |
| 3.3.1. L'impact de la fluence du laser | 79 |
| 3.3.2. Les fragmentations en mode CID par MALDI-TOF/TOF | 83 |
| 3.3.3. Les fragmentations en mode CID par ESI-ITMS ⁿ | 84 |
| 3.3.4. Le bilan des fragmentations en mode CID | 95 |
| 3.4. Les fragmentations en mode CID des chimères ARN/ADN supérieures à 4-mer par MALDI-TOF/TOF | 97 |
| 3.4.1. Les fragmentations en mode CID des chimères ARN/ADN de 5-mer | 97 |
| 3.4.2. Les fragmentations en mode CID des chimères ARN/ADN de 7-mer | 102 |
| 3.4.3. Les fragmentations en mode CID des chimères ARN/ADN de 10-mer | 104 |
| 4. Discussion | 106 |
| 4.1. Les conditions expérimentales | 106 |
| 4.1.1. La préparation de la chimère ARN/ADN | 106 |
| 4.1.2. L'analyse par MALDI-TOF/TOF | 107 |
| 4.2. Les fragmentations par MALDI-TOF-TOF de la chimère ARN/ADN | 109 |
| 4.2.1. La détection des ions | 109 |
| 4.2.2. Les fragmentations en mode CID | 110 |
| 4.2.3. Le séquençage | 114 |
| 4.3. Les fragmentations en mode CID de la chimère GCTA | 114 |
| 4.3.1. L'analyse par ESI-ITMS ⁿ | 115 |
| 4.3.2. Les ions multichargés | 115 |
| 4.3.3. Les fragmentations par spectrométrie de masse en tandem | 116 |
| 5. Conclusion | 117 |
| Chapitre 2. Multiplex microhaplotypage par clivage d'une chimère ARN/ADN simple-brin et par MALDI-TOF MS | 120 |
| 1. Introduction | 121 |
| 2. Matériels et Méthodes | 123 |
| 2.1. Matériels | 123 |
| 2.2. Méthodes | 125 |
| 2.2.1. PCR | 125 |

| | |
|---|------------|
| 2.2.2. L'élongation linéaire par l'ADN polymérase <i>G46E CS6R</i> | 126 |
| 2.2.3. Le clivage | 126 |
| 2.2.4. Le dessalage | 126 |
| 2.2.5. L'analyse par spectrométrie de masse MALDI-TOF | 127 |
| 3. Résultats | 127 |
| 3.1. La PCR | 128 |
| 3.2. La chimère ARN/ADN simple-brin | 129 |
| 3.3. Le clivage de la chimère ARN/ADN | 130 |
| 3.4. Le dessalage des fragments de clivage | 130 |
| 3.5. L'analyse des spectres de masse MALDI-TOF | 130 |
| 3.5.1. La matrice HCCA | 131 |
| 3.5.2. Les paramètres d'analyse | 131 |
| 3.5.3. Les microhaplotypes | 131 |
| 3.5.4. Le microhaplotype ASTW du locus <i>HLA-DRB1-227</i> | 132 |
| 3.5.5. Les microhaplotypes HCACH du locus <i>HLA-A-98</i> et MGAR du locus <i>HLA-A-453</i> | 134 |
| 3.6. L'analyse des individus | 135 |
| 4. Discussion | 137 |
| 4.1. La PCR | 137 |
| 4.2. La chimère ARN/ADN simple-brin | 138 |
| 4.2.1. L'ADN polymérase <i>G46E CS6R</i> | 138 |
| 4.2.2. Les amorces | 138 |
| 4.3. L'analyse des microhaplotypes | 139 |
| 4.3.1. Le clivage | 139 |
| 4.3.2. Le dessalage | 139 |
| 4.3.2. L'analyse par MALDI-TOF MS | 140 |
| 4.3.3. Les caractéristiques de la méthode | 140 |
| 5. Conclusion | 141 |
| Chapitre 3. Analyse de la méthylation de l'ADN par clivage d'une chimère ARN/ADN simple-brin et par MALDI-TOF MS | 143 |
| 1. Introduction | 144 |

| | |
|---|------------|
| 2. Matériels et Méthodes | 146 |
| 2.1. Matériels..... | 146 |
| 2.2. Méthodes..... | 148 |
| 2.2.1. La conversion au bisulfite..... | 148 |
| 2.2.2. Les mélanges d'ADN..... | 149 |
| 2.2.3. La PCR..... | 149 |
| 2.2.4. La purification de la PCR..... | 150 |
| 2.2.5. L'élongation linéaire par l'ADN polymérase <i>KB17</i> | 150 |
| 2.2.6. Le clivage..... | 151 |
| 2.2.7. Le dessalage..... | 151 |
| 2.2.8. L'analyse par spectrométrie de masse MALDI-TOF..... | 151 |
| 3. Résultats | 153 |
| 3.1. Le principe de la méthode..... | 153 |
| 3.1.1. La conversion au bisulfite..... | 154 |
| 3.1.2. La PCR..... | 154 |
| 3.1.3. L'élongation linéaire par l'ADN polymérase <i>KB17</i> | 155 |
| 3.1.4. Les fragments de clivage des CpGs..... | 157 |
| 3.1.5. L'analyse par spectrométrie de masse MALDI-TOF..... | 162 |
| 3.1.6. L'analyse de la méthylation de l'ADN..... | 162 |
| 3.2. L'analyse du locus <i>CDKN2A-msp1</i> | 163 |
| 3.2.1. Les fragments de clivage du locus..... | 163 |
| 3.2.2. L'élongation linéaire ATP..... | 166 |
| 3.2.3. L'élongation linéaire GTP..... | 168 |
| 3.3. La quantification de la méthylation de l'ADN..... | 169 |
| 3.3.1. La courbe d'étalonnage du locus <i>CDKN2A-msp1</i> | 170 |
| 3.3.2. La courbe d'étalonnage du locus <i>CDKN2A-msp3b</i> | 171 |
| 3.4. L'analyse des individus..... | 175 |
| 3.4.1. Les loci des gènes <i>CDKN2A</i> et <i>RASSF1A</i> | 175 |
| 3.4.2. L'analyse des loci des gènes <i>CDKN2A</i> et <i>RASSF1A</i> | 175 |

| | |
|--|------------|
| 3.4.3. Le pourcentage de méthylation de l'ADN des loci <i>CDKN2A</i> - <i>msp1</i> et <i>msp3b</i> | 176 |
| 4. Discussion | 178 |
| 4.1. La conversion au bisulfite..... | 179 |
| 4.2. La PCR..... | 179 |
| 4.2.1. Les amorces..... | 179 |
| 4.2.2. Le biais de PCR..... | 179 |
| 4.3. L'analyse de la méthylation de l'ADN..... | 181 |
| 4.3.1. Le re-séquençage..... | 181 |
| 4.3.2. L'analyse des CpGs..... | 182 |
| 4.3.3. La quantification de la méthylation de l'ADN..... | 183 |
| 4.4. L'analyse des individus..... | 186 |
| 4.5. L'avantage de cette méthode..... | 186 |
| 5. Conclusion | 187 |
| Chapitre 4. Re-séquençage de l'ADN par clivage d'une <i>ribo-PCR</i> et par MALDI-TOF MS | 189 |
| 1. Introduction | 190 |
| 2. Matériels et Méthodes | 192 |
| 2.1. Matériels..... | 192 |
| 2.2. Méthodes..... | 193 |
| 2.2.1. La <i>ribo-PCR</i> par l'ADN polymérase KB17..... | 193 |
| 2.2.2. Le clivage..... | 195 |
| 2.2.3. Le dessalage..... | 195 |
| 2.2.4. L'analyse par spectrométrie de masse MALDI-TOF..... | 195 |
| 3. Résultats | 196 |
| 3.1. La <i>ribo-PCR</i> par l'ADN polymérase KB17..... | 197 |
| 3.2. Les fragments de clivage..... | 197 |
| 3.3. L'analyse des spectres de masse MALDI-TOF..... | 201 |
| 3.3.1. L'analyse du locus <i>NOS1</i> | 201 |
| 3.3.2. L'analyse du locus <i>H19</i> | 204 |
| 3.3.3. L'analyse du duplex des loci <i>NOS1</i> et <i>H19</i> | 207 |

| | |
|--|------------|
| 3.3.4. L'analyse du locus <i>SLCO1B1</i> | 208 |
| 3.4. L'analyse des individus..... | 210 |
| 4. Discussion | 213 |
| 4.1. L'ADN polymérase <i>KBI7</i> | 213 |
| 4.1.1. Les propriétés..... | 213 |
| 4.1.2. La séquence spécifique d'amplification..... | 213 |
| 4.1.2. Le multiplexage des <i>ribo-PCRs</i> | 214 |
| 4.2. L'analyse par MALDI-TOF MS..... | 215 |
| 4.2.1. Le mode linéaire..... | 215 |
| 4.2.2. L'étalonnage..... | 215 |
| 4.2.3. L'analyse du double-brin..... | 216 |
| 4.2.4. La détection des fragments de clivage..... | 216 |
| 4.3. Les avantages de la méthode..... | 217 |
| 5. Conclusion | 218 |
| Chapitre 5. Multiplex allèle-spécifique génotypage par clivage d'une <i>riboPAP-PCR</i> et par MALDI-TOF MS | 220 |
| 1. Introduction | 221 |
| 2. Matériels et Méthodes | 223 |
| 2.1. Matériels..... | 223 |
| 2.2. Méthodes..... | 225 |
| 2.2.1. La <i>riboPAP-PCR</i> par l'ADN polymérase <i>FP-1</i> | 225 |
| 2.2.2. Le clivage..... | 226 |
| 2.2.3. Le dessalage..... | 226 |
| 2.2.4. L'analyse par spectrométrie de masse MALDI-TOF..... | 226 |
| 3. Résultats | 227 |
| 3.1. Le principe de la méthode..... | 227 |
| 3.2. La <i>riboPAP-PCR</i> par l'ADN polymérase <i>FP-1</i> | 229 |
| 3.3. Les fragments de clivage..... | 230 |
| 3.4. L'analyse des spectres de masse MALDI-TOF..... | 232 |
| 3.4.1. L'analyse du locus <i>NOS1</i> | 232 |

| | |
|---|------------|
| 3.4.2. L'analyse du locus <i>H19</i> | 234 |
| 3.4.3. L'analyse du duplex des loci <i>SLCO1B1</i> | 235 |
| 3.4.4. Les <i>flags</i> du duplex des loci <i>SLCO1B1</i> | 236 |
| 3.5. L'analyse des individus..... | 243 |
| 4. Discussion | 244 |
| 4.1. La réaction <i>PAP</i> | 245 |
| 4.2. Les amorces..... | 245 |
| 4.3. L'ADN polymérase <i>FP-1</i> | 246 |
| 4.4. Les <i>tags</i> et les <i>flags</i> | 247 |
| 4.5. L'avantage de la méthode..... | 248 |
| 5. Conclusion | 248 |
| Conclusion générale et perspectives | 250 |
| Annexes | 259 |
| Annexe 1. SNP genotyping using alkali cleavage of RNA/DNA chimeras and MALDI-TOF MS..... | 260 |
| Annexe 2. DNA sequencing by MALDI-TOF MS using alkali cleavage of RNA/DNA chimeras..... | 271 |
| Annexe 3. Ribo-PCR. A facile method for the preparation of chimeric RNA/DNA applied to DNA sequencing..... | 283 |
| Annexe 4. Ribonucleotide tag nucleic acid detection..... | 297 |
| Annexe 5. High specificity single tube multiplex genotyping using ribo-PAP PCR, tag primers, alkali cleavage of RNA/DNA chimeras and MALDI-TOF MS..... | 319 |
| Références bibliographiques | 331 |

Abréviations

| | |
|---|---|
| A : Adénine | MALDI : Matrix-assisted laser desorption/ionization |
| ADN : Acide désoxyribonucléique | MS : Spectrométrie de masse |
| ATP: Adénosine 5'triphosphate | MS/MS : Spectrométrie de masse en tandem |
| ARN : Acide ribonucléique | NTP : ribonucléoside 5'triphosphate |
| C : Cytosine | <i>NOS1</i> :oxyde nitrique acide neuronale |
| C ^{Me} : 5-méthylecytosine | PAP : Polymérisation activée par pyrophosphorolyse |
| <i>CDKN2A</i> :cyclin-dependent kinase inhibitor 2A | Pb : paire de bases |
| CpG : dinucléotide cytosine guanine | PCR : Réaction de polymérisation en chaîne |
| CEPH : Centre d'Etude du polymorphisme Humain - Fondation Jean DAUSET | R : Adénine ou Guanine |
| CID : Collision-induced dissociation | <i>RASSF1A</i> :Rass association domain-containing protein 1 |
| CMH : Complexe d'histocompatibilité chez l'humain | <i>riboPAP-PCR</i> : PCR activée par pyrophosphorolyse de chimère ARN/ADN |
| CTP : Cytidine 5'triphosphate | <i>ribo-PCR</i> : PCR de chimère ARN/ADN |
| Da : Dalton unité de masse moléculaire des biologistes | S : Cytosine ou Guanine |
| dATP : désoxyadénosine 5'triphosphate | <i>SCLO1B1</i> : Solute carrier organic anion transporter family member 1B1 |
| dCTP : désoxycytidine 5'triphosphate | SNP : Single nucleotide polymorphisms |
| dGTP : désoxyguanosine 5'triphosphate | T : Thymine |
| dNTP : désoxynucléoside 5' triphosphate | <i>Tag</i> : séquence répétitive |
| dTTP : désoxythymidine 5'triphosphate | THAP : Trihydroxyacétophénone |
| ESI : désorption/ionisation par électronebulisation | TOF : Time of Flight : temps de vol |
| <i>Flag</i> : brin d'ADN complémentaire de la <i>Tag</i> | Tris : tris(hydroxyméthyl)aminométhane |
| G : Guanine | U : Uracile |
| GTP : Guanosine 5'triphosphate | U ^{Me} : 5-méthyluracile |
| H : Adénine ou Cytosine ou Thymine | UNG : Uracile-N-glycosylase |
| <i>H19</i> :gène situé sur le chromosome 11 | UTP : Uridine 5'triphosphate |
| HCCA : acide α -cyano 4-hydroxycinnamique | W : Adénine ou Thymine |
| HLA : Antigène des leucocytes humains | Y : Thymine ou Cytosine |
| HPA : acide 3-hydroxypicolinique | |
| IT : Trappe ionique | |
| kV : kilovolt | |
| M : Adénine ou Cytosine | |

Introduction générale

Introduction générale

En 1990, le projet du génome humain a eu pour but le séquençage du génome entier. Il s'est achevé en 2004 grâce à un consortium public international (Lander et al. 2001) et une compagnie privée Celera Genomics (Venter et al. 2001). Il a permis l'étude de l'information génétique portée par l'ensemble du génome humain qui contient environ 20 000 gènes.

Depuis, les projets de recherche se focalisent sur l'analyse des modifications de la séquence de l'ADN, notamment l'étude des polymorphismes ou de la méthylation de l'ADN. Les méthodes de séquençage actuellement développées visent au re-séquençage de l'ADN afin d'étudier les polymorphismes pouvant servir de marqueurs pour d'identification de maladies héréditaires. Le développement de nouvelles méthodes efficaces, rapides, rentables, est donc une nécessité pour l'avancée de la biologie moléculaire.

Les travaux réalisés au cours de ce projet de thèse ont pour but la mise au point de méthodes d'analyse de l'ADN par clivage d'une chimère ARN/ADN et par spectrométrie de masse MALDI-TOF. Ce projet rentre dans le cadre du projet européen, *READNA (REvolutionary Approaches and Devices for Nucleic Acid analysis)*, 2008-2012, visant à développer de nouvelles méthodes d'analyse de l'ADN.

Tout d'abord, les modifications de l'ADN seront exposées. Puis, les principales méthodes d'analyses de l'ADN et celles par spectrométrie de masse MALDI-TOF seront décrites. Enfin, le principe de ce travail, qui est le développement de méthodes d'analyse des modifications de l'ADN par clivage chimique d'une chimère ARN/ADN et par spectrométrie de masse MALDI-TOF sera détaillé.

1. Les modifications de l'ADN

1.1. L'ADN

L'ADN est une macromolécule qui est composée de nucléotides. Sa structure a été découverte en 1953 par Watson et Crick (Watson et Crick. 1953). L'ADN génomique humain est constitué de 23 paires de chromosomes et contient 3,3 milliards de nucléotides.

Les nucléotides sont composés d'un groupement phosphate (ou acide phosphorique), d'un aldopentose (2'-déoxy-D-ribose pour l'ADN et D-ribose pour l'ARN) et d'une base hétérocyclique azotée (purine ou pyrimidine).

La molécule d'ADN est constituée de quatre bases. La cytosine (C), la thymine (T) (pour l'ADN et l'uracile (U) pour l'ARN) dérivent de la pyrimidine tandis que l'adénine (A) et la guanine (G) dérivent de la purine (figure 1).

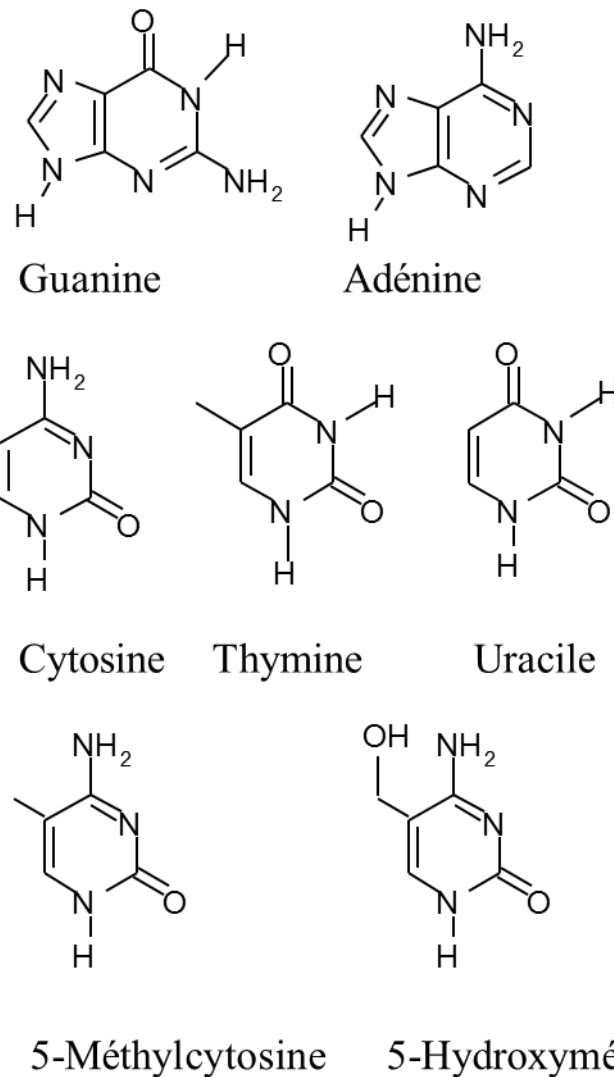


Figure 1 : Les différentes bases de l'ADN: l'adénine, la guanine, la thymine (l'uracile pour l'ARN) et la 5-méthylcytosine et la 5-hydroxyméthylcytosine qui dérivent de la cytosine.

De plus, chez les mammifères, la cytosine adjacente à base guanine formant ainsi un dinucléotide CpG, peut être méthylée. La 5-méthylcytosine est considérée comme étant la *cinquième base* mais sa fréquence dans le génome humain n'est que d'environ 1%. Enfin, la 5-hydroxyméthylcytosine a été récemment découverte dans les neurones et les cellules

souches embryonnaires de mammifères et elle est synthétisée à partir de la 5-méthylcytosine par une réaction d'oxydation catalysée par la protéine *TET* (Kriaucionis et Heintz. 2009; Tahiliani et al. 2009).

Les bases sont reliées par la liaison entre la liaison β N osidique entre le 9-N des purines ou 1-N des pyrimidines et le 1'- du sucre. Une liaison ester relie le groupement phosphate au sucre en position 3'. Les nucléotides sont liés entre eux par l'intermédiaire du groupement phosphate en position 5' du nucléoside adjacent.

L'ADN est composé d'un double-brin formant une hélice. Ces deux brins, orientés dans le sens 5' vers 3', sont reliés entre eux par des liaisons hydrogènes entre les bases purines d'un brin et les bases pyrimidines de l'autre brin. Les bases sont complémentaires deux à deux, deux liaisons hydrogènes relient l'adénine et la thymine tandis que trois liaisons hydrogènes relient la cytosine et la guanine. Les deux brins sont complémentaires et antiparallèles.

1.2. Les modifications de l'ADN

La nature, l'ordre de la séquence des bases A, C, G, T constitue un code génétique qui détermine le caractère phénotypique de chaque individu. Il existe deux catégories de modifications de l'ADN : des modifications de la séquence et les modifications épigénétiques.

1.2.1. Les modifications de la séquence

Tout d'abord, un SNP (*Single Nucleotide Polymorphism*) est une modification d'un nucléotide à une position donnée du génome, comprenant en générale deux allèles. Il représente la forme la plus abondante de variations génétiques : sa fréquence est supérieure à 1 % dans la population et il est présent environ toutes les 1000 pb dans le génome humain. Il est considéré comme un marqueur idéal pour étudier la diversité des populations, les maladies et le traitement adéquat à administrer au patient.

Un microsatellite est une séquence répétitive de 2 à 10 nucléotides. Un motif donné peut être présent à des milliers d'exemplaire dans le génome. Il est présent dans tout le génome et le plus fréquemment au niveau des introns et des exons du gène. La localisation des séquences étant bien conservée, le nombre de répétition constitue un marqueur génétique.

Il peut y avoir également une délétion ou insertion d'un ou plusieurs nucléotides.

Enfin, les mutations d'une ou plusieurs bases sont généralement spontanées mais des facteurs mutagènes, comme la radioactivité, l'environnement et certains produits chimiques peuvent les provoquer. Leurs fréquences sont inférieures à 1 % dans le génome.

1.2.2. Les modifications épigénétiques

1.2.2.1. La méthylation de l'ADN

L'ADN peut être également modifié autour de sa séquence, c'est une modification épigénétique. Ces modifications interviennent sur la structure de l'ADN et n'impliquent aucun changement de la séquence primaire. Ces mutations épigénétiques sont plus fréquentes que les mutations classiques de l'ADN.

Dans le génome humain, l'ADN est méthylé sous la forme de la 5-méthycytosine d'un dinucléotide CpG. Cette modification post-réplivative, est un transfert, par l'ADN méthyltransférase (*DNMT*), du méthyle de la S-adénosyl-L-méthionine (*SAM*) alors convertie en S-adénosylhomocystéine (*SAH*) sur une cytosine d'un dinucléotide CpG (Smith et al. 1992; Svedruzić 2008).

1.2.2.2. Les îlots CpGs

De 60 à 90% des dinucléotides CpGs sont méthylés et sont localisés dans les régions de chromatines condensées. Ils sont par conséquent inaccessibles au processus de transcription. Cependant, des regroupements de dinucléotide CpGs d'environ 1 à 4 kb, sont appelés îlots CpG et constituent des marqueurs épigénétiques.

Les îlots CpGs ont une teneur en dinucléotide CpGs de plus de 50%, une longueur supérieure à 200 pb et un ratio supérieur à 0,6 du nombre observé de dinucléotides CpG par rapport au nombre de bases C et G présentes dans le segment (Gardiner-Garden et Frommer. 1987). Ils sont concentrés en majorité dans les régions promotrices et celles du premier exon de la plupart des gènes. Environ 30 000 îlots CpG, se répartissent en moyenne tous les 100 kb recouvrant ainsi environ 1 à 2 % du génome (Antequera et Bird. 1993). 75 % des sites de

transcription et 88 % des promoteurs actifs sont associés à des îlots CpGs et pourraient être régulés par la méthylation de l'ADN.

Les régions différenciellement méthylées (DMR) vont essentiellement agir sur la régulation de l'expression du gène. Elles peuvent se produire spontanément ou en réponse à l'environnement et conduisent à des changements des états d'activation ou d'inhibition des gènes. Leurs études suscites depuis plusieurs années un grand intérêt car elles permettent notamment l'étude de maladies comme les syndromes liés à la perte d'empreinte parentale et certains cancers d'étiologie non identifiés (Robertson 2005).

1.2.2.3. Les marqueurs épigénétiques potentiels

Par ailleurs, les régions situées à proximité (environ 2 kb) des îlots CpGs qui contiennent une plus faible densité sont appelés îlots CpG *Shore*. Ils seraient en relation dans certains cancers avec les DMR (Irizarry et al. 2009). Ces régions présentent une forte corrélation inverse entre la méthylation et l'état de l'expression des gènes. Les DMR associés à ces îlots CpGs *Shore* peuvent distinguer différents types de cancers (Hansen et al. 2011).

L'étude des marqueurs épigénétiques pourrait donc être étendue à l'étude du méthylome entier, plutôt de se concentrer que sur l'étude seule des îlots CpGs. Le méthylome contient des régions de moins en moins denses d'îlots CpGs : les îlots CpGs *Shore* (adjacents aux îlots CpGs), les îlots CpGs *Shelf* (adjacents aux îlots CpGs *shore*) et les îlots *Open-sea* ou *Ocean* (isolés dans le génome).

Enfin, des études récentes suggèrent que la 5-hydroxyméthylcytosine pourrait être également un marqueur épigénétique (Ruzov et al. 2011) et avoir un rôle biochimique en tant qu'intermédiaire dans le procédé de déméthylation de l'ADN (He et al. 2011; Maiti et Drohat 2011).

2. Les méthodes d'analyse de l'ADN

Différentes méthodes de séquençage de l'ADN ont été développées : le séquençage de *première génération* : par la méthode de *Sanger* et par la méthode de *Maxam et Gilbert* ; le séquençage de *deuxième génération* : sur phase solide, par hybridation, par spectrométrie de

masse (partie 3.2), cyclique sur puce, par microélectrophorèse et par nanopore et le séquençage de *nouvelle génération* NGS (*Next Generation Sequencing*) est également en cours de développement car ce domaine est en pleine mutation (Shendure et al. 2011).

Par ailleurs, il est difficile d'affirmer avec certitude quelle sera la méthode la plus adaptée pour une problématique donnée. Toutefois, certains paramètres sont importants pour le choix de la méthode adéquate, comme le coût, l'exactitude des données, la fiabilité de la méthode, la capacité haut-débit et la longueur des séquences. Le développement de protocoles robustes, simples pour la construction de bibliothèque et la création d'outils bio-informatiques pour l'interprétation des quantités massives de données sont des défis majeurs pour l'essor de nouvelles méthodes d'analyse de l'ADN.

2.1. Le séquençage de *première génération*

Dans les années 1970, les deux premières méthodes de séquençage de l'ADN ont été développées indépendamment. L'approche de *Sanger* est une synthèse enzymatique tandis que celle de *Maxam et Gilbert* est une dégradation chimique. Frederik Sanger et Walter Gilbert ont obtenu le prix Nobel de chimie en 1980 pour cette découverte.

2.1.1. Le séquençage de *Sanger*

Le séquençage par didésoxy de Sanger ou séquençage enzymatique (Sanger et al. 1977; Sanger et al. 1977; Sanger 1988) implique une ADN polymérase, ADN dépendant, synthétisant une copie complémentaire du simple-brin d'ADN à partir de l'extrémité 3' de l'amorce. Le principe de la méthode est illustré dans la figure 2.

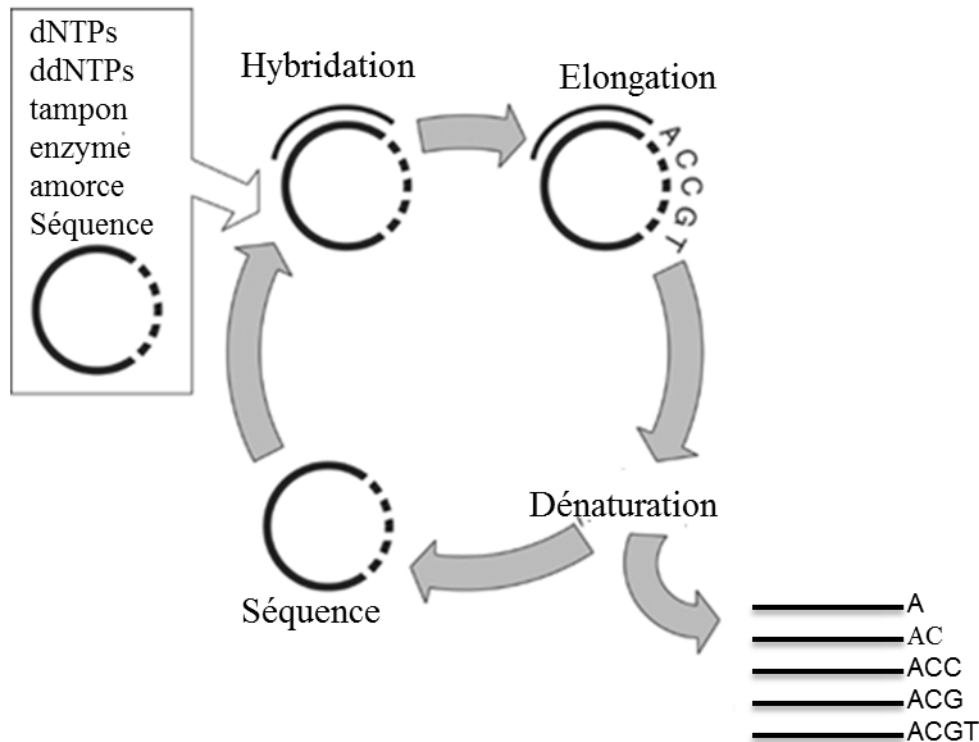


Figure 2 : Schéma de principe du séquençage didésoxy de Sanger, reproduit et modifié d'Applied Biosystems à l'adresse internet suivant : (http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_041003.pdf).

Lors de l'élongation linéaire, le désoxynucléotide ajouté est complémentaire du nucléotide du brin d'ADN. La création d'un pont phosphodiester entre le groupement 3'-OH de l'extrémité de l'amorce et le groupement 5'-phosphate du désoxynucléotide incorporé allonge la chaîne. Les quatre désoxyribonucléotides (dATP, dCTP, dGTP et dTTP) sont ajoutés, ainsi qu'une faible concentration de l'un des quatre didésoxyribonucléotides (ddATP, ddCTP, ddGTP ou ddTTP). Les didésoxyribonucléotides incorporés dans le brin synthétisé empêchent la poursuite de l'élongation. La faible concentration du didésoxynucléotide par rapport au désoxynucléotide entraîne statistiquement un mélange de fragments d'ADN de tailles différentes se terminant tous par un didésoxynucléotide. Cette terminaison permet d'identifier la position de la base du nucléotide dans la séquence d'ADN. Plusieurs élongations sont réalisées en parallèle avec les quatre didésoxynucléotides différents.

Ces fragments sont ensuite séparés par électrophorèse sur gel de polyacrylamide ou sur séquenceur capillaire. La détection des fragments est réalisée par l'intermédiaire d'un marqueur radioactif ou fluorescent. Le marqueur radioactif (isotopes ^{32}P , ^{33}P ou ^{35}S) est porté

par le désoxynucléotide ou le didésoxynucléotide (Tabor et al. 1987). Le marqueur fluorescent est fixé soit sur l'amorce soit sur les didésoxynucléotides.

Le séquençage enzymatique a été le plus répandu pendant de nombreuses années et a permis le séquençage du projet du génome humain.

2.1.2. Le séquençage de *Maxam et Gilbert*

La méthode de *Maxam et Gilbert* est une méthode de dégradation chimique de l'ADN. Chaque base A, C, G et T possèdent des réactivités différentes et peuvent donc être modifiées afin d'être clivées sélectivement (Maxam et Gilbert 1977). La séquence du brin d'ADN est déterminée à l'aide de l'assemblage de l'ordre de coupure des bases des fragments de clivage. Cette méthode est constituée de 6 étapes. Un marquage radioactif par le ^{32}P est réalisé sur les extrémités des deux brins d'ADN. Le fragment d'ADN à étudier est alors sélectionné et les deux brins d'ADN sont ensuite séparés. Les bases d'ADN sont modifiées spécifiquement et parallèlement : les bases G sont alkylées, les bases A et G sont dépurinées et les bases C et T sont hydrolysées. La pipéridine permet alors de cliver le brin d'ADN après chaque base modifiée. L'analyse des fragments de clivage est similaire à celle de la méthode de Sanger. Les fragments d'ADN sont séparés par électrophorèse capillaire afin de reconstituer la séquence d'ADN.

Les inconvénients de cette méthode sont l'analyse de fragments d'ADN de moins de 250 pb et l'utilisation de réactifs chimiques toxiques.

2.2. Le séquençage de *seconde génération*

2.2.1. Le séquençage sur phase solide

L'innovation de cette méthode est de générer des simple-brins d'ADN (Hultman et al. 1989; Jones et al. 1991; Kaneoka et al. 1991; Zimmerman et al. 1992.) par la capture sur une phase solide. Un des deux brins d'ADN est biotinylé via une réaction de PCR pour laquelle une des deux amorces est biotinylée. La molécule d'ADN hémi-biotinylé est alors fixée sur des billes

magnétiques de streptavidine, formant un complexe spécifique avec la biotine. Les brins sont séparés par lavage des billes par une solution basique. Les brins biotinylés sont alors séparés par l'intermédiaire d'un aimant. Les réactions de séquençage peuvent être effectuées indépendamment : par le brin biotinylé, par le brin non biotinylé ou par les deux.

Le séquençage chimique sur phase solide par capture au niveau des terminaisons didésoxynucléotides élimine les inconvénients du marquage chimique de l'amorce et des terminaisons, les amorces fluorescentes étant couplées à la phase solide par capture des terminaisons didésoxynucléotides (Ju. 1999). Après la réaction de séquençage, les fragments d'ADN biotinylés sont capturés par les billes magnétiques recouvertes de streptavidine, tandis que les autres composants de la réaction de séquençage sont lavés. Seuls les brins constitués de la terminaison didésoxynucléotide sont libérés des billes magnétiques et donc séquencés.

Cette méthode permet d'éviter l'élongation possible de l'autre brin complémentaire de l'ADN par l'amorce et ainsi d'améliorer la fiabilité des résultats.

2.2.2. Le séquençage par hybridation sur puce

Le principe du séquençage par hybridation (*SBH*) repose sur l'utilisation de puces à ADN. Elles sont constituées de multiples sondes, utilisées pour décoder la séquence d'ADN. L'ADN est fractionné en de multiples fragments et hybridé sur la puce à l'aide de sondes complémentaires. La puce d'ADN est ensuite analysée par la détection des oligonucléotides hybridés. L'analyse par informatique des différentes portions de la séquence d'ADN permet ensuite de reconstituer la séquence entière de l'ADN.

Le principal avantage de cette approche est le re-séquençage du génome de l'individu par la comparaison de sa séquence avec celle d'une séquence de référence. Le même concept est utilisé par plusieurs plates-formes de génotypage (*Affymetrix et Perlegen-Pfizer*) sur puce mais seuls les polymorphismes prédéfinis peuvent être détectés alors qu'avec cette méthode tous les polymorphismes sont identifiés.

L'absence de recouvrements redondants de la séquence pour assurer la fiabilité des analyses contrairement à d'autres méthodes de séquençage constitue une limitation de cette méthode. Il peut en résulter un taux élevé de faux positif (3 %) (Patil et al. 2001).

Cette méthode est très précise pour l'étude de petits génomes haploïdes (Gresham et al. 2006) contrairement à l'étude des grands génomes diploïdes du fait de leur hétérogénéité. Une approche en deux étapes a été développée pour l'étude des génomes bactériens (Albert et al. 2005; Herring et al. 2006) : la découverte de l'emplacement approximatif des mutations et suivie de la cartographie plus précise des mutations.

2.2.3. Le séquençage cyclique sur puce

Les plates-formes actuelles de séquençage (*de 454/Roche, Illumina, Applied Biosystems / Life Technologies, Dover, Helicos Biosciences, et Ion Torrent-Life Technologie*) ne reposent pas sur le séquençage de Sanger mais sur le principe d'un séquençage cyclique sur puce (figure 3) (Shendure et Ji 2008; Metzker 2009).

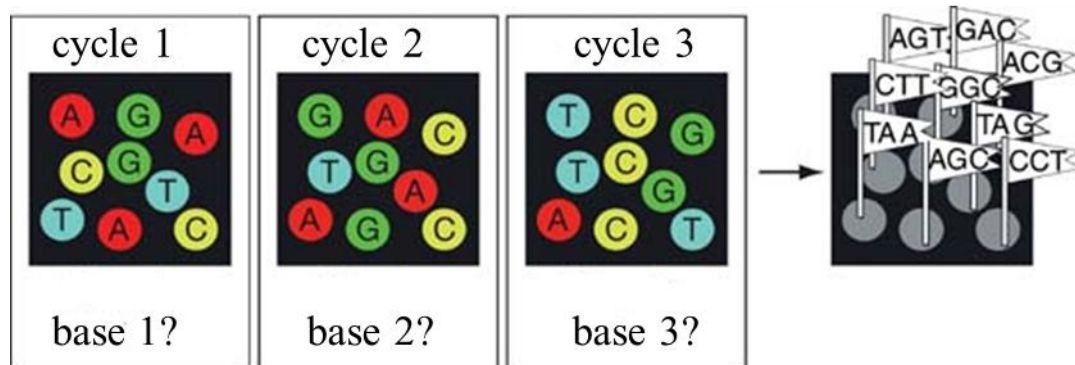


Figure 3 : Schéma de principe du séquençage cyclique sur puce, reproduit et modifié du site internet à l'adresse suivante : <http://www.currentprotocols.com/protocol/mb0701>.

Il permet de faibles coûts par le décodage bidimensionnel sur puce de millions (potentiellement de milliards) de bases de différents fragments d'ADN. Ces fragments d'ADN, de séquences individuelles, peuvent être séquencés.

Le séquençage est *clonal* car chaque fragment contient une seule espèce d'ADN (une seule molécule ou plusieurs identiques) immobilisée sur la puce. Les fragments peuvent être disposés de façon ordonné ou aléatoire. Chaque séquence d'ADN comprend généralement une séquence inconnue d'intérêt corrélée avec celles des adaptateurs universels. Un point important de cette approche est de pouvoir utiliser un volume de réactif unique sur la puce et de tout réaliser en parallèle grâce aux fragments analysés sur une même surface.

Le séquençage est cyclique car à chaque cycle une étape enzymatique identifie la nature et la position de la base de chaque séquence en parallèle. L'identification s'effectue par fluorescence d'un groupement fixé, ou par la mesure de la libération de proton. À la fin de chaque cycle, les données sont acquises par un capteur CCD (*Charged Coupled Device*) ou par une puce à semi-conducteur agissant comme un pH-mètre (*Torrent Ion*). Les cycles suivants déterminent les positions des bases. Une séquence pour chaque fragment d'ADN peut être déterminée par l'analyse de la série complète des données couvrant sa position.

Les différentes plates-formes du séquençage cyclique sur puce reposent sur le même principe de base mais diffèrent par la méthode utilisée pour générer les fragments d'ADN, par la biochimie utilisée pour effectuer le séquençage cyclique et par la méthode de détection.

2.2.4. Le séquençage par microélectrophorèse

Le séquençage classique de Sanger est réalisé avec des volumes de réactifs de l'ordre du microlitre. Un des objectifs de la microélectrophorèse est l'utilisation de la miniaturisation développée dans l'industrie des semi-conducteurs pour permettre une miniaturisation du séquençage classique didésoxy (Paegel et al. 2003). Par exemple, une nanoélectrophorèse sur gel par canaux sur la surface d'une plaquette de silicium (Emrich et al. 2002) a été développée.

L'objectif le plus important est l'intégration d'une série d'étapes de séquençage liées (par exemple, l'amplification par PCR, la purification du produit, et le séquençage) dans un *lab-on-a-chip* (Blazej et al. 2006; Forster et al. 2008.).

L'avantage de cette approche est l'application du séquençage classique ayant prouvé sa robustesse et son efficacité pour des séquences de plus de 1000 bases. Cependant, le séquençage par microélectrophorèse peut s'avérer critique pour la poursuite de la tendance à la réduction des coûts.

2.2.5. Le séquençage par nanopore

Une nouvelle approche de séquençage, initialement proposé dans les années 1980, consiste dans le passage d'un seul brin d'ADN à travers un nanopore (Deamer et Akeson 2000) (figure 4).

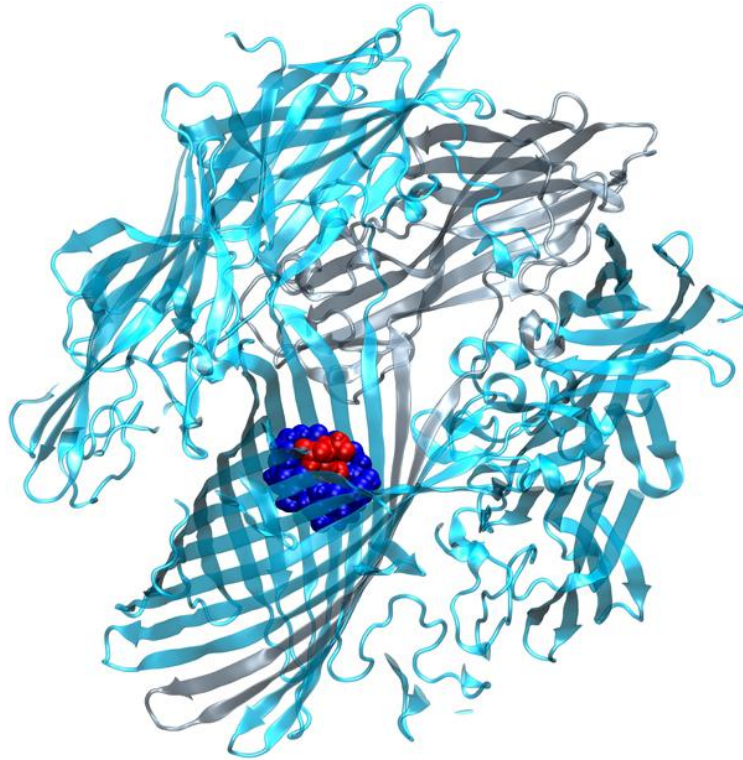


Figure 4 : Image d'un nanopore reproduite avec l'accord d'Oxford Nanopore Technologies.

Un nanopore est une protéine membranaire : l' α -hémolysine (Kasianowicz et al. 1996), la *MspA* (*Mycobacterium smegmatis* porin A) (Derrington et al. 2010) ou une protéine synthétique (Fologea et al. 2005). Il est modifié chimiquement par fixation covalente d'une cyclodextrine sur sa surface intérieure. Cette modification constitue un site de liaison pour les bases de l'ADN simple-brin et permet une mesure précise de son passage à travers le nanopore.

Deux méthodes sont en cours de développement, une méthode exonucléase (Hornblower et al. 2007) et une méthode de polymérisation (Lieberman et al. 2010). Les quatre nucléotides obstruent le nanopore de manière spécifique. La variation de la conductance à travers le nanopore peut alors être mesurée pour en déduire la séquence d'ADN.

Le séquençage par nanopore présente un grand potentiel dans le séquençage rapide et rentable de mélanges d'ADN par l'intermédiaire de préparations d'échantillons relativement simple. Cependant, le développement technologique de cette méthode sera nécessaire pour garantir l'exactitude de l'analyse de grandes séquences d'ADN à partir d'un mélange complexe d'ADN.

2.3. Le séquençage de nouvelle génération NGS

De nouvelles méthodes sont en cours de développement. Il s'agit notamment du séquençage par microscopie électronique à transmission (*TEM*) (Thomas et al. 2008.; Tanaka et Kawai 2009) et du séquençage optique (Xiao et Kwok. 2008; Zhou et al. 2008).

Dans la méthode de séquençage *TEM*, de longs fragments d'ADN sont disposés sur les grilles des puces ayant incorporés différents nucléotides contenant un nombre différents de protons. Le différentiel de charge est détecté et analysé.

Dans le séquençage optique, de grands fragments d'ADN sont disposés sur des lames de verre. Des étapes enzymatiques permettent la création de *trou* de 20 à 50 pb dans l'ADN ou l'ADN peut être également *tendu* en raison de son élasticité. Les trous peuvent être comblés par une ADN polymérase et des nucléotides marqués par fluorescence qui sont ensuite analysés par microscopie.

3. Les méthodes d'analyse de l'ADN par spectrométrie de masse

Au cours de la dernière décennie, la spectrométrie de masse s'est imposée pour l'analyse du domaine émergent de la protéomique mais elle est également utilisée dans le domaine de la génomique.

Les acides nucléiques, sont des biopolymères polaires et sont analysés notamment par une source ESI (*Electrospray Ionization*) ou par une source MALDI (*Matrix-Assisted Desorption/Ionisation*). Ces deux techniques d'ionisation/désorption permettent d'analyser des molécules de quelques dizaines de femtomoles à quelques picomoles. Les oligonucléotides sont plus difficiles à analyser que les protéines. Tout d'abord, ils ont tendance à former des adduits cationiques avec les ions sodium ou potassium présents dans l'échantillon, formant des liaisons ioniques avec les groupements phosphodiester. De plus, ils ont tendance à se fragmenter facilement en perdant les bases nucléiques. La stabilité des oligonucléotides est plus importante par ionisation MALDI que par ionisation ESI. Enfin, les oligonucléotides peuvent être analysés en mode positif et négatif mais, en mode négatif, le pouvoir résolutif et la sensibilité sont bien meilleurs surtout par ESI.

3.1. La spectrométrie de masse ESI-IT et MALDI-TOF

3.1.1. La spectrométrie de masse ESI-IT

La source ESI est généralement couplée à un analyseur à piège ionique (ou trappe) (*IT : Ion Trap*) (figure 5).

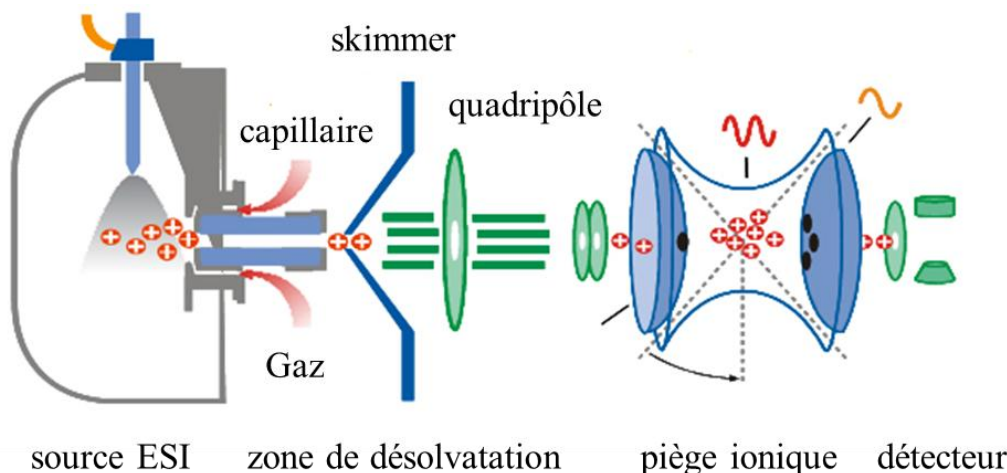


Figure 5 : Schéma de principe d'un Esquire™ ESI-ITMSⁿ reproduit et modifié de Bruker Daltonics®.

3.1.1.1. La source ESI

La source ESI a été développée par Fenn (Yamashita et Fenn 1984; Fenn et al. 1989). Elle permet d'analyser des macromolécules telles que les polymères et les biopolymères mais aussi des petites molécules polaires. Elle est la plus répandue car elle est très sensible et qu'elle peut-être couplée à la chromatographie liquide ou à l'électrophorèse capillaire permettant ainsi l'étude de mélanges complexes.

3.1.1.1.1. Le principe

Lors de l'ionisation/désorption par ESI, une différence de potentiel de 3 à 6 kV entre le capillaire et la contre-électrode, séparés de 0,3 à 2 cm, génère un champ électrique. Ce champ est appliqué, à pression atmosphérique, sur une solution nébulisée à partir d'un tube capillaire à faible débit (1 à 100 $\mu\text{L}/\text{min}$). Il engendre une accumulation de charges à la surface du liquide à l'extrémité du capillaire, produisant ainsi un brouillard (ou spray) (Yamashita et Fenn 1984; Smith et al. 1990; Loo et al. 1989; Ikonomou et al. 1990).

Lorsque la tension électrique est faible les gouttelettes semblent sphériques mais lorsque la force exercée est supérieure à la tension superficielle, la surface se rompt pour former un *cône de Taylor* et un spray de gouttelettes. Fortement chargées, elles sont constituées notamment de

l'analyte et de l'électrolyte de la solution introduite dans le capillaire et des ions en excès. La charge de ces ions dépend du potentiel appliqué au capillaire par rapport à la contre-électrode (ions positifs correspond à une différence de potentiel positif...) et des propriétés des solvants.

L'évaporation du solvant de chaque gouttelette provoque leur rétrécissement et augmente leur densité de charge jusqu'à leur explosion coulombienne conduisant à la désorption des ions (Kearle et Tang 1993). Ces ions solvatés possèdent un grand nombre de charges, dépendant des sites ionisables de la molécule qui permettent les échanges de protons ou la formation d'ions cationiques ou anioniques.

Le flux d'ions produit dans l'électrospray doit être compensé par un courant d'électrons au niveau du capillaire (Kearle et Tang 1993). L'électrospray est donc un processus électrochimique qui produit un courant d'oxydation ou de réduction à l'extrémité du capillaire permettant ainsi de fixer la quantité d'ions formés. Le courant ionique total atteint généralement environ 1 μA .

3.1.1.1.2. Les facteurs influençant l'ionisation/désorption ESI

L'électrospray est une source d'ions à courant constant (Van Berkel et Zhou. 1995). Le courant est donc partagé par tous les constituants de la solution engendrant ainsi une diminution du signal de l'analyte. Par conséquent, la solution doit être dessalée afin de permettre une meilleure ionisation.

Le flux et la concentration d'ions doivent être suffisants afin d'assurer un courant ionique constant. Si le flux n'est pas suffisant, il faudra produire plus d'ions par oxydation (en mode positif) ou par réduction (en mode négatif). S'il est trop important, le signal diminue car le courant est constant et il devient difficile de maintenir un bon spray dans ce cas.

Le courant dû à l'analyte varie linéairement avec sa concentration jusqu'à se rapprocher de celui du courant constant. Comme le courant reste constant, le courant de l'analyte rentre en compétition avec celui des autres ions pour le partage de ce même courant à cause du facteur cinétique de chaque espèce. Le courant des autres ions va donc décroître quand la concentration de l'analyte augmente.

De plus, s'il y a compétition entre deux espèces, celle qui est la plus hydrophobe permet une meilleure désorption. Il peut en résulter une disparition d'une des deux espèces du spectre de masse mais en diminuant la concentration elle pourra être de nouveau détectable.

Ce même phénomène de compétition entre deux espèces est observé en maintenant la concentration mais à un très faible flux (nanospray).

3.1.1.1.3. Les ions multichargés

L'ionisation/désorption par ESI de grosses molécules possédant des sites ionisables produit des ions multichargés (une charge pour environ 500 Da). La formation de ces ions de rapport m/z faible permet d'améliorer la sensibilité et ainsi analyser des masses moléculaires très importantes allant au-delà des limites en masse (m/z avec $z=5$) de l'instrument.

Les spectres de masse ESI de grosses molécules correspondent généralement à une distribution de pics consécutifs caractérisant des espèces moléculaires multichargées par protonation $(M+zH)^{z+}$ ou déprotonation $(M-zH)^{z-}$ de rapport m/z . Les rapports m/z de deux ions d'une même molécule, multichargée mais qui diffèrent par un état de charge, permettent de déterminer la masse de la molécule. Des algorithmes permettent la déconvolution du spectre de masse en transformant les pics des ions qui correspondent aux molécules multichargées en formes moléculaires monochargées afin d'en déterminer la masse moléculaire.

De plus, seuls les mélanges simples peuvent être directement analysés à cause des nombreux ions multichargés qui se superposent et rendent le spectre de masse ininterprétable. Seuls les instruments à très hautes résolutions permettent de lever les ambiguïtés. Les mélanges plus complexes sont analysés par couplage avec la chromatographie liquide afin de séparer les différentes molécules.

3.1.1.2. L'analyseur à piège ionique

Il existe trois types d'analyseur à piège ionique (ou trappe) : le piège ionique classique (ou 3D), le piège ionique linéaire (ou 2D) et le piège ionique quadrolarithmique à transformée de Fourier (FT) électrostatique (ou orbitrap).

3.1.1.2.1. Le piège ionique classique

3.1.1.2.1.1. Le principe

Le piège ionique classique, 3D, est le plus ancien et le plus répandu. Il est constitué d'une électrode circulaire fermée aux extrémités par deux calottes sphériques (Stafford et al. 1984) (figure 5). La superposition de tensions continue et alternative (RF) permet de créer un champ quadripolaire à trois dimensions où les ions sont piégés sur des trajectoires complexes (3 D).

Tous les ions présents simultanément dans le piège ionique, sont ensuite éjectés en fonction de leur masse par augmentation de l'amplitude de la tension RF. L'éjection résonante des ions se fait par application sur les calottes de la fréquence de l'ion. L'amplitude du mouvement de l'ion augmente donc jusqu'à absorber une quantité d'énergie suffisante pour être éjecté du piège ionique. Lors de l'enregistrement du spectre de masse, les ions sont toujours éjectés à une fréquence résonante qui correspond à un champ particulier. Pour une faible gamme de rapport m/z, ils sont éjectés au champ quadripolaire qui correspond en générale à la moitié de la fréquence de la tension RF tandis que pour une gamme de rapport m/z plus élevée, ils sont éjectés au champ hexapolaire qui correspond à une fréquence d'éjection plus faible.

Deux approches mathématiques complémentaires sont proposées pour expliquer le fonctionnement du piège ionique. L'approche basée sur la recherche des solutions de l'équation de Mathieu qui conduit aux équations suivantes (valables si $r_0^2 = 2z_0^2$) :

$$a_z = \frac{-8zeU}{m\omega^2 r_0^2} \quad q_z = \frac{4zeV}{m\omega^2 r_0^2}$$

Une autre approche mathématique basée sur les méthodes de Floquet et de Fourier décrit le mouvement des ions à l'aide d'un terme e^μ (avec $\mu = \alpha + i\beta$) pour trouver les conditions d'une trajectoire stable. Les conditions $\alpha = 0$ et β compris entre 0 et 1 (β nombre non entier pour maintenir les ions dans le piège ionique) doivent être remplies simultanément suivant r et z. Si $\beta = 1$, c'est l'éjection quadripolaire.

3.1.1.2.1.2. La source ESI et le piège ionique 3D

L'utilisation d'une source ESI et de l'éjection résonante permet d'analyser des ions de rapports m/z maximum de 6000 Th pour un *Esquire*TM de Bruker et d'étudier les fragmentations en phase gazeuse des ions par expériences MSⁿ séquentielles.

Pour que les ions restent stockés dans le piège ionique, ils doivent être ramenés au centre du piège ionique malgré les répulsions coulombiennes au sein du nuage d'ions. Si les ions sont trop éloignés du centre, la rencontre de résonance non linéaire apparaît et conduit à l'éjection non contrôlée des ions. Le gaz, d'hélium est introduit dans le piège ionique pour engendrer de multiples collisions avec les ions afin de les ramener au centre du piège ionique par une relaxation cinétique.

La présence d'hélium dans le piège ionique permet également d'exister un ion donné en lui appliquant sa fréquence propre au niveau des calottes. L'amplitude des ions augmentent sans modifier leur fréquence permettant ainsi d'augmenter leur énergie cinétique. Les ions acquièrent donc de l'énergie interne afin de favoriser le processus CID (*Collision Induced Dissociation*).

En les expulsant sélectivement, l'analyse MS/MS d'un ion de rapport m/z par processus consécutifs choisis est réalisée. Des ions précurseurs peuvent être sélectionnés. La spectrométrie de masse en tandem est donc réalisée par séquences temporelles et pas dans l'espace comme pour l'analyseur à temps de vol ou multiple quadripôle.

3.1.1.2.2. Les autres pièges ioniques

Le piège ionique linéaire ou 2D est constitué de 4 barres analogues à celle d'un quadripôle, fermé par des électrodes d'entrée et de sortie permettant de repousser les ions vers l'intérieur du quadripôle afin de diminuer les répulsions coulombiennes (Douglas et al. 2005). Il possède les mêmes caractéristiques que le piège ionique classique avec quelques modifications des tensions d'entrée et de sortie afin d'éviter les effets de champs aux limites (*fringe fields*).

L'orbitrap est composé d'une électrode externe en forme de tonneau qui est séparée par un espace étroit en deux parties égales (Makarov. 2000). L'électrode centrale a une forme de fuseau. Les ions sont injectés dans le piège ionique au niveau d'une des extrémités par un champ quadrol logarithmique. Ce champ conduit à la rotation des ions autour du fuseau avec un mouvement axial d'aller et retour formant ainsi une spirale. Pour une gamme de rapport

m/z élevée, la spirale est constituée d'un grand rayon et d'une faible amplitude tandis que pour une gamme de rapport m/z faible, la spirale possède un plus faible rayon et une amplitude plus grande. Les ions de rapport m/z faible ont donc une fréquence plus grande que ceux de rapport m/z plus élevés.

Le signal est analysé au niveau du signal transitoire. Le signal transitoire doit durer le plus de temps possible afin d'obtenir une meilleure résolution. Le pouvoir résolutif, après transformée de fourrier, peut être supérieur à 100000 Rs. Ce piège ionique, peu sensible aux effets de charge de l'espace permet d'analyser un grand nombre d'ions.

3.1.2. La spectrométrie de masse MALDI-TOF

La source MALDI est une désorption-ionisation laser assistée par matrice (Karas et al. 1985; Karas et al. 1987; Karas et Hillenkamp 1988; Karas et Bahr. 1990). Historiquement, elle est principalement associée à un analyseur à temps de vol (TOF) car il nécessite une source d'ionisation pulsée.

Cette technique de désorption et d'ionisation douce permet l'analyse des biomolécules (oligonucléotides, peptides, protéines et oligosaccharides) de haute masse moléculaire (MDa) (Tanaka et al. 1988). Facile à mettre en œuvre, elle présente une certaine tolérance aux sels et aux tampons utilisés en biologie moléculaire. Elle est donc adaptée à l'étude de biomolécules comme l'ADN et aux mélanges.

Le spectrométrie de masse MALDI-TOF se compose: d'une source MALDI produisant les ions en phase gazeuse, d'un analyseur à temps de vol séparant les ions en fonction de leur rapport m/z , d'un détecteur convertissant un signal ionique en signal électrique et du traitement du signal pour obtenir un spectre de masse des ions en fonction du rapport m/z (figure 6).

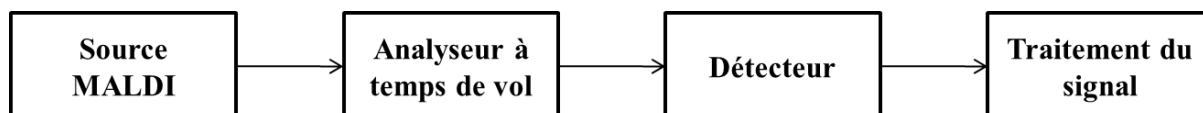


Figure 6 : Schéma de principe de la spectrométrie de masse MALDI-TOF.

3.1.2.1. La source MALDI

Le principe de l'ionisation/désorption MALDI repose sur le mélange en solution de l'analyte et de la matrice. La matrice, molécule organique de faible masse moléculaire, absorbe à la longueur d'onde du laser utilisé. Le mélange est déposé sur une cible métallique sur laquelle le mélange sèche afin de co-cristalliser (figure 7).

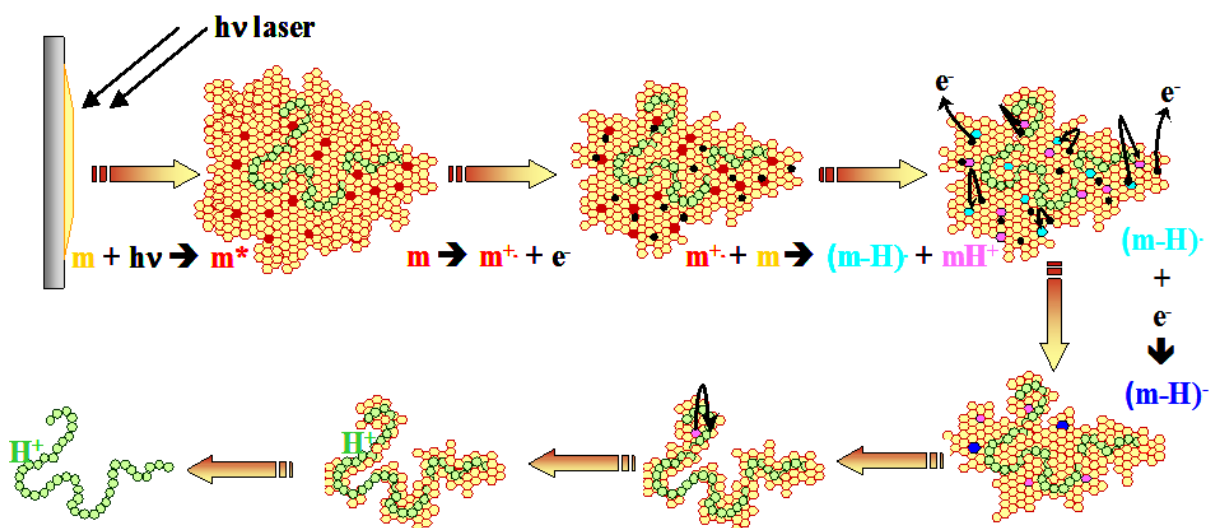


Figure 7 : Schéma de principe de la source MALDI (Livadaris 2002).

La cible est ensuite placée dans la source MALDI, sous vide (10^{-6} à 10^{-7} torr). La matrice et l'analyte sont irradiés par un court impact (environ 5 ns) d'un faisceau laser à impulsion afin de les désorber de la surface de la cible et de les éjecter en phase gazeuse. Il se forme alors, en phase gazeuse, des agrégats de molécules de matrice et d'analyte appelé panache au sein duquel se forment les ions. L'analyte est donc ionisé, sous forme protonée ou déprotonée ou de cations alcalins. Il se forme majoritairement des ions monochargés.

3.1.2.1.1. La matrice

La matrice est un élément essentiel car elle intervient dans le processus de formation des ions en phase gazeuse. Elle absorbe à la longueur d'onde du laser par excitation électronique à l'aide d'un rayonnement avec un laser ultraviolet d'impulsion ou par excitation vibrationnelle par utilisation d'un laser infrarouge. Le choix de la matrice appropriée est fonction de l'analyte à étudier car les interactions matrice-analyte jouent un rôle dans le processus d'ionisation et de désorption.

Le choix de la matrice repose sur différents critères empiriques. Elle doit absorber à la longueur d'onde du laser, être chimiquement inerte et ne doit pas sublimer dans les conditions de vide poussé. La préparation du mélange matrice-analyte est aussi très importante afin d'obtenir une bonne cristallisation. L'analyte doit être dilué dans une grande quantité de matrice (rapport molaire de matrice/analyte de 10^3 à 10^5) afin que les interactions entre les molécules d'analyte soient réduites, facilitant ainsi leur transfert en phase gazeuse (Horneffer et al. 2006). Elle doit posséder une fonction acide permettant le transfert de proton dans l'étape d'ionisation.

La matrice peut jouer un rôle plus ou moins important dans la fragmentation des ions métastables mode PSD (*Post Source Decay*) (Karas et al. 2000). Il existe des matrices dites *froides* induisant facilement des fragmentations à faible fluence du laser et des matrices dites *chaudes* induisant peu de fragmentation (Karas et al. 1995; Luo et al. 2002).

Le choix de la matrice la mieux adaptée à l'analyte étudié est crucial. Elles peuvent être solides, solides ioniques ou liquides. Les matrices solides sont les plus utilisées. Ce sont des composés simples de petite taille portant un chromophore UV et qui possèdent le plus souvent des groupements aromatiques. Cependant, elles engendrent une mauvaise reproductibilité du spectre de masse entre deux impacts du laser consécutifs dû à la cristallisation du dépôt. Il est donc très important de choisir la méthode de préparation de l'échantillon la mieux adaptée. De plus, les dépôts ont une très courte durée de vie. Les matrices solides ioniques sont constituées d'une matrice solide et d'une base organique comme la pyridine (Armstrong et al. 2001; Li & Gross 2004; Tholey & Heinzle 2006). Les solutions de matrice sont donc plus homogènes que celles des matrices solides, permettant ainsi une meilleure reproductibilité des spectres de masse obtenus entre deux impacts du laser sur le même dépôt. Les matrices liquides montrent également une bonne reproductibilité des spectres de masse obtenus entre deux impacts de laser consécutifs. Cependant, elles n'absorbent pas ou peu dans l'ultraviolet et le visible, elles sont donc utilisées avec un laser absorbant à des longueurs d'ondes inférieures à 300 nm.

Les matrices utilisées pour l'analyse des oligonucléotides ne le sont généralement pas pour l'analyse des peptides et protéines et des polymères de synthèse (tableau 1).

| Matrices | Abréviations |
|---|---------------|
| Acide-3 hydroxypicolinique | HPA |
| 2,4,6 trihydroxyacétophénone | THAP |
| Acide anthranilique | AA |
| Acide picolinique | PA |
| Acide nicotinique | NA |
| α -cyano-4-hydroxycinnamate de méthyle | α CNMe |
| Acide 2,5-dihydroxybenzoïque | DHB |
| Acide quinaldique | QA |
| 3,4-diaminobenzophénone | DABP |
| Acide 5-méthoxysalicylique | MSA |
| 6-aza-2-thiothymine | ATT |
| 3-Hydroxycoumarin | 3-HC |
| Acide α -cyano-4-hydroxycinnamique | HCCA |

Tableau 1 : Quelques matrices utilisées en MALDI-UV pour l'analyse des oligonucléotides (Wu & McLuckey 2004; Tang et al. 1993; Krause et al. 1996; Zhang et al. 2006; Song 2003; Fu et al. 2006; Distler & Allison 2001).

Pour l'analyse des oligonucléotides, les matrices 3-HPA et THAP sont les plus utilisées. Les acides nucléiques, étant riches en groupements phosphodiester, favorisent la formation d'adduits de métaux alcalins (Na^+ , K^+ par exemple) perturbant la mesure de la masse de l'oligonucléotide. Pour parer à ce phénomène, les solutions de matrice contiennent généralement un additif (par exemple le citrate d'ammonium) afin d'améliorer la détection des ions. L'ion ammonium permet d'une part, de dessaler l'analyte et d'autre part, de régénérer les formes protonées dans l'agrégat.

3.1.2.1.2. La préparation de l'échantillon

La préparation de l'échantillon est donc une étape cruciale lors de l'analyse par MALDI-TOF MS. Elle dépend de la nature de l'analyte, de la sensibilité requise et du type d'information recherché. Le choix de la matrice, des solvants et du type de dépôt, doivent donc être optimisés pour chaque préparation d'échantillon. Différentes méthodes de préparation existent selon la matrice et l'analyte : la méthode de la goutte séchée, la méthode de la couche mince et la méthode du dépôt en sandwich.

Lors de la méthode de la goutte séchée, la matrice est solubilisée le plus souvent dans l'acétonitrile. La solution de matrice est ensuite mélangée à l'échantillon et une goutte de cette préparation est déposée sur la surface de la cible métallique. Le mélange est soit réalisé

au préalable (pre-mix), soit effectué sur la cible elle-même. La cristallisation obtenue par séchage de la goutte à température ambiante peut-être accélérée par chauffage, ventilation forcée ou aspiration sous vide. Cette méthode est la plus simple à mettre en œuvre et fournit de bons résultats pour de nombreux types d'échantillons, en particulier pour l'analyse des oligonucléotides, peptides et protéines. Le principal inconvénient de cette méthode de dépôt est le phénomène de ségrégation. La distribution irrégulière des cristaux à la périphérie de la goutte, *effet Marangoni* (Amado et al. 1997) conduit à une grande variabilité des ions détectés sur le dépôt. L'accélération du séchage sous vide permet de réduire l'effet de ségrégation en réduisant la taille et en augmentant l'homogénéité des cristaux.

Lors de la méthode de la couche mince, la matrice est solubilisée dans un solvant volatil (acétone ou méthanol) (Vorm et al. 1994). Une goutte de solution de matrice est déposée sur la cible et l'évaporation rapide du solvant produit une couche mince et homogène de cristaux. Puis, l'échantillon est déposé sur cette préparation, permettant ainsi un dépôt de faible dimension. Les cristaux de matrice de la première couche servent ainsi de site de nucléation pour l'échantillon déposé ensuite. Ce type de dépôt peut être lavé afin d'éliminer les contaminants (sels, détergents..). Un solvant aqueux est ajouté sur le dépôt et rapidement ré-aspiré à l'aide d'une pipette. Cette méthode permet d'obtenir des cristaux plus petits et plus homogènes que ceux de la méthode de la goutte séchée (Dai et al. 1996; Onnerfjord et al. 1999). Cette homogénéité permet une meilleure reproductibilité (sensibilité, rapport signal sur bruit, résolution) des analyses. Cependant, la couche mince de matrice et d'échantillons est rapidement détruite après quelques tirs du laser. La reproductibilité des dépôts d'échantillons est rendu difficile par l'utilisation de solvants volatiles.

Lors de la méthode du dépôt en sandwich, une première couche de matrice est déposée selon la méthode de la couche mince (Li et al. 1996) . Une goutte d'analyte est déposée sur la couche mince et enfin une goutte de solution de matrice est déposée sur la deuxième couche. Le dépôt est alors séché à température ambiante. Un lavage du dépôt peut également être effectué avec ce type de méthode. L'analyte se retrouve disposé entre deux couches de matrice.

3.1.2.1.3. Le laser

Le laser le plus répandu est un laser à azote émettant à 337 nm. Les autres lasers les plus courants sont les lasers Nd:YAG (1060nm, fréquence triplet, 353 nm) et les lasers infrarouges comme celui à dioxyde de carbone à 10,6 μm (Zenobi & Knochenmuss 1998).

Comme les matrices absorbent généralement moins bien dans l'infrarouge, la puissance du laser IR appliquée à l'analyte doit être plus importante que pour celle d'un laser UV. Les spectres de masse obtenus par un laser IR ou UV sont globalement similaires. Le laser IR montre juste une plus grande tendance à former des ions multichargés et des adduits.

L'impact du laser sur l'analyte n'étant que de quelques nanosecondes, le phénomène de désorption/ionisation dépend donc de la fluence du laser (J/cm^2) (Zenobi & Knochenmuss 1998; Dreisewerd et al. 1995).

3.1.2.1.4. Les modèles d'ionisation

Le mécanisme de formation des ions issus d'une source MALDI est composé de deux étapes : la désorption et l'ionisation. Tout d'abord, l'absorption de photons du laser par le mélange co-cristallisé de matrice et d'analyte entraîne la désorption. Il se forme alors un panache par décompression du solide vers la phase gazeuse. Puis, les molécules d'analyte et la matrice s'ionisent. L'ionisation MALDI reste encore un phénomène mal connu aujourd'hui car il ne peut s'expliquer par un phénomène unique. Un grand nombre de paramètres expérimentaux influencent la formation des ions tels que la longueur d'onde du laser, les conditions de préparation des échantillons (et la cible), la nature de l'analyte, le mélange de matrice-analyte et le mode de détection des ions (positif ou négatif).

Trois modèles décrivent le processus MALDI. Le premier modèle est celui du contrôle thermodynamique de Zenobi et Knochenmuss (Zenobi & Knochenmuss 1998; Breuker et al. 2003; Knochenmuss et al. 2000.; Knochenmuss 2002; Knochenmuss 2003; Knochenmuss & Zenobi 2003). L'ionisation est composée de deux événements distincts : l'ionisation primaire et secondaire où se déroulent différentes réactions ion-molécule. Les processus en phase gazeuse ont lieu à un équilibre thermodynamique au sein du panache après un délai de plusieurs centaines de nanosecondes d'expansion de celui-ci.

Le second modèle, est celui proposé par Karas, du *Lucky Survivor Ions* (Karas et al. 2000; Karas et al. 2003; Karas & Krüger 2003; Krüger & Karas 2002; Glückmann et al. 2001). Il est

le premier à considérer les agrégats comme des précurseurs des ions formés en MALDI. Les espèces sont des survivants de processus de neutralisation par capture d'électrons.

Le troisième modèle, est celui proposé par le groupe Tabet, des *espèces agrégats précurseurs*, proche de celui de Karas mais qui présente une analogie avec les processus d'ionisation ESI (Fournier et al. 2002; Fournier et al. 2003; Livadaris et al. 2000). Ce modèle considère des réactions de désolvatation par perte de matrice neutre ou chargée.

L'ensemble de ces modèles, surtout les deux derniers, impliquent un processus de production d'agrégats qui désorbent de la surface au moins pendant les premières 6 nanosecondes de la fluence du laser. La taille de ses agrégats est variable mais elle peut grossir ou se former dans le panache après l'impact laser. Les ions issus de ces agrégats apparaissent donc à des temps et à des endroits différents de la source. Il en résulte une perte du pouvoir résolutif de l'instrument. Les ions n'ont pas exactement la même énergie cinétique (dépend de leur position dans la source) et peuvent être décalés dans le temps car ils sont émis à des temps différents à partir de la surface. L'énergie interne des ions ainsi libérée dépend de la taille des agrégats. Plus la taille des agrégats est grande plus l'énergie sera relaxée. Si cette énergie interne est suffisante, elle peut être la cause de fragmentation. Comme il s'agit de distribution de l'énergie des ions, les ions riches en énergie vont conduire à des fragmentations promptes dans la source tandis que les ions de plus faible énergie se décomposent dans la région de libre champ du tube de vol.

3.1.2.2. L'analyseur à temps de vol

La source MALDI est le plus généralement couplée à un analyseur à temps de vol. Cependant, elle peut être également couplée à un piège ionique (Jonscher & Yates 1997), à un analyseur à résonance ionique cyclotronique à transformée de Fourier (Li et al. 1994) ou à un analyseur à secteur magnétique (Bordoli et al. 1994). Il existe aussi des instruments hybrides : quadripôle/TOF, piège ionique/TOF ou TOF/TOF.

Le principe de l'analyseur à temps de vol (TOF Time-of-Flight) a été décrit par Stephens en 1946 et largement étudié par la suite par Cotter (Stephens 1946; Cotter 1989; Cotter 1992; Cotter 1999).

L'analyseur à temps de vol mesure la durée de transfert des ions après le tir du laser pour arriver au détecteur au travers de la région libre de champ. Le rapport m/z est donc

directement déduit de la durée de vol de l'ion. L'utilisation d'une impulsion laser permet l'émission d'ions très localisés dans l'espace et le temps. Le couplage du MALDI avec l'analyseur de temps de vol est donc idéal. Le temps de vol possède une transmission élevée permettant l'analyse simultanée de tous les ions désorbés dans une gamme de masse potentiellement illimitée.

3.1.2.2.1. L'analyseur à temps de vol linéaire

L'analyseur à temps de vol linéaire est composé d'une zone d'accélération et d'une zone libre de champ (figure 8).

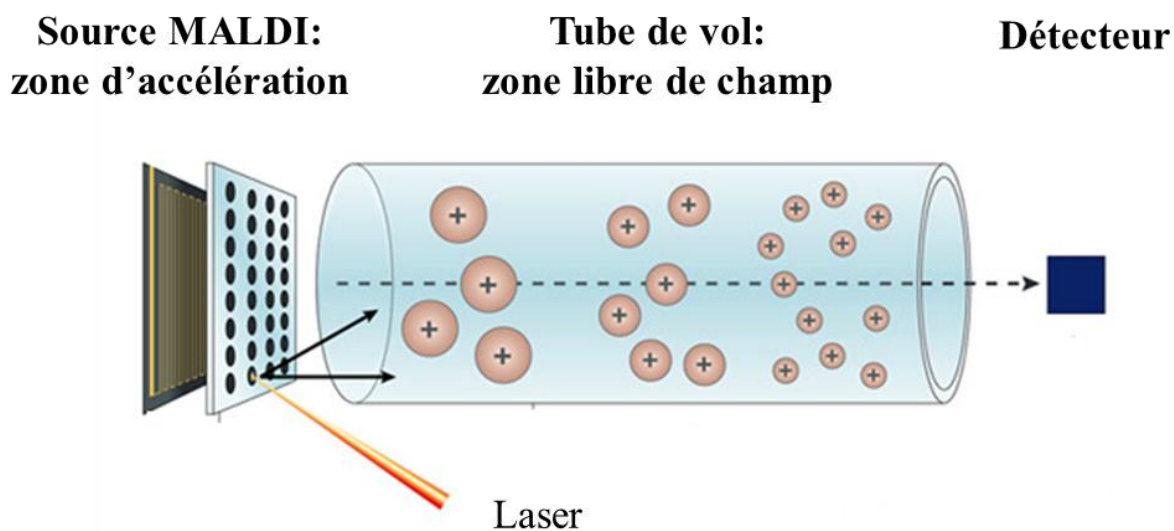


Figure 8 : Schéma d'un analyseur à temps de vol linéaire reproduit et modifié de Bruker Daltonics®.

Les ions produits dans la source MALDI, sont accélérés par une forte différence de potentiel appliquée entre la cible et une électrode d'extraction. Ils atteignent alors leur énergie cinétique finale commune à tous les ions qui parcourent la zone libre de champ afin d'être séparés temporellement.

Le tube de vol mesure le temps mis par un ion pour atteindre le détecteur à partir du moment où il est désorbé. Le rapport m/z est proportionnel au temps de vol des ions. Le temps, caractérisant l'ion, pour parcourir le tube de vol est donc (z : le nombre de charge de l'ion ; e : charge élémentaire de l'électron $1,602 \cdot 10^{-19}$ C ; v : différence de potentiel entre la cible et la zone d'extraction et L : longueur du tube de vol) : $t = L \sqrt{\frac{m}{z2ev}}$. Cette valeur est approximative car elle dépend des conditions d'analyse.

La principale limitation de l'analyseur à temps de vol linéaire est son pouvoir résolutif (Russell & Edmondson 1997; Vestal & Juhasz 1998) qui peut s'expliquer par une variation du moment de formation des ions, de leur localisation dans la zone d'accélération ou de leur énergie cinétique initiale différente. Si deux ions de même rapport m/z et de même énergie cinétique sont formés à des temps différents, ils gardent cette différence de temps lorsqu'ils parcourent la zone de libre champ. Cet inconvénient peut être limité en augmentant le temps de vol de l'ion par application d'un faible potentiel d'accélération ou en utilisant un tube de vol plus long. Lorsque deux ions de même rapport m/z atteignent le détecteur à des temps différents cela entraîne une perte du pouvoir résolutif et donc un élargissement des signaux. Ces phénomènes peuvent être expliqués par l'apparition des ions à des temps différents et de leur distribution d'énergie cinétique différente. Les ions se désorbent de la surface à des temps différents entraînant un décalage dans l'échelle temporelle. Les ions peuvent être formés à des endroits différents de la source et donc avoir des potentiels d'accélération différents. Enfin, l'énergie interne des ions peut conduire à des fragmentations dans la région libre de champ conduisant ainsi à des ions fragments de même vitesse que celle de l'ion précurseur. Ces ions sont accompagnés d'une grande dispersion de l'énergie cinétique qui conduit encore plus à un élargissement des pics des ions. Pour résoudre ce phénomène temporel et énergétique, deux solutions complémentaires sont possibles : l'extraction retardée et le réflectron.

Afin d'améliorer le pouvoir résolutif de l'analyseur à temps de vol linéaire, un délai temporel, appelé extraction retardée, est intégré entre la formation de l'ion (étape de désorption/ionisation) et son accélération afin de le focaliser sur le plan parallèle à celui du détecteur (Tabet & Cotter 1983; Brown & Lennon 1995; Vestal et al. 1995). Pendant ce délai (quelques nanosecondes à plusieurs centaines de microsecondes), les ions évoluent librement à leur vitesse initiale dans des zones différentes de la source. Un potentiel répulsif est appliqué sur une des grilles pour empêcher la sortie des ions qui sont entraînés par la haute tension. Lorsque la tension de la grille est supprimée, les ions sont tous éjectés en même temps. Les premiers ions ont été bloqués au niveau de la grille et les autres ions ont eu le temps de les rejoindre. Les ions sont donc au même endroit, en même temps sous le même champ électrique. Ils sont donc focalisés du point de vue temporel. Cependant, la distribution de l'énergie cinétique des ions restent encore relativement large due à l'effet des grilles, quand elles existent, mais surtout à la dispersion de l'énergie cinétique lors de la dissociation en vol dans la région libre de champ. Pour corriger ce phénomène, un miroir électrostatique, le réflectron, peut être utilisé.

3.1.2.2.1. L'analyseur à temps de vol en mode réflecteur

3.1.2.2.1.1. Le principe

Le réflecteur électrostatique ou réflectron est placé à l'extrémité du tube de vol (figure 9).

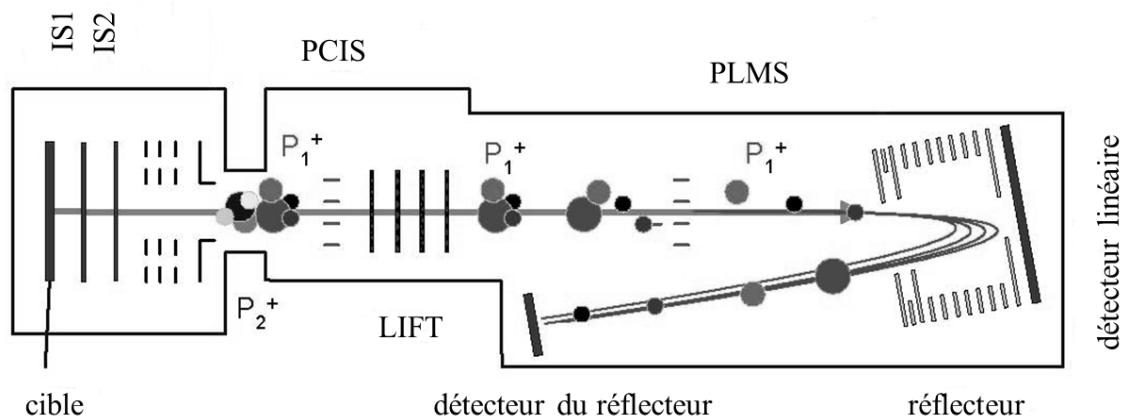


Figure 9 : Schéma d'un MALDI TOF/TOF LIFT reproduit et modifié de Bruker Daltonics®.

Il est constitué d'une succession d'électrodes annulaires alignées selon un même axe. Des potentiels croissants appliqués aux électrodes engendrent un champ électrique répulsif. Ce champ électrique freine et réfléchit les ions à la même vitesse initiale mais en direction opposée vers le détecteur.

Il existe deux sortes de réflecteur. Le réflecteur à un étage est constitué d'un seul champ électrique entre les deux grilles parallèles d'entrée et de sortie du réflecteur. Ce réflecteur est le plus simple mais il peut y avoir un problème d'homogénéité dans le champ du réflecteur. Le réflecteur à deux étages est constitué de deux régions de champ homogène séparées. La première possède un champ fort et la deuxième un champ faible. Ceci permet d'avoir une plus forte décélération des ions au niveau du premier étage et donc d'avoir un tube de vol plus compact.

Le réflecteur conduit à un pouvoir résolutif important. Les ions de même rapport m/z et ceux possédant une énergie cinétique plus importante pénètrent plus profondément dans le réflecteur et donc y restent plus longtemps. Ils rattrapent ensuite les ions de plus faible énergie cinétique durant le trajet retour. Grâce au réflecteur, la résolution isotopique est nettement

améliorée, mais au détriment de la gamme de masse limitée à des rapports m/z inférieurs ou égaux à 10 000 Da (Mamyrin 1994; Mamyrin et al. 1973).

3.1.2.2.1.2. L'analyse MS/MS

Le second avantage du réflecteur est qu'il donne accès à l'analyse en mode MS/MS. Lorsque la durée de vie de l'ion est supérieure à son temps de séjour dans la source mais inférieure à son temps de vol, il se produit des fragmentations métastables dans la zone libre de champ du tube de vol. L'ion parent et ces ions fragments possèdent la même énergie cinétique, ils sont donc analysés sous forme d'un seul pic au niveau du détecteur en mode linéaire mais sont séparés au niveau du réflecteur et détectés à des temps différents en mode réflecteur.

L'ion parent est tout d'abord sélectionné, lui et ses fragments sont ensuite ré-accélérés par une tension d'accélération de quelques kV afin de diminuer la gamme d'énergie cinétique de ces ions. Le spectre métastable de l'ion parent et de ses ions fragments déjà formés avant sélection est ainsi obtenu en une seule étape.

La société Bruker a mis au point une cellule *LIFT* (ascenseur à potentiel) (figure 9). Après fragmentation, l'ion parent et ses ions fragments, possédant la même énergie cinétique, sont sélectionnés en mode PCIS (*Pre CuRSOR Ion Selector*) et sont ré-accélérés à l'aide d'un potentiel d'accélération dans la cellule *LIFT* (Suckau et al. 2003). L'ion parent est supprimé dans le PLMS (*Post Lift Metastable Suppressor*) afin d'éviter une saturation du détecteur et des fragmentations métastables sont produites dans le deuxième tube de vol. Les ions ayant une vitesse dépendante de leur masse, sont alors refocalisés. Pour réaliser des fragmentations en mode CID de haute énergie, un gaz de collision peut être introduit dans la source.

3.1.2.3. Le détecteur et le traitement des données

Après avoir traversé l'analyseur, les ions arrivent au détecteur qui transforme le signal ionique en un signal électrique (figure 10).

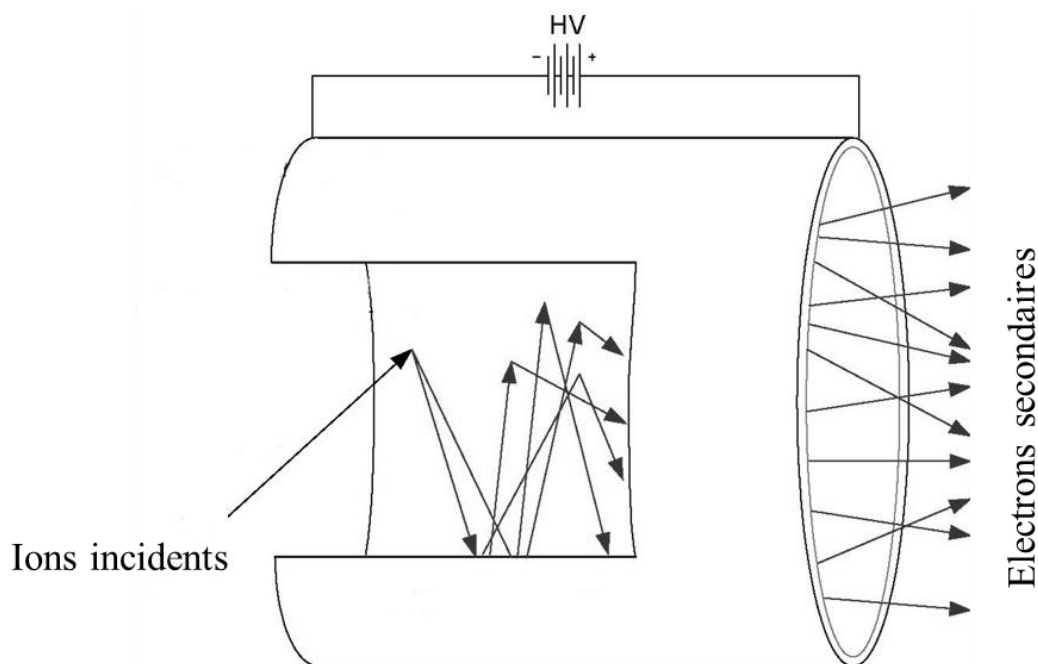


Figure 10 : Schéma d'un microcanal reproduit de Bruker Daltonics®.

Le détecteur est composé d'une galette de microcanaux pour l'analyse d'une large gamme de masse (jusqu'à 300 000 Da) en mode linéaire et de deux galettes de microcanaux montés en chevron en mode réflecteur. La galette est constituée d'une plaque poreuse percée de millions de petits canaux et recouverte d'une surface semi-conductrice. Chaque canal est un amplificateur d'électrons indépendant et connecté en parallèle aux autres. Ce type de détecteur sature rapidement notamment, pour l'analyse des polymères, car les électrons ne sont pas encore tous transmis quand les ions suivants arrivent au niveau du détecteur.

La mesure du temps de vol des ions s'effectue à l'aide d'un oscilloscope numérique et les données sont transmises à un ordinateur pour le stockage et le traitement des spectres.

3.2. Les méthodes d'analyse de l'ADN

Les oligonucléotides peuvent donc être analysés notamment par spectrométrie de masse ESI-IT ou MALDI-TOF. Cependant, les méthodes d'analyses de séquençage de l'ADN, *de seconde génération*, par spectrométrie de masse, sont basées que sur l'utilisation du MALDI-TOF MS. Au début du projet du génome humain, la spectrométrie de masse a été envisagée pour réaliser plus rapidement son séquençage. D'une capacité haut débit importante, la

spectrométrie de masse MALDI-TOF permet l'analyse rapide d'un grand nombre d'échantillons. Elle est facile à mettre en œuvre et les mélanges complexes d'ADN sont analysés directement sans séparation chromatographique préalable.

L'analyse de l'ADN par spectrométrie de masse MALDI-TOF repose sur deux approches différentes. La première est la détection directe de l'ADN et de ses modifications. Le principal désavantage de cette méthode est une faible sensibilité due à la mauvaise désorption des oligonucléotides. Le séquençage de novo est généralement limité à moins de 100 pb. La deuxième approche est une détection indirecte, plus facile des oligonucléotides, par marquage moléculaire. Cette méthode permet un grand degré d'analyse en multiplexage.

Les différentes méthodes de d'analyse de l'ADN par spectrométrie de masse sont : le génotypage des SNPs, l'haplotypage, l'analyse quantitative, l'analyse de l'expression des gènes, l'analyse des insertions et délétions, l'analyse de la méthylation de l'ADN, le séquençage de novo et le re-séquençage (Ragoussis et al. 2006).

3.2.1. Les méthodes directes

Les méthodes directes reposent toutes sur le même principe d'un clivage spécifique de la séquence afin d'obtenir des fragments plus petits donc plus facilement analysables (Edwards et al. 2005). La fragmentation peut être obtenue par extension d'amorce avec terminaison didésoxy ou bien par fragmentation d'ARN ou d'ADN. La mesure précise des masses des fragments présents dans le mélange permet de réaliser le re-séquençage de l'ADN.

Les avantages par rapport à l'électrophorèse capillaire sont le séquençage direct de l'ADN et l'ARN (y compris pour l'identification des modifications post-traductionnelles de l'ARN ribosomal), la robustesse de la détection des mutations hétérozygotes et l'analyse de la méthylation d'ADN (Ragoussis et al. 2006).

La méthode directe la plus répandue est celle de *MassARRAY*[®] de Sequenom. L'analyse de l'ADN est réalisée à l'aide du clivage enzymatique des bases contenues dans l'ARN transcript (figure 11) (Ehrich, Bocker, et al. 2005; Krebs et al. 2003).

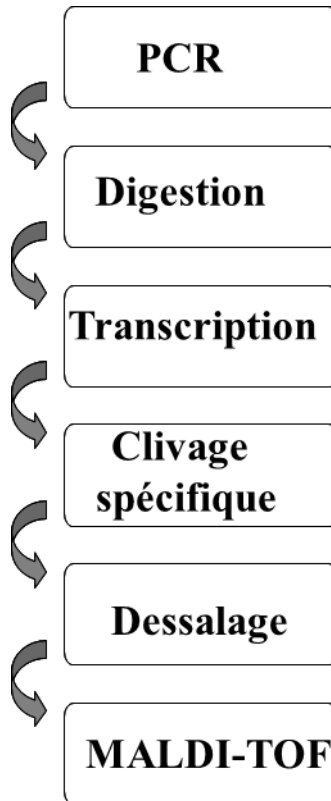


Figure 11 : Schéma de principe de la méthode MassARRAY[®] de Sequenom.

La partie du génome à analyser est amplifiée par PCR à l'aide d'amorces spécifiques du promoteur *T7* indispensable à la transcription. Le produit de PCR est ensuite purifié par la *Shrimp Alkaline Phosphatase (SAP)* qui digère les dNTPs n'ayant pas réagi. Il est alors transcrit en ARN simple-brin dans le sens et l'anti-sens. Puis, l'ARN est alors clivé par une enzyme spécifique des bases de l'ARN. La combinaison du clivage spécifique des bases cytosines et uraciles dans le sens et l'anti-sens permet d'obtenir une information virtuelle sur les quatre bases. Les fragments de clivage sont alors dessalés à l'aide d'une résine échangeuse de cations, et ensuite analysés par MALDI-TOF. Les spectres de masse sont ensuite analysés par un logiciel afin de déterminer la séquence d'ADN par comparaison avec la séquence de référence.

Une région allant jusqu'à 1 Kb peut ainsi être étudiée mais le taux de détection dépend de la nature et de la longueur de la séquence d'ADN. Pour l'étude d'une séquence de 500 bp, le taux de détection d'une seule base d'ADN peut atteindre 99 % pour une séquence haploïde et 95 % pour une séquence diploïde mais dans 85 % des cas elle est déterminée sans aucune ambiguïté.

Cette méthode possède plusieurs applications : le génotypage (Buetow et al. 2001), l'analyse de la méthylation de l'ADN (Ehrich, Nelson, et al. 2005), le séquençage *de novo* ou le re-séquençage (Stanssens et al. 2004; von Wintzingerode et al. 2002; Böcker 2004).

3.2.2. Les méthodes indirectes

Les méthodes indirectes reposent sur le même principe consistant à marquer l'analyte afin d'améliorer sa détection. Le marquage est largement utilisé en spectrométrie de masse et notamment pour la quantification. Le marqueur peut être moléculaire, atomique ou isotopique. Les méthodes de quantification protéomique, les plus courantes sont la méthode *SILAC* (*Stable Isotope Labeling by amino acids in cell culture*) (Ong et al. 2002) incorporant des isotopes, ou celles induisant une modification chimique comme la méthode *ICAT* (*Isotope Coded Affinity Tag*) (Gygi et al. 1999; Tao & Aebersold 2003), la méthode *iTRAQTM* (*Isotope Tag for Relative and Absolute Quantification*) et la méthode *ICPL* (*Isotope Coded Protein Labelling*) (Ross et al. 2004; Schmidt et al. 2005).

Les méthodes de marquage, notamment moléculaire, sont également utilisées pour l'analyse génomique. La méthode *GenoSNIP* utilise une molécule photoclivable pour le multiplexage des génotypes (Wenzel et al. 2003).

La méthode du *GOODASSAY* a été mise au point par Gut dans le laboratoire (Sauer et al. 2000). Elle repose sur le marquage moléculaire de l'oligonucléotide par une molécule dite *Charge Tag* et une étape d'alkylation dans le but d'améliorer sa détection par MALDI-TOF MS (figure 12) (Gut & Beck 1995; Gut et al. 1997).

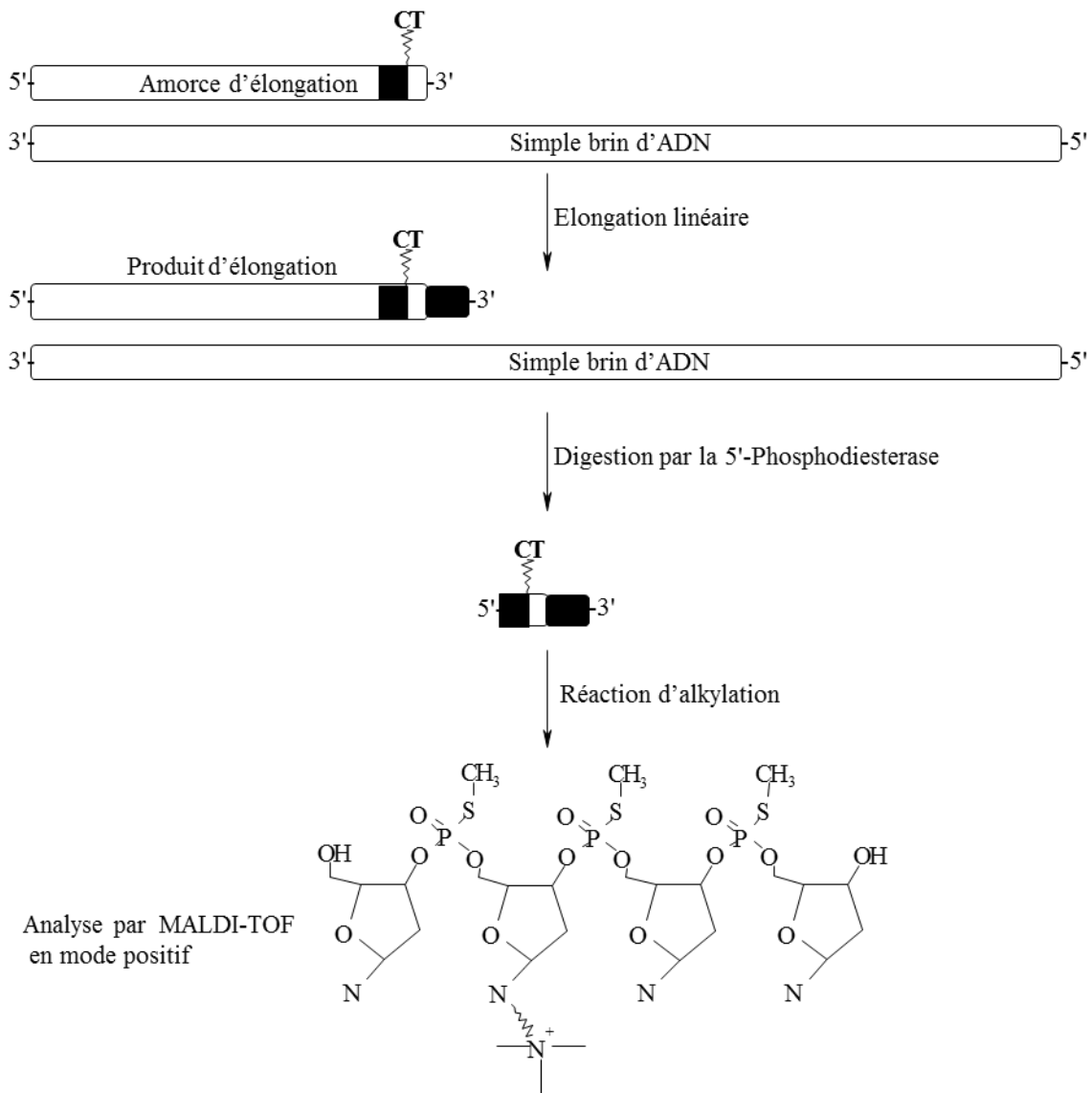


Figure 12 : Schéma de principe de la méthode du GOODASSAY.

La partie du génome à analyser est amplifiée par PCR. Le produit de PCR est ensuite purifié par la *SAP* afin d'enlever les dNTPs n'ayant pas réagi. La troisième étape du protocole est l'utilisation de l'amorce d'élongation linéaire modifiée. L'amorce d'élongation linéaire possède une base modifiée contenant une amine et des ponts phosphorothioates au niveau de cette base. L'amine réagit avec des groupements N-hydroxysuccinimidyl en formant une liaison covalente entre le *Charge Tag* et l'amorce. Le *Charge Tag* est chargé positivement à l'aide d'un groupement de triméthylammonium hexyryl-N-hydroxysuccinimidyl ou de ses dérivés. Cette amorce est adjacente à la position de la base à analyser. L'amorce est étendue juste d'une base avec l'utilisation de α -S-ddNTP. Les ponts phosphates contenus dans le produit d'élongation sont digérés par la 5-Phosphodiesterase. Les ponts phosphorothioates sont ensuite alkylés afin de neutraliser les charges négatives. Il ne persiste alors que la seule

charge positive du *Charge Tag*. La neutralisation des charges des groupements phosphates permet une meilleure désorption de l'oligonucléotide à analyser. L'introduction d'un groupement ammonium quaternaire permet un meilleur rendement de production des ions que les protons. La combinaison de ces deux étapes permet donc un meilleur signal (sensibilité) des oligonucléotides et donc une étude au MALDI-TOF en mode positif.

La méthode du *GOODASSAY* a été développée pour le génotypage des SNPs (Sauer et al. 2000; Sauer et al. 2003). Puis, elle a été utilisée pour le microhaplotypage du système d'histocompatibilité du gène *HLA* (Kucharzak 2006) et l'analyse de la méthylation de l'ADN (Tost et al. 2003). Cette méthode est bien adaptée pour l'étude d'une position ou d'un microhaplotype mais elle ne permet pas l'étude d'une séquence d'ADN.

4. Développement de méthodes d'analyse de l'ADN par clivage d'une chimère ARN/ADN et par spectrométrie de masse MALDI-TOF

Le but de ce travail est de développer des méthodes d'analyse de l'ADN reposant sur le clivage d'une chimère ARN/ADN et l'analyse par spectrométrie de masse MALDI-TOF.

4.1. Le principe

Le principe de la méthode est d'analyser les masses des fragments de clivage de la chimère ARN/ADN par spectrométrie de masse MALDI-TOF afin d'analyser la séquence d'ADN (figure 13).

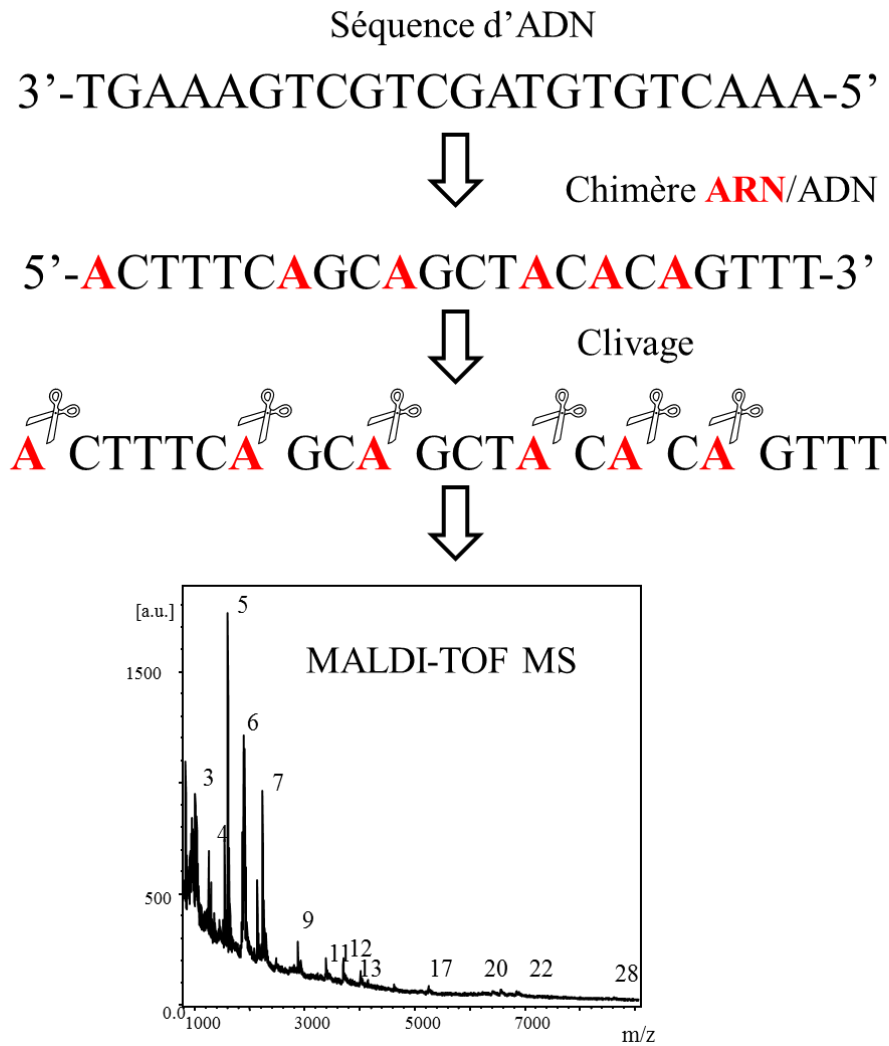


Figure 13 : Principe de la méthode : la synthèse de la chimère ARN/ADN, le clivage de la chimère et l'analyse des fragments de clivage par MALDI-TOF MS.

La chimère ARN/ADN, constituée de ribonucléotides et de désoxynucléotides est synthétisée et ensuite clivée chimiquement après chaque ribonucléotide. Chaque fragment de clivage se termine par un ribonucléotide et possède un groupement 3'-phosphate terminal. Les fragments générés, sont alors dessalés et ensuite analysés par spectrométrie de masse MALDI-TOF en mode négatif. Les masses des fragments de clivage sont comparées à celles de la séquence de référence afin d'en déduire les différences. Ces différences de masse sont significatives d'une ou plusieurs modifications dans la séquence d'ADN. C'est une méthode de cartographie massique.

4.1.1. La chimère ARN/ADN

Cette méthode tire avantage d'une nouvelle génération d'ADN polymérase (figure 14).

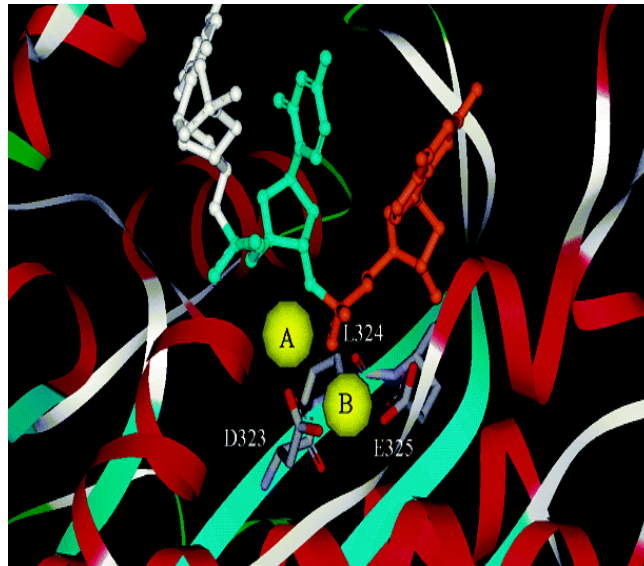


Figure 14 : Représentation du site des ADN polymérase permettant l'incorporation des NTPs.

Généralement, les ADN polymérase amplifient l'ADN en n'incorporant que des dNTPs tandis que les ARN polymérase transcrivent le brin d'ADN en ARN en n'incorporant que des NTPs. Dans cette méthode, une nouvelle classe d'ADN polymérase incorpore des dNTPs et des NTPs et amplifie le ou les deux brins d'ADN permettant ainsi la création d'une chimère ARN/ADN.

Plusieurs générations d'ADN polymérase, élaborées par Roche Molecular Systems, existent : l'ADN polymérase *G46E CS6R* (Mauger et al. 2006), l'ADN polymérase *KB17* (Mauger et al. 2007) et l'ADN polymérase *FP-1*.

4.1.2. Le clivage de la chimère ARN/ADN

La chimère ARN/ADN est ensuite clivée, en milieu basique, au niveau de la liaison P-O du phosphodiester de la chaîne nucléotidique situé après chaque ribonucléotide (figure 15).

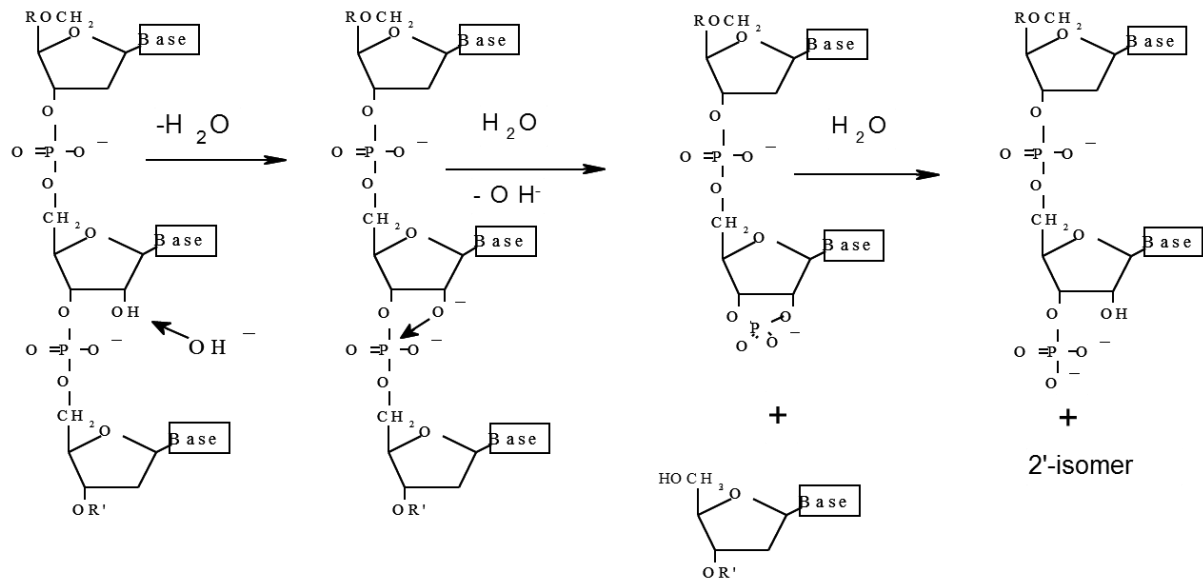


Figure 15 : Réaction de clivage de la chaîne nucléotidique située après un ribonucléotide, en milieu basique au niveau de la liaison P-O du phosphodiester: substitution nucléophile au niveau de l'atome de phosphore suivant le ribonucléotide (Shabarova & Bogdanov 1994).

Le groupement 2'-OH du ribonucléotide peut être attaqué en milieu basique par l'ion hydroxyde. Il se forme alors un alcoolate O^- en position 2' attaquant ensuite le groupement phosphodiester en position 3' du ribose. Il en résulte une coupure de la chaîne nucléotidique entre la liaison P-O du phosphodiester et le nucléotide suivant. L'information sur la masse du fragment ne permet pas de distinguer ces deux espèces complémentaires: ($2'\text{-H}_2\text{PO}_4$, $3'\text{-OH}$) et ($2'\text{-OH}$, $3'\text{-H}_2\text{PO}_4$). Il est probable que les fragments de clivage soient un mélange de ces deux espèces.

Il en résulte que tous les fragments de clivage de la chimère ARN/ADN sont constitués de désoxynucléotides et d'un seul ribonucléotide terminal qui possède un groupement 3'-phosphate.

4.1.3. L'analyse par MALDI-TOF MS

Les fragments de clivage de la chimère ARN/ADN doivent être dessalés afin d'éliminer les cations présents en solution afin de pouvoir les analyser par MALDI-TOF MS. L'analyse des fragments de clivage s'effectue en mode négatif grâce à la charge négative portée par l'atome d'oxygène des groupements phosphates.

Les masses des fragments de clivage sont ensuite comparées à celles des fragments de clivage de la séquence de référence. Les différences de masse sont significatives de la présence de

modification dans la séquence d'ADN. Les modifications peuvent être des polymorphismes, des délétions, des insertions, des mutations ou la méthylation de l'ADN.

Plusieurs chimères ARN/ADN peuvent être utilisées, en parallèle, pour réaliser cette cartographie massive afin d'apporter des informations complémentaires.

4.2. Les méthodes

Le but de ce travail est de développer un arsenal de méthodes d'analyse des modifications de l'ADN basé sur le clivage d'une chimère ARN/ADN et par MALDI-TOF MS.

4.2.1. Les méthodes actuelles

Deux méthodes ont été préalablement publiées à ce travail (annexe 1 et 2). La première méthode est le génotypage des SNPs (Mauger et al. 2006) et la deuxième est le séquençage de l'ADN mitochondrial humain (Mauger et al. 2007).

4.2.1.1. Le génotypage de SNPs

Le but de la méthode de génotypage des SNPs (annexe 1) est de développer une méthode d'analyse d'une position d'une seule base connue du génome (SNP) afin de déterminer les individus homozygotes A ou B et hétérozygote AB (Mauger et al. 2006).

La séquence d'ADN, contenant le SNP, est préalablement amplifiée par PCR. Le produit de PCR n'est pas purifié car l'ADN polymérase de l'élongation linéaire n'incorpore que les NTPs. Puis, la chimère ARN/ADN est synthétisée par l'ADN polymérase *G46E CS6R*, des NTPs (A, C, G et U) et d'une amorce contenant des désoxynucléotides et un ribonucléotide. Le ribonucléotide de l'amorce sert à diminuer la masse du fragment de clivage contenant le génotype afin d'améliorer sa détection par MALDI-TOF MS. La chimère ARN/ADN est ensuite clivée par l'hydroxyde de sodium, l'hydroxyde de potassium ou l'hydroxyde d'ammonium. L'hydroxyde de sodium et l'hydroxyde de potassium permettent un clivage rapide et total tandis qu'il est moins rapide avec l'hydroxyde d'ammonium mais l'ammonium a l'avantage d'améliorer la détection des ions. Les fragments de clivage sont ensuite purifiés.

Les fragments de clivage sont ensuite analysés à l'aide de la matrice HCCA car sa cristallisation est homogène et elle est adaptée à l'étude des petits oligonucléotides. L'analyse s'effectue en mode négatif linéaire par MALDI-TOF MS.

Dans cette méthode, deux procédures de préparation de l'échantillon ont été utilisées : l'une est basée sur le dessalage par une résine échangeuse de cations, comme la méthode de *MassARRAY*[®], et l'autre est basée sur une réaction d'alkylation sur les ponts phosphorothioates de l'amorce d'élongation, comme la méthode du *GOODASSAY*, (figure 12). Le dessalage par les billes échangeuses de cations est rapide et facile à mettre en œuvre. De plus, les billes sont simplement ajoutées à la solution, il n'y a donc pas de perte de fragments de clivage. Au contraire, la méthode d'alkylation nécessite l'extraction de phase liquide ou des fragments de clivage peuvent être perdus. De plus, la réaction d'alkylation est toxique contrairement à la résine. Après l'alkylation, le pH de la solution doit être diminué avec du chlorure d'hydrogène alors qu'avec l'utilisation des billes échangeuses de cations la diminution du pH se fait en même temps par l'intermédiaire des protons présents à la surface des billes.

Les positions hétérozygotes et homozygotes des polymorphismes étudiés ont pu être déterminées. Pour l'étude d'un SNP Y (T/C), l'uracile et la cytosine possèdent la même masse, elles ne permettent donc pas d'analyser un SNP Y. Dans ce cas, elles peuvent être remplacées par la 5-méthylcytosine ou la 5-méthyluracile afin de pouvoir analyser le SNP.

4.2.1.2. Le séquençage de l'ADN mitochondrial

Le but de cette méthode (annexe 2) est de développer une méthode de séquençage de la région hypervariable *HVI* (environ 450 pb) de l'ADN mitochondrial de 22 individus (Mauger et al. 2007).

Le génome de l'ADN mitochondrial est beaucoup moins complexe que celui du génome humain et beaucoup plus concentré (environ 20 fois). De plus, il n'y a pas de position hétérozygote car l'ADN mitochondrial est d'origine exclusivement maternelle.

La région hypervariable *HVI* est préalablement amplifiée et purifiée afin de synthétiser la chimère ARN/ADN. Plusieurs élongations linéaires sont utilisées en parallèle avec ATP, CTP et GTP. La chimère ARN/ADN est synthétisée à l'aide d'une amorce d'élongation

linéaire anti-sens, de l'ADN polymérase *KB17* et d'un mélange d'une base ribonucléotide et des trois autres bases complémentaires désoxynucléotides. Les chimères d'ARN/ADN sont ensuite clivées par l'hydroxyde de sodium et dessalées par une résine échangeuse de cations. Les fragments de clivage sont analysés à l'aide de la matrice THAP car elle permet la détection, en mode linéaire, avec une résolution suffisante des oligonucléotides de masse allant de 900 à 10 000 Da. L'analyse par MALDI-TOF MS en mode négatif s'effectue, en mode linéaire et en mode réflecteur.

Les différences de masse entre les fragments de clivage de l'individu et celle de la séquence de référence sont significatives de la présence de positions polymorphiques, de délétions ou d'insertions.

La complémentarité des informations des trois chimères ARN/ADN et l'utilisation des deux modes de détection par MALDI-TOF MS a permis le séquençage de ses individus sans ambiguïté.

4.2.2. Le développement de nouvelles méthodes

Le but de ce travail est de développer d'autres méthodes d'analyse des modifications de l'ADN, basées sur ce concept, tout en diminuant le nombre d'étapes du protocole afin de le rendre plus compétitif.

Tout d'abord, dans le chapitre 1, la chimère ARN/ADN est analysée par spectrométrie de masse en tandem. Les fragmentations (métastables ou en mode CID) de la chimère GCTA par spectrométrie de masse en tandem (MALDI/TOF/TOF ou ESI-IT) et les fragmentations en mode CID par spectrométrie de masse MALDI-TOF/TOF de plusieurs chimères ARN/ADN de 4 à 10-mer sont étudiées.

La méthode de génotypage des SNPs (Mauger et al. 2006) doit être adaptée à l'étude des régions hypervariables afin de pouvoir déterminer les haplotypes. Une méthode de microhaplotypage pour l'étude du système d'histocompatibilité du gène *HLA*, région hypervariable, est développée dans le chapitre 2.

L'analyse de la méthylation de l'ADN est actuellement en pleine essor. Une méthode de ciblage des régions hyperméthylées et hypométhylées des îlots CpG du gène basée sur celle

du séquençage de l'ADN mitochondrial (Mauger et al. 2007) mais aussi sur la quantification relative de la méthylation de l'ADN, est développée dans le chapitre 3.

Une autre amélioration de ce concept est la synthèse d'une chimère ARN/ADN double-brin qui est beaucoup plus complexe à réaliser qu'une chimère ARN/ADN simple-brin. L'utilisation de cette chimère permet de réduire le nombre d'étapes du protocole mais aussi de développer une méthode de re-séquençage dans le chapitre 4 et une méthode de multiplex allèle-spécifique génotypage dans le chapitre 5.

Enfin, un bilan de ce travail est exposé dans une conclusion générale ainsi que ses perspectives.

Chapitre 1

Analyse de la chimère ARN/ADN par spectrométrie de masse en tandem

Chapitre 1. Analyse de la chimère ARN/ADN par spectrométrie de masse en tandem

1. Introduction

En 2006 et 2007, une méthode de génotypage des SNPs et une méthode de re-séquençage de l'ADN mitochondrial par clivage d'une chimère ARN/ADN simple-brin et par spectrométrie de masse MALDI-TOF (Mauger et al. 2006; Mauger et al. 2007), ont été mises au point. Dans la méthode de re-séquençage, une chimère simple-brin d'ARN/ADN est composée de ribonucléotides (A, C ou G) et de désoxynucléotides. La base du ribonucléotide (A, C ou G) est différente de celles des désoxyucléotides (CGT, AGT ou ACT). La chimère ARN/ADN est ensuite clivée par hydrolyse après chaque ribonucléotide. Les fragments de clivage sont constitués des désoxyucléotides et se terminent en 3' par un seul ribonucléotide et un groupement 3'-phosphate. Les échantillons sont ensuite dessalés et analysés par spectrométrie de masse MALDI-TOF à l'aide de la matrice THAP en mode négatif.

Dans ce chapitre, l'analyse de la chimère ARN/ADN par spectrométrie de masse en tandem est développée. L'étude est composée de trois étapes : (i) la formation de la chimère ARN/ADN simple-brin par clivage chimique ; (ii) le dessalage et (iii) l'analyse en mode CID par spectrométrie de masse en tandem. Le but de cette analyse est d'étudier les fragmentations en phase gazeuse de l'ion parent $[M-H]^-$ de la chimère ARN/ADN dans les mêmes conditions d'analyse que celles des méthodes de re-séquençage de l'ADN mitochondrial (Mauger et al. 2007), de l'analyse de la méthylation de l'ADN (chapitre 3), de re-séquençage par *ribo-PCR* (chapitre 4) et de multiplex allèle-spécifique génotypage par *riboPAP-PCR* (chapitre 5).

La spectrométrie de masse en tandem est utilisée, entre autre, pour caractériser la séquence des oligonucléotides. L'analyse en phase gazeuse des oligonucléotides peut être réalisée par une source ESI ou MALDI couplée à un analyseur de type FT-ICR (*Fourier Transform Ion Cyclotron Resonance*) ou par ESI couplée à un triple quadripôle (Stemmler et al. 1995; Boschenok & Sheil 1996) mais le plus souvent, ils sont analysés en mode CID par ESI-ITMSⁿ (Little et al. 1996; McLuckey et al. 1992; McLuckey & Habibigoudarzi 1993; Wu et McLuckey 2004). Lors du mode CID, un gaz inerte est introduit dans la cellule de collision afin d'induire des fragmentations de l'ion parent par activation collisionnelle avec le gaz cible et d'obtenir ainsi des ions fragments. Dans cette étude, un spectromètre de masse composé

d'une source MALDI est couplé à un analyseur TOF possédant une cellule *LIFT* (Suckau et al. 2003; Vestal & Campbell 2005) est utilisé pour sélectionner un ion et produire les fragmentations des oligonucléotides monochargés positivement ou négativement.

Deux nomenclatures ont été proposées pour caractériser les ions obtenus lors de l'analyse par spectrométrie de masse en tandem des oligonucléotides. Une nomenclature proposée par McLuckey (McLuckey et al. 1992), pour les ions négatifs ($[M-H]^-$), caractérise le clivage de la chaîne nucléotidique au niveau des liaisons du phosphodiester par l'intermédiaire de 8 séries d'ions : a_n^- , b_n^- , c_n^- , d_n^- constitués de l'extrémité 5' de l'ion précurseur et w_n^- , x_n^- , y_n^- et z_n^- constitués de l'extrémité 3' de l'ion précurseur. Une autre nomenclature a été également proposée par Viari et complétée par Nordhoff pour les ions fragments MALDI (Nordhoff et al. 1995; Viari et al. 1988). Ne caractérisant pas l'ensemble des ions possibles mais seulement trois types d'ions X^- , X^{*-} et Y^- qui correspondent respectivement à ceux des séries de type d_n^- , $[a_n-B_nH]^-$ et w_n^- , elle ne sera donc pas employée.

Les fragmentations des oligonucléotides déprotonés en phase gazeuse sont similaires bien que les techniques d'ionisation/désorption soient différentes. En effet, elles dépendent de l'état de charge de l'oligonucléotide et de son énergie interne. L'orientation de leur dissociation est fonction du mode d'activation et du déplacement cinétique dépendant de l'instrument. Il existe des ions internes qui résultent d'un double clivage de la chaîne nucléotidique. Toutefois, les fragmentations les plus fréquentes sont la perte de la base nucléique et le clivage de la chaîne nucléotidique pour donner les séries d'ions a_n^- , $[a_n-B_nH]^-$, d_n^- et w_n^- . Les ions $[M-H]^-$ préparés par MALDI conduisent sous CID aux ions de type d_n^- et w_n^- abondants tandis que préparés par ESI conduisent sous CID aux ions $[a_n-B_nH]^-$ et w_n^- abondants.

Des études proposent deux mécanismes de fragmentation pour la formation de ces 4 séries d'ions complémentaires (Grotjahn L 1986; Nordhoff et al. 1995; Barry et al. 1995; McLuckey & Habibigoudarzi 1993). Le premier mécanisme nécessite une charge voisine au site de fragmentation et la perte de la base nucléique comme intermédiaire tandis que le deuxième ne nécessite ni la perte préalable de la base nucléique ni la charge à proximité du site de fragmentation. La première voie admise comme étant composée de deux étapes d'élimination 1,2. La première étape est la perte de la base nucléique sous forme neutre (ou anionique) par une élimination 1,2 qui forme les ions fragments intermédiaires $[M-H-B_nH]^-$. La deuxième étape est la rupture de la liaison 3'-CO du phosphodiester adjacent induite par la charge. Cette réaction est favorisée par la formation d'un furane stabilisé par résonance. Si les ions $[a_n-$

$B_nH]^-$ possèdent un excès d'énergie interne, ils peuvent alors se fragmenter pour donner les ions d_n^- (Hettich & Stemmler 1996). La deuxième voie ne nécessite ni la perte préalable de la base nucléique ni la présence d'une charge adjacente au site de clivage. Les ions a_n^- et w_n^- sont formés par la rupture directe de la liaison entre le 3'C-O du groupement phosphodiester. Les ions $[a_n-B_nH]^-$ sont formés ensuite à partir de la perte de la base nucléique des ions a_n^- mais aussi à partir de $[M-H-BH]^-$. Deux mécanismes ont été proposés cette voie de fragmentation. Il s'agit d'une élimination 1,2 où le proton en 4' est transféré sur un groupement phosphodiester chargé négativement éloigné ou sur l'oxygène du groupement phosphate en 5' non chargé à l'aide d'un état de transition cyclique à 6 atomes.

Cette étude est réalisée sur l'analyse des chimères ARN/ADN simple-brins issues du clivage d'oligonucléotides synthétiques qui contiennent deux ribonucléotides et des désoxynucléotides. La base nucléique du ribonucléotide (A, C, G ou U) est différente de celle des désoxynucléotides (CGT, AGT, ACT ou ACG). Les fragmentations en mode CID par MALDI-TOF/TOF des chimères ARN/ADN de 4, 5, 7 et 10 bases sont analysées afin de déterminer l'influence des bases nucléiques et du nombre de nucléotides sur l'orientation des fragmentations. Les dissociations de quelques chimères de 4-mer sont également comparées à celles obtenues sur des oligonucléotides de même séquence mais qui contiennent un groupement 3'-OH ou un groupement 3'-PO₄H₂ et/ou se terminent par un ribonucléotide ou par un désoxynucléotide. Enfin, la chimère GCTA est étudiée par spectrométrie de masse MALDI-TOF/TOF et ESI-IT pour comparer les dissociations obtenues.

2. Matériels et Méthodes

2.1. Matériels

Les oligonucléotides ont été synthétisés par Eurogentec (Angers, France) (tableau 1.1).

| Nb bases | Oligonucléotide | Chimère | Oligonucléotide | Chimère | Oligonucléotide | Chimère | Oligonucléotide | Chimère |
|----------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|---------------|
| 4 | TAGTTAT | GTT <u>A</u> | TCATTCT | ATTC | TGCTTGT | CTT <u>G</u> | AUACCUA | ACCU |
| | TATGTAT | TGT <u>A</u> | TCTATCT | TATC | TGTCGT | TCT <u>G</u> | AUCACUA | CACU |
| | TATTGAT | TTG <u>A</u> | TCCTACT | TTAC | TGTTCTG | TTC <u>G</u> | AUCCUA | CCAU |
| | TACTTAT | CTT <u>A</u> | TCGTTCT | GTTC | TGATTGT | ATT <u>G</u> | AUGCCUA | GCCU |
| | TATCTAT | TCT <u>A</u> | TCTGTCT | TGTC | TGTATGT | TAT <u>G</u> | AUCGCUA | CGCU |
| | TATTCAT | TTC <u>A</u> | TCTTGCT | TTGC | TGTTAGT | TTA <u>G</u> | AUCCGUA | CCGU |
| | TAGCTAT | GCT <u>A</u> | TCGATCT | GATC | TGCATGT | CAT <u>G</u> | AUGACUA | GACU |
| | TAGTCAT | GTC <u>A</u> | TCGACT | GTAC | TGCTAGT | CTA <u>G</u> | AUGCAUA | GCAU |
| | TATCGAT | TCG <u>A</u> | TCTAGCT | TAGC | TGACTGT | ACT <u>G</u> | AUCAGUA | CAGU |
| | TATGCAT | TGC <u>A</u> | TCTGACT | TGAC | TGATCGT | ATC <u>G</u> | AUCGAUA | CGAU |
| | TACGTAT | CGT <u>A</u> | TCATGCT | ATGC | TGTACGT | TAC <u>G</u> | AUACGUA | ACGU |
| | TACTGAT | CTG <u>A</u> | TCAGTCT | AGTC | TGTCAGT | TCA <u>G</u> | AUAGCUA | AGCU |
| | TAGCGAT | GCG <u>A</u> | TCAGGCT | AGGC | TGACCGT | ACC <u>G</u> | AUAGGUA | AGGU |
| | TACGGAT | CGG <u>A</u> | TCGAGCT | GAGC | TGCACGT | CAC <u>G</u> | AUGAGUA | GAGU |
| | TAGGCAT | GGC <u>A</u> | TCGGACT | GGAC | TGCCAGT | CCA <u>G</u> | AUGGAUA | GGAU |
| | TATTTAT | TTT <u>A</u> | TCTTTCT | TTTC | TGTTTGT | TTT <u>G</u> | AUCCCUA | CCCU |
| | TACCCAT | CCC <u>A</u> | TCAAACCT | AAAC | TGAAAAGT | AAA <u>G</u> | AUAAAUA | AAAU |
| | TAGGGAT | GGG <u>A</u> | TCGGGCT | GGGC | TGCCCGT | CCC <u>G</u> | AUGGGUA | GGGU |
| | TACGCAT | CGC <u>A</u> | TCGAACT | GAA <u>C</u> | TGCAAGT | CAA <u>G</u> | AUCGGUA | CGGU |
| | TAGCCAT | GCC <u>A</u> | TCAGACT | AGAC | TGACAGT | ACA <u>G</u> | AUGCGUA | GCGU |
| | TACCGAT | CCG <u>A</u> | TCAAAGCT | AAGC | TGAACGT | AAC <u>G</u> | AUGGCUA | GGCU |
| | TACTCAT | CTC <u>A</u> | TCATACT | ATAC | TGAATGT | AAT <u>G</u> | AUCAUA | CAAU |
| | TATCCAT | TCC <u>A</u> | TCTAACT | TAA <u>C</u> | TGATAGT | ATA <u>G</u> | AUACAUA | ACAU |
| | TACCTAT | CCT <u>A</u> | TCAAATCT | AATC | TGTAAGT | TAAG | AUAACUA | AACU |
| | TAGTGAT | GTG <u>A</u> | TCGTGCT | GTGC | TGCCTGT | CCT <u>G</u> | AUGAAUA | GAUU |
| | TATGGAT | TGG <u>A</u> | TCTGGCT | TGGC | TGCTCGT | CTC <u>G</u> | AUAGUA | GAU |
| | TAGGTAT | GGT <u>A</u> | TCGGTCT | GGTC | TGTCCGT | TCC <u>G</u> | AUAAGUA | AAGU |
| | | GCTA | GCT <u>A</u> | GCTA | GCTA | GCTA | | |
| | | GTCA | GTC <u>A</u> | GTCA | GTCA | GTCA | | |
| | | TCGA | TCG <u>A</u> | TCGA | TCGA | TCGA | | |
| | | TGCA | TGC <u>A</u> | TGCA | TGCA | TGCA | | |
| | | CGTA | CGT <u>A</u> | CGTA | CGTA | CGTA | | |
| | | CTGA | CTG <u>A</u> | CTGA | CTGA | CTGA | | |
| | | GATC | GATC | GATC | GATC | GATC | | |
| | | GTAC | GTAC | GTAC | GTAC | GTAC | | |
| | | TAGC | TAGC | TAGC | TAGC | TAGC | | |
| | TGAC | TGAC | TGAC | TGAC | TGAC | | | |
| | ATGC | ATGC | ATGC | ATGC | ATGC | | | |
| | AGTC | AGTC | AGTC | AGTC | AGTC | | | |
| 5 | TATTTCGAT | TTCG <u>A</u> | TATCTGAT | TCTG <u>A</u> | TACGTTAT | CGTT <u>A</u> | TATGCTAT | TGCT <u>A</u> |
| | TATTGCAT | TTGC <u>A</u> | TATGTCAT | TGTC <u>A</u> | TAGCTTAT | GCTT <u>A</u> | TAGTTCAT | GTTC <u>A</u> |
| | TATCGTAT | TCGT <u>A</u> | TACTTGAT | CTTG <u>A</u> | TACTGTAT | CTGT <u>A</u> | TAGTCTAT | GTCT <u>A</u> |
| 7 | TGAAACTTGT | AAACTT <u>G</u> | TGAAATCTGT | AAATCT <u>G</u> | TGAAATTCGT | AAATTC <u>G</u> | | |
| 10 | TCAGTAGTTAGCT | AGTAGTTAGC | TCTAGTAAGGTCT | TAGTAAGGTCT | TCGTAGTGATACT | GTAGTGATAC | | |

Tableau 1.1: Séquence des oligonucléotides et des chimères ARN/ADN de 4, 5, 7 et 10-mer. Dans les séquences, les ribonucléotides sont en gras et les nucléotides qui sont soulignés contiennent un groupement 3'-PO₄H₂. La première base des oligonucléotides est constituée d'un groupement 5'-OH.

La résine échangeuse de cations provient de Biorad (Marnes la coquette, France), le kit *Clean Resin Tool* de *MassARRAY*[®] de chez Sequenom (Hamburg, Allemagne) et les produits chimiques d'Aldrich (Steinheim, Allemagne).

Le clivage a été réalisé sur des thermocycleurs gradients Eppendorf (Eppendorf, Allemagne). La centrifugeuse sous vide, *SpeedVac*[®] provient de Thermo Fisher (Les Ulis, France).

Les analyses en phase gazeuse ont été réalisées par un MALDI-TOF/TOF *Ultraflex*TM II et par un ESI-ITMSⁿ *Esquire*TM 3000 de chez Bruker (Bremen, Allemagne). La cible MALDI est une *Anchorchip*TM Bruker (Bremen, Allemagne).

2.2. Méthodes

2.2.1. Le clivage

De 0,2 nmol à 10 nmol d'oligonucléotide sont évaporés à l'aide d'une centrifugeuse sous vide. Puis, ils sont dissous dans 20 µL de la solution de clivage qui contient 2 µL d'hydroxyde de sodium à 3M et 18µL d'eau milli-Q. La réaction de clivage est mise sous incubation pendant 1h30 à 70°C.

2.2.2. Le dessalage

A la réaction de clivage, 6 mg de résines échangeuses de cations sont ajoutées à l'aide du kit *Clean Resin Tool de MassARRAY*[®]. Les échantillons sont incubés pendant 20 min à température ambiante sous agitation et ensuite centrifugés sous 134 xg afin de récupérer le surnageant.

2.2.3. L'analyse par spectrométrie de masse MALDI-TOF/TOF

2.2.3.1. La préparation de la cible MALDI

La solution de matrice trihydroxyacetophenone (THAP) est préparée avec 0,2 M de 2,4,6 THAP dans 50 % d'acétonitrile, 0,2 M de 2,3,4 THAP dans 50% d'acétonitrile et 0,3 M de citrate d'ammonium dans les proportions 6/3/2 (v/v) (Zhu et al. 1996).

Sur une cible *Anchorchip*TM Bruker de 384 spots de porosité 400 µm, sont déposés 0,5 µL de la solution de matrice THAP et 0,5 µL du surnageant ou de l'analyte à 10 pmol/µL par spot. La cible est séchée à température ambiante.

2.2.3.2. La spectrométrie de masse MALDI-TOF/TOF

La cible MALDI est introduite dans un spectromètre de masse MALDI/TOF-TOF.

L'analyse est réalisée à l'aide du logiciel *FlexControl* en mode négatif et en mode réflecteur avec les paramètres suivants : IS1=8 kV; IS2=7,2 kV; lens=3,6 kV; Reflectron=29,5 kV; Reflectron2=13,8 kV ; Lift1=19 kV ; Lift2=3,2 kV et en mode continu.

La fluence du laser est de 64% pour la sélection de l'ion parent et 100% pour les expériences de MS/MS. Les fragmentations en mode CID s'effectuent avec le gaz argon. Le spectre de masse de l'ion parent est obtenu par l'accumulation de 200 tirs de laser tandis que celui des ions fragments est obtenu par l'accumulation de 1000 tirs de laser. L'étalonnage de la méthode est réalisé par le constructeur en mode positif et il est transféré en mode négatif.

Ces conditions permettent d'analyser des échantillons d'une quantité minimum de 5 pmol par spot.

Les spectres de masse sont analysés à l'aide du logiciel *FlexAnalysis*. L'annotation des pics caractérisés par le rapport m/z des ions, la soustraction de la ligne de base ainsi qu'un lissage du signal sont réalisés.

2.2.4. L'analyse par spectrométrie de masse ESI-IT

La chimère GCTA est introduite à la concentration de 5 μ M dans 50% de méthanol dans la source electrospray sous le débit de 200 μ L/h.

L'analyse est réalisée à l'aide du logiciel *EsquireControl* en mode négatif, en mode standard (de 50 à 3000 Th) selon les paramètres suivants : source capillary=3500 V ; end plate offset=-500 V ; cap exit offset=-68,7 V ; skim₁=-22,1 V ; TD=80; dry temp=250 C ; nebulizer=7 psi et dry gas=5 L/min. Les fragmentations en mode CID s'effectuent avec le gaz hélium.

Les spectres de masse et CID sont obtenus par la moyenne de 5 balayages. L'étalonnage de l'instrument est externe.

Les spectres de masse sont analysés à l'aide du logiciel *DataAnalysis*. L'annotation des pics caractérisés par le rapport m/z des ions, la soustraction de la ligne de base ainsi qu'un lissage du signal sont réalisés.

2.2.5. Les ions fragments

Les masses monoisotopiques des ions ont été calculées à l'aide du logiciel *Mongo oligo Mass Calculator v2.06* développée par Jef Rozenski en 1999 à l'adresse internet suivante : <http://library.med.utah.edu/masspec/mongo.htm> ainsi qu'à partir des masses des atomes.

3. Résultats

3.1. La chimère ARN/ADN

Différentes chimères ARN/ADN, de 4 à 10-mer ont été étudiées.

3.1.1. La formation de la chimère ARN/ADN

Les oligonucléotides de 7 à 13-mer contiennent 2 ribonucléotides, constitués de la même base, et de désoxyribonucléotides contenant des bases différentes de celle du ribonucléotide (tableau 1.1). Ils sont clivés après chaque ribonucléotide par l'hydroxyde de sodium (Mauger et al. 2006; Mauger et al. 2007).

La chimère ARN/ADN, de n bases, est composée de $(n-1)$ désoxynucléotides de la position B_1 à B_{n-1} et se termine par un ribonucléotide B_n en 3'. La chimère possède un groupement 5'-OH et un groupement 3'-PO₄H₂. La base B_n peut être A, C, G ou U et les autres bases de la chimère doivent être différentes de la base du ribonucléotide B_n .

La chimère ARN/ADN est ensuite dessalée par une résine échangeuse de cations et analysée en mode négatif par MALDI-TOF/TOF avec la matrice THAP ou par ESI-ITMSⁿ.

3.1.2. La nomenclature des ions fragments

La nomenclature pour décrire les ions fragments des oligonucléotides déprotonés, en phase gazeuse, proposée par McLuckey en 1992 (McLuckey et al. 1992) a été utilisée car elle caractérise toutes les dissociations possibles telles qu'elles sont rencontrées (figure 1.1).

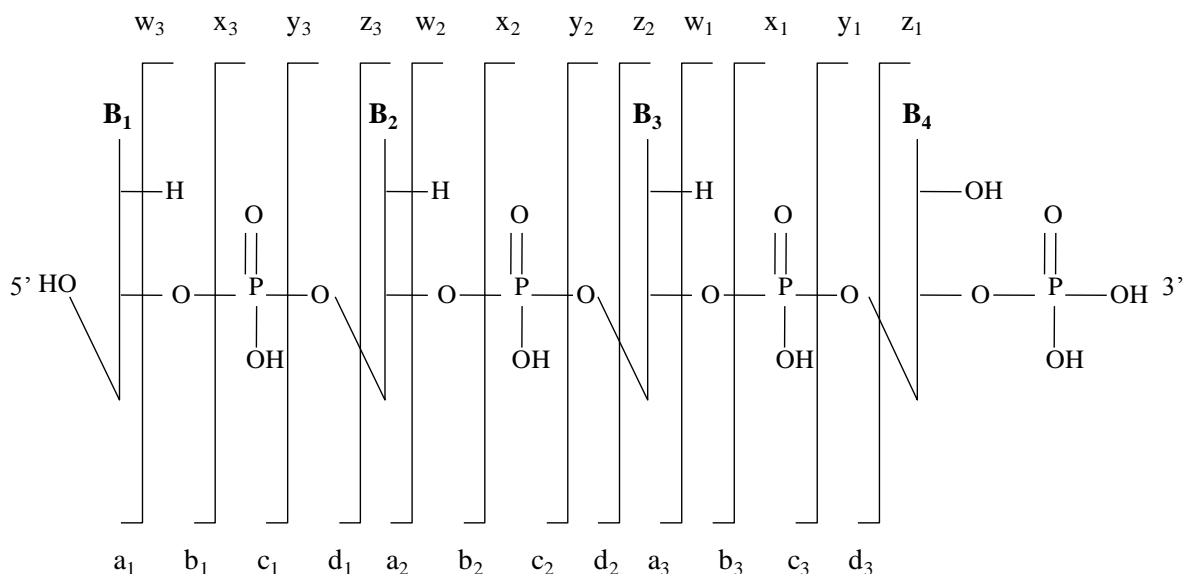


Figure 1.1: Nomenclature des fragments de dissociation de la chimère ARN/ADN de 4-mer proposée par McLuckey selon la représentation des oligonucléotides (McLuckey et al. 1992).

La figure 1.1 montre les différentes ruptures de la chaîne nucléotidique au niveau des liaisons du groupement phosphodiester qui donne lieu à huit types d'ions fragments. Les ions fragments constitués de l'extrémité 5'-OH de l'ion parent sont : a_n^- , b_n^- , c_n^- et d_n^- tandis que ceux constitués de l'extrémité 3'- PO_4H_2 de l'ion parent sont : w_n^- , x_n^- , y_n^- et z_n^- . De plus, la perte d'une base nucléique est également une des fragmentations les plus importantes qui engendre notamment les ions fragments suivants : les formes complémentaires B^- et $[M-H-B_nH]^-$ ainsi qu'un clivage consécutif qui conduit aux ions $[a_n-B_nH]^-$.

3.1.3. Le seuil de détection des chimères ARN/ADN par MALDI-TOF/TOF

Afin de pouvoir sélectionner l'ion parent $[M-H]^-$ de la chimère ARN/ADN et de détecter facilement ses ions fragments, la quantité minimum de chimère nécessaire a été déterminée expérimentalement. Le tableau 1.2 présente, le seuil de détection des chimères ARN/ADN de 4 à 10-mer pour l'analyse par MALDI-TOF/TOF.

| Chimères | $B_n = A$ | | $B_n = C$ | | $B_n = G$ | | $B_n = U$ | | |
|---------------|-----------|----------|-------------------|----------|-------------|----------|------------|--------|-----------|
| 5-mer | T T C G | 200 pmol | | | | | | | |
| 7-mer | | | | | T T A A A C | 250 pmol | | | |
| 10-mer | | | T T T A A A G G G | 250 pmol | | | | | |
| 4-mer | TTT | 5 pmol | TTT | 5 pmol | TTT | 5 pmol | GGG | 5 pmol | |
| | TTC | | TTA | | TTA | | GGA | | |
| | TTGG | | TTG | | TTG | | GGC | | |
| | TCC | | TAA | | TAA | | GAA | | |
| | TGG | | TGG | | TAC | | GCC | | |
| | TCC | TAG | TCC | GAC | 12,5 pmol | | | | |
| | GGC | 50 pmol | GGA | 50 pmol | AAA | 50 pmol | AAC | 5 pmol | |
| | CCC | | GGG | | CCC | | CAC et CCA | | |
| | GGG | | AAA | | AAC | | AAA | | 12,5 pmol |
| | CCG | | AAG | | CCA | | ACC | | 50 pmol |
| | 100 pmol | | 200 pmol | | 400 pmol | CCC | 200 pmol | | |

Tableau 1.2: Seuil de détection par MALDI-TOF/TOF en pmol/spot pour l'analyse des chimères de 4, 5, 7 et 10-mer.

Il montre que les chimères ARN/ADN, de 4 à 10-mer, sont analysables bien que leur quantité varie de 5 à 400 pmol par spot. Plus le nombre de nucléotides des chimères ARN/ADN augmente, plus elles doivent être concentrées pour être obtenir un signal suffisamment abondant.

Par ailleurs, les chimères de 4-mer constituées de la base $B_4(A, C \text{ ou } G)$ et d'une (ou plusieurs) base(s) T ont un seuil de détection inférieur à celles sans thymine. Les chimères contenant des thymines sont donc plus facilement détectables que les autres.

Inversement, les chimères de 4-mer, sans thymine, constituées d'une base $B_4(A \text{ ou } C)$, et d'une seule guanine sont plus difficilement détectables que celles contenant plusieurs guanines. En l'absence de thymine, la présence des guanines améliorent donc la détection de la chimère. De plus, les chimères de 4-mer, sans thymine, constituées d'une base $B_4(G)$, sont plus difficilement détectables lorsque le nombre d'adénines diminue. En l'absence de thymine, la présence des adénines améliorent également la détection des chimères. Les chimères constituées de 3 adénines sont analysées à une concentration deux fois plus élevée que les chimères qui contiennent 3 guanines. Les guanines améliorent plus la détection des chimères que les adénines.

Enfin, les chimères de 4-mer constituées de la base $B_4(U)$ et de guanines sont détectées à des concentrations inférieures à celles sans guanine. Pour les chimères qui contiennent une seule guanine et celles ne contenant pas de guanine, plus le nombre de cytosines augmente plus les chimères sont analysées à des concentrations plus élevées. Les guanines améliorent la

détection des chimères tandis que les cytosines la diminuent. Les chimères constituées de la base B₄(U) et ne contenant pas de guanine sont plus difficilement détectables lorsque le nombre d'adénines diminuent et que le nombre de cytosines augmentent.

Certaines chimères sont donc plus facilement détectées que d'autres. La détection des chimères dépend de la nature des bases et du nombre de nucléotides présents mais moins de la position des bases dans la séquence. La facilité de détection des bases est donc la suivante : T, U ≥ G ≥ A ≥ C. Il suffit donc d'augmenter la concentration des chimères ARN/ADN pour toutes les détecter.

3.2. L'analyse des chimères ARN/ADN de 4-mer par MALDI-TOF/TOF

Les fragmentations en mode CID des chimères ARN/ADN de 4-mer ont été étudiées par MALDI-TOF/TOF.

3.2.1. Les fragmentations en mode CID des chimères avec une base B₄(A)

Les ions fragments obtenus en mode CID de l'ion parent [M-H]⁻ (m/z 1268,808) des chimères isomères GCTA, GTCAA, TCGA, TGCAA, CGTA et CTGA sont indiqués dans le tableau 1.3.

Ils ont été comparés à ceux observés dans les spectres CID des oligonucléotides déprotonés de même séquence contenant à la place du ribonucléotide B₄ et du groupement 3'-PO₄H₂: (i) un ribonucléotide et un groupement 3'-OH (M=1189,836 Da) ; (ii) un désoxynucléotide et un groupement 3'-PO₄H₂ (M=1253,817 Da) et (iii) un désoxynucléotide et un groupement 3'-OH (M=1173,837 Da) (tableau 1.3).

| Fragments des chimères 5'OH-B ₁ B ₂ B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} | | | Masse | | | | | |
|---|--|---|----------|----------|----------|----------|----------|----------|
| Ion m/z | Spécifique | Séquence | GCTA | GTCA | TCGA | TGCA | CGTA | CTGA |
| PO ₃ ⁻ | | PO ₃ | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 |
| PO ₄ H ₂ ⁻ | | PO ₄ H ₂ | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 |
| C ⁻ | B ₁ =C | C | | | | | 110,094 | 110,094 |
| T ⁻ | | T | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 |
| A ⁻ | | A | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 |
| G ⁻ | | G | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 |
| [B, PO-H ₂] ⁻ | | B, PO-H ₂ | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 |
| [B, PO ₃] ⁻ | | B, PO ₃ | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 |
| [B, PO ₄ H] ⁻ | | B, PO ₄ H | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 |
| [B, PO ₄ H ₂ , PO ₃] ⁻ | | B, PO ₄ H ₂ , PO ₃ | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 |
| [B, PO ₄ H ₂ , PO ₄ H] ⁻ | | B, PO ₄ H ₂ , PO ₄ H | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 |
| [B(C) PO ₄] ⁻ | B ₁ ≠C | B(C) PO ₄ | 288,177 | 288,177 | 288,177 | 288,177 | | |
| c ₁ ⁻ | | ⁵ OH-B ₁ -PO ₃ | 328,202 | 328,202 | 303,189 | 303,189 | 288,177 | 288,177 |
| d ₁ ⁻ | | ⁵ OH-B ₁ -PO ₄ H | 346,217 | 346,217 | 321,204 | 321,204 | 306,192 | 306,192 |
| z ₁ ⁻ | | <u>A</u> -PO ₄ H ₂ ^{3'} | 328,201 | 328,201 | 328,201 | 328,201 | 328,201 | 328,201 |
| y ₁ ⁻ | | O- <u>A</u> -PO ₄ H ₂ ^{3'} | 346,216 | 346,216 | 346,216 | 346,216 | 346,216 | 346,216 |
| [w ₂ , a ₃] ⁻ | | PO ₄ H-B ₃ | 303,190 | 288,177 | 328,202 | 288,177 | 303,190 | 328,202 |
| [w ₂ , c ₃] ⁻ | | PO ₄ H-B ₃ -PO ₃ | 383,169 | 368,157 | 408,182 | 368,157 | 383,169 | 408,182 |
| [a ₂ -B ₂ H] ⁻ | B ₂ ≠T | ⁵ OH-B ₁ -désoxy | 426,302 | | 401,289 | 401,289 | 386,277 | |
| x ₁ ⁻ | | <u>A</u> -PO ₄ H ₂ ^{3'} | 408,181 | 408,181 | 408,181 | 408,181 | 408,181 | 408,181 |
| w ₁ ⁻ | | O- <u>A</u> -PO ₄ H ₂ ^{3'} | 426,196 | 426,196 | 426,196 | 426,196 | 426,196 | 426,196 |
| c ₂ ⁻ | | ⁵ OH-B ₁ B ₂ -PO ₃ | 617,387 | 632,399 | 592,374 | 632,399 | 617,387 | 592,374 |
| d ₂ ⁻ | | ⁵ OH-B ₁ B ₂ -PO ₄ H | 635,402 | 650,414 | 610,389 | 650,414 | 635,402 | 610,389 |
| z ₂ ⁻ | | B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} | 632,398 | 617,386 | 657,411 | 617,386 | 632,398 | 657,411 |
| y ₂ ⁻ | | O-B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} | 650,413 | 635,401 | 675,426 | 635,401 | 650,413 | 675,426 |
| x ₂ ⁻ | | PO ₃ -B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} | 712,378 | 697,366 | 737,391 | 697,366 | 712,378 | 737,391 |
| w ₂ ⁻ | | PO ₄ H-B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} | 730,393 | 715,381 | 755,406 | 715,381 | 730,393 | 755,406 |
| [c ₃ -B ₃ H] ⁻ | B ₃ ≠T et 3'-PO ₄ H ₂ | ⁵ OH-B ₁ B ₂ -désoxy-PO ₃ | | 810,482 | 770,458 | 810,482 | | 770,458 |
| a ₃ ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ | 841,604 | 841,604 | 841,604 | 841,604 | 841,604 | 841,604 |
| b ₃ ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ -O | 859,619 | 859,619 | 859,619 | 859,619 | 859,619 | 859,619 |
| [a ₃ -B ₃ H] ⁻ | B ₃ ≠T et 3'-PO ₄ H ₂ ^{3'} | ⁵ OH-B ₁ -B ₂ -désoxy | | 730,499 | 690,474 | 730,499 | | 690,474 |
| c ₃ ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ -PO ₃ | 921,584 | 921,584 | 921,584 | 921,584 | 921,584 | 921,584 |
| d ₃ ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ -PO ₄ H | 939,599 | 939,599 | 939,599 | 939,599 | 939,599 | 939,599 |
| z ₃ ⁻ | | B ₂ B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} | 921,583 | 921,583 | 946,596 | 946,596 | 961,608 | 961,608 |
| y ₃ ⁻ | | O-B ₂ B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} | 939,598 | 939,598 | 964,611 | 964,611 | 979,623 | 979,623 |
| x ₃ ⁻ | | PO ₃ -B ₂ B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} | 1001,563 | 1001,563 | 1026,576 | 1026,576 | 1041,588 | 1041,588 |
| w ₃ ⁻ | | PO ₄ H-B ₂ B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} | 1019,578 | 1019,578 | 1044,591 | 1044,591 | 1059,603 | 1059,603 |
| [M-H-B(G)H] ⁻ | | chimère - base G | 1117,682 | 1117,682 | 1117,682 | 1117,682 | 1117,682 | 1117,682 |
| [M-H-B(A)H] ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ -désoxy-PO ₄ H ₂ ^{3'} | 1133,681 | 1133,681 | 1133,681 | 1133,681 | 1133,681 | 1133,681 |
| [M-H-B(C)H] ⁻ | | chimère - base C | 1157,706 | 1157,706 | 1157,706 | 1157,706 | 1157,706 | 1157,706 |
| b ₄ ⁻ | 3'-PO ₄ H ₂ | ⁵ OH-B ₁ B ₂ B ₃ <u>A</u> -OH | 1188,828 | 1188,828 | 1188,828 | 1188,828 | 1188,828 | 1188,828 |
| [M-H] ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} | 1268,808 | 1268,808 | 1268,808 | 1268,808 | 1268,808 | 1268,808 |

Tableau 1.3: Séquences et masses monoisotopiques des spectres CID, par MALDI-TOF/TOF, en mode négatif, de l'ion parent [M-H]⁻ des chimères GCTA, GTCAA, TCGA, TGCAA, CGTA et CTGAA. La base A, soulignée et en gras, est un ribonucléotide possédant un groupement 3'-PO₄H₂.

Les fragmentations obtenues en mode CID des ions [M-H]⁻ montrent le clivage de la chaîne nucléotidique au niveau des liaisons du phosphodiester donnant lieu à la formation de huit séries d'ions fragments : ceux qui sont composés de l'extrémité 5'-OH de l'ion parent (a_n⁻, b_n⁻,

c_n^- et d_n^-) et ceux qui sont composés de l'extrémité 3'-PO₄H₂ de l'ion parent (w_n^- , x_n^- , y_n^- et z_n^-).

Ils montrent également la perte des bases (A, C ou G) qui engendre les ions fragments suivants : $[M-H-BH_n]^-$ et $[a_n-B_nH]^-$. De plus, l'ion fragment b_4^- correspond à la perte de PO₃H₂ du groupement 3'-PO₄H₂. Cette dissociation est spécifique de la présence groupement 3'-PO₄H₂. Les ions fragments $[a_3-B_3H]^-$ et $[c_3-B_3H]^-$ sont absents si la base B₃ est un T et qu'il y a un groupement 3'-phosphate. Enfin, si la cytosine est en position B₁, l'anion C⁻ est observée tandis que si elle est en position B₂ ou B₃ l'ion fragment $[B(C) PO_4]^-$ du désoxynucléotide est présent.

L'analyse des ions fragments de la chimère permet donc d'analyser sa séquence (tableau 1.3). Par ailleurs, les ions fragments obtenus en mode CID n'ont pas la même abondance suivant la séquence étudiée. La figure 1.2 présente les ions fragments les plus abondants de l'ion parent $[M-H]^-$ des 6 isomères.

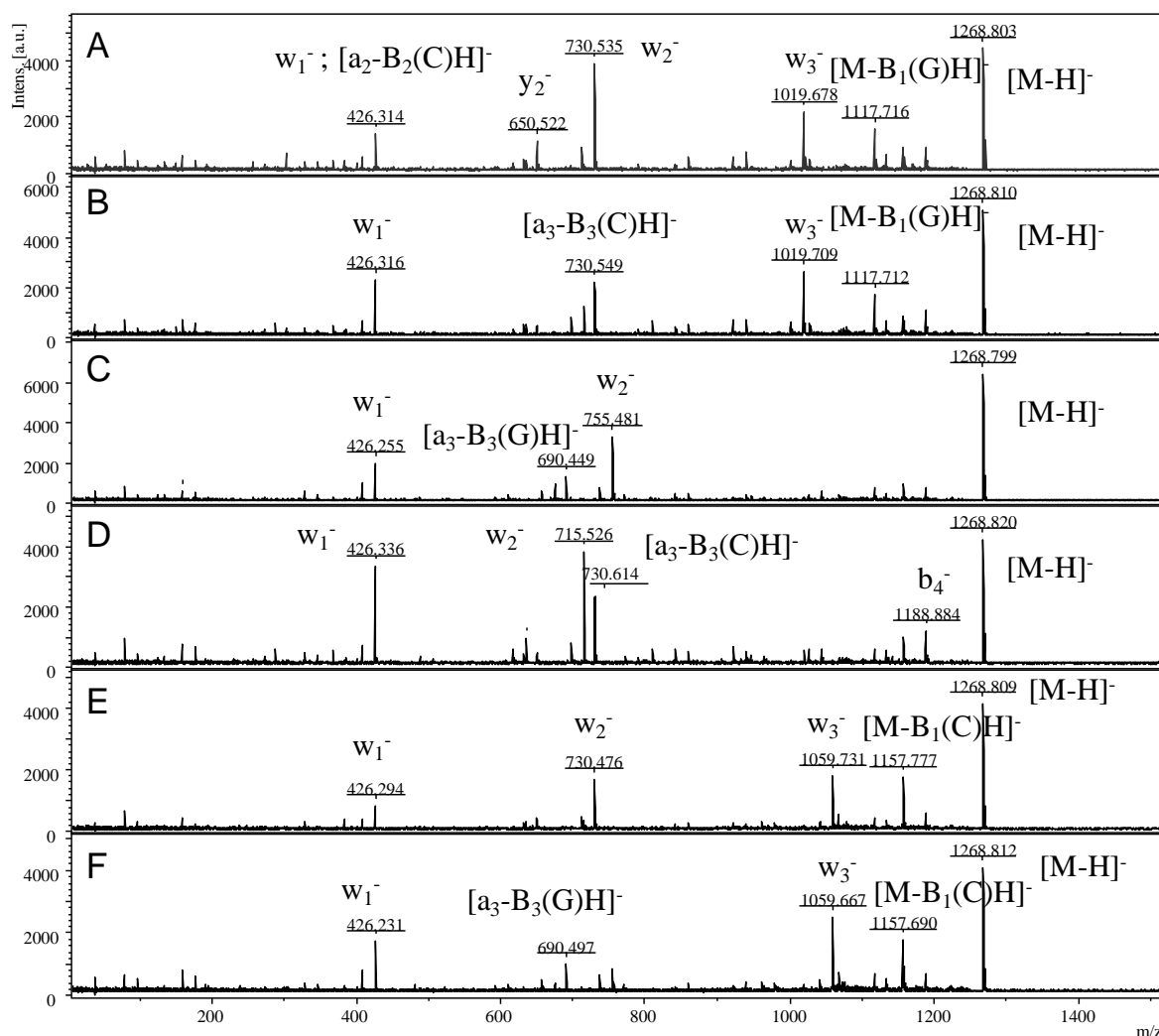


Figure 1.2: Spectres CID, par MALDI-TOF/TOF, en mode négatif, de l'ion parent $[M-H]^-$ m/z 1268,808 des isomères : (A) GCTA; (B) GTCA; (C) TCGA; (D) TGCA; (E) CGTA et (F) CTGA. La base A, soulignée et en gras, est un ribonucléotide possédant un groupement 3'-PO₄H₂.

L'ion w_1^- est toujours un des ions les plus abondants (figure 1.2). Par contre, l'abondance des ions w_3^- , w_2^- , $[M-H-B_1H]^-$ et $[a_3-B_3H]^-$ dépend de la présence de la thymine. Les ions w_3^- et $[M-H-B_1H]^-$ sont abondants sauf si la thymine est en position B₁. L'ion w_2^- est abondant sauf si la thymine est en position B₂. L'ion $[a_3-B_3H]^-$ est abondant sauf si la thymine est en position B₃. Enfin deux autres ions suivants sont abondants que dans le cas d'une seule chimère : l'ion y_2^- est également un des ions les plus abondants dans le cas de la chimère GCTA tandis que l'ion b_4^- est également un des ions les plus abondants dans le cas de la chimère TGCA.

L'abondance des ions fragments de la chimère dépend donc de sa séquence. Par ailleurs, les ions w_n^- et les ions issus de la perte de la base sont les plus abondants. Toutefois, les spectres CID des oligonucléotides déprotonés de même séquence mais contenant soit un ribonucléotide et un groupement 3'-OH, soit un désoxynucléotide B₄ et du groupement 3'-

PO_4H_2 ou soit un désoxynucléotide B_4 et un groupement 3'-OH, donnent également les ions abondants w_n^- , et ceux issus de la perte de la base abondants.

3.2.2. Les fragmentations en mode CID des chimères avec une base $\text{B}_4(\text{C})$

Les ions fragments obtenus en mode CID de l'ion parent $[\text{M}-\text{H}]^-$ (m/z 1268,808) des isomères : $\text{GAT}\underline{\text{C}}$, $\text{GTAC}\underline{\text{C}}$, $\text{TAG}\underline{\text{C}}$, $\text{TGAC}\underline{\text{C}}$, $\text{ATGC}\underline{\text{C}}$ et $\text{AGTC}\underline{\text{C}}$ sont indiqués dans le tableau 1.4.

Ils ont été comparés à ceux observés dans les spectres CID des oligonucléotides de même séquence contenant à la place du ribonucléotide B_4 et du groupement 3'- PO_4H_2 : (i) un ribonucléotide et un groupement 3'-OH ($M=1189,836$ Da) ; (ii) un désoxynucléotide et un groupement 3'- PO_4H_2 ($M=1253,817$ Da) et (iii) un désoxynucléotide et un groupement 3'-OH ($M=1169,816$ Da) (tableau 1.4).

| Fragments des chimères 5'OH-B ₁ B ₂ B ₃ <u>C</u> -PO ₄ H ₂ ^{3'} | | | Masse | | | | | |
|---|--|---|----------|----------|----------|----------|----------|----------|
| Ion m/z | Spécifique | Séquence | GATC | GTAC | TAGC | TGAC | ATGC | AGTC |
| PO ₃ ⁻ | | PO ₃ | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 |
| PO ₄ H ₂ ⁻ | | PO ₄ H ₂ | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 |
| C ⁻ | B ₄ =CTP et 3'-PO ₄ H ₂ | C | 110,094 | 110,094 | 110,094 | 110,094 | 110,094 | 110,094 |
| T ⁻ | | T | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 |
| A ⁻ | | A | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 |
| G ⁻ | | G | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 |
| [B, PO-H ₂] ⁻ | | B, PO-H ₂ | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 |
| [B, PO ₃] ⁻ | | B, PO ₃ | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 |
| [B, PO ₄ H] ⁻ | | B, PO ₄ H | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 |
| [B, PO ₄ H ₂ , PO ₃] ⁻ | | B, PO ₄ H ₂ , PO ₃ | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 |
| [B, PO ₄ H ₂ , PO ₄ H] ⁻ | | B, PO ₄ H ₂ , PO ₄ H | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 |
| [B ₄ PO ₄] ⁻ | B ₄ =C | B ₄ PO ₄ | 304,176 | 304,176 | 304,176 | 304,176 | 304,176 | 304,176 |
| c ₁ ⁻ | | ⁵ OH-B ₁ -PO ₃ | 328,202 | 328,202 | 303,189 | 303,189 | 312,202 | 312,202 |
| d ₁ ⁻ | | ⁵ OH-B ₁ -PO ₄ H | 346,217 | 346,217 | 321,204 | 321,204 | 330,217 | 330,217 |
| z ₁ ⁻ | | <u>C</u> -PO ₄ H ₂ ^{3'} | 304,176 | 304,176 | 304,176 | 304,176 | 304,176 | 304,176 |
| y ₁ ⁻ | | O- <u>C</u> -PO ₄ H ₂ ^{3'} | 322,191 | 322,191 | 322,191 | 322,191 | 322,191 | 322,191 |
| [w ₂ , a ₃] ⁻ | | PO ₄ H-B ₃ | 303,189 | 312,202 | 328,203 | 312,202 | 328,203 | 303,189 |
| [w ₂ , c ₃] ⁻ | | PO ₄ H-B ₃ -PO ₃ | 383,169 | 392,182 | 408,182 | 392,182 | 408,182 | 383,169 |
| [a ₂ -B ₂ H] ⁻ | B ₂ ≠ T | ⁵ OH-B ₁ -désoxy | 426,302 | | 401,289 | 401,289 | | 410,302 |
| x ₁ ⁻ | | <u>C</u> -PO ₄ H ₂ ^{3'} | 384,156 | 384,156 | 384,156 | 384,156 | 384,156 | 384,156 |
| w ₁ ⁻ | | O- <u>C</u> -PO ₄ H ₂ ^{3'} | 402,171 | 402,171 | 402,171 | 402,171 | 402,171 | 402,171 |
| c ₂ ⁻ | | ⁵ OH-B ₁ B ₂ -PO ₃ | 721,392 | 632,399 | 616,399 | 632,399 | 616,399 | 721,392 |
| d ₂ ⁻ | | ⁵ OH-B ₁ B ₂ -PO ₄ H | 739,407 | 650,414 | 634,414 | 650,414 | 634,414 | 739,407 |
| z ₂ ⁻ | | B ₃ <u>C</u> -PO ₄ H ₂ ^{3'} | 608,373 | 617,386 | 633,386 | 617,386 | 633,386 | 608,373 |
| y ₂ ⁻ | | O-B ₃ <u>C</u> -PO ₄ H ₂ ^{3'} | 626,388 | 635,401 | 651,401 | 635,401 | 651,401 | 626,388 |
| x ₂ ⁻ | | PO ₃ -B ₃ <u>C</u> -PO ₄ H ₂ ^{3'} | 688,353 | 697,366 | 713,366 | 697,366 | 713,366 | 688,353 |
| w ₂ ⁻ | | PO ₄ H-B ₃ <u>C</u> -PO ₄ H ₂ ^{3'} | 706,368 | 715,381 | 731,381 | 715,381 | 731,381 | 706,368 |
| [c ₃ -B ₃ H] ⁻ | B ₃ ≠ T et 3'-PO ₄ H ₂ | ⁵ OH-B ₁ B ₂ -ribose-PO ₃ | | 810,482 | 794,483 | 810,482 | 794,483 | |
| a ₃ ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ | 865,629 | 865,629 | 865,629 | 865,629 | 865,629 | 865,629 |
| b ₃ ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ -O | 883,644 | 883,644 | 883,644 | 883,644 | 883,644 | 883,644 |
| [a ₃ -B ₃ H] ⁻ | B ₃ ≠ T et 3'-PO ₄ H ₂ | ⁵ OH-B ₁ -B ₂ -désoxy | | 730,499 | 714,499 | 730,499 | 714,499 | |
| c ₃ ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ -PO ₃ | 945,609 | 945,609 | 945,609 | 945,609 | 945,609 | 945,609 |
| d ₃ ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ -PO ₄ H | 963,624 | 963,624 | 963,624 | 963,624 | 963,624 | 963,624 |
| z ₃ ⁻ | | B ₂ B ₃ <u>C</u> -PO ₄ H ₂ ^{3'} | 921,583 | 921,583 | 946,596 | 946,596 | 937,583 | 937,583 |
| y ₃ ⁻ | | O-B ₂ B ₃ <u>C</u> -PO ₄ H ₂ ^{3'} | 939,598 | 939,598 | 964,611 | 964,611 | 955,598 | 955,598 |
| x ₃ ⁻ | | PO ₃ -B ₂ B ₃ <u>C</u> -PO ₄ H ₂ ^{3'} | 1001,563 | 1001,563 | 1026,576 | 1026,576 | 1017,563 | 1017,563 |
| w ₃ ⁻ | | PO ₄ H-B ₂ B ₃ <u>C</u> -PO ₄ H ₂ ^{3'} | 1019,578 | 1019,578 | 1044,591 | 1044,591 | 1035,578 | 1035,578 |
| [M-H-B(G)H] ⁻ | | chimère - base G | 1117,682 | 1117,682 | 1117,682 | 1117,682 | 1117,682 | 1117,682 |
| [M-H-B(A)H] ⁻ | | chimère - base A | 1133,682 | 1133,681 | 1133,681 | 1133,681 | 1133,681 | 1133,681 |
| [M-H-B(C)H] ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ -désoxy-PO ₄ H ₂ ^{3'} | 1157,706 | 1157,706 | 1157,706 | 1157,706 | 1157,706 | 1157,706 |
| a ₄ ⁻ | B ₄ =C et 3'-PO ₄ H ₂ | ⁵ OH-B ₁ B ₂ B ₃ <u>C</u> | 1170,813 | 1170,813 | 1170,813 | 1170,813 | 1170,813 | 1170,813 |
| b ₄ ⁻ | 3'-PO ₄ H ₂ | ⁵ OH-B ₁ B ₂ B ₃ <u>C</u> -OH | 1188,828 | 1188,828 | 1188,828 | 1188,828 | 1188,828 | 1188,828 |
| [M-H] ⁻ | | ⁵ OH-B ₁ B ₂ B ₃ <u>C</u> -PO ₄ H ₂ ^{3'} | 1268,808 | 1268,808 | 1268,808 | 1268,808 | 1268,808 | 1268,808 |

Tableau 1.4: Séquences et masses monoisotopiques des spectres CID, par MALDI-TOF/TOF, en mode négatif de l'ion parent [M-H]⁻ des chimères GATC, GTAC, TAGC, TGAC, ATGC et AGTC. La base C, soulignée et en gras, est un ribonucléotide possédant un groupement 3'-PO₄H₂.

Les différents types d'ions fragments obtenus en mode CID des chimères B₄(A ou C) déprotonées sont comparables et permettent l'analyse des séquences (tableaux 1.3 et 1.4). Cependant, l'ion fragment a₄⁻, spécifique de la dissociation de la chimère B₄(C), est observé. Il

correspond à la perte du groupement 3'-PO₄H₂ de la chimère B₄(C). De plus, si la base est B₄(C), la perte de l'anion C⁻ est observée comme dans le cas où la base est B₁(C). Lorsque la base est B₄(C), l'ion fragment [**B**₄ PO₄]⁻ du ribonucléotide, est observé tandis que lorsqu'elle est B₂(C) ou B₃(C), l'ion fragment [B(C) PO₄]⁻ du désoxynucléotide, est observé.

Par ailleurs, tout comme les chimères possédant une base B₄(A), les ions fragments des chimères B₄(C), n'ont pas la même abondance suivant les séquences. La figure 1.3 présente les ions fragments obtenus en mode CID les plus abondants de l'ion parent [M-H]⁻ des 6 chimères isomères.

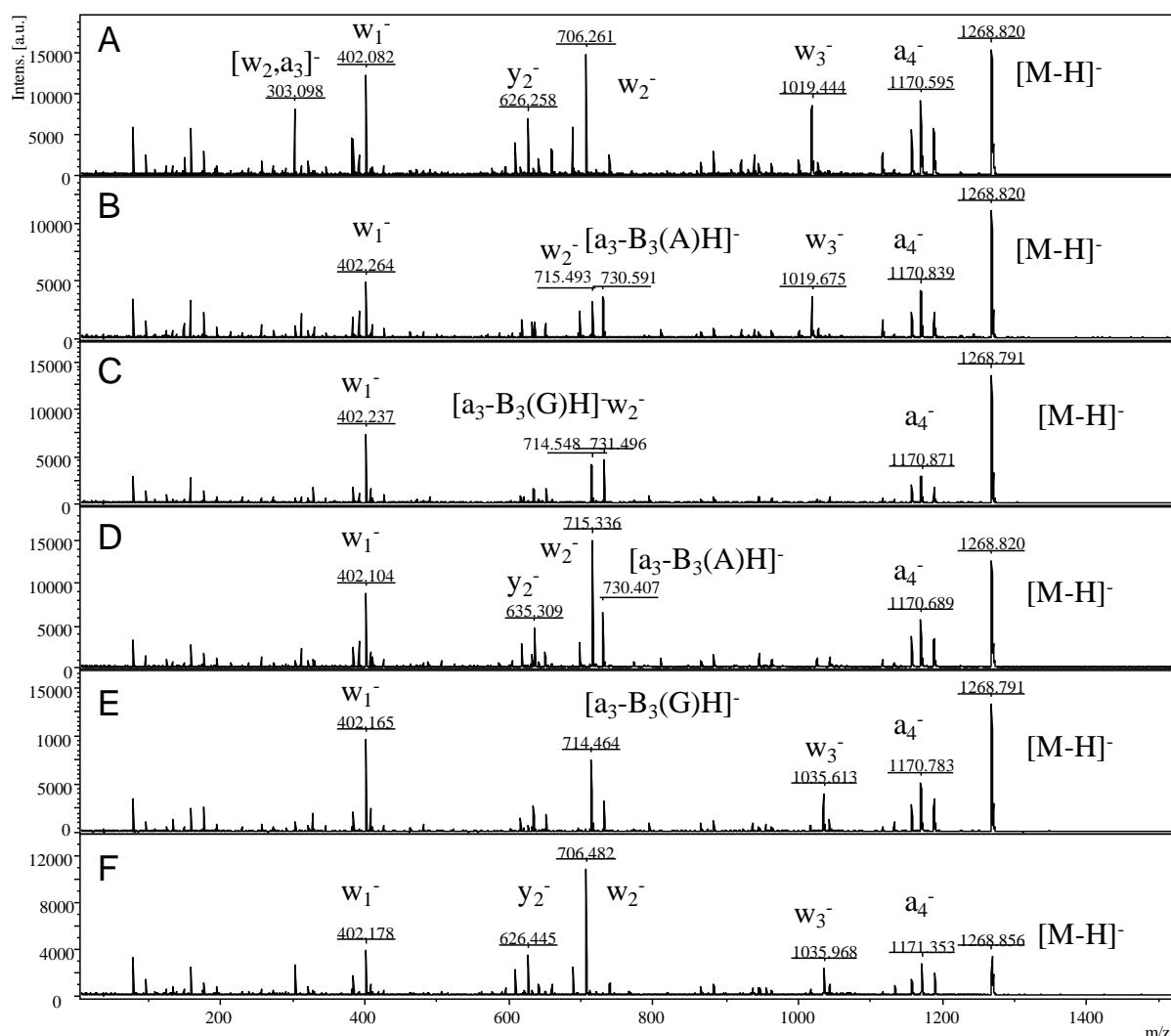


Figure 1.3: Spectres CID par MALDI-TOF-TOF, en mode négatif, de l'ion parent [M-H]⁻ m/z 1268,808 des chimères : (A) GATC; (B) GTAC; (C) TAGC; (D) TGAC; (E) ATGC et (F) AGTC. La base C, soulignée et en gras, est un ribonucléotide possédant un groupement 3'-PO₄H₂.

Les ions w_1^- et a_4^- sont toujours abondants tandis que l'ion w_2^- est abondant sauf pour la chimère ATGC. Par contre, l'abondance des ions w_3^- et $[a_3-B_3H]^-$ dépend de la présence de la thymine. L'ion w_3^- est abondant sauf si la thymine est en position B_1 tandis que l'ion $[a_3-B_3H]^-$ est abondant sauf si elle est en position B_3 . Enfin, deux autres ions sont abondants que dans le cas de certaines chimères. L'ion y_2^- est également un des ions fragments les plus abondants des chimères GATC et TGACC tandis que l'ion $[w_2,a_3]^-$ est un des ions les plus abondants de la chimère GATC. Dans la région des 4-mer, les ions $[M-H-B_1H]^-$ sont moins abondants que l'ion a_4^- et ne font donc pas partie des ions fragments les plus abondants des chimères qui contiennent une base $B_4(C)$.

L'étude des ions fragments et de ceux les plus abondants permet également l'analyse de la séquence de ces chimères.

Enfin, les ions fragments les plus abondants sont les ions w_n^- , les ions issus de la perte de la base et l'ion a_4^- . L'abondance de l'ion a_4^- est spécifique de ces chimères $B_4(C)$ car les spectres CID des oligonucléotides de même séquence mais contenant soit un ribonucléotide et un groupement 3'-OH, soit un désoxynucléotide B_4 et du groupement 3'- PO_4H_2 ou soit un désoxynucléotide B_4 et un groupement 3'-OH donnent les ions abondants w_n^- , et ceux issus de la perte de la base abondants.

3.2.3. Les fragmentations en mode CID des chimères avec une base $B_4(G)$

Les ions fragments obtenus en mode CID de l'ion parent $[M-H]^-$ (m/z 1268,808) des chimères : CATG, CTAGG, ACTG, ATCG, TACG et TCAG sont indiqués dans le tableau 1.5.

| Fragments des chimères 5'OH-B ₁ B ₂ B ₃ G-PO ₄ H ₂ ^{3'} | | | Masse | | | | | |
|---|--|--|----------|----------|----------|----------|----------|----------|
| Ion m/z | Spécifique | Séquence | CATG | CTAG | ACTG | ATCG | TACG | TCAG |
| PO ₃ ⁻ | | PO ₃ | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 |
| PO ₄ H ₂ ⁻ | | PO ₄ H ₂ | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 |
| C ⁻ | B ₁ =C | C | 110,094 | 110,094 | | | | |
| T ⁻ | | T | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 |
| A ⁻ | | A | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 |
| G ⁻ | | G | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 |
| [B, PO-H ₂] ⁻ | | B, PO-H ₂ | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 |
| [B, PO ₃] ⁻ | | B, PO ₃ | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 |
| [B, PO ₄ H] ⁻ | | B, PO ₄ H | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 |
| [B, PO ₄ H ₂ , PO ₃] ⁻ | | B, PO ₄ H ₂ , PO ₃ | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 |
| [B, PO ₄ H ₂ , PO ₄ H] ⁻ | | B, PO ₄ H ₂ , PO ₄ H | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 |
| [B(C) PO ₄] ⁻ | B ₁ ≠C | B(C) PO ₄ | | | 288,177 | 288,177 | 288,177 | 288,177 |
| c ₁ ⁻ | | 5'OH-B ₁ -PO ₃ | 288,177 | 288,177 | 312,202 | 312,202 | 303,189 | 303,189 |
| d ₁ ⁻ | | 5'OH-B ₁ -PO ₄ H | 306,192 | 306,192 | 330,217 | 330,217 | 321,204 | 321,204 |
| z ₁ ⁻ | | <u>G</u> -PO ₄ H ₂ ^{3'} | 344,201 | 344,201 | 344,201 | 344,201 | 344,201 | 344,201 |
| y ₁ ⁻ | | O- <u>G</u> -PO ₄ H ₂ ^{3'} | 362,216 | 362,216 | 362,216 | 362,216 | 362,216 | 362,216 |
| [w ₂ ,a ₃] ⁻ | | PO ₄ H-B ₃ | 303,189 | 312,202 | 303,189 | 288,177 | 288,177 | 312,202 |
| [w ₂ ,c ₃] ⁻ | | PO ₄ H-B ₃ -PO ₃ | 383,169 | 392,182 | 383,169 | 368,157 | 368,157 | 392,182 |
| [a ₂ -B ₂ H] ⁻ | B ₂ ≠T | 5'OH-B ₁ -désoxy | 386,277 | | 410,302 | | 401,289 | 401,289 |
| x ₁ ⁻ | | <u>G</u> -PO ₄ H ₂ ^{3'} | 424,201 | 424,201 | 424,201 | 424,201 | 424,201 | 424,201 |
| w ₁ ⁻ | | O- <u>G</u> -PO ₄ H ₂ ^{3'} | 442,216 | 442,216 | 442,216 | 442,216 | 442,216 | 442,216 |
| c ₂ ⁻ | | 5'OH-B ₁ B ₂ -PO ₃ | 601,387 | 592,374 | 601,387 | 616,399 | 616,399 | 592,374 |
| d ₂ ⁻ | | 5'OH-B ₁ B ₂ -PO ₄ H | 619,402 | 610,389 | 619,402 | 634,414 | 634,414 | 610,389 |
| z ₂ ⁻ | | B ₃ <u>G</u> -PO ₄ H ₂ ^{3'} | 648,398 | 657,411 | 648,398 | 633,386 | 617,387 | 657,411 |
| y ₂ ⁻ | | O-B ₃ <u>G</u> -PO ₄ H ₂ ^{3'} | 666,413 | 675,426 | 666,413 | 651,401 | 635,402 | 675,426 |
| x ₂ ⁻ | | PO ₃ -B ₃ <u>G</u> -PO ₄ H ₂ ^{3'} | 728,378 | 737,391 | 728,378 | 713,366 | 713,366 | 737,391 |
| w ₂ ⁻ | | PO ₄ H-B ₃ <u>G</u> -PO ₄ H ₂ ^{3'} | 746,393 | 755,406 | 746,393 | 731,381 | 731,381 | 755,406 |
| [c ₃ -B ₃ H] ⁻ | B ₃ ≠T et 3'-PO ₄ H ₂ | 5'OH-B ₁ B ₂ -désoxy-PO ₃ | | 770,457 | | 794,482 | 794,482 | 770,457 |
| a ₃ ⁻ | | 5'OH-B ₁ B ₂ B ₃ | 825,604 | 825,604 | 825,604 | 825,604 | 825,604 | 825,604 |
| b ₃ ⁻ | | 5'OH-B ₁ B ₂ B ₃ -O | 843,619 | 843,619 | 843,619 | 843,619 | 843,619 | 843,619 |
| [a ₃ -B ₃ H] ⁻ | B ₃ ≠T et 3'-PO ₄ H ₂ | 5'OH-B ₁ -B ₂ -désoxy | | 690,474 | | 714,499 | 714,499 | 690,474 |
| c ₃ ⁻ | | 5'OH-B ₁ B ₂ B ₃ -PO ₃ | 905,584 | 905,584 | 905,584 | 905,584 | 905,584 | 905,584 |
| d ₃ ⁻ | | 5'OH-B ₁ B ₂ B ₃ -PO ₄ H | 923,599 | 923,599 | 923,599 | 923,599 | 923,599 | 923,599 |
| z ₃ ⁻ | | B ₂ B ₃ <u>G</u> -PO ₄ H ₂ ^{3'} | 961,608 | 961,608 | 937,583 | 937,583 | 946,596 | 946,596 |
| y ₃ ⁻ | | O-B ₂ B ₃ <u>G</u> -PO ₄ H ₂ ^{3'} | 979,623 | 979,623 | 955,598 | 955,598 | 964,611 | 964,611 |
| x ₃ ⁻ | | PO ₃ -B ₂ B ₃ <u>G</u> -PO ₄ H ₂ ^{3'} | 1041,588 | 1041,588 | 1017,563 | 1017,563 | 1026,576 | 1026,576 |
| w ₃ ⁻ | | PO ₄ H-B ₂ B ₃ <u>G</u> -PO ₄ H ₂ ^{3'} | 1059,603 | 1059,603 | 1035,578 | 1035,578 | 1044,591 | 1044,591 |
| [M-H-B(G)H] ⁻ | | 5'OH-B ₁ B ₂ B ₃ -désoxy-PO ₄ H ₂ ^{3'} | 1117,682 | 1117,682 | 1117,682 | 1117,682 | 1117,682 | 1117,682 |
| [M-H-B(A)H] ⁻ | | chimère - base A | 1133,682 | 1133,681 | 1133,681 | 1133,681 | 1133,681 | 1133,681 |
| [M-H-B(C)H] ⁻ | | chimère - base C | 1157,706 | 1157,706 | 1157,706 | 1157,706 | 1157,706 | 1157,706 |
| b ₄ ⁻ | 3'-PO ₄ H ₂ | 5'OH-B ₁ B ₂ B ₃ <u>G</u> -OH | 1188,828 | 1188,828 | 1188,828 | 1188,828 | 1188,828 | 1188,828 |
| [M-H] ⁻ | | 5'OH-B ₁ B ₂ B ₃ <u>G</u> -PO ₄ H ₂ ^{3'} | 1268,808 | 1268,808 | 1268,808 | 1268,808 | 1268,808 | 1268,808 |

Tableau 1.5: Séquences et masses monoisotopiques des spectres CID, obtenus par MALDI-TOF/TOF, en mode négatif, de l'ion parent [M-H]⁻ des chimères CATG, CTAG, ACTG, ATCG, TACG et TCAG. La base G, soulignée et en gras, est un ribonucléotide possédant un groupement 3'-PO₄H₂.

Les différents types d'ions fragments obtenus en mode CID des chimères B₄(A, C ou G) déprotonées sont comparables (tableaux 1.3 à 1.5). L'analyse, des ions fragments et des plus abondants, permet de déterminer la séquence de ces chimères isomères. Lorsque la chimère est B₄(C), l'ion fragment a₄⁻ est observé tandis qu'il n'est pas observé lorsque B₄(A ou G).

Tout comme les chimères B₄(A, C), les ions fragments les plus abondants sont les ions w_n⁻ et les ions issus de la perte de la base.

3.2.4. Les fragmentations en mode CID des chimères avec une base B₄(U)

Les ions fragments obtenus en mode CID de l'ion parent [M-H]⁻ (m/z 1254,781) des chimères : GACU, GCAU, CAGU, CGAU, ACGU et AGCU sont indiqués dans le tableau 1.6.

| Fragments des chimères 5'OH-B ₁ B ₂ B ₃ <u>U</u> -PO ₄ H ₂ ^{3'} | | | Masse | | | | | |
|---|--|--|----------|----------|----------|----------|----------|----------|
| Ion m/z | Spécifique | Séquence | GACU | GCAU | CAGU | CGAU | ACGU | AGCU |
| PO ₃ ⁻ | B ₁ =C | PO ₃ | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 |
| PO ₄ H ₂ ⁻ | | PO ₄ H ₂ | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 |
| C ⁻ | | C | | | 110,094 | 110,094 | | |
| U ⁻ | | U | 111,079 | 111,079 | 111,079 | 111,079 | 111,079 | 111,079 |
| A ⁻ | | A | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 |
| G ⁻ | | G | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 |
| [B, PO-H ₂] ⁻ | | B, PO-H ₂ | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 |
| [B, PO ₃] ⁻ | | B, PO ₃ | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 |
| [B, PO ₄ H] ⁻ | | B, PO ₄ H | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 |
| [B, PO ₄ H ₂ , PO ₃] ⁻ | | B, PO ₄ H ₂ , PO ₃ | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 |
| [B, PO ₄ H ₂ , PO ₄ H] ⁻ | B, PO ₄ H ₂ , PO ₄ H | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | |
| [B(C) PO ₄] ⁻ | B ₁ ≠C | B(C) PO ₄ | 288,177 | 288,177 | | | 288,177 | 288,177 |
| c ₁ ⁻ | | 5'OH-B ₁ -PO ₃ | 328,202 | 328,202 | 288,177 | 288,177 | 312,202 | 312,202 |
| d ₁ ⁻ | | 5'OH-B ₁ -PO ₄ H | 346,217 | 346,217 | 306,192 | 306,192 | 330,217 | 330,217 |
| z ₁ ⁻ | | <u>U</u> -PO ₄ H ₂ ^{3'} | 305,161 | 305,161 | 305,161 | 305,161 | 305,161 | 305,161 |
| y ₁ ⁻ | | O- <u>U</u> -PO ₄ H ₂ ^{3'} | 323,176 | 323,176 | 323,176 | 323,176 | 323,176 | 323,176 |
| [w ₂ ,a ₃] ⁻ | | PO ₄ H-B ₃ | 288,178 | 312,203 | 328,203 | 312,203 | 328,203 | 288,178 |
| [w ₂ ,c ₃] ⁻ | | PO ₄ H-B ₃ -PO ₃ | 368,157 | 392,182 | 408,182 | 392,182 | 408,182 | 368,157 |
| [a ₂ -B ₂ H] ⁻ | | 5'OH-B ₁ -désoxy | 426,302 | 426,302 | 386,277 | 386,277 | 410,302 | 410,302 |
| x ₁ ⁻ | | <u>U</u> -PO ₄ H ₂ ^{3'} | 385,141 | 385,141 | 385,141 | 385,141 | 385,141 | 385,141 |
| w ₁ ⁻ | | O- <u>U</u> -PO ₄ H ₂ ^{3'} | 403,156 | 403,156 | 403,156 | 403,156 | 403,156 | 403,156 |
| c ₂ ⁻ | 5'OH-B ₁ B ₂ -PO ₃ | 641,412 | 617,387 | 601,387 | 617,387 | 601,387 | 641,412 | |
| d ₂ ⁻ | 5'OH-B ₁ B ₂ -PO ₄ H | 659,427 | 635,402 | 619,402 | 635,402 | 619,402 | 659,427 | |
| z ₂ ⁻ | B ₃ <u>U</u> -PO ₄ H ₂ ^{3'} | 594,346 | 618,371 | 634,371 | 618,371 | 634,371 | 594,346 | |
| y ₂ ⁻ | O-B ₃ <u>U</u> -PO ₄ H ₂ ^{3'} | 612,361 | 636,386 | 652,386 | 636,386 | 652,386 | 612,361 | |
| x ₂ ⁻ | PO ₃ -B ₃ <u>U</u> -PO ₄ H ₂ ^{3'} | 674,326 | 698,351 | 714,351 | 698,351 | 714,351 | 674,326 | |
| w ₂ ⁻ | PO ₄ H-B ₃ <u>U</u> -PO ₄ H ₂ ^{3'} | 692,341 | 716,366 | 732,366 | 716,366 | 732,366 | 692,341 | |
| [c ₃ -B ₃ H] ⁻ | 5'OH-B ₁ B ₂ -désoxy-PO ₃ | 819,495 | 795,470 | 779,471 | 795,470 | 779,471 | 819,495 | |
| a ₃ ⁻ | 5'OH-B ₁ B ₂ B ₃ | 850,617 | 850,617 | 850,617 | 850,617 | 850,617 | 850,617 | |
| b ₃ ⁻ | 5'OH-B ₁ B ₂ B ₃ -O | 868,632 | 868,632 | 868,632 | 868,632 | 868,632 | 868,632 | |
| [a ₃ -B ₃ H] ⁻ | 5'OH-B ₁ -B ₂ -désoxy | 739,512 | 715,487 | 699,487 | 715,487 | 699,487 | 739,512 | |
| c ₃ ⁻ | 5'OH-B ₁ B ₂ B ₃ -PO ₃ | 930,597 | 930,597 | 930,597 | 930,597 | 930,597 | 930,597 | |
| d ₃ ⁻ | 5'OH-B ₁ B ₂ B ₃ -PO ₄ H | 948,612 | 948,612 | 948,612 | 948,612 | 948,612 | 948,612 | |
| z ₃ ⁻ | B ₂ B ₃ <u>U</u> -PO ₄ H ₂ ^{3'} | 907,556 | 907,556 | 947,581 | 947,581 | 923,556 | 923,556 | |
| y ₃ ⁻ | O-B ₂ B ₃ <u>U</u> -PO ₄ H ₂ ^{3'} | 925,571 | 925,571 | 965,597 | 965,597 | 941,571 | 941,571 | |
| x ₃ ⁻ | PO ₃ -B ₂ B ₃ <u>U</u> -PO ₄ H ₂ ^{3'} | 987,536 | 987,536 | 1027,561 | 1027,561 | 1003,536 | 1003,536 | |
| w ₃ ⁻ | PO ₄ H-B ₂ B ₃ <u>U</u> -PO ₄ H ₂ ^{3'} | 1005,551 | 1005,551 | 1045,577 | 1045,577 | 1021,551 | 1021,551 | |
| [M-H-B _(G) H] ⁻ | chimère - base G | 1103,655 | 1103,655 | 1103,655 | 1103,655 | 1103,655 | 1103,655 | |
| [M-H-B _(A) H] ⁻ | chimère - base A | 1119,655 | 1119,655 | 1119,655 | 1119,655 | 1119,655 | 1119,655 | |
| [M-H-B _(C) H] ⁻ | chimère - base C | 1143,679 | 1143,679 | 1143,679 | 1143,679 | 1143,679 | 1143,679 | |
| b ₄ ⁻ | 3'-PO ₄ H ₂ | 5'OH-B ₁ B ₂ B ₃ <u>U</u> -OH | 1174,801 | 1174,801 | 1174,801 | 1174,801 | 1174,801 | 1174,801 |
| [M-H] ⁻ | | 5'OH-B ₁ B ₂ B ₃ <u>U</u> -PO ₄ H ₂ ^{3'} | 1254,781 | 1254,781 | 1254,781 | 1254,781 | 1254,781 | 1254,781 |

Tableau 1.6: Séquences et masses monoisotopiques des spectres CID, par MALDI-TOF/TOF, en mode négatif, de l'ion parent [M-H]⁻ des chimères GACU, GCAU, CAGU, CGAU, ACGU et AGCU. La base U, soulignée et en gras, est un ribonucléotide possédant un groupement 3'-PO₄H₂.

Les différents types d'ions fragments obtenus en mode CID des chimères B₄(A, C, G ou U) déprotonées sont comparables (tableaux 1.3 à 1.6). L'analyse, des ions fragments et des plus

abondants, permet de déterminer la séquence de ces chimères isomères. L'ion fragment a_4^- n'est observé que lorsque la chimère est $B_4(C)$. Cette dissociation est donc spécifique des chimères de 4-mer qui contiennent une base $B_4(C)$. Tout comme les chimères $B_4(A, C, G)$, les ions fragments les plus abondants sont les ions w_n^- et les ions issus de la perte de la base.

3.2.5. Le bilan des fragmentations en mode CID

Le schéma des fragmentations en mode CID des chimères de 4-mer est donc résumé dans le tableau 1.7.

| Fragments des chimères | | |
|--|--|--|
| Ion m/z | Spécifique | Séquence |
| PO ₃ ⁻ | | PO ₃ |
| PO ₄ H ₂ ⁻ | | PO ₄ H ₂ |
| C ⁻ | B ₁ =C ou B ₄ = <u>C</u> 3'-PO ₄ H ₂ | C |
| U ⁻ | B ₄ =U | U |
| T ⁻ | | T |
| A ⁻ | | A |
| G ⁻ | | G |
| [B, PO-H ₂] ⁻ | | B, PO-H ₂ |
| [B, PO ₃] ⁻ | | B, PO ₃ |
| [B, PO ₄ H] ⁻ | | B, PO ₄ H |
| [B, PO ₄ H ₂ , PO ₃] ⁻ | | B, PO ₄ H ₂ , PO ₃ |
| [B, PO ₄ H ₂ , PO ₄ H] ⁻ | | B, PO ₄ H ₂ , PO ₄ H |
| [B(C) PO ₄] ⁻ | B ₁ ou B ₄ =C | B(C) PO ₄ |
| c ₁ ⁻ | | ⁵ 'OH-B ₁ -PO ₃ |
| d ₁ ⁻ | | ⁵ 'OH-B ₁ -PO ₄ H |
| z ₁ ⁻ | | B₄ -PO ₄ H ₂ ^{3'} |
| y ₁ ⁻ | | O- B₄ -PO ₄ H ₂ ^{3'} |
| [w ₂ , a ₃] ⁻ | | PO ₄ H-B ₃ |
| [w ₂ , c ₃] ⁻ | | PO ₄ H-B ₃ -PO ₃ |
| [a ₂ -B ₂] ⁻ | B ₂ ≠ T | ⁵ 'OH-B ₁ -désoxy |
| x ₁ ⁻ | | B₄ -PO ₄ H ₂ ^{3'} |
| w ₁ ⁻ | | O- B₄ -PO ₄ H ₂ ^{3'} |
| c ₂ ⁻ | | ⁵ 'OH-B ₁ B ₂ -PO ₃ |
| d ₂ ⁻ | | ⁵ 'OH-B ₁ B ₂ -PO ₄ H |
| z ₂ ⁻ | | B ₃ B₄ -PO ₄ H ₂ ^{3'} |
| y ₂ ⁻ | | O-B ₃ B₄ -PO ₄ H ₂ ^{3'} |
| x ₂ ⁻ | | PO ₃ -B ₃ B₄ -PO ₄ H ₂ ^{3'} |
| w ₂ ⁻ | | PO ₄ H-B ₃ B₄ -PO ₄ H ₂ ^{3'} |
| [c ₃ -B ₃] ⁻ | B ₃ ≠ T et 3'-PO ₄ H ₂ | ⁵ 'OH-B ₁ B ₂ -désoxy-PO ₃ |
| a ₃ ⁻ | | ⁵ 'OH-B ₁ B ₂ B ₃ |
| b ₃ ⁻ | | ⁵ 'OH-B ₁ B ₂ B ₃ -O |
| [a ₃ -B ₃] ⁻ | | ⁵ 'OH-B ₁ -B ₂ -désoxy |
| c ₃ ⁻ | | ⁵ 'OH-B ₁ B ₂ B ₃ -PO ₃ |
| d ₃ ⁻ | | ⁵ 'OH-B ₁ B ₂ B ₃ -PO ₄ H |
| z ₃ ⁻ | | B ₂ B ₃ B₄ -PO ₄ H ₂ ^{3'} |
| y ₃ ⁻ | | O-B ₂ B ₃ B₄ -PO ₄ H ₂ ^{3'} |
| x ₃ ⁻ | | PO ₃ -B ₂ B ₃ B₄ -PO ₄ H ₂ ^{3'} |
| w ₃ ⁻ | | PO ₄ H-B ₂ B ₃ B₄ -PO ₄ H ₂ ^{3'} |
| [M-H-B _(G) H] ⁻ | | chimère - base G |
| [M-H-B _(A) H] ⁻ | | chimère - base A |
| [M-H-B _(C) H] ⁻ | | chimère - base C |
| a ₄ ⁻ | B ₄ =C et 3'-PO ₄ H ₂ | ⁵ 'OH-B ₁ B ₂ B ₃ C |
| b ₄ ⁻ | 3'-PO ₄ H ₂ | ⁵ 'OH-B ₁ B ₂ B ₃ B₄ -OH |
| [M-H] ⁻ | | ⁵ 'OH-B ₁ B ₂ B ₃ B₄ -PO ₄ H ₂ ^{3'} |

Tableau 1.7: Séquences des spectres CID, par MALDI-TOF/TOF, en mode négatif, de l'ion parent [M-H]⁻ des chimères 4-mer.

Les fragmentations en mode CID de toutes les chimères de 4-mer, constituées d'une base B₄ (A, C, G ou U) et des 3 autres bases différentes de celle de B₄, ont été étudiées et confirment les résultats obtenus précédemment (tableau 1.7). Toutefois, les chimères sans thymine sont plus difficilement détectables que les autres chimères (tableau 1.2). Par ailleurs, l'ion a₄⁻ représente la seule dissociation spécifique des chimères de 4-mer B₄(C).

L'étude des fragmentations en mode CID des chimères de 4-mer permettent donc de déterminer la séquence. L'abondance des ions donnent également une information sur la position des bases. Tout d'abord, l'étude de l'abondance des ions, notamment les ions les plus abondants (figure 1.2 et 1.3) et des ions correspondant à la perte de la base nucléique (figure 1.4) permet de déterminer la séquence. La figure 1.4 présente la variation de l'abondance des ions fragments [M-H-BH]⁻ des chimères : GCTA, GTCAA, TCGA, TGCAA, CGTA, et CTGA.

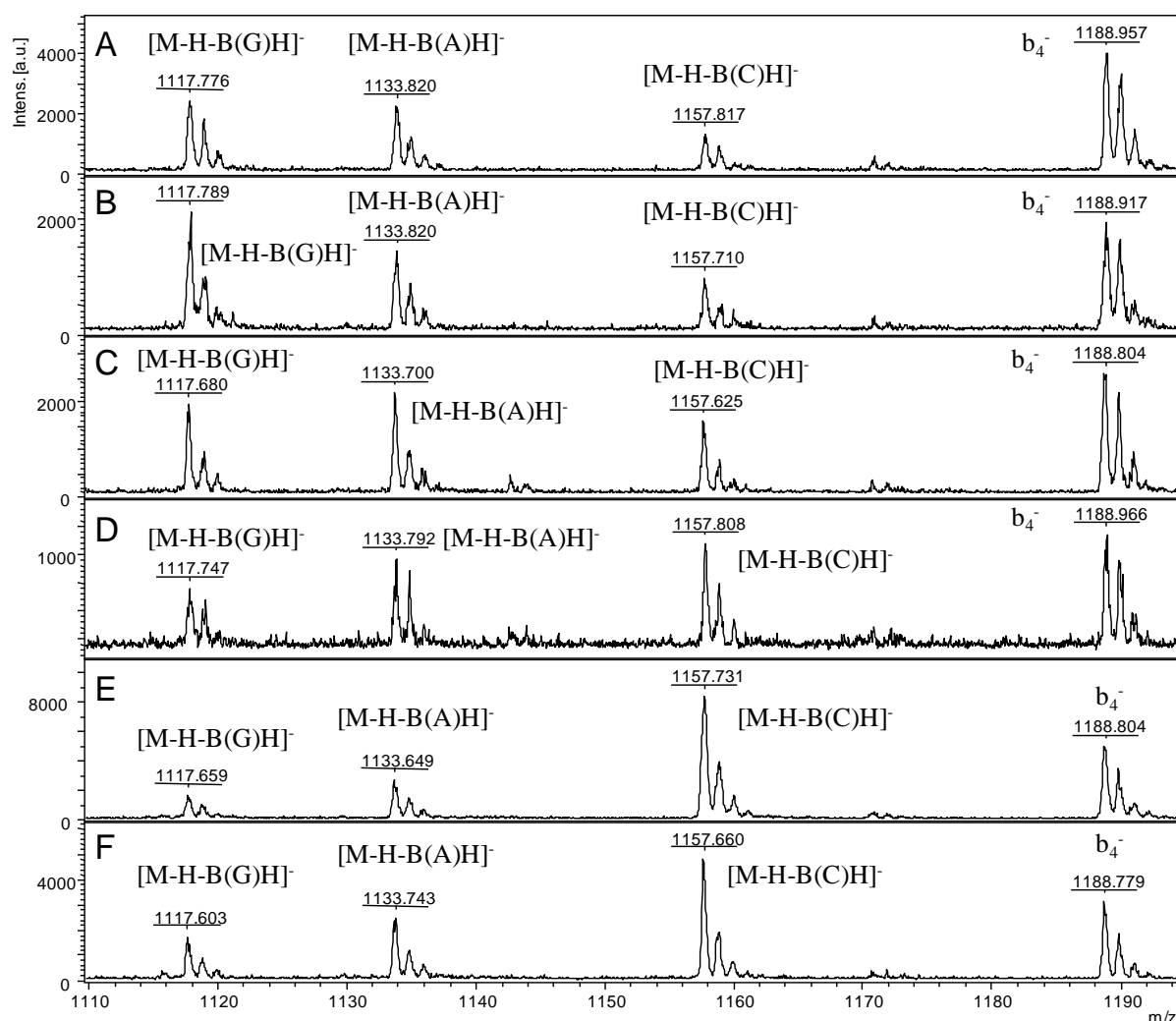


Figure 1.4: Spectres CID, de la région des 4-mer, par MALDI-TOF/TOF, en mode négatif, de l'ion parent [M-H]⁻ m/z 1268,808 des chimères : (A) GCTA; (B) GTCAA; (C) TCGA; (D) TGCAA; (E) CGTA et (F) CTGA. La base A, soulignée et en gras, est un ribonucléotide possédant un groupement 3'-PO₄H₂. Les ions fragments suivants sont présents: [M-H-GH]⁻, [M-H-AH]⁻, [M-H-CH]⁻ et b₄⁻.

Dans la région des 4-mer, ils montrent à m/z 1117,682 l'ion $[M-H-B(G)H]^-$, à m/z 1133,681 l'ion $[M-H-B(A)H]^-$, à m/z 1157,706 l'ion $[M-H-B(C)H]^-$ et à m/z 1188,828 l'ion b_4^- . De plus, l'ion fragment $[M-H-B(T)H]^-$, correspondant à la perte de la base T n'est pas observé. La thymine semble moins se fragmenter en phase gazeuse que les autres bases.

Les 6 chimères isomères sont composées de la base $B_4(A)$ qui entraîne la présence de l'ion fragment $[M-H-B_4(A)H]^-$ et des bases C, G et T couvrant les positions B_1 , B_2 et B_3 . Lorsque la base B_1 est un C, l'abondance de l'ion $[M-H-B_1(C)H]^-$ est supérieure à celles de la perte des autres bases (figure 1.4F et G). La perte de la cytosine en position B_1 semble être favorisée par rapport aux pertes des autres bases. Lorsque la base B_1 est un T, l'abondance de l'ion $[M-H-B_2H]^-$ est la plus faible (figure 1.4D et E). La thymine *semble* empêcher partiellement la perte de la base adjacente suivante. Par contre, lorsque la base B_1 est un G, l'abondance de l'ion varie et dépend de la séquence de la chimère (figure 1.4A et B). La perte de la guanine en position B_1 semble être favorisée. L'abondance des ions fragments $[M-H-BH]^-$, permettent donc d'apporter approximativement des informations relatives sur la séquence de la chimère et sur la stabilité des bases en phase gazeuse.

Enfin, les ions obtenus en mode CID des chimères ARN/ADN de 4-mer sont les ions w_n^- , et les ions issus de la perte de la base tandis que dans la littérature ce sont les ions w_n^- et d_n^- qui sont abondants. En effet, les ions fragments d_n^- obtenus par MALDI-TOF/TOF sont moins abondants que prévu. Ce phénomène peut provenir des conditions d'analyses par MALDI-TOF/TOF ou de la chimère ARN/ADN de 4-mer.

3.3. Les fragmentations en phase gazeuse de la chimère GCTA

Les fragmentations métastables de la chimère GCTA ont été étudiées par MALDI-TOF/TOF tandis que les fragmentations en mode CID ont été analysées par MALDI-TOF/TOF et ESI-ITMSⁿ afin de les comparer.

3.3.1. L'impact de la fluence du laser

La variation de la fluence du laser sur les fragmentations métastables et sur celles obtenues en mode CID a été étudiée sur la chimère GCT**A**. La fluence du laser pour la sélection de l'ion parent est de 64%.

3.3.1.1. Les fragmentations métastables

Tout d'abord, la variation de la fluence en MS/MS a été expérimentée pour montrer son rôle sur les fragmentations métastables (figure 1.5).

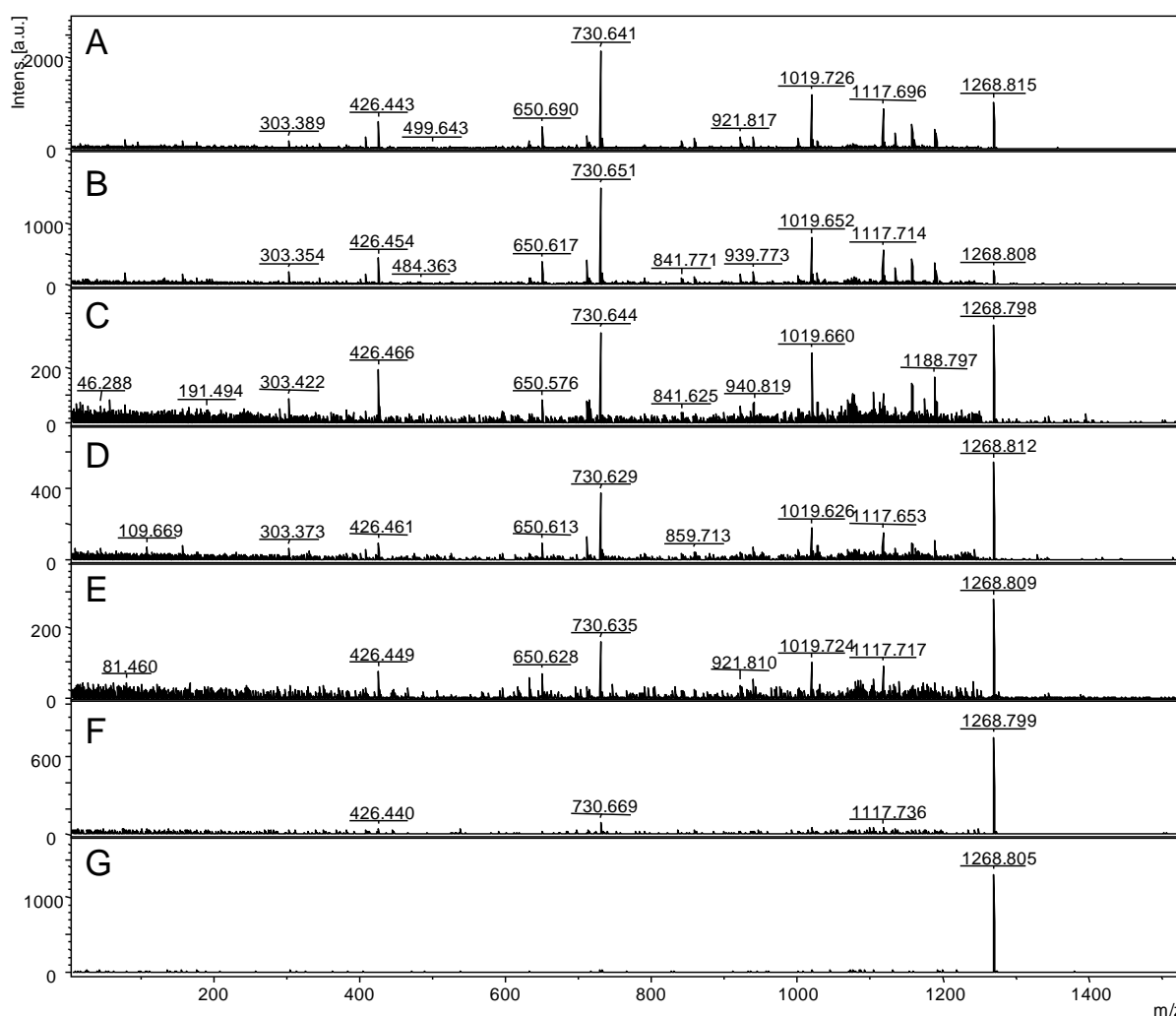


Figure 1.5: Spectres métastables, par MALDI-TOF-TOF, en mode négatif, de l'ion parent $[M-H]^-$ m/z 1268,808 de la chimère GCT**A** obtenus avec une fluence de : (A) 100% ; (B) 90% ; (C) 80% ; (D) 70% ; (E) 64% ; (F) 60% et (G) 50%. La base **A**, soulignée et en gras, est un ribonucléotide possédant un groupement 3'- PO_4H_2 .

Ils montrent l'ion parent $[M-H]^-$ à m/z 1268,808 de la chimère GCT**A** ainsi que ses ions fragments métastables. Lorsque la fluence du laser diminue l'abondance des ions fragments

diminue. Ils sont de moins en moins détectés jusqu'à leur disparition pour une fluence inférieure à 64% (figure 1.5F et 1.5G). A partir de 80% de fluence, l'intensité du pic de l'ion parent devient plus importante que celle de l'ion fragment le plus intense (figure 1.5C).

L'augmentation de la fluence du laser conduit donc à une augmentation du nombre et de l'abondance des ions fragments. Ces fragmentations sont dues à l'élargissement de l'énergie interne des espèces désorbées, s'étendant sur plusieurs eV, suivant une distribution gaussienne (Gabelica & Pauw 2005). Dans la première partie de la distribution, les ions se décomposent dans la source alors que dans la deuxième partie, ils se décomposent en dehors de la source et notamment dans la cellule *LIFT*. Cette très faible variation d'empreinte digitale est indépendante de la fluence du laser.

3.3.1.2. Les fragmentations en mode CID

La variation de la fluence en MS/MS a également été expérimentée sur les fragmentations en mode CID (figure 1.6).

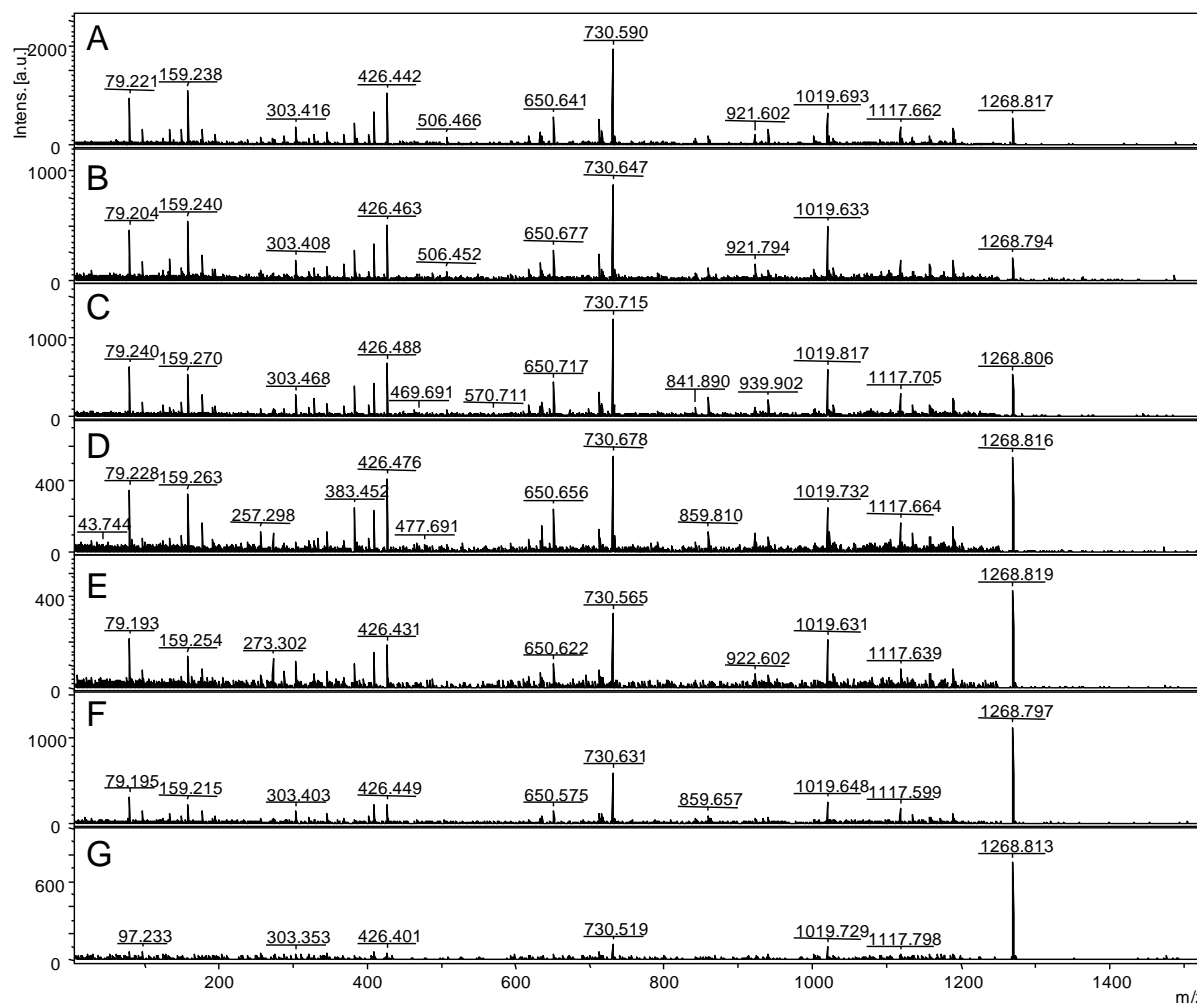


Figure 1.6: Spectres CID, par MALDI-TOF/TOF, en mode négatif, de l'ion parent $[M-H]^-$ m/z 1268,808 de la chimère GCTA obtenus avec une fluence de : (A) 100% ; (B) 90% ; (C) 80% ; (D) 70% ; (E) 64% ; (F) 60% et (G) 50%. La base A, soulignée et en gras, est un ribonucléotide possédant un groupement 3'- PO_4H_2 .

Ils montrent l'ion parent $[M-H]^-$ à m/z 1268,808 de la chimère GCTA ainsi que ses ions fragments obtenus en mode CID.

Les spectres des ions métastables (figure 1.5) et ceux obtenus en mode CID (figure 1.6) sont quasiment identiques. Les fragmentations métastables sont également observées en mode CID.

Le nombre et l'abondance des ions fragments diminuent également lorsque la fluence diminue. Pour les fragmentations métastables, l'abondance de l'ion parent devient plus importante que celle de l'ion fragment le plus intense pour une fluence de 80% (figure 1.5C), tandis qu'en mode CID la fluence est de 64% (figure 1.6E).

Les ions fragment de faible masse sont détectés en mode CID alors qu'ils ne sont pas analysés pour les fragmentations métastables. De plus, en mode CID, l'abondance des ions fragments est plus grande.

Les fragmentations en mode CID permettent d'analyser plus facilement les ions fragments. En mode CID, une fluence de 80 à 100% pourrait donc être utilisée (figure 1.6A, B et C). Cependant, la fluence de 100% a été utilisée car les ions fragments sont plus facilement détectables et reproductibles.

3.3.2. Les fragmentations en mode CID par MALDI-TOF/TOF

Les fragmentations en mode CID de la chimère GCTA_A par MALDI-TOF/TOF sont indiquées dans le tableau 1.3. Le spectre CID de la chimère est présenté dans la figure 1.2A et la figure 1.7 suivante montre la région des 2-mer.

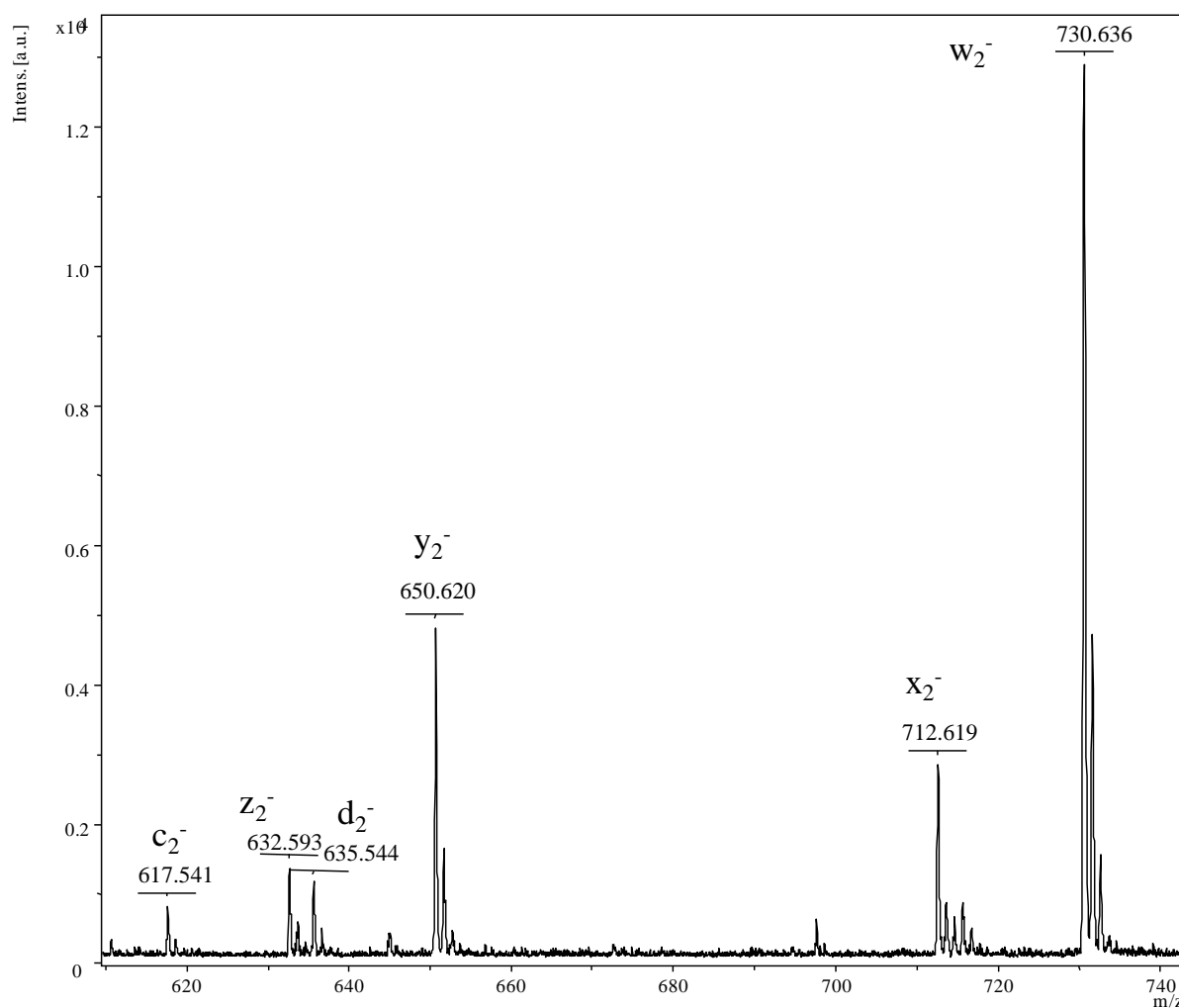


Figure 1.7: Spectre CID, de la région des 2-mer, par MALDI-TOF/TOF, en mode négatif, de l'ion

parent $[M-H]^-$ m/z 1268,808 de la chimère GCTA. La base A, est soulignée et en gras, est un ribonucléotide possédant un groupement 3'- PO_4H_2 .

La région des 2-mer du spectre CID montrent les ions fragments qui contiennent l'extrémité 5'-OH de l'ion parent : m/z 617,541 de l'ion c_2^- (m/z 617,387) et m/z 635,544 de l'ion d_2^- (m/z 635,402). Les ions fragments qui contiennent l'extrémité 3'- PO_4H_2 de l'ion parent sont également présents : m/z 632,593 de l'ion z_2^- (m/z 632,393), m/z 650,620 de l'ion y_2^- (m/z 650,413), m/z 712,619 de l'ion x_2^- (m/z 712,378) et l'ion m/z 730,636 de l'ion w_2^- (m/z 730,393).

Par ailleurs, dans cette région, l'ion w_2^- est le plus abondant et l'ion y_2^- est abondant tandis que l'ion d_2^- est moins abondant. Dans le spectre CID, les ions fragments les plus abondants sont w_1^- , w_2^- , w_3^- , y_2^- et $[M-B_1(G)H]^-$.

3.3.3. Les fragmentations en mode CID par ESI-ITMSⁿ

La chimère GCTA a été étudiée à l'aide de processus de décompositions métastables ou de plus hautes énergies de collision par MALDI-TOF/TOF mais elle a été également analysée par processus, de plus basses énergies de collision par ESI-ITMSⁿ. De plus, ces dissociations se produisent sous un plus faible déplacement cinétique (~100 ns) que celui du *LIFT* (de 10 à 100 μ s). L'avantage des expériences avec un piège ionique réside en la possibilité de produire des processus consécutifs après sélection de chacun des ions précurseurs qui sont également des ions fragments.

3.3.3.1. Le spectre de masse de la chimère

La figure 1.8 présente le spectre de masse par ESI-IT de la chimère GCTA.

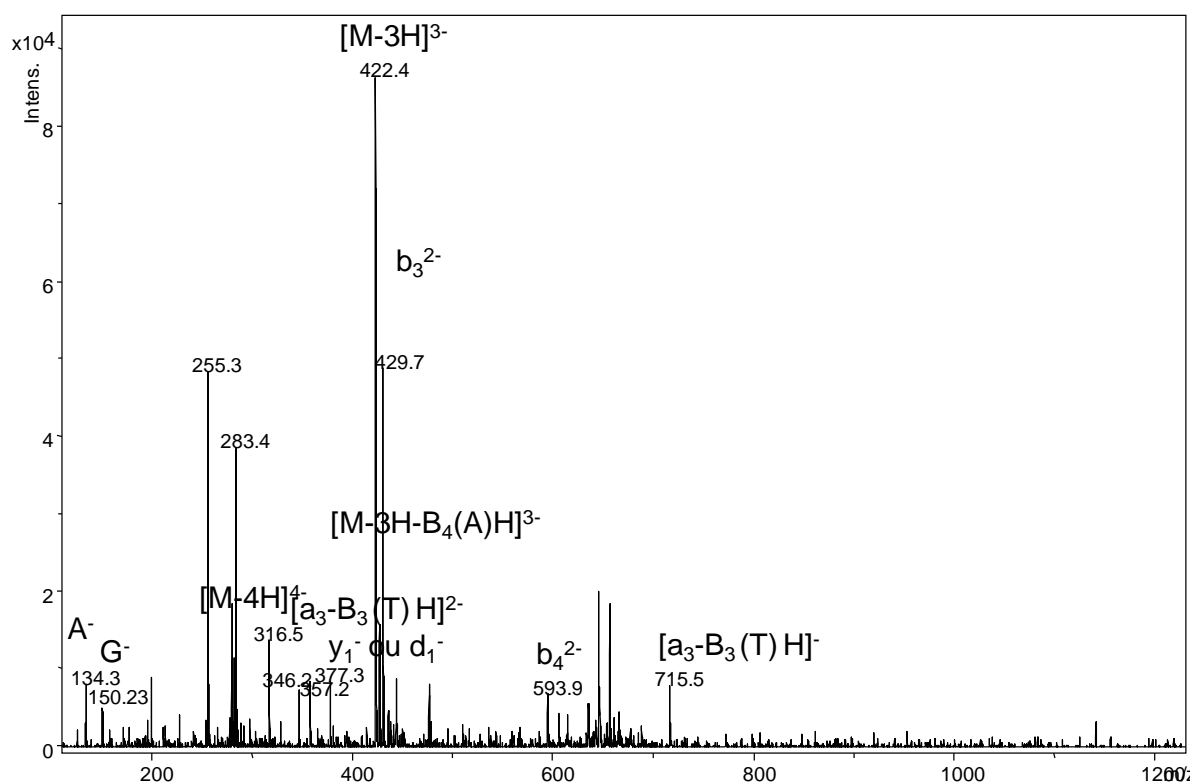


Figure 1.8: Spectre de masse, par ESI-ITMS, en mode négatif, de la chimère GCTA ($M=1269,816$ Da). La base A, soulignée et en gras, est un ribonucléotide possédant un groupement $3'-PO_4H_2$.

Le spectre de masse montre deux ions à m/z 422,4 et à m/z 316,5 qui correspondent respectivement à l'ion parent sous sa forme $[M-3H]^{3-}$ et $[M-4H]^{4-}$ de la chimère ($M=1269,816$ Da). L'ion $[M-3H]^{3-}$ étant le plus abondant, sa formation semble favorisée devant celle de l'ion $[M-4H]^{4-}$ bien que la chimère contient 4 sites déprotonables : trois phosphodiester et un 3'-phosphate. L'analyse MS^n pourrait permettre de déterminer éventuellement les positions des charges de l'ion trichargé.

Il montre également des ions fragments provenant de la dissociation de la chimère multidéprotonée : à m/z 593,9 l'ion b_4^{2-} et à m/z 429,7 l'ion b_3^{2-} . Ces deux ions doublement chargés, sont constitués d'un 3'-OH. Les 2 charges de l'ion b_3^{2-} sont présentes sur les deux phosphodiester entre les bases $B_1(G)-B_2(C)-B_3(T)$. Par contre, l'ion b_4^{2-} également doublement chargé, contient trois phosphodiester entre les bases $B_1(G)-B_2(C)-B_3(T)-B_4(A)$. Ces charges doivent être probablement les plus éloignées possibles à cause de l'effet d'acidité en phase gazeuse.

Les ions à m/z 715,5 et à m/z 357,2 correspondent respectivement à la forme simplement et doublement chargée de l'ion fragment $[a_3-B_3(T)H]^-$ (m/z 715,487). L'ion est monochargé

alors qu'il est constitué de 2 groupements phosphodiester entre B₁(G)-B₂(C)-B₃(T). La perte de la thymine est totalement inattendue, elle pourrait être due à la présence du 3'-phosphate qui pourrait induire sa dissociation. L'analyse de la localisation des charges de la chimère pourrait permettre d'expliquer ce phénomène.

De plus, l'ion issu de la perte de la base A de la chimère est observée à m/z 377,3 de l'ion fragment [M-3H-B₄(A)H]³⁻ (M=1134,689 Da) tandis qu'à m/z 134,3 l'anion de la base A (M=135,127 Da) perdue est également présent.

Par ailleurs, l'ion fragment à m/z 150,3 correspond à l'anion de la base G⁻ (151,126 Da). L'ion fragment à m/z 346,2 peut correspondre à l'ion d₁⁻ (m/z 346,217) ou à l'ion y₁⁻ (m/z 346,216). L'analyse par spectrométrie de masse ne permet pas d'attribuer cet ion à m/z 346,2 mais l'analyse des ions fragments de cet ion pourrait permettre sa détermination. Enfin, les autres ions n'ont pu être identifiés.

3.3.3.2. Le spectre CID de l'ion m/z 422,4

Le spectre CID de l'ion parent [M-3H]³⁻, m/z 422,4, de la chimère GCTA a été obtenu à l'aide d'activation collisionnelle en mode résonant dans le piège ionique (figure 1.9).

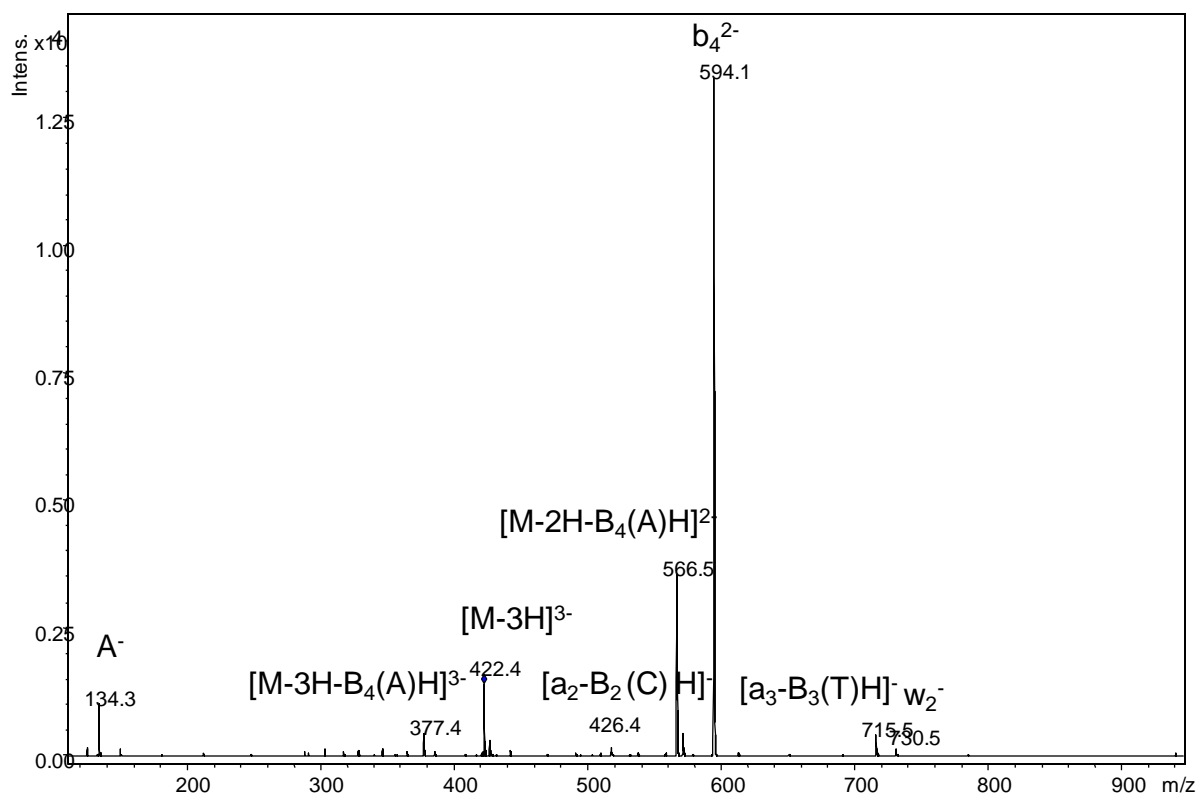


Figure 1.9: Spectre CID, par ESI-ITMS², en mode négatif, de l'ion parent $[M-3H]^{3-}$, m/z 422,4, de la chimère GCTA.

Il montre à m/z 422,4 l'ion parent $[M-3H]^{3-}$ de la chimère GCTA (M=1269,816 Da).

Il montre également des ions fragments qui sont aussi présents dans le spectre de masse de la chimère. L'ion fragment, le plus abondant, à m/z 594,1 correspond à l'ion b_4^{2-} de séquence GCTA-OH^{3'}. Il est caractéristique de la rupture de la chaîne nucléotidique au niveau du groupement 3'-PO₄H₂. Les ions fragments à m/z 566,5 et à m/z 377,4 correspondent respectivement à la forme doublement et triplement chargée de l'ion issu de la perte de l'adénine de la chimère (M=1134,689 Da) tandis que l'ion à m/z 134,3 correspond à l'anion A⁻ (M=135,127 Da) éliminée. L'ion doublement chargée est le second ion le plus abondant. L'ion fragment à m/z 715,5 de l'ion $[a_3-B_3(T)H]^-$ est également présent dans le spectre de masse alors que sa forme doublement chargée est observée que dans le spectre de masse.

Enfin, les deux ions fragments suivants, peu abondants, ne sont pas présents dans le spectre de masse. Les ions fragments à m/z 730,5 de l'ion w_2^- (m/z 730,393) et à m/z 426,4 de l'ion fragment $[a_2-B_2(C)H]^-$ (m/z 426,302) sont issus de la même dissociation.

Le spectre CID de la chimère montre surtout l'abondance de l'ion fragment b_4^{2-} qui est caractéristique de la présence du 3'-phosphate. Ces ions fragments confirment la structure de la chimère.

3.3.3.3. Le spectre CID en MS³ de l'ion m/z 594,1 du spectre CID de l'ion m/z 422,4

Le spectre CID en expérience MS³ séquentielle a été obtenu à partir de l'ion précurseur b_4^{2-} , m/z 594,1, présent dans le spectre CID issu de l'ion parent $[M-3H]^{3-}$, m/z 422,4 (figure 1.10).

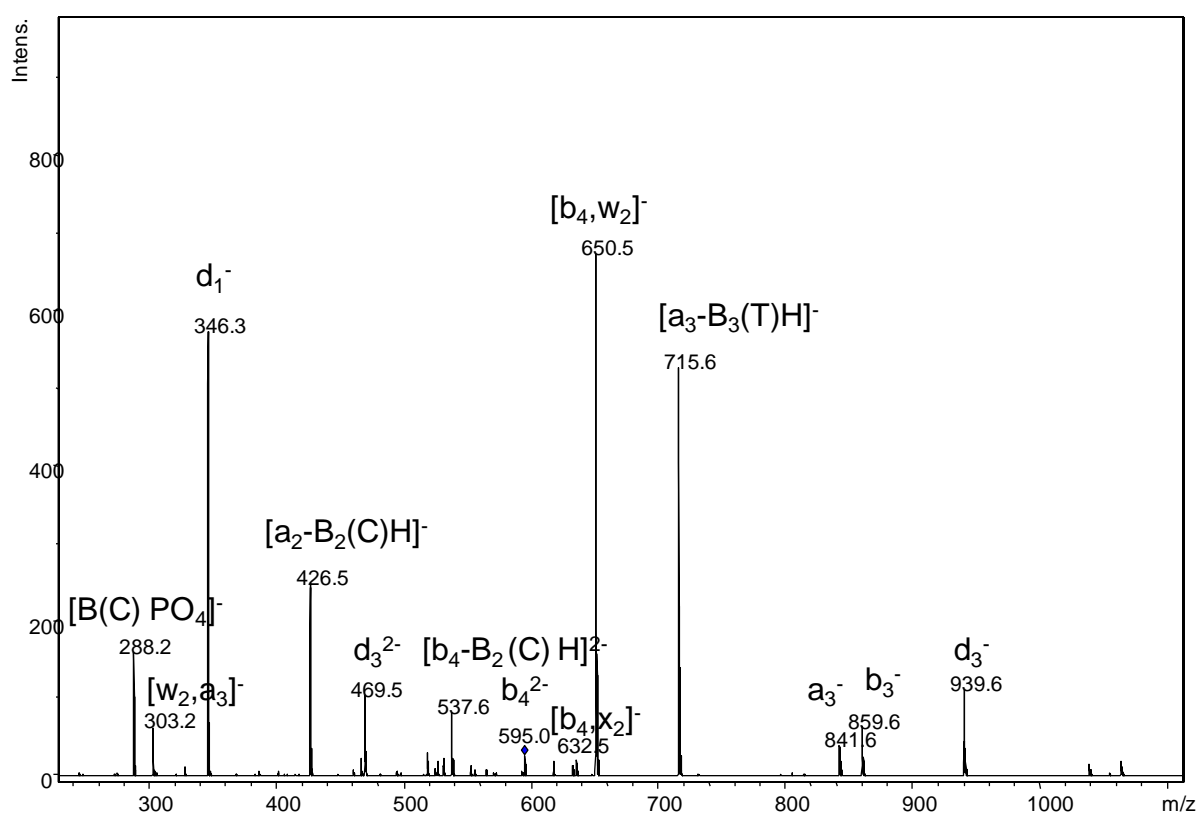


Figure 1.10: Spectre CID, par ESI-ITMS³, en mode négatif, de l'ion précurseur b_4^{2-} , m/z 594,1 du spectre CID de l'ion parent $[M-3H]^{3-}$, m/z 422,4.

Il montre à m/z 595,0 l'ion précurseur b_4^{2-} de séquence GCTA-OH-3'. Le rapport m/z de l'ion fragment est décalé de 0,9 par rapport à l'ion précurseur (m/z 594,1) correspondant probablement à son isotope à cause de la dissociation totale de l'ion précurseur.

Dans cette expérience séquentielle MS³, les types d'ions fragments constitués de l'extrémité 5'-OH (a_n^- , b_n^- , d_n^- et $[a-B_nH]^-$) de l'ion précurseur b_4^{2-} (GCTA-OH-3'), sont donc identiques en terme d'annotation et de séquence à ceux des ions fragments de la chimère (GCTA). Les autres ions fragments constitués de l'extrémité 3'-OH de l'ion précurseur b_4^{2-} sont différents

de ceux issus de la chimère qui sont constitués de l'extrémité 3'-PO₄H₂. Par conséquent, les dissociations obtenues, à partir de l'ion précurseur, sont rajoutées après le crochet. La numérotation du fragment perdu est notée en fonction de celle de la séquence de l'ion précurseur.

Les ions fragments suivants sont constitués de l'extrémité 5'-OH de l'ion précurseur. Les ions à m/z 939,6 et à m/z 469,5 correspondent respectivement à la forme simplement et doublement chargée de l'ion fragment d₃⁻ (m/z 939,589). L'ion fragment à m/z 346,3 correspond à l'ion d₁⁻ (m/z 346,217). Le rapport m/z de cet ion est identique à celui de l'ion au même rapport observé du spectre de masse de la chimère qui n'avait pas pu être identifié (figure 1.8). De plus, l'ion d₁⁻ est le deuxième ion le plus abondant du spectre CID.

L'ion fragment à m/z 859,6 de l'ion b₃⁻ (m/z 859,619) est également présent dans le spectre de masse de l'ion précurseur sous sa forme doublement chargée. Cependant, l'ion simplement chargée est issu de la dissociation de l'ion b₄²⁻ alors que l'ion doublement chargé est issu de celle de la chimère sous sa forme [M-3H]³⁻ ou [M-4H]⁴⁻.

L'ion fragment à m/z 841,6 correspond à l'ion a₃⁻ (m/z 841,604). Les ions à m/z 715,6 et à m/z 426,5 correspondent respectivement aux ions fragments [a₃-B₃(T)H]⁻ et [a₂-B₂(C)H]⁻ qui sont également présents dans le spectre CID. La perte de la thymine est très abondante bien que cette dissociation ne soit pas attendue.

De plus, les deux ions fragments suivants sont constitués de l'extrémité 3' de l'ion précurseur. L'ion à m/z 650,6 est le plus abondant et il correspond à l'ion fragment [b₄,w₂]⁻ (m/z 650,413). L'ion à m/z 632,5 correspond à l'ion fragment [b₄,x₂]⁻ (m/z 632,398).

L'ion fragment à m/z 537,6 correspond à l'ion issu de la perte de cytosine [b₄-B₂(C)H]²⁻ (m/z 1077,726).

Enfin, les deux ions fragments suivants sont issus de dissociations internes qui sont également observées par MALDI : à m/z 303,2 l'ion fragment [w₂,a₃]⁻ (m/z 303,190) et à m/z 288,2 l'ion fragment [B(C) PO₄]⁻ (m/z 288,177) de la perte du nucléotide C.

Ce spectre CID présente donc un grand nombre d'ions fragments. Tous les ions fragments confirment la structure de l'ion précurseur. Les ions fragments abondants sont notamment les ions w_n⁻, d_n⁻, [a_n-B_nH]⁻.

3.3.3.4. Le spectre CID en MS⁴ de l'ion m/z 650,5 du spectre CID en MS³

Le spectre CID en expérience MS⁴ séquentielle a été obtenu à partir de l'ion précurseur $[b_4, w_2]^-$, m/z 650,5, du spectre CID en expérience MS³ séquentielle, de l'ion précurseur b_4^{2-} , m/z 594,1, présent dans le spectre CID issu de l'ion parent $[M-3H]^{3-}$, m/z 422,4 (figure 1.11).

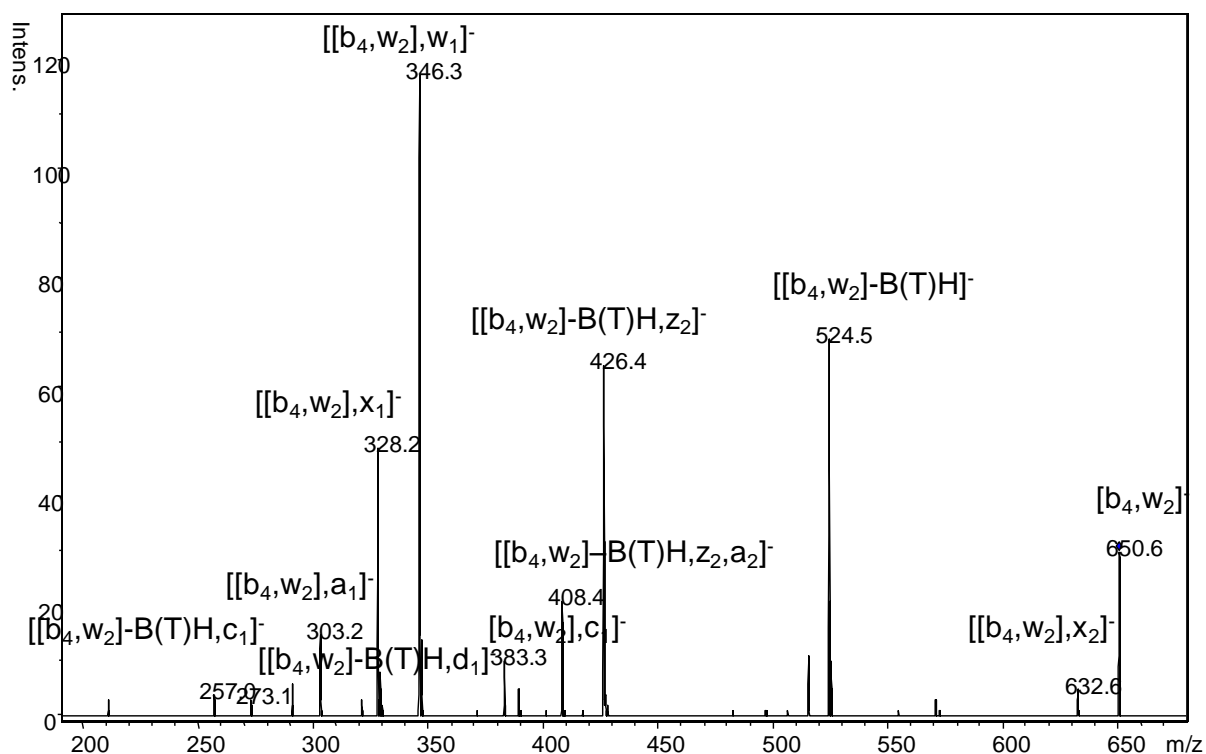


Figure 1.11: Spectre CID, par ESI-ITMS⁴ en mode négatif, de l'ion précurseur $[b_4, w_2]^-$, m/z 650,6, du spectre CID en MS³ de l'ion précurseur b_4^{2-} , m/z 594,1, du spectre CID issu de l'ion parent $[M-3H]^{3-}$, m/z 422,4.

Il montre l'ion précurseur $[b_4, w_2]^-$ (m/z 650,413) à m/z 650,6 de séquence 5'-PO₄H₂-TA-OH-3'.

Les ions fragments suivants de type x_n^- et w_n^- sont constitués de l'extrémité 3' de l'ion précurseur. L'ion fragment à m/z 346,3 de l'ion $[b_4, w_2], w_1^-$ (m/z 346,216) est le plus abondant. Les ions à m/z 632,6 et à m/z 328,2 correspondent respectivement aux ions fragments $[b_4, w_2], x_2^-$ (m/z 632, 393) et $[b_4, w_2], x_1^-$ (m/z 328,201). L'ion fragment $[b_4, w_2], x_1^-$ est abondant tandis que l'ion fragment $[b_4, w_2], x_2^-$ est moins abondant. La dissociation au niveau du phosphodiester entre B₁(T)-B₂(A) est plus importante qu'au niveau du 5'-phosphate.

Les ions fragments suivants de type a_n^- et c_n^- sont constitués de l'extrémité 5' de l'ion précurseur. Les ions à m/z 383,3 et à m/z 303,2 correspondent respectivement aux ions fragments $[[b_4, w_2], c_1]^-$ (m/z 383,167) et $[[b_4, w_2], a_1]^-$ (303,190). Ces ions sont moins abondants que ceux possédant l'extrémité 3' de l'ion précurseur.

Dans ce spectre CID, beaucoup d'ions fragments sont issus de la perte de la thymine de l'ion précurseur alors que ceux issus de la perte de l'adénine ne sont pas observés. La charge au niveau du phosphodiester de l'ion précurseur favorise probablement la perte de la thymine. L'ion à m/z 524,5 est, le deuxième ion le plus abondant, de l'ion fragment $[[b_4, w_2]-B(T)H]^-$ (m/z 524,300) de séquence 5'-PO₄H₂-désoxyribo-A-OH-3'. De plus, les ions fragments suivants sont issus d'une part, de la perte de la thymine et d'autre part, de la rupture de la chaîne au niveau des liaisons du phosphodiester et/ou du 5'-phosphate. Les ions fragments à m/z 273,1 et à m/z 257,0 correspondent respectivement aux ions fragments $[[b_4, w_2]-B(T)H, d_1]^-$ (m/z 273,051) et $[[b_4, w_2]-B(T)H, c_1]^-$ (m/z 257,052) et sont constitués de l'extrémité 5'-PO₄H₂ de l'ion précurseur tandis que l'ion fragment, abondant, à m/z 426,4, de à l'ion $[[b_4, w_2]-B(T)H, z_2]^-$ (m/z 426,305) est constitué de l'extrémité 3'-OH de l'ion précurseur. L'ion fragment à m/z 408,4 correspond à l'ion $[[b_4, w_2]-B(T)H, z_2, a_2]^-$ (m/z 408,304). Il est issu de la perte de la thymine et d'une double rupture de la chaîne.

L'analyse des ions fragments de ce spectre CID a permis de confirmer la séquence de l'ion précurseur $[b_4, w_2]^-$ (m/z 650,413) de séquence 5'-PO₄H₂-TA-OH-3'. Le spectre CID montre l'abondance des ions issus de la perte de la thymine.

Le spectre CID en MS⁵ de l'ion m/z 524,5 du spectre CID en MS⁴

Le spectre CID en expérience MS⁵ séquentielle a été obtenu à partir de l'ion précurseur $[[b_4, w_2]-B(T)H]^-$, m/z 524,5, du spectre CID en expérience MS⁴ séquentielle de l'ion précurseur $[b_4, w_2]^-$, m/z 650,5, du spectre CID en expérience MS³ séquentielle, de l'ion précurseur b_4^{2-} , m/z 594,1, présent dans le spectre CID issu de l'ion parent $[M-3H]^{3-}$ à m/z 422,4 (figure 1.12).

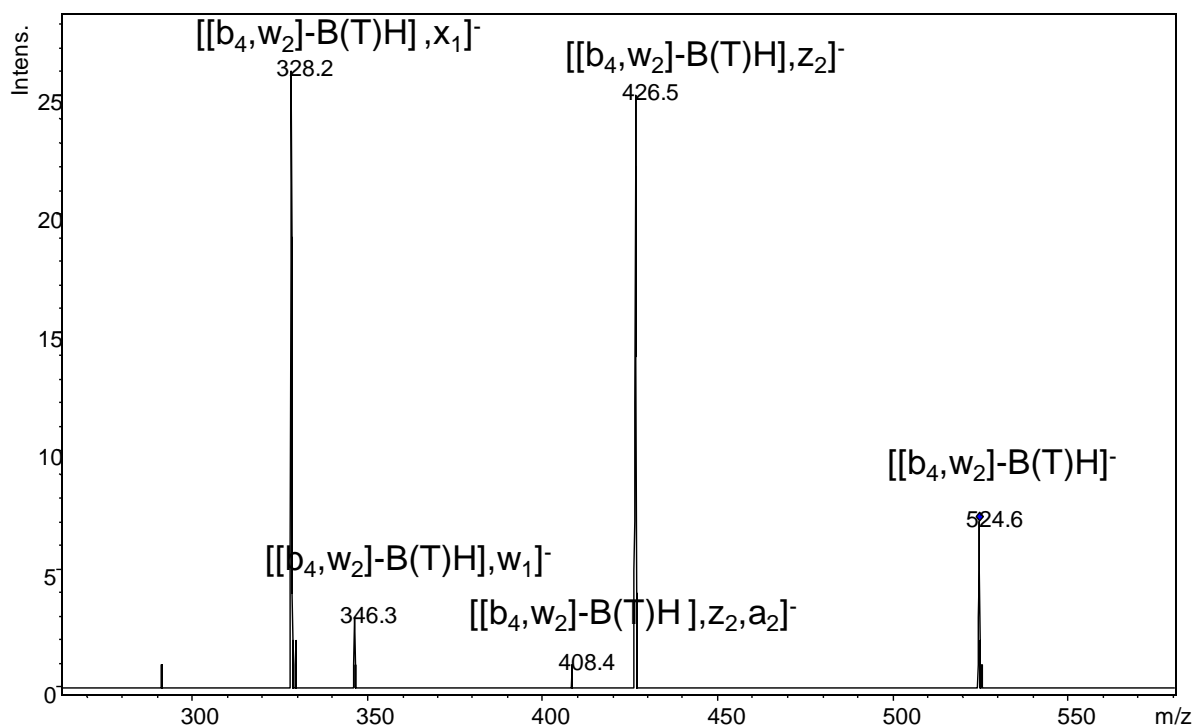


Figure 1.12: Spectre CID, par ESI-ITMS⁵, en mode négatif, de l'ion précurseur $[[b_4, w_2], B(T)H]^-$, m/z 524,6, du spectre CID en MS⁴ de l'ion précurseur $[b_4, w_2]^-$, m/z 650,5, du spectre CID en MS³, de l'ion précurseur b_4^{2-} , m/z 594,1, du spectre CID issu de l'ion parent $[M-3H]^{3-}$, m/z 422,4.

Il montre l'ion précurseur $[[b_4, w_2]-B(T)H]^-$ (m/z 524,300) à m/z 524,6 de séquence 5'-PO₄H₂-ribo-A-OH-3'.

Les ions fragments à m/z 426,5, à m/z 346,3 et à m/z 328,3 correspondent respectivement aux ions $[[[b_4, w_2]^-B(T)H], z_2]^-$ (m/z 426,305) $[[[b_4, w_2]^-B(T)H], w_1]^-$ (m/z 346,216) et $[[[b_4, w_2]^-B(T)H], x_1]^-$ (m/z 328,201) et sont constitués de l'extrémité 3'-OH de l'ion précurseur. Les ions $[[[b_4, w_2]^-B(T)H], z_2]^-$ et $[[[b_4, w_2]^-B(T)H], x_1]^-$ sont les plus abondants. L'ion fragment à m/z 408,4 correspond à l'ion $[[[b_4, w_2]^-B(T)H], z_2, a_2]^-$ (m/z 408,304). Il est issu d'une double rupture de la chaîne.

Tous ces ions fragments sont également présents dans le spectre MS⁴ et ils confirment la structure de l'ion précurseur $[[b_4, w_2]-B(T)H]^-$ ainsi que sa dissociation.

3.3.3.5. Le spectre CID en MS⁴ de l'ion m/z 715,5 du spectre CID en MS³

Le spectre CID en expérience MS⁴ séquentielle a été obtenu à partir de l'ion précurseur [a₃-B₃(T)H]⁻, m/z 715,5, du spectre CID en expérience MS³ séquentielle, de l'ion précurseur b₄²⁻, m/z 594,1, présent dans le spectre CID issu de l'ion parent [M-3H]³⁻, m/z 422,4 (figure 1.13).

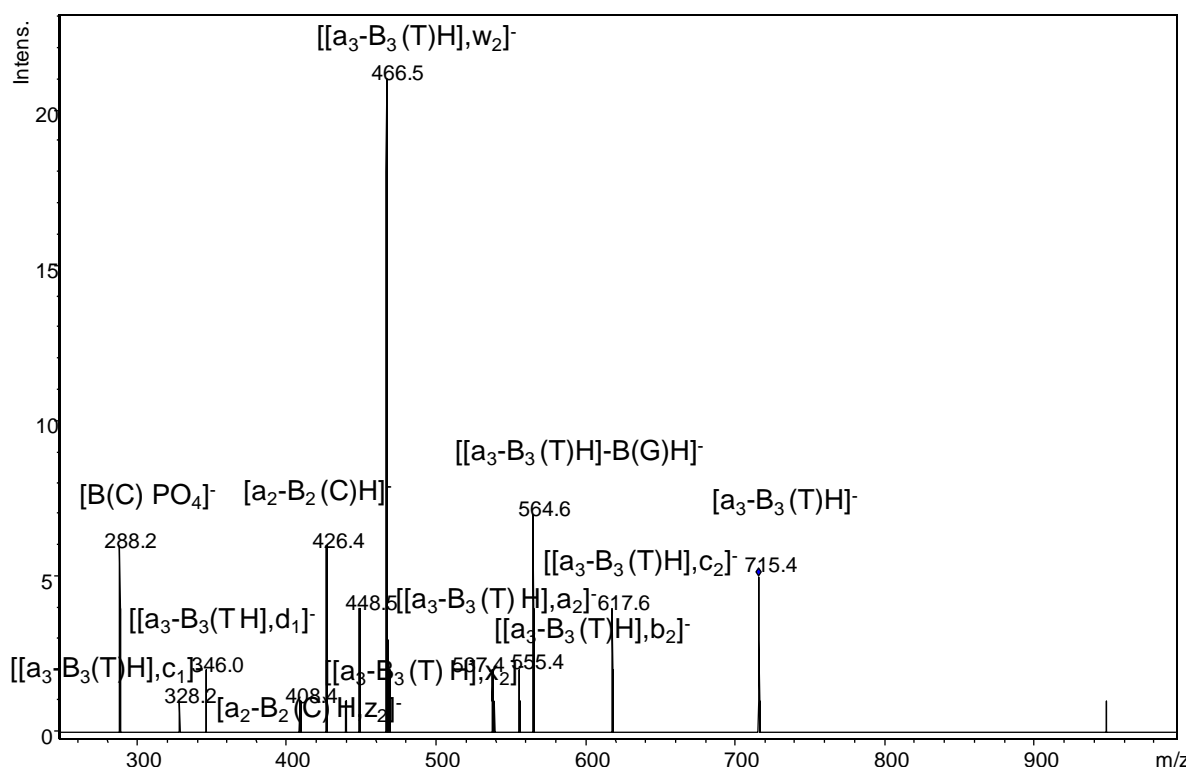


Figure 1.13: Spectre CID, par ESI-ITMS⁴, en mode négatif, de l'ion précurseur [a₃-B₃(T)H]⁻, m/z 715,5, du spectre CID en MS³, de l'ion précurseur b₄²⁻, m/z 594,1, du spectre CID issu de l'ion parent [M-3H]³⁻, m/z 422,4.

Il montre l'ion précurseur [a₃-B₃(T)H]⁻ à m/z 715,4 de séquence 5'-OH-GC-désoxyribo.

Les ions fragments suivants correspondent à la rupture de la chaîne nucléotidique. Tout d'abord, les ions constitués de l'extrémité 5'-OH de l'ion précurseur à m/z 617,6, à m/z 555,4 et à m/z 537,4 correspondent respectivement aux ions [[a₃-B₃(T)H],c₂]⁻ (m/z 617,387), [[a₃-B₃(T)H],b₂]⁻ (m/z 555,422) et [[a₃-B₃(T)H],a₂]⁻ (m/z 537,407). Les ions fragments à m/z 346,0 et 328,2 correspondent respectivement aux ions [[a₃-B₃(T)H],d₁]⁻ (m/z 346,217) et [[a₃-B₃(T)H],c₁]⁻ (m/z 328,207).

Ensuite, les ions suivants sont constitués de l'extrémité 3' de l'ion précurseur. L'ion le plus abondant à m/z 466,5 correspond à l'ion fragment [[a₃-B₃(T)H],w₂]⁻ (m/z 466,257) et celui à

m/z 448,5 à l'ion fragment $[[a_3-B_3(T)H],x_2]^-$ (m/z 448,241). Enfin, l'ion fragment à m/z 288,2 correspond à l'ion $[B(C) PO_4]^-$ (m/z 288,177) de la perte du nucléotide de la base C.

Les ions fragments suivants sont des ions issus de la perte de base. L'ion fragment à m/z 564,5 correspond à l'ion $[[a_3-B_3(T)H]-B(G)H]^-$ (m/z 564,274) de la perte de la base G. Les ions fragments à m/z 426,5 et à m/z 408,4 correspondent respectivement à l'ion $[a_2-B_2(C)H]^-$ (m/z 426,302) de la perte de la base C et l'ion $[a_2-B_2(C)H,z_2]^-$ (m/z 408,287) de la perte de la base C et de la rupture de la chaîne.

Tous ces ions fragments confirment la structure de l'ion précurseur $[a_3-B_3(T)H]^-$.

3.3.3.6. Le spectre en CID MS⁴ de l'ion m/z 346,3 du spectre CID MS³

Le spectre CID en expérience MS⁴ séquentielle a été obtenu à partir de l'ion précurseur d_1^- , m/z 346,3, du spectre CID en expérience MS³ séquentielle, de l'ion précurseur b_4^{2-} , m/z 594,1, présent dans le spectre CID issu de l'ion parent $[M-3H]^{3-}$, m/z 422,4 (figure 1.14).

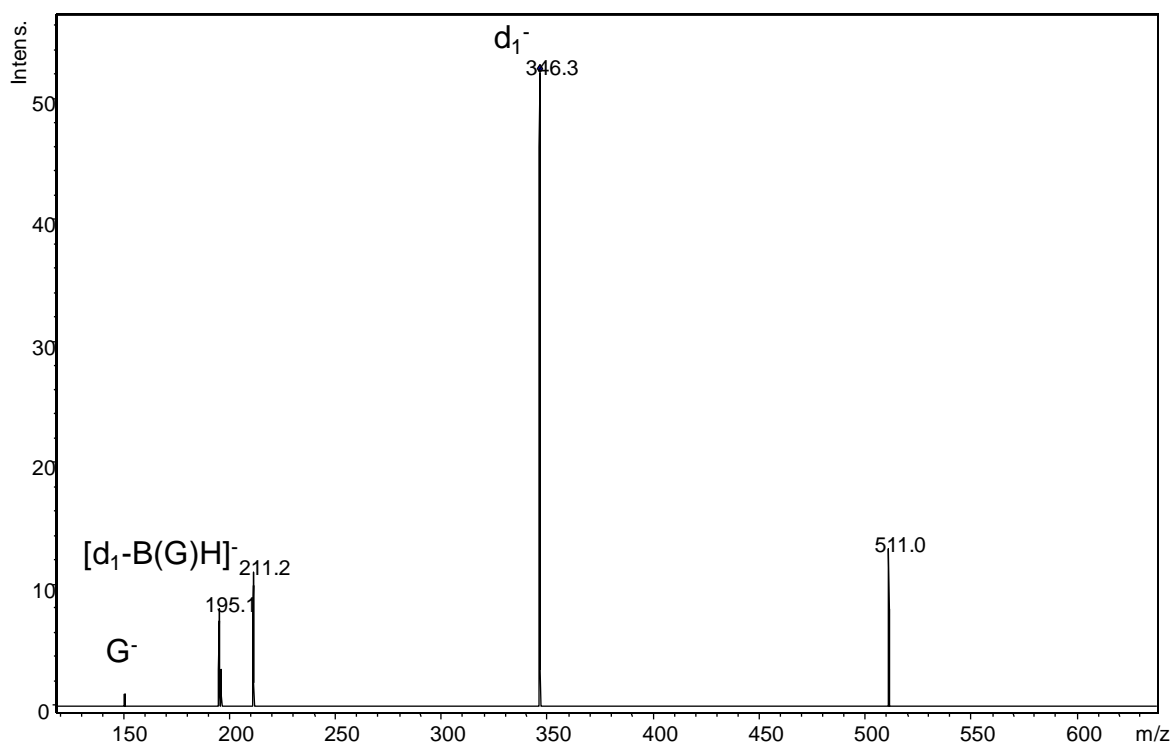


Figure 1.14: Spectre CID, par ESI-ITMS⁴, en mode négatif, de l'ion précurseur d_1^- , m/z 346,3, du spectre CID en MS³, de l'ion précurseur b_4^{2-} , m/z 594,1, du spectre CID issu de l'ion parent $[M-3H]^{3-}$, m/z 422,4.

Il montre l'ion précurseur d_1^- (m/z 346,217) à m/z 346,3 de séquence 5'-OH-G-PO₄H₂-3'.

L'ion fragment à m/z 195,1 est l'ion $[d_1-B(G)H]^-$ (m/z 195,087) de la perte de la base G tandis que l'ion fragment à m/z 150,0 de l'ion G⁻ (m/z 150,118) de la base perdue.

Les deux ions à m/z 511 et à m/z 211,2 n'ont pas pu être interprétés.

Les ions fragments confirment la structure de l'ion précurseur d_1^- .

3.3.4. Le bilan des fragmentations en mode CID

Les fragmentations en mode CID par spectrométrie de masse en tandem de la chimère GCTA sont résumées dans le tableau 1.8.

| Maldi | ESI | Séquence |
|--|--|--|
| PO_3^- | | PO_3 |
| PO_4H_2^- | | PO_4H_2 |
| T^- | | T |
| A^- | A^- | A |
| G^- | | G |
| $[\text{B}, \text{PO}-\text{H}_2]^-$ | | B, PO-H ₂ |
| $[\text{B}, \text{PO}_3]^-$ | | B, PO ₃ |
| $[\text{B}, \text{PO}_4\text{H}]^-$ | | B, PO ₄ H |
| $[\text{B}, \text{PO}_4\text{H}_2, \text{PO}_3]^-$ | | B, PO ₄ H ₂ , PO ₃ |
| $[\text{B}, \text{PO}_4\text{H}_2, \text{PO}_4\text{H}]^-$ | | B, PO ₄ H ₂ , PO ₄ H |
| $[\text{B}(\text{C}) \text{PO}_4]^-$ | | B(C) PO ₄ |
| c_1^- | | ^{5'} OH-C-PO ₃ |
| d_1^- | | ^{5'} OH-C-PO ₄ H |
| z_1^- | | <u>A</u> -PO ₄ H ₂ ^{3'} |
| y_1^- | | O- <u>A</u> -PO ₄ H ₂ ^{3'} |
| $[\text{w}_2, \text{a}_3]^-$ | | PO ₄ H-T |
| $[\text{w}_2, \text{c}_3]^-$ | | PO ₄ H-T-PO ₃ |
| $[\text{a}_2-\text{B}_2(\text{C})\text{H}]^-$ | $[\text{a}_2-\text{B}_2(\text{C})\text{H}]^-$ | ^{5'} OH-C-désoxy |
| x_1^- | | <u>A</u> -PO ₄ H ₂ ^{3'} |
| w_1^- | | O- <u>A</u> -PO ₄ H ₂ ^{3'} |
| c_2^- | | ^{5'} OH-CG-PO ₃ |
| d_2^- | | ^{5'} OH-CG-PO ₄ H |
| z_2^- | | <u>TA</u> -PO ₄ H ₂ ^{3'} |
| y_2^- | | O- <u>TA</u> -PO ₄ H ₂ ^{3'} |
| x_2^- | | PO ₃ - <u>TA</u> -PO ₄ H ₂ ^{3'} |
| w_2^- | w_2^- | PO ₄ H- <u>TA</u> -PO ₄ H ₂ ^{3'} |
| $[\text{c}_3-\text{B}_3(\text{T})\text{H}]^-$ | | ^{5'} OH-GC-désoxy-PO ₃ |
| a_3^- | | ^{5'} OH-GCT |
| b_3^- | | ^{5'} OH-GCT-O |
| $[\text{a}_3-\text{B}_3(\text{T})\text{H}]^-$ | $[\text{a}_3-\text{B}_3(\text{T})\text{H}]^-$ | ^{5'} OH-CG-désoxy |
| c_3^- | | ^{5'} OH-GCT-PO ₃ |
| d_3^- | | ^{5'} OH-GCT-PO ₄ H |
| z_3^- | | <u>CTA</u> -PO ₄ H ₂ ^{3'} |
| y_3^- | | O- <u>CTA</u> -PO ₄ H ₂ ^{3'} |
| x_3^- | | PO ₃ - <u>CTA</u> -PO ₄ H ₂ ^{3'} |
| w_3^- | | PO ₄ H- <u>CTA</u> -PO ₄ H ₂ ^{3'} |
| $[\text{M}-\text{H}-\text{B}_2(\text{G})\text{H}]^-$ | | ^{5'} OH-désoxy- <u>CTA</u> -PO ₄ H ₂ ^{3'} |
| $[\text{M}-\text{H}-\text{B}_4(\text{A})\text{H}]^-$ | $[\text{M}-2\text{H}-\text{B}_4(\text{A})\text{H}]^{2-}$ et $[\text{M}-3\text{H}-\text{B}_4(\text{A})\text{H}]^{3-}$ | ^{5'} OH-GCT-désoxy-PO ₄ H ₂ ^{3'} |
| $[\text{M}-\text{H}-\text{B}_2(\text{C})\text{H}]^-$ | | ^{5'} OH-G-désoxy- <u>A</u> -PO ₄ H ₂ ^{3'} |
| b_4^- | b_4^{2-} | ^{5'} OH-GCT <u>A</u> -OH |
| $[\text{M}-\text{H}]^-$ | $[\text{M}-3\text{H}]^{3-}$ | ^{5'} OH-GCT <u>A</u> -PO ₄ H ₂ ^{3'} |

Tableau 1.8: Séquences des spectres CID, par ionisation MALDI et ESI en mode négatif, de l'ion parent de la chimère GCT**A**. La base **A**, soulignée et en gras, est un ribonucléotide possédant un groupement 3'-PO₄H₂. Les ions parents sont en gras.

Toutes les fragmentations en mode CID de la chimère GCTA par ESI-ITMSⁿ sont également présentes par MALDI-TOF/TOF. Cependant, les ions fragments issus par MALDI-TOF/TOF avec la matrice THAP sont beaucoup plus nombreux bien que cette technique d'ionisation/désorption soit douce. Toutefois, les fragmentations par MALDI ont été réalisées sous une fluence de 100 % (haute énergie) favorisant de nombreuses dissociations de la chimère. Les ions fragments par MALDI-TOF/TOF sont monochargés tandis que les ions par ESI-ITMSⁿ sont multichargés.

Les ions fragments communs sont issus de la perte : de la base A ($[M-H-B_4(A)H]^-$ et A^-), de la base C ($[a_2-B_2(C)H]^-$ et w_2^-), de la base T ($[a_3-B_3(T)H]^-$), du groupement 3'-PO₄H₂ (b_4^-) et de la rupture de la chaîne (w_2^-).

Les ions fragments abondants obtenus en mode CID de la chimère sont les ions w_1^- , w_2^- , w_3^- , y_2^- et $[M-B_1(G)H]^-$ par MALDI-TOF/TOF et les ions b_4^{2-} , la perte de la base A et les ions $[a_n-B_nH]^-$ par ESI-ITMSⁿ. Ces deux techniques d'ionisation/désorption produisent généralement des ions a_n^- , $[a_n-B_nH]^-$, d_n^- et w_n^- abondants mais les ions les plus abondants par MALDI-TOF/TOF sont les ions w_n^- , d_n^- tandis que par ESI-ITMSⁿ ce sont les ions $[a_n-B_nH]^-$ et w_n^- .

Les abondances des ions obtenus en mode CID par ESI-ITMSⁿ sont donc semblables à la littérature contrairement à ceux obtenus par MALDI-TOF/TOF. Des chimères ARN/ADN constituées d'un plus grand nombre de nucléotides ont donc été étudiées par MALDI-TOF/TOF afin de déterminer l'influence de la chimère ARN/ADN sur les dissociations en phase gazeuse.

3.4. Les fragmentations en mode CID des chimères ARN/ADN supérieures à 4-mer par MADI-TOF/TOF MS

Les fragmentations en mode CID des chimères ARN/ADN de 5, 7 et 10-mer ont été étudiées par MALDI-TOF/TOF.

3.4.1. Les fragmentations en mode CID des chimères ARN/ADN de 5-mer

Les ions fragments obtenus en mode CID de l'ion parent $[M-H]^-$ (m/z 1573,005) des chimères de 5-mer suivants: TTCGA, TTGCA, TCGTA, TGCTA, TCTGA, TGTCA, CTTGA, CTTCA, CGTTA, GCTTA, CTGTA et GTCTA sont présentés dans le tableau 1.9.

1.9A

| Fragments des chimères 5'OH-B ₁ B ₂ B ₃ B ₄ <u>A</u> -PO ₄ H ₂ ^{3'} | | |
|--|--------------------------------------|---|
| Ion m/z | Spécifique | Séquence |
| PO ₃ ⁻ | | PO ₃ |
| PO ₄ H ₂ ⁻ | | PO ₄ H ₂ |
| C ⁻ | B ₁ = C | C |
| T ⁻ | | T |
| A ⁻ | | A |
| G ⁻ | | G |
| [B, PO-H ₂] ⁻ | | B, PO-H ₂ |
| [B, PO ₃] ⁻ | | B, PO ₃ |
| [B, PO ₄ H] ⁻ | | B, PO ₄ H |
| [B, PO ₄ H ₂ , PO ₃] ⁻ | | B, PO ₄ H ₂ , PO ₃ |
| [B, PO ₄ H ₂ , PO ₄ H] ⁻ | | B, PO ₄ H ₂ , PO ₄ H |
| [B(C) PO ₄] ⁻ | B ₁ ≠ C | B(C) PO ₄ |
| x ₁ ⁻ | | <u>A</u> -PO ₄ H ₂ ^{3'} |
| w ₁ ⁻ | | O- <u>A</u> -PO ₄ H ₂ ^{3'} |
| z ₂ ⁻ | | B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} |
| y ₂ ⁻ | | O-B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} |
| x ₂ ⁻ | | PO ₃ -B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} |
| w ₂ ⁻ | | PO ₄ H-B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} |
| [c ₃ -B ₃ H] ⁻ | B ₃ et B ₄ ≠ T | 5'OH-B ₁ B ₂ -désoxy-PO ₃ |
| c ₃ ⁻ | | 5'OH-B ₁ B ₂ B ₃ -PO ₃ |
| d ₃ ⁻ | | 5'OH-B ₁ B ₂ B ₃ -PO ₄ H |
| z ₃ ⁻ | | B ₂ B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} |
| y ₃ ⁻ | | O-B ₂ B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} |
| [a ₄ -B ₄ H] ⁻ | B ₄ ≠ T | 5'OH-B ₁ B ₂ B ₃ -désoxy |
| x ₃ ⁻ | | PO ₃ -B ₂ B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} |
| w ₃ ⁻ | | PO ₄ H-B ₂ B ₃ <u>A</u> -PO ₄ H ₂ ^{3'} |
| [c ₄ -B ₄ H] ⁻ | B ₄ ≠ T | 5'OH-B ₁ B ₂ B ₃ -désoxy-PO ₃ |
| a ₄ ⁻ | | 5'OH-B ₁ B ₂ B ₃ B ₄ |
| b ₄ ⁻ | | 5'OH-B ₁ B ₂ B ₃ B ₄ -O |
| c ₄ ⁻ | | 5'OH-B ₁ B ₂ B ₃ -PO ₃ |
| d ₄ ⁻ | | 5'OH-B ₁ B ₂ B ₃ -PO ₄ H |
| z ₄ ⁻ | | B ₂ B ₃ B ₄ <u>A</u> -PO ₄ H ₂ ^{3'} |
| y ₄ ⁻ | | O-B ₂ B ₃ B ₄ <u>A</u> -PO ₄ H ₂ ^{3'} |
| x ₄ ⁻ | | PO ₃ -B ₂ B ₃ B ₄ <u>A</u> -PO ₄ H ₂ ^{3'} |
| w ₄ ⁻ | | PO ₄ H-B ₂ B ₃ B ₄ <u>A</u> -PO ₄ H ₂ ^{3'} |
| [M-H-B(G)H] ⁻ | | chimère -GH |
| [M-H-B ₄ (A)H] ⁻ | | chimère - base AH |
| [M-H-B(C)H] ⁻ | | chimère - CH |
| b ₅ ⁻ | | 5'OH-B ₁ B ₂ B ₃ B ₄ <u>A</u> -OH ^{3'} |
| [M-H] ⁻ | | 5'OH-B ₁ B ₂ B ₃ B ₄ <u>A</u> -PO ₄ H ₂ ^{3'} |

1.9B

| Masse | | | | | | | | | | | |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| <u>TTCGA</u> | <u>TTGCA</u> | <u>TCGTA</u> | <u>TGCTA</u> | <u>TCTGA</u> | <u>TGTC A</u> | <u>CTTGA</u> | <u>GTTC A</u> | <u>CGTTA</u> | <u>GCTTA</u> | <u>CTGTA</u> | <u>GTCTA</u> |
| 78,972 | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 | 78,972 |
| 97,092 | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 | 97,092 |
| | | | | | | 110,094 | | 110,094 | | 110,094 | |
| 125,105 | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 | 125,105 |
| 134,119 | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 | 134,119 |
| 150,118 | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 | 150,118 |
| 159,057 | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 | 159,057 |
| 177,072 | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 | 177,072 |
| 195,087 | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 | 195,087 |
| 257,052 | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 | 257,052 |
| 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 | 273,051 |
| 288,177 | 288,177 | 288,177 | 288,177 | 288,177 | 288,177 | | 288,177 | | 288,177 | | 288,177 |
| 408,181 | 408,181 | 408,181 | 408,181 | 408,181 | 408,181 | 408,181 | 408,181 | 408,181 | 408,181 | 408,181 | 408,181 |
| 426,196 | 426,196 | 426,196 | 426,196 | 426,196 | 426,196 | 657,411 | 617,386 | 632,398 | 632,398 | 632,398 | 632,398 |
| 657,411 | 617,386 | 632,398 | 632,398 | 657,411 | 617,386 | 675,426 | 635,401 | 650,413 | 650,413 | 650,413 | 650,413 |
| 675,426 | 635,401 | 650,413 | 650,413 | 675,426 | 635,401 | 426,196 | 426,196 | 426,196 | 426,196 | 426,196 | 426,196 |
| 737,391 | 697,366 | 712,378 | 712,378 | 737,391 | 697,366 | 737,391 | 697,366 | 712,378 | 712,378 | 712,378 | 712,378 |
| 755,406 | 715,381 | 730,393 | 730,393 | 755,406 | 715,381 | 755,406 | 715,381 | 730,393 | 730,393 | 730,393 | 730,393 |
| 785,469 | 785,469 | 770,457 | 810,482 | 770,457 | 810,482 | 785,469 | 785,470 | | | 770,457 | 810,482 |
| 896,574 | 936,596 | 921,584 | 921,584 | 896,574 | 936,596 | 896,574 | 936,596 | 921,584 | 921,584 | 921,584 | 921,584 |
| 914,590 | 954,611 | 939,599 | 939,599 | 914,590 | 954,611 | 914,590 | 954,611 | 939,599 | 939,599 | 939,599 | 939,599 |
| 946,596 | 946,596 | 961,608 | 921,583 | 961,608 | 921,583 | 961,608 | 921,583 | 936,595 | 936,595 | 961,608 | 921,583 |
| 964,611 | 964,611 | 979,623 | 939,598 | 979,623 | 939,598 | 979,623 | 939,598 | 954,610 | 954,610 | 979,623 | 939,598 |
| 994,671 | 1034,696 | | | 994,671 | 1034,696 | 994,671 | 1034,696 | | | | |
| 1026,576 | 1026,576 | 1041,588 | 1001,563 | 1041,588 | 1001,563 | 1041,588 | 1001,563 | 1016,575 | 1016,575 | 1041,588 | 1001,563 |
| 1044,591 | 1044,591 | 1059,603 | 1019,578 | 1059,603 | 1019,578 | 1059,603 | 1019,578 | 1034,590 | 1034,590 | 1059,603 | 1019,578 |
| 1074,655 | 1114,679 | | | 1074,676 | 1114,700 | 1074,676 | 1114,700 | | | | |
| 1145,801 | 1145,801 | 1145,801 | 1145,801 | 1145,801 | 1145,801 | 1145,801 | 1145,801 | 1145,801 | 1145,801 | 1145,801 | 1145,801 |
| 1163,816 | 1163,816 | 1163,816 | 1163,816 | 1163,816 | 1163,816 | 1163,816 | 1163,816 | 1163,816 | 1163,816 | 1163,816 | 1163,816 |
| 1225,781 | 1225,781 | 1225,781 | 1225,781 | 1225,781 | 1225,781 | 1225,781 | 1225,781 | 1225,781 | 1225,781 | 1225,781 | 1225,781 |
| 1243,796 | 1243,796 | 1243,796 | 1243,796 | 1243,796 | 1243,796 | 1243,796 | 1243,796 | 1243,796 | 1243,796 | 1243,796 | 1243,796 |
| 1250,793 | 1250,793 | 1250,793 | 1250,793 | 1250,793 | 1250,793 | 1265,805 | 1225,780 | 1265,805 | 1225,780 | 1265,805 | 1225,780 |
| 1268,808 | 1268,808 | 1268,808 | 1268,808 | 1268,808 | 1268,808 | 1283,820 | 1243,795 | 1283,820 | 1243,795 | 1283,820 | 1243,795 |
| 1330,768 | 1330,768 | 1330,768 | 1330,768 | 1330,768 | 1330,768 | 1345,785 | 1305,760 | 1345,785 | 1305,760 | 1345,785 | 1305,760 |
| 1348,783 | 1348,783 | 1348,783 | 1348,783 | 1348,783 | 1348,783 | 1363,800 | 1323,775 | 1363,800 | 1323,775 | 1363,800 | 1323,775 |
| 1421,879 | 1421,879 | 1421,879 | 1421,879 | 1421,879 | 1421,879 | 1421,879 | 1421,879 | 1421,879 | 1421,879 | 1421,879 | 1421,879 |
| 1437,879 | 1437,879 | 1437,879 | 1437,879 | 1437,879 | 1437,879 | 1437,879 | 1437,879 | 1437,879 | 1437,879 | 1437,879 | 1437,879 |
| 1461,903 | 1461,903 | 1461,903 | 1461,903 | 1461,903 | 1461,903 | 1461,903 | 1461,903 | 1461,903 | 1461,903 | 1461,903 | 1461,903 |
| 1493,025 | 1493,025 | 1493,025 | 1493,025 | 1493,025 | 1493,025 | 1493,025 | 1493,025 | 1493,025 | 1493,025 | 1493,025 | 1493,025 |
| 1573,005 | 1573,005 | 1573,005 | 1573,005 | 1573,005 | 1573,005 | 1573,005 | 1573,005 | 1573,005 | 1573,005 | 1573,005 | 1573,005 |

Tableau 1.9 : (A) Séquences des spectres CID, par MALDI-TOF/TOF, en mode négatif, de l'ion parent [M-H] (m/z 1573,005) des chimères TTCGA, TTGCA, TCGTA, TGCTA, TCTGA, TGTC A, CTTGA, CTTCA, CGTTA, GCTTA, CTGTA et GTCTA. (B) Masses monoisotopiques des ions fragments. La base A, soulignée et en gras, est un ribonucléotide possédant un groupement 3'-PO₄H₂. Les masses des ions en gras sont les ions les plus abondants.

Les différents types d'ions fragments obtenus en mode CID des chimères de 4 et 5-mer sont identiques et permettent l'analyse des séquences (tableaux 1.4 et 1.9). Cependant, pour les

chimères de 5-mer, l'ion fragment $[c_3-B_3H]^-$ est observé sauf dans le cas où les bases B_3 et B_4 contiennent une base T alors que son équivalent pour les chimères de 4-mer $[c_2-B_2H]^-$ n'est jamais observé.

De plus, pour les chimères de 5-mer, les spectres CID montrent la présence d'ions fragments beaucoup plus abondants que ceux des 4-mer (figure 1.2 et 1.3). L'abondance de ces ions fragments est en relation directe avec la séquence. Les spectres CID de la figure 1.15 présentent les ions les plus abondants des chimères : $TTCG\mathbf{A}$, $TCTG\mathbf{A}$, $TGTCA$, $CTTGA$, $GTTC\mathbf{A}$ et $CGTT\mathbf{A}$.

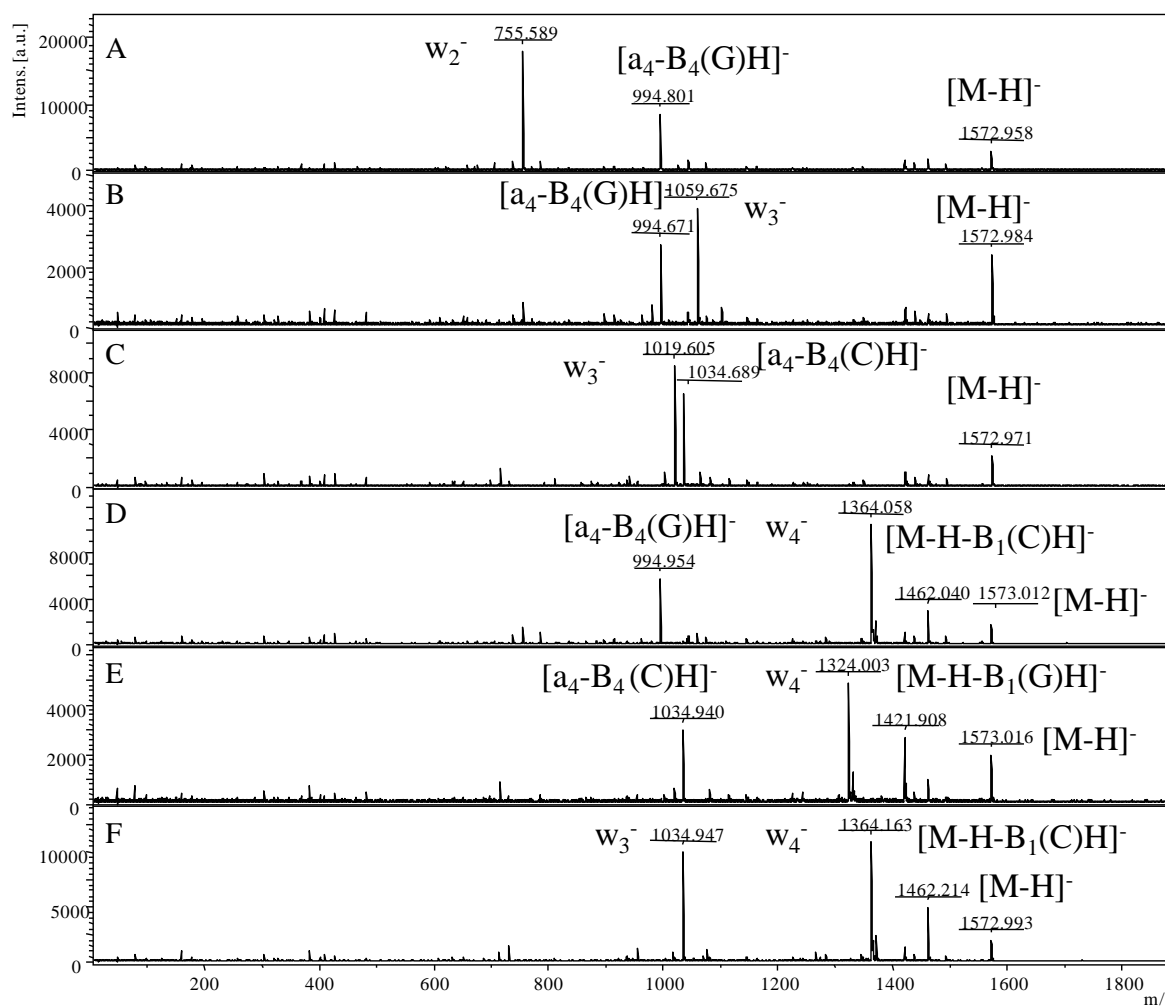


Figure 1.15: Spectres CID, par MALDI-TOF/TOF, en mode négatif, de l'ion parent $[M-H]^-$ (m/z 1573,005) des chimères : (A) $TTCG\mathbf{A}$; (B) $TCTG\mathbf{A}$; (C) $TGTCA$; (D) $CTTGA$; (E) $GTTC\mathbf{A}$ et (F) $CGTT\mathbf{A}$. La base \mathbf{A} , soulignée et en gras est un ribonucléotide possédant un groupement $3'-PO_4H_2$.

Ils montrent l'ion parent $[M-H]^-$ (m/z 1573,005) des chimères de 5-mer.

Lorsque la base est B_1 (C ou G), l'ion $[M-H-B_1H]^-$ et l'ion w_4^- font partie des ions les plus abondants du spectre CID (figure 1.15D à F). Lorsque la base est B_2 (C ou G), l'ion w_3^- est un des ions les plus abondants (figure 1.15B et C). Lorsque la base est B_3 (C ou G), l'ion w_2^- est également un des ions les plus abondants (figure 1.15.A). Enfin, lorsque la base est B_4 (C ou G), l'ion $[a_4-B_4H]^-$ est un des ions les plus abondants (figure 1.15.A à E).

Tout comme les chimères de 4-mer, les ions fragments d_n^- issus de chimères de 5-mer ne font également pas partie des ions les plus abondants tandis que les ions w_n^- et les ions issus des pertes de bases sont les plus abondants.

Plus le nombre de nucléotides de la chimère augmente plus les fragmentations sont dépendantes de la séquence. L'analyse par MS/MS permet donc le séquençage des chimères de 5-mer. Cependant, les chimères de 5-mer doivent être analysées à des concentrations plus élevées que les chimères de 4-mer afin de pouvoir sélectionner l'ion parent et détecter tous les ions fragments.

3.4.2. Les fragmentations en mode CID des chimères ARN/ADN de 7-mer

Les ions fragments obtenus en mode CID de l'ion parent $[M-H]^-$ (m/z 2199,425) des chimères de 7-mer suivantes: AAACTTG, AAATCTG et AAATTCTG sont présentés dans la figure 1.16.

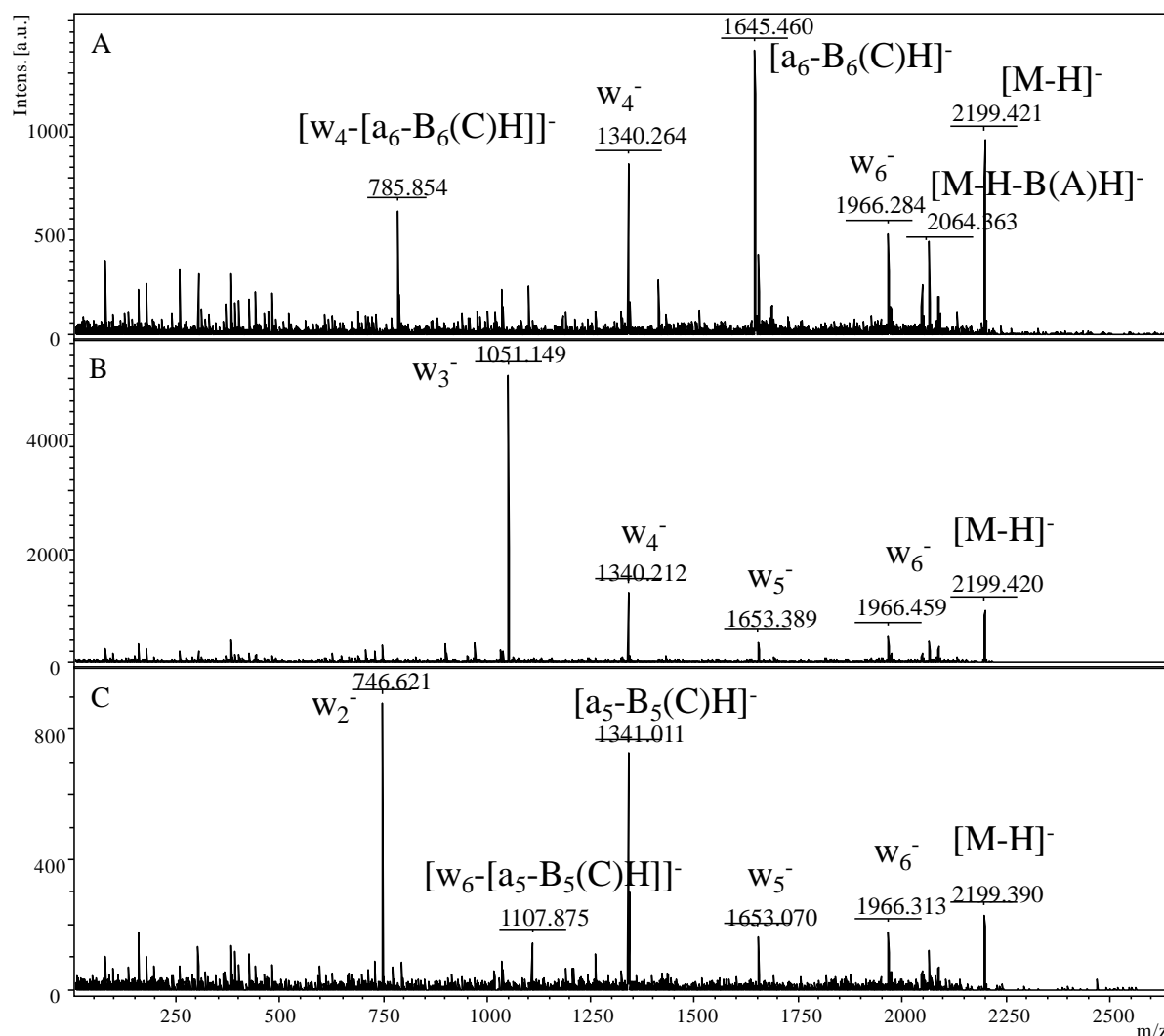


Figure 1.16: Spectres CID, par MALDI-TOF/TOF, en mode négatif, de l'ion parent $[M-H]^-$ (m/z 2199,425) des chimères : (A) AAATTCG ; (B) AAACTTG et (C) AAATCTG. La base G, soulignée et en gras, est le ribonucléotide possédant un groupement $3'-PO_4H_2$.

Ils montrent à m/z 2199,425 l'ion parent $[M-H]^-$ des chimères de 7-mer.

Les spectres CID montrent la présence des ions fragments issus de la perte de la base $[M-H-B(C)H]^-$ (m/z 2088,323), $[M-H-B(A)H]^-$ (m/z 2064,299) et $[M-H-B(G)H]^-$ (m/z 2048,299). Ils ne sont pas très abondants sauf dans le cas de la chimère AAATTCG ou l'ion fragment $[M-H-B(A)H]^-$ est abondant (figure 1.16A). Cette chimère est la seule dont les deux bases T sont adjacentes aux trois bases A.

L'ion w_6^- (m/z 1966,195) est abondant pour les trois chimères (figure 1.16). Lorsque la base C est en position B₄ ou B₅, l'ion w_5^- (m/z 1652,986) est abondant (figure 1.16B et C). Lorsque la base C est en position B₄ ou B₆, l'ion w_4^- (m/z 1339,7754) est abondant (figure 1.16A et B).

Pour la chimère AAATTCG, l'ion fragment $[a_6-B_6(C)H]^-$ (m/z 1645,121) et l'ion fragment interne $[w_4-[a_6-B_6(C)H]]^-$ (m/z 785,466) alors que pour la chimère AAATCTG, l'ion fragment $[a_5-C_5H]^-$ (m/z 1340,919) et l'ion fragment interne $[w_6-[a_5-B_5(C)H]]^-$ (m/z 1107,959) sont abondants (figure 1.16A et C). Par contre, lorsque la base est B₄(C), l'ion fragment $[a_4-B_4(C)H]^-$ et celui interne ne sont pas présents. Lorsque la base est B₅(C), l'ion fragment w_2^- (m/z 746,393) est le plus abondant tandis que lorsque la base est B₄(C), l'ion fragment w_3^- (m/z 1050,5904) est le plus abondant. La coupure prépondérante semble avoir lieu avant la base C (figure 1.6B et C). Par contre, lorsque la base est B₆(C), l'ion fragment $[a_6-B_6(C)H]^-$ (m/z 1645,121) est le plus abondant.

Plus le nombre de nucléotides de la chimère augmente plus les fragmentations sont dépendantes et favorisées par sa séquence. L'analyse par MS/MS permet donc le séquençage des chimères de 7-mer. Cependant, les chimères de 7-mer doivent être analysées à des concentrations plus élevées que les chimères de 4 et 5-mer afin de pouvoir sélectionner l'ion parent et détecter tous les ions fragments.

Enfin, les ions fragments abondants des chimères de 7-mer sont les ions w_n^- et ceux issus de la perte de la base, tout comme les chimères de 4 et 5-mer.

3.4.3. Les fragmentations en mode CID des chimères ARN/ADN de 10-mer

Les ions fragments obtenus en mode CID de l'ion parent $[M-H]^-$ (m/z 3162,042) des chimères de 10-mer: AGTAGTTAGC, TAGTAAGGTC et GTAGTGATACC sont présentés dans la figure 1.17.

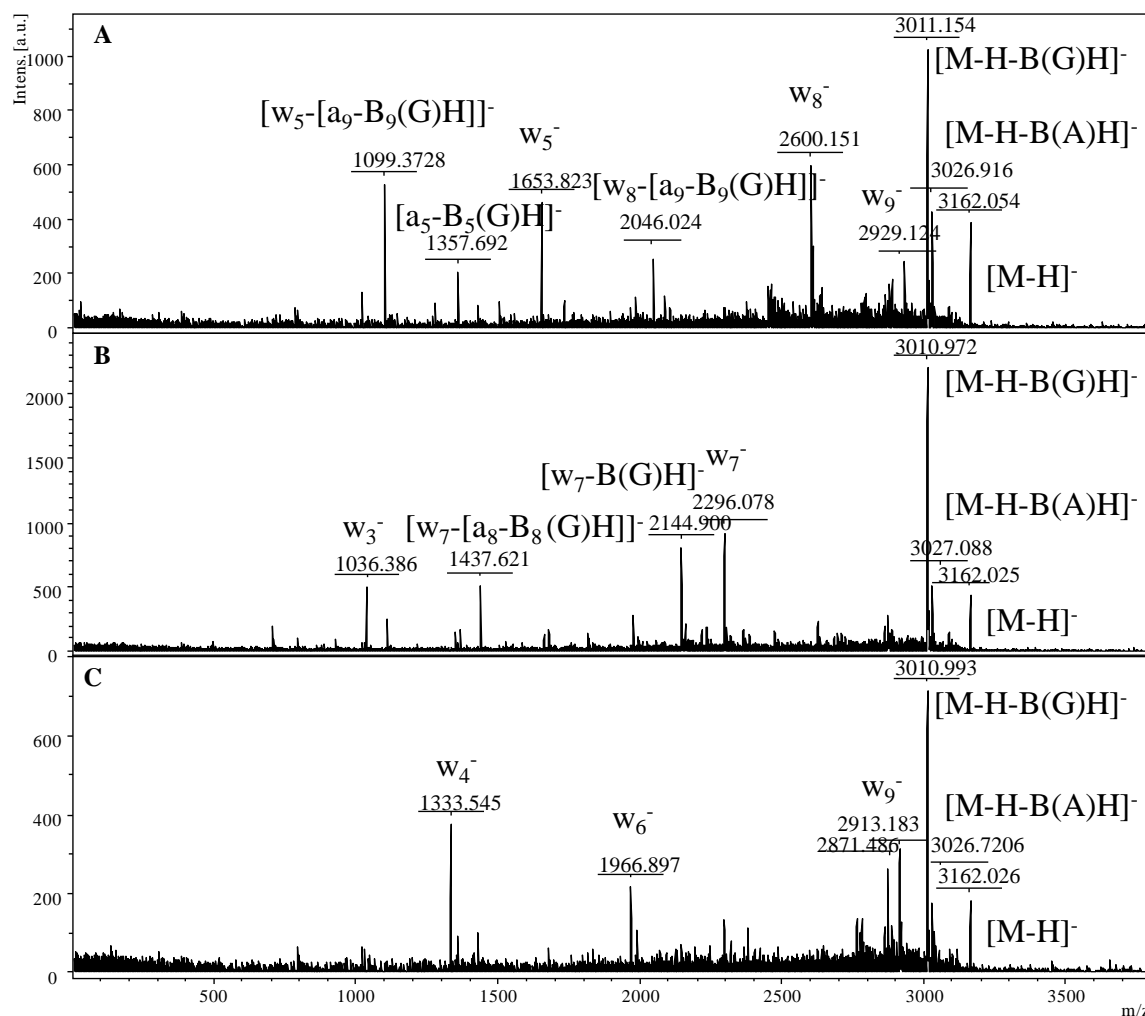


Figure 1.17: Spectres CID, par MALDI-TOF-TOF, en mode négatif, de l'ion parent $[M-H]^-$ (m/z 3162,042) des chimères : (A) AGTAGTTAGC; (B) TAGTAAGGTC et (C) GTAGTGATAC. La base C qui est soulignée et en gras, est le ribonucléotide qui possède un groupement 3'- PO_4H_2 .

Pour les trois chimères de 10-mer, l'ion parent $[M-H]^-$ est celui à m/z 3162,042.

Les ions fragments des chimères de 10-mer sont plus difficiles à détecter que ceux des chimères plus petites. Par conséquent, la détermination du rapport m/z des ions est plus difficile, augmentant ainsi l'écart entre le rapport m/z théorique et expérimental.

Dans les spectres CID, l'ion fragment $[M-H-B(G)H]^-$ (m/z 3010,916) est le plus abondant et l'ion fragment $[M-H-B(A)H]^-$ (m/z 3026,916) est beaucoup moins abondant bien que le nombre de base de G et de A soit identique. La perte de la base G est donc prépondérante par rapport à celle de la base A.

Les spectres CID des ions fragments des 3 chimères montrent des ions fragments abondants qui sont caractéristiques de leur séquence. La figure 1.17A de la chimère

AGTAGTTAGC montre 3 ions de la série d'ions w_n^- : à m/z 2929,124 l'ion w_9^- (m/z 2928,8124), à m/z 2600,151 l'ion w_8^- (m/z 2599,6024) et à m/z 1853,823 l'ion fragment w_5^- (m/z 1652,985). L'ion fragment à m/z 1357,692 correspond à l'ion $[a_5-B_5(G)H]^-$ (m/z 1356,919) qui est issu de la même dissociation de la chimère que l'ion w_5^- . Enfin, les deux ions fragments suivants sont des ions internes : à m/z 2046,024 l'ion $[w_8-[a_9-B_9(G)H]]^-$ (m/z 2045,2927) et à m/z 1099,3728 à l'ion $[w_5-[a_9-B_9(G)H]]^-$ (m/z 1098,6757).

La figure 1.17B de la chimère TAGTAAGGTC montre deux ions de la série d'ions w_n^- : à m/z 2296,078 à l'ion w_7^- (m/z 2295,405) et à m/z 1036,386 l'ion w_3^- (m/z 1035,578). Enfin, les deux ions suivants sont des ions fragments internes : à m/z 2144,900 l'ion fragment $[w_7-GH]^-$ (m/z 2144,280) et à m/z 1437,621 à l'ion $[w_7-[a_8-B_8(G)H]]^-$ (m/z 1436,899).

La figure 1.15C de la chimère GTAGTGATAC montre 3 ions de la série w_n^- : à m/z 2913,183 à l'ion w_9^- (m/z 2912,812), à m/z 1966,897 à l'ion w_6^- (m/z 1966,195) et à m/z 1333,545 l'ion w_4^- (m/z 1332,788). Par ailleurs, l'ion à m/z 2871,486 n'est pas un ion qui possède un massif isotopique, il n'a pas pu être interprété.

Plus le nombre de nucléotides de la chimère augmente plus les fragmentations sont dépendantes et favorisées par sa séquence. De plus, les chimères de 7 et 10-mer doivent être analysées à des concentrations plus élevées que les chimères de 4 et 5-mer afin de pouvoir sélectionner l'ion parent et détecter tous les ions fragments. L'analyse par MS/MS permet donc le séquençage des chimères de 10-mer mais il devient de plus en plus difficile dû à la diminution du pouvoir résolutif de l'instrument lorsque les masses des ions augmentent.

Enfin, les ions fragments abondants des chimères de 10-mer sont les ions w_n^- et ceux issus de la perte de la base, tout comme les chimères de 4, 5 et 7-mer.

4. Discussion

4.1. Les conditions expérimentales

4.1.1. La préparation de la chimère ARN/ADN

Une chimère ARN/ADN synthétique peut être synthétisée commercialement mais le but de cette étude est d'analyser la chimère en phase gazeuse dans les mêmes conditions de préparation que celles des méthodes présentées dans ce travail. La chimère ARN/ADN a donc

été formée à partir du clivage par l'hydroxyde de sodium d'un oligonucléotide contenant deux ribonucléotides et des désoxynucléotides. Les bases des ribonucléotides sont identiques et différentes de celles des désoxynucléotides. La chimère ARN/ADN est composée de plusieurs désoxynucléotides et se termine par un ribonucléotide constitué d'un groupement 3'-PO₄H₂.

La chimère ARN/ADN est ensuite dessalée à l'aide d'une résine échangeuse de cations afin d'une part, d'échanger les ions sodium de l'échantillon avec les protons et d'autre part, de neutraliser les ions sodium provenant de la soude.

Les étapes de clivage par la soude, de dessalage et d'analyses par MALDI-TOF MS ont été décrites précédemment (Mauger et al. 2007) et sont identiques dans les chapitres 3, 4 et 5.

Toutefois, quelques chimères de 4-mer synthétiques ont été utilisées. Elles donnent les mêmes résultats de fragmentation en phase gazeuse que leurs homologues qui ont été clivées. Les conditions de préparation en solution de la chimère ARN/ADN n'impacte donc pas sur ces fragmentations en phase gazeuse.

4.1.2. L'analyse par MALDI-TOF/TOF

La matrice THAP a été utilisée car elle est utilisée dans les méthodes présentées dans ce travail. La matrice THAP présente une meilleure reproductibilité des résultats des tirs du laser sur les agrégats de matrice-analyte et une meilleure résolution que la matrice HPA qui est la plus couramment utilisée pour l'analyse des oligonucléotides (Zhu et al. 1996). De plus, l'analyse de la chimère ARN/ADN a été réalisée en mode négatif car l'atome d'oxygène du groupement phosphate porte naturellement une charge négative.

Le spectre de l'ion parent est sélectionné à l'aide d'une fluence de 64 % à partir de la somme de 200 tirs de laser tandis que celui de ses ions fragments en mode CID est obtenu à l'aide d'une fluence de 100 % à partir de la somme de 1000 tirs de laser et il superposé à celui de l'ion parent.

La variation de la fluence du laser en MS/MS a été étudiée de manière expérimentale sur la chimère GCTA car elle impacte sur les fragmentations métastables et les fragmentations en mode CID. Pour l'analyse, des ions métastables, la fenêtre de temps de dissociation des ions dans la cellule *LIFT* doit être très courte (quelques microsecondes) car le déplacement cinétique (différence entre le seuil d'énergie de dissociation dans la cellule *LIFT* et l'énergie

d'activation) est grand. Dans la deuxième partie de la distribution, les ions se dissocient dans le deuxième TOF et ne peuvent donc pas être analysés *normalement* car ils apparaissent en même temps que leur précurseur. L'augmentation de la fluence du laser fait déplacer la distribution de l'énergie interne vers les plus hautes énergies entraînant une augmentation de l'abondance des ions fragments et la diminution de l'ion parent. Ces conditions de sensibilité de détection des ions produits à partir des ions métastables, permettent de mieux observer la distribution des ions fragments dont l'abondance relative n'est pas fortement modifiée. Si la sensibilité diminue, l'ion parent a tendance à ne plus être détecté. Lorsque la fluence augmente, certaines décompositions observées avec la cellule *LIFT* sont déjà présentes dans le spectre de masse et sont dues à des décompositions promptes de type ISD (*Ion Source Decay*).

Par ailleurs, les fragmentations des oligonucléotides sont principalement celles observées en mode CID produisant ainsi de nombreuses dissociations (Bartlett et al. 1996) à l'aide de la collision avec un gaz inerte. Les fragmentations métastables et celles obtenues en mode CID de la chimère GCTA sont les mêmes sauf pour les petits fragments de rapport m/z inférieurs à celui d'un nucléotide. Les fragmentations en mode CID semblent donc être surtout des fragmentations métastables renforcées. Sous CID, les ions se sont décomposés juste après la source mais avant la cellule *LIFT*. Ils ont donc approximativement la même vitesse que celle de l'ion précurseur. Ils sont donc sélectionnés dans la cellule *LIFT* et subissent des décompositions consécutives tout comme les décompositions qui auront lieu dans la cellule. Le mode CID permet donc de renforcer les dissociations consécutives bien que les collisions soient statistiquement peu importantes compte-tenu de la faible pression qui règne dans la cellule *LIFT*.

Les ions fragments ont donc été analysés à l'aide d'une fluence de 100% en mode CID afin de pouvoir détecter tous les ions fragments et d'obtenir ainsi une bonne reproductibilité des mesures car l'abondance des ions fragments en mode CID est plus grande que celle des fragmentations métastables et nécessite donc moins de tirs de laser. Les spectres CID sont plus reproductibles que ceux métastables surtout pour la comparaison de la chimère avec les oligonucléotides sans groupement 3'-PO₄H₂. En effet, ces oligonucléotides présentent moins de décompositions métastables et semblent donc plus difficiles à fragmenter que la chimère correspondante. Le groupement 3'-PO₄H₂ semble rendre la chimère plus fragile en induisant des fragmentations supplémentaires.

4.2. Les fragmentations par MALDI-TOF-TOF de la chimère ARN/ADN

4.2.1. La détection des ions

Les fragmentations en mode CID des chimères ARN/ADN de 4, 5, 7 et 10 bases ont pu être réalisées car la chimère ARN/ADN est stable en phase gazeuse.

Plus le nombre de nucléotides des chimères est importante plus le *seuil de détection* des ions augmente également. Les chimères ARN/ADN, qui possèdent des thymines, de 4-mer sont analysées à 5 pmol, les chimères de 5-mer sont analysées à 200 pmol et les chimères de 7 et 10-mer sont analysées à 250 pmol. La limite de détection de la gamme de masse en mode réflecteur se situe aux alentours des 3000 Da. Plus le nombre de nucléotides de la chimère augmente plus la possibilité de fragmentations en phase gazeuse augmente. Il est donc plus difficile de sélectionner l'ion parent afin de le fragmenter. De plus, les ions fragments des chimères 7 et 10-mer sont plus difficiles à détecter ce qui pose des problèmes pour l'interprétation de leur séquence. Plus le nombre de nucléotides de la chimère augmente plus la chimère doit être concentrée.

Par ailleurs, l'étude de la détection des chimères ARN/ADN de 4-mer a montré la stabilité des bases suivantes : $T, U \geq G \geq A \geq C$ qui s'explique facilement par leur affinité protonique. Toutefois, le schéma des fragmentations des chimères est identique pour toutes les bases. Les thymines semblent stabiliser les chimères. La thymine est la plus acide en phase gazeuse donc la charge se forme préférentiellement sur son nucléotide (Hunter & Lias 1998). De plus, lors de la formation des ions zwitterions, la thymine possède une affinité protonique beaucoup plus faible que les autres bases (figure 1.18) (Klassen et al. 1998).

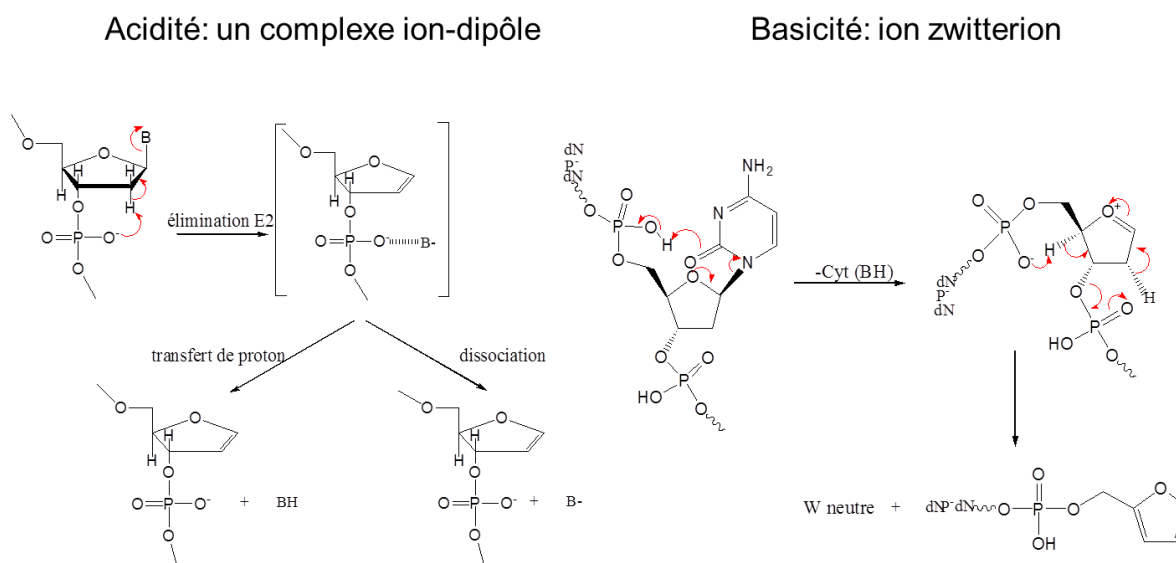


Figure 1.18 : Mécanismes de fragmentations en phase gazeuse : l'acidité des bases par la formation d'un complexe ion-dipôle et la basicité des bases par la formation d'un ion zwitterion. (Rodgers et al. 1994; Klassen et al. 1998).

Elle est donc beaucoup moins facilement protonée que les autres bases et donc plus stable en phase gazeuse. De plus, la présence de cette base dans la chimère diminue donc également la déprotonation des groupements phosphodiester à cause des charges positives, portées par les bases, de l'ion zwitterion qui sont moins neutralisées. Par conséquent, le nombre des fragmentations en phase gazeuse de la chimère diminue, rendant ainsi la chimère plus stable et détectable.

Par contre, lors de la formation des ions zwitterions, les bases A, C et G ont un meilleur rendement de protonation par rapport à la base T car leur affinité protonique est plus forte (Hunter & Lias 1998). Elles ont donc une constante de vitesse de dissociation plus grande que celle obtenue à partir de la thymine. Les fragmentations métastables de l'ion parent peuvent donc avoir lieu avant qu'il soit isolé diminuant ainsi sa stabilité et sa détection. Les chimères ARN/ADN ne contenant pas de thymine doivent donc être analysées à des concentrations plus élevées que les autres chimères afin de pallier à ce phénomène de dissociation prompt.

4.2.2. Les fragmentations en mode CID

4.2.2.1. Les ions fragments

Les schémas de dissociation des chimères ARN/ADN de 4-mer et de 5-mer en mode CID par MALDI-TOF/TOF sont identiques. Les ions fragments sont nombreux et tous monochargés. Tout d'abord, les huit séries d'ions qui correspondent à la dissociation de la chaîne

nucléotidique au niveau des quatre liaisons du phosphodiester sont présentes. Les séries d'ions fragments a_n^- , b_n^- , c_n^- et d_n^- contiennent l'extrémité 5'-OH de la chimère tandis que celles w_n^- , x_n^- , y_n^- , z_n^- contiennent l'extrémité 3'-PO₄H₂ de la chimère. Les ions issus de l'élimination des bases sont également présents. D'autres types de dissociations internes qui résultent d'un double clivage de la chaîne nucléique sont observés. La formation des ions internes augmente lorsque l'énergie interne, ainsi que la facilité à former des ions zwitterions, augmentent.

L'utilisation d'une fluence de 100 % du laser, de haute énergie, en mode CID peut engendrer toutes ces fragmentations bien que l'ionisation/désorption MALDI soit douce. Ce grand nombre d'ions peut être expliqué par des dissociations consécutives. Avec une fluence de 100 % du laser, de nombreuses fragmentations sont favorisées d'une part, dans la source et d'autre part, dans le tube de vol et plus particulièrement dans la région entre la source et la cellule *LIFT*. Les ions métastables, produits dans cette région, ont la même vitesse que leurs ions précurseurs. Ils subissent également les mêmes collisions que les ions précurseurs favorisant de nombreuses dissociations. De plus, le mode CID permet de favoriser les fragmentations non accessibles en basse énergie en MALDI augmentant ainsi encore les dissociations.

4.2.2.2. L'abondance des ions

Généralement, les spectres CID par MALDI-TOF/TOF ont tendance à présenter des ions d_n^- et w_n^- abondants (Hettich & Stemmler 1996). Par contre, dans ce cas, les ions w_n^- et les ions issus de la perte de bases sont très abondants tandis que les ions d_n^- ne le sont pas. Les ions fragments abondants des chimères de 4 et 5-mer étudiées sont: w_n^- , $[M-H-B_nH]^-$ et $[a_n-B_nH]^-$. Pour les chimères de 4-mer qui contiennent une base B₄(C), l'ion a_4^- est également très abondant et beaucoup plus que ceux des pertes de bases. Pour les chimères de 7-mer, les ions les plus abondants sont également les ions fragments de la série w_n^- , $[M-H-B_nH]^-$ et $[a_n-B_nH]^-$ mais aussi les ions fragments internes $[w-[a-BH]]^-$. Pour les chimères de 10-mer étudiées, les ions les plus abondants sont également les fragments de la série w_n^- , $[M-H-B_nH]^-$ et $[a_n-B_nH]^-$ mais aussi des fragments internes $[w-[a-BH]]^-$ et $[w-BH]^-$.

Par conséquent, plus le nombre de nucléotides des chimères augmente (5 à 10 bases) plus les ions w_n^- et les ions issus de la perte de bases deviennent prépondérants. Plus le nombre de nucléotides augmentent, plus la probabilité qu'un ion zwitterion se forme est importante.

Lorsque la chaîne nucléotidique est petite (de 2 à 4 bases) la présence d'une charge ne peut pas s'accompagner d'une seconde, à cause du groupement phosphate. En effet, l'acidité d'un groupement phosphate est grande en phase gazeuse mais la présence d'un deuxième groupement phosphate à proximité du premier engendre, par l'effet de charges, la diminution de l'acidité du second. Lorsque la chaîne nucléotidique est plus grande, la formation d'ions multichargés est donc plus grande. Cependant, les ions observés en MALDI sont monochargés car les charges négatives peuvent être équilibrées par des charges positives formant ainsi un ion zwitterion. Les charges positives sont portées par les bases tandis que les charges négatives sont portées par les groupements phosphates. La formation d'un ion zwitterion permet la protonation d'une base (selon sa basicité) et donc son élimination, favorisant ainsi les ions issus de pertes de bases ainsi que les ions w_n^- . De plus, les ions zwitterions conduisent à de grandes constantes de vitesse de dissociation permettant ainsi l'observation d'un grand nombre d'ions issus de la perte de la base préalable.

Par ailleurs, la perte de la base nucléique est observée sous forme d'anion B^- qui est en compétition avec BH (figure 1.18). Plus le nombre de nucléotide est important et plus la probabilité de former des ions zwitterions augmente (Klassen et al. 1998). Par conséquent, la base BH est éliminée par formation d'un ion-dipôle selon son acidité en phase gazeuse (figure 1.18). L'échange du proton avec la chimère fragmentée peut être favorisé. De plus, lors de la formation d'un ion zwitterion, l'anion B^- permet de neutraliser une charge négative de la chimère et son complémentaire est neutre (le nombre de charges positives et négatives est identique).

4.2.2.3. Le 3'-phosphate

La chimère ARN/ADN de 4 mer est composée de 3 désoxynucléotides et se termine par un seul ribonucléotide contenant un groupement 3'- PO_4H_2 . La comparaison de ces fragmentations avec celle d'un oligonucléotide : avec ou sans groupement 3'- PO_4H et avec ou sans ribonucléotide a montré que les ions fragments $[a_{n-1}-B_{n-1}H]^-$, $[c_{n-1}-B_{n-1}H]^-$ ($B_{n-1} \neq T$), C^- (pour la chimère $B_4(C)$), et b_4^- sont spécifiques de la présence du groupement 3'- PO_4H_2 . La charge négative de la chimère serait donc située sur le 3'-phosphate favorisant ainsi ces ions fragments. La charge du phosphate induit une élimination par une substitution intramoléculaire favorisant la formation des ions $[a_{n-1}-B_{n-1}H]^-$, $[c_{n-1}-B_{n-1}H]^-$.

Toutefois, pour les chimères de 4-mer, l'ion $[c_{n-1}-B_{n-1}H]^-$ ($B_{n-1} \neq T$) est présent alors que pour les chimères de 5-mer, les ions $[c_{n-1}-B_{n-1}H]^-$ ($B_{n-1} \neq T$), $[c_{n-2}-B_{n-2}H]^-$ (B_{n-2} et $B_{n-1} \neq T$) sont formés. La chimère de 5-mer contient plus de nucléotide que la chimère de 4-mer. Un ion zwitterion est donc probablement formé engendrant une charge négative sur le phosphodiester. Cette charge a permis la formation de l'ion $[c_{n-2}-B_{n-2}H]^-$.

4.2.2.4. La cytosine

Les fragmentations des chimères de 4-mer ont montrés que l'anion C^- est présent que lorsqu'il est en position B_1 ou B_4 (B_4 avec un 3'-phosphate). Lorsque la cytosine est en position B_2 ou B_3 , l'ion fragment $[B(C) PO_3]^-$ du désoxynucléotide est observé, et de même, si elle est en position B_4 , l'ion fragment $[B(C) PO_3]^-$ du ribonucléotide est produit. Pour la cytosine, il y a donc une dissociation spécifique. L'élimination des bases est donc favorisée aux extrémités par la présence de la charge négative sur le phosphate adjacent et la formation d'un ion-dipôle (figure 1.18). La charge négative serait donc située sur les phosphates situés aux extrémités 5' ou 3'.

De plus, dans le cas des chimères de 4-mer dont la B_4 (C), l'ion a_4^- est abondant. Il est le seul ion fragment spécifique de la dissociation des chimères de 4-mer. Tous les autres ions fragments obtenus sont similaires pour les différentes chimères. La charge négative se situant sur le 3'-phosphate, un transfert de proton est effectué entre le phosphate et le 2'-OH, puis, suivi d'un transfert entre le 2'-H du ribonucléotide et le CO de la cytosine dû à sa grande affinité protonique. L'ion O^- de la cytosine attaque donc l'atome 3'-C du ribonucléotide afin d'éliminer son groupement 3'-phosphate.

4.2.2.5. La thymine

Les ions fragments $[M-H-B_nH]^-$, $[a_n-B_nH]^-$ et $[c_{n-1}-B_{n-1}H]^-$ sont présents sauf ceux qui correspondent à la perte de thymine. Ce phénomène est expliqué grâce à l'affinité protonique des bases qui impliquent que les bases éliminées sont sous forme pré-protonée. Les ions zwitterions ont été formés où les bases ont été perdues. La thymine possède une plus faible affinité protonique, elle est donc faiblement protonée. Les ions fragments qui devraient être

issus de la perte de la thymine ne sont donc pas présents. La formation des ions zwitterions favorisent donc potentiellement les dissociations de ces ions.

4.2.3. Le séquençage

L'analyse de tous les ions fragments, produits sous CID, des chimères deprotonées de 4 ou 5-mer permet de les séquencer. De plus, l'analyse des ions fragments abondants permet de séquencer les chimères de 4 à 10-mer. Plus le nombre de nucléotides de la chimère augmente, plus l'abondance de certains ions est renforcée et est caractéristique de la séquence grâce à la formation des ions zwitterions qui s'accroît.

L'analyse des fragmentations en mode CID est donc une méthode de séquençage potentiellement très rapide. Elle permet d'avoir une information directe sur la séquence de la chimère. Cependant, il faut une certaine quantité de chimère pour pouvoir détecter l'ion parent et tous les ions fragments d'intérêt. Toutefois, plus le nombre de nucléotides de la chimère augmente plus les ions fragments sont difficiles à détecter précisément et les spectres CID deviennent plus complexes à étudier. L'interprétation des ions fragments devient difficile car elle dépend du pouvoir résolutif de l'instrument suffisant afin d'identifier précisément les masses des ions. Cette méthode est limitée à l'étude des chimères de 10 à 13-mer car au-delà le pouvoir résolutif de l'instrument ne permet pas de déterminer précisément les masses des ions fragments engendrant ainsi des ambiguïtés d'interprétation.

Par ailleurs, les chimères ARN/ADN sans thymine sont plus difficiles à détecter. Elles doivent donc être plus concentrées que les chimères ARN/ADN qui en contiennent pour être analysées. Cependant, les chimères ARN/ADN des méthodes présentées dans les chapitres suivants sont issues du clivage de la même chimère. Les fragments de clivage obtenus ont tous la même concentration que la chimère ARN/ADN dont ils sont issus. Les ions fragments sans T seront donc probablement plus difficiles à détecter dans ces conditions.

4.3. Les fragmentations en mode CID de la chimère GCTA

Contrairement au MALDI-TOF/TOF, Les ions fragments obtenus par ESI-ITMSⁿ sont produits sous faible collision et sont multichargés.

4.3.1. L'analyse par ESI-ITMSⁿ

Pour exemple, les fragmentations en mode CID de la chimère de 4-mer GCTA ont été étudiées par ESI-ITMSⁿ jusqu'à l'expérience MS⁵ séquentielle.

Dans le spectre de masse, la chimère est sous les formes [M-3H]³⁻, (m/z 422,4) et [M-4H]⁴⁻, (m/z 316,5). Il y a également quelques ions fragments qui sont également présents : [a₃-B₃(T)H]⁻, b₄²⁻, [M-3H-B₄(A)H]³⁻, [a₃-B₃(T)H]²⁻, A⁻ et G⁻, et y₁⁻ et/ou d₁⁻. Le spectre CID a été réalisé à partir de l'ion parent, m/z 422,4, de la forme trichargée de la chimère. Le spectre CID de l'expérience MS³ séquentielle a été réalisé à partir de l'ion précurseur, m/z 594,1, de l'ion b₄²⁻ de même séquence que la chimère mais qui se termine par un groupement 3'-OH. Trois spectres CID des expériences MS⁴ séquentielles différents ont été réalisés à partir des ions de rapports m/z suivants : à m/z 346,3 de l'ion précurseur d₁⁻, à m/z 650,5 de l'ion précurseur [b₄,w₂]⁻ et à m/z 715,5 de l'ion précurseur [a₃-B₃(T)H]⁻. Enfin, un spectre CID de l'expérience MS⁵ séquentielle a été réalisée à partir de l'ion précurseur [[b₄,w₂]-B(T)H]⁻, m/z 524,5, du spectre CID de l'expérience MS⁴ séquentielle de l'ion précurseur [b₄,w₂]⁻, m/z 650,5. Ces expériences séquentielles ont confirmées la séquence de l'ion précurseur.

Les ions produits par ESI-ITMSⁿ sont monochargés et multichargés dans le spectre de masse et dans les spectres CID des expériences séquentielles MS² et MS³ tandis que les ions des spectres des expériences séquentielles MS⁴ et MS⁵ sont simplement chargés car le nombre de charges des ions sélectionnés diminuent lorsque le nombre d'étapes consécutives de fragmentation augmentent.

Toutes ces expériences séquentielles en mode CID ont permis d'identifier tous les ions fragments et de confirmer la séquence de l'ion précurseur. C'est une bonne méthode pour le séquençage des chimères.

4.3.2. Les ions multichargés

L'étude des ions multichargés par ESI-ITMSⁿ permet d'identifier les positions des charges sur les ions. Dans le spectre de masse de la chimère, l'ion parent de la chimère est triplement chargé GCTA alors que l'ion b₄²⁻ qui a pour séquence GCTA est doublement chargé ce qui laisse à penser que 1 des 3 charges négatives de la chimère se trouve sur le 3'-PO₄H₂. Dans le

spectre CID de l'ion précurseur de la chimère triplement chargée, la perte de la base B₄(A) est observée sous forme [M-H-B₄(A)H]⁻ doublement et triplement chargée tandis que celles des bases B₃(T) et B₂(C) sont observées sous la série d'ions [a_n-B_nH]⁻. Le mécanisme de fragmentation nécessite l'ion intermédiaire [M-BH]⁻ et nécessite une charge adjacente à la perte de la base (figure 1.18). L'ion intermédiaire [M-2H-BH]²⁻ a perdu une charge par rapport à la chimère tandis que l'ion final [a_n-B_nH]⁻ en a perdu une autre. Les deux autres charges négatives de la chimère semblent donc être situées sur les deux phosphodiesters entre B₁(G)-B₂(C) et B₂(C)-B₃(T) permettant ainsi la formation des ions fragments [a₃-B₃(T)H]⁻ et [a₂-B₂(C)H]⁻. Par conséquent, le groupement phosphodiester situé entre B₃(T)-B₄(A) n'est pas chargé et donc il n'y a pas de formation de l'ion fragment [a₄-B₄(A)H]⁻ à partir de l'ion intermédiaire [M-B₄(A)H]⁻.

Dans le spectre CID de l'expérience séquentielle MS³ de l'ion précurseur b₄²⁻ de séquence GCTA-OH-3', les ions fragments sont : d₃ simplement et doublement chargé, a₃⁻, b₃⁻, [a₃-B₃(T)H]⁻, [b₄,w₂]⁻, [b₄,x₂]⁻, [a₂-B₂(C)H]⁻, [b₄-B₂(C)H]²⁻, [C PO₄]⁻ et [w₂,a₃]⁻. Le spectre CID de la chimère triplement chargée et celui de l'ion précurseur b₄²⁻, de même séquence mais se terminant par le groupement 3'-OH au lieu du groupement 3'-PO₄H₂ n'induisent donc pas les mêmes fragmentations. La perte de la base A n'est pas observée lorsque le groupement 3'-OH est présent mais cependant d'autres ions fragments sont observés. Par contre, les ions [a₂-B₂(C)H]⁻, [a₃-B₃(T)H]⁻ et w₂⁻ sont présents dans les deux cas. Le groupement 3'-PO₄H₂ semble empêcher certaines fragmentations qui sont présentes dans le cas où le groupement 3'-OH est présent.

L'ion fragment [a₃-B₃(T)H]⁻ de la perte de la thymine est présent et abondant tandis que les ions fragments [a₂-B₂(C)H]⁻, [b₄-B₂(C)H]²⁻ issus de la perte de la cytosine sont moins abondants et la perte de l'adénine n'est pas observée. La perte de la thymine ne provient pas de l'ion intermédiaire correspondant à la perte de la thymine de l'ion précurseur (comme pour la perte de la cytosine) mais plutôt directement de l'ion a₃⁻. Le nombre de nucléotides de la chimère est faible et l'ion est triplement chargée, il est donc difficile d'ajouter une autre charge négative pour former un ion zwitterion. Les bases ne peuvent donc pas être protonées contrairement au MALDI où les ions parents sont simplement chargés. La perte de la thymine devient donc compétitive à celle des autres bases.

Les deux charges négatives de l'ion b₄²⁻ semblent donc être situées sur les deux phosphodiesters entre les nucléotides B₁(G)-B₂(C) et B₃(T)-B₄(A) permettant ainsi la

formation de l'ion fragment $[a_3-B_3(T)H]^-$ et de l'ion fragment $[a_2-B_2(C)H]^-$ (figure 1.18). De plus, l'ion d_3^{2-} est doublement chargé alors que les ions a_3^- et b_3^- sont simplement chargés car ils ont perdu le phosphodiester entre les nucléotides $B_3(T)-B_4(A)$. L'ion fragment $[b_4-H-B_2(C)H]^{2-}$ n'a pas perdu son nombre de charge par rapport à celle de l'ion précurseur. La charge présente sur le phosphodiester entre $B_1(G)-B_2(C)$, favoriserait ainsi la perte du désoxynucléotide en position 5' et la formation de l'ion d_1^- .

4.3.3. Les fragmentations par spectrométrie de masse en tandem

La comparaison des fragmentations métastables et celles produites en mode CID de la chimère GCTA par MALDI-TOF/TOF a montré que les fragmentations en mode CID sont principalement des conséquences des fragmentations métastables.

Les ions fragments obtenus en mode CID par ESI-ITMSⁿ de la chimère sont : $[M-3H]^{3-}$, b_4^{2-} , w_2^- , $[M-B_4(A)H]$ doublement et triplement chargé, A^- , $[a_3-B_3(T)H]^-$ et $[a_2-B_2(C)H]^-$ tandis que les plus abondants obtenus par MALDI-TOF/TOF sont w_n^- , y_2^- et $[M-H-B_1(G)H]^-$. Les fragmentations en mode CID des oligonucléotides par ESI-ITMSⁿ donnent généralement des ions fragments $[a_n-B_nH]^-$ et w_n^- abondants (Wan et al. 2001) tandis que par MALDI ce sont les ions w_n^- et d_n^- qui sont les plus abondants. Les fragmentations en mode CID par ESI-ITMSⁿ sont donc conformes aux résultats attendus contrairement à celle par MALDI-TOF/TOF. Cette différence est due à la formation d'ions monochargés en MALDI et à la formation d'ions multichargés en ESI. Comme le nombre de nucléotides est faible, la formation d'un ion zwitterion est plus favorable au sein d'un ion simplement chargé que multichargé. En effet, l'acidité du phosphate est diminuée par la présence d'un deuxième à proximité et empêche la formation d'un ion zwitterion. La diminution de la contribution des ions zwitterions conduit aux dissociations classiques qui sont moins discriminatives que la thymine. Par contre, les nombreuses dissociations obtenues en mode CID par MALDI-TOF/TOF sont donc spécifiques aux conditions d'analyses, en haute énergie et de la formation des ions zwitterions mais pas de la chimère.

5. Conclusion

Dans ce chapitre, l'analyse en phase gazeuse de la chimère ARN/ADN par MALDI-TOF/TOF et par ESI-ITMSⁿ a été réalisée.

La chimère ARN/ADN est obtenue à partir du clivage d'un oligonucléotide qui contient deux ribonucléotides contenant la même base et des désoxynucléotides constitués des 3 autres bases complémentaires. Le clivage de la chaîne nucléotidique est réalisé à partir de l'hydroxyde de sodium qui coupe après chaque ribonucléotide. La chimère ARN/ADN est donc composée de désoxynucléotides et se termine par un ribonucléotide avec un groupement 3'-phosphate. Dans ce travail, différentes chimères ont été étudiées. Les 108 chimères de 4-mer couvrent toutes les combinaisons des 4 bases possibles. Les 12 chimères de 5-mer sont constituées d'un ribonucléotide qui contient une base A et 4 autres désoxynucléotides qui contiennent 2 bases T et 1 base C et G. Les 3 chimères de 7-mer sont constituées d'un ribonucléotide qui contient 1 base G et 6 autres désoxynucléotides qui contiennent 3 bases A, 2 bases T et 1 base C. Enfin, les chimères de 10-mer sont constituées d'un ribonucléotide qui contient une base C et 9 autres désoxynucléotides qui contiennent 3 bases A, G et T. Les chimères ont été ensuite dessalées et analysées en mode CID par spectrométrie de masse en tandem. Les conditions de clivage, de dessalage et d'analyse par MALDI-TOF MS sont les mêmes que celles utilisées dans les chapitres 3, 4 et 5 afin que la chimère ARN/ADN soit étudiée dans les mêmes conditions.

L'analyse des fragmentations en mode CID par MALDI-TOF/TOF des chimères de 4 à 10-mer a permis l'étude des chimères ARN/ADN. Les fragmentations en mode CID de la chimère déprotonée produisent beaucoup d'ions fragments permettant ainsi son séquençage. Le seul ion fragment caractéristique de la chimère ARN/ADN est l'ion a_4^- des chimères de 4-mer qui sont constitués d'un ribonucléotide C. Par ailleurs, plus le nombre de nucléotides de la chimère augmente plus des ions w_n^- , $[a_n-B_nH]^-$, $[M-H-BH]^-$ et $[w_n-[a_n-B_nH]]^-$ sont abondants et spécifiques de la séquence. Les fragmentations en mode CID par MALDI-TOF/TOF devient alors comparable à celle par l'ESI-ITMSⁿ. Les nombreuses fragmentations observées en mode CID par MALDI-TOF/TOF sont expliquées d'une part, par les conditions de collision en haute énergie et d'autre part, par la formation des ions zwitterions qui facilitent la perte de la base.

L'étude des fragmentations en mode CID par MALDI-TOF/TOF est donc une méthode rapide et efficace pour réaliser le séquençage complet des chimères ARN/ADN. Cependant, la chimère doit être suffisamment concentrée afin de détecter l'ion parent et d'identifier les

masses des ions fragments sans ambiguïté. Cette méthode est donc limitée à l'étude des chimères de 10-mer pour avoir une couverture complète de leur séquence car au-delà le pouvoir résolutif de l'instrument n'est pas suffisant. Pour l'analyse des échantillons des chapitres suivants, cette méthode ne pourra être utilisée que si les échantillons sont assez concentrés et/ou que les chimères contiennent moins de 10 bases. De plus, la stabilité des bases T en phase gazeuse améliore la détection des chimères ARN/ADN qui les contiennent. Les chimères ne contenant pas de thymine doivent donc être analysées à des concentrations plus élevées.

Enfin, les fragmentations en phase gazeuse de la chimère GCTA ont été étudiées. L'étude des fragmentations par MALDI-TOF/TOF de la chimère GCTA a montré que les fragmentations en mode CID sont principalement des fragmentations métastables et la présence d'ions zwitterions favorisent ainsi l'abondance des ions issus de la perte de la base ainsi que les ions w_n^- . L'étude des fragmentations en mode CID de la chimère GCTA par ESI-ITMSⁿ a permis d'analyser la séquence de la chimère et de localiser des charges des ions multichargées. Les fragmentations de la chimère par ESI-IT MSⁿ donnent des ions abondants b_4^{2-} , w_n^- , $[a_n-B_nH]^-$ et $[M-H-B_nH]^-$ tandis que par MALDI-TOF/TOF les ions abondants sont w_n^- , y_2^- et $[M-H-B_nH]^-$. Les ions fragments obtenus par MALDI-TOF-TOF sont beaucoup plus nombreux que ceux obtenus par ESI-IT MSⁿ. L'abondance des ions fragments de la chimère par ESI-ITMSⁿ est comparable à celle des oligonucléotides de la littérature : w_n^- et $[a_n-B_nH]^-$. L'ion b_4^{2-} est caractéristique de la présence du groupement 3'-phosphate de la chimère ARN/ADN.

Les fragmentations en mode CID par ESI-ITMSⁿ ont été réalisées sur une seule chimère de 4-mer. Une étude comparable à celle réalisée par MALDI-TOF/TOF pourrait également être réalisée par ESI-ITMSⁿ

Chapitre 2

Multiplex microhaplotypage par clivage d'une chimère ARN/ADN simple-brin et par MALDI-TOF MS

Chapitre 2. Multiplex microhaplotypage par clivage d'une chimère ARN/ADN simple-brin et par MALDI-TOF MS

1. Introduction

En 2006, une méthode de génotypage des SNPs par clivage d'une chimère ARN/ADN simple-brin et par spectrométrie de masse MALDI-TOF a été mise au point (Mauger et al. 2006). Dans cette méthode (annexe 1), une chimère simple-brin d'ARN/ADN est synthétisée à l'aide d'une amorce contenant un ribonucléotide dans les dernières bases du côté 3' terminal et d'une ADN polymérase qui incorpore des NTPs. Cette amorce est adjacente au SNP étudié. La chimère ARN/ADN est ensuite clivée après chaque ribonucléotide. Il en résulte un fragment de clivage qui contient des désoxynucléotides et un seul ribonucléotide à l'extrémité 3' qui contient le SNP a étudié. Il est ensuite analysé par spectrométrie de masse MALDI-TOF MS en mode linéaire.

Dans ce chapitre, une méthode de multiplex microhaplotypage par clivage d'une chimère ARN/ADN simple-brin et d'analyse par spectrométrie de masse MALDI-TOF est développée. Elle est basée sur la méthode précédente (Mauger et al. 2006) mais permet en plus l'étude de plusieurs SNPs dans une région très polymorphe. La méthode de microhaplotypage utilise plusieurs amorces d'élongation linéaire afin d'analyser plusieurs SNPs consécutifs du même allèle sous forme de microhaplotype. Les amorces sont toujours adjacentes aux SNPs à étudier mais elles contiennent également un second SNP qui est proche du premier.

Par ailleurs, la méthode du *GOODASSAY* utilise également plusieurs amorces d'élongation pour l'étude des microhaplotypes. Dans cette méthode, les amorces contiennent une *Charge Tag* qui attribue à chaque microhaplotype étudié une masse unique (Kucharzak 2006).

Dans la méthode présentée ici, le clivage de la chimère ARN/ADN génère le microhaplotype à étudier. Des amorces permettent d'étudier différents loci dans une même réaction d'élongation à condition qu'elles soient situées dans le locus amplifié préalablement par PCR. La méthode est composée de cinq étapes : (i) une PCR, (ii) une élongation linéaire à l'aide de plusieurs amorces, (iii) un clivage chimique, (iv) un dessalage et (v) l'analyse par MALDI-TOF MS. L'analyse en MALDI-TOF MS des masses des fragments de clivage donne

directement le microhaplotype de l'individu. La méthode est mise au point sur l'étude de microhaplotype du système *HLA* (antigènes des leucocytes humains).

Le système *HLA* est le complexe majeur d'histocompatibilité (*CMH*) chez l'humain (Nelson & Fremont 1999). Il est impliqué dans l'identification du système immunitaire. Les gènes qui codent pour le système *HLA* sont localisés sur le bras court du chromosome 6 (6p21.3) et contiennent environ 4 à 6 Mb. Il y a deux classes de molécules de *CMH*. La molécule de *CMH* de classe I est composée d'une chaîne α constituée de trois domaines extracellulaires, d'une petite chaîne polypeptidique et de la β -2-microglobuline. Elle est présente à la surface de toutes les cellules, permettant ainsi la reconnaissance de la cellule par son propre organisme. Elle est impliquée dans la présentation des peptides principalement dérivés des protéines intracellulaires des lymphocytes T CD8+ cytotoxiques. Les lymphocytes T8 ciblent les cellules infectées par un virus ou un agent pathogène intracellulaire. Les gènes *HLA-A*, *B* et *C* du *CMH* de classe I sont situés sur le télomère et contiennent environ 2 Mb. Ils codent pour la chaîne α de l'antigène. Ces gènes font également partie de la famille des gènes d'immunoglobulines. La molécule *CMH* de classe II est composée de deux chaînes α et β constituées chacune par deux domaines extracellulaires qui forment des hétérodimères non-covalents de la glycoprotéine membranaire. Elle est présente uniquement à la surface des cellules présentatrices de l'antigène soit essentiellement les macrophages. Elle contrôle à la fois la sélection et l'activation périphérique intrathymique des lymphocytes T CD4+ (Pan et al. 1998). Les gènes *HLA-DP*, *DQ* et *DR* du *CMH* de classe II sont situés en grande partie sur le centromère. Ils codent pour les chaînes α et β .

Le système *HLA* est très polymorphe et de très dense. Une base de données IMGT/*HLA* sur internet (www.ebi.ac.uk/imgt/hla) répertorie donc tous ces haplotypes. Dans la version 3.8 datant de Juillet 2011, 7527 allèles du gène *HLA* sont listés dont 5880 allèles de classe I et 1647 allèles de classe II. Il y a donc un grand nombre de combinaisons théoriques possibles pour chaque phénotype. De plus, les différents phénotypes de *HLA* sont plus ou moins fréquents suivants le groupe ethnique.

Le système *HLA* intervient d'une part, dans le système immunitaire de l'organisme et d'autre part, permet de réaliser des greffes d'organes. Pour les greffes, le donneur et le receveur doivent avoir une compatibilité la plus proche possible afin d'éviter les rejets. Le typage du gène *HLA* est donc utilisé afin de trouver l'individu le plus compatible pour une greffe

d'organe. Son étude nécessite de développer de nouvelles méthodes d'analyse des haplotypes (Albert et al. 1984; Middleton 1999; Saiki et al. 1985; Erlich et al. 1991; Erlich et al. 2001).

Cette méthode a été développée en utilisant une sélection de microhaplotypes de la méthode du *GOODASSAY* (Kucharzak 2006). Le microhaplotype ASTW du locus *HLA-DRB1-227*, le microhaplotype HCACH du locus *HLA-A-98* et le microhaplotype MGAR du locus *HLA-A-453* ont été validés sur des individus témoins.

2. Matériels et Méthodes

2.1. Matériels

Les échantillons d'ADN utilisés dans cette méthode proviennent de la collection du CEPH (www.cephb.fr) des familles suivantes : 1333, 1329-2 1416 et 1418. Elles ont été choisies en fonction de leurs haplotypes. Ces individus ont été analysés préalablement par la méthode de microhaplotypage du *GOODASSAY* (Kucharzak 2006).

Les amorces de PCR ont été synthétisées par BIOTEZ (Berlin, Allemagne) tandis que celles de l'élongation linéaire par Eurogentec (Angers, France). Les séquences des amorces sont listées dans le tableau 2.1.

| <i>HLA</i> | Amorces de PCR | Amorces d'élongation |
|-----------------------------|--|---|
| <i>HLA-A-98</i> HCACH | GGCCTCTGYGGGGAGAAGCAA GTCCCAATTGTCTCCCCTCCTT | GGCCGGGACACGGG <u>A</u> GGTG GGCCGGGACACGGAT <u>T</u> GTG GGCCGGGACACGG <u>A</u> AGTG |
| <i>HLA-A-453</i> MGAR | GGCCTCTGYGGGGAGAAGCAA GTCCCAATTGTCTCCCCTCCTT | GTCCAAGAGCGCAGGT <u>C</u> TTC GTCCAAGAGCGCAGGT <u>C</u> CCTC GTCCAAGAGCTCAGGT <u>C</u> CCTC |
| <i>HLA-DRB1-227</i> ASTW | TTCTTGTGGSAGCTTAAGTT TTCCTGTGGCAGCCTAAGAGG CACGTTTCTTGGAGTACTCTAB CGTTTCTTGGAGTACTCTACGGG CGTTTCTTGGAGTACTCTACGTC GTTTCTTGGAGCAGGTAAAC CCTGTGGCAGGGTAARTATA CCCAACCACGTTTCTTGA AGACCACGTTTCTTGGAGG TCGCCGCTGCACYGTGAA | CGACAGCGACGTGGGG <u>G</u> ACT CGACAGCGACGTGVGG <u>G</u> AGT |

Tableau 2.1 : Amorces de PCR et d'élongation des microhaplotypes *HLA-A-98*, *HLA-A-453* et *HLA-DRB1-227*. Les bases qui sont en gras et soulignées sont des bases ribonucléotides.

Les dNTPs (N=A, C, G et T) et les NTPs (N=A, C, G et U) proviennent de GE Healthcare (Saclay, France) et les produits chimiques d'Aldrich (Steinheim, Allemagne). La résine échangeuse de cations provient de Biorad (Marnes la coquette, France) et le kit *Clean Resin Tool* de *MassArray*[®] de chez Sequenom (Hamburg, Allemagne). Les étalons externes d'oligonucléotides de 4 à 11-mer proviennent de chez BRUKER (Bremen, Allemagne). L'ADN *Taq* polymérase *High-Fidelity Platinum*TM provient d'Invitrogen (Karlsruhe, Allemagne) et l'enzyme *G46E CS6R* a été développée par Roche Molecular Systems (Pleasanton, CA, USA).

L'ADN polymérase *G46E CS6R* n'est pas commercialisée, elle peut être obtenue par Dr Thomas Myers (thomas.myers@roche.com). Cette chimère d'ADN polymérase contient un domaine 5'-nucléase de l'ADN polymérase *Thermus sp. ZO5* et les domaines 3'-nucléase et 5'-3'-nucléase de l'ADN polymérase *Thermotoga maritima (Tma)*. L'activité 5'-nucléase a été éliminée par la mutation G46E (Mauger et al. 2006). L'activité *haute-fidélité* de l'ADN polymérase non-mutée *Tma* a été supprimée par l'introduction de la mutation du 'Motif I' du domaine 3'-5' exonucléase. L'enzyme contient une mutation E678G qui élimine la spécificité

de l'enzyme non-mutée contre l'incorporation des nucléotides substitués en position 2'-C, favorisant ainsi l'incorporation des ribonucléosides triphosphates (Gelfand 1999).

Les procédures de thermocyclage ont été réalisées sur des thermocycleurs gradients Eppendorf (Eppendorf, Allemagne). Les analyses ont été réalisées par un MALDI-TOF MS *AutoflexTM* de chez Bruker (Bremen, Allemagne).

2.2. Méthodes

2.2.1. PCR

Les gènes *HLA-DRB1* et *HLA-A* sont amplifiés. (Kucharzak 2006).

2.2.1.1. PCR *HLA-DRB1*

Le mélange de 50 μL de la solution de PCR *HLA-DRB1* contient 4,2 μL de tampon d'enzyme de PCR, 1,7 μL de sulfate de magnésium à 50 mM, 1,25 μL de dNTPs à 4 mM, 1,5 μL de chaque amorce de PCR (tableau 2.1 ; 3 μL pour GGCCTCTGYGGGGAGAAGCAA) à 10 pmol/ μL , 1 μL d'enzyme *High-Fidelity PlatinumTM* à 5 U/ μL , 20 μL d'ADN à 10 ng/ μL et 6,97 μL d'eau milli-Q. La réaction de PCR est dénaturée pendant 3 min à 95°C, puis suivie de 30 cycles de 20 s à 95°C, 20 s à 58,5°C et 20 s à 72°C et se termine par une élongation de 5 min à 72°C.

2.2.1.2. PCR *HLA-A*

Le mélange de 50 μL de la solution de PCR *HLA-A* contient 5 μL de tampon d'enzyme de PCR, 1,5 μL de sulfate de magnésium à 50 mM, 0,63 μL de dNTPs à 4 mM, 7,5 μL de chaque amorce de PCR (tableau 2.1) à 7,5 pmol/ μL , 0,90 μL d'enzyme *High-Fidelity PlatinumTM* à 5 U/ μL , 20 μL d'ADN à 10 ng/ μL et 6,97 μL d'eau milli-Q. La réaction de PCR est dénaturée pendant 3 min à 94°C, puis suivie de 30 cycles de 40 s à 94°C, 45 s à 70°C et 1 min à 72°C et se termine par une élongation de 4 min à 72°C.

2.2.2. L'élongation linéaire par l'ADN polymérase *G46E CS6R*

2.2.2.1. Le locus *HLA-DRB1-227*

Le mélange de 10 μL de la solution d'élongation de *HLA-DRB1-227* contient 0,5 μL de tris-HCl (pH 8,3) à 200 mM, 0,5 μL de chlorure d'ammonium à 500 mM, 0,5 μL de chlorure de magnésium à 50 mM, 1 μL de NTPs à 4 mM, 0,5 μL de chaque amorce d'élongation (tableau 2.1) à 20 pmol/ μL , 0,25 μL d'enzyme *G46E CS6R* 4 U/ μL , 5 μL de solution de PCR *HLA-DRB1* et 0,75 μL d'eau milli-Q.

2.2.2.2. Le duplex des loci *HLA-A-98* et *HLA-A-453*

Le mélange de 10 μL de la solution d'élongation du duplex *HLA-A* contient 0,5 μL de tris-HCl (pH 8,3) à 200 mM, 0,5 μL de chlorure d'ammonium à 500 mM, 0,5 μL de chlorure de magnésium à 50 mM, 1 μL de NTPs à 4 mM, 0,25 μL de chaque amorce d'élongation (tableau 2.1) à 20 pmol/ μL , 0,25 μL d'enzyme *G46E CS6R* 4 U/ μL , 5 μL de solution de PCR *HLA-A* et 0,75 μL d'eau milli-Q.

Les réactions d'élongation sont dénaturées pendant 2 min à 95°C, puis suivies de 40 cycles de 15 s à 95°C, 2 min à 65°C et 20 s à 72°C et se terminent par une élongation de 4 min à 72°C.

2.2.3. Le clivage

Au 10 μL de la solution d'élongation est ajouté un mélange de 1,5 μL d'hydroxyde de sodium à 3 M et 3,5 μL d'eau milli-Q. La réaction de clivage est mise sous incubation pendant 1h30 à 70°C.

2.2.4. Le dessalage

A la réaction de clivage, 6 mg de résines échangeuses de cations sont ajoutées à l'aide du kit *Clean Resin Tool* de *MassARRAY*[®]. Les échantillons sont incubés pendant 20 min à

température ambiante sous agitation et ensuite centrifugés sous 134 xg afin de récupérer le surnageant.

2.2.5. L'analyse par spectrométrie de masse MALDI-TOF

2.2.5.1. La préparation de la cible MALDI

La solution de matrice α -cyano-4-hydroxycinnamic (HCCA) est saturée dans l'acétone. Un mélange de 5 μ L du surnageant, 7,5 μ L de 60 % d'acétonitrile et de 2,5 μ L de solution de matrice est préparé (Mauger et al. 2006).

Sur une cible MALDI, constituée d'un alliage d'aluminium et de nickel de 384 spots, est déposé 1 μ L du mélange de l'analyte-matrice par spot. La cible est séchée à température ambiante.

2.2.5.2. La spectrométrie de masse MALDI-TOF

La cible MALDI est introduite dans un MALDI-TOF MS. L'analyse est réalisée en mode linéaire, négatif à l'aide du logiciel *FlexControl* et des paramètres du MALDI-TOF MS suivants : IS1=20 kV ; IS2=18,5 kV ; lens=7,5 kV et un délai d'extraction de 200 ns. Les spectres de masse sont obtenus par l'accumulation de 500 tirs de laser.

Un étalonnage externe est réalisé à l'aide du kit d'étalon externe d'oligonucléotides de 4 à 11-mer de chez BRUKER.

Ces conditions permettent d'analyser des masses moyennes et de détecter des échantillons d'environ 0,7 pmol.

Les spectres de masse sont analysés à l'aide du logiciel *FlexAnalysis*. L'annotation des pics caractérisés par le rapport m/z des ions, la soustraction de la ligne de base ainsi qu'un lissage du signal sont réalisés.

3. Résultats

Le principe de la méthode est illustré dans la figure 2.1.

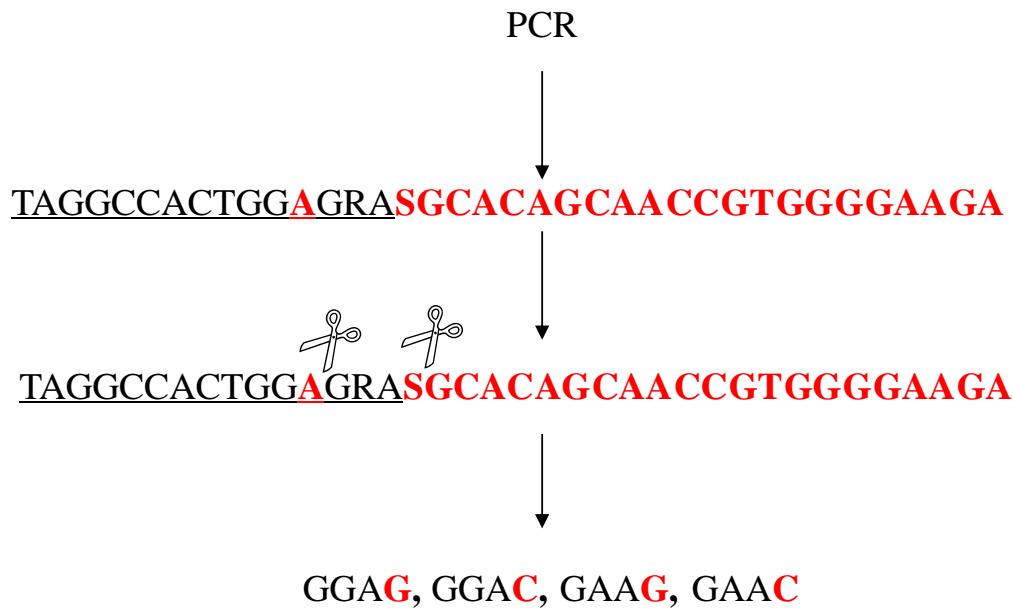


Figure 2.1 : Principe de la méthode de microhaplotypage pour l'étude d'un microhaplotype GRAS : l'amplification de la séquence à étudier par PCR, la synthèse de la chimère ARN/ADN simple-brin, le clivage de la chimère ARN/ADN, l'analyse des microhaplotypes GGAG, GGAC, GAAG ou GAAC par MALDI-TOF MS. Les bases ADN sont en noir tandis que les bases ARN sont en gras et en rouge. Les amorces d'élongation sont soulignées.

La méthode est constituée de cinq étapes : (i) une PCR, (ii) une élongation linéaire, (iii) un clivage chimique, (iv) un dessalage et (v) l'analyse des masses des microhaplotypes par MALDI-TOF MS.

3.1. La PCR

Une PCR est réalisée à l'aide de plusieurs amorces de PCR afin d'amplifier la région d'intérêt ainsi que tous les allèles *HLA* présents dans la base de données IMGT-*HLA* sur internet (www.ebi.ac.uk/imgt/hla). Cette base de données est mise à jour par le groupe de Bioinformatique de Steven Marsh (Antony Nolan Trust, London, UK).

Les deux PCR étudiées *HLA-A* et *HLA-DRB1* ont été mises au point par rapport à la base de données de Janvier 2005 et pour la méthode du *GOODASSAY* (Kucharzak 2006). Les microhaplotypages étudiés servent uniquement pour valider la méthode en les comparant aux

résultats obtenus par la méthode du *GOODASSAY* réalisée dans le laboratoire (Kucharzak 2006).

3.2. La chimère ARN/ADN simple-brin

L'ADN polymérase *G46E CS6R* n'incorpore que les NTPs, il n'est donc pas nécessaire d'enlever les dNTPs en excès provenant de la PCR. L'élongation linéaire s'effectue donc à l'aide de cette enzyme, des quatre NTPs et de plusieurs amorces d'élongation.

Ces amorces d'élongation comportent une seul ribonucléotide et sont adjacentes à la base du SNP étudié dans le microhaplotype. Elles contiennent également un autre SNP qui est situé près de celui à étudier. Elles ont été choisies pour couvrir tous les microhaplotypes possibles de la base de données IMGT-HLA datant de Janvier 2005 pour la méthode du *GOODASSAY* (Kucharzak 2006). La position du ribonucléotide contenue dans les amorces varie afin d'analyser plusieurs microhaplotypes dans une même réaction. Les masses des microhaplotypes sont donc choisies en fonction de leurs séquences mais aussi en fonction de la masse des amorces d'élongation qui ne réagissent pas. Les amorces qui n'ont pas subi d'élongation pendant la réaction doivent avoir une masse différente de celle des microhaplotypes.

Pour le locus *HLA-DRB1*, un seul microhaplotype a été étudié *ASTW (HLA-DRB1-227)* à l'aide de deux amorces d'élongation (tableau 2.1).

Pour le locus *HLA-A*, deux microhaplotypes *HCACH (HLA-A-98)* et *MGAR (HLA-A-453)* ont été étudiés sous forme d'un duplex à partir de la même PCR de *HLA-A* (tableau 2.1). Le microhaplotype *HCAH* du locus *HLA-A-98* a été analysé à partir de deux amorces d'élongation tandis que le microhaplotype *MGAR* du locus *HLA-A-453* a été analysé à partir de trois amorces d'élongation afin de couvrir tous les microhaplotypes présents.

La chimère ARN/ADN simple-brin est donc synthétisée d'une part, à partir de l'amorce d'élongation contenant un seul ribonucléotide et plusieurs désoxynucléotides et d'autre part, des ribonucléotides qui ont été incorporées par l'ADN polymérase *G46E CS6R* durant l'élongation linéaire.

3.3. Le clivage de la chimère ARN/ADN

Le clivage est réalisé par l'hydroxyde de sodium à 0,3 M durant 1h30 à température ambiante (Mauger et al. 2006). Cette réaction est rapide, simple à mettre en œuvre et totale.

L'ion hydroxyde attaque chaque ribonucléotide car il possède un 2'-OH sur le ribose. Il en résulte une coupure en 3' de la chaîne nucléotidique au niveau du phosphodiester adjacente au ribonucléotide.

Les fragments de clivage sont composés de désoxyribonucléotides et à l'extrémité 3' d'un seul ribonucléotide et d'un groupement 3'-phosphate. Chaque microhaplotype ainsi formé contient un seul ribonucléotide qui contient le SNP incorporé et l'autre SNP contenu dans l'amorce délongation

3.4. Le dessalage des fragments de clivage

Les microhaplotypes sont formés à partir du clivage par l'hydroxyde de sodium. Les ions sodium, présents dans les échantillons, ne favorisent pas une bonne désorption des ions en phase gazeuse du MALDI. Ils doivent donc être enlevés des échantillons.

L'utilisation de la résine échangeuse de cations est une bonne méthode de dessalage car les échantillons ne peuvent pas être perdus pendant le dessalage contrairement à d'autres méthodes de purification comme l'alkylation. Elles permettent d'échanger les ions sodium de l'échantillon avec les protons présents sur les billes échangeuses de cations.

Dans la méthode de *MassARRAY*[®], les billes échangeuses de cations (ammonium) sont utilisées (Ehrich et al. 2005) car l'ion ammonium facilite une bonne désorption en MALDI. Dans cette méthode, les billes échangeuses de cations (proton) sont utilisées. Le clivage par l'hydroxyde de sodium à 0,3M a entraîné un pH très basique ce qui est défavorable à une bonne cristallisation des échantillons avec la matrice HCCA acide. Le pH doit donc être diminué par l'addition d'acide. La résine échangeuse de cations permet d'une part, d'enlever les ions sodium présents dans les échantillons et d'autre part, de diminuer le pH de la solution.

3.5. L'analyse des spectres de masse MALDI-TOF

3.5.1. La matrice HCCA

La matrice HCCA est une matrice couramment utilisée pour l'analyse des peptides jusqu'à 12 kDa (Beavis & Chait 1989), elle peut être également utilisée pour l'étude des petits oligonucléotides (Mauger et al. 2006). Cette matrice a une faible basicité, les ions zwitterions sont donc faiblement protonés. Il n'y a donc pas de fragmentation en phase gazeuse. De plus, elle possède une cristallisation homogène qui facilite la détection des échantillons et l'automatisation de la mesure en MALDI-TOF MS.

3.5.2. Les paramètres d'analyse

L'analyse s'effectue en mode négatif car les fragments de clivage de la chimère ARN/ADN sont naturellement chargés négativement grâce aux atomes d'oxygène des groupements phosphates.

L'analyse s'effectue en mode linéaire par l'analyse des masses moyennes des microhaplotypes. Le mode réflecteur n'est pas nécessaire car les masses des microhaplotypes sont suffisamment distinctes pour être détectées en mode linéaire.

3.5.3. Les microhaplotypes

La formule suivante permet de calculer les masses moyennes des microhaplotypes :

$$M = 313,20658 * N(A) + 289,18188 * N(C) + 329,20598 * N(G) + 304,19322 * N(T) + 291,16621 * N(U) + 34,01468$$

Les masses des microhaplotypes des locus *HLA-A-98*, *HLA-A-453* et *HLA-DRB1* sont indiquées dans le tableau 2.2.

| HLA | Microhaplotype | Fragment | Masse |
|-----------------------------|-----------------------|-----------------|--------------|
| <i>HLA-A-98</i> HCACH | ACACC | GGTGU | 1616,0 |
| | TCACA | GTGA | 1309,8 |
| | CCACA | GTGG | 1325,8 |
| | ACACT | AGTGU | 1600,0 |
| <i>HLA-A-453</i> MGAR | AGAA | TTCU | 1221,7 |
| | CGAG | CTCG | 1245,8 |
| | AGAG | CTCU | 1206,7 |
| | CGAG | CTCG | 1245,8 |
| | AGAG | CTCU | 1206,7 |
| <i>HLA-DRB1-227</i> ASTW | ACTA | ACTA | 1253,8 |
| | AGTA | AGTA | 1293,8 |
| | AGTT | AGTU | 1270,7 |

Tableau 2.2 : Masses moyennes des microhaplotypes étudiés HCACH pour le locus *HLA-A-98*, MGAR pour le locus *HLA-A-453* et ASTW pour le locus *HLA-DRB1-227*.

Les masses moyennes des microhaplotypes ont été choisies pour être dans une gamme de masse facilement détectable par MALDI-TOF MS. Ils sont constitués de 4 ou 5 nucléotides.

3.5.4. Le microhaplotype ASTW du locus *HLA-DRB1-227*

Les spectres de masse du microhaplotype ASTW du locus *HLA-DRB1-227* sont présentés dans la figure 2.2.

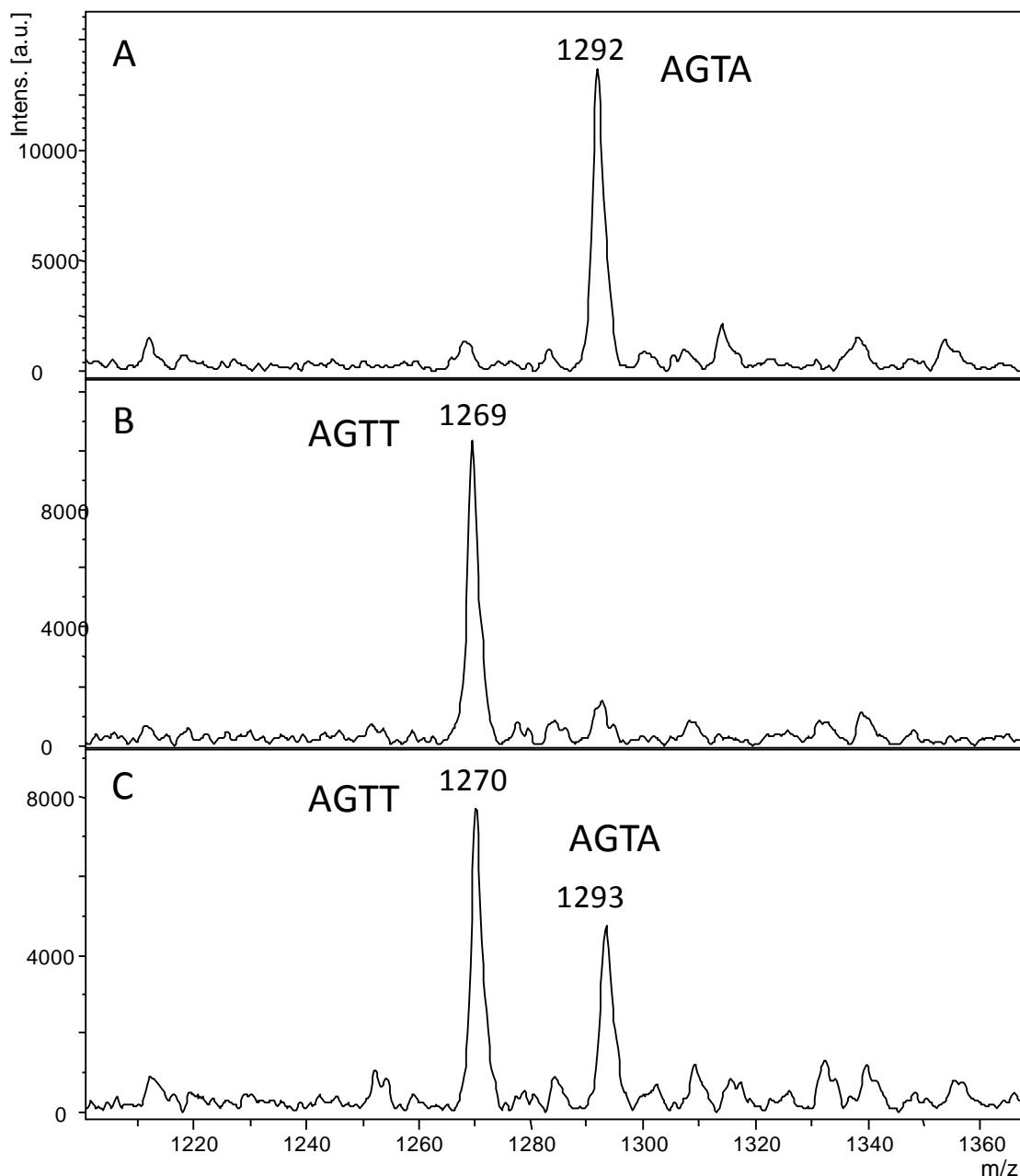


Figure 2.2 : Spectres de masse MALDI-TOF de la région des 4-mer des fragments sens $[M-H]^-$ du microhaplotype ASTW du locus *HLA-DRB1-227* d'un microhaplotype : (A) homozygote AGTA ; (B) homozygote AGTT et (C) hétérozygote AGTA et AGTT. L'ion $[M-H]^-$ AGTA représente le microhaplotype AGTA et l'ion $[M-H]^-$ AGTU représente le microhaplotype AGTT.

Les spectres de masse montrent à m/z 1293 l'ion $[M-H]^-$ du fragment AGTA ($M=1293,8$ Da) du microhaplotype AGTA et à m/z 1269 l'ion $[M-H]^-$ du fragment AGTU ($M=1270,7$ Da) du microhaplotype AGTT.

Le microhaplotype homozygote AGTA (figure 2.2.A), le microhaplotype homozygote AGTT (figure 2.2.B) ainsi que les microhaplotypes hétérozygotes AGTA et AGTT (figure 2.2.C) du microhaplotype ASTW du locus *HLA-DRB1-227* sont identifiés par cette méthode.

3.5.5. Les microhaplotypes HCACH du locus *HLA-A-98* et MGAR du locus *HLA-A-453*

Les spectres de masse du duplex des microhaplotypes HCACH et MGAR des loci *HLA-A-98* et *HLA-A-453* sont présentés dans la figure 2.3.

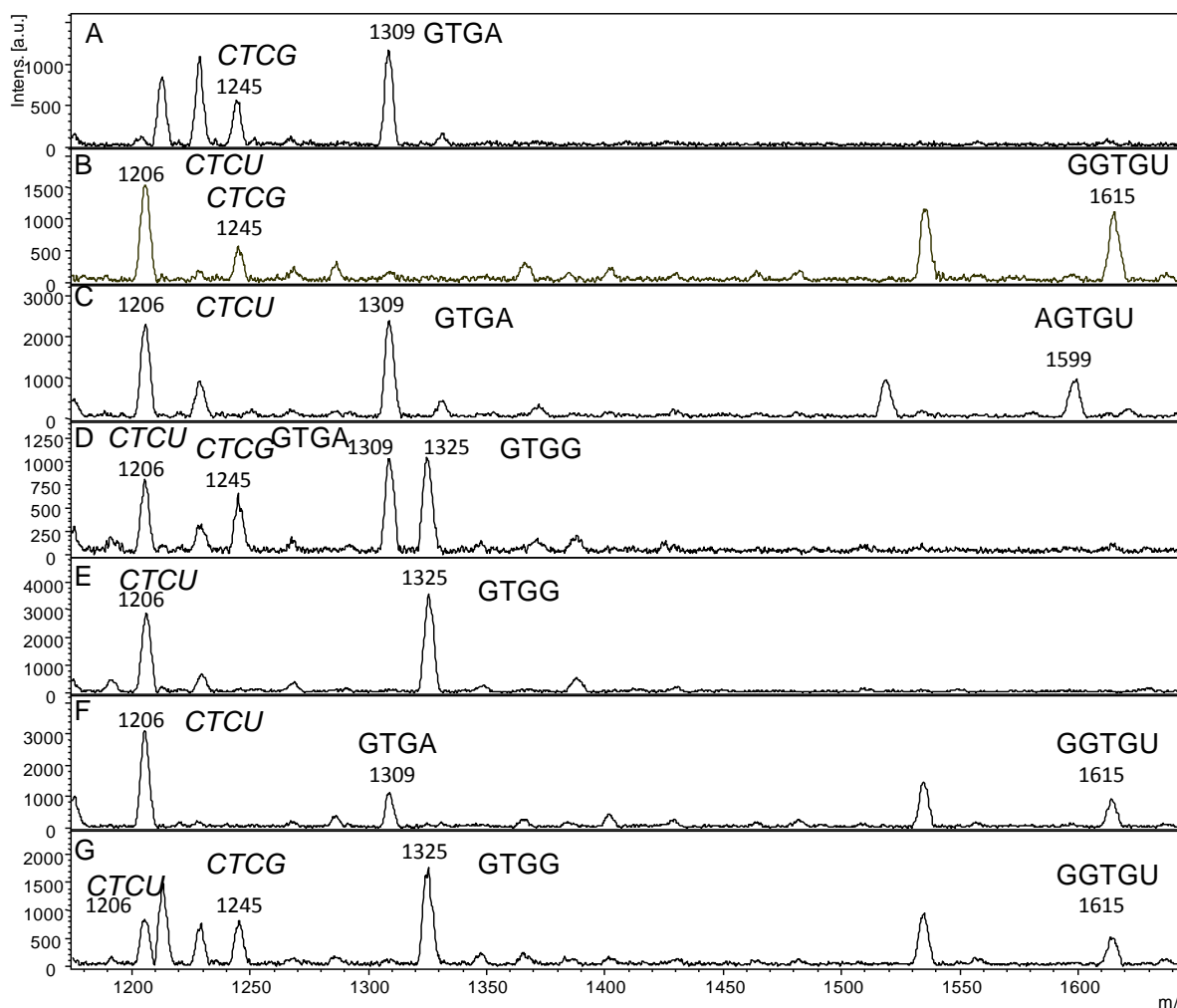


Figure 2.3 : Spectres de masse MALDI-TOF de la région des 4 à 5-mer des fragments sens $[M-H]^-$ du microhaplotype HCACH du locus *HLA-A-98* et des fragments anti-sens $[M-H]^-$ du microhaplotype MGAR du locus *HLA-A-453* d' : (A) un homozygote TCACA et CGAG ; (B) un homozygote ACACC et un hétérozygote MGAG ; (C) un hétérozygote TACACA, AACT et homozygote AGAG ; (D) un hétérozygote YCACA et hétérozygote AGAG, CGAG ; (E) un homozygote CCACA et AGAG ; (F) un hétérozygote ACACA, TCACC et homozygote AGAG et (G) un hétérozygote ACACC, CCACA et hétérozygote MGAG. Les ions $[M-H]^-$ GTGA, GTGG, AGTU et GGTGU sont les microhaplotypes de HCACH tandis que les ions $[M-H]^-$, en italique, CTCU et CTCG sont les microhaplotypes de MGAR.

Pour le microhaplotype HCACH du locus *HLA-A-98*, les spectres de masse montrent à m/z 1309 l'ion $[M-H]^-$ du fragment GTGA ($M=1309,8$ Da) du microhaplotype TCACA, à m/z 1325 l'ion $[M-H]^-$ du fragment GTGG ($M=1325,8$ Da) du microhaplotype CCACA, à m/z 1599 l'ion $[M-H]^-$ du fragment AGTGU ($M=1600$ Da) du microhaplotype ACACT et à m/z 1615 l'ion $[M-H]^-$ du fragment GGTGU ($M=1616,0$ Da) du microhaplotype ACACC.

Pour le microhaplotype MGAR du locus *HLA-A-453*, les spectres de masse montrent à m/z 1206 l'ion $[M-H]^-$ du fragment CTCU ($M=1206,7$ Da) du microhaplotype AGAG et à m/z 1245 l'ion $[M-H]^-$ du fragment CTCG ($M=1245,8$ Da) du microhaplotype AGAG.

Un individu homozygote TCACA et homozygote CGAG (figure 2.3.A), un individu homozygote ACACC et hétérozygote MGAG (figure 2.3.B), un individu hétérozygote TCACA, ACACT et homozygote AGAG (figure 2.3.C), un individu hétérozygote YCACA et hétérozygote MGAG (figure 2.3D), un individu homozygote CCACA et homozygote AGAG (figure 2.3E), un individu hétérozygote ACACA, TCACC et homozygote AGAG (figure 2.3.F) et un individu hétérozygote ACACC, CCACA et hétérozygote MGAG (figure 2.3.G) du duplex des microhaplotypes HCACH du locus *HLA-A-98* et MGAR du locus *HLA-A-98* sont analysés par cette méthode.

3.6. L'analyse des individus

La méthode a été validée pour l'analyse du microhaplotype ASTW du locus *HLA-DRB1-227* sur les 23 individus témoins du tableau 2.3 des familles CEPH : 1333, 1416 et 1418.

| Individus | HLA-A-98 | HLA-A-453 | HLA-DRB1-227 |
|-----------|---------------|-------------|--------------|
| 1333 01 | | | AGTA |
| 1333 02 | TCACA | CGAG | AGTA |
| 1333 04 | TCACA - CCACA | CGAG - AGAG | AGTA |
| 1333 09 | | | AGTA |
| 1333 11 | | | AGTA - AGTT |
| 1333 12 | | | AGTA |
| 1333 13 | | | AGTA - AGTT |
| 1333 14 | | | AGTA |
| 1329-2-01 | TCACA - CCACA | CGAG | |
| 1329-2-02 | TCACA - ACACT | CGAG - AGAG | |
| 1329-2-10 | TCACA - CCACA | CGAG - AGAG | |
| 1329-2-11 | CCACA - ACACT | CGAG | |
| 1329-2-12 | TCACA - ACACT | AGAG | |
| 1329-2-13 | TCACA | CGAG - AGAG | |
| 1416 01 | | | AGTA - AGTT |
| 1416 02 | ACACC - TCACA | AGAG | AGTA - AGTT |
| 1416 05 | | | AGTA - AGTT |
| 1416 06 | | | AGTA - AGTT |
| 1416 11 | CCACA | AGAG | AGTA - AGTT |
| 1416 12 | | | AGTA - AGTT |
| 1416 13 | ACACC | CGAG - AGAG | |
| 1416 14 | | | AGTA |
| 1416 15 | | | AGTA - AGTT |
| 1418 01 | | | AGTA - AGTT |
| 1418 02 | ACACC - CCACA | CGAG - AGAG | AGTT |
| 1418 03 | | | AGTA - AGTT |
| 1418 04 | | | AGTT |
| 1418 12 | | | AGTT |
| 1418 13 | | | AGTT |
| 1418 14 | | | AGTT |

Tableau 2.3 : Microhaplotypes des loci HLA-A 98 et 453 et HLA-DRB1 227 des individus de la collection CEPH.

Les résultats obtenus sont les mêmes que ceux attendus : 7 individus de microhaplotype homozygote AGTA, 5 individus de microhaplotype homozygote AGTT et 11 individus de microhaplotype hétérozygote AGTA et AGTT.

La méthode de microhaplotypage a été également validée pour l'analyse en duplex des microhaplotypes HCACH et MGAR des loci HLA-A (98 et 453) sur les 12 échantillons

d'ADNs témoins du tableau 2.3 des familles CEPH : 1333, 1416, 1329-2, 1416 et 1418. Les résultats obtenus sont donc les mêmes que ceux attendus.

Pour le microhaplotypage HCACH du locus *HLA-A-98*, il y a 2 individus homozygotes TCACA, 1 individu homozygote CCACA, 1 individu homozygote ACACC, 3 individus hétérozygotes TCACA CCACA, 2 individus hétérozygotes TCACA ACACT, 1 individu hétérozygote CCACA ACACT, 1 individu hétérozygote ACACC TCACA et 1 individu hétérozygote ACACC CCACA (tableau 2.3).

Pour le microhaplotype MGAR du locus *HLA-A-453*, il y a 3 individus homozygotes CGAG, 3 individus homozygotes AGAG, 6 individus hétérozygotes CGAG et AGAG (tableau 2.3).

4. Discussion

Dans ce chapitre, une nouvelle méthode de multiplex microhaplotypage qui est basée sur la méthode de génotypage des SNPs (Mauger et al. 2006) utilisant le clivage d'une chimère ARN/ADN et analyse par MALDI-TOF MS, a été développée. Elle diffère de la méthode de génotypage des SNPs par l'analyse de plusieurs SNPs dans la même réaction permettant ainsi d'analyser des microhaplotypes. Pour cela, plusieurs amorces d'élongation sont utilisées contrairement au protocole de génotypage des SNPs.

4.1. La PCR

Les gènes *HLA-A* et *HLA-DRB1* ont été amplifiés à l'aide d'une ADN polymérase couramment utilisée et de plusieurs amorces de PCR qui sélectionnent tous les allèles *HLA* présents dans la base de données IMGT-HLA sur internet (www.ebi.ac.uk/imgt/hla) de Janvier 2005. Ces gènes sont très polymorphes, la base de données est donc mise à jour régulièrement grâce aux nouveaux allèles référencés.

Il est donc nécessaire de vérifier régulièrement que les deux PCRs amplifient bien tous les allèles présents dans la base de données et dans le cas échéant de nouvelles amorces de PCR devront être utilisées.

4.2. La chimère ARN/ADN simple-brin

La chimère ARN/ADN simple-brin a été synthétisée par l'ADN polymérase *G46E CS6R*.

4.2.1. L'ADN polymérase *G46E CS6R*

L'ADN polymérase *G46E CS6R* incorpore uniquement des NTPs contrairement aux autres enzymes qui transcrivent le brin d'ADN en brin d'ARN ou qui incorporent des dNTPs à partir du brin d'ADN (Mauger et al. 2006).

Elle permet la synthèse d'une chimère ARN/ADN simple-brin à partir d'une amorce linéaire. D'autre part, cette enzyme est très sélective. Il n'est donc pas nécessaire d'enlever les dNTPs en excès de la PCR ce qui est un avantage par rapport aux autres méthodes de microhaplotypage (Mauger et al. 2006). L'étape d'élongation linéaire peut donc être réalisée directement à partir d'une PCR non purifiée.

4.2.2. Les amorces

Dans cette méthode de microhaplotypage, l'utilisation de l'ADN polymérase et de plusieurs amorces d'élongation pour l'étude de différents microhaplotypes permettent ainsi la création de plusieurs chimères ARN/ADN simple-brin dans une même réaction. Cette méthode est adaptée pour l'étude d'une région très polymorphe.

Les amorces d'élongation sont adjacentes au premier SNP du microhaplotype à étudier. Afin d'étudier un microhaplotype, un deuxième SNP, aux alentours du premier, a été analysé en même temps en utilisant différentes séquences d'amorces. Ces amorces possèdent plusieurs désoxynucléotides et une seul ribonucléotide. La position du ribonucléotide peut varier afin d'obtenir des microhaplotypes de masse unique. Les masses des microhaplotypes doivent également être différentes de celles des amorces d'élongation qui ne réagissent pas. Cette méthode permet le multiplexage de plusieurs microhaplotypes en même temps, ce qui est un avantage.

Un duplex de deux microhaplotypes du locus *HLA-A* a été étudié avec au total 6 amorces d'élongation qui couvrent les allèles de la base de données *IMGT-HLA* sur internet (www.ebi.ac.uk/imgt/hla) de Janvier 2005.

Tout comme pour la PCR, de nouveaux allèles peuvent être découverts. Par conséquent, de nouvelles amorces couvrant les nouveaux microhaplotypes devront être ajoutées.

4.3. L'analyse des microhaplotypes

4.3.1. Le clivage

La chimère ARN/ADN est ensuite clivée par une simple réaction chimique afin d'obtenir le microhaplotype à étudier.

La méthode de génotypage des SNPs par clivage de la chimère ARN/ADN et MALDI-TOF MS a montré que l'hydroxyde de sodium, l'hydroxyde de potassium ou l'hydroxyde d'ammonium pouvaient être utilisés (Mauger et al. 2006). Le clivage avec l'hydroxyde d'ammonium est lent et incomplet. Le clivage par l'hydroxyde de sodium ou l'hydroxyde de potassium est plus rapide et total. Ils peuvent être utilisés tous les deux. Dans cette méthode, le clivage est réalisée avec de l'hydroxyde de sodium.

La réaction de clivage chimique est facile à mettre en œuvre, rapide, totale et peu coûteuse contrairement au clivage enzymatique qui est utilisé notamment dans la méthode *MassARRAY*[®].

4.3.2. Le dessalage

La méthode de génotypage des SNPs par clivage de la chimère ARN/ADN et MALDI-TOF MS a montré que deux méthodes d'échantillonnages peuvent être utilisées : une réaction d'alkylation des ponts phosphorothioates ou un dessalage par une résine échangeuse de cations (Mauger et al. 2006).

Dans cette méthode de microhaplotypage, la méthode de dessalage de cations est utilisée car il n'y a pas de perte de fragments de clivage car les billes échangeuses de cations sont simplement ajoutées à la solution. De plus, elles permettent également de réduire le pH de la solution en neutralisant les ions OH⁻ provenant du clivage par la soude à l'aide des protons introduits.

De plus, afin d'automatiser la quantité de billes ajoutées, le kit *Clean Resin Tool* de *MassARRAY*[®] a été utilisé par rapport aux protocoles précédents (Mauger et al. 2006; Mauger et al. 2007).

4.3.2. L'analyse par MALDI-TOF MS

Comme dans la méthode de génotypage des SNPs par clivage de la chimère ARN/ADN et par MALDI-TOF MS, la matrice HCCA a été utilisée car elle est adaptée pour l'analyse des petits oligonucléotides et elle cristallise de manière homogène dans le mélange avec l'analyte. Une simple cible en alliage d'aluminium est nécessaire.

Les fragments de clivage possèdent une charge négative provenant de l'atome d'oxygène du groupement phosphate, l'analyse au MALDI-TOF MS s'effectue donc en mode négatif.

Les microhaplotypes ont été choisis pour que leurs masses soient analysables en mode linéaire et dans une région où le pouvoir résolutif de l'instrument soit suffisant pour distinguer la différence de masse entre les microhaplotypes.

De plus, le MALDI-TOF MS permet une analyse facile et rapide d'un grand nombre d'échantillons.

4.3.3. Les caractéristiques de la méthode

Les microhaplotypes étudiées ici font partie d'un ensemble de plusieurs microhaplotypes sélectionnés afin de pouvoir caractériser un haplotype *HLA* avec seulement l'étude de quelques microhaplotypes (Kucharzak 2006). D'autres microhaplotypes doivent être étudiées afin d'analyser les différents haplotypes.

Les amorces de PCR et d'élongation sont choisies en fonction de la table de la base de données IMGT-*HLA* sur internet (www.ebi.ac.uk/imgt/hla) datant de Janvier 2005 et des fréquences d'allèles propres à chaque population. Il existe des haplotypes *HLA* fréquents et d'autres rares et leurs proportions varient suivant la population étudiée. Le choix des amorces de PCR et/ou d'élongation dépendent d'une part de la mise à jour de la base de données et d'autre part de la population étudiée. Le même problème est rencontré dans la méthode de microhaplotypage du *GOODASSAY* (Kucharzak 2006). Cette méthode doit donc être

actualisée avec les nouveaux haplotypes découverts et les populations étudiées afin de pouvoir analyser les haplotypes des individus.

Enfin, cette méthode permet l'étude des SNPs connus et séparés de quelques bases. Elle ne permet pas d'analyser des SNPs inconnus ou isolés dans la séquence. Seule une méthode de re-séquençage pourrait permettre de les analyser.

5. Conclusion

Dans ce chapitre, une nouvelle méthode de multiplex microhaplotypage par clivage chimique d'une chimère ARN/ADN simple-brin et l'analyse par MALDI-TOF MS a été développée. Cette méthode est basée sur celle du génotypage des SNPs par clivage chimique d'une chimère ARN/ADN simple-brin et l'analyse par MALDI-TOF MS (Mauger et al. 2006) mais elle diffère par l'utilisation d'un ensemble d'amorces d'élongation. Ces amorces d'élongation permettent l'étude d'une région polymorphe par l'intermédiaire de microhaplotype.

Une PCR amplifie et sélectionne la région du gène à étudier ainsi que tous ces allèles. L'ADN polymérase *G46E CS6R* synthétise une chimère ARN/ADN simple-brin à l'aide d'amorces d'élongation et de NTPs. Tout comme la méthode de génotypage, les amorces sont adjacentes au SNP à étudier et elles contiennent un ribonucléotide vers l'extrémité 3' qui peut varier d'une amorce à l'autre. Par contre, la séquence de l'amorce intègre également un autre SNP à étudier qui est situé près du premier. Plusieurs amorces d'élongation peuvent être utilisées afin de couvrir tous les microhaplotypes. De plus, plusieurs microhaplotypes du même gène peuvent être étudiés en même temps par l'intermédiaire d'autres amorces d'élongation. La chimère ARN/ADN simple-brin est ensuite clivée après chaque ribonucléotide pour former un microhaplotype. Chaque microhaplotype se termine par un ribonucléotide et un groupement 3'-phosphate. Ils sont ensuite dessalés par une résine échangeuse de cations et analysés par MALDI-TOF MS en mode linéaire négatif. Les masses sont significatives des microhaplotypes, permettant ainsi de déterminer les individus de microhaplotypes homozygotes ou hétérozygotes.

Cette méthode a été validée sur le système *HLA* hypervariable, pour l'analyse du microhaplotype ASTW du locus *HLA-DRB1-227* de 23 individus et pour l'analyse d'un

duplex deux microhaplotypes HCACH du locus *HLA-A-98* et MGAR du locus *HLA-A-453* de 12 individus.

L'analyse de ses seuls microhaplotypes ne suffit pas pour déterminer l'allèle des individus. Afin de distinguer les allèles de ces deux gènes, d'autres microhaplotypes doivent être ajoutés. Le multiplexage d'un ensemble de microhaplotypes du même gène *HLA* dans une même réaction permettrait de réaliser une méthode directe d'haplotypage.

Cependant, cette méthode a été développée par rapport à la base de données IMGT-HLA sur internet (www.ebi.ac.uk/imgt/hla) datant de janvier 2005. Les amorces de PCR et d'élongation ainsi que le choix des microhaplotypes à étudier sont fonction des allèles présents à cette date et de la fréquence des allèles dans la population à étudier (Kucharzak 2006). Cette méthode doit être mise à jour en fonction des nouveaux allèles référencés et de la population à étudier.

Cependant, le choix des microhaplotypes est fonction des SNPs connus de la région polymorphique. Cette méthode ne permet donc pas l'identification des SNPs inconnus, ni de nouveaux allèles. Pour l'analyse de SNPs inconnus, une méthode de re-séquençage doit être utilisée pour identifier de nouveaux allèles.

Cette méthode peut être utilisée pour l'étude des microhaplotypes des autres gènes du système *HLA* ou d'autres gènes très polymorphes.

Cette chimère ARN/ADN simple-brin pourrait également avoir d'autres usages que le couplage avec le MALDI-TOF MS présenté ici.

Chapitre 3

Analyse de la méthylation de l'ADN par clivage d'une chimère ARN/ADN simple-brin et par MALDI-TOF MS

Chapitre 3. Analyse de la méthylation de l'ADN par clivage d'une chimère ARN/ADN simple-brin et par MALDI-TOF MS

1. Introduction

En 2007, une méthode de re-séquençage de l'ADN mitochondrial par clivage d'une chimère ARN/ADN simple-brin et d'analyse par spectrométrie de masse MALDI-TOF a été développée (Mauger et al. 2007). Dans cette méthode (annexe 2), la chimère ARN/ADN est synthétisée à partir du produit de PCR purifié, d'une amorce d'élongation et d'un mélange d'un ribonucléotide (A, C ou G) et des désoxynucléotides complémentaires. La chimère est ensuite clivée. Les fragments de clivage sont dessalés et analysés par MALDI-TOF MS. Les masses des fragments sont ensuite comparées à ceux de la séquence de référence. Les différences sont caractéristiques d'une modification dans la séquence de l'ADN. Dans ce chapitre, une nouvelle méthode d'analyse de la méthylation de l'ADN par clivage d'une chimère ARN/ADN simple-brin et par spectrométrie de masse MALDI-TOF est développée. Elle est basée sur la méthode précédente (Mauger et al. 2007) mais elle a pour but l'analyse de la méthylation de l'ADN génomique des îlots CpGs.

Ces dernières années, l'étude de la méthylation de l'ADN n'a cessé de croître. Chez les mammifères, seules les cytosines contenues dans un dinucléotide CpG peuvent être méthylées sous la forme de la 5-méthylcytosine. Les CpGs forment un revêtement moléculaire qui confère un caractère unique et transmissible de l'ADN. Par conséquent, de nouvelles méthodes d'analyse de la méthylation de l'ADN se sont développées qui visent à étudier plus particulièrement les îlots CpGs.

Le pyroséquençage est une méthode couramment utilisée pour quantifier la méthylation de l'ADN avec une résolution de 5% (Dupont et al. 2004; Tost et al. 2006). Cependant, elle ne peut séquencer qu'environ 100 pb dans une réaction. Il existe également des méthodes d'analyse de la méthylation de l'ADN par MALDI-TOF MS : la méthode de génotypage du *GOODASSAY* (Tost et al. 2003) et notamment la méthode de re-séquençage de *MassArray*[®] (Ehrich et al. 2005).

La plupart des méthodes d'analyse de la méthylation de l'ADN reposent sur la conversion par le bisulfite de sodium des cytosines. La cytosine et la 5-méthylcytosine ne sont pas

différenciées par les ADN polymérase. Afin de les identifier, les cytosines sont converties en uraciles par le bisulfite de sodium alors que les 5-méthylcytosine ne le sont pas.

Par ailleurs, les méthodes par spectrométrie de masse permettent de quantifier les échantillons grâce à la nature semi-quantitative du MALDI-TOF MS (Szájli et al. 2008). Les tirs du laser sur le mélange d'agrégats de matrice-analyte sont tous différents car les agrégats sont différents au sein d'un même spot ou d'un spot à l'autre. La quantification est donc dépendante d'une cristallisation homogène de la matrice. Pour quantifier par MALDI-TOF MS, il est donc nécessaire de multiplier l'accumulation des spectres de masse et de déposer plusieurs fois les échantillons afin d'établir une courbe d'étalonnage.

Dans ce travail, la combinaison de la conversion au bisulfite et de la méthode de re-séquençage permet l'analyse et la quantification de la méthylation de l'ADN. La conversion au bisulfite permet de distinguer les fragments de clivage de la chimère ARN/ADN des cytosines des CpGs méthylées ou non. L'abondance des ions provenant des fragments de clivage des cytosines 0 et 100% méthylées est utilisée pour élaborer une courbe d'étalonnage afin de quantifier la méthylation de l'ADN des individus. Cette méthode est composée de 7 étapes : (i) la conversion au bisulfite, (ii) la PCR, (iii) la purification de la PCR, (iv) l'élongation linéaire, (v) le clivage, (vi) le dessalage et (vii) l'analyse par MALDI-TOF MS.

Cette méthode a été développée sur 7 individus sur l'étude des îlots CpGs des loci des gènes *CDKN2A* et *RASSF1A*. Le re-séquençage et l'étude de la méthylation de l'ADN des loci du gène permet de déterminer les régions hypométhylées et hyperméthylées.

Le gène *CDKN2A*, suppresseur de tumeur, est situé chez l'humain sur le chromosome 9p21. Il code pour la protéine *CDKN2A* (*cyclin-dependent kinase inhibitor 2A*) (Smith-Sørensen & Hovig 1996). Elle joue un rôle important dans la régulation du cycle cellulaire. Elle est surexprimée dans la sénescence cellulaire. Les variantes de ce gène augmentent le risque de développer certaines formes de cancer et notamment le mélanome. La méthylation des îlots CpGs de la région promotrice du gène est associée avec l'extinction de transcription de son expression dans plusieurs cancers comme celui du colon (Shima et al. 2011).

Le gène *RASSF1A* est situé chez l'humain sur le chromosome 3p21.3. Il code pour la protéine *RASSF1A* (*Rass association domain-containing protein 1*). Elle aurait une implication potentielle dans la signalisation apoptotique, la stabilisation des microtubules, et la progression du cycle cellulaire. L'inactivation de ce gène a été corrélée avec

l'hyperméthylation des îlots CpGs de sa région promotrice. L'hyperméthylation du gène *RASSF1A* est fréquent dans la plupart des tumeurs humaines tels que le poumon, le sein, la prostate, le pancréas, le rein, le foie, le col de l'utérus, la thyroïde.... La détection de la méthylation du gène *RASSF1A* dans les fluides corporels tels que le sérum, l'urine et la salive pourrait être un marqueur utile pour la détection précoce du cancer (Pfeifer & Dammann 2005).

2. Matériels et Méthodes

2.1. Matériels

Les 7 échantillons d'ADN utilisés pour l'étude des loci des gènes *CDKN2A* et *RASSF1A* ont été analysés par pyroséquençage (Tost et al. 2006).

Les amorces de PCR ont été synthétisées par Biotex (Berlin, Allemagne) et les amorces d'élongation par Eurogentec (Angers, France). Les séquences des amorces sont présentées dans le tableau 3.1. Les amorces d'élongation contiennent une modification 2'-OMe du ribose de l'avant dernière base. Cette modification permet d'inhiber la digestion 3'-5' de l'amorce par l'activité exonucléase *haute-fidélité* de l'ADN polymérase (Schonbrunner et al. 2006; Mauger et al. 2007). Elles contiennent également une modification 5'-6-FAM fluorescente afin de contrôler la taille de la chimère ARN/ADN simple-brin par électrophorèse capillaire.

| Locus | Amorce de PCR | Amorce d'élongation |
|----------------------|--|----------------------------------|
| <i>CDKN2A</i> _msp1 | GGGAGTAAGTTTTTTTTGGTTTTTTT CACCACCCTCCAACCTAAATA | FAM-GTTTTTTTAGGTTGGAGTGTAAAT(G)G |
| <i>CDKN2A</i> _msp2 | TTTTTGTTTTAGTTTTTTGAGTAGTTGG ACCTTAAAAACAATACCACACATCC | FAM-GTTTTTTGAGTAGTTGGAATTATA(T)A |
| <i>CDKN2A</i> _msp3a | GGTGGGGTTTTTATAATTAGGAAAG AAACTAACTCCTCCCCACCTAC | FAM-ATTTGGTAGTTAGGAAGGTTGT(A)T |
| <i>CDKN2A</i> _msp3b | GTAGGTGGGGAGGAGTTTAGTTT CACCTCTAATAACCAACCAAC | FAM-TTTTTTTTGTAAAYGTTGGTTTT(G)G |
| <i>CDKN2A</i> _msp4 | GTTTTTTTGTTTGGAAAGATAT TACAAACCCTCTACCCACCTAAAT | FAM-AGAGGATTTGAGGGATAGG(G)T |
| <i>CDKN2A</i> _msp5 | TTTAGGTGGTAGAGGGTTTGTAG TCCTAATAAAAAATAAACTCCATCC | FAM-GTAGYGGGAGTAGGGGAT(G)G |
| <i>CDKN2A</i> _msp6 | TGGAGTTTATTTTTTTTATTAGGAGG ATAATTACTTCTAAATAATTTTTCTCTC | FAM-GTGGTGTATGTTGGAATAAAT(A)T |
| <i>CDKN2A</i> _msp7 | TAATTAGATATTTGGGGTTTGTGTG AATCATCAATCCTCACCTAAAAAAC | FAM-GGTAGGTTATGATGATGGGT(A)G |
| <i>CDKN2A</i> _msp8 | GGGTATTAGAGGTAGTAATTATGTT ACAACCTAACCTAATTTAAAAAAA | FAM-GTAGTAATTATGTTYGTATAGAT(G)T |
| <i>CDKN2A</i> _msp9 | GTATTAGGTGTAAAAGATGTTTGGG CAAAACCTCTCTAATTCCTTCAATC | FAM-GATGYGGGAAGGGAAAGGTTA(T)A |
| <i>RASSF1A</i> _msp1 | ATTATTTATGTTAAGGGAATTTATTTAGAA CCTAAATACAAAACTATAAAACCC | FAM-AGTTTTTYGTAGTGGAGAGTAG(A)G |
| <i>RASSF1A</i> _msp2 | AGTTTTTGTATTTAGGTTTTTATTTG CTACCCCTTAACTACCCCTTCC | FAM-TAGTTTGGATTTTGGGGGA(G)G |
| <i>RASSF1A</i> _msp3 | GGAAGGGGTAGTTAAGGGGTAG CTTTCTCAAAAAAACCAAAACATT | FAM-TAGTTAAGGGGTAGYGTA(G)T |

Tableau 3.1 : Amorces de PCR et d'élongation des loci des gènes *CDKN2A* et *RASSF1A*. Les bases qui sont entre parenthèses contiennent la modification 2'OMe sur le ribose. La base à l'extrémité 5' contient une modification 5'-6-FAM.

Les dNTPs (N=A, C, G et T), les NTPs (N=A et G), le marqueur de taille 550R, le tampon de charge pour l'électrophorèse capillaire et la *Shrimp alkaline phosphatase (SAP)* proviennent de GE Healthcare (Saclay, France) et les produits chimiques d'Aldrich (Steinheim, Allemagne). La résine échangeuse de cations provient de Biorad (Marnes la coquette, France) et le kit *Clean ResinTool* de *MassARRAY*[®] de chez Sequenom (Hamburg, Allemagne). L'*éxonuclease I* provient de Ozyme (Saint-Quentin-en-Yvelines, France). La *Hot startTaq* ADN polymérase, le kit *EpiTect*[®] et les ADN 0 et 100% méthylés (bisulfités) de chez Qiagen (Courtaboeuf, France). Les étalons externes pour l'analyse en MALDI-TOF MS d'oligonucléotides de 4 à 11-mer et de 12 à 30-mer proviennent de chez BRUKER (Bremen, Allemagne).

L'ADN polymérase *KB17* a été développée par Roche Molecular Systems (Pleasanton, CA, USA). Elle n'est pas commercialisée, elle peut être obtenue par Dr Thomas Myers (thomas.myers@roche.com). Cette chimère d'ADN polymérase contient un domaine 5'-nucléase de l'ADN polymérase *Thermussp. ZO5* et les domaines 3'-nucléase et 5'-3'-nucléase de l'ADN polymérase *Thermotogamaritima (Tma)*. L'activité 5'-nucléase a été supprimée par

l'introduction de la mutation *G46E* (Mauger et al. 2006). L'activité *haute-fidélité* de l'ADN polymérase non-mutée *Tma* a été modulée par l'introduction d'une mutation *L329A* (Schonbrunner et al. 2006; Mauger et al. 2007) du 'Motif I' du domaine 3'-5' exonucléase. L'enzyme contient une mutation *E678G* qui élimine la spécificité de l'enzyme non-mutée contre l'incorporation des nucléotides substitués en position 2'-C, favorisant ainsi l'incorporation des ribonucléosides triphosphates. Enfin, l'enzyme contient trois mutations supplémentaires dans le domaine de la polymérase qui améliore l'accroche sur la matrice entraînant ainsi l'augmentation de la constante de vitesse d'élongation, particulièrement dans l'incorporation de ribonucléotides.

Les procédures de thermocyclage ont été réalisées sur des thermocycleurs gradients *Eppendorf* (Eppendorf, Allemagne). L'analyse du produit d'élongation linéaire a été effectuée sur un *MegaBace1000TM* de chez GE Healthcare (Saclay, France). Les analyses ont été réalisées par un MALDI-TOF MS *AutoflexTM* ou *UltraflexTM II* de chez Bruker (Bremen, Allemagne). La cible MALDI est une *AnchorchipTM* Bruker (Bremen, Allemagne).

2.2. Méthodes

2.2.1. La conversion au bisulfite

Le mélange de 140 μL de la solution de conversion au bisulfite par *EpiTect[®]* contient 20 μL d'ADN à 50ng/ μL , 85 μL du mélange de bisulfite de sodium et 35 μL du tampon de protection de l'ADN. La réaction de conversion est dénaturée pendant 5 min à 99°C, puis suivie de 25 min à 60°C, 5 min à 99°C, 1h25 à 60°C, 5 min à 99°C et se termine par 2h55 à 60°C.

Aux 140 μL de la solution de conversion sont ajoutés 560 μL de tampon BL. La solution est placée sur une colonne *EpiTect[®]* et elle est centrifugée pendant 1 min à vitesse maximale. Sur la colonne sont ajoutés 500 μL du tampon BW de lavage. La colonne est centrifugée pendant 1 min à vitesse maximale.

Sur la colonne sont ajoutés 500 μL du tampon BD de désulfonation et la réaction est mise sous incubation pendant 15 min à température ambiante. La colonne est ensuite centrifugée pendant 1 min à vitesse maximale.

Sur la colonne sont ajoutés 500 μ L du tampon BW de lavage. La colonne est centrifugée pendant 1 min à vitesse maximale. L'étape précédente est répétée une deuxième fois.

Au centre de la colonne sont ajoutés 20 μ L du tampon EB d'éluion. La colonne est centrifugée pendant 1 min à 12 000 rpm. L'éluat est récupéré et sa concentration est alors mesurée.

2.2.2. Les mélanges d'ADN

Afin de réaliser la courbe d'étalonnage, les ADN 0 et 100% méthylés et bisulfités de chez Qiagen sont mélangés suivant le tableau 3.2. La courbe d'étalonnage est réalisée avec quatre séries de cette gamme (Tost et al. 2003).

| % de méthylation | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|--------------------|----|----|----|----|----|----|----|----|----|----|-----|
| V μ l ADN 100% | 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
| V μ l ADN 0% | 20 | 18 | 16 | 14 | 12 | 10 | 8 | 6 | 4 | 2 | 0 |

Tableau 3.2 : Préparation des mélanges d'ADN méthylés et bisulfités de 0 à 100% pour la courbe d'étalonnage.

2.2.3. La PCR

Le mélange de 25 μ L de la solution de la PCR contient 2,5 μ L de tampon d'enzyme 10X de PCR, 1,6 μ L de chlorure de magnésium à 25 mM, 1,25 μ L de dNTPs à 8 mM, 1 μ L de chaque amorce de PCR (tableau 3.1) à 10 pmol/ μ L, 0,4 μ L d'enzyme *Taq Hot Start* à 5 U/ μ L, 2 μ L d'ADN à 10 ng/ μ L et 15,25 μ L d'eau milli-Q. La réaction de PCR est dénaturée pendant 15 min à 95°C, suivie de 50 cycles de 30 s à 95°C, 30 s à T°C (tableau 3.3) et 30 s à 72°C et se termine par une élongation de 5 min à 72°C. Le tableau 3.3 indique la température d'hybridation de chaque PCR.

| Locus | T °C |
|----------------------|-------------|
| <i>RASSF1A</i> _msp1 | 58 |
| <i>CDKN2A</i> msp2 | 60 |
| <i>CDKN2A</i> msp4 | |
| <i>CDKN2A</i> msp6 | |
| <i>CDKN2A</i> msp5 | |
| <i>CDKN2A</i> msp8 | |
| <i>RASSF1A</i> msp2 | 62 |
| <i>CDKN2A</i> msp1 | 63 |
| <i>CDKN2A</i> msp3a | |
| <i>CDKN2A</i> msp7 | |
| <i>CDKN2A</i> msp9 | |
| <i>CDKN2A</i> msp3b | 64 |
| <i>RASSF1A</i> msp3 | |

Tableau 3.3 : Températures d'hybridation de chaque PCR.

Le mélange, de 5µL de chaque PCR et de 5µL de tampon de charge, est ensuite déposé sur un gel d'agarose à 2 % afin de vérifier la taille du produit PCR.

2.2.4. La purification de la PCR

Au 10 µL de la réaction de PCR est ajoutée 1 µL de SAP à 1U/µL, 0,1µL d'*exonucléase I* à 20U/µL et 0,90µL d'eau milli-Q. La réaction est mise sous incubation pendant 1h à 95°C et se termine par 10 min à 90 °C.

2.2.5. L'élongation linéaire par l'ADN polymérase *KB17*

Le mélange de 20 µL de la solution d'élongation contient 1,5 µL de tricine (pH 7,75) à 200 mM, 1,6µL d'acétate d'ammonium à 500 mM, 0,67µL d'acétate de magnésium à 30 mM, 0,5 µL d'un mélange (ATP, dCTP, dGTP, dTTP ou GTP, dATP, dCTP, dTTP) à 4 mM , 2µL de l'amorce d'élongation (tableau 3.1) à 10 pmol/µL, 0,5µL d'enzyme *KB17* 10 U/µL, 12µL de la PCR purifiée et 1,23 µL d'eau milli-Q. La réaction d'élongation linéaire est dénaturée pendant 15 min à 95°C et suivie de 50 cycles de 30s à 95°C, 30 s à 62°C et 30 s à 72°C.

Afin de contrôler le nombre de base de l'élongation linéaire, 5 µL de la réaction d'élongation linéaire est déposée dans une électrophorèse capillaire *MegaBace1000TM*.

Sur une plaque de purification, sont déposées 400µL d'une solution de G50. La plaque est ensuite centrifugée pendant 2 min à 2500 rpm. Un mélange, de 5 µL de la solution d'élongation linéaire et de 15 µL d'eau, est déposé sur la plaque. La plaque est centrifugée pendant 2 min à 2200 rpm. Un mélange, de 2 µL de l'éluat et de 3 µL du marqueur de taille 550R (dilué à 0,25µL dans 2,25µL), est dénaturé pendant 2 min à 96°C. Il est alors refroidi sur de la glace et ensuite centrifugé pendant 1 min à 2000 rpm.

L'analyse s'effectue avec les paramètres suivants : siv=3 kV ; sit=45 s ; rv=10 kV ; rt 80 min et dye set 2. L'analyse dans l'électrophorèse capillaire s'effectue avec l'injection des 2 plaques de 96 puits contenant 100µL de tampon de charges, des 6 tubes de matrice, de l'eau et du mélange de 5µL.

2.2.6. Le clivage

Aux 20 µL de la réaction d'élongation linéaire est ajouté un mélange de 2,5 µL d'hydroxyde de sodium à 3 M et 2,5µL d'eau milli-Q. La réaction de clivage est mise sous incubation pendant 1h30 à 70°C.

2.2.7. Le dessalage

A la réaction de clivage, 6 mg de résines échangeuses de cations sont ajoutées à l'aide du kit *Clean ResinTool* de *MassArray*[®]. Les échantillons sont incubés pendant 20 min à température ambiante sous agitation et ensuite centrifugés sous 134 xg afin de récupérer le surnageant.

2.2.8. L'analyse par spectrométrie de masse MALDI-TOF

2.2.8.1. La préparation de la cible MALDI

La solution de matrice trihydroxyacetophenone (THAP) est préparée avec 0,2 M de 2,4,6 THAP dans 50 % d'acétonitrile, 0,2 M de 2,3,4 THAP dans 50% d'acétonitrile et 0,3 M de citrate d'ammonium dans les proportions 6/3/2 (v/v) (Zhu et al. 1996).

Sur une cible AnchorchipTM Bruker de 384 spots de porosité 400 µm, sont déposés 0,5 µL de la solution de matrice THAP et 0,5 µL du surnageant par spot. La cible est séchée à température ambiante.

2.2.8.2. La spectrométrie de masse MALDI-TOF

La cible MALDI est introduite dans un spectromètre de masse MALDI-TOF. L'analyse est réalisée en mode négatif à l'aide du logiciel *FlexControl*.

Les paramètres en mode linéaire sont : IS1=20 kV ; IS2=18,5 kV ; lens=7,5 kV et un délai d'extraction de 150ns.

Les paramètres en mode réflecteur sont: IS1=20 kV; IS2=17,25 kV; lens=6 KV; Reflectron=21kV; Reflectron2=11kV et un délai de 150 ns.

Les spectres de masse sont obtenus par l'accumulation de 500 tirs de laser. Un étalonnage externe est réalisé à l'aide des kits d'étalons externes d'oligonucléotides de 4 à 11-mer et de 12 à 30-mer.

Ces conditions permettent de détecter des échantillons d'environ 0,375 pmol.

Les spectres de masse sont analysés à l'aide du logiciel *FlexAnalysis*. L'annotation des pics caractérisés par le rapport m/z des ions, la soustraction de la ligne de base ainsi qu'un lissage du signal sont réalisés.

Un logiciel de calcul de masse des fragments de clivage a été développé par Jérémy Semhoun (Mauger et al. 2007). Il a été modifié afin de pouvoir calculer les masses des fragments de clivage des cytosines 0 et 100% méthylés et bisulfités.

Les rapports m/z des ions [M-H]⁻ sont ensuite comparés aux masses moléculaires des fragments de clivage de la séquence de référence et des fragments de clivage des CpGs des cytosines 0% et 100% méthylées. L'intensité des pics des ions de rapport m/z des cytosines méthylées et non méthylées des CpGs servent d'une part à déterminer les régions hypométhylées et hyperméthylées des îlots CpGs et d'autre part à réaliser une courbe d'étalonnage afin de quantifier la méthylation de l'ADN.

3. Résultats

3.1. Le principe de la méthode

Le principe de la méthode est illustré dans la figure 3.1.

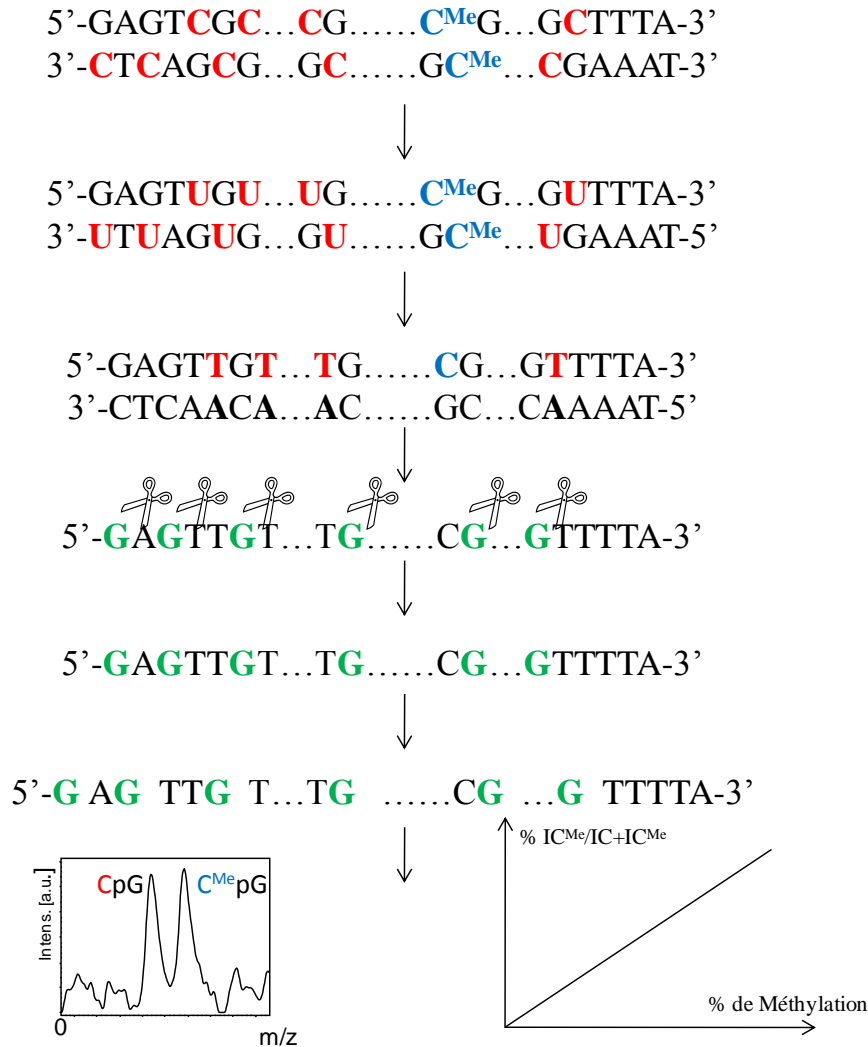


Figure 3.1 : Principe de la méthode. La première étape est une conversion au bisulfite de sodium. Puis, l'ADN est amplifié par PCR. La chimère ARN/ADN simple-brin est synthétisée et elle est ensuite clivée. Les fragments de clivage contenant les îlots CpGs sont analysés par MALDI-TOF MS. Une courbe d'étalonnage réalisée à partir des intensités des pics des ions des CpGs. Cette courbe permet de déterminer le pourcentage de méthylation des individus. Les bases qui correspondent aux cytosines sont rouges et les bases qui correspondent aux 5-méthylcytosines sont bleues. Les ribonucléotides sont en verts.

Cette méthode est composée de 7 étapes : (i) la conversion au bisulfite de l'ADN méthylé ; (ii) la PCR ; (iii) la purification de la PCR; (iv) l'élongation linéaire ; (v) le clivage ; (vi) le dessalage ; (vi) le MALDI-TOF MS.

3.1.1. La conversion au bisulfite

L'ADN polymérase de la réaction de PCR ne différencie pas d'une part, la cytosine de la 5-méthylcytosine et d'autre part l'uracile de la thymine. La conversion au bisulfite transforme la cytosine en uracile alors que la 5-méthylcytosine reste identique. La conversion au bisulfite est présentée dans la figure 3.2.

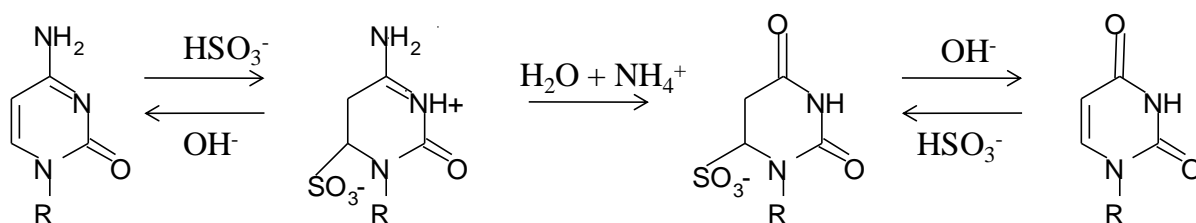


Figure 3.2 : Réaction de conversion au bisulfite. La première étape réversible est une réaction de sulfonation sur la cytosine pour former une cytosine sulfonate. La deuxième étape est une réaction totale de désamination de la cytosine sulfonate pour former un uracile sulfonate. La dernière étape réversible est une réaction de désulfonation par hydrolyse pour former l'uracile.

La conversion de la cytosine en uracile est composée de trois étapes : (i) une réaction de sulfonation réversible, (ii) une réaction de désamination totale et (iii) une réaction de désulfonation réversible.

3.1.2. La PCR

La réaction de PCR est réalisée sur de l'ADN bisulfité.

3.1.2.1. Les amorces de PCR

Les amorces de PCR amplifient le locus étudié. L'amorce sens ne contient pas de cytosine tandis que l'amorce anti-sens ne contient pas de guanine. Les amorces de PCR ne doivent pas contenir de dinucléotides CpGs afin de pouvoir amplifier tous les brins.

3.1.2.2. La purification de la PCR

La PCR est ensuite purifiée à l'aide de deux enzymes. Dans le protocole du re-séquençage de l'ADN mitochondrial (Mauger et al. 2007). La PCR est purifiée par un kit de purification mais cette procédure comporte beaucoup d'étapes et entraîne pas mal de perte de produit PCR.

Dans cette méthode, l'enzyme *SAP* digère les dNTPs en excès. L'ADN polymérase *KB17* utilisée lors de l'élongation linéaire incorpore les NTPs et les dNTPs. Dans cette étape, il n'y a qu'un seul ribonucléotide : ATP ou GTP. S'il reste des dATP ou dGTP dans la solution, l'ADN polymérase *KB17* pourra les incorporer durant l'élongation. Par conséquent, il peut y avoir des *sauts de clivage* dus à la présence de désoxynucléotides (A ou G).

L'*éxonucléase I* digère les amorces de PCR en excès. L'étape suivante est une élongation linéaire, elle amplifie donc un seul brin préférentiellement pour l'analyser. Les amorces de PCR pourraient être également utilisées et le re-séquençage du brin sélectionné serait alors difficile.

3.1.3. L'élongation linéaire par l'ADN polymérase *KB17*

La chimère ARN/ADN simple-brin a été synthétisée à partir de l'ADN polymérase *KB17*.

3.1.3.1. Les ribonucléotides

L'ADN polymérase *KB17* peut incorporer des ATP, CTP ou GTP lors d'une élongation linéaire (Mauger et al. 2007). Comme la conversion au bisulfite transforme toutes les cytosines en uraciles, la synthèse de la chimère ARN/ADN avec CTP n'est pas judicieuse car elle doit être ensuite clivée au niveau de ses cytosines. L'élongation linéaire est donc

effectuée sur chaque locus par un mélange d'ATP, dCTP, dGTP et dTTP ou par un mélange d'GTP, dATP, dCTP et dTTP et d'une amorce sélectionnant le brin sens de la PCR.

3.1.3.2. L'amorce

Tout comme l'amorce sens de la PCR, l'amorce d'élongation ne comporte pas de cytosine afin d'amplifier les brins méthylés ou non. Cependant, dans le cas des amorces de *CDNKN2A*-msp3b, *CDNKN2A*-msp5, *CDNKN2A*-msp8, *CDNKN2A*-msp9 et *RASSF1A*-msp1, il n'a pas été possible de trouver une séquence pour l'amorce ne contenant pas de dinucléotide CpG. Ces amorces contiennent donc à la place de la cytosine du CpG une base Y (C/T) afin d'amplifier les brins.

3.1.3.3. La chimère ARN/ADN simple-brin

L'analyse par électrophorèse capillaire des chimères ARN/ADN simple-brins de l'élongation linéaire ATP (411 bases) du locus *RASSF1A*-msp2, issus des ADNs 0, 20%, 50% et 70% et 100% méthylés sont présentées dans la figure 3.3.

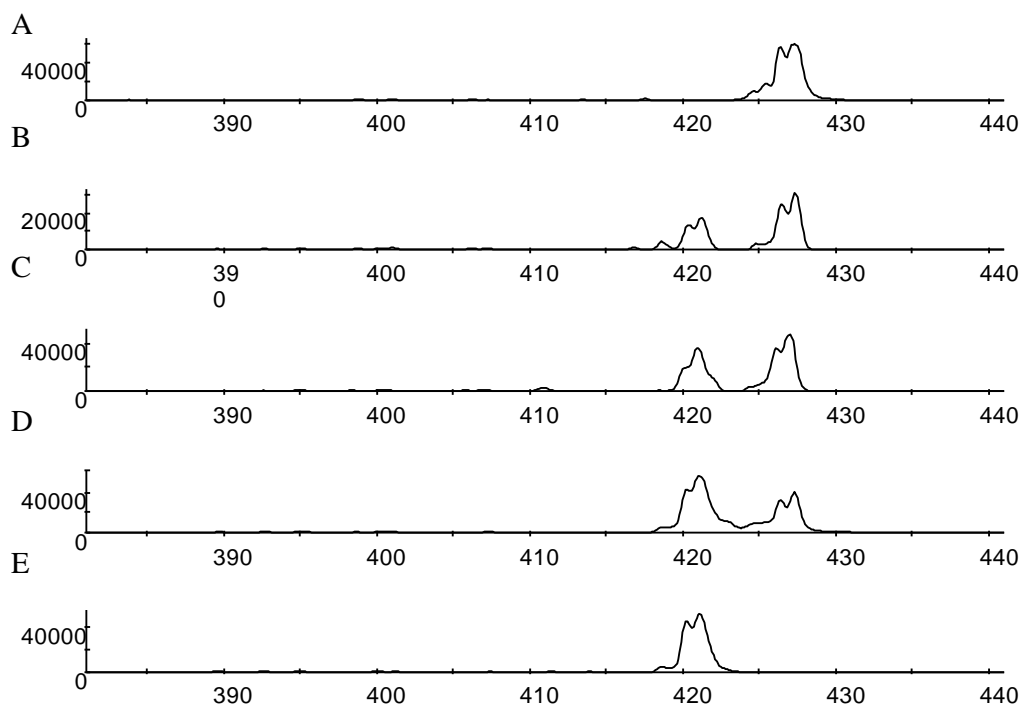


Figure 3.3 : Analyse de la chimère ARN/ADN simple-brin de 411 bases avec ATP du locus *RASSF1A*-msp2 par électrophorèse capillaire de l'ADN : (A) 0% méthylé ; (B) 20% méthylé ; (C) 50% méthylé ;

(D) 70% méthylé et (E) 100% méthylé. L'axe des abscisses représente le nombre de base et l'axe des ordonnées représente l'intensité de fluorescence.

L'analyse par électrophorèse capillaire montre que la chimère ARN/ADN du 0% méthylé est analysée à 427 bases (figure 3.3A) tandis que celle du 100% méthylé est analysée à 411 bases (figure 3.3E).

Elles devraient être analysées à 411 bases, mais les deux chimères ARN/ADN ne possèdent pas la même séquence, celle du 0% méthylé contient plus de bases T alors que celle du 100% méthylé contient plus de bases C. De plus, la chimère ARN/ADN est composée de ribonucléotides et de désoxynucléotides qui peut entraîner une différence de mobilité entre ces deux types de nucléotides.

Par ailleurs, les figures 3.3B à 3.3D montrent la présence des deux chimères ARN/ADN du 0% et 100% méthylés pour les mélanges d'ADN de 20%, 50% et 70%. La proportion des deux chimères ARN/ADN varie dans la même proportion que le taux de méthylation du mélange des ADN 0% et 100% méthylés (figure 3.3B à D).

3.1.4. Les fragments de clivage des CpGs

Pour chaque locus, deux élongations linéaires (ATP et GTP) ont été réalisées.

La chimère ARN/ADN simple-brin est clivée par la soude après chaque ribonucléotide. Il en résulte des fragments de clivage qui sont composés de désoxynucléotides et se terminent par un seul ribonucléotide et un groupement 3'-phosphate.

Certains fragments de clivage ne contiennent pas de CpG mais ils permettent le re-séquençage du locus. D'autres fragments de clivages contiennent des CpGs mais leur masse moyenne est commune à d'autres fragments de clivage. Ils ne peuvent donc pas servir à analyser le taux de méthylation de l'ADN mais ils permettent le re-séquençage du locus.

Seuls les fragments de clivage des CpGs qui possèdent une masse moyenne unique peuvent être utilisés pour déterminer le taux de méthylation de l'ADN. La hauteur des pics des fragments de clivage 0% et 100% méthylé du CpG est utilisée pour réaliser une courbe d'étalonnage externe. Cette courbe est réalisée à partir de différentes proportions de mélanges d'ADN 0 et 100% méthylés et elle permet de déterminer le taux de méthylation du CpG.

L'analyse des îlots CpGs des différents loci du gène permet d'identifier les régions hypométhylés et hyperméthylés.

Le tableau 3.4 présente les fragments de clivage des CpGs qui possèdent une masse moyenne distincte.

3.4A

| CDKN2A msp1 | 0 % méthylation | | | | 100% méthylation | | | |
|------------------------------|---------------------------|--------|-------|-----------------------------|---------------------------|--------|-------|-----|
| | Séquence | Masse | Début | Fin | Séquence | Masse | Début | Fin |
| ATP | tgtga | 1614,0 | 246 | 250 | cgtga | 1599,0 | 246 | 250 |
| | tttga | 1918,2 | 260 | 265 | ttcga | 1903,2 | 260 | 265 |
| | gttga | 1943,2 | 324 | 329 | gtcga | 1928,2 | 324 | 329 |
| | tgtgta | 2222,4 | 113 | 119 | tgcgta | 2207,4 | 113 | 119 |
| | gttttga | 2526,6 | 86 | 93 | gttttcga | 2511,6 | 86 | 93 |
| | gttttgta | 2830,8 | 124 | 132 | gtttcgta | 2815,8 | 124 | 132 |
| | ttttgta | 3135,0 | 30 | 39 | tttcggtta | 3120,0 | 30 | 39 |
| | gggtttgta | 3489,2 | 157 | 167 | gggtttcgta | 3474,2 | 157 | 167 |
| | ggttgtttga | 3793,4 | 177 | 188 | ggttgtttcga | 3778,4 | 177 | 188 |
| | ttgtttgtttggttttta | 6531,2 | 207 | 227 | ttcgtttgtttcggtttta | 6501,2 | 207 | 227 |
| | ttttgtttttttgggtta | 6835,4 | 46 | 67 | ttttcgtttttttgggtta | 6820,4 | 46 | 67 |
| gttttttaggtggagtgaatgg.tgtga | 9206,0 | 1 | 29 | gttttttaggtggagtgaatgg.cgca | 9175,9 | 1 | 29 | |
| GTP | ttttg | 1884,2 | 87 | 92 | ttttcg | 1869,2 | 87 | 92 |
| | gttttttaggtggagtgaatgg.tg | 8259,3 | 1 | 26 | gttttttaggtggagtgaatgg.cg | 8244,3 | 1 | 26 |

3.4B

| CDKN2A msp2 | 0 % méthylation | | | | 100% méthylation | | | |
|----------------|------------------------------------|---------|-------|-----|------------------------------------|---------|-------|-----|
| | Séquence | Masse | Début | Fin | Séquence | Masse | Début | Fin |
| ATP | tgggtgtta | 3210,1 | 71 | 80 | cggggtgta | 3195,1 | 71 | 80 |
| | gttggtttga | 3489,2 | 94 | 104 | gttggtttcga | 3474,2 | 94 | 104 |
| GTP | gtttttgagtagttggaattatata.tgtgtgta | 11008,2 | 1 | 35 | gtttttgagtagttggaattatata.cgtgcgta | 10978,1 | 1 | 35 |
| | gtttttgagtagttggaattatata.tg | 8819,7 | 1 | 28 | gtttttgagtagttggaattatata.cg | 8804,7 | 1 | 28 |

3.4C

| CDKN2A msp3a | 0 % méthylation | | | | 100% méthylation | | | |
|-----------------|------------------------------|--------|-------|-----|-------------------------------|--------|-------|-----|
| | Séquence | Masse | Début | Fin | Séquence | Masse | Début | Fin |
| ATP | tga | 1284,8 | 75 | 78 | cga | 1269,8 | 75 | 78 |
| | ggttga | 2272,5 | 129 | 135 | ggtcga | 2257,4 | 129 | 135 |
| | ggtgggga | 2955,9 | 101 | 109 | ggcgggga | 2940,9 | 101 | 109 |
| | gtgtggttggga | 3868,5 | 110 | 121 | gcgcggttggga | 3838,5 | 110 | 121 |
| | tgttttgttgga | 4401,8 | 87 | 100 | cgttttgttgga | 4386,8 | 87 | 100 |
| | tgggtggggtgga | 4881,1 | 40 | 54 | cggggcggggcgga | 4836,1 | 40 | 54 |
| | gggtgtgtgggggga | 5514,5 | 136 | 152 | gggcggtgtgggggga | 5499,5 | 136 | 152 |
| | fatttgtagttaggaaggttgat .gga | 9110,1 | 1 | 30 | fatttgtagttaggaaggttgat .cgga | 9399,2 | 1 | 30 |
| GTP | aatg | 1293,8 | 73 | 76 | aacg | 1278,8 | 73 | 76 |
| | aaatg | 1607,1 | 37 | 41 | aaacg | 1592,0 | 37 | 41 |
| | tatttaaatg | 3754,5 | 77 | 88 | tatttaaacg | 3739,4 | 77 | 88 |
| | fatttgtagttaggaaggttgat .g | 8467,6 | 1 | 28 | fatttgtagttaggaaggttgat .cg | 8756,8 | 1 | 28 |

3.4D

| CDKN2A msp3b | 0 % méthylation | | | | 100% méthylation | | | |
|-------------------------|----------------------|--------|-------|--------------------------|----------------------|--------|-------|-----|
| | Séquence | Masse | Début | Fin | Séquence | Masse | Début | Fin |
| ATP | ttgga | 1614,0 | 223 | 227 | tcgga | 1599,0 | 223 | 227 |
| | tttgta | 2197,4 | 93 | 99 | ttcgta | 2182,4 | 93 | 99 |
| | tttgta | 2501,6 | 133 | 140 | ttcgta | 2486,6 | 133 | 140 |
| | gggttga | 2601,7 | 196 | 203 | gggtcga | 2586,7 | 196 | 203 |
| | gttttga | 2880,9 | 101 | 109 | gtttcga | 2865,8 | 101 | 109 |
| | ttttttga | 3110,0 | 120 | 129 | tttttcga | 3095,0 | 120 | 129 |
| | ttggggtga | 3210,1 | 67 | 76 | ttggggcga | 3195,1 | 67 | 76 |
| | ttgtggtttta | 4376,8 | 166 | 179 | tcgcggtttta | 4346,8 | 166 | 179 |
| | gggggtttttgta | 5060,3 | 204 | 219 | gggggtttttcgta | 5045,2 | 204 | 219 |
| | tggtgtttttgttga | 5972,8 | 141 | 159 | cggcgtttttgttga | 5942,8 | 141 | 159 |
| fttttttgaatgttgg .a | 8669,8 | 1 | 29 | fttttttgaacgttgg .a | 8654,8 | 1 | 29 | |
| gggtgttttgggtgttttgggga | 9519,1 | 30 | 59 | gggtgttttcggtgttttcgggga | 9489,1 | 30 | 59 | |
| GTP | atattg | 1902,2 | 163 | 168 | atattc | 1887,2 | 163 | 168 |
| | ttatattg | 2814,8 | 88 | 96 | ttatattc | 2799,8 | 88 | 96 |
| | tattttttg | 3414,2 | 118 | 128 | tatttttc | 3399,2 | 118 | 128 |
| | fttttttgaatgttgg .ag | 8999,0 | 1 | 30 | fttttttgaacgttgg .ag | 8984,0 | 1 | 30 |

3.4E

| CDKN2A msp4 | 0 % méthylation | | | | 100% méthylation | | | |
|---------------------------|-------------------------|--------|-------|---------------------------|--------------------------|--------|-------|-----|
| | Séquence | Masse | Début | Fin | Séquence | Masse | Début | Fin |
| ATP | ttgga | 1614,0 | 45 | 49 | tcgga | 1599,0 | 45 | 49 |
| | ggttga | 1943,2 | 294 | 299 | ggtcga | 1928,2 | 294 | 299 |
| | tggttga | 2576,7 | 286 | 293 | cggtcgga | 2546,6 | 286 | 293 |
| | gttttggtga | 3793,4 | 184 | 195 | gttttcggtga | 3778,4 | 184 | 195 |
| | gggtgggtgga | 3918,5 | 83 | 94 | gggtggggcga | 3903,5 | 83 | 94 |
| | gttggtggtggga | 4526,9 | 160 | 173 | gtcggcggcggga | 4481,9 | 160 | 173 |
| | gtgggtggtggga | 4551,9 | 136 | 149 | gcggcggcggga | 4506,9 | 136 | 149 |
| | gggggtttttgta | 5060,3 | 26 | 41 | gggggtttttcgta | 5045,2 | 26 | 41 |
| | ggtgggggtgttga | 5439,5 | 253 | 269 | ggcgggggcgttga | 5409,5 | 253 | 269 |
| | ggtgtgggtgttga | 5768,7 | 235 | 252 | ggtgcgggcgttga | 5738,7 | 235 | 252 |
| | tggttgggttgggtggga | 7339,7 | 207 | 229 | cggtcgggtcgggtcggga | 7264,7 | 207 | 229 |
| | ttgtgtgttgggttga | 7618,9 | 95 | 118 | tcgcgtgcgttcggcgggtcggga | 7528,8 | 95 | 118 |
| agaggattgagggatagggt.tgga | 8032,2 | 1 | 25 | agaggattgagggatagggt.cgga | 8017,2 | 1 | 25 | |
| GTP | tttttg | 1884,2 | 185 | 190 | tttttc | 1869,2 | 185 | 190 |
| | ttttttg | 2492,6 | 31 | 38 | tttttcg | 2477,6 | 31 | 38 |
| | agaggattgagggatagggt.tg | 7389,8 | 1 | 23 | agaggattgagggatagggt.cg | 7374,8 | 1 | 23 |

3.4F

| CDKN2A msp5 | 0 % méthylation | | | | 100% méthylation | | | |
|----------------|-----------------------------|--------|-------|-----|----------------------------|--------|-------|-----|
| | Séquence | Masse | Début | Fin | Séquence | Masse | Début | Fin |
| ATP | tttttga | 2501,6 | 206 | 213 | tttttcga | 2486,6 | 206 | 213 |
| | tggtgta | 2551,6 | 194 | 201 | cgccgta | 2506,6 | 194 | 201 |
| | tttttga | 2830,8 | 77 | 85 | tttttcgga | 2815,8 | 77 | 85 |
| | gttttga | 2855,8 | 68 | 76 | gcgttcga | 2825,8 | 68 | 76 |
| | gtttggga | 2880,9 | 292 | 300 | cgtttggga | 2865,8 | 292 | 300 |
| | ttttttggtga | 4376,8 | 136 | 149 | ttttttggcga | 4361,8 | 136 | 149 |
| | tgtttgggggttggga | 5743,7 | 150 | 167 | cgtttgggggttggga | 5728,7 | 150 | 167 |
| | gtagtgggagtagggatgg.tgggtga | 8722,6 | 1 | 27 | gtagcgggagtagggatgg.cggcga | 8677,6 | 1 | 27 |
| GTP | atataaaattatg | 4380,9 | 276 | 289 | atataaaattacg | 4365,9 | 276 | 289 |
| | gtagtgggagtagggatgg.tg | 7117,6 | 1 | 22 | gtagcgggagtagggatgg.cg | 7087,6 | 1 | 22 |

3.4G

| CDKN2A msp6 | 0 % méthylation | | | | 100% méthylation | | | |
|----------------|---------------------------|--------|-------|-----|---------------------------|--------|-------|-----|
| | Séquence | Masse | Début | Fin | Séquence | Masse | Début | Fin |
| ATP | gtttga | 1918,2 | 136 | 141 | gttcga | 1903,2 | 136 | 141 |
| | ttgtttga | 2526,6 | 254 | 261 | tcgtttga | 2511,6 | 254 | 261 |
| | ggttggga | 2601,7 | 128 | 135 | ggtcggga | 2586,7 | 128 | 135 |
| | ggtgtttga | 3185,0 | 211 | 220 | ggcgtttga | 3170,0 | 211 | 220 |
| | ggtggggga | 3285,1 | 109 | 118 | ggcggggga | 3270,1 | 109 | 118 |
| | ttttgttta | 3414,2 | 166 | 176 | tttcgttta | 3399,2 | 166 | 176 |
| | gttgggggtggtgga | 5160,3 | 195 | 210 | gtcgggggtggtgga | 5145,3 | 195 | 210 |
| | gtggttgggtggtgttta | 6681,3 | 56 | 76 | gcggtcgggcgcggtgttta | 6621,2 | 56 | 76 |
| | gtggtgatgttgggaataaat.tga | 8245,4 | 1 | 26 | gtggtgatgttgggaataaat.cga | 8215,3 | 1 | 26 |
| GTP | attg | 1284,8 | 39 | 42 | atcg | 1269,8 | 39 | 42 |
| | aatttg | 1902,2 | 261 | 266 | aattcg | 1887,2 | 261 | 266 |
| | aaatttg | 2519,6 | 163 | 170 | aaattcg | 2504,6 | 163 | 170 |
| | gtggtgatgttgggaataaat.tg | 7932,2 | 1 | 25 | gtggtgatgttgggaataaat.cg | 7902,1 | 1 | 25 |

3.4H

| CDKN2A msp7 | 0 % méthylation | | | | 100% méthylation | | | |
|----------------|----------------------------|--------|-------|-----|----------------------------|--------|-------|-----|
| | Séquence | Masse | Début | Fin | Séquence | Masse | Début | Fin |
| ATP | tgga | 1614,0 | 237 | 241 | cgga | 1599,0 | 237 | 241 |
| | tttga | 1918,2 | 224 | 229 | tcgga | 1888,2 | 224 | 229 |
| | tgttga | 2222,4 | 230 | 236 | gttcga | 2207,4 | 230 | 236 |
| | ttttga | 2501,6 | 77 | 84 | tttcga | 2486,6 | 77 | 84 |
| | gtggtgga | 2601,7 | 30 | 37 | gtggcgga | 2586,7 | 30 | 37 |
| | tgtgtgga | 2905,9 | 53 | 61 | cggcgcgga | 2860,8 | 53 | 61 |
| | tgttgggtgttga | 5085,3 | 133 | 148 | cgttgggtgttga | 5070,3 | 133 | 148 |
| | ttgggtgggtgtggtgga | 6731,3 | 149 | 169 | tcgggtcgggcgcggttga | 6671,3 | 149 | 169 |
| | tttgtgtggttgggggga | 7339,7 | 242 | 264 | tttcgcgcggttgggggga | 7279,7 | 242 | 264 |
| | tgttgggtgttgggttga | 7593,9 | 179 | 202 | gttgggtgttgggttga | 7563,9 | 179 | 202 |
| | gtagggtatgatggtgtag.tgttga | 9240,0 | 1 | 29 | gtagggtatgatggtgtag.cgttga | 9210,0 | 1 | 29 |
| GTP | ttttatg | 2197,4 | 48 | 54 | ttttacg | 2182,4 | 48 | 54 |
| | ttattttattg | 4031,6 | 82 | 94 | ttattttattcg | 4016,6 | 82 | 94 |
| | gtagggtatgatggtgtag.tg | 7685,0 | 1 | 24 | gtagggtatgatggtgtag.cg | 7670,0 | 1 | 24 |

3.4I

| CDKN2A msp8 | 0 % méthylation | | | | 100% méthylation | | | |
|----------------|-------------------------------|--------|-------|-----|-------------------------------|--------|-------|-----|
| | Séquence | Masse | Début | Fin | Séquence | Masse | Début | Fin |
| ATP | tttgga | 2247,4 | 331 | 337 | ttcgga | 2232,4 | 331 | 337 |
| | tttggga | 2526,6 | 241 | 248 | tttcgga | 2511,6 | 241 | 248 |
| | tggttga | 2551,6 | 176 | 183 | cgttga | 2536,6 | 176 | 183 |
| | gtttgga | 2576,7 | 123 | 130 | gttcgga | 2561,6 | 123 | 130 |
| | gggtgga | 2601,7 | 209 | 216 | ggcgga | 2586,7 | 209 | 216 |
| | ttggtgga | 2905,9 | 167 | 175 | tcggtgga | 2890,9 | 167 | 175 |
| | ttgtttgttga | 4985,2 | 193 | 208 | tcgtttgttga | 4955,2 | 193 | 208 |
| | gttttgggttgttta | 5618,6 | 136 | 153 | gtttcgggttgttta | 5603,6 | 136 | 153 |
| | gtagtaattatgttataatgat.tgtgga | 9791,4 | 1 | 31 | gtagtaattatgttataatgat.cgtgga | 9746,3 | 1 | 31 |
| GTP | ttttg | 1884,2 | 137 | 142 | ttttcg | 1869,2 | 137 | 142 |
| | ttaattg | 2206,4 | 163 | 169 | ttaattcg | 2191,4 | 163 | 169 |
| | tttttattg | 3414,2 | 185 | 195 | tttttattcg | 3399,2 | 185 | 195 |
| | aaataatttg | 3450,3 | 223 | 233 | aaataattcg | 3435,2 | 223 | 233 |
| | gtagtaattatgttataatgat.tg | 8515,5 | 1 | 27 | gtagtaattatgttataatgat.cg | 8485,5 | 1 | 27 |

Tableau 3.4 : Séquences des fragments de clivage des CpGs de masse moyenne distincte des chimères ARN/ADN ATP et GTP: (A) *CDKN2A-msp1*; (B) *CDKN2A-msp2*; (C) *CDKN2A-msp3a*; (D) *CDKN2A-msp3b*; (E) *CDKN2A-msp4*; (F) *CDKN2A-msp5*; (G) *CDKN2A-msp6*; (H) *CDKN2A-msp7*; (I) *CDKN2A-msp8*; (J) *CDKN2A-msp9*; (K) *RASSF1A-msp1*; (L) *RASSF1A-msp2* et (M) *RASSF1A-msp3*.

Le tableau 3.4 montre que les élongations linéaires ATP et GTP sont complémentaires car les CpGs analysables ne sont pas toujours les mêmes.

Par contre, certains fragments de clivage des CpGs possèdent des masses supérieures à 6000 Da, ils sont donc difficiles à utiliser pour l'élaboration d'une courbe d'étalonnage à cause de leurs faibles intensités dû à la limite de détection de l'instrument.

Pour l'élaboration d'une courbe d'étalonnage, il vaut mieux privilégier les fragments de clivage des CpGs possédant une masse inférieure à 3000 Da qui ont une plus forte abondance.

3.1.5. L'analyse par spectrométrie de masse MALDI-TOF

Les fragments de clivage sont ensuite dessalés par une résine échangeuse de cations.

Ils sont ensuite analysés en mode négatif grâce à la charge négative portée par l'atome d'oxygène du groupement phosphate. L'analyse est réalisée en mode linéaire et en mode réflecteur par spectrométrie de masse MALDI-TOF MS.

3.1.6. L'analyse de la méthylation de l'ADN

Les fragments de clivage servent d'une part, au re-séquençage et d'autre part, à l'analyse du taux de méthylation de l'ADN par l'intermédiaire des fragments de clivage des CpGs qui possèdent une masse singulière.

Le re-séquençage et l'analyse de îlots CpGs permet de réaliser un criblage des différents loci du gène afin de dépister les zones hypométhylés et hyperméthylés. Cette méthode a été réalisée sur les deux gènes suivants : *CDKN2A* et *RASSF1A*. Elle a été mise au point à l'aide des ADN 0 et 100% méthylés et de leurs différents mélanges.

3.2. L'analyse du locus *CDKN2A*-msp1

Pour exemple, l'analyse du locus *CDKN2A*-msp1 est présentée.

3.2.1. Les fragments de clivage du locus

Le tableau 3.5 montre tous les fragments de clivage du re-séquençage du locus *CDKN2A*-msp1 des deux élongations linéaires : ATP et GTP.

3.5A

| CpG | 0 % méthylation | | | | 100% méthylation | | | |
|------------------|-----------------|--------|-------|-----|------------------|--------|-------|-----|
| | Séquence | Masse | Début | Fin | Séquence | Masse | Début | Fin |
| 70-71 | tta | 955,6 | 83 | 85 | tta | 955,6 | 83 | 85 |
| | tta | 955,6 | 104 | 106 | tta | 955,6 | 104 | 106 |
| | tta | 955,6 | 120 | 122 | tta | 955,6 | 120 | 122 |
| | tta | 955,6 | 239 | 241 | tta | 955,6 | 239 | 241 |
| | gta | 980,6 | 94 | 96 | gta | 980,6 | 94 | 96 |
| | gta | 980,6 | 148 | 150 | gta | 980,6 | 148 | 150 |
| | gta | 980,6 | 302 | 304 | gta | 980,6 | 302 | 304 |
| | tta | 1259,8 | 296 | 299 | tta | 1259,8 | 296 | 299 |
| | tta | 1259,8 | 344 | 347 | tta | 1259,8 | 344 | 347 |
| | gta | 1284,8 | 251 | 254 | gta | 1284,8 | 251 | 254 |
| | tta | 1284,8 | 277 | 280 | tta | 1284,8 | 277 | 280 |
| | gta | 1309,8 | 69 | 72 | gta | 1294,8 | 69 | 72 |
| | gta | 1309,8 | 109 | 112 | gta | 1309,8 | 109 | 112 |
| | gta | 1309,8 | 284 | 287 | gta | 1309,8 | 284 | 287 |
| 256-257 | ttta | 1564,0 | 291 | 295 | ttta | 1564,0 | 291 | 295 |
| | ttta | 1589,0 | 40 | 44 | ttta | 1574,0 | 255 | 259 |
| 246-247 | ttta | 1589,0 | 255 | 259 | ttta | 1589,0 | 40 | 44 |
| | ttta | 1614,0 | 246 | 250 | ttta | 1599,0 | 246 | 250 |
| 262-263 | ttta | 1639,1 | 202 | 206 | ttta | 1639,1 | 202 | 206 |
| | ttta | 1893,2 | 330 | 335 | ttta | 1893,2 | 330 | 335 |
| 326-327 | ttta | 1918,2 | 260 | 265 | ttta | 1903,2 | 260 | 265 |
| | ttta | 1943,2 | 324 | 329 | ttta | 1928,2 | 324 | 329 |
| 115-116 | ttta | 2172,4 | 141 | 147 | ttta | 2172,4 | 141 | 147 |
| | ttta | 2197,4 | 134 | 140 | ttta | 2197,4 | 134 | 140 |
| | ttta | 2222,4 | 113 | 119 | ttta | 2207,4 | 113 | 119 |
| | ttta | 2272,5 | 97 | 103 | ttta | 2272,5 | 97 | 103 |
| | ttta | 2272,5 | 348 | 354 | ttta | 2272,5 | 348 | 354 |
| | ttta | 2476,6 | 73 | 80 | ttta | 2476,6 | 73 | 80 |
| | ttta | 2501,6 | 336 | 343 | ttta | 2501,6 | 336 | 343 |
| | ttta | 2521,7 | 355 | 363 | ttta | 2511,6 | 86 | 93 |
| | ttta | 2526,6 | 86 | 93 | ttta | 2521,7 | 355 | 363 |
| | ttta | 2576,7 | 231 | 238 | ttta | 2576,7 | 231 | 238 |
| 91-92 | ttta | 2830,8 | 124 | 132 | ttta | 2815,8 | 124 | 132 |
| | ttta | 2855,8 | 168 | 176 | ttta | 2855,8 | 168 | 176 |
| 128-129 | ttta | 3135,0 | 30 | 39 | ttta | 3120,0 | 30 | 39 |
| | ttta | 3389,2 | 266 | 276 | ttta | 3389,2 | 266 | 276 |
| 33-34 | ttta | 3489,2 | 157 | 167 | ttta | 3474,2 | 157 | 167 |
| | ttta | 3743,4 | 190 | 201 | ttta | 3743,4 | 190 | 201 |
| 163-164 | ttta | 3793,4 | 177 | 188 | ttta | 3778,4 | 177 | 188 |
| | ttta | 5847,8 | 305 | 323 | ttta | 5847,8 | 305 | 323 |
| 186-187 | ttta | 5847,8 | 305 | 323 | ttta | 5847,8 | 305 | 323 |
| | ttta | 6531,2 | 207 | 227 | ttta | 6501,2 | 207 | 227 |
| 209-210; 218-219 | ttta | 6531,2 | 207 | 227 | ttta | 6501,2 | 207 | 227 |
| 50-51 | ttta | 6835,4 | 46 | 67 | ttta | 6820,4 | 46 | 67 |
| 25-26; 27-28 | ttta | 9206,0 | 1 | 29 | ttta | 9175,9 | 1 | 29 |

3.5B

| CpG | 0 % méthylation | | | | 100% méthylation | | | |
|---------|---------------------------|--------|-------|-----|---------------------------|--------|-------|-----|
| | Séquence | Masse | Début | Fin | Séquence | Masse | Début | Fin |
| 326-327 | ttg | 971,6 | 98 | 100 | tcg | 956,6 | 325 | 327 |
| | ttg | 971,6 | 170 | 172 | acg | 965,6 | 245 | 247 |
| | ttg | 971,6 | 233 | 235 | ttg | 971,6 | 98 | 100 |
| | ttg | 971,6 | 325 | 327 | ttg | 971,6 | 170 | 172 |
| | ttg | 971,6 | 350 | 352 | ttg | 971,6 | 233 | 235 |
| | tag | 980,6 | 95 | 97 | ttg | 971,6 | 350 | 352 |
| 246-247 | tag | 980,6 | 149 | 151 | tag | 980,6 | 95 | 97 |
| | atg | 980,6 | 245 | 247 | tag | 980,6 | 149 | 151 |
| | tttg | 1275,8 | 179 | 182 | tttg | 1275,8 | 179 | 182 |
| | tttg | 1275,8 | 211 | 214 | tttg | 1275,8 | 211 | 214 |
| | tatg | 1284,8 | 111 | 114 | tatg | 1284,8 | 111 | 114 |
| | ttag | 1284,8 | 174 | 177 | ttag | 1284,8 | 174 | 177 |
| 128-129 | atag | 1293,8 | 154 | 157 | atag | 1293,8 | 154 | 157 |
| | aatg | 1293,8 | 229 | 232 | aatg | 1293,8 | 229 | 232 |
| | ttttg | 1580,0 | 125 | 129 | tttcg | 1565,0 | 125 | 129 |
| | ttttg | 1580,0 | 160 | 164 | tttcg | 1565,0 | 160 | 164 |
| | ttttg | 1580,0 | 183 | 187 | tttcg | 1565,0 | 183 | 187 |
| | ttttg | 1580,0 | 215 | 219 | tttcg | 1565,0 | 215 | 219 |
| 163-164 | ttatg | 1589,0 | 165 | 169 | attcg | 1574,0 | 206 | 210 |
| | ttatg | 1589,0 | 206 | 210 | ttatg | 1589,0 | 165 | 169 |
| | ttttg | 1884,2 | 87 | 92 | ttttcg | 1869,2 | 87 | 92 |
| | ttttg | 1893,2 | 29 | 34 | atttcg | 1878,2 | 29 | 34 |
| | ttttg | 1893,2 | 197 | 202 | ttatcg | 1878,2 | 252 | 257 |
| | ttttg | 1893,2 | 252 | 257 | tattcg | 1878,2 | 258 | 263 |
| 186-187 | tatttg | 1893,2 | 258 | 263 | ttttag | 1893,2 | 197 | 202 |
| | ttaag | 1902,2 | 64 | 69 | tttaag | 1902,2 | 64 | 69 |
| | taatag | 1911,2 | 279 | 284 | taatag | 1911,2 | 279 | 284 |
| | ttatg | 2197,4 | 36 | 42 | ttatg | 2197,4 | 36 | 42 |
| | attatag | 2215,4 | 103 | 109 | attatag | 2215,4 | 103 | 109 |
| | attatag | 2215,4 | 238 | 244 | attatag | 2215,4 | 238 | 244 |
| 218-219 | tttttag | 2501,6 | 221 | 228 | tttttag | 2501,6 | 221 | 228 |
| | ttatttag | 2510,6 | 341 | 348 | ttatttag | 2510,6 | 341 | 348 |
| | aattttg | 2510,6 | 188 | 195 | aattttg | 2510,6 | 188 | 195 |
| | ttattaag | 2519,6 | 117 | 124 | ttattaag | 2519,6 | 117 | 124 |
| | taattttg | 2814,8 | 43 | 51 | taattttcg | 2799,8 | 43 | 51 |
| | ttaattttg | 2814,8 | 130 | 138 | ttaattttg | 2814,8 | 130 | 138 |
| 209-210 | tttttttg | 3101,0 | 52 | 61 | tttttttg | 3101,0 | 52 | 61 |
| | ttttatttg | 3110,0 | 331 | 340 | ttttatttg | 3110,0 | 331 | 340 |
| | tattttttg | 3119,0 | 139 | 148 | tattttttg | 3119,0 | 139 | 148 |
| | attttttttatg | 4335,8 | 265 | 278 | attttttttatg | 4335,8 | 265 | 278 |
| | atttttttatattag | 4658,0 | 72 | 86 | atttttttatattag | 4658,0 | 72 | 86 |
| | ataattttatataag | 4989,3 | 287 | 302 | ataattttatataag | 4989,3 | 287 | 302 |
| 91-92 | tatttttttttttttg | 6152,0 | 303 | 322 | tatttttttttttttg | 6152,0 | 303 | 322 |
| | gttttttaggtggagtgaatgg.tg | 8259,3 | 1 | 26 | gttttttaggtggagtgaatgg.cg | 8244,3 | 1 | 26 |
| | 33-34 | | | | | | | |
| | 256-257 | | | | | | | |
| | 262-263 | | | | | | | |
| | 50-51 | | | | | | | |

Tableau 3.5 : Masses moyennes des fragments de clivage du locus CDKN2A-*msp1* de l'élongation linéaire : (A) ATP et (B) GTP.

Grâce à l'information de l'analyse des masses des fragments de clivage des deux élongations ATP et GTP, le re-séquençage du locus *CDKN2A-msp1* est réalisé.

Ce locus contient 16 CpGs (tableau 3.5). Tous les CpGs sont potentiellement détectables par MALDI-TOF MS avec l'élongation ATP tandis que seulement 13 CpGs sont potentiellement détectables avec l'élongation GTP. Les trois CpGs restant sont présents dans des fragments de clivage de 1 à 2 bases, ils ne peuvent donc pas être détectés par MALDI-TOF MS.

De plus, avec l'élongation ATP, 12 CpGs possèdent une masse distincte tandis qu'avec l'élongation GTP, seulement 2 CpGs possèdent une masse distincte. Tous les CpGs ne sont donc pas utilisables pour la réalisation d'une courbe d'étalonnage à partir de l'abondance des ions issus des fragments de clivage des CpGs.

3.2.2. L'élongation linéaire ATP

Les spectres de masse de l'élongation linéaire ATP qui identifie le CpG 115-116 sont présentés dans la figure 3.4.

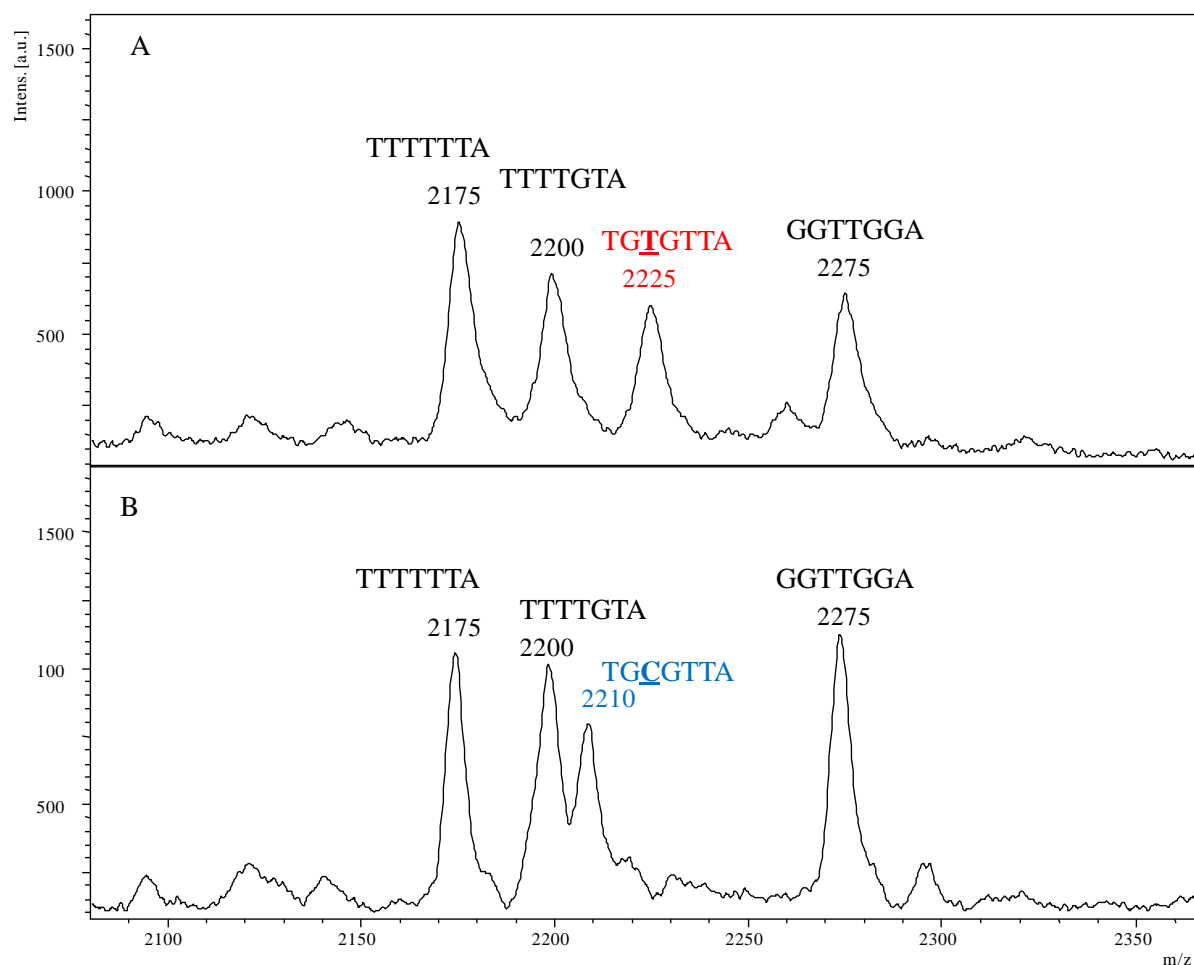


Figure 3.4 : Spectres de masse MALDI-TOF en mode linéaire de la région des 7-mer de l'élongation linéaire ATP des ions $[M-H]^-$ des fragments de clivage du locus *CDKN2A-*msp1** d'un ADN: (A) 0% méthylé et (B) 100% méthylé. L'ion $[M-H]^-$ TGTTGTTA, en rouge, représente le CpG 0% méthylé alors que l'ion $[M-H]^-$ TGCGTTA, en bleu, représente le CpG 100% méthylé. Les autres ions $[M-H]^-$ sont les fragments de clivage du re-séquençage du locus.

Dans la région des 7-mer, ils montrent à m/z 2175 l'ion $[M-H]^-$ du fragment 141-147 TTTTTTA ($M=2172,4$ Da), à m/z 2200 l'ion $[M-H]^-$ du fragment 134-140 TTTTGTA ($M=2197,4$ Da) et à m/z 2275 les ions $[M-H]^-$ des fragments 97-103 et 348-354 GGTTGGA ($M=2272,4$ Da). Ils correspondent aux fragments de re-séquençage du locus.

Les spectres de masse montrent également à m/z 2210 l'ion $[M-H]^-$ du fragment 113-119 TGCGTTA ($M=2207,4$ Da) et à m/z 2225 l'ion $[M-H]^-$ du fragment 113-119 TGTTGTTA ($M=2222,4$ Da). Ils correspondent respectivement à la cytosine méthylée et la cytosine non méthylée du CpG 115-116.

Les rapports m/z des ions sont légèrement différents de ceux attendus mais l'écart de masse est toujours constant. Cet écart est dû à l'analyse par MALDI-TOF MS en mode linéaire qui a

été étalonnée à l'aide d'un étalon externe. Les rapports m/z des ions sont comparés aux masses de la séquence de référence afin de re-séquencer le locus étudié de chaque individu. De plus, le CpG 115-116 qui possède une masse spécifique peut être utilisé pour réaliser une courbe d'étalonnage afin de déterminer le taux de méthylation d'un individu inconnu.

3.2.3. L'élongation linéaire GTP

Les spectres de masse de l'élongation linéaire GTP qui identifie les CpGs 128-129, 163-164, 186-187, 218-219 sont présentés dans la figure 3.5.

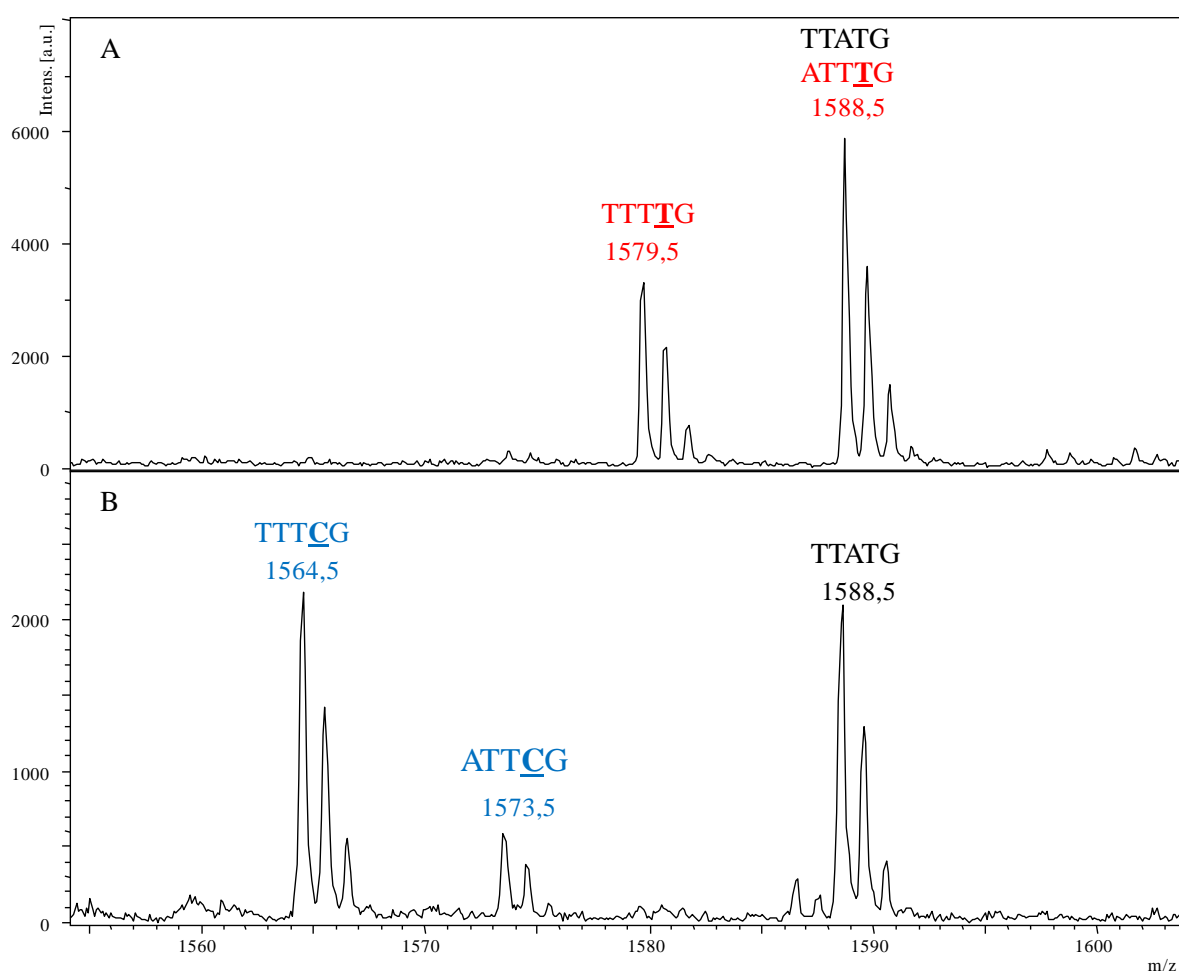


Figure 3.5 : Spectres de masse MALDI-TOF en mode réflecteur de la région des 5-mer de l'élongation linéaire GTP des ions $[M-H]$ des fragments de clivage du locus *CDKN2A-*msp1** d'un ADN: (A) 0% méthylé et (B) 100% méthylé. Les ions $[M-H]$ $ATTTG$, $TTTCG$, en rouge, représentent les CpGs 0% méthylé alors que les ions $[M-H]$ $ATTCG$, $TTTCG$, en bleu, représentent les CpGs 100% méthylé. L'ion $[M-H]$ $TTATG$ est un fragment de clivage du re-séquençage du locus.

Les masses indiquées sont la moyenne des masses isotopiques. Les spectres de masse en mode réflecteur peuvent identifier la présence de fragment de clivage de masses proches qui ne sont pas différenciées en mode linéaire et inférieures à 4000 Da.

Dans la région des 5-mer, ils montrent à m/z 1564,5 les ions [M-H]⁻ des fragments TTTCG (M=1565,0 Da) des CpGs méthylés 125-129, 160-164, 183-187, 215-219 et à m/z 1573,5 l'ion [M-H]⁻ du fragment ATTCG (M=1574,0 Da) du CpG méthylé 206-210.

Les spectres de masse montrent également à m/z 1579,5 les ions [M-H]⁻ des fragments TTTTG (M=1580,0 Da) des CpGs non méthylés 125-129, 160-164, 183-187, 215-219 et à m/z 1588,5 l'ion [M-H]⁻ du fragment TTATG (M=1589,0 Da) de re-séquençage et à l'ion [M-H]⁻ du fragment ATTTG (M=1589,0 Da) du CpG non méthylé 206-210.

Dans ce cas, les CpGs n'ont pas de masses propres, ils ne peuvent donc pas être utilisés pour réaliser des courbes d'étalonnage. Par contre, il y a d'autres CpGs qui peuvent être utilisés pour évaluer le taux de méthylation de l'ADN.

Les élongations ATP et GTP permettent donc d'une part, de re-séquencer le locus étudié et d'autre part, de déterminer le taux de méthylation de chaque locus.

3.3. La quantification de la méthylation de l'ADN

La quantification de la méthylation de l'ADN est réalisée par l'intermédiaire d'une courbe d'étalonnage.

La courbe d'étalonnage d'un locus est réalisée en fonction des ADNs méthylés de 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 à 100% et de l'intensité des pics des ions [M-H]⁻ du fragment non méthylé et du fragment méthylé qui contient le CpG étudié. La courbe d'étalonnage est réalisée à partir de 4 séries indépendantes et chaque série a été déposée 6 fois sur la cible MALDI (Tost et al. 2003).

Le rapport d'intensité des pics des ions est calculé à l'aide de la formule suivante :

$$y = \frac{100 * I \text{ méthylé}}{I \text{ méthylé} + I \text{ non méthylé}}$$

3.3.1. La courbe d'étalonnage du locus *CDKN2A*-*msp1*

La courbe d'étalonnage du CpG 91-92 du locus *CDKN2A*-*msp1* est réalisée à partir du clivage de l'élongation linéaire ATP. La figure 3.6 présente les différents spectres de masse des ADNs méthylés 0%, 20%, 50%, 70% et 100% pour l'étude de ce CpG.

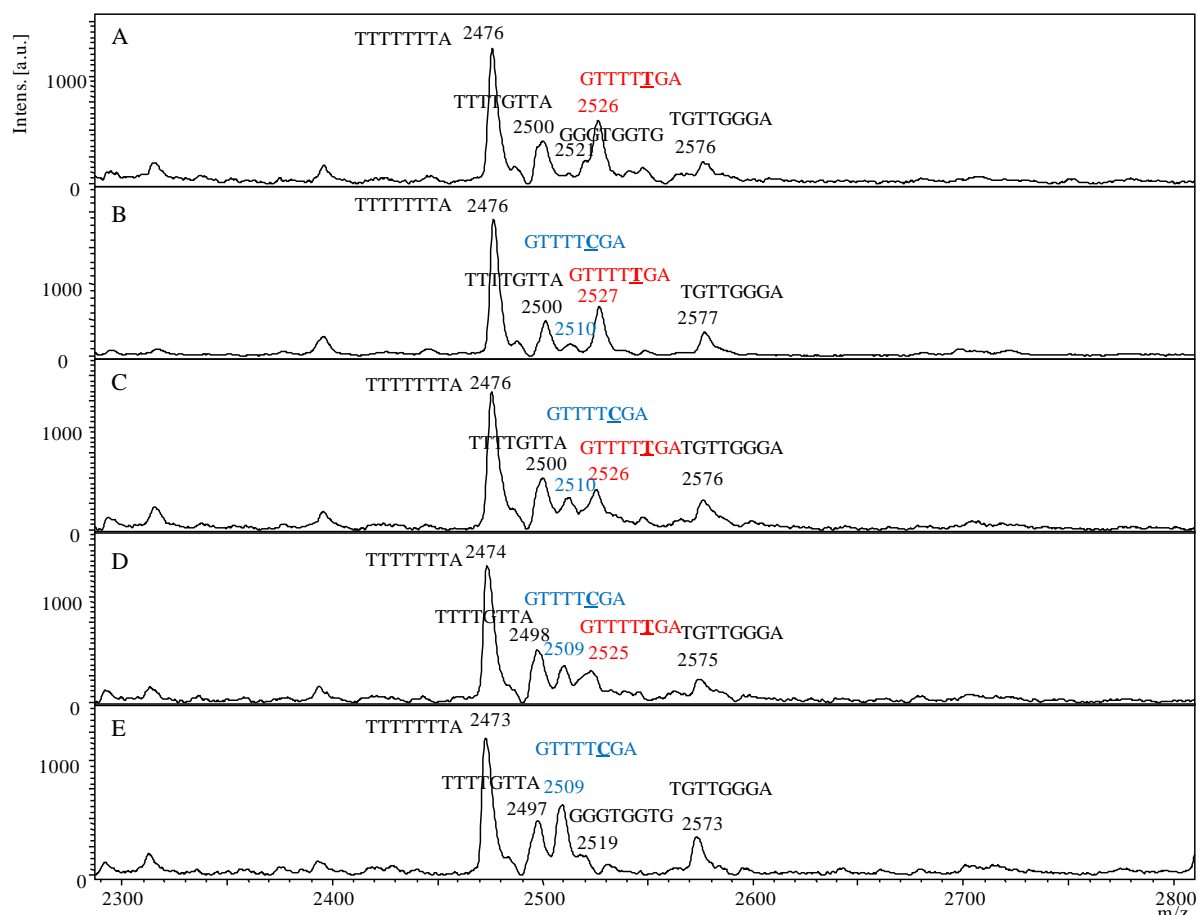


Figure 3.6 : Spectres de masse MALDI-TOF en mode linéaire de la région des 8-mer de l'élongation linéaire ATP des ions $[M-H]^-$ des fragments de clivage du locus *CDKN2A*-*msp1* d'un ADN: (A) 0% méthylé ; (B) 20% méthylé ; (C) 50% méthylé ; (D) 70% méthylé et (E) 100% méthylé. L'ion $[M-H]^-$ GTTTTTGA, en rouge, représente le CpG 0% méthylé alors que l'ion $[M-H]^-$ GTTTTTGA, en bleu, représente le CpG 100% méthylé. Les ions $[M-H]^-$ TTTTTTTA, TTTTCTTA, GGGTGGTG et TGTTGGGA sont des fragments de clivage du re-séquençage du locus.

Dans la région des 8-mer, ils montrent à m/z 2476 l'ion $[M-H]^-$ du fragment 73-80 TTTTTTTA ($M=2476,6$ Da), à m/z 2500 l'ion $[M-H]^-$ du fragment 336-343 TTTTGTTA ($M=2501,6$ Da), à m/z 2521 l'ion $[M-H]^-$ du fragment 355-363 GGGTGGTG ($M=2521,7$ Da) et à m/z 2576 l'ion $[M-H]^-$ du fragment 231-238 TGTTGGGA ($M=2576,7$ Da). Ils correspondent aux fragments de re-séquençage du locus.

Les spectres de masse montrent également à m/z 2510 l'ion $[M-H]^-$ du fragment GTTTTTGA ($M=2511,6$ Da) et à m/z 2525 l'ion $[M-H]^-$ du fragment GTTTTTGA ($M=2526,6$ Da). Ils

correspondent respectivement à la cytosine méthylée et la cytosine non méthylée du CpG 91-92.

L'intensité des deux pics des ions des fragments du CpG est donc utilisée pour réaliser une courbe d'étalonnage dans la figure 3.7.

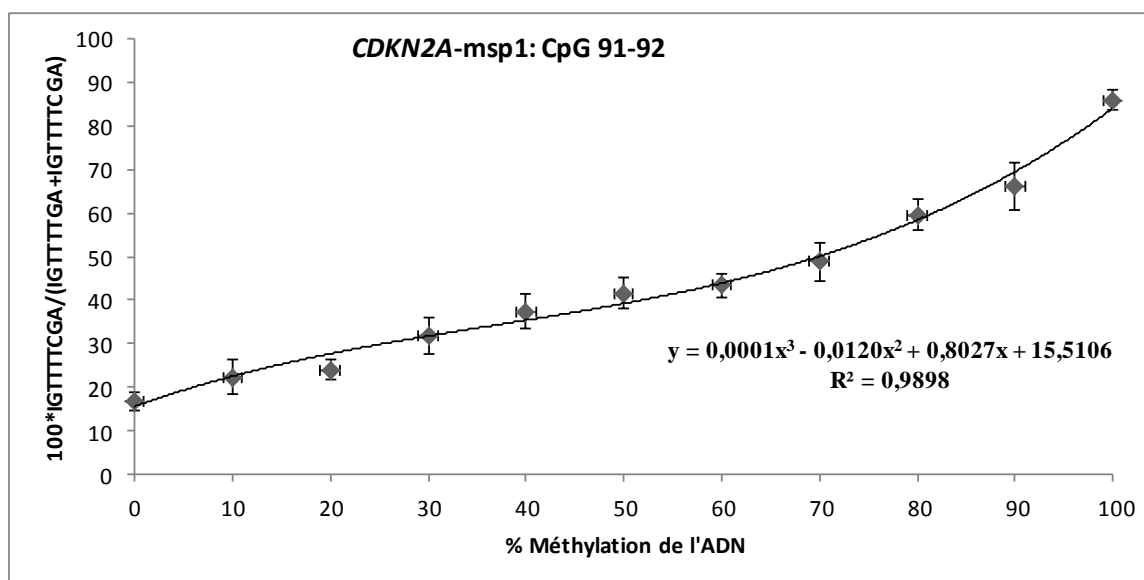


Figure 3.7 : Courbe d'étalonnage de l'abondance des ions [M-H] GTTTTGA et GTTTTCGA des fragments de clivage du CpG 91-92 de l'élongation linéaire ATP du locus CDKN2A-msp1. L'ion [M-H] GTTTTGA représente le CpG 0% méthylé alors que l'ion [M-H] GTTTTCGA représente le CpG 100% méthylé. L'axe des abscisses représente le pourcentage de méthylation de l'ADN. L'axe des ordonnées est le rapport d'intensité du pic de l'ion [M-H] GTTTTCGA sur la somme des intensités des pics des deux ions [M-H] GTTTTGA et GTTTTCGA. Les barres d'erreurs représentent l'écart type. La courbe d'étalonnage est une courbe polynomiale du troisième degré.

La courbe d'étalonnage est une courbe polynomiale du troisième degré d'équation:

$$y = 0,0001 x^3 + 0,0120 x^2 + 0,8027 x + 15,5106$$

Le coefficient de régression de la courbe est R^2 0,9898. Les écarts types varient de 2 à 5 % suivants les points. De plus, la courbe d'étalonnage ne passe pas par l'ordonnée à l'origine en raison des fluctuations de la ligne de base. Elles peuvent empêcher une bonne détection des ions de faibles concentrations.

3.3.2. La courbe d'étalonnage du locus CDKN2A-msp3b

Une autre courbe d'étalonnage a donc été réalisée sur le locus *CDKN2A*-msp3b

3.3.2.1. La courbe d'étalonnage à partir des mélanges des ADNs 0 et 100%

La courbe d'étalonnage du CpG135-136 du locus *CDKN2A*-msp3b est réalisée à partir du clivage de l'élongation linéaire ATP. La figure 3.8 suivante présente les différents spectres de masse des ADNs méthylés 0%, 20%, 50%, 70% et 100% pour l'étude de ce CpG.

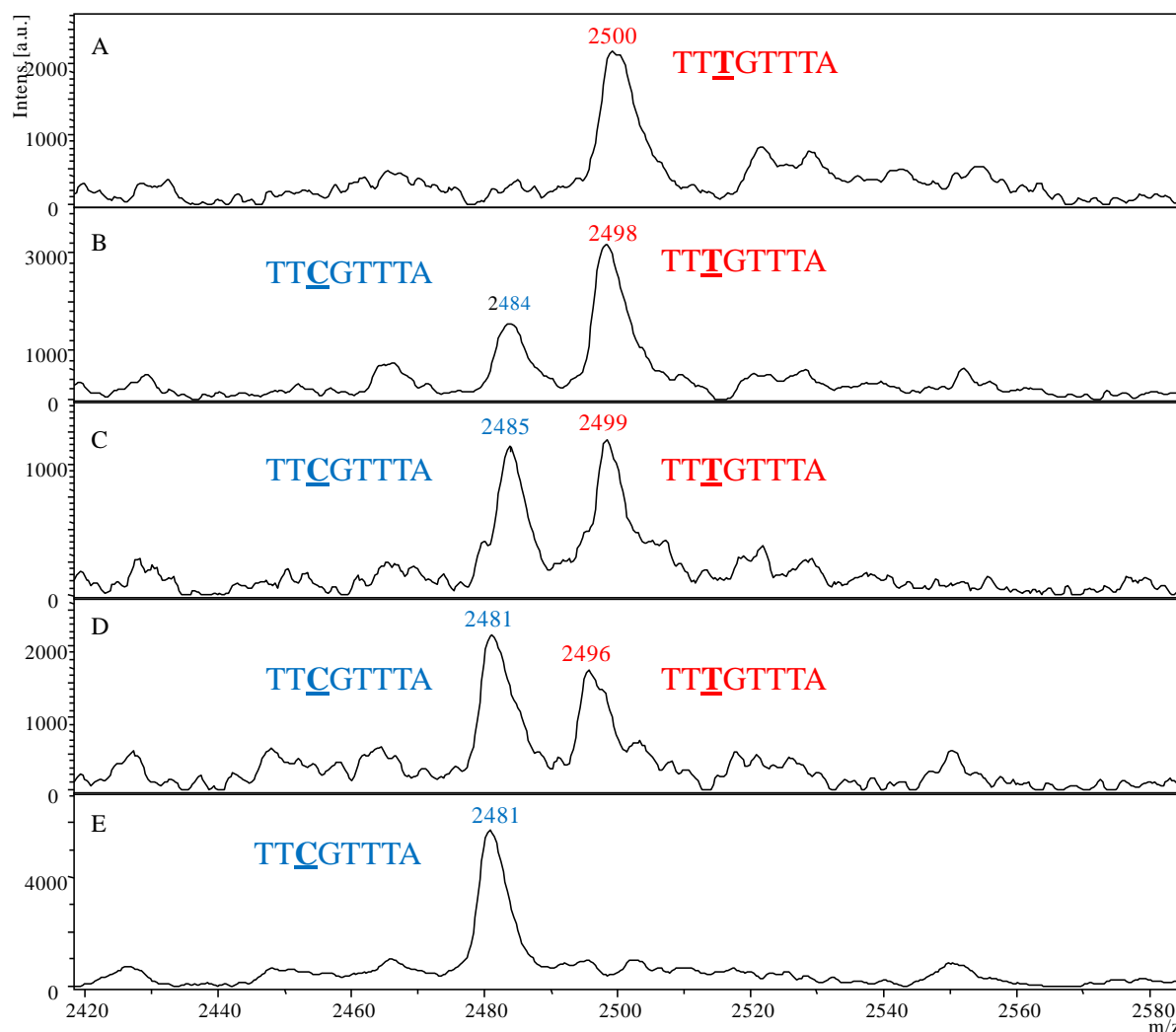


Figure 3.8 : Spectres de masse MALDI-TOF en mode linéaire de la région des 8-mer de l'élongation linéaire ATP des ions $[M-H]^+$ des fragments de clivage du locus *CDKN2A*-msp3b d'un ADN: (A) 0% méthylé ; (B) 20% méthylé ; (C) 50% méthylé ; (D) 70% méthylé et (E) 100% méthylé. L'ion $[M-H]^+$ TTTGTTTA, en rouge, représente le CpG 0% méthylé alors que l'ion $[M-H]^+$ TTCGTTTA, en bleu, représente le CpG 100% méthylé.

Dans la région des 8-mer, ils montrent à m/z 2500 l'ion $[M-H]^-$ du fragment $TTT\underline{T}GTTTA$ ($M=2501,6$ Da) et à m/z 2485 l'ion $[M-H]^-$ du fragment $TT\underline{C}GTTTA$ ($M=2486,6$ Da). Ils correspondent respectivement à la cytosine non méthylée et la cytosine méthylée du CpG 135-136.

L'intensité des deux pics des ions des fragments du CpG est donc utilisée pour réaliser une courbe d'étalonnage dans la figure 3.9.

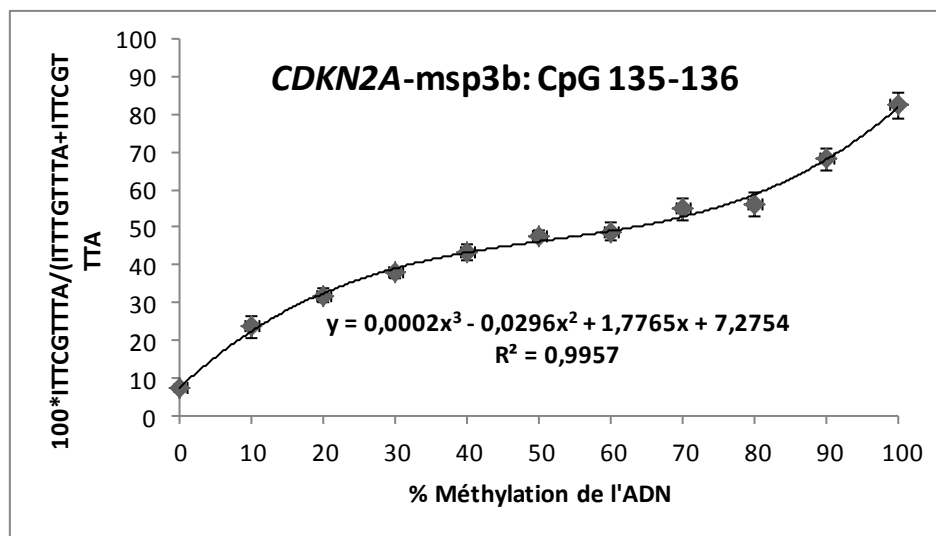


Figure 3.9 : Courbe d'étalonnage de l'abondance des ions $[M-H]^- TTT\underline{T}GTTTA$ et $TT\underline{C}GTTTA$ des fragments de clivage du CpG135-136 de l'élongation linéaire ATP du locus *CDKN2A-msp3b*. L'ion $[M-H]^- TTT\underline{T}GTTTA$ représente le CpG 0% méthylé alors que l'ion $[M-H]^- TT\underline{C}GTTTA$ représente le CpG 100% méthylé. L'axe des abscisses représente le pourcentage de méthylation de l'ADN. L'axe des ordonnées est le rapport d'intensité du pic de l'ion $[M-H]^- TT\underline{C}GTTTA$ sur la somme des intensités des pics des deux ions $[M-H]^- TTT\underline{T}GTTTA$ et $TT\underline{C}GTTTA$. Les barres d'erreurs représentent l'écart type. La courbe d'étalonnage est une courbe polynomiale du troisième degré.

La courbe d'étalonnage est une courbe polynomiale du troisième degré d'équation :

$$y = 0,0002 x^3 - 0,0296 x^2 + 1,7765 x + 7,2754$$

Le coefficient de régression de la courbe est $R^2 0,9957$. Les écarts types varient de 1 à 3,5 % suivants les points.

Les courbes d'étalonnage du CpG 91-92 du locus *CDKN2A-msp1* et du CpG 135-136 du locus *CDKN2A-msp3b* sont des polynomiales du troisième degré au lieu de droites.

L'équation de l'évolution non linéaire peut venir de la nature semi-quantitative du MALDI-TOF MS ou bien d'un biais de PCR. Une nouvelle courbe d'étalonnage a donc été élaborée à partir du mélange des produits de PCRs et non pas des mélanges des ADN.

3.3.2.2. La courbe d'étalonnage à partir des mélanges des PCRs 0 et 100%

La figure 3.10 présente la courbe d'étalonnage du CpG 135-136 du locus *CDKN2A-msp3b* à partir du mélange des PCR 0 et 100% ADN méthylé.

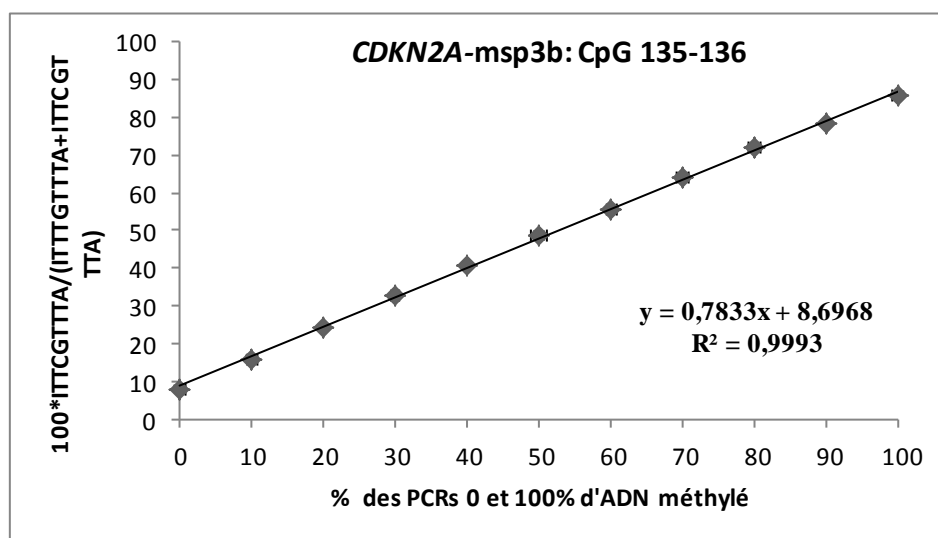


Figure 3.10 : Courbe d'étalonnage de l'abondance des ions $[M-H]$ $TTT\underline{G}TTTA$ et $TTC\underline{G}TTTA$ des fragments de clivage du CpG135-136 de l'élongation linéaire ATP du locus *CDKN2A-msp3b*. L'ion $[M-H]$ $TTT\underline{G}TTTA$ représente le CpG 0% méthylé alors que l'ion $[M-H]$ $TTC\underline{G}TTTA$ représente le CpG 100% méthylé. L'axe des abscisses représente le pourcentage des deux PCRs 0 et 100% d'ADN méthylé. L'axe des ordonnées est le rapport d'intensité du pic de l'ion $[M-H]$ $TTC\underline{G}TTTA$ sur la somme des intensités des pics des deux ions $[M-H]$ $TTT\underline{G}TTTA$ et $TTC\underline{G}TTTA$. Les barres d'erreurs représentent l'écart type. La courbe d'étalonnage est droite.

La courbe d'étalonnage est une droite d'équation:

$$y = 0,783 x + 8,696$$

Le coefficient de régression de la courbe est R^2 0,999. Les écarts types varient de 0,5 à 1 % suivant les points.

La courbe d'étalonnage étant linéaire, il y a donc bien un biais de PCR et pas un problème de reproductibilité des analyses au MALDI-TOF MS. Des expériences par PCR quantitative

pourraient infirmer ou confirmer cette hypothèse (Smith et al. 2009). L'analyse des différents mélanges d'ADN du locus *CDKN2A-msp3b* par PCR quantitative a confirmé le biais de PCR.

3.4. L'analyse des individus

3.4.1. Les loci des gènes *CDKN2A* et *RASSF1A*

L'étude du gène *CDKN2A* est composée de 9 loci. Les loci 1 et 2 sont situés en amont du site du début de transcription tandis que les loci 3a et 3b couvrent le site de début de transcription. Les loci 3b et 4 couvrent le premier exon tandis que les loci 5, 6 et 7 couvrent le premier intron. Le locus 8 couvre le second exon tandis que le locus 9 couvre le second intron.

L'étude du gène *RASSF1A* est composée de 3 loci. Le locus 1 est situé en amont du site du début de transcription tandis que le locus 2 couvre le site de début de transcription. Le locus 3 couvre le premier exon et le premier intron.

3.4.2. L'analyse des loci des gènes *CDKN2A* et *RASSF1A*

Le tableau 3.6 montre le re-séquençage et l'identification des loci hypométhylés ou hyperméthylés des deux gènes *CDKN2A* et *RASSF1A* des 7 individus.

| Locus | Gène | Echantillons d'ADN | | | | | | |
|---------------------|---------------------|--------------------|----|-------|-------|----|------|------|
| | | T3 | A1 | T2P37 | T2P28 | 23 | HMEC | T47D |
| <i>CDKN2A_msp1</i> | amont promoteur | X | X | X | X | X | | X |
| <i>CDKN2A_msp2</i> | amont promoteur | | X | X | X | | | X |
| <i>CDKN2A_msp3a</i> | promoteur | | | | | | | |
| <i>CDKN2A_msp3b</i> | promoteur et exon 1 | | | | | | X | X |
| <i>CDKN2A_msp4</i> | exon 1 | | | | | | | |
| <i>CDKN2A_msp5</i> | intron 1 | | | | | | | X |
| <i>CDKN2A_msp6</i> | intron 2 | | | | | | X | X |
| <i>CDKN2A_msp7</i> | intron 3 | X | X | X | X | X | | X |
| <i>CDKN2A_msp8</i> | exon 2 | X | X | X | X | X | X | X |
| <i>CDKN2A_msp9</i> | intron 2 | X | X | X | X | X | | X |
| <i>RASSF1A_msp1</i> | amont promoteur | | | | | | | |
| <i>RASSF1A_msp2</i> | promoteur | | | | | | X | X |
| <i>RASSF1A_msp3</i> | exon et intron 1 | X | | X | X | X | X | X |

Tableau 3.6 : Cartographie des loci des gènes *CDKN2A* et *RASSF1A* hypométhylés ou hyperméthylés des 7 individus. Les loci hyperméthylés sont notés par une croix tandis que les loci hypométhylés ne le sont pas. Les loci qui sont grisés indiquent qu'il n'y a pas eu de résultats.

Chaque individu possède des loci hypométhylés et hyperméthylés pour les deux gènes *CDKN2A* et *RASSF1A*.

Pour l'étude du gène *CDKN2A*, les individus T3 et 23 sont hyperméthylés sur les loci 1, 7, 8 et 9. Les individus A1, T2P37 et T2P28 sont hyperméthylés sur les loci 1, 2, 7, 8 et 9. L'individu HMEC est hyperméthylé sur les loci 3b, 6, 8. L'individu T47D est hyperméthylé sur les loci 1, 2, 3b, 5, 6, 7, 8 et 9. Seuls les individus HMEC et T47D sont hyperméthylés pour la fin du site début du site de transcription et du début du premier exon.

Pour l'étude du gène *RASSF1A*, les individus T3, T2P37, T2P28 et 23 sont hyperméthylés sur le locus 3 qui couvre le premier exon et intron. Pour les individus A1, HMEC et T47D, l'analyse du locus 1 du gène *RASSF1A* n'a pas pu être réalisée à cause d'un problème expérimental et par manque d'ADNs pour refaire l'expérience. Les ADNs utilisés, proviennent d'autres projets, ils sont donc en faible quantité. Pour l'étude des loci 2 et 3, l'individu A1 est hypométhylé tandis que les individus HMEC et T47D sont hyperméthylés. Seul l'individu A1 est hypométhylé au niveau du premier exon et intron. De plus, seuls les individus HMEC et T47D sont hyperméthylés sur la partie de début de transcription.

Les individus HMEC et T47D pourraient avoir une réduction de l'expression des deux gènes *CDKN2A* et *RASSF1A*.

3.4.3. Le pourcentage de méthylation de l'ADN des loci *CDKN2A*-msp1et msp3b

Les spectres de masse du CpG 91-92 du locus *CDKN2A*-msp1 et du CpG 135-136 du locus *CDKN2A*-msp3b de 2 individus HMEC et T2P8 sont montrés dans la figure 3.11.

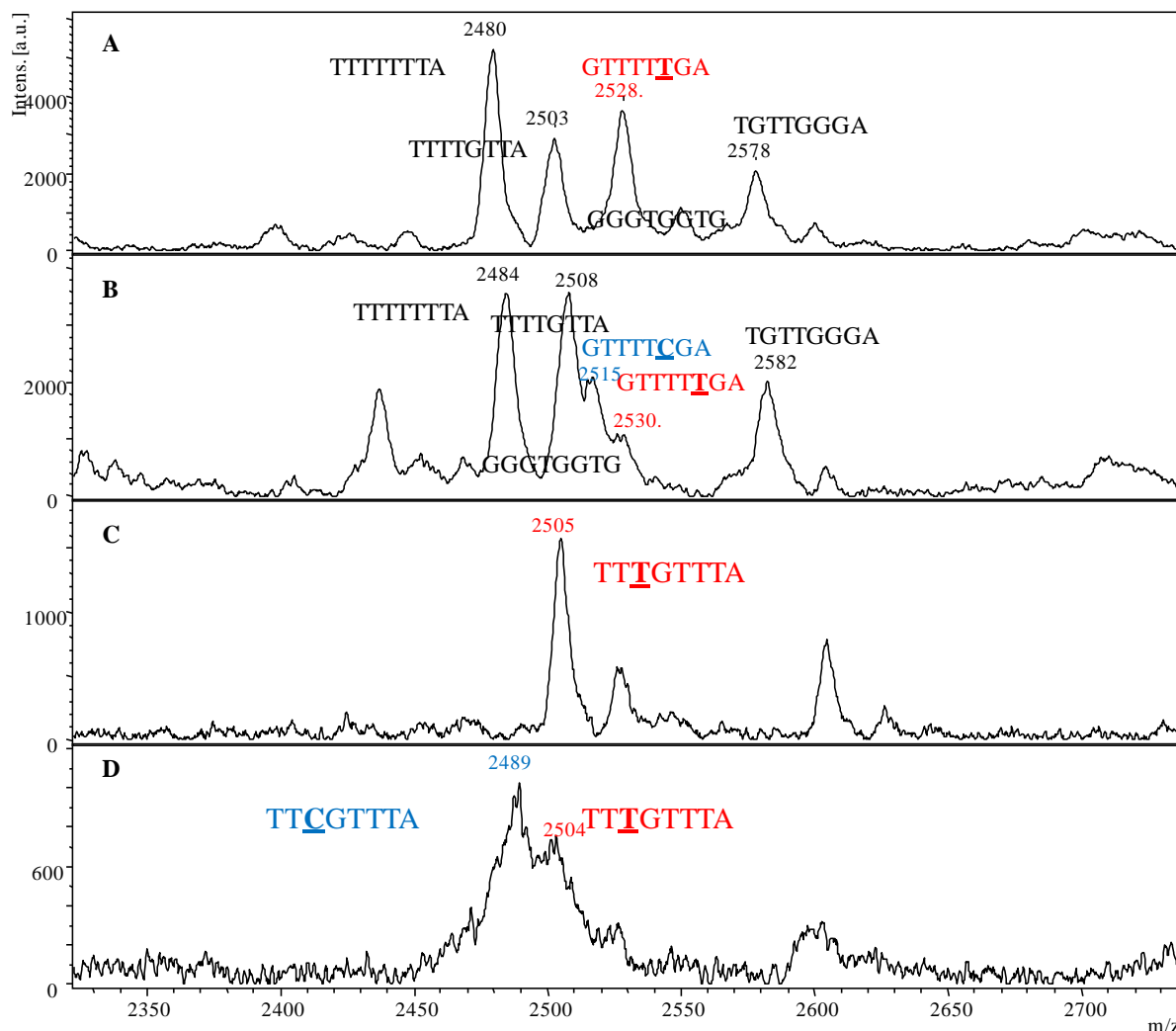


Figure 3.11 : Spectres de masse MALDI-TOF en mode linéaire de la région des 8-mer de l'élongation linéaire ATP des ions $[M-H]^-$ des fragments de clivage du locus : (A) *CDKN2A-msp1* de l'individu HMEC; (B) *CDKN2A-msp1* de l'individu T2P28; (C) *CDKN2A-msp3b* de l'individu T2P28 et (D) *CDKN2A-msp3b* de l'individu HMEC. Les ions $[M-H]^-$ $TTTTGTTA$ et $GTTTTTGA$, en rouge, représente le CpG 0% méthylé alors que les ions $[M-H]^-$ $TTCGTTA$ et $GTTTTTGA$, en bleu, représente le CpG 100% méthylé. Les ions $[M-H]^-$ $TTTTTTTA$, $TTTTCTTA$, $GGGTGGTG$ et $TGTTGGGA$ sont des fragments de clivage du re-séquençage du locus *CDKN2A-msp1*.

Le spectre de masse (figure 3.11A) de l'individu HMEC montre à m/z 2528 l'ion $[M-H]^-$ du fragment $GTTTTTGA$ ($M=2526,6$ Da). Il correspond à la cytosine non méthylée du CpG 91-92 du locus *CDKN2A-msp1*. Le spectre de masse (figure 3.11D) de l'individu HMEC montrent à m/z 2504 l'ion $[M-H]^-$ du fragment $TTTGTGTTA$ ($M=2501,6$ Da) et à m/z 2489 l'ion $[M-H]^-$ du fragment $TTCGTTA$ ($M=2486,6$ Da). Ils correspondent respectivement à la cytosine méthylée et la cytosine non méthylée du CpG 135-136 du locus *CDKN2A-msp3b*. L'ADN de cet individu étant en très faible quantité, il n'a pas été possible d'obtenir un spectre de masse de meilleure qualité (figure 3.11D).

Le spectre de masse (figure 3.11B) de l'individu T2P28 montre à m/z 2515 l'ion [M-H]⁻ du fragment GTTTTCGA (M=2511,6 Da) et à m/z 2530 l'ion [M-H]⁻ du fragment GTTTTTGA (M=2526,6 Da). Ils correspondent respectivement à la cytosine méthylée et non méthylée du CpG 91-92 du locus *CDKN2A*-msp1. Le spectre de masse (figure 3.11D) de l'individu T2P28 montrent à m/z 2505 l'ion [M-H]⁻ du fragment TTTTGTTA (M=2501,6 Da). Il correspond à la cytosine méthylée du CpG 135-136 du locus *CDKN2A*-msp3b.

L'intensité des pics de ces ions est utilisée pour calculer les rapports d'intensité et ainsi déterminer le taux de méthylation par l'intermédiaire de l'équation des courbes d'étalonnages des deux CpGs (tableau 3.7).

| | | Individus | | | | | | |
|--|---|------------|-----------|-----------|-----------|-----------|-----------|-----------|
| | | T3 | A1 | T2P37 | T2P28 | 23 | HMEC | T47D |
| <i>CDKN2A</i>_msp1 CpG 91-92 | Intensité GTTTT <u>C</u> GA | 1164,71 | 1410,39 | 1977,87 | 2095,24 | 2117,20 | | 2623,00 |
| | Intensité GTTTT <u>T</u> GA | | 718,78 | 1520,47 | 1035,52 | 1125,91 | 3611,05 | 1455,19 |
| | $100 \cdot \frac{I_{\text{GTTTTCGA}}}{(I_{\text{GTTTTCGA}} + I_{\text{GTTTTTGA}})}$ | 1 | 66 | 57 | 67 | 65 | 0 | 64 |
| | % de méthylation de l'ADN | 100 | 93 | 83 | 93 | 92 | 0 | 91 |
| <i>CDKN2A</i>_msp3b CpG 135-136 | Intensité GGGT <u>C</u> GGA | | | | | | 1028,85 | 318,09 |
| | Intensité GGGT <u>T</u> GGA | 4248,28 | 3078,22 | 2577,23 | 1573,05 | 2280,31 | 755,28 | 129,25 |
| | $100 \cdot \frac{I_{\text{GGGTGGA}}}{(I_{\text{GGGTGGA}} + I_{\text{GGGTTCGA}})}$ | 0 | 0 | 0 | 0 | 0 | 58 | 71 |
| | % de méthylation de l'ADN | 0 | 0 | 0 | 0 | 0 | 74 | 88 |

Tableau 3.7 : Détermination du taux de méthylation de l'ADN pour le CpG 91-92 du locus *CDKN2A*-msp1 et du CpG 135-136 du locus *CDKN2A*-msp3b.

Pour le CpG 91-92 du locus *CDKN2A*-msp1, situé en amont du site de début de transcription, 6 individus sont hyperméthylés de 83 à 100% et seul l'individu HMEC n'est pas méthylé. Pour le CpG 135-136 du locus *CDKN2A*-msp3b, situé à la fin du site de début de transcription, 5 individus ne sont pas méthylés et les individus HMEC et T47D sont hyperméthylés (74 à 88%).

4. Discussion

Cette méthode de re-séquençage pour l'étude de la méthylation de l'ADN est basée sur celle du re-séquençage de l'ADN mitochondrial (Mauger et al. 2007) mais elle diffère d'une part, par une étape supplémentaire préalable de conversion de l'ADN méthylé et d'autre part, par la quantification de la méthylation de l'ADN. De plus, elle permet l'analyse de l'ADN génomique qui est beaucoup plus complexe que l'analyse de l'ADN mitochondrial.

L'ADN est tout d'abord converti par le bisulfite de sodium afin de différencier les cytosines des 5-méthylcytosines des CpGs. Le locus est amplifié par PCR à l'aide des amorces de PCR. Les chimères ARN/ADN sont synthétisées à l'aide d'une amorce d'élongation, d'une ADN polymérase qui incorpore les bases dNTPs et NTPs et le produit de PCR purifié. Elles sont ensuite clivées par la soude et dessalées par de résine échangeuse de cations. Les fragments de clivage sont analysés par MALDI-TOF MS en mode négatif. L'analyse des masses des fragments permet de re-séquencer le locus du gène et d'analyser les CpGs. La courbe d'étalonnage réalisée à partir de l'intensité des pics des ions fragments 0 et 100% méthylés du CpG permet de quantifier la méthylation de l'ADN des individus.

4.1. La conversion au bisulfite

Les cytosines et les 5-méthylcytosines des CpGs ne peuvent pas être différenciées par l'ADN polymérase de la PCR. Il est donc nécessaire de pouvoir les différencier à l'aide d'une réaction au bisulfite de sodium. La cytosine est convertie en uracile alors que la 5'-méthylcytosine n'est pas modifiée.

Cette réaction de conversion est couramment utilisée dans les méthodes d'analyse de la méthylation de l'ADN.

4.2. La PCR

4.2.1. Les amorces

La réaction de PCR sélectionne et amplifie le locus à étudier à l'aide des amorces de PCR. L'analyse de la méthylation de l'ADN entraîne une certaine restriction dans le choix des amorces de PCR par rapport à l'analyse de l'ADN. A cause de la conversion au bisulfite, les cytosines sont maintenant des uraciles alors que les 5-méthylcytosine n'ont pas été modifiées. Les deux brins d'ADN ne sont donc plus complémentaires. Les amorces ne doivent donc pas contenir de CpGs afin d'amplifier les deux brins d'ADN non complémentaires.

4.2.2. Le biais de PCR

L'ADN polymérase doit pouvoir amplifier de l'ADN bisulfite qui contient des bases uraciles. De plus, les conditions de PCR doivent être optimales afin d'amplifier tous les brins avec la même efficacité. En effet, les brins issus des CpGs méthylés contiennent plus de C alors que les brins issus des CpGs non méthylés contiennent plus de U. Les brins non méthylés, pauvres en GC, se dénaturent donc plus vite que les brins méthylés riches en GC. Il pourrait en résulter une amplification préférentielle ou spécifique du brin non méthylé (Warnecke et al. 1997). Cette amplification spécifique du brin méthylé peut poser un problème lors de la quantification de l'ADN. Le nombre de paires de bases et de CpGs présents dans chaque PCR est montré dans le tableau 3.8.

| Locus | PCR | |
|---------------------|-----|-----|
| | pb | CpG |
| <i>CDKN2A_msp1</i> | 432 | 17 |
| <i>CDKN2A_msp2</i> | 305 | 9 |
| <i>CDKN2A_msp3a</i> | 269 | 13 |
| <i>CDKN2A_msp3b</i> | 288 | 16 |
| <i>CDKN2A_msp4</i> | 356 | 33 |
| <i>CDKN2A_msp5</i> | 414 | 15 |
| <i>CDKN2A_msp6</i> | 436 | 18 |
| <i>CDKN2A_msp7</i> | 442 | 38 |
| <i>CDKN2A_msp8</i> | 413 | 16 |
| <i>CDKN2A_msp9</i> | 438 | 29 |
| <i>RASSF1A_msp1</i> | 436 | 25 |
| <i>RASSF1A_msp2</i> | 435 | 51 |
| <i>RASSF1A_msp3</i> | 251 | 17 |

Tableau 3.8 : Nombre de paires de bases et nombre de CpGs des PCR des loci *CDKN2A* et *RASSF1A*.

Les PCRs des loci *RASSF1A*-mps2, *CDKN2A*-msp7 et *CDKN2A*-msp4 contiennent respectivement 51, 38 et 33 CpGs. Ces PCRs pourraient avoir un biais de PCR important (Warnecke et al. 1997).

Dans le cas de l'étude de la quantification du locus *CDKN2A*-msp3b qui contient 16 CpGs un biais de PCR a été démontré par PCR quantitative et par l'élaboration de la courbe d'étalonnage à partir du mélange des PCRs. La courbe d'étalonnage du locus *CDKN2A*-msp1 qui contient 17 CpGs possède la même allure que celle de *CDKN2A*-msp3b. Il se pourrait qu'il y ait également un biais de PCR pour la PCR du locus *CDKN2A*-msp1.

Les PCRs ont été optimisées sur des ADN 0, 50 et 100% méthylés et contrôlées par la méthode. Afin de minimiser ce biais de PCR, le temps d'élongation, d'appariement des

amorces, la température de dénaturation, la concentration de chlorure de magnésium, l'addition de diméthylsulfoxyde ou de formamide ou une autre ADN polymérase avec un fragment *Stoffel* pourraient être optimisés (Warnecke et al. 1997; Saiki et al. 1988). Quelques expériences ont été réalisées sur *CDKN2A-msp1* et *mps3b*, mais n'ont pas permis d'améliorer significativement les courbes d'étalonnage. L'analyse des courbes de fusion des PCRs par PCR quantitative pourrait permettre de mieux contrôler et d'optimiser les PCRs.

Récemment, une approche mathématique de correction des données de quantification a été développée pour améliorer le biais de PCR des ADNs méthylés (Moskalev et al. 2011). Elle inclut en parallèle de l'analyse des échantillons un étalonnage. L'écart entre les données théoriques et expérimentales sert à calculer une courbe de régression. Puis, les données des individus sont normalisées à l'aide de cette courbe afin de pouvoir les quantifier.

4.3. L'analyse de la méthylation de l'ADN

4.3.1. Le re-séquençage

Les loci de chaque gène se recouvrent partiellement en séquence afin de couvrir le gène et d'avoir le maximum d'informations complémentaires.

Deux chimères ARN/ADN simple-brin ont été synthétisées en parallèle (ATP et GTP) pour chaque locus. Par ailleurs, les chimères sont constituées d'environ 400 à 500 bases. Les informations complémentaires des masses des fragments de clivage des deux chimères ARN/ADN permettent d'obtenir le maximum d'informations sur les séquences.

L'analyse en MALDI-TOF MS s'effectue tout d'abord en mode linéaire car d'une part, l'analyse des masses moyennes est suffisante et d'autre part, elle permet une bonne détection des ions de 1 à 10 kDa.

Dans un deuxième temps, l'analyse en MALDI-TOF MS s'effectue en mode réflecteur mais seules les masses inférieures à 4 kDa peuvent être détectées. Cette analyse est utilisée dans le cas où les masses moyennes des fragments de clivage sont très proches et ne peuvent pas être différenciées par le mode linéaire car le pouvoir résolutif de l'instrument est trop faible.

L'utilisation du recouvrement de séquence des loci, les deux élongations et les deux modes de détection par MALDI-TOF MS ont permis de réaliser une cartographie massive et par conséquent le re-séquençage des loci des deux gènes.

4.3.2. L'analyse des CpGs

L'étude de la présence et de l'intensité des pics des ions des fragments de clivage qui contiennent des CpGs 0 ou 100% méthylés permet de mettre en évidence la présence de régions hypométhylées ou hyperméthylées pour les individus étudiés. Cependant, tous les CpGs présents ne peuvent pas être utilisés pour quantifier la méthylation de l'ADN car leurs masses moléculaires sont identiques à d'autres fragments de clivage. Le tableau 3.9 montre le nombre de CpGs des élongations des loci et le nombre de CpGs possédant une masse des ions des fragments de CpGs 0 et 100 % méthylés utilisable pour la quantification de la méthylation.

| Locus | CpG de l'élongation | |
|----------------------|---------------------|------------------|
| | CpG | CpG masse unique |
| <i>CDKN2A</i> _msp1 | 16 | 14 |
| <i>CDKN2A</i> _msp2 | 9 | 5 |
| <i>CDKN2A</i> _msp3a | 12 | 11 |
| <i>CDKN2A</i> _msp3b | 16 | 15 |
| <i>CDKN2A</i> _msp4 | 31 | 29 |
| <i>CDKN2A</i> _msp5 | 14 | 13 |
| <i>CDKN2A</i> _msp6 | 18 | 14 |
| <i>CDKN2A</i> _msp7 | 35 | 23 |
| <i>CDKN2A</i> _msp8 | 16 | 13 |
| <i>CDKN2A</i> _msp9 | 29 | 15 |
| <i>RASSF1A</i> _msp1 | 23 | 17 |
| <i>RASSF1A</i> _msp2 | 47 | 31 |
| <i>RASSF1A</i> _msp3 | 17 | 14 |

Tableau 3.9 : Nombre de CpGs des élongations et nombre de CpGs qui possèdent une masse moyenne unique des loci *CDKN2A* et *RASSF1A*.

Certains fragments contenant les CpGs possèdent une masse identique à celles des fragments de re-séquençage, ils ne sont donc pas utilisables pour déterminer le taux de méthylation de l'ADN. De plus, tous les fragments de clivage qui possèdent une masse distincte ne sont également pas tous utilisables pour la détermination du taux de méthylation de l'ADN. Plus les masses des fragments augmentent, plus les pics des ions s'élargissent et leurs intensités

diminuent. Les fragments sont donc moins détectables par MALDI-TOF MS. Par ailleurs, plus la gamme de masse est importante, plus la différence de masse entre les ions du 0 et 100% méthylés est inférieure au pouvoir résolutif de l'instrument ne permettant pas ainsi l'élaboration d'une courbe d'étalonnage. Par contre, la présence et l'intensité relative des ions des fragments de clivage du 0 et 100% méthylés du CpG peuvent permettre de cibler les régions hypométhylées ou hyperméthylées.

Pour l'élaboration des courbes d'étalonnage, il est donc préférable de choisir des masses des fragments des CpGs 0 et 100% méthylés qui possèdent une masse moléculaire inférieure à 3 kDa afin d'avoir une intensité et une résolution suffisante. Tous les CpGs ne sont donc pas tous utilisables pour la quantification de l'ADN méthylé. Cependant, cette méthode est une méthode de cartographie des régions hypométhylées et hyperméthylées. Dans chaque locus, il n'est donc pas nécessaire de quantifier précisément chaque CpG. De plus, les îlots CpGs d'un même locus, proches dans la séquence, ont probablement un taux de méthylation similaire. Pour chaque locus, l'élaboration de quelques courbes d'étalonnage pourrait donc être suffisante surtout que si les CpGs sont éloignés dans la séquence.

4.3.3. La quantification de la méthylation de l'ADN

La quantification de la méthylation de l'ADN est réalisée à l'aide d'une courbe d'étalonnage.

4.3.3.1. Le nombre de points de la courbe d'étalonnage

Les courbes d'étalonnage ont été réalisées à partir des ADNs 0 et 100 % méthylés et bisulfités. Ces ADNs ont été mélangés afin d'obtenir des incréments de 10 en 10. Pour la réalisation de la courbe d'étalonnage, 4 gammes de 0 à 100% ont été utilisées. Chaque échantillon a été déposé 6 fois, soit 24 points pour un échantillon (Tost et al. 2003). Ces dépôts et la somme des tirs des lasers permettent de s'affranchir de l'aspect semi-quantitatif du MALDI-TOF MS et de normaliser les données.

4.3.3.2. L'intensité ou l'aire des pics

L'intensité et l'aire des pics des ions ont été testées mais les courbes d'étalonnage réalisées à partir des aires ont un coefficient de régression plus faible. Il est plus facile de déterminer précisément la hauteur du pic de l'ion que son aire. En effet, plus la masse de l'ion est importante plus le pic de l'ion s'élargisse et donc l'analyse de son aire est difficile.

De plus, les fluctuations de la ligne de base empêchent une bonne détection des ions faiblement concentrés. Ce phénomène se rencontre surtout aux extrémités de la courbe d'étalonnage ce qui peut expliquer que les courbes ne passent pas par l'ordonnée à l'origine.

4.3.3.3. Le mode linéaire ou le mode réflecteur

Les courbes d'étalonnage ont été réalisées en mode réflecteur et en mode linéaire mais les courbes d'étalonnage réalisées en mode réflecteur ont un coefficient de régression plus faible que celle réalisée en mode linéaire. Les pics monoisotopiques des ions sont utilisés pour la courbe d'étalonnage en mode réflecteur. Elle est réalisée à l'aide de l'ion le plus intense du massif isotopique qui est dans ce cas le pic monoisotopique. Cependant, le pic monoisotopique possède une intensité moindre que celui du centroïde du massif isotopique non résolue en mode linéaire.

De plus, en mode réflecteur, la gamme de masse de détection des ions est de 4 kDa tandis que celle en mode linéaire est de 10 kDa. Cette différence est due à la correction temporelle des ions à leur entrée dans le réflecteur. Pour l'élaboration d'une courbe d'étalonnage en mode réflecteur, il faut donc privilégier l'étude de CpGs de plus petite masse ce qui restreint encore le choix des CpGs à étudier contrairement au mode linéaire.

4.3.3.4. Le choix de fragments de clivage des CpGs

Les courbes d'étalonnage ont été donc réalisées à partir du mode linéaire et avec l'intensité des pics des ions.

Plus la masse des ions des CpGs augmentent plus le coefficient de régression diminue. Les courbes d'étalonnage du locus *CDKN2A_msp1* et du locus *CDKN2A-msp3b* ont été réalisées avec des fragments de clivage des CpGs de 8-mer.

Par ailleurs, les fragments de clivage de CpGs qui contiennent plus de T ont un coefficient de régression plus important. Les bases T sont moins labiles que les autres bases car elles sont faiblement protonées lors de la formation d'un ion zwitterion en phase gazeuse (chapitre 1). Par conséquent, les ions qui contiennent plus de T fragmentent moins en phase gazeuse et donc sont plus facilement détectés que les autres car ils sont plus abondants. Les fragments de clivage du CpGs des courbes d'étalonnage sont GTTTTCGA et GTTTTTGA pour le locus *CDKN2A_msp1* et TTTGTTTA et TTCGTTTA pour le locus *CDKN2A_msp3b*.

4.3.3.5. Les courbes d'étalonnage

Les courbes d'étalonnage pour les deux loci étudiés sont des polynomiales du troisième degré. Le coefficient de régression R^2 est de 0,9898 pour le locus *CDKN2A_msp1* et il est de 0,9957 pour le locus *CDKN2A_msp3b*. L'écart type est aux alentours de 1 à 5% suivant les positions. L'écart type est plus important au niveau des extrémités 0 à 10% et 90-100% méthylés car les pics des ions se confondent avec les fluctuations de la ligne de base. La limite de détection est aux alentours de 5%.

La courbe d'étalonnage du locus *CDKN2A_msp3b* réalisée à partir du mélange des PCRs 0 et 100% est une droite linéaire et son coefficient de régression R^2 est de 0,999. L'écart type de chaque mesure est inférieur à 1%. La polynomiale du troisième degré provient donc plus du biais de PCR que de la répétabilité des mesures du MALDI-TOF MS. Le MALDI-TOF MS est semi-quantitatif car le laser tire sur des cristaux de matrice et la reproduction des tirs est difficile. Il dépend donc de la cristallisation homogène de la matrice. De plus, le spectre de masse est une somme de plusieurs tirs de laser.

Dans la méthode du *GOODASSAY* (Tost et al. 2003), les courbes d'étalonnage sont également des polynomiales du troisième degré avec un coefficient de régression R^2 supérieur à 0,99 et un écart type des mesures de 2%. La limite de détection est de 5%. Dans cette méthode, le biais de PCR peut également poser un problème. De plus, cette méthode n'est pas une méthode de re-séquençage mais seulement une méthode de génotypage.

La méthode du *MassARRAY*[®] (Ehrich et al. 2005; Coolen et al. 2007; Bellido et al. 2010) est également une méthode de re-séquençage mais les courbes d'étalonnage sont des évolutions linéaires et la limite de détection est de 5%. Dans ce protocole, plusieurs procédures ont été mises en place pour l'analyse haut-débit et notamment l'optimisation des PCRs et l'usage

d'automates de distribution. Les PCRs sont analysées à l'aide des courbes de fusion par PCR quantitative. Toutes ces procédures de contrôle permettent d'assurer une meilleure reproductibilité des données et ainsi mieux quantifier la méthylation de l'ADN. Afin d'améliorer les courbes d'étalonnage de la méthode, les PCRs devraient être optimisées à l'aide des courbes de fusion.

D'autre part, bien que les courbes d'étalonnage soient des polynomiales du troisième degré, la méthylation de l'ADN peut être quand même déterminée. C'est une méthode de cartographie de la méthylation de l'ADN. Elle permet de quantifier la méthylation de l'ADN de chaque locus du gène. Une fois les régions hypométhylées ou hyperméthylées analysées, les régions d'intérêt peuvent être quantifiées par le pyroséquenceur (Dupont et al. 2004; Tost et al. 2006).

4.4. L'analyse des individus

Lors du re-séquençage des loci des deux gènes *CDKN2A* et *RASSF1A* des 7 individus, aucun polymorphisme (SNP, délétion ou insertion) n'a été observés. Seules des modifications épigénétiques ont été observées. Les cytosines et les 5-méthylcytosines des îlots CpGs ont pu être analysées et quantifiées.

La cartographie des loci de ces deux gènes a permis d'établir les régions hypométhylées et les régions hyperméthylées des 7 individus.

La corrélation entre les données biologiques des gènes et des résultats obtenus a permis d'émettre l'hypothèse que les individus HMEC et T47D pourraient avoir une réduction de l'expression des deux gènes *CDKN2A* et *RASSF1A*. Cette hypothèse est à confirmer ou à infirmer par d'autres méthodes comme la PCR quantitative (Smith et al. 2009).

De plus, la capacité haut débit du MALDI-TOF MS permet d'analyser un grand nombre d'échantillons. Cette méthode est donc une méthode efficace pour identifier rapidement les régions importantes du gène pour chaque individu pour se focaliser ensuite sur ces régions par d'autres méthodes comme la PCR quantitative ou le pyroséquençage.

4.5. L'avantage de cette méthode

Cette méthode pourrait également servir à étudier le méthylome qui comprend les îlots CpGs, les îlots CpGs *Shore*, les îlots CpGs *Shelf* et les îlots *Open-sea* ou *Ocean*. Cependant, le nombre d'îlots présents sera plus faible. Par conséquent, le choix des CpGs sera plus restreint pour réaliser une courbe d'étalonnage optimale.

5. Conclusion

Dans ce chapitre, une nouvelle méthode d'analyse de la méthylation de l'ADN utilisant le clivage chimique d'une chimère ARN/ADN simple-brin et l'analyse des fragments par MALDI-TOF MS a été développée.

Une réaction de conversion par le bisulfite de sodium est réalisée afin de différencier les cytosines des 5-méthylcytosines des dinucléotides CpGs. Les cytosines sont converties en uraciles alors que les 5-méthylcytosine ne le sont pas. Les îlots CpGs des gènes *CDKN2A* et *RASSF1A* sont analysés à l'aide de 9 loci pour *CDKN2A* et 3 loci pour *RASSF1A*. La PCR amplifie le locus étudié. La chimère ARN/ADN simple-brin est synthétisée à partir de la PCR purifiée, d'une amorce d'élongation, d'un mélange d'un ribonucléotide (A ou G) et des trois autres désoxynucléotides complémentaires et de l'ADN polymérase *KB17*. La chimère est ensuite clivée par la soude. Les fragments de clivage sont dessalés et analysés par MALDI-TOF MS en mode négatif. Les masses des fragments sont utilisés pour re-séquencer les loci des gènes et pour analyser le taux de méthylation de l'ADN afin de déterminer des régions hypométhylées ou hyperméthylées. Le taux de méthylation de l'ADN est déterminé à l'aide de courbe d'étalonnage et de l'intensité des pics des ions des fragments de clivage des CpGs 0 et 100% méthylés. Les courbes d'étalonnage des loci *CDKN2A-msp1* et *CDKN2A-msp3b* sont des polynomiales du troisième degré avec un coefficient de régression R^2 de 0,99 et un écart type inférieur de 1 à 5%.

La chimère ARN/ADN simple-brin a été utilisée pour re-séquencer les loci des deux gènes à partir d'ADN génomique bisulfité. Cette méthode a identifié les régions hypométhylées ou hyperméthylées des îlots CpGs de 7 individus. Elle a permis de quantifier la méthylation de l'ADN des loci *CDKN2A-msp1* et *CDKN2A-msp3b* de ces individus. En association avec les données biologiques des gènes étudiés, elle a permis de déterminer des régions importantes à étudier pour certains individus.

Elle constitue un autre ajout dans les méthodes d'analyse de la méthylation de l'ADN par spectrométrie de masse mais son principal avantage réside dans la cartographie du gène et de l'analyse du taux de méthylation de l'ADN. Grâce à la capacité haut-débit du MALDI-TOF MS, cette méthode est bien adaptée à l'étude de gènes pour un grand nombre d'individus. Elle pourrait être utilisée pour analyser d'autres gènes que ceux présentés dans ce travail.

Cette méthode a également confirmé la nature semi-quantitative du MALDI-TOF MS avec la nécessité de multiplier les dépôts et les tirs de laser pour normaliser les données.

Cependant, dans cette méthode, les PCRs sont de nouveau à optimiser afin d'éviter le biais de PCR pour la quantification du taux de méthylation de l'ADN. Cette étape est probablement la plus critique de la méthode pour la détermination du taux de méthylation. Une fois toutes les PCRs optimisées, toutes les courbes d'étalonnage pourront être réalisées afin de quantifier plus précisément la méthylation de l'ADN des individus.

Enfin, cette méthode a permis l'étude des îlots CpGs mais elle pourrait être également utilisée pour l'étude du méthylome.

Chapitre 4

**Re-séquençage de l'ADN
par clivage
d'une *ribo-PCR*
et
par MALDI-TOF MS**

Chapitre 4. Re-séquençage de l'ADN par clivage d'une *ribo-PCR* et par MALDI-TOF MS

1. Introduction

Dans le chapitre 2, une méthode de microhaplotypage a été développée. Elle permet le génotypage de seulement quelques SNPs connus sous formes de microhaplotypes. De plus, la chimère ARN/ADN est synthétisée sous forme simple-brin ce qui nécessite une PCR préalable afin d'amplifier le locus d'ADN à étudier. Dans le chapitre 3, une méthode d'analyse de la méthylation de l'ADN a été développée. C'est une méthode de re-séquençage de l'ADN mais elle a pour but l'étude des cytosines et des 5-méthylcytosines des CpG. De plus, la chimère ARN/ADN est synthétisée sous forme simple-brin ce qui nécessite une PCR purifiée préalable afin d'amplifier le locus d'ADN à étudié.

Dans ce chapitre, une nouvelle méthode de re-séquençage de l'ADN par clivage d'une chimère ARN/ADN double-brin (*ribo-PCR*) et l'analyse par spectrométrie de masse MALDI-TOF est développée (annexe 3).

Depuis 25 ans, la PCR est devenue une des techniques les plus puissantes en biologie moléculaire (Saiki et al. 1985). L'ADN polymérase de *Thermus aquaticus*, la *Taq* polymérase, est largement utilisée dans la PCR pour ses propriétés thermorésistantes (Saiki et al. 1988). Elle a également été rapidement utilisée dans les méthodes de séquençage de l'ADN tout d'abord dans la réaction PCR (Innis et al. 1988) puis, dans la réaction de séquençage (Carothers et al. 1989). Dans les années 1990, l'utilisation de la PCR et des mutants de la *Taq* polymérase couplée à l'analyse par électrophorèse capillaire a permis de réaliser le séquençage du génome humain (Lander et al. 2001; Venter et al. 2001).

La spectrométrie de masse et plus particulièrement le MALDI-TOF MS (Karas & Hillenkamp 1988), est bien adaptée pour l'analyse de l'ADN du fait de sa fiabilité, son efficacité et son haut-débit. Le développement de méthodes d'analyse de l'ADN en combinaison avec le MALDI-TOF MS est dû aux étapes simples de biologie moléculaire et la grande capacité haut-débit du MALDI-TOF MS. Il a été utilisé pour l'analyse de l'ADN et notamment pour le génotypage des SNPs ainsi que le séquençage de l'ADN (Hahner et al. 1997; Nordhoff 1996; Sauer et al. 2000). Plus récemment, des nouvelles méthodes pour déterminer les variations plus complexes dans la séquence ont été décrites (Ehrich et al. 2005; Krebs et al. 2003). La

plus importante des méthodes de séquençage utilisant le MALDI-TOF est la méthode *MassARRAY*[®] qui utilise une PCR avec des amorces qui contiennent le site d'initiation de transcription à leurs extrémités 5'. Après la transcription du produit de PCR, des ARN endonucléases de bases spécifiques sont utilisées pour générer des fragments se terminant par la base spécifique clivée. L'analyse des fragments de clivage par MALDI-TOF MS, permet l'identification des différences par rapport à la séquence de référence.

La méthode présentée dans ce chapitre, est basée sur la méthode de génotypage des SNPs (Mauger et al. 2006) et de séquençage de l'ADN mitochondrial (Mauger et al. 2007) par clivage d'une chimère ARN/ADN simple-brin par spectrométrie de masse MALDI-TOF. L'innovation de cette méthode réside d'une part, dans la synthèse d'une chimère ARN/ADN double-brin qui permet d'avoir des informations complémentaires sur les deux brins d'ADN dans la même réaction et d'autre part, et de diminuer le nombre d'étapes de la méthode. Cette méthode est composée de quatre étapes : (i) une *ribo-PCR*, (ii) un clivage, (iii) un dessalage et (iv) l'analyse par MALDI-TOF MS. Les masses moléculaires des fragments de clivage sont comparées à celles de la séquence de référence afin de déterminer les différences. Ces différences dans la séquence de l'ADN sont significatives des polymorphismes de l'ADN tel que les SNPs, les délétions, les insertions ou les mutations. L'intérêt de développer une méthode de séquençage par rapport à une méthode de génotypage est de pouvoir détecter et analyser des variations connues et inconnues de la séquence étudiée.

Cette méthode est mise au point sur trois loci d'ADN génomique : *NOS1* *163731 et *H19* *103280 sur 30 individus ainsi que *SLCO1B1* *604843 sur 95 individus.

L'oxyde nitrique synthétase (NOS) est une protéine enzymatique qui possède trois isoformes : NOS neuronale (nNOS, aussi appelé NOS1), NOS inductible (iNOS ou NOS2), et la NOS endothéliale (eNOS ou NOS3). La forme neuronale est codée par le gène *NOS1* qui est situé sur le chromosome 12 chez l'homme. Cette enzyme synthétise l'oxyde nitrique (NO) dans une réaction qui convertit l'arginine et le dioxygène en citrulline et NO. C'est une molécule à courte durée de vie qui est un neurotransmetteur et communicateur présent dans le système nerveux jouant le rôle de modulateur. Les variantes du gène *NOS1* sont associées à certains cancers ainsi qu'à des maladies chroniques (Kanwar et al. 2009).

Le gène *H19* est situé sur le chromosome 11 chez l'homme et il code pour un long ARN non traduit. Le gène *H19* fait partie d'une région soumise à l'empreinte contenant un ensemble de gènes différentiellement méthylés appartenant à une zone appelée *IRC* (*Imprinted Control*

Region). Cette *IRC* permet l'expression de l'allèle paternel ou maternel. Le gène *H19* semble avoir un rôle dans certaines formes de cancer. De plus, le locus *H19*, soumis à l'empreinte, est impliqué dans la croissance. (Edwards & Ferguson-Smith 2007).

La protéine *SCLO1B1* (*Solute carrier organic anion transporter family member 1B1*) est codée par le gène *SCLO1B1* chez l'humain. Il est situé sur le chromosome 12. La protéine *SCLO1B1* est un récepteur transmembranaire et elle est également impliquée dans l'élimination de médicament tel que les statines. Les patients possédant les variantes de *SLCO1B1* ont un risque considérablement accru de développer des myopathies lorsqu'ils sont traités par la simvastatine qui est un médicament anti-cholestérol (Link et al. 2008).

2. Matériels et Méthodes

2.1. Matériels

Les 30 échantillons d'ADN utilisés pour l'étude des loci *NOS1* et *H19* proviennent d'individus caucasiens. Les 95 échantillons d'ADN utilisés pour l'étude du locus de *SLCO1B1* proviennent d'individus ayant des antécédents d'infarctus du myocarde. Ils ont été choisis en fonction de leurs génotypes. Les 30 échantillons d'ADNs ont été génotypés par la méthode d'*Amplifluor*[®] Assay et les 95 ADNs par *Illumina*[®] *Infinium*TM II Assay (Link et al. 2008).

Les amorces ont été synthétisées par Eurogentec (Angers, France). Les séquences des amorces sont listées dans le tableau 4.1. Elles contiennent une modification 2'-OMe du ribose de l'avant-dernière base. Cette modification permet d'inhiber la digestion 3'-5' de l'exonucléase de l'amorce de l'activité *haute-fidélité* de l'ADN polymérase (Schonbrunner et al. 2006; Mauger et al. 2007).

| Locus | Amorce sens | Amorce anti-sens |
|----------------|-------------------------|-----------------------------|
| <i>NOS1</i> | TTGATGGCTCACCATTGA(A)A | GGCTAGTGGTGAGGATGG(A)A |
| <i>H19</i> | GCCAGTGAGGAGTGTGGAG(T)A | TTCTTTCCAGCCCTAGCT(C)A |
| <i>SLCO1B1</i> | CAGGATGATTCTGGCTTCA(A)A | GCCAAATTCTACTGGAAATACAA(A)A |

Tableau 4.1 : Amorces de la ribo-PCR des loci *NOS1*, *H19* et *SLCO1B1*. Les bases qui sont entre parenthèses contiennent la modification 2'OMe sur le ribose.

Les dNTPs (N=A, C, G et T) et les NTPs (N=A, C, G et U) proviennent de GE Healthcare (Saclay, France) et les produits chimiques d'Aldrich (Steinheim, Allemagne). La résine échangeuse de cations provient de Biorad (Marnes la coquette, France) et le kit *Clean Resin Tool* de *MassARRAY*[®] de chez Sequenom (Hamburg, Allemagne). Les étalons externes pour l'analyse en MALDI-TOF MS d'oligonucléotides de 4 à 11-mer et de 12 à 30-mer proviennent de chez BRUKER (Bremen, Allemagne).

L'ADN polymérase *KB17* a été développée par Roche Molecular Systems (Pleasanton, CA, USA). Elle n'est pas commercialisée, elle peut être obtenue par Dr Thomas Myers (thomas.myers@roche.com). Cette chimère d'ADN polymérase contient un domaine 5'-nucléase de l'ADN polymérase *Thermus sp. ZO5* et les domaines 3'-nucléase et 5'-3'-nucléase de l'ADN polymérase *Thermotoga maritima (Tma)*. L'activité 5'-nucléase a été supprimée par l'introduction de la mutation *G46E* (Mauger et al. 2006). L'activité *haute-fidélité* de l'ADN polymérase non-mutée *Tma* a été modulée par l'introduction d'une mutation *L329A* (Schonbrunner et al. 2006; Mauger et al. 2007) du 'Motif I' du domaine 3'-5' exonucléase. L'enzyme contient une mutation *E678G* qui élimine la spécificité de l'enzyme non-mutée contre l'incorporation des nucléotides substitués en position 2'-C, favorisant ainsi l'incorporation des ribonucléosides triphosphates. Enfin, l'enzyme contient trois mutations supplémentaires dans le domaine de la polymérase qui améliore l'accroche sur la matrice entraînant ainsi l'augmentation de la rapidité d'élongation, particulièrement dans l'incorporation de ribonucléotides.

Les procédures de thermocyclages ont été réalisées sur des thermocycleurs gradients Eppendorf (Eppendorf, Allemagne). Les analyses ont été réalisées par un spectromètre de masse MALDI-TOF *Autoflex*TM de chez Bruker (Bremen, Allemagne). La cible MALDI est une *Anchorchip*TM Bruker (Bremen, Allemagne).

2.2. Méthodes

Les trois loci sont amplifiés par ribo-PCR : *NOS1*, *H19* et *SLCO1B1*.

2.2.1. La ribo-PCR par l'ADN polymérase *KB17*

La ribo-PCR du locus *NOS1*

Le mélange de 20 μL de la solution de la *ribo-PCR* du locus *NOS1* contient 4 μL de tricine à 500 mM (pH 8), 4 μL d'acétate de potassium 500 mM (pH 7,5), 2 μL d'acétate de magnésium à 30 mM, 1 μL d'un mélange de ATP, dCTP, dGTP et dTTP à 4 mM chacun, 0,8 μL de chaque amorce (tableau 4.1) à 10 pmol/ μL , 0,5 μL d'enzyme *KB17* à 10 U/ μL , 2 μL d'ADN à 10 ng/ μL et 4,9 μL d'eau milli-Q.

La ribo-PCR du locus *H19*

Le mélange de 20 μL de la solution de la *ribo-PCR* du locus *H19* contient 4 μL de tricine à 500 mM (pH 8), 4 μL d'acétate de potassium 500 mM (pH 7,5), 2 μL d'acétate de magnésium à 30 mM, 1 μL d'un mélange de ATP, dCTP, dGTP et dTTP ou UTP, dATP, dCTP et dGTP de à 4 mM chacun, 0,8 μL de chaque amorce (tableau 4.1) à 10 pmol/ μL , 0,5 μL d'enzyme *KB17* à 10 U/ μL , 2 μL d'ADN à 10 ng/ μL et 4,9 μL d'eau milli-Q.

Le duplex des *ribo-PCRs* des loci *NOS1* et *H19*

Le mélange de 20 μL de la solution des *ribo-PCRs* du duplex des loci *NOS1* et *H19* contient 4 μL de tricine à 500 mM (pH 8), 4 μL d'acétate de potassium 500 mM (pH 7,5), 2 μL d'acétate de magnésium à 30 mM, 1 μL d'un mélange de ATP, dCTP, dGTP et dTTP à 4 mM chacun, 0,8 μL de chaque amorce (tableau 4.1) à 10 pmol/ μL , 1 μL d'enzyme *KB17* à 10 U/ μL , 2 μL d'ADN à 10 ng/ μL et 2,8 μL d'eau milli-Q.

La ribo-PCR du locus *SLCO1B1*

Le mélange de 20 μL de la solution de la *ribo-PCR* du locus *SLCO1B1* contient 4 μL de tricine à 500 mM (pH 8), 4 μL d'acétate de potassium 500 mM (pH 7,5), 2 μL d'acétate de magnésium à 30 mM, 1 μL d'un mélange de CTP, dATP, dGTP et dTTP ou GTP, dATP, dCTP et dTTP de à 4 mM chacun, 0,8 μL de chaque amorce (tableau 4.1) à 10 pmol/ μL , 0,5 μL d'enzyme *KB17* à 10 U/ μL , 2 μL d'ADN à 10 ng/ μL et 4,9 μL d'eau milli-Q.

Les réactions de *ribo-PCR* sont dénaturées pendant 4 min à 92°C, puis suivies de 46 cycles de 15 s à 92°C et 5 min à 62°C.

Un mélange, de 5 µL de la réaction de la *ribo-PCR* et de 5 µL du tampon de charge, est ensuite déposé sur un gel d'agarose à 2 % afin de vérifier la taille du produit PCR.

2.2.2. Le clivage

Au 15 µL de la solution de la *ribo-PCR* est ajouté un mélange de 2 µL d'hydroxyde de sodium à 3 M et 3 µL d'eau milli-Q. La réaction de clivage est mise sous incubation pendant 1h30 à 70°C.

2.2.3. Le dessalage

A la réaction de clivage, 6 mg de résines échangeuses de cations sont ajoutées à l'aide du kit *Clean Resin Tool* de *MassARRAY*[®]. Les échantillons sont incubés pendant 20 min à température ambiante sous agitation et ensuite centrifugés sous 134 xg afin de récupérer le surnageant.

2.2.4. L'analyse par spectrométrie de masse MALDI-TOF

2.2.4.1. La préparation de la cible MALDI

La solution de matrice trihydroxyacetophenone (THAP) est préparée avec 0,2 M de 2,4,6 THAP dans 50 % d'acétonitrile, 0,2 M de 2,3,4 THAP dans 50% d'acétonitrile et 0,3 M de citrate d'ammonium dans les proportions 6/3/2 (v/v) (Zhu et al. 1996).

Sur une cible *Anchorchip*TM Bruker de 384 spots de porosité 400 µm, sont déposés 0,5 µL de la solution de matrice THAP et 0,5 µL du surnageant par spot. La cible est séchée à température ambiante.

2.2.4.2. La spectrométrie de masse MALDI-TOF

La cible MALDI est introduite dans un spectromètre de masse MALDI-TOF MS. L'analyse est réalisée en mode linéaire, négatif à l'aide du logiciel *FlexControl* et des paramètres du MALDI-TOF MS suivants : IS1=20 kV ; IS2=18,5 kV ; lens=7,5 kV et un délai d'extraction de 150 ns. Les spectres de masse sont obtenus par l'accumulation de 500 tirs de laser.

Un étalonnage externe est réalisé à l'aide des kits d'étalons externes d'oligonucléotides de 4 à 11-mer et de 12 à 30-mer.

Ces conditions permettent d'analyser des masses moyennes et de détecter des échantillons d'environ 0,15 pmol.

Les spectres de masse sont analysés à l'aide du logiciel *FlexAnalysis*. L'annotation des pics caractérisés par le rapport m/z des ions, la soustraction de la ligne de base ainsi qu'un lissage du signal sont réalisés.

Un logiciel de calcul de masse des fragments de clivage a été développé par Jérémy Semhoun (Mauger et al. 2007). Il a été modifié afin de pouvoir calculer les masses des fragments de clivage sens et anti-sens des *ribo-PCRs*.

Les rapports m/z des ions $[M-H]^-$ sont ensuite comparés aux masses des fragments de clivage de la séquence de référence. Les différences de masse observées sont caractéristiques d'une modification de la séquence de référence et donc de la présence de polymorphismes.

3. Résultats

Dans ce chapitre, l'ADN polymérase thermostable *KB17* est utilisée. Elle incorpore efficacement les ribonucléotides et les deoxyribonucléotides dans une réaction de *ribo-PCR* sur de l'ADN génomique. La chimère ARN/ADN double-brin est générée à partir de l'ADN génomique, deux amorces sens et anti-sens, l'ADN polymérase et un mélange d'un ribonucléotide et des trois autres deoxyribonucléotides complémentaires dans une seule réaction de *ribo-PCR*. Ensuite, elle est fragmentée à l'aide d'une simple réaction de clivage par la soude et dessalée par une résine échangeuse de cations pour son analyse par MALDI-TOF MS. Les rapports m/z des ions $[M-H]^-$ des spectres de masse sont comparés aux masses des fragments de clivage de la séquence de référence afin de re-séquencer et d'identifier sans ambiguïté des polymorphismes. L'information complémentaire de la *ribo-PCR* permet

d'analyser la séquence ainsi que les polymorphismes sur les deux brins dans une seule et même réaction. Cette méthode a été mise au point par l'étude de trois loci génomiques : *NOS1* rs3741476 (SNP R) et *H19* rs2839698 (SNP Y) et *SLCO1B1* rs4363657 (SNP Y) (Mauger et al. 2012).

3.1. La *ribo-PCR* par l'ADN polymérase *KB17*

Comme indiqué précédemment (Mauger et al. 2007), l'ADN polymérase *KB17* incorpore facilement 100 % d'ATP, 100 % d'CTP ou 100 % d'GTP dans une réaction d'élongation linéaire d'environ 450 pb sur de l'ADN mitochondrial. Elle incorpore également 100 % d'ATP, ou 100 % d'GTP dans une réaction d'élongation linéaire d'environ 500 pb sur de l'ADN génomique (chapitre 3).

Dans ce cas, l'ADN polymérase *KB17* a incorporé 100% d'ATP, 100 % d'CTP, 100 % d'GTP ou 100 % d'UTP dans une réaction de *ribo-PCR* d'environ 100 pb sur de l'ADN génomique. Une chimère ARN/ADN double-brin a donc été synthétisée. La *ribo-PCR* de *NOS1* a été générée à partir de 100 % ATP. La *ribo-PCR* de *H19* a été générée à partir de 100 % ATP ou de 100 % UTP. La *ribo-PCR* de *SLCO1B1* a été générée à partir de 100 % CTP ou de 100% GTP. L'enzyme incorpore facilement une succession de ribonucléotides de 3 à 4 bases, par contre au-delà il peut y avoir une baisse de l'efficacité de la *ribo-PCR*.

3.2. Les fragments de clivage

Les fragments de clivage pour l'analyse par spectrométrie de masse sont obtenus à l'aide d'un clivage par la soude après chaque ribonucléotide (Mauger et al. 2006; Mauger et al. 2007). Les fragments de clivage sont composés de désoxynucléotides et se terminent par un ribonucléotide qui est composé d'un groupement 3'-phosphate.

Les masses moyennes des fragments de clivage des *ribo-PCRs* sont listées dans les tableaux 4.2 : A, B, C, D et E.

4.2A

| <i>NOS1</i> rs3741476 | Séquences | Masse | Début | Fin | Sens |
|-----------------------|----------------------------------|--------|-------|-----|-----------|
| | gcc | 845,6 | 103 | 105 | sens |
| | cca | 925,6 | 97 | 99 | sens |
| | cta | 940,6 | 49 | 51 | sens |
| | cta | 940,6 | 100 | 102 | sens |
| | tca | 940,6 | 4 | 2 | anti-sens |
| | gca | 965,6 | 65 | 67 | sens |
| | gca | 965,6 | 32 | 30 | anti-sens |
| | gta | 980,6 | 68 | 70 | sens |
| | tga | 980,6 | 44 | 42 | anti-sens |
| | gga | 1005,6 | 40 | 38 | anti-sens |
| | gcca | 1254,8 | 71 | 74 | sens |
| | gcca | 1254,8 | 8 | 5 | anti-sens |
| | gtca | 1269,8 | 49 | 46 | anti-sens |
| | ttga | 1284,8 | 45 | 48 | sens |
| | cgga | 1294,8 | 56 | 59 | sens |
| | gtga | 1309,8 | 84 | 81 | anti-sens |
| | ggta | 1309,8 | 36 | 33 | anti-sens |
| | cttta | 1549,0 | 60 | 64 | sens |
| | tcctca | 1823,2 | 91 | 96 | sens |
| | ctgcta | 1863,2 | 68 | 63 | anti-sens |
| | tggtca | 1903,2 | 74 | 69 | anti-sens |
| | tggtga | 1943,2 | 14 | 9 | anti-sens |
| | ctttcca | 2127,4 | 84 | 90 | sens |
| | ttctttca | 2750,8 | 75 | 83 | sens |
| | ccttcctca | 3009,9 | 35 | 44 | sens |
| | gtccgtgcta | 3419,2 | 60 | 50 | anti-sens |
| | gggggctgcta | 3509,2 | 24 | 34 | sens |
| C | gcccctt <u>c</u> tttca | 4221,7 | 29 | 16 | anti-sens |
| T | gcccctt <u>t</u> tttca | 4236,7 | 29 | 16 | anti-sens |
| A | ttgatggctcaccattga(a) <u>a</u> | 6555,3 | 1 | 21 | sens |
| | ggctagtggtaggatgg(a)a a | 6711,3 | 105 | 85 | anti-sens |
| G | ttgatggctcaccattga(a) <u>g</u> a | 6884,5 | 1 | 22 | sens |

4.2B

| <i>H19</i> | rs2839698 | Séquence | Masse | Début | Fin | Sens |
|------------|-----------|---|--------|-------|-----|-----------|
| | | tca | 940,6 | 53 | 55 | sens |
| | | cga | 965,6 | 64 | 66 | sens |
| | | cga | 965,6 | 36 | 34 | anti-sens |
| A | | ccta | 1229,8 | 23 | 20 | anti-sens |
| | | gcta | 1269,8 | 76 | 79 | sens |
| | | ggca | 1294,8 | 30 | 33 | sens |
| | | ctggc | 1479,0 | 5 | 1 | anti-sens |
| | | ctcca | 1519,0 | 19 | 15 | anti-sens |
| | | gctga | 1599,0 | 57 | 53 | anti-sens |
| | | gggcga | 1953,2 | 45 | 50 | sens |
| | | ctcctca | 2112,4 | 12 | 6 | anti-sens |
| | | tcgtgca | 2192,4 | 34 | 40 | sens |
| | | gctgggga | 2586,7 | 56 | 63 | sens |
| | | gggctgga | 2586,7 | 80 | 87 | sens |
| | | tcgtccca | 2730,8 | 66 | 58 | anti-sens |
| | | tggcctga | 2850,8 | 67 | 75 | sens |
| A | | tcctgggca | 3140,0 | 33 | 24 | anti-sens |
| G | | tcctgggc <u>g</u> ccta | 4351,8 | 33 | 20 | anti-sens |
| | | tgctgccctgtctgca | 4920,2 | 52 | 37 | anti-sens |
| | | <i>ttctttccagccctagct(c)a</i> ggccca | 7914,1 | 92 | 67 | anti-sens |
| C | | <i>gccagtgaggagtgtggag(t)a</i> <u>ggc</u> gccca | 9174,9 | 1 | 29 | sens |
| T | | <i>gccagtgaggagtgtggag(t)a</i> <u>ggt</u> gccca | 9189,9 | 1 | 29 | sens |

4.2C

| <i>H19</i> | rs2839698 | Séquence | Masse | Début | Fin | Sens |
|------------|-----------|---|---------|-------|-----|-----------|
| | | ggc | 885,6 | 3 | 1 | anti-sens |
| | | ccu | 902,6 | 10 | 8 | anti-sens |
| | | acu | 926,6 | 20 | 18 | anti-sens |
| | | cgu | 942,6 | 35 | 37 | sens |
| | | cgu | 942,6 | 65 | 63 | anti-sens |
| | | gau | 966,6 | 54 | 52 | anti-sens |
| | | cacu | 1215,8 | 7 | 4 | anti-sens |
| | | gccu | 1231,8 | 32 | 29 | anti-sens |
| | | cagcu | 1545,0 | 54 | 58 | sens |
| | | gagcu | 1585,0 | 74 | 78 | sens |
| | | cgccu | 1810,2 | 49 | 44 | anti-sens |
| | | ggccu | 1890,2 | 68 | 73 | sens |
| | | aggcu | 1914,2 | 79 | 84 | sens |
| | | ccacacu | 2107,4 | 17 | 11 | anti-sens |
| | | gcacgau | 2187,4 | 39 | 33 | anti-sens |
| | | ccccagcu | 2412,6 | 62 | 55 | anti-sens |
| A | | gggc <u>a</u> ccu | 2492,6 | 28 | 21 | anti-sens |
| G | | gggc <u>g</u> ccu | 2508,6 | 28 | 21 | anti-sens |
| | | ggaaagaa | 2587,7 | 85 | 92 | sens |
| | | ggggacgau | 2885,9 | 59 | 67 | sens |
| T | | gcccaggcau | 3095,0 | 25 | 34 | sens |
| | | gcagacaggcgacau | 5022,3 | 38 | 53 | sens |
| T | | <i>gccagtgaggagtgtggag(t)a</i> <u>ggu</u> | 7666,0 | 1 | 24 | sens |
| | | <i>ttctttccagccctagct(c)a</i> ggcccau | 8204,3 | 92 | 66 | anti-sens |
| C | | <i>gccagtgaggagtgtggag(t)a</i> <u>ggc</u> gccaggcau | 10725,9 | 1 | 34 | sens |

4.2D

| <i>SCLO1B1</i> rs4363657 | Séquence | Masse | Début | Fin | Sens |
|--------------------------|---|---------|-------|-----|-----------|
| | ttc | 931,6 | 44 | 46 | sens |
| | atc | 940,6 | 6 | 4 | anti-sens |
| | ttc | 1235,8 | 47 | 50 | sens |
| | ttc | 1235,8 | 37 | 34 | anti-sens |
| | ttaac | 1558,0 | 33 | 29 | anti-sens |
| | tgtac | 1574,0 | 76 | 72 | anti-sens |
| | agaac | 1896,2 | 12 | 7 | anti-sens |
| | aattttc | 2470,6 | 53 | 60 | sens |
| | agtggtac | 2874,9 | 76 | 84 | sens |
| | taaattattc | 3097,0 | 71 | 62 | anti-sens |
| | tttttgatttc | 4007,6 | 91 | 103 | sens |
| | agtagaatttggc | 4109,7 | 105 | 117 | sens |
| | tgattttttgaagc | 4684,0 | 28 | 14 | anti-sens |
| | agaataatttagtac | 4695,1 | 61 | 75 | sens |
| | agttaagaagagtc | 4745,1 | 28 | 42 | sens |
| | tgaaaaattgagaagaaggac | 6969,6 | 61 | 40 | anti-sens |
| | caggatgattctggcttca(a)a aaaatc | 8417,5 | 1 | 27 | sens |
| A | gccaaattctactggaatacaa(a)a aagagt <u>g</u> gtac | 11527,5 | 117 | 81 | anti-sens |
| G | gccaaattctactggaatacaa(a)a aagagt <u>g</u> gtac | 11543,5 | 117 | 81 | anti-sens |

4.2E

| <i>SCLO1B1</i> rs4363657 | Séquence | Masse | Début | Fin | Sens |
|--------------------------|----------------------------------|--------|-------|-----|-----------|
| | tag | 980,6 | 107 | 109 | sens |
| | aag | 989,6 | 45 | 43 | anri-sens |
| | aag | 989,6 | 17 | 15 | anri-sens |
| | ccag | 1254,8 | 14 | 11 | anri-sens |
| | aaag | 1302,9 | 35 | 38 | sens |
| | aaag | 1302,9 | 49 | 46 | anri-sens |
| | tacag | 1583,0 | 73 | 77 | sens |
| | ttaag | 1598,0 | 30 | 34 | sens |
| | aatttg | 1902,2 | 110 | 115 | sens |
| | aaaaattg | 2537,7 | 59 | 52 | anri-sens |
| | taccactg | 2754,8 | 83 | 75 | anri-sens |
| | tattccag | 2784,8 | 98 | 106 | sens |
| | atTTTTTg | 2805,8 | 26 | 18 | anri-sens |
| | aatcatcctg | 2987,0 | 10 | 1 | anri-sens |
| | aataatttag | 3146,1 | 63 | 72 | sens |
| | actctttctaactg | 4589,0 | 41 | 27 | anri-sens |
| | tactaaattattctg | 4637,0 | 74 | 60 | anri-sens |
| C | tacc <u>c</u> actctTTTTg | 4869,1 | 82 | 97 | sens |
| T | tact <u>c</u> actctTTTTg | 4884,2 | 82 | 97 | sens |
| | tccttctttc <a>aa ttttcag | 6688,3 | 41 | 62 | sens |
| | gccaaattctactggaatacaa(a)a aag | 8702,7 | 117 | 90 | anri-sens |
| | caggatgattctggcttca(a)a aaaatcag | 9059,9 | 1 | 29 | sens |

Tableau 4.2 : Masses moyennes des fragments des ribo-PCRs des loci de : (A) *NOS1* rs3741476 avec ATP ; *H19* rs2839698 avec (B). ATP et (C) UTP ; *SLCO1B1* rs4363657 avec (D) CTP et (E) GTP. Les

amorces des ribo-PCRs sont en italique. Les bases qui sont entre parenthèses contiennent la modification 2'OMe sur le ribose. Le SNP est mis en gras et surligné.

3.3. L'analyse des spectres de masse MALDI-TOF

En raison de la charge négative portée par l'atome d'oxygène du groupement phosphate, les molécules cationisées peuvent être observées. Les échantillons sont donc dessalés par l'ajout d'une résine échangeuse de cations avant d'être analysés par MALDI-TOF MS.

Comme les échantillons possèdent une charge négative et grâce à la distribution des masses des fragments de clivage (tableau 4.2), ils sont analysés en mode linéaire négatif. La combinaison des masses des fragments de clivage des deux brins complémentaires sens et anti-sens permet l'attribution de la séquence. Cette méthode comporte seulement une simple étape enzymatique, une réaction chimique, un dessalage et une analyse par MALDI-TOF MS. Le clivage chimique de la *ribo-PCR* couplé avec le MALDI-TOF MS constitue probablement le plus simple des protocoles de re-séquençage. Cette méthode a été appliquée à différents loci génomiques : *NOS1* rs3741476 (SNP R) et *H19* rs2839698 (SNP Y) sur 30 individus et *SLCO1B1* rs4363657 (SNP Y) sur 95 individus.

3.3.1. L'analyse du locus *NOS1*

Le re-séquençage du locus *NOS1* rs3741476 du SNP R est montré dans la figure 4.1 avec l'interprétation des spectres de masses.

La séquence est complètement déterminée à l'aide de la complémentarité des informations des brins sens et anti-sens. Dans la région de 1 à 2-mer, de nombreux fragments ont une masse identique et ne peuvent donc pas être distingués par MALDI-TOF MS. Dans la région supérieure à 20-mer, il y a peu de fragments mais ils restent difficiles à détecter à cause du faible pouvoir résolutif du MALDI-TOF MS. La plus petite différence de masse entre deux bases est de 9 Da entre l'adénine et la thymine. Dans les hautes masses, le pouvoir résolutif de l'instrument n'est pas suffisant pour pouvoir distinguer cette variation de masse.

Les fragments de clivage GCA, GCCA et CGGA ne sont pas présents dans le figure 4.1 ainsi que GGGCGA dans de la *ribo-PCR* ATP de *H19* (tableau 4.2B) et AAG, le CCAG, AAAG de la ribo-PCR GTP de *SLCO1B1* (tableau 4.2E). Ces fragments ont la particularité de ne pas contenir de base thymine. Les fragments de clivage contenant des bases thymine sont plus facilement détectables dans ces conditions d'analyses par MALDI-TOF MS que ceux qui n'en contiennent pas (chapitre 1). Les conditions d'analyses ont été modifiées afin de détecter les fragments de clivage sans thymine. Tout d'abord, le mode positif a été utilisé mais le spectre de masse montre la présence d'un polymère qui couvre une bonne partie du spectre de masse. L'analyse de la séquence devient alors impossible. Ce polymère provient du tampon de l'enzyme, il ne peut donc pas être supprimé du protocole. Une autre matrice a été utilisée l'acide 3-hydroxypicolinique (HPA). La matrice HPA est la plus utilisée pour l'analyse des oligonucléotides (Tang et al. 1993). Les fragments de clivage sans thymine sont également absents du spectre de masse. L'absence de détection des fragments de clivage sans thymine est donc liée à la stabilité des bases en phase gazeuse (chapitre 1). Les fragments sans thymine doivent être plus concentrés pour être détectés.

Les spectres de masse de la *ribo-PCR* ATP qui identifient le SNP R sont présentés dans la figure 4.2.

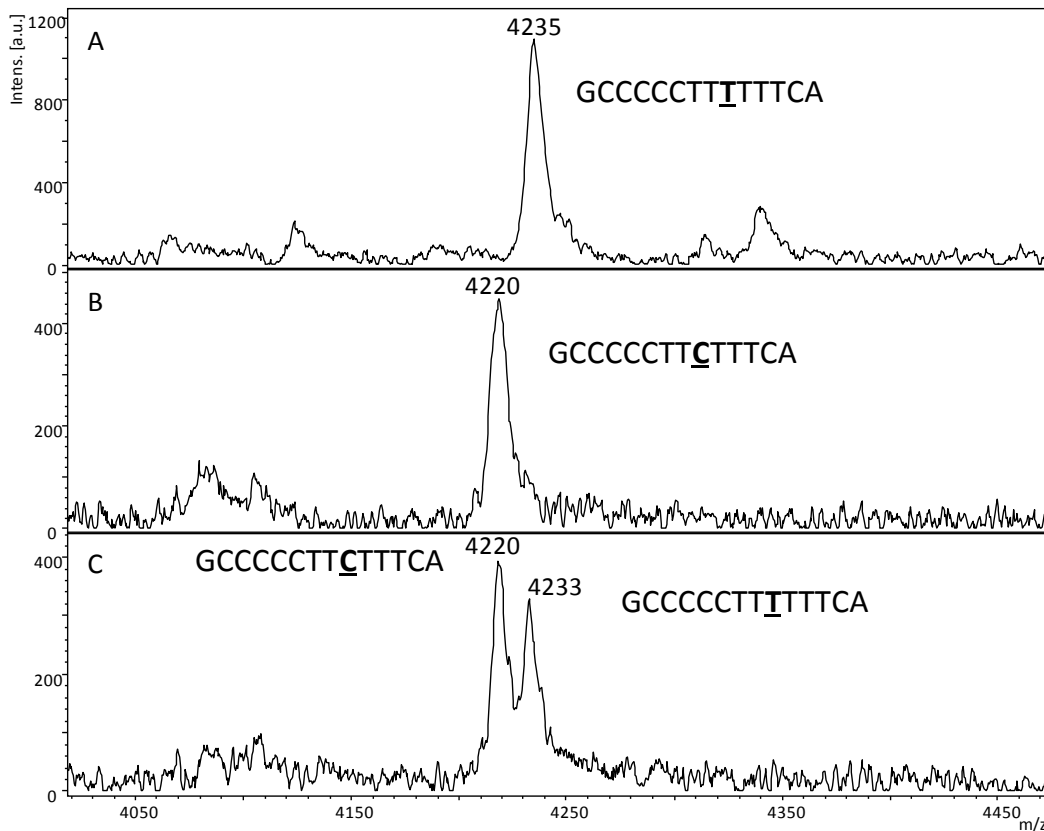


Figure 4.2 : Spectres de masse MALDI-TOF de la région des 14-mer de la ribo-PCR ATP des fragments anti-sens $[M-H]^-$ du locus *NOS1* rs3741476 d' : (A) un homozygote A; (B) un homozygote G et (C) un hétérozygote R. L'ion $[M-H]^-$ GCCCCCTTCTTTCA représente l'allèle G et l'ion $[M-H]^-$ GCCCCCTTTTTTCA représente l'allèle A.

Dans la région des 14-mer, ils montrent à m/z 4220 l'ion $[M-H]^-$ du fragment GCCCCCTTCTTTCA ($M=4221,7$ Da) et à m/z 4235 l'ion $[M-H]^-$ du fragment GCCCCCTTTTTTCA ($M=4236,7$ Da). Ils sont caractéristiques respectivement de la présence des allèles G et A du SNP rs3741476. Un homozygote A (figure 4.2A), un homozygote G (figure 4.2B) et un hétérozygote R (figure 4.2C) du SNP rs3741476 sont identifiés par le clivage de la *ribo-PCR* ATP du locus *NOS1*.

3.3.2. L'analyse du locus *H19*

Pour le locus *H19* rs2839698, deux *ribo-PCRs* sont possibles avec ATP ou UTP.

Pour la *ribo-PCR* UTP du locus *H19* (tableau 4.2C), tous les fragments de clivage sont détectés mais il y a également des ions supplémentaires qui proviennent de la dépurination d'une base adénine ou guanine (Gut 1997). Lorsque le nombre de bases des fragments de

clivage augmente, l'abondance de l'ion dépuriné devient plus grande que celle de l'ion attendu. Dans la *ribo-PCR* UTP, tous les fragments de clivage contiennent une base uracile 3' terminale. Dans une *ribo-PCR* UTP, tous les fragments de clivage sont donc détectés contrairement aux autres *ribo-PCR* qui possèdent des fragments de clivage sans thymine ni uracile (chapitre 1). Il n'y a donc pas de perte d'information avec la *ribo-PCR* UTP. Des ions supplémentaires de dépurination sont donc présents et perturbent l'interprétation du spectre de masse. Le spectre de masse de la *ribo-PCR* UTP du locus *H19* est donc plus difficile à analyser que celui de la *ribo-PCR* ATP. Toutefois, dans les deux cas le SNP a pu être analysé.

Les spectres de masse de la *ribo-PCR* ATP qui identifient le SNP sont présentés dans la figure 4.3.

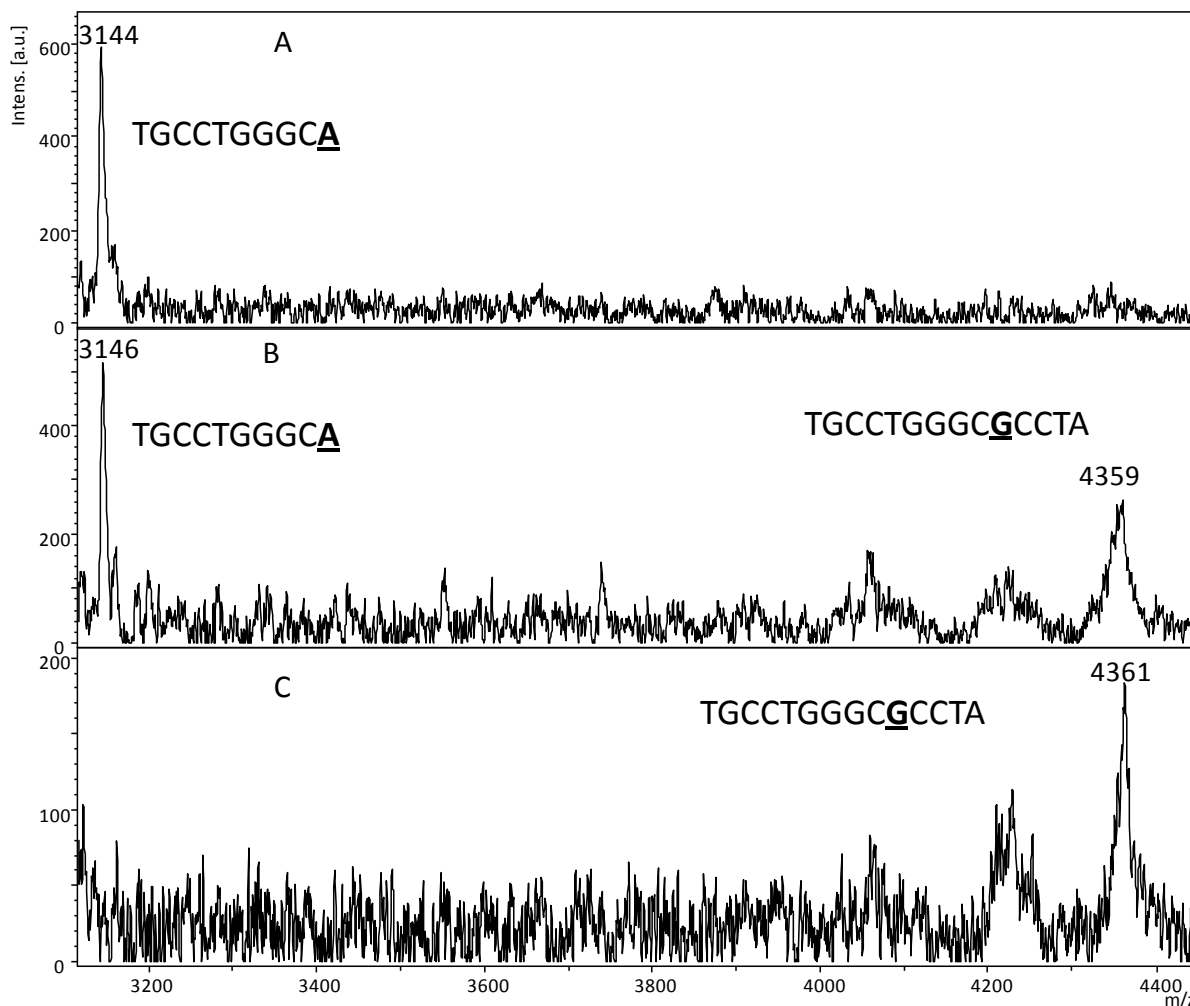


Figure 4.3 : Spectres de masse MALDI-TOF de la région des 10 à 14-mer de la ribo-PCR ATP des fragments anti-sens [M-H] du locus *H19* rs2839698 d' : (A) un homozygote T; (B) un hétérozygote Y et (C) un homozygote C. L'ion [M-H] TGCCTGGGCA représente l'allèle T et l'ion [M-H] TGCCTGGGCGCCTA représente l'allèle C.

Dans la région des 10 à 14-mer, ils montrent à m/z 3144 l'ion $[M-H]^-$ du fragment TGCCTGGGCA (M=3140,0 Da) et à m/z 4359 $[M-H]^-$ du fragment TGCCTGGGCGCCTA (M=4351,8 Da). Ils sont caractéristiques respectivement de la présence des allèles T et C du SNP rs2839698. Un homozygote T (figure 4.3A), un homozygote C (figure 4.3C) et un hétérozygote Y (figure 4.3B) du SNP rs2839698 sont identifiés par le clivage de la ribo-PCR ATP du locus *H19*.

Les spectres de masse de la ribo-PCR UTP qui identifie le SNP sont présentés dans la figure 4.4.

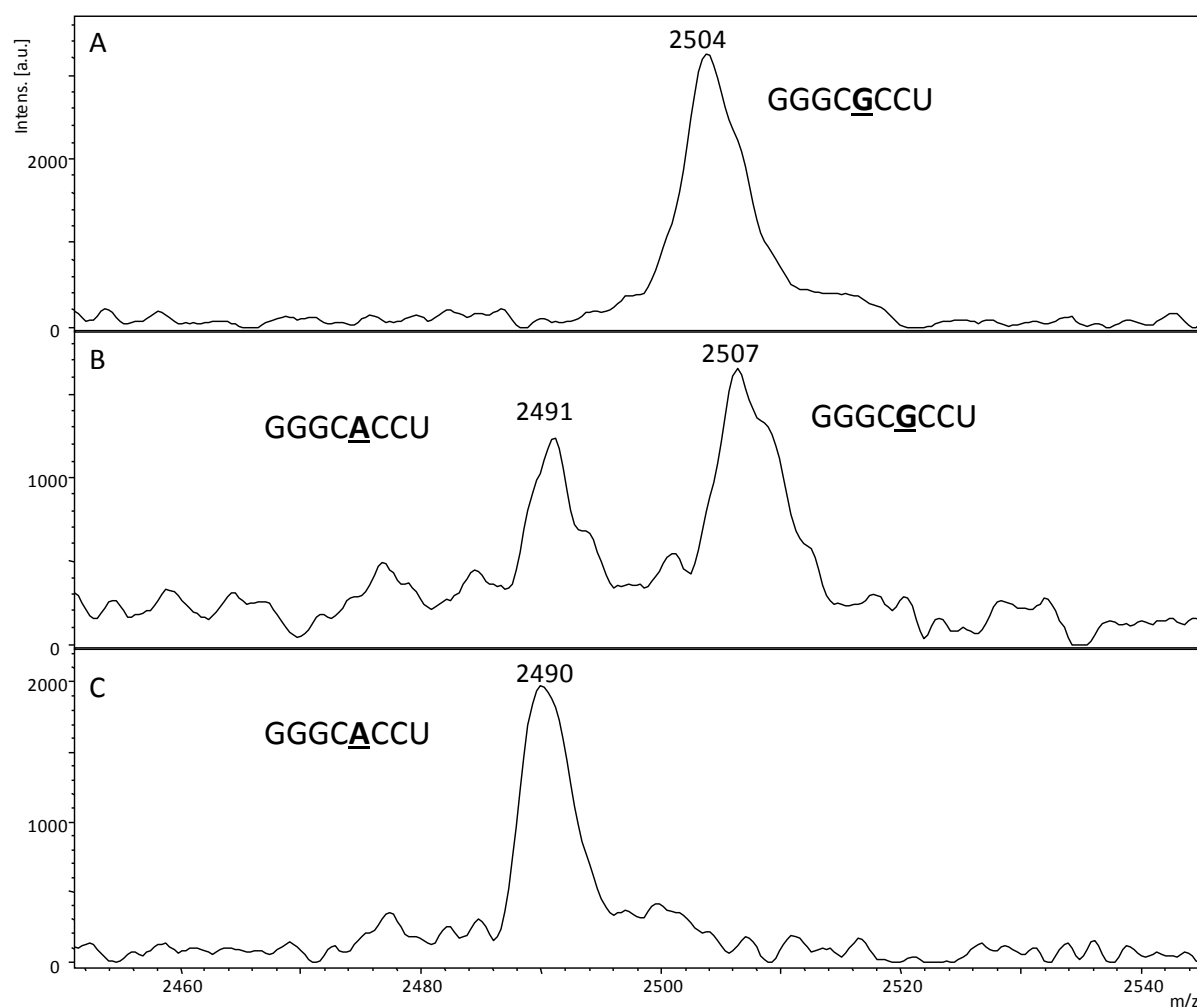


Figure 4.4 : Spectres de masse MALDI-TOF de la région des 8-mer de la ribo-PCR UTP des fragments anti-sens $[M-H]^-$ du locus *H19* rs2839698 d' : (A) un homozygote C; (B) un hétérozygote Y et (C) un homozygote T. L'ion $[M-H]^-$ GGGCACCU représente l'allèle T et l'ion $[M-H]^-$ GGGCGCCU représente l'allèle C.

Dans la région des 8-mer, ils montrent à m/z 2491 l'ion $[M-H]^-$ du fragment GGGCACCU (M=2492,6 Da) et à m/z 2507 l'ion $[M-H]^-$ du fragment GGGCGCCU (M=2508,6 Da). Ils

sont caractéristiques respectivement de la présence des allèles T et C du SNP rs2839698. Un homozygote C (figure 4.4a), un homozygote T (figure 4.4c) et un hétérozygote Y (figure 4.4b) du SNP rs2839698 sont identifiés par le clivage de la ribo-PCR UTP du locus *H19*.

3.3.3. L'analyse du duplex des loci *NOS1* et *H19*

Comme les loci *NOS1* et *H19* sont analysés à l'aide d'une *ribo-PCR* ATP, un duplex des deux *ribo-PCRs* a été réalisé dans une même réaction permettant ainsi le multiplexage de la méthode.

Les spectres de masse du duplex des *ribo-PCRs* ATP qui identifient chaque SNP des locis *NOS1* et *H19* sont présentés dans la figure 4.5.

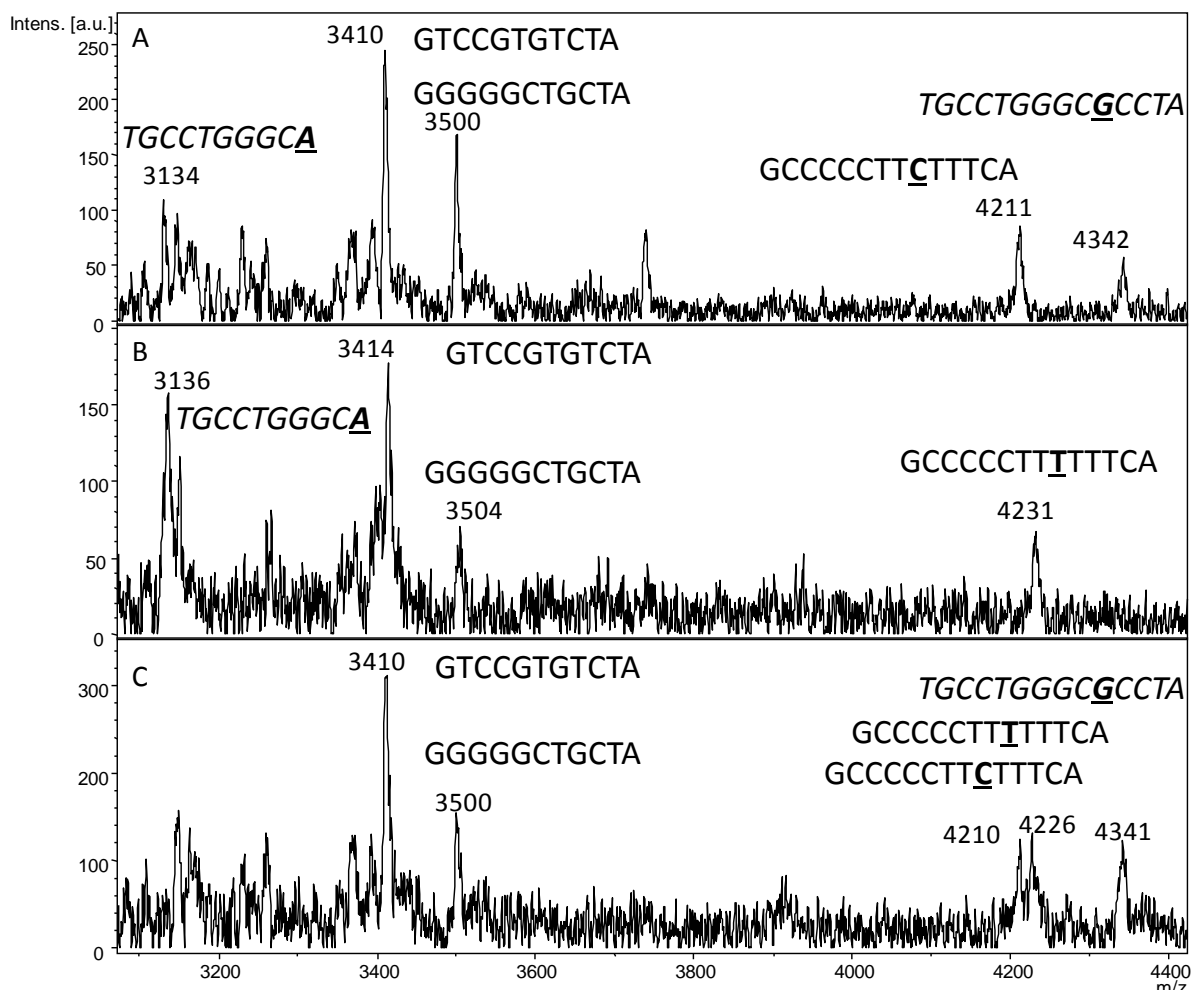


Figure 4.5 : Spectres de masse MALDI-TOF de la région des 10 à 14-mer du duplex de la ribo-PCR ATP des fragments anti-sens [M-H] des loci *NOS1* rs3741476 et *H19* rs2839698 d': (A) un homozygote G et un hétérozygote Y; (B) un homozygote A et un homozygote T et (C) un hétérozygote R et un homozygote C. L'ion [M-H] TGCCTGGGCA représente l'allèle T et l'ion [M-H] TGCCTGGGCCCTA représente l'allèle C du locus *H19*. Ils sont en italique. L'ion [M-H]

GCCCCCTTCTTTCA représente l'allèle G et l'ion $[M-H]^-$ *GCCCCCTTTTTTCA* représente l'allèle A du locus *NOS1*. L'ion $[M-H]^-$ *GTCCGTGTCTA* et l'ion $[M-H]^-$ *GGGGGCTGCTA* sont des fragments de re-séquençage du locus *NOS1*.

Dans la région des 10 à 14-mer, ils montrent à m/z 3136 l'ion $[M-H]^-$ du fragment *TGCCTGGGCA* (M= 3140,0 Da) et à m/z 4342 l'ion $[M-H]^-$ du fragment *TGCCTGGGCGCCTA* (M= 4351,8 Da). Ils sont caractéristiques respectivement de la présence des allèles T et C du SNP rs2839698 du locus *H19*. Les spectres de masse montrent également à m/z 4211 l'ion $[M-H]^-$ du fragment *GCCCCCTTCTTTCA* (M= 4221,7 Da) et à m/z 4231 l'ion $[M-H]^-$ du fragment *GCCCCCTTTTTTCA* (M= 4236,7 Da). Ils sont caractéristiques respectivement de la présence des allèles G et A du SNP rs3741476 du locus *NOS1*.

Enfin, les spectres de masse montrent à m/z 3414 l'ion $[M-H]^-$ du fragment *GTCCGTGTCTA* (M= 3420,2 Da) et à m/z 3504 l'ion $[M-H]^-$ du fragment *GGGGGCTGCTA* (M= 3510,2 Da). Ils sont caractéristiques du re-séquençage du locus *NOS1* (tableau 4.2a).

Le clivage du duplex des *ribo-PCR* ATP permet d'identifier respectivement le SNP R rs3741476 du locus *NOS1* et le SNP Y rs2839698 du locus *H19* : un homozygote G et un hétérozygote Y (figure 4.5A), un homozygote A et un homozygote T (figure 4.5B) et un hétérozygote R et un homozygote C (figure 4.5C). Elle permet également le re-séquençage des deux séquences.

3.3.4. L'analyse du locus *SLCO1B1*

Pour le locus *SLCO1B1*, deux *ribo-PCRs* sont possibles avec CTP ou GTP.

Pour celle de CTP (tableau 4.2d), le SNP Y est identifié par deux fragments de clivage de masse 11527,5 Da pour l'allèle T et 11543,5 Da pour l'allèle C. La différence de masse entre les deux fragments de clivage est de 16 Da qui correspond à la substitution de la base A par G. Dans la région des 37-mer, les pics des ions sont difficilement détectables à cause du faible pouvoir résolutif de l'instrument. Les masses ne peuvent donc pas être déterminées de façon précise. L'écart de masse de 16 Da est au-delà du pouvoir résolutif de l'instrument. La substitution d'un seul nucléotide ne peut pas être déterminée mais la présence des fragments

de clivage est identifiée. Le SNP ne peut donc pas être analysé avec cette *ribo-PCR* CTP mais un re-séquençage partiel reste possible.

Par contre, pour la *ribo-PCR* de GTP (tableau 4.2e), le SNP Y est identifié par deux fragments de clivage de masse 4869,1 Da pour l'allèle C et 4884,2 Da pour l'allèle T. Dans la région des 16-mer, la différence de masse entre les deux fragments de clivage est de 15 Da qui correspond à la substitution de la base C par T. Le pouvoir résolutif de l'instrument permet de détecter cette différence de masse. Le re-séquençage ainsi que le génotype du SNP peuvent donc être analysés avec la *ribo-PCR* GTP.

Les spectres de masse de la *ribo-PCR* GTP qui identifie le SNP sont présentés dans la figure 4.6.

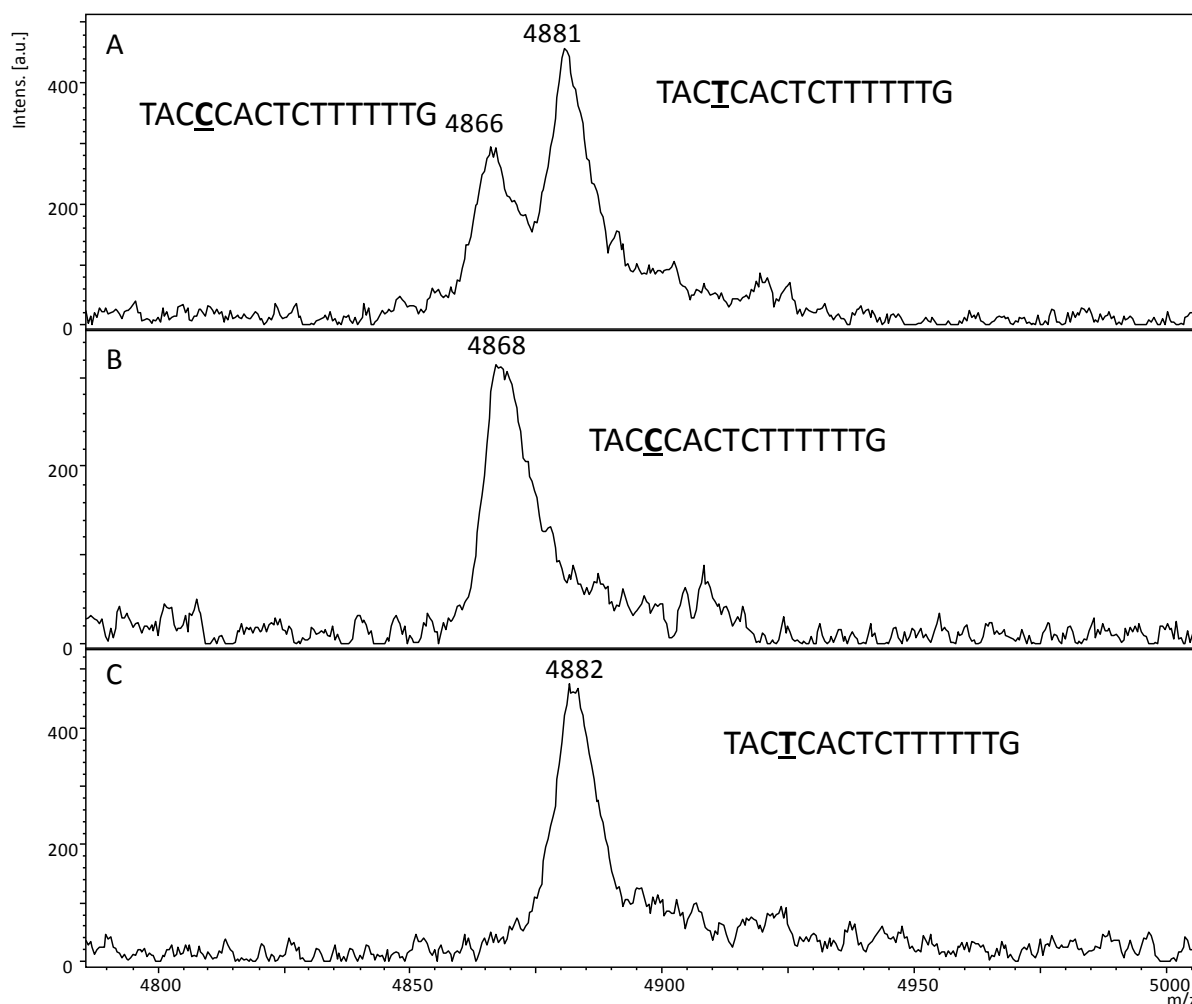


Figure 4.6 : Spectres de masse MALDI-TOF de la région des 16-mer de la *ribo-PCR* GTP des fragments sens $[M-H]^-$ du locus *SLCO1B1* rs4363657 d': (A) un hétérozygote Y; (B) un homozygote C et (C) un homozygote T. L'ion $[M-H]^-$ TACCCACTCTTTTTTG représente l'allèle C et l'ion $[M-H]^-$ TACTTACTCTTTTTTG représente l'allèle T.

Dans la région des 16-mer, ils montrent à m/z 4868 l'ion $[M-H]^-$ du fragment TACCCACTCTTTTTTG ($M= 4869,1$ Da) et à m/z 4882 l'ion $[M-H]^-$ du fragment TACTICACTCTTTTTTG ($M= 4884,2$ Da). Ils sont caractéristiques respectivement de la présence des allèles C et T du SNP rs4363657. Un homozygote C (figure 4.6B), un homozygote T (figure 4.6C) et un hétérozygote Y (figure 4.6A) du SNP rs4363657 sont identifiés par le clivage de la *ribo-PCR* GTP du locus *SLCO1B1*.

3.4. L'analyse des individus

En raison de la couverture par les deux brins complémentaires, la séquence est complètement interprétable pour les loci étudiés ici.

Le re-séquençage du locus *NOS1* rs3741476 du SNP R des 30 individus a permis d'identifier : 12 individus homozygotes A, 3 individus homozygotes G et 15 individus hétérozygotes R (tableau 4.3).

| Individu | NOS1 rs3741476 | H19 rs2839698 |
|-----------------|-----------------------|----------------------|
| 310003201 | AG | C |
| 310016201 | A | CT |
| 110013201 | A | CT |
| 210015201 | A | CT |
| 710180201 | AG | CT |
| 210131201 | AG | T |
| 710214201 | AG | CT |
| 310038201 | AG | CT |
| 710320201 | A | CT |
| 710315201 | G | CT |
| 310048201 | AG | CT |
| 410017201 | A | T |
| 710322201 | G | CT |
| 710313201 | AG | C |
| 210309201 | AG | CT |
| 210001201 | AG | CT |
| 210010201 | A | CT |
| 210111201 | A | T |
| 710173201 | A | CT |
| 210113201 | AG | CT |
| 210115201 | AG | C |
| 710246201 | G | CT |
| 710248201 | AG | CT |
| 210181201 | A | CT |
| 210175201 | A | CT |
| 210180201 | A | CT |
| 710281201 | A | C |
| 310050201 | AG | CT |
| 210187201 | AG | CT |
| 410127201 | AG | T |

Tableau 4.3 : Génotypes des 30 individus du locus NOS1 rs3741476 et du locus H19 rs2839698.

Le re-séquençage du locus H19 rs2839698 du SNP Y des 30 individus a permis d'identifier : 4 individus homozygotes C, 4 individus homozygotes T et 22 individus hétérozygotes Y (tableau 4.3).

Le re-séquençage du locus SLCO1B1 rs4363657 du SNP Y des 95 individus a permis d'identifier : 53 individus homozygotes T, 12 individus homozygotes C et 30 individus hétérozygotes Y (tableau 4.4).

| Individu | rs4363657 | Individu | rs4363657 |
|-----------------|------------------|-----------------|------------------|
| A002324 | TT | B002S17 | TC |
| A00230Z | CC | B002S1Q | TT |
| A00231K | TC | B002S20 | TT |
| A0023VU | TC | B002S2M | TT |
| B000XFE | TT | B002S2Y | TC |
| B002RZD | TT | B002S38 | CC |
| B002S0T | TC | A0023VQ | TT |
| B002S1H | TT | A00231G | TT |
| B002S1V | TC | A00231S | TC |
| B002S24 | TC | B000USM | TT |
| B002S2T | CC | B002RZA | TT |
| B002S32 | TT | B002SDD | TT |
| A002327 | TC | B002S18 | TT |
| A0023SA | TC | B002S1R | TT |
| A00231N | TT | B002S21 | TC |
| A00234T | TC | B002S2Q | CC |
| B000XK5 | CC | B002S2Z | TC |
| B002SDB | TT | B002S39 | TC |
| B002S0U | TT | A00230S | TT |
| B002S1I | TT | A00231H | TT |
| B002S1W | TT | A00231U | CC |
| B002S26 | TT | B000UWG | TT |
| B002S2U | TC | B002RZ5 | TC |
| B002S34 | TT | B002SDE | TT |
| A002329 | TT | B002SDI | TC |
| A002318 | TT | B002S1S | TT |
| A00231O | TC | B002S22 | TT |
| A00234V | TC | B002S2R | TT |
| B002RYZ | TT | B002S30 | TT |
| B002SDC | TT | B002S3E | TT |
| B002SDH | TT | A00230W | TC |
| B002S1M | TT | A00231I | TT |
| B002S1X | TT | A0023VT | TT |
| B002S2J | TT | B000X8P | TC |
| B002S2W | TC | B002RZA | TC |
| B002S35 | TC | B002RZM | TC |
| A00232Z | CC | B002SDJ | CC |
| A00231D | TT | B002S1U | TT |
| B000UI3 | TC | B002S23 | TT |
| B002SD9 | TT | B002S2S | TC |
| B002RZG | CC | B002S31 | TT |
| B002S11 | TT | B002S3F | TC |
| B002S1P | TT | A00231E | CC |
| B002S1Y | TC | A00231Q | TT |
| B002S2K | TT | B000Y2Y | CC |
| B002S2X | TC | B002SDA | TC |
| B002S37 | TT | B002RZH | TT |
| A002331 | CC | | |

Tableau 4.4 : Génotypes des 95 individus du locus *SLCO1B1* rs4363657.

Les génotypes obtenus par cette méthode sont identiques à ceux obtenus par la méthode d'Amplifluor[®] Assay sur les 30 individus pour les locus *NOS1* et *H19* et par la méthode Illumina[®] Infinium[™] II sur les 95 individus pour le locus *SLCO1B1*. De plus, aucun polymorphisme non attendu n'a été identifié dans ces séquences étudiées.

4. Discussion

Une nouvelle méthode de re-séquençage par *ribo-PCR* pour la préparation d'une chimère ARN/ADN double-brin à partir de l'ADN génomique a pu être établie (Mauger et al. 2012).

Cette méthode est basée sur la méthode de génotypage des SNPs (Mauger et al. 2006) ainsi que celle de séquençage de l'ADN mitochondriale (Mauger et al. 2007) qui analysent les fragments de clivage d'une chimère ARN/ADN simple-brin par MALDI-TOF MS.

Le re-séquençage de l'ADN génomique est effectué à l'aide du clivage par la soude de la *ribo-PCR* et de l'analyse des fragments de clivage par MALDI-TOF MS. La comparaison des masses moléculaires des fragments de clivage avec ceux de l'empreinte de la séquence de référence permet d'identifier les différences de masse afin d'identifier les changements dans la séquence. Les polymorphismes (SNP, délétion, insertion ou mutation) connus ou inconnus peuvent être détectés par cette méthode.

4.1. L'ADN polymérase *KB17*

4.1.1. Les propriétés

Contrairement aux autres enzymes utilisées en séquençage, l'ADN polymérase *KB17* contient un domaine 3'-5' exonucléase qui permet de maintenir la *haute-fidélité* dans la réaction de *ribo-PCR*. Elle est également compatible avec les dUTP (Schonbrunner et al. 2006), ce qui permet de se prévenir des contaminations avec l'utilisation de l'uracile N-glycosylase (UNG) contrairement à la plupart des ADN polymérases.

4.1.2. La séquence spécifique d'amplification

La chimère ARN/ADN double-brin a été synthétisée respectivement avec l'échange de 100% d'un dNTP (dATP, dCTP, dGTP ou dTTP) par 100% d'un NTP (ATP, CTP, GTP ou UTP) à partir d'une réaction de *ribo-PCR*. Cette chimère ARN/ADN double-brin est stable et peut atteindre environ 100 pb. Il peut être possible d'augmenter sa taille lors d'optimisations plus poussées.

Toutefois, une séquence spécifique d'amplification a été déterminée. Les séquences riches en GC sont amplifiables par une *ribo-PCR* ATP et/ou UTP alors que les séquences pauvres en GC sont amplifiables par une *ribo-PCR* CTP et/ou GTP.

Pour la *ribo-PCR* ATP du locus *NOS1*, le brin synthétisé sens contient 21-22 % d'A, 16-18% de G et 32 % C tandis que le brin synthétisé anti-sens contient 31 % d'A, 25% de G et 21-22 % de C. Dans le brin synthétisé sens, 5 G sont consécutifs tandis que dans le brin synthétisé anti-sens 4 A et 5 C sont consécutifs.

Pour les *ribo-PCRs* ATP ou UTP du locus *H19*, le brin synthétisé sens contient 24% d'A, 16-41% de G et 22-23 % de C tandis que le brin synthétisé anti-sens contient 12-14 % d'A, 22-24% de G et 43 % de C. Dans les brins synthétisés, seulement 3 A sont consécutifs tandis que 4 G et 4 C sont consécutifs.

Pour les *ribo-PCRs* GTP ou CTP du locus *SLCO1B1*, le brin synthétisé sens contient 29% d'A, 16% de G et 16-17 % de C tandis que le brin synthétisé anti-sens contient 34-35 % d'A, 17-18% de G et 16 % de C. Dans les brins synthétisés, seulement 3 G et 3 C sont consécutifs tandis que 5 A sont consécutifs.

Le succès de la *ribo-PCR* amplifié par l'ADN polymérase *KB17* réside également en l'absence de séquence répétitive de plus de 4 bases du ribonucléotide d'incorporation dans le brin synthétisé.

4.1.2. Le multiplexage des *ribo-PCRs*

Cette ADN polymérase permet de réaliser un multiplexage des *ribo-PCRs*. Pour cela, les séquences doivent être à peu près de même composition afin d'être amplifiées par le même ribonucléotide.

Plusieurs *ribo-PCRs* peuvent être analysés lors d'une réaction de multiplexage mais il faut que les masses des fragments de clivage des séquences soient différentes pour ne pas perdre de l'information par rapport à chaque *ribo-PCR* seule. Plus le nombre de multiplex des *ribo-PCRs* est élevé, plus il sera difficile d'analyser les différentes séquences et donc de pouvoir identifier des polymorphismes. La possibilité de pouvoir amplifier plusieurs séquences dans une réaction de multiplexage de *ribo-PCRs* est intéressante mais son application dans cette méthode devient vite inexploitable.

L'ADN polymérase *KB17* peut donc synthétisée d'une part, une chimère ARN/ADN simple-brin d'environ 500 pb (Mauger et al. 2007) et d'autre part, une chimère ARN/ADN double-brin d'environ 100 pb. Elle permet également l'étude de différentes séquences dans une réaction de multiplexage des *ribo-PCRs* (Mauger et al. 2012).

4.2. L'analyse par MALDI-TOF MS

4.2.1. Le mode linéaire

Les fragments de clivage sont analysés par MALDI-TOF MS en mode linéaire. Le mode linéaire permet de couvrir une gamme large de masse (1-10 kDa) mais il possède un faible pouvoir résolutif. L'instrument a été optimisé spécialement pour l'étude de l'ADN, permettant donc qu'un maximum d'ions pénètre dans le TOF. Cependant, plus la gamme de masse augmente, plus les pics des ions s'élargissent et diminuent en intensité. Les ions de haute masse restent plus longtemps dans le tube de vol et donc fragmentent plus avant d'être détectés. Le pouvoir résolutif de l'instrument devient donc de plus en plus faible quand les masses augmentent. Dans les hautes masses, le pouvoir résolutif de l'instrument est inférieur à la différence entre la substitution d'un seul nucléotide. Il est donc impossible de détecter un polymorphisme dans les hautes masses. Cependant, les fragments de haute masse sont statistiquement peu nombreux et le brin complémentaire peut également donner l'information.

4.2.2. L'étalonnage

L'étalonnage est externe et réalisé dans une gamme de masse 1000 à 10000 Da. Il peut donc y avoir des écarts de masse suivant les masses des fragments de clivage étudiés. Ce phénomène est d'autant plus important que les pics des ions s'élargissent quand les masses augmentent.

Par ailleurs, la matrice THAP ne cristallise pas de manière homogène. Il y a de nombreux cristaux de différentes tailles. Suivant la position du tir du laser sur les cristaux de matrice, il peut y avoir un écart de temps qui se traduit par un écart de masse.

Au problème de mauvaise homogénéité de la cristallisation de la matrice s'ajoute la mauvaise détection des ions de hautes masses. Il peut donc y avoir des écarts de plus en plus importants du rapport m/z des ions par rapport à la masse attendue des fragments de clivage.

4.2.3. L'analyse du double-brin

Le mode réflecteur peut également être utilisé mais seulement les fragments de clivage de masse inférieure à 4000 Da sont détectés. Toute la séquence ne peut donc pas être entièrement analysée en mode réflecteur. En revanche, il peut être utilisé dans le cas de masses proches de 1 à 5 Da. Cependant, les fragments de clivage qui peuvent avoir une différence de masse inférieure à 5 Da sont statistiquement peu nombreux dans le cas de l'étude d'une séquence de 100 pb. Dans les 3 loci étudiés, il n'y a pas de fragments de clivage qui ont une différence de masse de moins de 5 Da (tableau 4.2).

Par ailleurs, l'analyse en mode réflecteur n'est pas nécessaire car cette méthode tire l'avantage de la complémentarité des informations des deux brins sens et anti-sens de la *ribo-PCR*. Pour chaque séquence étudiée, l'assemblage des informations des fragments de clivage de la *ribo-PCR* a permis de donner des résultats sans ambiguïté de la séquence de l'individu. Les changements par rapport à la séquence de référence ont été clairement établis et a permis d'attribuer ainsi les génotypes des individus. Comme la séquence des deux brins est analysée simultanément, chaque *ribo-PCR* de 100 pb a permis l'analyse de la séquence de 200 bases ainsi que 400 bases pour le duplex.

Enfin, deux *ribo-PCRs* peuvent être synthétisées en parallèle (ATP et UTP ou CTP et GTP) pour une même séquence permettant ainsi d'apporter encore plus d'informations sur la séquence.

4.2.4. La détection des fragments de clivage

Les fragments sans base thymine n'ont pas pu être détectés dans ces conditions d'analyse. Les bases thymine uracile ont une affinité protonique beaucoup plus faible que celle de l'adénine, la cytosine et la guanine. Lors de la formation, des ions zwitterion, la thymine ou l'uracile sont beaucoup moins protonées que les autres bases et elle sont donc plus stables en phase gazeuse (chapitre 1). De plus, l'introduction d'un délai d'extraction a permis d'augmenter le temps de présence des ions dans la source favorisant ainsi leur fragmentation en phase gazeuse. Les fragments de clivage sans thymine sont plus protonés que ceux qui en contiennent et par conséquent fragmentent plus facilement en phase gazeuse. Ils sont donc moins détectables dans ces conditions d'analyses par MALDI-TOF MS. En augmentant la quantité analysée de ces fragments, ils deviennent détectables (chapitre 1).

L'analyse des spectres de masse des *ribo*-PCR UTP a montré des fragments de dépurination. En effet, l'uracile est également plus stable en phase gazeuse (chapitre 1). Comme tous les fragments de clivage contiennent une base uracile, ils sont donc tous détectables. Cependant, la seule base uracile en position 3' terminale ne stabilise que partiellement le fragment de clivage. Les autres bases adjacentes du fragment de clivage peuvent être protonées et donc dépurinées en phase gazeuse (chapitre 1). Ce phénomène est accentué lorsque le nombre de bases des fragments de clivage augmentent (chapitre 1). Par conséquent, les spectres de masse des *ribo*-PCR UTP sont plus difficile à analyser que ceux des *ribo*-PCR, ATP, CTP ou GTP.

4.3. Les avantages de la méthode

Notre méthode est rapide, facile à mettre en œuvre et peu coûteuse. L'ADN polymérase *KB17* incorpore efficacement et rapidement des NTPs peu coûteux. Les nucléotides modifiés onéreux tel que α -S-CTP, α -S-UTP ou les nucléotides qui possèdent un groupement fluorescent ne sont pas nécessaires.

La méthode de séquençage *MassARRAY*[®] utilise une ARN polymérase modifiée T7 (Hahner et al. 1997) afin de transcrire l'ADN en ARN. Les amorces ont une séquence spécifique afin de contenir le promoteur. Contrairement à ces approches, notre méthode ne nécessite pas de séquence spécifique des amorces ce qui augmente les possibilités de choix des amorces.

La réaction chimique de clivage est rapide, totale et peu coûteuse. L'utilisation d'enzyme spécifique de clivage onéreuse n'est pas nécessaire.

L'étape de dessalage est également rapide et facile à mettre en œuvre.

Le MALDI-TOF MS est rapide, facile d'utilisation et permet l'analyse d'un grand nombre d'échantillons.

La complémentarité de l'analyse des 2 deux brins de la *ribo-PCR* permet de s'affranchir des limites du pouvoir résolutif de l'instrument.

Grâce à l'utilisation de l'ADN polymérase *KB17*, cette méthode comporte moins d'étapes que dans les autres approches. Le nombre d'étapes est également réduit par rapport à notre protocole de re-séquençage de l'ADN mitochondrial (Mauger et al. 2007) et celui de l'analyse de la méthylation de l'ADN (chapitre 3). Le protocole de re-séquençage comporte une étape de PCR afin d'amplifier la séquence à étudier, une purification afin d'enlever les dNTPs et les amorces de PCR en excès de la PCR et une étape d'élongation linéaire pour former la chimère ARN/ADN simple-brin. Plusieurs élongations linéaires (ATP, CTP et GTP), en parallèle, sont nécessaires afin d'apporter des informations complémentaires sur la séquence. Le protocole d'analyse de la méthylation de l'ADN contient une étape supplémentaire de conversion de l'ADN. Au contraire, dans cette méthode, une seule *ribo-PCR* est nécessaire et elle fournit toutes les informations nécessaires pour bien re-séquencer la région étudiée.

5. Conclusion

Dans ce chapitre, une nouvelle méthode de re-séquençage de l'ADN utilisant le clivage chimique d'une chimère ARN/ADN double-brin et l'analyse des fragments par MALDI-TOF MS a été développée (Mauger et al. 2012). La chimère a été préparée lors d'une réaction de *ribo-PCR* à partir d'une séquence d'ADN génomique, de deux amorces, d'un mélange d'un ribonucléotide (A, C, G ou U) et des trois autres désoxynucléotides complémentaires et de l'ADN polymérase *KB17*. Cette méthode (annexe 3) a permis de re-séquencer l'ADN en fragmentant directement la chimère ARN/ADN double-brin par un clivage par la soude et en analysant les fragments de clivage par MALDI-TOF MS. Les masses des ions sont utilisées pour analyser les déviations par rapport à la séquence de référence qui sont caractéristiques de

modification de la séquence. Cette méthode de re-séquençage permet de distinguer les positions homozygotes ou hétérozygotes de la séquence.

Toutes les *ribo-PCRs* sont composées d'environ 100 pb. Des optimisations pourraient permettre d'étudier des séquences plus grandes. De plus, une séquence spécifique d'amplification a été déterminée, les séquences riches en GC sont amplifiables par ATP et/ou UTP tandis que les séquences pauvres en GC sont amplifiables par CTP et/ou GTP. La réussite de l'amplification dépend également de l'absence de plus de quatre bases consécutives du ribonucléotide d'incorporation dans le brin synthétisé. L'étude de plusieurs loci dans la même réaction est possible par l'intermédiaire du multiplexage des *ribo-PCRs* amplifiables par le même ribonucléotide.

Grâce à la capacité haut-débit du MALDI-TOF MS, cette méthode est bien adaptée pour l'analyse de petites séquences d'ADN spécifique du génome pour un grand nombre d'individus. Elle a permis d'analyser 30 individus pour le re-séquençage des loci *NOS1* rs3741476 et *H19* rs2839698 et 95 individus pour le re-séquençage du locus *SLCO1B1* rs4363657.

Cette méthode de re-séquençage identifie les polymorphismes connus. Elle peut être également utilisée afin de détecter des polymorphismes inconnus de la séquence. Elle constitue une nouvelle alternative dans l'arsenal des méthodes d'analyse de l'ADN qui ne cesse de croître. La méthode présentée ici n'est pas plus compliquée qu'un protocole de génotypage classique mais elle permet également de détecter des mutations inconnues.

Son principal avantage, par rapport aux autres méthodes de séquençage utilisant la spectrométrie de masse, réside dans sa réduction du nombre d'étapes de biologie moléculaire, sa simplicité de mise en œuvre et son coût probablement inférieur aux méthodes actuelles.

La *ribo-PCR* pourrait également être utilisée pour d'autres applications que le MALDI-TOF MS présenté ici. Dans le chapitre 5, une autre application de l'utilisation de la chimère ARN/ADN double-brin et du MALDI-TOF MS est présentée.

Chapitre 5

**Multiplex allèle-spécifique
géotypage
par clivage
d'une *riboPAP-PCR*
et
par MALDI-TOF MS**

Chapitre 5. Multiplex allèle-spécifique génotypage par clivage d'une *riboPAP-PCR* et par MALDI-TOF MS

1. Introduction

Dans le chapitre 4, une méthode de re-séquençage de l'ADN par clivage d'une chimère ARN/ADN double-brin (*ribo-PCR*) et d'analyse par spectrométrie de masse MALDI-TOF a été développée (Mauger et al. 2012). Elle permet le séquençage de polymorphismes connus et inconnus. Cependant, l'identification du polymorphisme connu dépend de la séquence du fragment de clivage qui le contient. Il peut y avoir des fragments de clivage de masse identique ou des fragments de clivage de haute masse qui peuvent être également difficiles à détecter par MALDI-TOF MS. Le pouvoir résolutif de l'instrument peut être au-delà de la différence de masse des bases du polymorphisme ou bien la masse du fragment de clivage peut être supérieure au seuil de détection de l'instrument. La masse du fragment de clivage peut donc être difficile à détecter par MALDI-TOF MS. Il serait donc plus facile de choisir la masse du fragment de clivage à analyser et par conséquent sa séquence.

Dans ce chapitre, une nouvelle méthode de multiplex allèle-spécifique génotypage et re-séquençage de l'ADN par clivage d'une chimère ARN/ADN double-brin (*riboPAP-PCR*) et d'analyse par spectrométrie de masse MALDI-TOF est développée (annexe 4 et 5). Cette méthode est basée sur la méthode de re-séquençage du chapitre 4 (Mauger et al. 2012). C'est une méthode de multiplex génotypage allèle-spécifique par l'intermédiaire d'amorces qui sont spécifiques de l'allèle et qui contiennent une séquence répétitive, appelé *tag*, à leur extrémité 5' afin d'attribuer une masse différente à chaque allèle (Gelfand et al. 2009).

D'autres méthodes de génotypage allèle-spécifique existent qui utilisent notamment des amorces allèle-spécifiques avec un *tag* 5'-GC sur l'une d'entre elles ce qui entraîne une température de fusion différente des produits de PCR (Germer & Higuchi 1999) permettant ainsi de les discriminer. L'allèle-spécificité est déterminée par la complémentarité du génotype du côté 3' des amorces et l'élongation de la partie complémentaire de la séquence.

Une autre méthode est basée sur l'utilisation d'amorces qui ne possèdent pas d'extrémité 3'-OH indispensable à l'élongation de la séquence. La dernière base de l'amorce est enlevée par une réaction de pyrophosphorolyse en fonction de la complémentarité de la séquence avec

celle de l'amorce. L'ADN polymérase incorpore alors un didésoxynucléotide à la place de la dernière base de l'amorce. La réaction de pyrophosphorolyse active donc l'élongation de l'amorce allèle-spécifique. Cette procédure est appelée polymérisation activée par pyrophosphorolyse (*PAP*) (Liu & Sommer 2000). La principale différence entre ces deux méthodes est que dans le cas d'une réaction *PAP*, une erreur dans la sélection de l'amorce n'engendre pas une sélection de la mauvaise séquence.

La méthode, présentée dans ce chapitre, est basée sur le même concept *PAP* mais elle diffère par l'utilisation d'une part, d'amorce plus facile à synthétiser et d'autre part, par l'utilisation d'une ADN polymérase qui est compatible avec la réaction *riboPAP-PCR*. Par ailleurs, avec les méthodes antérieures, le multiplexage des génotypes allèle-spécifique est difficile alors que cette méthode le permet. L'utilisation de différentes amorces de PCR, contenant des *tags* spécifiques, permet un multiplexage de la méthode.

La réaction de polymérisation activée par pyrophosphorolyse, sur les amorces qui possèdent à leur extrémité 3' un groupement 2'-PO₄, permet d'augmenter l'allèle spécificité de chaque *ribo-PAP PCR*. L'ADN polymérase *Thermotoga maritima* (*Tma*) est compatible avec ces amorces modifiées et incorpore des NTPs et dNTPs. Le groupement 2'-PO₄ empêche l'élongation de l'amorce par l'ADN polymérase (Gelfand & Gupta 2008). Par ailleurs, une méthode de synthèse de ses amorces a été développée (Bodepudi et al. 2011). En outre, la substitution empêche l'activité exonucléase 3'-5' de relecture de l'ADN polymérase (Gelfand 1999; Gelfand & Gupta 2008). En présence d'ADN génomique et de pyrophosphate, l'ADN polymérase est capable de supprimer le nucléotide substitué par le groupement 2'-PO₄ de l'amorce par pyrophosphorolyse et incorpore à la place un nucléotide afin de démarrer l'élongation (Gelfand & Gupta 2008). L'utilisation de ces amorces *Hot-Start* qui possèdent un *tag* améliore donc la sélectivité et la spécificité de la réaction et constitue un moyen simple de déterminer l'amorce qui a participé à la PCR et ainsi le génotype de l'individu.

Cette méthode de génotypage allèle-spécifique est également une méthode de re-séquençage. Elle est composée de quatre étapes : (i) une *riboPAP-PCR* allèle-spécifique, (ii) un clivage, (iii) un dessalage et (iv) l'analyse par MALDI-TOF MS. La *riboPAP-PCR* génère une chimère ARN/ADN double-brin spécifique de l'allèle présent. Son clivage engendre d'une part, des fragments de re-séquençage et d'autre part, des fragments complémentaires de la séquence répétitive du *tag* de l'amorce (*flag*). La masse moyenne de chaque *flag* est spécifique de l'amorce allèle-spécifique et permet donc d'identifier l'allèle présent. La

présence de plusieurs amorces contenant des *tags* différentes permet l'étude de plusieurs loci dans une même réaction ainsi que l'analyse des génotypes homozygotes ou hétérozygotes des individus.

Tout comme dans le chapitre 4, cette méthode est mise au point sur les trois loci d'ADN génomique : *NOS1* *163731 et *H19* *103280 sur 30 individus ainsi que *SLCO1B1* *604843 sur 95 individus. Pour l'étude du locus *SLCO1B1*, les deux SNP rs4149056 et rs4363657 qui sont corrélés avec les patients à risque d'un traitement par statines à hautes doses (Link et al. 2008) ont été étudiés lors d'une même réaction.

2. Matériels et Méthodes

2.1. Matériels

Les 30 échantillons d'ADN utilisés pour l'étude des loci *NOS1* et *H19* proviennent d'individus caucasiens. Les 95 échantillons d'ADNs utilisés pour l'étude du locus de *SLCO1B1* proviennent d'individus ayant des antécédents d'infarctus du myocarde. Ils ont été choisis en fonction de leurs génotypes. Les 30 échantillons d'ADNs ont été génotypés par la méthode d'Amplifluor[®] Assay et les 95 ADNs par Illumina[®] Infinium[™] II Assay (Link et al. 2008).

Les amorces de *riboPAP-PCR* ont été synthétisées par Roche Molecular Systems (Pleasanton, CA, USA). Les séquences des amorces allèles-spécifiques sont listées dans le tableau 5.1. Elles contiennent un 5'-*tag* qui est une séquence répétitive spécifique pour chaque allèle ainsi qu'une modification 2'-PO₄ sur la dernière base. La séquence 5'-*tag* est composée de 5 ou 6 bases.

| Locus | Amorces |
|-----------------------------|--|
| <i>NOS1</i> rs3741476 | <u>CCTAGAAACTAGAAACTAGAACTCTGATGGCTCACCATTGAAAA*</u> GTCAATGAAGGAAGGTAGCA* |
| <i>H19</i> rs2839698 | <u>CCTAAGACTAAGACTAAGACTCGTGAGGAGTGTGGAGTAGGC*</u> |
| <i>SLCO1B1</i> rs4149056 | TTTAAAGGAATCTGGGTCATA* <u>CCTAGAGACTAGAGACTAGAGACTCAAGCATATTACCCATGAACA*</u> |
| <i>SLCO1B1</i> rs4363657 | CAATTTTTCAGAATAATTTAGTACA* <u>CCTAGAGCTAGAGCTAGAGCTCCTGGAAATACAAAAAAGAGTGA*</u> <u>CCTGGAGCTGGAGCTGGAGCTCCTGGAAATACAAAAAAGAGTGG*</u> |

Tableau 5.1 : Amorces de la riboPAP-PCR des loci *NOS1*, *H19* et *SLCO1B1*. L'astérisque sur la dernière base indique la modification 2'PO₄ et les séquences surlignées correspondent au tag.

Les dNTPs (N=A, C, G et T) et les NTPs (N=A, et G) proviennent de GE Healthcare (Saclay, France) et les produits chimiques d'Aldrich (Steinheim, Allemagne). La résine échangeuse de cations provient de Biorad (Marnes la coquette, France) et le kit *Clean Resin Tool* de *MassARRAY*[®] de chez Sequenom (Hamburg, Allemagne). Les kits d'étalons externes d'oligonucléotides de 4 à 11-mer et de 12 à 30-mer proviennent de chez Bruker (Bremen, Allemagne).

L'ADN polymérase *FP-1* a été développée par Roche Molecular Systems (Pleasanton, CA, USA). Elle n'est pas commercialisée, elle peut être obtenue sur demande auprès du Dr Thomas Myers (thomas.myers@roche.com). Cette chimère d'ADN polymérase contient un domaine 5'-nucléase de l'ADN polymérase *Thermus sp. ZO5* et les domaines 3'-nucléase et 5'-3'-nucléase de l'ADN polymérase *Thermotoga maritima (Tma)*. L'activité 5'-nucléase a été supprimée par la mutation *G46E* (Mauger et al. 2006). L'activité *haute-fidélité* de l'ADN polymérase non-mutée *Tma* a été modulée par l'introduction d'une mutation *L329A* (Schonbrunner et al. 2006; Mauger et al. 2006) du 'Motif I' du domaine 3'-5' exonucléase. L'enzyme contient une mutation qui élimine la spécificité de l'enzyme non-mutée contre l'incorporation des nucléotides substitués en position 2'-C, favorisant ainsi l'incorporation des NTPs. Enfin, l'enzyme contient trois mutations supplémentaires dans le domaine de la polymérase qui améliore l'accroche sur la matrice entraînant ainsi l'augmentation de la rapidité d'élongation, particulièrement dans l'incorporation de ribonucléotides et améliore le taux de pyrophosphorolyse (Bauer 2009).

Les procédures de thermocyclage ont été réalisées sur des thermocycleurs gradients Eppendorf (Eppendorf, Allemagne). Les analyses ont été réalisées par un spectromètre de masse MALDI-TOF *Autoflex*TM de chez Bruker (Bremen, Allemagne). La cible MALDI est une *Anchorchip*TM Bruker (Bremen, Allemagne).

2.2. Méthodes

2.2.1. La riboPAP-PCR par l'ADN polymérase *FP-1*

Les trois locus sont amplifiés par *riboPAP-PCR* : *NOS1*, *H19* et *SLCO1B1*.

La *riboPAP-PCR* du locus *NOS1*

Le mélange de 20 µL de la solution de *riboPAP-PCR* du locus *NOS1* contient 4 µL de tricine à 500 mM (pH 7,3), 4 µL d'acétate de potassium 500 mM (pH 7,5), 0,75 µL de pyrophosphate de sodium à 4 mM, 2 µL d'acétate de magnésium à 30 mM, 1 µL d'un mélange de ATP, dCTP, dGTP et dTTP à 4 mM chacun, 0,8 µL de chaque amorce (tableau 5.1) à 10 pmol/µL, 0,5 µL d'enzyme *FP-1* à 10 U/µL, 2 µL d'ADN à 10 ng/µL et 3,35 µL d'eau milli-Q.

La *riboPAP-PCR* du locus *H19*

Le mélange de 20 µL de la solution de la *riboPAP-PCR* de *H19* contient 4 µL de tricine à 500 mM (pH 7,3), 4 µL d'acétate de potassium 500 mM (pH 7,5), 0,75 µL de pyrophosphate de sodium à 4 mM, 2 µL d'acétate de magnésium à 30 mM, 1 µL d'un mélange de dATP à 0,4 mM, ATP à 3,6 mM, dCTP, dGTP et dTTP à 4 mM chacun, 0,8 µL de chaque amorce (tableau 5.1) à 10 pmol/µL, 0,5 µL d'enzyme *FP-1* à 10 U/µL, 2 µL d'ADN à 10 ng/µL et 4,9 µL d'eau milli-Q.

Le duplex des *riboPAP-PCRs* des loci *SLCO1B1*

Le mélange de 20 µL de la solution des *riboPAP-PCR* des deux loci *SLCO1B1* contient 4 µL de tricine à 500 mM (pH 7,3), 4 µL d'acétate de potassium 500 mM pH 7,5, 0,75 µL de pyrophosphate de sodium à 4 mM, 2 µL d'acétate de magnésium à 30 mM, 1 µL d'un mélange de dGTP à 0,4 mM, GTP à 3,6 mM, dATP, dCTP et dTTP de à 4 mM chacun, 0,8 µL de chaque amorce du locus *SLCO1B1* rs4149056 et 0,8 µL du locus *SLCO1B1* rs4363657 (tableau 5.1) à 10 pmol/ µL, 0,5 µL d'enzyme *FP-1* à 10 U/µL, 2 µL d'ADN à 10 ng/µL et 4,9 µL d'eau milli-Q.

Les réactions de *riboPAP-PCR* sont dénaturées pendant 4 min à 92°C, puis elles sont suivies de 46 cycles de 15s à 92°C et 5 min à 63°C pour le locus *NOS1*, 66°C pour le locus *HI9* et 60°C pour le locus *SLCO1B1*.

Un mélange de 5 µL de la réaction de *riboPAP-PCR* et de 5 µL de la solution de tampon de charge est ensuite déposé sur un gel d'agarose à 2 % afin de vérifier la taille du produit PCR.

2.2.2. Le clivage

Au 15 µL de la réaction de *riboPAP-PCR* est ajouté un mélange de 2 µL d'hydroxyde de sodium à 3 M et 3 µL d'eau milli-Q. La réaction de clivage est mise sous incubation pendant 1h30 à 70 °C.

2.2.3. Le dessalage

A la réaction de clivage, 6 mg de résines échangeuses de cations sont ajoutées à l'aide du kit *Clean Resin Tool* de *MassARRAY*[®]. Les échantillons sont incubés pendant 20 min à température ambiante sous agitation et ensuite centrifugés sous 134 xg afin de récupérer le surnageant.

2.2.4. L'analyse par spectrométrie de masse MALDI-TOF

2.2.4.1. La préparation de la cible MALDI

La solution de matrice trihydroxyacetophenone (THAP) est préparée avec 0,2 M de 2,4,6 THAP dans 50 % d'acétonitrile, 0,2 M de 2,3,4 THAP dans 50% d'acétonitrile et 0,3 M de citrate d'ammonium dans les proportions 6/3/2 (v/v) (Zhu et al. 1996).

Sur une cible *Anchorchip*TM Bruker de 384 spots de porosité 400 µm, sont déposés 0,5 µL de la solution de matrice THAP et 0,5 µL du surnageant par spot. La cible est séchée à température ambiante.

2.2.4.2. L'analyse

La cible MALDI est introduite dans un spectromètre de masse MALDI-TOF. L'analyse est réalisée en mode linéaire, négatif, à l'aide du logiciel *FlexControl* et des paramètres du MALDI-TOF MS suivants : IS1=20 kV ; IS2=18,5 kV ; lens=7,5 kV et un délai d'extraction de 150 ns. Les spectres de masse sont obtenus par l'accumulation de 200 tirs de laser.

Un étalonnage externe est réalisé à l'aide des kits d'étalons externes d'oligonucléotides de 4 à 11-mer et de 12 à 30-mer de chez BRUKER.

Ces conditions permettent d'analyser des masses moyennes et de détecter des échantillons d'environ 0,15 pmol.

Les spectres de masse sont analysés à l'aide du logiciel *FlexAnalysis*. L'annotation des pics caractérisés par le rapport m/z des ions, la soustraction de la ligne de base ainsi qu'un lissage du signal sont réalisés.

Un logiciel de calcul de masse des fragments de clivage a été créé par Jérémy Semhoun (Mauger et al. 2007). Il a été modifié afin de pouvoir calculer les masses moyennes des fragments de clivage des *riboPAP-PCRs*.

Les rapports m/z des ions sont comparés aux masses des *flags* afin d'identifier les allèles correspondants pour chaque individu et celles de la séquence de référence.

3. Résultats

3.1. Le principe de la méthode

Le principe de la méthode de multiplex allèle-spécifique génotypage est illustré dans la figure 5.1 (Gelfand et al. 2009).

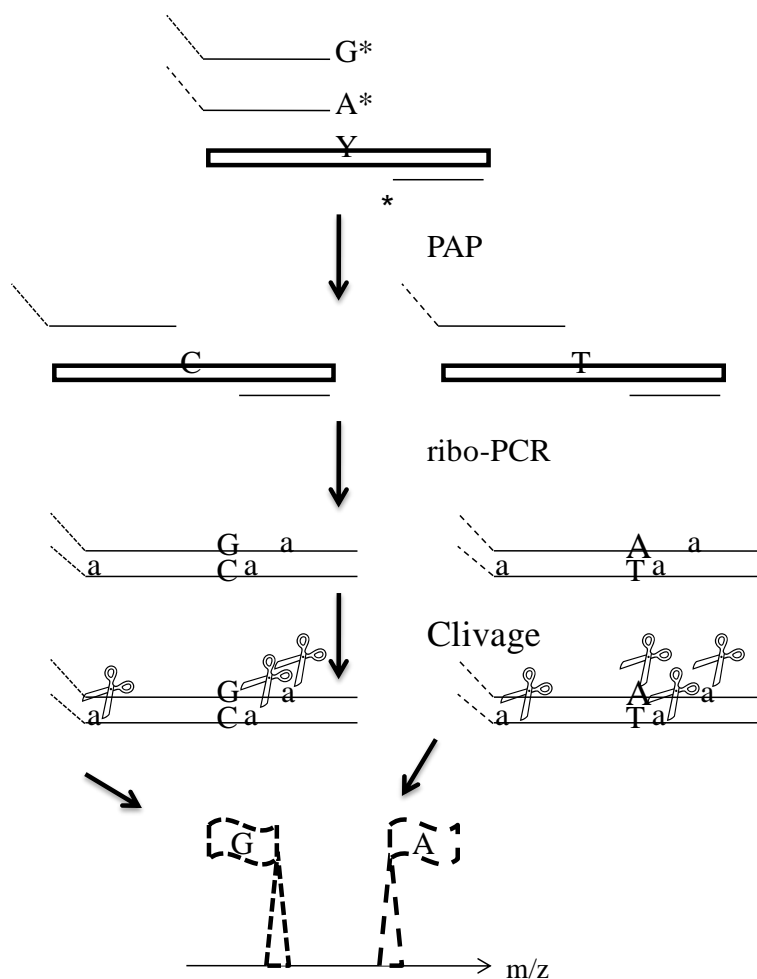


Figure 5.1 : Principe de la riboPAP-PCR avec ATP pour l'étude d'un SNP Y. Tout d'abord, une réaction de pyrophosphorolyse, PAP, a lieu lors de l'accroche des amorces allèles-spécifiques contenant les tags sur l'ADN. La dernière base qui contient la modification * 2' PO₄ est enlevée. Puis, la réaction de ribo-PCR ATP a lieu en incorporant de nouveau la dernière base des amorces. Elle est ensuite clivée afin d'analyser les séquences allèles-spécifiques flag par MALDI-TOF MS.

Elle comprend la formation d'une chimère ARN/ADN double-brin lors d'une réaction de *riboPAP-PCR* à l'aide des amorces sens et anti-sens contenant une modification 2'-PO₄ sur l'extrémité 3'. Les amorces sens, allèle-spécifique, contiennent également une séquence répétitive spécifique en position 5', appelée *tag*. La chimère est ensuite clivée chimiquement. Les fragments de clivage sont dessalés par une résine échangeuse de cations et ensuite analysés par MALDI-TOF MS en mode linéaire négatif. Chaque allèle possède une masse différente, appelée *flag*, qui permet d'identifier le génotype de l'individu. Cette méthode a été mise au point par l'étude de trois loci génomique : *NOS1*, *H19* et *SLCO1B1* en duplex.

3.2. La *riboPAP-PCR* par l'ADN polymérase *FP-1*

L'ADN polymérase *KB17* incorpore facilement des ribonucléotides (100 % A, 100 % C ou 100 % G) dans une réaction d'élongation linéaire (Mauger et al. 2007) d'environ 500 pb (chapitre 3) ainsi que dans une réaction de *ribo-PCR* (100 % A, 100 % C, 100 % G ou 100 % U) d'environ 100 pb (chapitre 4).

Dans ce travail, la préparation d'une *riboPAP-PCR* d'environ 80 bp a été réalisée avec l'ADN polymérase *FP-1* dans une réaction de simplex ou de duplex. La *riboPAP-PCR* du locus *NOS1* a été générée à partir de 100 % ATP et 8 bases sont présentes entre les amorces sens et anti-sens. La *riboPAP-PCR* du locus *H19* a été générée à partir de 90 % ATP et 10 % dATP et 15 bases sont présentes entre les amorces sens et anti-sens. Le duplex de *riboPAP-PCR* des loci de *SLCO1B1* a été généré à partir de 90 % GTP et 10 % dGTP et respectivement 8 et 14 bases sont présentes entre les deux amorces sens et anti-sens des loci rs4363657 et rs4149056. Dans le duplex, la concentration des amorces de la *ribo-PAP PCR* du loci rs4363657 est le double de celle de rs4149056.

Les amorces allèles-spécifiques sont constituées d'une séquence, *5'-tag*, répétitive et spécifique de chaque allèle (tableau 5.2).

| Locus | Tag | Flag |
|-----------------------------|-----------------------------------|--------------------|
| <i>NOS1</i> rs3741476 | CCT(AGAAACT)3C CCT(AAAAACCT)3C | GTTTCTA GTTTTTA |
| <i>H19</i> rs2839698 | CCT(AAGACT)3C CCT(AGAGCT)3C | GTCTTA GCTCTA |
| <i>SLCO1B1</i> rs4149056 | CCT(AGAGACT)3C CCT(GAGGACT)3C | TCTCTAG TCCTCAG |
| <i>SLCO1B1</i> rs4363657 | CCT(AGAGCT)3C CCT(GGAGCT)3C | CTCTAG CTCCAG |

Tableau 5.2 : Séquences des tags et des flags des loci *NOS1* rs3741476 avec ATP, de *H19* rs2839698 avec ATP et du duplex *SLCO1B1* rs4363657 et rs4149056 avec GTP. Les tags sont les séquences spécifiques des amorces et les flags sont les séquences complémentaires des tags générés après clivage de la chimère ARN/ADN.

Le *tag* a été choisi en fonction de la masse du fragment complémentaire du *flag* généré après clivage de la chimère, qui doit être différente de celles des fragments de re-séquençage du locus étudié. Chaque *tag* a été conçu afin d'être utilisé lors d'une réaction de *riboPAP-PCR* avec ATP ou GTP. Elles sont donc constituées de différentes proportions de A et G ainsi que

d'un seul C et T à la fin d'une répétition. Chaque motif du *tag* est répété 3 fois et il contient 6 à 7 bases afin d'être facilement détectable par MALDI-TOF MS.

3.3. Les fragments de clivage

Les fragments pour l'analyse par spectrométrie de masse sont générés par un clivage chimique par la soude (Mauger et al. 2006; Mauger et al. 2007) après chaque ribonucléotide. Les fragments de clivage contiennent des désoxynucléotides et se terminent par un ribonucléotide et un groupement 3'-phosphate. Les masses moyennes des fragments de clivage des *riboPAP-PCRs* sont listées dans les tableaux 5.3 : A, B et C.

5.3A

| <i>NOSI</i> rs3741476 | Séquence | Masse | Début | Fin | Sens |
|-----------------------|--|---------|-------|-----|-----------|
| | tca | 940,6 | 28 | 26 | anti-sens |
| | gcca | 1254,8 | 32 | 29 | anti-sens |
| | ttga | 1284,8 | 69 | 72 | sens |
| | tggtga | 1943,2 | 38 | 33 | anti-sens |
| Flag A | 3 gtttcta | | 23 | 17 | anti-sens |
| | | | 16 | 10 | anti-sens |
| | | | 9 | 3 | anti-sens |
| Flag G | 3 gttttta | | 23 | 17 | anti-sens |
| | | | 16 | 10 | anti-sens |
| | | | 9 | 3 | anti-sens |
| | ccttcttca | 3009,9 | 59 | 68 | sens |
| | gggggctgcta | 3509,2 | 48 | 58 | sens |
| G | gcccccttcttca | 4221,7 | 53 | 40 | anti-sens |
| A | gccccctttttca | 4236,7 | 53 | 40 | anti-sens |
| | gtcaatgaaggaaggtagc a | 6335,1 | 73 | 54 | anti-sens |
| A | cctagaaactagaaactagaaactctgatggctcaccattgaaa a | 13919,1 | 1 | 45 | sens |
| G | cctaaaaactaaaaactaaaaactctgatggctcaccattgaaa ga | 14200,3 | 1 | 46 | sens |

5.3B

| H19 rs2839698 | Séquence | Masse | Début | Fin | Sens |
|----------------------|--|--------------|--------------|------------|-------------|
| | tca | 940,6 | 71 | 73 | sens |
| | cga | 965,6 | 54 | 52 | anti-sens |
| | cga | 965,6 | 23 | 21 | anti-sens |
| T | ccta | 1229,8 | 41 | 38 | anti-sens |
| | ggca | 1294,8 | 48 | 51 | sens |
| | ctcca | 1519,0 | 37 | 33 | anti-sens |
| | | | 20 | 15 | anti-sens |
| Flag T | 3 gctcta | 1863,2 | 14 | 9 | anti-sens |
| | | | 8 | 3 | anti-sens |
| | | | 20 | 15 | anti-sens |
| Flag C | 3 gtctta | 1878,2 | 14 | 9 | anti-sens |
| | | | 8 | 3 | anti-sens |
| | ggcga | 1953,2 | 63 | 68 | sens |
| | ctcctca | 2112,4 | 30 | 24 | anti-sens |
| | tcgtgca | 2192,4 | 52 | 58 | sens |
| T | tgccctgggca | 3140,0 | 51 | 42 | anti-sens |
| C | tgccctgggcgcca | 4351,8 | 51 | 38 | anti-sens |
| | gctgatgctgcctgtc tgca | 6485,2 | 75 | 55 | anti-sens |
| C | cctaagactaagactaagactcgtgaggagtgtggagtagg cgccca | 14682,5 | 1 | 47 | sens |
| T | cctagagctagagctagagctcgtgaggagtgtggagtagg tgccca | 14745,5 | 1 | 47 | sens |

5.3C

| SLCO1B1 | Séquence | Masse | Début | Fin | Sens |
|------------------|---|--------------|--------------|------------|-------------|
| rs4149056 | cttg | 1260,8 | 53 | 56 | sens |
| rs4149056 | tatg | 1284,8 | 21 | 18 | anti-sens |
| | | | 59 | 64 | sens |
| rs4363657 Flag C | 3 ctccag | 1848,2 | 65 | 70 | sens |
| | | | 71 | 76 | sens |
| rs4149056 | accag | 1857,2 | 17 | 12 | anti-sens |
| | | | 59 | 64 | sens |
| rs4363657 Flag T | 3 ctctag | 1863,2 | 65 | 70 | sens |
| | | | 71 | 76 | sens |
| rs4149056 | ttcatg | 1878,2 | 38 | 43 | sens |
| | | | 59 | 65 | sens |
| rs4149056 Flag T | 3 tcctcag | 2152,4 | 66 | 72 | sens |
| | | | 73 | 79 | sens |
| | | | 59 | 65 | sens |
| rs4149056 Flag C | 3 tcctcag | 2167,4 | 66 | 72 | sens |
| | | | 73 | 79 | sens |
| rs4149056 | atatatg | 2215,4 | 29 | 35 | sens |
| rs4149056 | taatatg | 2215,4 | 46 | 52 | sens |
| rs4363657 | aaaaattg | 2441,7 | 8 | 1 | anti-sens |
| rs4363657 | taccactg | 2754,8 | 32 | 24 | anti-sens |
| rs4363657 | tattccag | 2784,8 | 47 | 55 | sens |
| rs4149056 | attcctttaa | 3290,2 | 11 | 1 | anti-sens |
| rs4149056 C | catatatccacatg | 4302,8 | 35 | 22 | anti-sens |
| rs4363657 | tactaaattattctg | 4637,0 | 23 | 9 | anti-sens |
| rs4363657 C | taccactctttttg | 4869,1 | 31 | 46 | sens |
| rs4363657 T | taccactctttttg | 4884,2 | 31 | 46 | sens |
| rs4149056 | tttaaaggaatctgggcat acatg | 7846,1 | 1 | 25 | sens |
| rs4363657 | caattttcagaataatttagtac ag | 8093,3 | 1 | 26 | sens |
| rs4363657 C | cctggagctggagctggagctcctggaatacaaaaaagagtg g | 13800,0 | 77 | 34 | anti-sens |
| rs4149056 C | cctagagctagagactagactcaagcatattaccatgaac g | 13936,1 | 80 | 36 | anti-sens |
| rs4363657 T | cctagagctagagctcctggaatacaaaaaagagtg ag | 14065,2 | 77 | 33 | anti-sens |
| rs4149056 T | cctgaggactgaggactgaggactcaagcatattaccatgaac acatatccacatg | 18236,9 | 80 | 22 | anti-sens |

Tableau 5.3 : Masses moyennes des fragments de clivage des riboPAP-PCRs : (A) de *NOS1* rs3741476 avec ATP; (B) de *H19* rs2839698 avec ATP; (C) du duplex *SLCO1B1* rs4363657 et rs4149056 avec GTP. Les amorces des riboPAP-PCRs sont en italique. Le SNP est mis en gras et souligné.

3.4. L'analyse des spectres de masse MALDI-TOF

Tout comme dans les méthodes précédentes, les échantillons sont dessalés à l'aide d'une résine échangeuse de cations.

Les échantillons possèdent une charge négative portée par l'atome d'oxygène du groupement phosphate, ils sont donc analysés en mode négatif.

De plus, grâce à la distribution des masses moyennes des fragments de re-séquençage (tableau 5.3) et des *flags* dans les régions de 6 à 7-mer, ils sont analysés en mode linéaire.

Cette méthode a été appliquée à différents loci génomiques : *NOS1* rs3741476 (SNP R) et *H19* rs2839698 (SNP Y) sur 30 individus et *SLCO1B1* rs4363657 (SNP Y) et rs4149056 (SNP Y) sur 95 individus.

3.4.1. L'analyse du locus *NOS1*

Les spectres de masse des *flags* de la riboPAP-PCR ATP, identifiant le SNP R, sont présentés dans la figure 5.2.

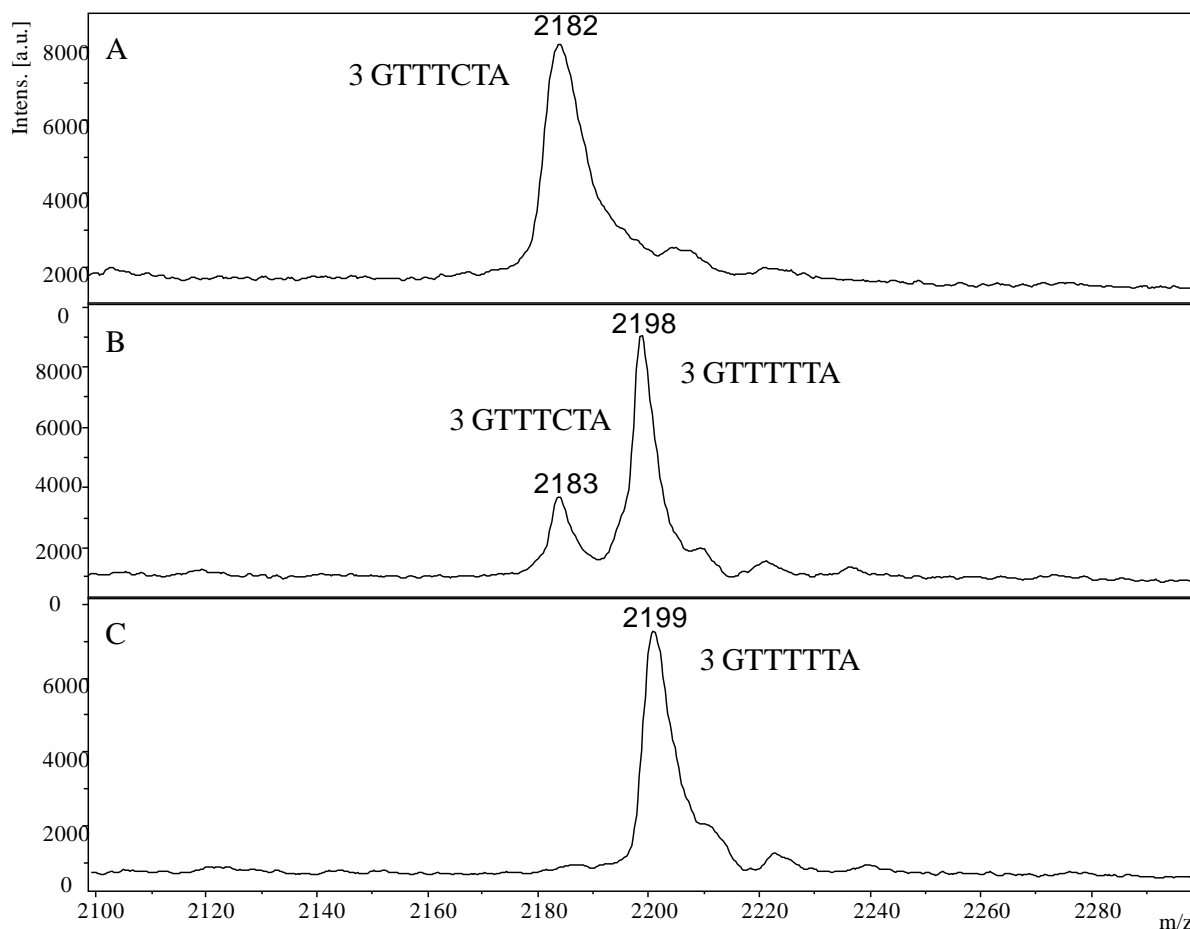


Figure 5.2 : Spectres de masse MALDI-TOF de la région 7-mer de la riboPAP-PCR ATP des ions des flags $[M-H]^-$ du locus *NOS1* rs3741476 d': (A) un homozygote A; (B) un hétérozygote R et (C) un homozygote G. L'ion $[M-H]^-$ GTTTTTA représente l'allèle A et l'ion $[M-H]^-$ GTTTCTA représente l'allèle G.

Dans la région des 7-mer, ils montrent à m/z 2182 les trois ions $[M-H]^-$ du flag GTTTCTA ($M=2182,4$ Da) et à m/z 2197 les 3 ions $[M-H]^-$ du flag GTTTTTA ($M=2197,4$ Da). Ils sont caractéristiques respectivement de la présence des allèles A et G du SNP R rs3741476 du locus *NOS1*. Un homozygote A (figure 5.2A), un hétérozygote R (figure 5.2B) et un homozygote G (figure 5.2C) du SNP rs3741476 sont identifiés par le clivage de la *riboPAP-PCR* ATP du locus *NOS1*.

D'autre part, dans la région des 14-mer du spectre de masse, le SNP est également identifié par la présence à m/z 4220,7 de l'ion $[M-H]^-$ du fragment GCCCCCTTCTTTCA ($M=4221,7$ Da) et à m/z 4235,7 de l'ion $[M-H]^-$ du fragment GCCCCCTTTTTTCA ($M=4236,7$ Da). Ils sont caractéristiques respectivement de la présence des allèles G et A du SNP rs3741476 du locus *NOS1*.

3.4.2. L'analyse du locus *H19*

Les spectres de masse des *flags* de la *riboPAP-PCR* 90 % ATP/10% dATP, identifiant le SNP Y, sont présentés dans la figure 5.3.

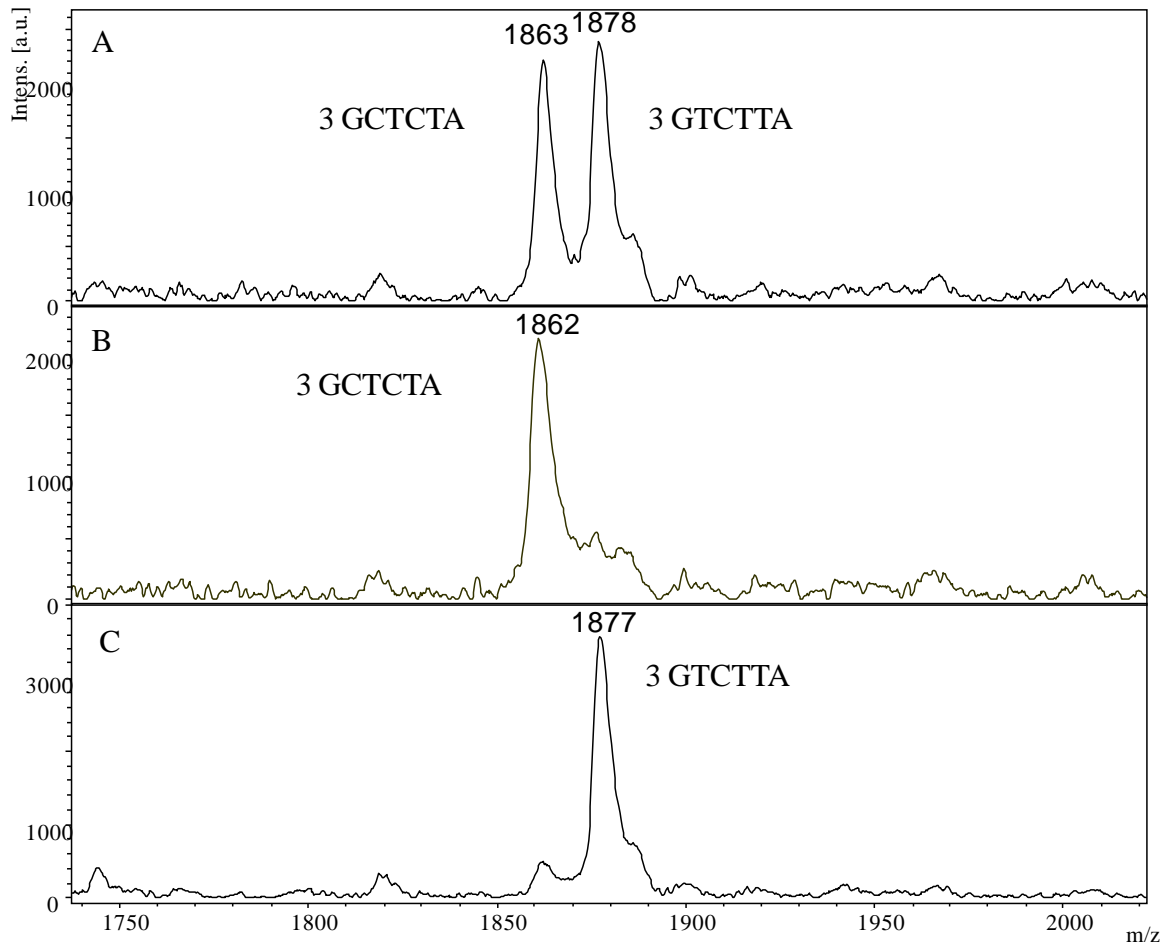


Figure 5.3 : Spectres de masse MALDI-TOF de la région des 6-mer de la *riboPAP-PCR* ATP des *flags* $[M-H]^-$ du locus *H19* rs2839698 d': (A) un hétérozygote R; (B) un homozygote T et (C) un homozygote C. L'ion $[M-H]^-$ GCTCTA représente l'allèle T et l'ion $[M-H]^-$ GTCTTA représente l'allèle C.

Dans la région des 6-mer, ils montrent à m/z 1862 les trois ions $[M-H]^-$ du *flag* GCTCTA ($M=1863,8$ Da) et à m/z 1877 les trois ions $[M-H]^-$ du *flag* GTCTTA ($M=1878,2$ Da). Ils sont caractéristiques respectivement de la présence des allèles T et C du SNP Y rs2839698 du locus *H19*. Un hétérozygote R (figure 5.3A), un homozygote T (figure 5.3B) et un homozygote C (figure 5.3C) du SNP rs2839698 sont identifiés par le clivage de la *riboPAP-PCR* ATP du locus *H19*.

D'autre part, dans la région des 10 à 14-mer du spectre de masse, le SNP est également identifié par la présence à m/z 3139,0 de l'ion $[M-H]^-$ du fragment TGCCTGGGCA ($M=3140,0$ Da) et à m/z 4350,8 de l'ion $[M-H]^-$ du fragment TGCCTGGGCCCTA ($M=4351,8$ Da). Ils sont caractéristiques respectivement de la présence des allèles T et C du SNP rs3741476 du locus *H19*.

3.4.3. L'analyse du duplex des loci *SLCO1B1*

Les spectres de masse des *flags* du duplex *riboPAP-PCR* 90%GTP/10% dGTP, identifiant les deux SNPs, sont présentés dans la figure 5.4.

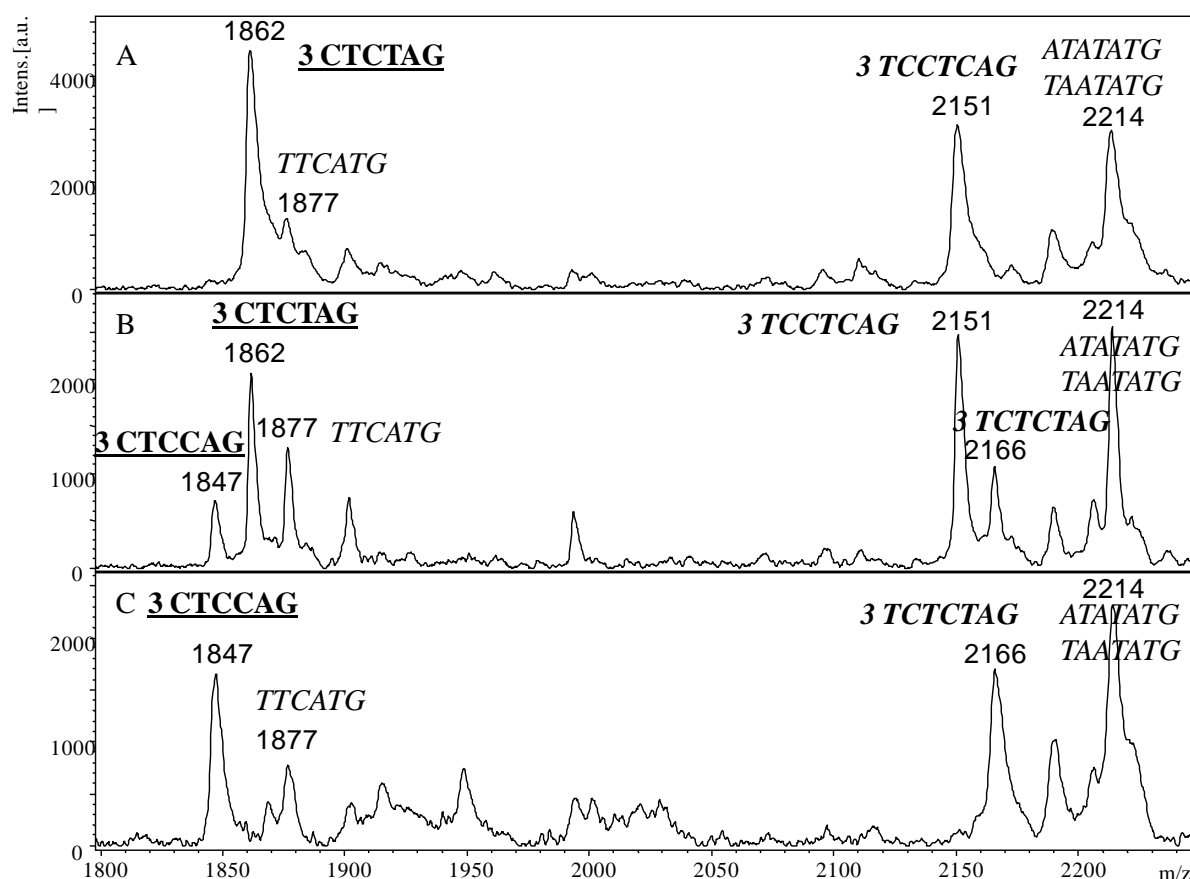


Figure 5.4: Spectres de masse MALDI-TOF de la région des 6 à 7-mer du duplex de la riboPAP-PCR GTP des flags $[M-H]^-$ des loci *SLCO1B1* rs4363657 et rs4149056 d' : (A) un homozygote TT-IT; (B) un hétérozygote TC-TC et (C) un homozygote CC-CC. Les flags du SNP rs4363657 sont en gras et soulignés tandis que les flags du SNP rs4149056 sont en gras et en italique. Les autres fragments proviennent du re-séquençage de la riboPAP-PCR du locus rs4149056. L'ion $[M-H]^-$ CTCTAG représente l'allèle C et l'ion $[M-H]^-$ CTCCAG représente l'allèle T du SNP rs4363657. L'ion $[M-H]^-$ TCCTCAG représente l'allèle T et l'ion $[M-H]^-$ TCTCTAG représente l'allèle C du SNP rs4149056.

Dans la région des 6-mer, ils montrent à m/z 1847 les trois ions $[M-H]^-$ du *flag* CTCCAG ($M=1848,3$ Da) et à m/z 1862 les trois ions $[M-H]^-$ du *flag* CTCTAG ($M=1863,2$ Da). Ils sont caractéristiques respectivement de la présence des allèles C et T du SNP Y rs4363657 du loci *SLCO1B1*.

Dans la région des 7-mer, ils montrent à m/z 2151 les trois ions $[M-H]^-$ du *flag* TCCTCAG ($M=2152,4$ Da) et à m/z 2166 les trois ions $[M-H]^-$ du *flag* TCTCTAG ($M=2167,4$ Da). Ils sont caractéristiques respectivement de la présence des allèles T et C du SNP Y rs4149056 du locus *SLCO1B1*.

Un homozygote TT-TT (figure 5.4A), un hétérozygote TC-TC (figure 5.4B) et un homozygote CC-CC (figure 5.4C) des SNPs rs4363657 et rs4149056 sont identifiés par le clivage des *riboPAP-PCRs* GTP des deux locus *SLCO1B1*.

De plus, dans la région des 6-mer, les spectres de masse montrent également à m/z 1877 l'ion $[M-H]^-$ du fragment TTCATG ($M= 1878,2$ Da) et dans la région des 7-mer, à m/z 2214,4 les ions $[M-H]^-$ des fragments ATATATG et TAATATG ($M=2215,4$ Da) (figure 5.4). Ces fragments proviennent du re-séquençage de la *riboPAP-PCR* GTP du locus *SLCO1B1* rs4149056. Par contre, l'ion $[M-H]^-$ du fragment ACCCAG ($M=1857,2$ Da) provenant également du re-séquençage du locus du SNP rs4149056 n'est pas observé dans le spectre de masse. Ce fragment ne contient pas de thymine, il est donc plus difficile à détecter (chapitre 1).

D'autre part, dans la région des 16-mer, le SNP rs4363657 est également identifié par les ions à m/z 4868,1 de l'ion $[M-H]^-$ du fragment TACCCACTCTTTTTTG ($M=4869,1$ Da) et à m/z 4883,2 de l'ion $[M-H]^-$ du fragment TACCTACTCTTTTTTG ($M=4884,2$ Da). Ils sont caractéristiques respectivement de la présence des allèles C et T du SNP rs4363657 du locus *SLCO1B1*.

Enfin, dans la région des 12-mer, à m/z 4301,8 l'ion $[M-H]^-$ du fragment CATATCCACATG ($M=4302,8$ Da) est caractéristique de la présence de l'allèle C du SNP rs4149056.

3.4.4. Les *flags* du duplex des loci *SLCO1B1*

Les conditions du duplex des deux loci de *SLCO1B1* ont été établies avec les deux couples de *flags* suivants : CTCTAG pour l'allèle C et CTCCAG pour l'allèle T du SNP rs4363657 et

TCCTCAG pour l'allèle T et TCTCTAG pour l'allèle C du SNP rs4149056. Cependant, d'une part, d'autres couples de *tags* peuvent être utilisés et d'autre part, les *tags* peuvent être associés aux autres allèles (tableau 5.4).

| <i>SLCO1B1</i> | <i>Tag</i> | <i>Flag</i> | <i>Masse</i> |
|------------------|-----------------|-------------|--------------|
| rs4363657 | CCT(GGAGCT)3C | CTCCAG | 1848,3 |
| | CCT(AGAGCT)3C | CTCTAG | 1863,2 |
| rs4149056 | CCT(GGAGGCT)3C | CCTCCAG | 2137,4 |
| | CCT(GAGGACT)3C | TCCTCAG | 2152,4 |
| | CCT(AGAGACT)3C | TCTCTAG | 2167,4 |
| | CCT(AGAAACT)3C | TTTCTAG | 2182,4 |
| | CCT(AAAA ACT)3C | TTTTTAG | 2197,4 |
| | | | |

Tableau 5.4 : Séquences des tags des amorces allèles-spécifiques des loci *SLCO1B1* rs4363657 et rs4149056 et séquences et masses moyennes des flags correspondantes.

Des expériences ont été réalisées afin d'associer les *tags* de 6-mer avec le SNP rs4149056 et les *tags* 7-mer avec le SNP rs4363657 mais elles n'ont pas permis d'obtenir les résultats souhaités.

Dans cette étude, les *tags* de 6-mer sont associés au SNP rs4363657 tandis que les *tags* de 7-mer sont associés au SNP rs4149056 des loci de *SLCO1B1* (tableau 5.4). Par conséquent, pour l'étude du SNP rs4363657 il n'y a que le couple de *flag* CTCTAG, CTCCAG qui est possible. Par contre, pour l'étude du SNP rs4149056, il y a en plus des deux *flags* TCCTCAG, TCTCTAG, 3 autres *flags* possibles : CCTCCAG, TTTCTAG et TTTTTAG.

3.4.4.1. L'association des *flags* et des allèles des loci *SCLO1B1*

Les spectres de masse présentent les 4 combinaisons possibles pour les deux couples de *flags* suivants : CTCTAG et CTCCAG pour le SNP Y rs4363657 et TCCTCAG et TCTCTAG pour le SNP Y rs4149056 (figure 5.5).

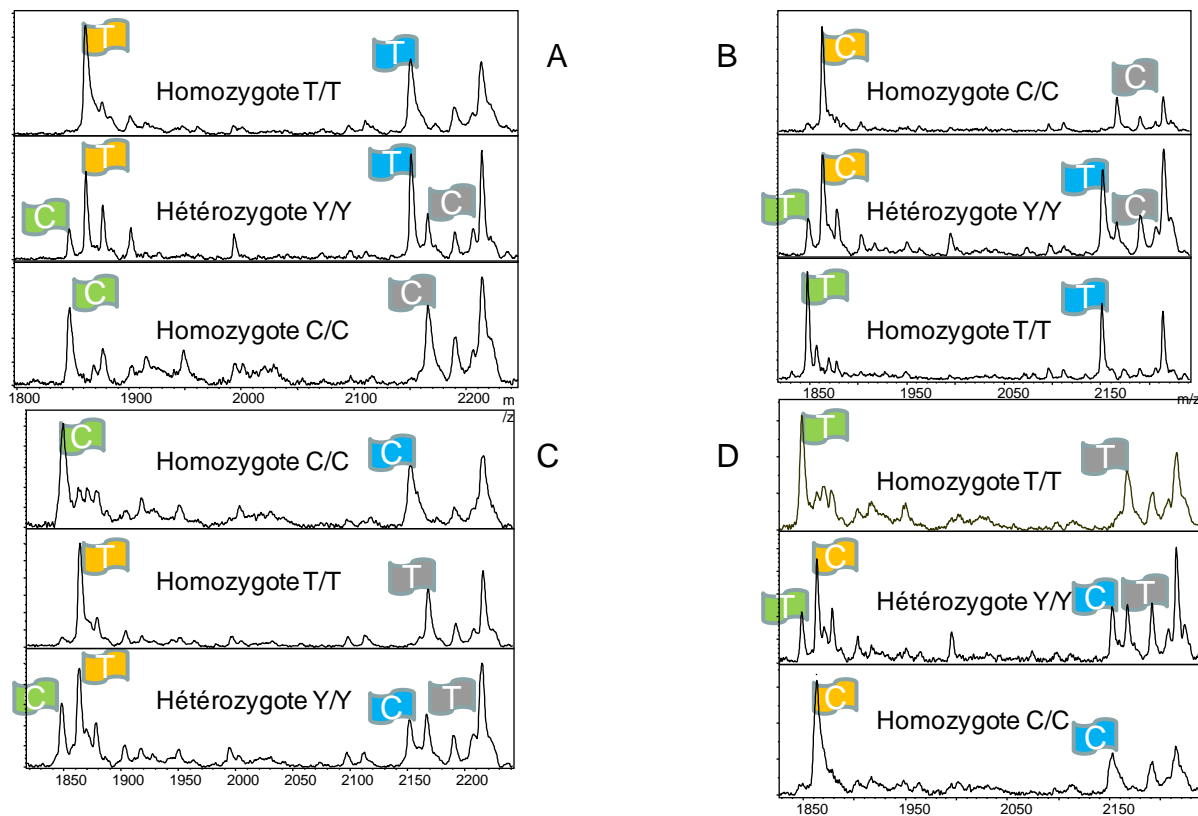


Figure 5.5 : Spectres de masse MALDI-TOF de la région des 6 à 7-mer du duplex de la riboPAP-PCR GTP des flags [M-H]⁻ des loci *SLCO1B1* rs4363657 et rs4149056: (A) CTCTAG pour l'allèle C et CTCCAG pour l'allèle T du SNP rs4363657 et TCCTCAG pour l'allèle T et TCTCTAG pour l'allèle C du SNP rs4149056; (B) CTCTAG pour l'allèle T et CTCCAG pour l'allèle C du SNP rs4363657; (C) TCCTCAG pour l'allèle C et TCTCTAG pour l'allèle T du SNP rs4149056; (D) CTCTAG pour l'allèle T et CTCCAG pour l'allèle C du SNP rs4363657 et TCCTCAG pour l'allèle C et TCTCTAG pour l'allèle T du SNP rs4149056. Le flag orange correspond à CTCTAG, le flag vert correspond à CTCCAG, le flag bleu correspond à TCCTCAG et le flag gris correspond à TCTCTAG. Les autres fragments proviennent du re-séquençage de la riboPAP-PCR du locus rs4149056.

La figure 5.5A montre les *flags* de référence du duplex. Dans le cas d'un individu hétérozygote, elle montre que l'abondance de l'ion [M-H]⁻ du *flag* CTCTAG (M=1848,3 Da) de l'allèle C est inférieure à celle du *flag* CTCCAG (M=1863,2 Da) de l'allèle T du SNP rs4363657 tandis que l'abondance de l'ion [M-H]⁻ du *flag* TCCTCAG (M=2152,4 Da) de l'allèle T est supérieure à celle du *flag* TCTCTAG (M=2167,4 Da) de l'allèle C du SNP rs4149056.

Dans le cas d'un individu hétérozygote, la figure 5.5B montre que l'abondance de l'ion [M-H]⁻ du *flag* CTCTAG (M=1848,3 Da) de l'allèle T est inférieure à celle du *flag* CTCCAG (M=1863,2 Da) de l'allèle C du SNP rs4363657 tandis que l'abondance de l'ion [M-H]⁻ du

flag TCCTCAG (M=2152,4 Da) de l'allèle T est toujours supérieure à celle du *flag* TCTCTAG (M=2167,4 Da) de l'allèle C du SNP rs4149056.

Dans le cas d'un individu hétérozygote, la figure 5.5C montre que l'abondance de l'ion [M-H]⁻ du *flag* CTCTAG (M=1848,3 Da) de l'allèle C est inférieure à celle du *flag* CTCCAG (M=1863,2 Da) de l'allèle T du SNP rs4363657 tandis que l'abondance de l'ion [M-H]⁻ du *flag* TCCTCAG (M=2152,4 Da) de l'allèle C et celle du *flag* TCTCTAG (M=2167,4 Da) de l'allèle T du SNP rs4149056 sont identiques.

La figure 5.5D, montre que dans le cas d'un individu hétérozygote, que l'abondance de l'ion [M-H]⁻ du *flag* CTCTAG (M=1848,3 Da) de l'allèle T est inférieure à celle du *flag* CTCCAG (M=1863,2 Da) de l'allèle C du SNP rs4363657 tandis que l'abondance de l'ion [M-H]⁻ du *flag* TCCTCAG (M=2152,4 Da) de l'allèle C et celle du *flag* TCTCTAG (M=2167,4 Da) de l'allèle T du SNP rs4149056 sont identiques.

La figure 5.5 montre donc que les *tags* peuvent être permutés pour l'étude d'un même locus. Cependant, les abondances des ions [M-H]⁻ des *flags* doivent être la plus proche possible afin de détecter facilement un individu hétérozygote. Pour l'étude du SNP rs4149056, il semble que la meilleure combinaison soit le *flag* TCTCTAG pour l'allèle T et le *flag* TCCTCAG pour l'allèle C car l'abondance des 2 des ions [M-H]⁻ des *flags* est quasiment identique. En revanche, pour l'étude du SNP rs4363657, l'abondance de l'ion [M-H]⁻ du *flag* CTCCAG est toujours inférieure à celle du *flag* CTCTAG dans le cas d'un hétérozygote indépendamment de la présence du *tag* sur l'amorce allèle-spécifique. Le *flag* CTCCAG contient seulement un T alors que le *flag* CTCTAG en contient deux. Le *flag* CTCCAG est donc moins détectable en phase gazeuse que le *flag* CTCTAG par MALDI-TOF MS (chapitre 1).

3.4.4.2. Les *tags* pour l'allèle C du SNP rs4149056 du locus *SCLO1B1*

Les spectres de masse de la figure 5.6 présentent le remplacement du *flag* de l'allèle C du SNP rs rs4149056 par les *flags* suivants: TCTCTAG (figure 5.6A), TTTCTAG (figure 5.6B), TTTTCTAG (figure 5.6C), CCTCCAG (figure 5.6D). Les *flags* restent inchangés : CTCTAG pour l'allèle C et CTCCAG pour l'allèle T du SNP rs4363657 et TCCTCAG pour l'allèle T du SNP rs4149056.

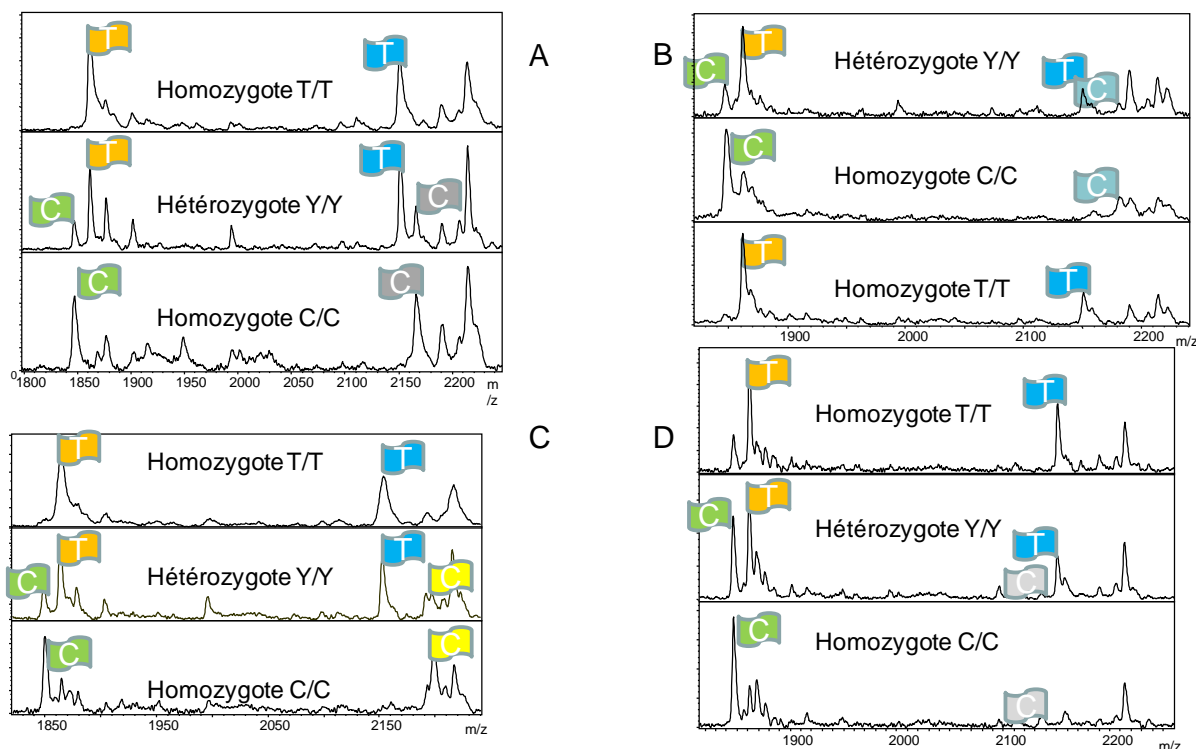


Figure 5.6 : Spectres de masse MALDI-TOF de la région des 6 à 7-mer du duplex de la riboPAP-PCR GTP des flags [M-H]⁻ des loci *SLCO1B1* rs4363657 et rs4149056. Le flag orange CTCTAG correspond à l'allèle T et le flag vert CTCCAG correspond à l'allèle C du SNP rs4363657 tandis que le flag bleu TCCTCAG correspond à l'allèle T du SNP rs4149056. Les flags suivants correspondent à l'allèle C du SNP rs4149056 : (A) gris TCTCTAG; (B) bleu clair TTTCTAG; (C) jaune TTTTTAG et (D) gris clair CCTCCAG. Les autres fragments proviennent du re-séquençage de la riboPAP-PCR du locus rs4149056.

La figure 5.6A montre les *flags* de référence du duplex. Dans le cas d'un individu hétérozygote, elle montre que l'intensité du pic de l'ion [M-H]⁻ du *flag* TCCTCAG (M=2152,4 Da) de l'allèle T est supérieure à celle du *flag* TCTCTAG (M=2167,4 Da) de l'allèle C du SNP rs4149056.

Dans le cas d'un individu hétérozygote, la figure 5.6B montre que l'abondance de l'ion [M-H]⁻ du *flag* TCCTCAG (M=2152,4 Da) de l'allèle T est supérieure à celle du *flag* TTTCTAG (M=2182,4 Da) de l'allèle C du SNP rs4149056.

Dans le cas d'un individu hétérozygote, la figure 5.6C montre que l'abondance de l'ion [M-H]⁻ du *flag* TCCTCAG (M=2152,4 Da) de l'allèle T est supérieure à celle du *flag* TTTTTAG (M=2197,4Da) de l'allèle C du SNP rs4149056.

Dans le cas d'un individu hétérozygote, la figure 5.6D montre que l'abondance de l'ion [M-H]⁻ du *flag* TCCTCAG (M=2152,4 Da) de l'allèle T est très supérieure à celle du *flag*

CCTCCAG (M=2137,4 Da) de l'allèle C du SNP rs4149056. Dans le cas d'un individu homozygote C, l'abondance de l'ion [M-H]⁻ du *flag* CCTCCAG (M=2137,4 Da) de l'allèle C du SNP rs4149056 est également très faible. Le *flag* CCCTCCAG possède qu'un seul T, il est donc moins détectable par MALDI-TOF MS que les autres *flags* (chapitre 1).

Dans tous les spectres de masse des individus hétérozygotes de la figure 5.6, l'abondance de l'ion [M-H]⁻ du *flag* TCCTCAG (M=2152,4 Da) de l'allèle T est toujours supérieure à celle des différents *flags* de l'allèle C du SNP rs4149056. De plus, les *flags* de l'allèle C CCTCCAG, TCTCTAG, TTTCTAG, TTTTCTAG contiennent respectivement 1, 3, 4 et 5 thymines tandis que le *flag* TCCTCAG de l'allèle T ne contient que 2 thymines (chapitre 1). Pour l'étude d'un hétérozygote, les 4 *flags* peuvent être utilisés. Cependant, le *flag* CCTCCAG est moins adapté que les autres.

3.4.4.3. Les *flags* pour l'allèle T du SNP rs4149056 du locus *SCLO1B1*

Les spectres de masse de la figure 5.7 présentent le remplacement du *flag* de l'allèle T du SNP rs rs4149056 par les *flags* suivants: TCCTCAG (figure 5.7A), TTTCTAG (figure 5.7B), TTTTCTAG (figure 5.7C), CCTCCAG (figure 5.7D). Les *flags* restent inchangés : CTCTAG pour l'allèle C et CTCCAG pour l'allèle T du SNP rs4363657 et TCTCTAG pour l'allèle C du SNP rs4149056.

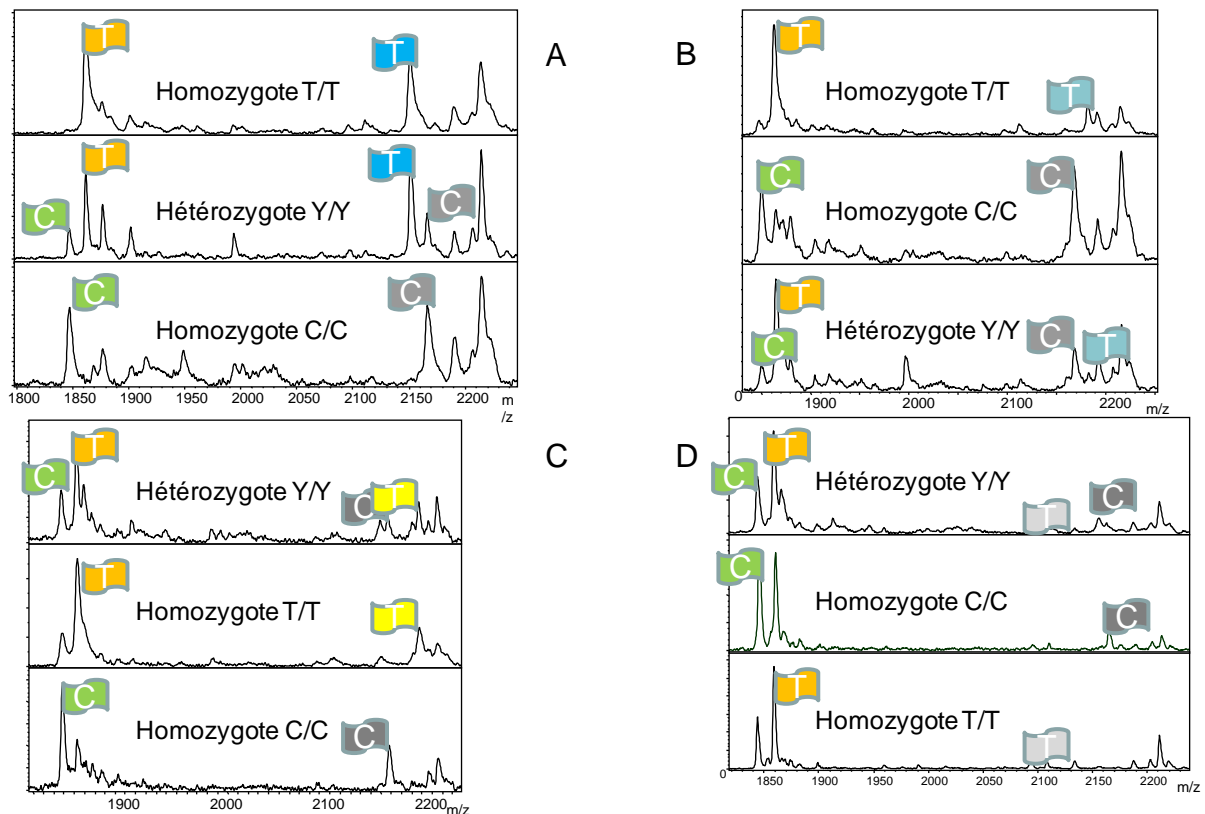


Figure 5.7 : Spectres de masse MALDI-TOF de la région des 6 à 7-mer du duplex de la riboPAP-PCR GTP des flags $[M-H]^-$ des locus *SLCO1B1* rs4363657 et rs4149056. Le flag orange CTCTAG correspond à l'allèle T et le flag vert CTCCAG correspond à l'allèle C du SNP rs4363657 tandis que le flag gris TCTCTAG correspond à l'allèle C du SNP rs4149056. Les flags suivants correspondent à l'allèle T du SNP rs4149056 (A) bleu TCCTCAG; (B) bleu clair TTTCTAG; (C) jaune TTTTTAG et (D) gris clair CCTCCAG. Les fragments proviennent du re-séquençage de la riboPAP-PCR du locus rs4149056.

La figure 5.7A montre les *flags* de référence du duplex. Dans le cas d'un individu hétérozygote, elle montre que l'abondance de l'ion $[M-H]^-$ du *flag* TCCTCAG ($M=2152,4$ Da) de l'allèle T est supérieure à celle du *flag* TCTCTAG ($M=2167,4$ Da) de l'allèle C du SNP rs4149056.

Dans le cas d'un individu hétérozygote, la figure 5.7B montre que l'abondance de l'ion $[M-H]^-$ du *flag* TTTCTAG ($M=2182,4$ Da) de l'allèle T est à peu près égale à celle du *flag* TCTCTAG ($M=2167,4$ Da) de l'allèle C du SNP rs4149056.

Dans le cas d'un individu hétérozygote, la figure 5.7C montre que l'abondance de l'ion $[M-H]^-$ du *flag* TTTTTAG ($M=2197,4$ Da) de l'allèle T est à peu près égale à celle du *flag* TCTCTAG ($M=2167,4$ Da) de l'allèle C du SNP rs4149056.

Dans le cas d'un individu hétérozygote, la figure 5.7D montre que l'abondance de l'ion [M-H]⁻ du *flag* CCTCCAG (M=2137,4 Da) de l'allèle T est très inférieure à celle du *flag* TCTCTAG (M=2167,4 Da) de l'allèle C du SNP rs4149056. De plus, comme pour l'expérience précédente, le cas d'un individu homozygote T, l'abondance de l'ion [M-H]⁻ du *flag* CCTCCAG (M=2137,4 Da) de l'allèle T du SNP rs4149056 est également très faible. Le *flag* CCTCCAG possède qu'un seul T, il est donc moins détectable par MALDI-TOF MS que les autres *flags* (chapitre 1).

De plus, les *flags* de l'allèle T CCTCCAG, TCCTCAG, TTTCTAG, TTTTTAG contiennent respectivement 1, 2, 4 et 5 thymines alors que le *flag* TCTCTAG de l'allèle C ne contient que 3 thymines. Les 4 *flags* peuvent être utilisés. Cependant, le *flag* CCTCCAG est moins adapté que les autres.

3.5. L'analyse des individus

Le génotypage allèle-spécifique de la *riboPAP-PCR* 100% ATP du locus *NOS1* rs3741476 du SNP R des 30 individus a permis d'identifier : 12 individus homozygotes A, 3 individus homozygotes G et 15 individus hétérozygotes R (tableau 4.3).

Le génotypage allèle-spécifique de la *riboPAP-PCR* 90% ATP du locus *H19* rs2839698 du SNP Y des 30 individus a permis d'identifier : 4 individus homozygotes C, 4 individus homozygotes T et 22 individus hétérozygotes Y (tableau 4.3).

Le génotypage par multiplexage allèle-spécifique des *riboPAP-PCR* 90% GTP en duplex des loci *SLCO1B1* rs4363657 et rs4149056 de 95 individus a permis d'identifier : 53 individus homozygotes TT-TT, 12 individus homozygotes CC-CC et 30 individus hétérozygotes Y-Y (tableau 5.5).

| Individu | rs4363657 | rs4149056 | Individu | rs4363657 | rs4149056 |
|----------|-----------|-----------|----------|-----------|-----------|
| A002324 | TT | TT | B002S17 | TC | TC |
| A00230Z | CC | CC | B002S1Q | TT | TT |
| A00231K | TC | TC | B002S20 | TT | TT |
| A0023VU | TC | TC | B002S2M | TT | TT |
| B000XFE | TT | TT | B002S2Y | TC | TC |
| B002RZD | TT | TT | B002S38 | CC | CC |
| B002S0T | TC | TC | A0023VQ | TT | TT |
| B002S1H | TT | TT | A00231G | TT | TT |
| B002S1V | TC | TC | A00231S | TC | TC |
| B002S24 | TC | TC | B000USM | TT | TT |
| B002S2T | CC | CC | B002RZ4 | TT | TT |
| B002S32 | TT | TT | B002SDD | TT | TT |
| A002327 | TC | TC | B002S18 | TT | TT |
| A0023SA | TC | TC | B002S1R | TT | TT |
| A00231N | TT | TT | B002S21 | TC | TC |
| A00234T | TC | TC | B002S2Q | CC | CC |
| B000XK5 | CC | CC | B002S2Z | TC | TC |
| B002SDB | TT | TT | B002S39 | TC | TC |
| B002S0U | TT | TT | A00230S | TT | TT |
| B002S1I | TT | TT | A00231H | TT | TT |
| B002S1W | TT | TT | A00231U | CC | CC |
| B002S26 | TT | TT | B000UWG | TT | TT |
| B002S2U | TC | TC | B002RZ5 | TC | TC |
| B002S34 | TT | TT | B002SDE | TT | TT |
| A002329 | TT | TT | B002SDI | TC | TC |
| A002318 | TT | TT | B002S1S | TT | TT |
| A00231O | TC | TC | B002S22 | TT | TT |
| A00234V | TC | TC | B002S2R | TT | TT |
| B002RYZ | TT | TT | B002S30 | TT | TT |
| B002SDC | TT | TT | B002S3E | TT | TT |
| B002SDH | TT | TT | A00230W | TC | TC |
| B002S1M | TT | TT | A00231I | TT | TT |
| B002S1X | TT | TT | A0023VT | TT | TT |
| B002S2J | TT | TT | B000X8P | TC | TC |
| B002S2W | TC | TC | B002RZA | TC | TC |
| B002S35 | TC | TC | B002RZM | TC | TC |
| A00232Z | CC | CC | B002SDJ | CC | CC |
| A00231D | TT | TT | B002S1U | TT | TT |
| B000UI3 | TC | TC | B002S23 | TT | TT |
| B002SD9 | TT | TT | B002S2S | TC | TC |
| B002RZG | CC | CC | B002S31 | TT | TT |
| B002S11 | TT | TT | B002S3F | TC | TC |
| B002S1P | TT | TT | A00231E | CC | CC |
| B002S1Y | TC | TC | A00231Q | TT | TT |
| B002S2K | TT | TT | B000Y2Y | CC | CC |
| B002S2X | TC | TC | B002SDA | TC | TC |
| B002S37 | TT | TT | B002RZH | TT | TT |
| A002331 | CC | CC | | | |

Tableau 5.5 : Génotypes des 95 individus des loci *SLCO1B1* rs4363657 et rs4149056.

Pour les loci *NOS1* et *H19*, les génotypes obtenus par cette méthode sont identiques à ceux obtenus d'une part par la méthode d'*Amplifluor*[®] Assay et d'autre part par la méthode de *ribo-PCR* (chapitre 4).

Pour les deux loci *SLCO1B1*, les génotypes obtenus par cette méthode sont identiques à ceux obtenus par la méthode *Illumina*[®] *Infinium*[™] II et par la méthode de la *ribo-PCR* (chapitre 4) pour l'étude du locus *SLCO1B1* rs4363657 et par séquençage par électrophorèse capillaire pour l'étude du locus *SLCO1B1* rs4149056 (tableau 5.5).

4. Discussion

Une nouvelle méthode de multiplex génotypage allèle-spécifique et de re-séquençage par *riboPAP-PCR* a été développée (Gelfand et al. 2009). Elle est basée sur celle de la *ribo-PCR* du chapitre 4 (Mauger et al. 2012). La *riboPAP-PCR* est réalisée à partir des amorces allèles-spécifiques du SNP à étudier, un nucléotide sous sa forme ribonucléotide (90 ou 100%) et les autres sous leurs formes désoxyribonucléotides, l'ADN polymérase *FP-I* et des tampons de PCR sur de l'ADN génomique. La chimère ARN/ADN double-brin est ensuite clivée après chaque ribonucléotide et dessalée. Les fragments de clivage sont ensuite analysés par MALDI-TOF MS. La masse moléculaire des fragments de clivage (*flag*) est caractéristique de l'empreinte du *tag* de l'amorce allèle-spécifique est donc du génotype de l'individu étudié.

4.1. La réaction *PAP*

Cette méthode tire avantage de l'utilisation de la réaction de polymérisation activée par pyrophosphorolyse (*PAP*) (Liu & Sommer 2000; Liu & Sommer 2002; Qin et al. 2007). La réaction de *PAP* confère d'une part, une très haute spécificité à la réaction de PCR et d'autre part, améliore considérablement la sélectivité des allèles (Liu & Sommer 2002; Qin et al. 2007).

Les amorces de PCR sont *bloquées* par la modification 2'-PO₄ de l'extrémité 3' et ne peuvent donc pas être élongées par l'ADN polymérase. Lors de l'accroche sur la séquence complémentaire de l'ADN et en présence de pyrophosphate, la dernière base qui contient la modification 2'-PO₄ est enlevée par la réaction de pyrophosphorolyse. L'ADN polymérase *FP-I* est alors en mesure d'ajouter un nouveau nucléotide à la place de la dernière base de l'amorce et de démarrer ainsi la réaction de polymérisation en chaîne. Cette étape primaire d'activation par pyrophosphorolyse est répétée à chaque cycle, permettant ainsi à chaque cycle une très haute allèle-spécificité de la réaction.

4.2. Les amorces

Les premières descriptions de la réaction de *PAP* utilisent des amorces qui contiennent un didésoxynucléotide en 3' terminale et une ADN polymérase thermostable mutante qui incorpore efficacement des didésoxynucléotides. En présence de pyrophosphate, elle élimine efficacement un didésoxynucléotide 3' terminale (Liu & Sommer 2000; Liu & Sommer 2002;

Qin et al. 2007). La réaction de pyrophosphorolyse a une très forte sélectivité pour la complémentarité entre les amorces et l'ADN (Liu & Sommer 2002; Qin et al. 2007). Les amorces contenant des didésoxynucléotides sont synthétisées par voie enzymatique par la terminale transférase (Liu & Sommer 2000) ou synthétisées de 5' vers 3' par la 5'phosphoramidite (Liu & Sommer 2002) car toutes les bases didésoxynucléotides ne sont pas commercialisées.

La découverte des ADN polymérase qui peuvent intégrer efficacement des ribonucléotides a permis d'une part, qu'ils soient incorporés par l'ADN polymérase en laissant le ribose intact et d'autre part, que la substitution de l'amorce 2'-PO₄ arrête toute élongation même si une fraction 3'-OH est présente sur le ribose (Gelfand & Gupta 2008).

Les amorces ayant une extrémité 2'-PO₄ et 3'-OH (Bodepudi et al. 2011) ont pu donc être synthétisées, ce qui a permis d'évaluer la capacité de l'ADN polymérase à incorporer ses amorces *hot start* (Gelfand & Gupta 2008). Les amorces pour la *riboPAP-PCR* sont facilement synthétisées par des synthétiseurs automatiques de l'extrémité 3' vers 5' à l'aide de colonnes CPGs[®] (*Controlled Pore Glass*) commerciales et de 3'phosphoramidite.

Enfin, contrairement au protocole *bi-PAP* (Liu & Sommer 2004; Qin et al. 2007), toutes les amorces ne sont pas allèle-spécifiques. La méthode permet donc d'une part, le génotypage par l'intermédiaire des amorces allèle-spécifiques et d'autre part, le re-séquençage par l'intermédiaire des fragments internes de la *riboPAP-PCR* qui peuvent confirmer les résultats obtenus en génotypage.

Par ailleurs, grâce à l'utilisation de plusieurs amorces allèle-spécifiques contenant des *tags* différents, la *riboPAP-PCR* est une méthode de multiplexage. Les deux SNPs du locus *SCLO1B1* associés au traitement des statines à forte dose ont pu être analysés lors d'une même réaction de multiplexage des deux *riboPAP-PCRs* (Link et al. 2008). Dans une seule et même réaction, le génotypage et le re-séquençage de plusieurs loci peuvent donc être réalisés.

4.3. L'ADN polymérase *FP-1*

L'ADN polymérase *FP-1* incorpore efficacement des NTPs et des dNTPs. Les autres nucléotides coûteux modifiés tel que α -S-CTP, α -S-UTP, ou des nucléotides fluorescents ne sont pas nécessaires.

Dans le protocole *didésoxy-PAP*, la quantité de dNTPs est assez faible augmentant ainsi la probabilité de réincorporer un ddNTP et donc de terminer prématurément l'élongation. Dans

la *riboPAP-PCR*, la quantité dNTP et/ou de NTP est beaucoup plus élevée (200 uM de chacun ou plus) minimisant ainsi la probabilité de réincorporer une terminaison de chaîne nucléosidique tétraphosphate.

Pour la *riboPAP-PCR* du locus *H19*, 90% ATP et 10% dATP sont utilisés et pour le duplex des *riboPAP-PCRs* des loci *SLCO1B1* 90% GTP et 10% dGTP sont utilisés mais la substitution partielle avec le dNTP correspondant n'est pas observée dans les deux cas. L'enzyme *FP-1* contient un domaine 3'-5' exonucléase modulé qui permet d'enlever les mésappariements alors que les autres ADN polymérases utilisées dans les méthodes de génotypage ou de séquençage incorporant les ddNTPs n'en contiennent pas (Schonbrunner et al. 2006). La *haute-fidélité* de l'ADN polymérase *FP-1* augmente considérablement l'efficacité de chaque cycle permettant ainsi de réduire le nombre de cycles.

Enfin, contrairement aux autres ADN polymérases qui contiennent un domaine *Tma* actif, elle est compatible avec l'utilisation de dUTP dans les PCRs permettant d'être utilisé avec dUTP et l'uracile-N-glycosylase (UNG) pour éviter les contaminations (Gelfand 1999).

Par ailleurs, l'ADN polymérase *FP-1* et l'ADN polymérase *KB17* (chapitre 4) donnent les mêmes résultats d'incorporation des NTPs pour les loci étudiés. Les loci *NOS1* rs3741476 et *H19* rs2839698 sont amplifiés par *ribo-PCR* et *riboPAP-PCR* avec ATP tandis que le locus *SCLO1B1* rs4363657 est amplifié par *ribo-PCR* et *riboPAP-PCR* avec GTP. Ces ADN polymérases sont de même nature, elles possèdent la même mutation pour l'incorporation des nucléotides.

La *riboPAP-PCR* et la *ribo-PCR* ont une séquence spécifique pour l'amplification. Les séquences riches en GC sont amplifiables par une *ribo-PCR* ou une *riboPAP-PCR* ATP et/ou UPT alors que les séquences pauvres en GC sont amplifiables par une *ribo-PCR* ou une *riboPAP-PCR* GTP et/ou CTP.

4.4. Les *tags* et les *flags*

Les amorces allèle-spécifiques de la *riboPAP-PCR* contiennent une séquence répétitive à leur extrémité 5' qui permet d'attribuer une masse, *flag*, différente pour chaque allèle dans la région des 6 ou 7-mer. En effet, les fragments de clivage sont analysés par MALDI-TOF MS en mode linéaire. Certains fragments peuvent avoir une masse identique ou une différence de masse entre deux fragments qui ne peut pas être détectée due à un trop faible pouvoir de résolutif de l'instrument.

Le *flag* améliore l'identification de l'allèle par MALDI-TOF MS en sélectionnant la séquence du fragment contenant le SNP à étudier. Sa détection est également améliorée par l'augmentation de sa quantité par rapport aux autres fragments de clivage par l'intermédiaire des trois séquences répétitives du *tag* des amorces. Par ailleurs, afin d'assurer une meilleure détection des *flags* par MALDI-TOF MS, il vaut mieux utiliser des *flags* contenant plusieurs thymines. Dans le cas de l'étude du duplex des loci de *SCLO1B1*, plusieurs *tags* et associations de *tags* peuvent être utilisés.

4.5. L'avantage de la méthode

Cette méthode constitue une autre alternative à l'arsenal sans cesse croissant des méthodes d'analyses de l'ADN. Le clivage chimique est rapide, facile et totale. Des enzymes de coupures de bases spécifiques, coûteuses, ne sont pas nécessaires.

Son principal avantage par rapport aux autres méthodes utilisant la spectrométrie de masse réside dans sa capacité de déterminer le multiplex génotypage par l'analyse de plusieurs *riboPAP-PCR* de différents loci simultanément dans une seule réaction.

De plus, l'utilisation d'une masse spécifique pour déterminer le génotype de l'individu permet de s'affranchir des limites du pouvoir résolutif du MALDI-TOF MS.

Elle permet dans une même réaction le multiplex allèle-spécifique génotypage et le re-séquençage de l'ADN. Elle est également facile d'utilisation, rapide et avec un coût probablement inférieur à celui des méthodes actuelles.

5. Conclusion

Dans ce chapitre (annexe 4 et 5), une nouvelle méthode de multiplex allèle-spécifique et re-séquençage de l'ADN utilisant le clivage chimique d'une chimère ARN/ADN double-brin et l'analyse des fragments par MALDI-TOF a été développée (D. Gelfand et al. 2009). Cette méthode est spécifique d'un allèle grâce à une polymérisation activée par pyrophosphorolyse (*PAP*). Une chimère ARN/ADN double-brin est générée lors d'une *riboPAP-PCR* à partir d'un ADN génomique, des amorces sens allèle-spécifiques, d'une amorce anti-sens, un seul nucléotide sous sa forme ribonucléotide (100 ou 90 %) et les autres sous leurs formes désoxynucléotides et une ADN polymérase *FP-I*. La chimère ARN/ADN est ensuite clivée par l'hydroxyde de sodium. Les fragments de clivage sont ensuite dessalés par des billes échangeuses de cations et les fragments sont analysés par MALDI-TOF MS. Les amorces contiennent un groupement 2'-PO₄ à l'extrémité 3' et les amorces allèle-spécifiques

contiennent une séquence répétitive et spécifique en 5' (*tag*). La séquence complémentaire du *tag* donne après fragmentation une séquence (*flag*) et une masse unique. Cette masse identifie la présence de l'amorce allèle-spécifique utilisée durant la *riboPAP-PCR* et, par conséquent, de l'allèle présent. Le *flag* doit être suffisamment concentré, contenir que quelques bases ainsi que plusieurs bases thymines afin d'être facilement détectable par MALDI-TOF MS. D'autres *tags* pourront être également utilisés lors de prochaines études.

Cette méthode a permis d'analyser 30 individus pour le génotypage et le re-séquençage des loci *NOS1* rs3741476 et *H19* rs2839698 et 95 individus pour le génotypage et re-séquençage en duplex des loci *SLCO1B1* rs4363657 et rs4149056. Les génotypes des individus pour les loci *NOS1* rs3741476, *H19* rs2839698 et *SLCO1B1* rs4363657 sont identiques à ceux obtenus par la méthode du chapitre 4. Par ailleurs, l'ADN polymérase *FP-1* possède les mêmes séquences préférentielles d'amplification que l'ADN polymérase *KB17*.

Cette méthode identifie facilement les positions homozygotes et hétérozygotes. Plusieurs allèles peuvent être analysés spécifiquement et simultanément. Cette méthode permet d'analyser plusieurs séquences dans une seule réaction, il constitue probablement le plus simple protocole de multiplex allèle-spécifique génotypage et de re-séquençage de l'ADN.

Cette chimère ARN/ADN double-brin pourrait également avoir d'autres usages que celui présenté ici. Elle pourrait être utilisée, par exemple, pour l'étude des mutations somatiques du codon 12 du gène *k-ras*.

**Conclusion générale
et
perspectives**

Conclusion générale et perspectives

Le travail de thèse présenté ici a eu pour but le développement de méthodes d'analyse de l'ADN par clivage d'une chimère ARN/ADN, par spectrométrie de masse MALDI-TOF MS et l'analyse de la chimère ARN/ADN par spectrométrie de masse en tandem (chapitre 1). Un arsenal de méthodes a été mis au point lors de ce travail : une méthode de multiplex microhaplotypage (chapitre 2), une méthode d'analyse de la méthylation de l'ADN (chapitre 3), une méthode de multiplex re-séquençage par *ribo-PCR* (chapitre 4 et annexe 3) et une méthode de multiplex allèle-spécifique génotypage par *riboPAP-PCR* (chapitre 5, annexes 4 et 5).

Conclusion générale

Concept de la méthode

Cette méthode repose sur la synthèse d'une chimère ARN/ADN par une nouvelle classe d'ADN polymérases qui ont été développées par Roche Molecular Systems.

Les enzymes *G46E CS6R* et *KB17* permettent la synthèse d'une chimère ARN/ADN simple-brin. L'enzyme *G46E CS6R* est utilisée dans la méthode de génotypage (Mauger et al. 2006) et celle de microhaplotypage (chapitre 2) tandis que l'enzyme *KB17* est utilisée dans la méthode de re-séquençage (Mauger et al. 2007) et celle d'analyse de la méthylation de l'ADN (chapitre 3). L'ADN polymérase *G46E CS6R* synthétise une chimère ARN/ADN simple-brin à partir d'une PCR, d'une ou plusieurs amorces d'élongation et de tous les NTPs. Plusieurs chimères peuvent être synthétisées dans la même réaction à l'aide d'amorces différentes permettant ainsi un multiplexage. L'ADN polymérase *KB17* génère une chimère ARN/ADN simple-brin à partir d'une PCR purifiée, d'une amorce d'élongation, des NTPs (A, C ou G) et des trois bases complémentaires dNTPs. Elle peut contenir jusqu'à environ 500 pb. Plusieurs chimères ARN/ADN simple-brin peuvent être synthétisées parallèlement afin d'obtenir des informations complémentaires de la séquence.

Par ailleurs, les ADN polymérases *KB17* et *FP-1* permettent la synthèse de chimères ARN/ADN double-brins. L'enzyme *KB17* est utilisée dans la méthode de re-séquençage (Mauger et al. 2012) alors que l'enzyme *FP-1* est utilisée dans celle d'allèle-spécifique

génotypage (chapitre 5). La base du ribonucléotide est différente de celle des trois autres bases désoxynucléotides. Une séquence spécifique d'amplification a été déterminée pour la synthèse de la chimère ARN/ADN double-brin. Les séquences riches en GC sont préférentiellement amplifiées par ATP et/ou UTP tandis que les séquences pauvres en GC sont préférentiellement amplifiées par CTP et/ou GTP. La base d'incorporation du ribonucléotide ne doit pas également être présente plus de quatre fois successives dans la séquence du brin synthétisé. L'ADN polymérase *KB17* permet de synthétiser une chimère ARN/ADN double-brin, d'environ 120 pb, à partir d'un NTP (A, C, G ou U) et des 3 autres bases complémentaires dNTPs dans une réaction de *ribo-PCR*. Plusieurs séquences peuvent être analysées dans une même réaction de *ribo-PCR* par l'intermédiaire de plusieurs amorces permettant ainsi un multiplexage. Enfin, l'ADN polymérase *FP-1* génère une chimère ARN/ADN double-brin, d'environ 80 pb, à partir d'un NTPs (A ou G) et des 3 autres bases complémentaires dNTPs dans une réaction de *riboPAP-PCR*. Elle est compatible avec l'utilisation d'amorces contenant un groupement 2'-PO₄H₂ sur l'extrémité 3' dans une réaction de polymérisation activée par pyrophosphorolyse. La réaction *PAP* sur les amorces permet d'augmenter la spécificité de la réaction de génotypage (chapitre 5, annexe 4 et 5). Plusieurs chimères ARN/ADN double-brins peuvent être synthétisées dans la même réaction à l'aide d'amorces différentes permettant ainsi un multiplexage.

La chimère ARN/ADN est ensuite clivée après chaque ribonucléotide par l'hydroxyde de sodium. Les fragments de clivage sont constitués de désoxynucléotides et se terminent par un ribonucléotide à l'extrémité 3' contenant un groupement 3'-PO₄H₂. La réaction est totale, rapide et peu coûteuse contrairement au clivage enzymatique.

Un logiciel de calcul de masse moyenne a été élaboré par Jérémy Semhoun afin de calculer les masses de la séquence de référence des chimères ARN/ADN. Il est également adapté aux contraintes de chaque méthode.

Les fragments de clivage sont ensuite dessalés par une résine échangeuse de cations. Elle permet d'échanger les ions sodium présents dans la solution avec les protons présents sur les billes. Les protons servent ensuite à neutraliser les ions hydroxyde en diminuant le pH de l'échantillon afin de permettre une meilleure cristallisation de la matrice.

Les fragments de clivage sont analysés par MALDI-TOF MS en mode négatif car les chimères ARN/ADN sont naturellement chargées négativement au niveau de l'atome d'oxygène du groupement phosphate. La spectrométrie de masse MALDI-TOF a été utilisée

car elle est rapide, simple à mettre en œuvre et sa capacité haut-débit permet l'analyse d'un grand nombre d'échantillons. La matrice HCCA a été utilisée pour l'analyse des fragments de clivage de la chimère ARN/ADN synthétisée par l'enzyme *G46E CS6R* tandis que la matrice THAP a été utilisée dans les autres méthodes. La matrice HCCA a une cristallisation homogène mais elle est adaptée à l'étude des oligonucléotides de 3 à 5-mer tandis que la matrice THAP est plus adéquate à l'étude des oligonucléotides jusqu'à environ 10000 Da. De plus, une cible composée d'une partie hydrophobe et hydrophile a été utilisée avec la matrice THAP afin de concentrer les échantillons. Les analyses ont été réalisées en mode linéaire afin de pouvoir détecter les ions jusqu'à 10000 Da. Le faible pouvoir résolutif du mode linéaire est suffisant dans la plupart des cas grâce à la distribution des masses des fragments de clivage et à l'utilisation des informations complémentaires. En effet, plusieurs chimères ARN/ADN simple-brins (A, C ou G) doivent être synthétisées en parallèle tandis que pour les brins sens et anti-sens de la chimère ARN/ADN double-brin donnent les informations complémentaires dans une seule réaction. Le mode réflecteur peut être également utilisé pour l'analyse de fragments de clivage de masses proches ne pouvant pas être détectés par le pouvoir résolutif du mode linéaire. Cependant, elle est limitée à l'étude des ions inférieurs à 4000 Da.

Les masses des fragments de clivage sont comparées à celles de la séquence de référence. Une déviation par rapport à la séquence de référence est caractéristique d'une modification dans la séquence de l'ADN. Cette modification peut être une mutation, un SNP, une délétion, une insertion ou des cytosines ou des 5-méthylcytosines des dinucléotides CpGs. Cependant, les fragments sans thymine ne sont pas détectés dans ces conditions d'analyse ainsi que les fragments de grande masse. Le pouvoir résolutif de l'instrument ne permet pas de distinguer des différences de masse de l'ordre d'une base pour les grands fragments. Cependant, l'utilisation des informations des chimères ARN/ADN simple-brins complémentaires ou des chimères double-brins peuvent pallier à ce phénomène.

L'analyse de la chimère ARN/ADN par spectrométrie de masse en tandem

Les fragmentations en mode CID des chimères de 4, 5, 7 et 10-mer ont été analysées par MALDI-TOF/TOF MS tandis que les fragmentations métastables (PSD) et en mode CID de la chimère GCTA ont été étudiées par spectrométrie de masse MALDI-TOF-TOF et l'ESI-IT (chapitre 1). L'analyse des chimères ARN/ADN en phase gazeuse est une approche rapide de séquençage. Cependant, les chimères sans thymine doivent être analysées à des concentrations

plus élevées ainsi que les chimères qui contiennent un grand nombre de nucléotides. Cette approche est limitée au séquençage des chimères de 10-mer.

Les fragmentations en mode CID, de haute énergie, des chimères de 4 et 5-mer par MALDI-TOF-TOF donnent un grand nombre d'ions fragments : les 8 séries d'ions du clivage de la chaîne nucléotidique au niveau des 4 liaisons du groupement phosphodiester, les ions issus de la perte de la base et des ions de clivage interne. L'ion a_4^- est la seule fragmentation caractéristique des chimères ARN/ADN de 4-mer dont la base du ribonucléotide est une cytosine. Les ions les plus abondants ne sont pas les ions d_n^- comme dans la littérature mais des ions w_n^- , $[a_n-B_nH]^-$, $[M-H-B_nH]^-$ comme par ESI-ITMSⁿ. Plus le nombre de nucléotides de la chimère augmente, plus les ions w_n^- , $[a_n-B_nH]^-$, $[M-H-B_nH]^-$ et les ions internes de ces ions deviennent abondants et spécifique de la séquence. Ce phénomène peut être expliqué par la formation d'ion zwitterion qui favorise la déprotonation des groupements phosphates tout en maintenant la protonation des bases.

Les fragmentations en mode CID de la chimère GCTA par MALDI-TOF/TOF MS sont essentiellement des fragmentations métastables. Les fragmentations en mode CID de la chimère GCTA par ESI-ITMSⁿ sont similaires à celles de la littérature. Elle donne des ions w_n^- , $[a_n-B_nH]^-$, $[M-H-B_nH]^-$ abondants. Grâce à l'analyse jusqu'à l'expérience MS⁵ séquentielle, les séquences des ions et la localisation des charges des ions multichargés ont pu être déterminées.

La méthode de multiplex microhaplotypage

La méthode de multiplex microhaplotypage du chapitre 2 est basée sur la méthode de génotypage des SNPs (Mauger et al. 2006) mais elle diffère par l'utilisation de plusieurs amorces permettant ainsi le multiplexage de la méthode et l'étude d'une région hypervariable. La méthode de microhaplotypage a été mise au point pour l'analyse des microhaplotypes des gènes *HLA-DRB1* et *HLA-A*. Elle a permis de déterminer le microhaplotype ASTW du locus *HLA-DRB1-227* de 23 individus et ceux en duplex des microhaplotypes HCACH du locus *HLA-A-98* et MGAR du locus *HLA-A-453* de 12 individus. Plusieurs microhaplotypes doivent être sélectionnés afin de pouvoir déterminer les haplotypes.

Enfin dans cette méthode, seuls les microhaplotypes connus peuvent être détectés. Afin d'analyser les nouveaux haplotypes référencés dans la base de données IMGT-HLA, d'autres

amorces de PCR et d'élongation identifiant les nouveaux microhaplotypes doivent être ajoutées à cette méthode.

La méthode d'analyse de la méthylation de l'ADN

La méthode d'analyse de la méthylation de l'ADN est basée sur celle du re-séquençage de l'ADN mitochondrial (Mauger et al. 2007) mais elle diffère par l'étude du génome humain qui est beaucoup plus complexe que celui de l'ADN mitochondrial. Elle a permis le re-séquençage et l'analyse des cytosines et des 5-méthylcytosines des îlots CpGs des loci des gènes *CDKN2A* et *RASSF1A* de 7 individus (chapitre 3). Cette méthode nécessite une étape préalable indispensable à la différenciation des cytosines, des 5-méthylcytosines. La réaction de conversion au bisulfite de sodium transforme les cytosines en uraciles mais pas les 5-méthylcytosines. Cette méthode possède donc le plus long protocole des méthodes présentées ici.

Cette méthode a permis de quantifier la méthylation de l'ADN par l'intermédiaire d'une courbe d'étalonnage réalisée à partir de l'intensité des pics des ions de la cytosine et de la 5-méthylcytosine du CpG. Les 2 courbes d'étalonnages sont des polynomiales du troisième degré qui sont dues à un problème de biais de PCR. Afin d'améliorer la linéarité des courbes d'étalonnage, des optimisations de PCRs doivent être réalisées.

Enfin, l'identification des îlots CpGs hypométhylés ou hyperméthylés par cette méthode et l'association avec les données biologiques des gènes étudiés a donc permis la cartographie de régions importantes à étudier pour certains individus. De plus, cette méthode est également une méthode de re-séquençage permettant ainsi l'identification des modifications de la séquence d'ADN.

La méthode de re-séquençage par *ribo-PCR*

Cette méthode (Mauger et al. 2012) a permis le re-séquençage de 30 individus pour les loci *NOS1* rs3741476 et *H19* rs2839698 et 95 individus pour le locus *SLCO1B1* rs4363657 (chapitre 4 et annexe 3). L'utilisation d'une chimère ARN/ADN double-brin engendre la réduction du nombre d'étapes de la méthode et la complémentarité de l'information portées

par les deux brins. Le multiplexage de la méthode est possible bien que l'analyse de multiples séquences dans une même réaction devient alors difficile. Elle est probablement la plus simple des méthodes de séquençage.

Cette méthode de re-séquençage permet la détection des polymorphismes connus, inconnus, rares ou fréquents. Par contre, elle est limitée d'une part, à une séquence d'amplification spécifique et d'autre part, à l'étude des séquences constituées d'une 100 pb. Toutefois, des optimisations peuvent augmenter le nombre de base de la chimère ARN/ADN double-brin.

La méthode de multiplex allèle-spécifique génotypage par *riboPAP-PCR*

Cette méthode a permis le multiplex allèle-spécifique génotypage et le re-séquençage de 30 individus pour les loci *NOS1* rs3741476 et *H19* rs2839698 et de 95 individus duplex des loci *SLCO1B1* rs4363657 et rs4149056 (chapitre 5, annexe 4 et 5). Les résultats des génotypes sont les mêmes que ceux obtenus avec la méthode de re-séquençage par *ribo-PCR* (Mauger et al. 2012).

Elle est basée sur l'utilisation d'amorces spécifiques de l'allèle étudié, permettant ainsi le multiplexage. Elles contiennent à leur extrémité 5' une séquence répétitive spécifique (*tag*) de chaque allèle et possède à leur extrémité 3' un groupement 2'-PO₄. Ce groupement permet d'augmenter la spécificité de la réaction à l'aide d'une réaction de polymérisation activée par pyrophosphorolyse (*PAP*). Les *tags* permettent de choisir les séquences et les masses étudiées afin de s'affranchir du pouvoir résolutif de l'instrument. L'analyse de la masse moléculaire du fragment complémentaire du *tag* (*flag*) permet de déterminer directement le génotype de l'individu. Le *flag* doit être suffisamment concentré, contenir des bases thymines et ne contenir que quelques nucléotides afin d'être facilement détectable.

Enfin, cette méthode pourrait également être utilisée pour l'étude des mutations somatiques du codon 12 du gène *k-ras* en utilisant différentes amorces.

Perspectives

Ces méthodes d'analyse de l'ADN par clivage d'une chimère ARN/ADN et par spectrométrie de masse MALDI-TOF MS permettent de détecter des modifications de la séquence de

l'ADN mais aussi des modifications épigénétiques. Ces méthodes mettent en évidence des polymorphismes fréquents, rares, connus et inconnus. La méthode d'analyse de la méthylation de l'ADN pourrait également servir à étudier le méthylome.

Ce concept pourrait être utilisé dans les laboratoires possédant déjà un spectromètre de masse MALDI-TOF afin d'analyser un grand nombre d'échantillons et pour l'étude d'une séquence spécifique ou de problème spécifique. Les principaux avantages de cette méthode par rapport aux autres méthodes d'analyse de l'ADN par spectrométrie de masse est son clivage chimique et l'utilisation d'une chimère ARN/ADN double-brin qui diminue le nombre d'étapes de la méthode et par conséquent son coût. Contrairement à la méthode du *GOODASSAY*, elle permet le re-séquençage, est donc la détection de polymorphismes inconnus. Cette méthode est en concurrence directe avec celle de *MassARRAY*[®] de Sequenom car elle possède les mêmes approches.

Les méthodes développées dans ce travail doivent être optimisées : le problème de biais de PCR pour l'analyse de la méthylation de l'ADN, l'augmentation du nombre de paires de bases de la chimère ADN/ADN double-brin, la détection des mutations somatiques... De plus, d'autres applications que celles présentées ici peuvent être faites.

Afin d'analyser un grand nombre d'individus et de séquences, un logiciel d'analyse doit être mis au point. Il est en cours de réalisation par Vincent Picaud du CEA/LIST. Ce logiciel devra intégrer toutes les méthodes, le calcul des masses des fragments de clivage, l'analyse des spectres de masse, la recherche des modifications de l'ADN et la détermination des modifications de la séquence de l'ADN. Il permettra l'automatisation à haut-débit de ce concept qui est nécessaire pour son développement à grande échelle.

Les méthodes de séquençage de l'ADN, de seconde génération, par spectrométrie de masse, sont complémentaires aux méthodes classiques d'analyse de l'ADN. L'avantage de ces méthodes est la détection directe des séquences d'ADN par l'intermédiaire de l'analyse de la masse et la capacité haut-débit de MALDI-TOF MS qui permet l'étude d'un grand nombre d'échantillons en série. Cependant, les progrès de la biologie moléculaire et en particulier de méthodes de séquençage prometteuses comme le séquençage par nanopore rendent les méthodes par spectrométrie de masse moins compétitives.

Il est difficile aujourd'hui de prédire quelle méthode sera utilisée dans les années à venir. Les méthodes de séquençage visent à pouvoir analyser directement le génome en entier à partir d'une simple molécule d'ADN. En outre, la combinaison de l'avancée des méthodes de biologie moléculaire et du domaine de la bioinformatique permettra sûrement une étude

rapide et fiable du génome afin de déterminer sans ambiguïté les modifications de l'ADN des gènes responsables de toutes les maladies.

ANNEXES

Annexe 1: SNP genotyping using alkali cleavage of RNA/DNA chimeras and MALDI time-of-flight mass spectrometry

Nucleic Acids Research, 2006, Vol. 34, No. 3 Published online 2006 February 9. doi: e18

doi:10.1093/nar/gnj021

Florence Mauger, Olivier Jaunay, Valérie Chamblain, Fred Reichert¹, Keith Bauer¹, Ivo G. Gut* and David H. Gelfand¹

Centre National de Génotypage, Bâtiment G2, 2 Rue Gaston Crémieux, 91057 Evry Cedex, France and ¹Roche Molecular Systems Inc., 1145 Atlantic Avenue, Alameda, CA 94501, USA

Received November 21, 2005; Revised and Accepted January 19, 2006

Abstract

Single nucleotide polymorphisms (SNPs) are now widely used for many DNA analysis applications such as linkage disequilibrium mapping, pharmacogenomics and traceability. Many methods for SNP genotyping exist with diverse strategies for allele-distinction. Mass spectrometers are used most commonly in conjunction with primer extension procedures with allele-specific termination. Here we present a novel concept for allele-preparation for SNP genotyping. Primer extension is carried out with an extension primer positioned immediately upstream of the SNP that is to be genotyped, a complete set of four ribonucleotides and a ribonucleotide incorporating DNA polymerase. The allele-extension products are then treated with alkali, which results in the cleavage immediately after the first added ribonucleotide. In addition, to obtain fragments easily detectable by mass spectrometry, we have included a ribonucleotide in the primer usually at the fourth nucleotide from the 3' terminus. The method was tested on four SNPs each with a different combination of nucleotides. The advantage over other mass spectrometry-based SNP genotyping assays is that this one only requires a PCR, a primer extension reaction with a universal extension mix and an inexpensive facile cleavage reaction, which makes it overall very cost effective and easy in handling.

Introduction

Single nucleotide polymorphisms (SNPs) are thought to be the ideal markers for the dissection of complex traits in association studies and linkage disequilibrium mapping (1). Furthermore, SNP analysis will quite likely become a powerful tool for pharmacogenetics (2) and traceability in the agro-alimentary sector.

Many methods for SNP genotyping with detection by DNA microarrays, electrophoresis, mass spectrometry and plate-readers have been introduced (3). Particularly for mass spectrometry-based SNP assays, matrix-assisted laser desorption/ionization mass spectrometry (MALDI) is the method of choice due to its reliability, accuracy and speed (4). DNA analysis by mass spectrometry was reviewed extensively and for a general overview we refer the reader to

publications cited therein (5,6). In order to analyze a SNP with a mass spectrometer, allele-specific products have to be prepared. Primer extension is the most commonly used format in conjunction with mass spectrometric detection. A primer is annealed upstream of a SNP and a combination of dNTP and/or ddNTP is used to synthesize an allele-specific product with the aid of a DNA polymerase. Two formats are used. One is multiple-nucleotide primer extension, such as in the Mass Extend assay (7); www.sequenom.com. The other is single-base primer extension, such as the GOOD assay (8). For the GOOD assay, single-base primer extension suffices because by including a cleavage step into its protocol the product mass is moved into a mass range where resolution of the mass spectrometer allows distinguishing even the smallest mass difference

between the two alleles of a SNP (the smallest difference is 9 Da between thymine and adenine). However, the additional processing and enzymatic cleavage step requires the use of a 5'phosphodiesterase, which is a fairly expensive and somewhat temperamental enzyme (cleavage of cytidine is less efficient). Recently a method for the removal of primers via alkali induced RNA cleavage of chimeric primers was presented (9).

Here we present a novel concept for allele-preparation for SNP genotyping combined with MALDI mass spectrometry for detection. Rather than extending a primer immediately upstream of the polymorphism to be genotyped with deoxynucleotides (dNTP) and dideoxynucleotides (ddNTP), four ribonucleotides (NTP) are used in conjunction with a modified thermostable DNA polymerase that readily incorporates NTPs. Ribonucleotide incorporating DNA polymerases that manage to incorporate a limited percentage of ribonucleotides have been described in the literature (10). However, complete transcription by this mechanism is novel. Further, we have included a ribose sugar at the fourth nucleotide from the 3' end of the primer. After the ribonucleotide primer extension reaction, cleavage is carried out using alkali. This results in the cleavage 3' of the first added ribonucleotide and 3' of the ribonucleotide in the primer leaving a product that carries the allele-information (Figure 1). We have applied this allele-distinction procedure with two conditioning procedures: the first uses a regular DNA backbone and ion exchange desalting for conditioning, while the second one uses phosphorothioate bridges in the primer that are charge-neutralized by alkylation and analyzed without desalting as is done in the GOOD assay. The method has been tested on four SNPs each with different combinations of nucleotides. Owing to the 1 Da mass difference between C and U, distinction between the two alleles of a Y SNP is achieved by replacing either cytidine by 5-methyl-cytidine or uridine by 5-methyl-uridine. This separates the two alleles by an easily measurable mass difference.

Materials and Methods

Sixteen DNA samples from the CEPH collection (www.cephb.fr) were used for this work. They were selected based on representing all possible genotypes

with the four SNPs used here. All samples were genotyped for these four SNPs by other methods for reference. Primers for PCR were synthesized by Biotex (Berlin, Germany) or MWG (Ebersberg, Germany), extension primers were from Eurogentec (Liège, Belgium) and PAGE purified. Sequences of the primers used in the four systems are listed in Table 1. SNPs with different nucleotide combinations were selected. Deoxynucleoside triphosphates, dNTPs, (N = A, C, G, T) and ribonucleoside triphosphates, NTPs, (N = A, C, C^{Me}, G, U, U^{Me}) were purchased from Roche Diagnostics (Mannheim, Germany) and Trilink Biotechnologies (San Diego). Platinum™ *Taq* HiFi DNA polymerase high fidelity was obtained from Invitrogen (Karlsruhe, Germany). General chemical reagents and α -cyano-4-hydroxycinnamic acid were purchased from Aldrich (Steinheim, Germany). G46E CS6R DNA Polymerase was provided by Roche Molecular Systems (Alameda). This Designer DNA Polymerase is a chimeric DNA polymerase comprising the 5'-nuclease domain of *Thermus* sp. Z05 DNA polymerase and the 3'-nuclease and 5'-3' DNA polymerase domains of *Thermotoga maritima* (*Tma*) DNA polymerase. The 5'-nuclease activity is eliminated in this enzyme by the introduced G46E mutation. The wild-type *Tma* proof reading activity has been eliminated by mutation of 'Motif I' of the 3'-5' exonuclease domain. Finally, the enzyme contains an E678G mutation that eliminates the selectivity of the wild-type enzyme against incorporation of C'-2-substituted nucleotides and which facilitates efficient incorporation of ribonucleoside triphosphates ['R' designation (11), Research samples of G46E CS6R DNA Polymerase may be obtained from Dr Thomas Myers (thomas.myers@roche.com), Director, Program in Core Research, Roche Molecular Systems, Inc., 1145 Atlantic Avenue, Alameda, CA 95401, USA]. Ion exchange resin AG 50W-X8 H⁺ was purchased from BioRad (Marmes la Coquette, France). Thermocycling procedures were carried out in Eppendorf Gradient Thermocyclers (Eppendorf, Germany).

PCR amplifications were in 5 μ l volumes with 10 ng of genomic DNA, 3.75 pmol of each primer, 0.5 U of DNA polymerase (Platinum HiFi *Taq* DNA polymerase), 1 mM MgSO₄, 0.1 mM dNTP (each),

0.5 μ l of 10 \times buffer [600 mM Tris-SO₄, pH 8.9, and 180 mM (NH₄)₂SO₄]. Thermal cycling for PCR is carried out with the following profile: 4 min at 94°C, followed by 35 cycles of 30 s at 94°C, 45 s at 60°C and 30 s at 72°C. This is concluded with 4 min at 72°C for final extension.

The cycled extension reaction follows without isolation of the PCR product or degradation of residual PCR primers or dNTPs. For a final volume of 10 μ l, 0.25 μ l of G46E CS6R DNA polymerase (4 U/ μ l), 1 μ l of NTP (4 mM each), 0.5 μ l of 200 mM Tris-HCl, pH 8.3, 0.5 μ l of 500 mM NH₄Cl, 0.5 μ l of 50 mM MgCl₂, 0.25 μ l of extension primer (40 pmol/ μ l) and 2 μ l of water are added to the amplification reactions. The composition of the chosen NTP mix is a function of the SNP that is to be genotyped. In the case of a Y SNP, the two alleles are not resolvable by mass as C and U are different in mass by only 1 Da. Accordingly, by using a mix of AC^{Me}GU triphosphate or ACGU^{Me} triphosphate, Y SNPs are readily assigned. The thermal cycling profile for the extension reaction is 2 min at 95°C followed by 40 cycles of 15 s at 95°C, 2 min at 65°C and 20 s at 72°C with a final extension of 4 min at 72°C.

For alkali cleavage, 1 μ l of 3.3 M KOH or 3.3 M NaOH is added for a final concentration of 0.3 M and incubated at 70°C for 1.5 h. Concentrated NH₄OH (40 μ l of 25 M) was also evaluated. For this, the incubation had to be carried out at 55°C overnight and the solvent removed in the SpeedVac at the end.

Following alkali treatment, two different methods for sample conditioning were used. The first was desalting the samples, which was done in the case of the phosphate backbone fragments. The second was alkylation in the case of phosphorothioate backbone fragments. Desalting is carried out by the addition of cation exchange resin charged with H⁺. Four microliters of water were added for a final volume of 15 μ l. One-third of the total reaction volume is added in resin and incubated for 20 min at room temperature under agitation. Thereafter, the sample was centrifuged for 30 s at 1000 r.p.m. 134 \times g to sediment the resin. Five microliters of the supernatant was mixed with 7.5 μ l of 60% acetonitrile and 2.5 μ l of saturated α -cyano-4-hydroxycinnamic acid in acetone. One microliter was transferred to a nickel-coated aluminium MALDI target with 384 positions and dried at room temperature. In the case of cleavage with NH₄OH, the residual was dissolved in 10 μ l of water and mixed with the matrix without treatment with ion exchange resin.

For the alkylation of phosphorothioate backbone products 1 μ l of 3.3 M HCl was added to neutralize the alkali used for cleavage. Alkylation was achieved by adding 28 μ l of a mixture of 12 μ l of acetonitrile,

0.5 μ l of 2 M triethylamine (TEA, pH 7.5 with CO₂), 2.5 μ l 2 M Tris-HCl, pH 8.0, and 6.0 μ l of methyl iodide. After incubation at 40°C for 30 min, 10 μ l of water were added. A biphasic system formed upon standing. Five microliters were sampled from the upper layer and added to 7.5 μ l of 60% acetonitrile and 2.5 μ l of a saturated solution of α -cyano-4-hydroxy-cinnamic acid in acetone. One microliter was transferred onto the MALDI target and dried.

The target is introduced into the MALDI mass spectrometer (Bruker, Reflex III or Bruker Autoflex, Bremen, Germany) for analysis. Analysis is carried out in negative ion mode with an acceleration voltage of +18–20 kV using pulsed ion extraction with a delay of 200 ns. Mass spectra were recorded in linear mode with external calibration.

Results

The principle of the SNP genotyping procedure developed here is illustrated in [Figure 1](#). A PCR is used to prepare a template of sufficient amount and reduced complexity for the following extension step. It was not necessary to remove residual primers and dNTPs from the PCR prior to the extension reaction. Extension products with deoxynucleotides were never observed under these extension conditions. Further, if residual PCR primers were extended, these products would be outside the observed mass ranges.

A primer is annealed immediately upstream of the SNP that is to be genotyped. Primer extension is carried out with a reaction mix containing four ribonucleotides and a thermostable, ribonucleotide incorporating chimeric DNA polymerase. This DNA polymerase was derived from *Thermus* sp. Z05 and *T.maritima* DNA polymerases. It has no proof reading activity and no 5'-nuclease activity. The primer extension requires transcription of the SNP sequence and incorporation of at least one additional ribonucleotide. The product is chimeric DNA/RNA where the DNA part is the primer and the RNA part is neosynthetic.

Allele-specific products for the mass spectrometric analysis are generated by alkali cleavage. This results in the cleavage 3' of the first added ribonucleotide. Under these conditioning and analysis methods, a residual 2' or 3' phosphate group remains at the 3' terminal nucleotide. From the mass we cannot distinguish 2'-HPO₄ + 3'-OH and 2'-OH + 3'-HPO₄.

Quite likely products are a mixture of these two species. Products corresponding to a termination of the extension reaction after addition of the first ribonucleotide, which would manifest themselves by the lack of a terminal phosphate group, were never observed. Resolution of the MALDI mass spectrometer is best in the low mass range and not sufficient in the mass range where 20mers lie to distinguish between two allele-products generated by the outlined procedure. By including a ribonucleotide at the fourth position from the 3' end of the extension primers, the mass of the products is shifted into the mass range of 1000–1500 Da where baseline resolution of products separated by as little as 5 Da is possible. The application of this procedure of SNP genotyping is shown in [Figure 2](#). Different conditions for cleavage were investigated (NaOH, KOH, NH₄OH). A problem arising from the substantial amount of Na⁺ or K⁺ added is the difficulty to remove it again prior to mass spectrometric analysis. A limited degree of bleed through was observed ([Figure 2](#)). However, as adducts appear at defined positions relative to the real allele-masses in mass spectra possible confusion of an allele-peak and an adduct peak can be avoided by the choice of the alkali. We carried out this genotyping procedure with SNPs with different base composition ([Figures 3](#) and [and4](#)). No cases of premature termination, preferential incorporation of bases or easier cleavage due to base context was observed. In no instances were the 3-base products of cleaved non-extended primers observed. This might be due to efficient primer extension or low efficiency of ionization of 3mers with two phosphate groups compared with 4mers with three interbase phosphate groups and a terminal phosphate group. Under the cycled extension conditions used one would assume that the extension primer was completely consumed.

Moving from deoxyribonucleotides to ribonucleotides results in an only 1 Da mass difference for a Y SNP due to the mass difference between cytidine and uridine rather than cytidine and thymidine. To resolve the two alleles of a Y SNP, we explored the possibility of replacing either cytidine by 5-methyl-cytidine or uridine by 5-methyl-uridine. Both 5-methyl-cytidine triphosphate and 5-methyl-uridine triphosphate are incorporated with nearly the same efficiency as cytidine triphosphate and uridine

triphosphate by G46E CS6R DNA polymerase ([Figure 4](#)).

In the past, two major strategies for sample conditioning were developed in order to master the mass spectrometers inherent sensitivity to impurities. Owing to the negative charges on the sugar phosphate backbone, adducts of the phosphate groups with Na⁺ and K⁺ are observed when samples are not freed of these components. The most common method for desalting reactions is by the addition of an ion exchange resin. Alternatively, the negative backbone charges can be removed chemically by an alkylation reaction as is done in the GOOD assay (8). This requires the replacement of phosphate groups of the backbone by phosphorothioate groups. As shown here, a negative ion mode variant of the GOOD assay can be used where a single negative charge is introduced with the terminal nucleotide in analogy to what was shown by Sauer and Gut (12). Both methods for sample conditioning were implemented here. Genotyping results with desalting are shown in [Figures 2](#) and [and3a,3a](#), while [Figures 3b](#) and [and4](#) show results using an extension primer containing phosphorothioates and alkylation (no desalting). Phosphorothioate is susceptible to oxidation under harsh alkali conditions. The conditions for the cleavage reaction were selected to cause minimal oxidation. However, peaks with -30 Da (-SCH₃ + OH) relative to the allele-peaks were observed ([Figures 3b](#) and [and4](#)). Again these peaks appeared at defined masses and are thus easy to identify.

The four SNPs shown here were selected from SNPs that had been genotyped previously at the CNG by the GOOD assay ([Table 1](#)). Repetition of the genotyping of all four SNPs under the optimized conditions provided in Materials and Methods always resulted in the correct genotypes relative to the control genotyping. The DNA samples of the CEPH collection were selected to be able to verify genotypes also based on the transmission of alleles from one generation to the next. Masses of the products are listed in [Table 2](#). The deviation of calculated and measured mass was within 1 Da throughout.

Discussion

Here we describe a novel principle for allele-discrimination. A cycled primer extension reaction is

carried out using four ribonucleotides and a ribonucleotide incorporating thermostable DNA polymerase. In contrast to an RNA polymerase, this polymerase is used with a sequence specific primer rather than a transcription start site. This allows positioning primers freely. After the extension reaction, the product is cleaved with alkali, which results in an extension primer product that only retains the first added nucleotide and effectively thus constitutes a single-base extension product. As primers are positioned immediately next to the position of the SNP that is to be genotyped, this added base carries the allele-information. Products were analyzed by MALDI mass spectrometry. Ribonucleotides were also included in the extension primers at the fourth base from the 3' end. This allows removal of a large part of the extension primer during the alkali cleavage and shifts the allele-products into a more favorable detection range of the mass spectrometer. The allele-distinction method was combined with two versions of sample conditioning, desalting with ion exchange resin and alkylation as is used in the GOOD assay. Both methods were equally effective and allowed unambiguous genotyping. Phosphorothioates in the primers used in the alkylation version are susceptible to oxidation under harsh alkali conditions. However, the conditions chosen here for the alkali cleavage of the ribonucleotides was sufficiently mild to only slightly compromise the phosphorothioates. Alkali cleavage can be carried out with NaOH or KOH. As traces of Na⁺ and K⁺ tend to survive the desalting procedure it is desirable to be able to choose where adducts are detected. Na⁺ adducts are 22 Da larger than their mother molecule, while K⁺ adducts are 38 Da larger. When genotyping, for example, a T/A SNP (which is U/A for the allele-products, e.g. FANCA 10162) where the mass difference between the two alleles is 24 Da, a potential Na⁺ contamination is undesirable as it can compromise allele-calling. The same applies for a C/G (FANCA 10191) where the mass difference is 40 Da. Here a K⁺ adduct contamination is undesirable. The alkali for cleavage can be chosen as a function of the SNP that is being genotyped. However, the desalting procedure is less efficient for K⁺ than Na⁺. [Figure 2](#) shows genotyping of the SNP KIAA 0604_18550 cleaved with NaOH and KOH. We also tested NH₄OH for cleavage. However, in order to achieve efficient cleavage, the high

concentrations and long incubations required seemed impractical even though NH₄⁺ is a substrate that is very compatible with mass spectrometry. An advantage of the transition from DNA to RNA is that the mass difference of T/A SNP increases from 9 to 24 Da due to U replacing T. On the other hand this is a disadvantage for C/T SNPs that is remedied with the inclusion of C^{Me} or U^{Me} to replace cytidine or uridine, respectively. Accordingly, sufficient mass resolution is achieved to distinguish all base combinations of SNPs.

The presented procedure constitutes another addition to the ever-growing arsenal of SNP genotyping methods. Its main advantage over other SNP genotyping methods using mass spectrometric detection lies in the cost and ease of operation. As the G46E CS6R DNA polymerase efficiently incorporates NTPs, an intermediate enzymatic clean-up step with, for example, shrimp alkaline phosphatase (and ExoI) can be omitted. Using alkali to cleave both immediately following the allele-specific base and in the primer is very cost effective. Further, ribonucleotides are orders of magnitude less expensive than the ddNTPs that are commonly used in primer extension assays.

Clearly, the mechanism of primer directed transcription described here could also be of use for applications other than genotyping.

Acknowledgments

This work was supported by the French Ministry of Research (MENRT), the European Commission in the project 'EUMORPHIA' under contract QLRT-2001-00930 and Hoffmann-la Roche AG. Funding to pay the Open Access publication charges for this article was provided by the French Ministry Research and the European Commission.

Conflict of interest statement. None declared.

References

1. Collins F.S., Guyer M.S., Chakravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science*. 1997;278:1580–1581. [[PubMed](#)]
2. McCarthy J.J., Hilfiker R. The use of single nucleotide polymorphism maps in pharmacogenomics. *Nat. Biotechnol.* 2000;18:505–508. [[PubMed](#)]

3. Landegren U., Nilsson M., Kwok P.Y. Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. *Genome Res.* 1998;8:769–776. [[PubMed](#)]
4. Karas M., Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons. *Anal. Chem.* 1988;60:2299–2301. [[PubMed](#)]
5. Nordhoff E., Kirpekar F., Roepstorff P. Mass spectrometry of nucleic acids. *Mass Spectrom. Rev.* 1996;15:67–138.
6. Tost J., Gut I.G. Genotyping single nucleotide polymorphisms by mass spectrometry. *Mass Spectrom. Rev.* 2002;21:388–418. [[PubMed](#)]
7. Little D.P., Braun A., O'Donnell M.J., Köster H. Mass spectrometry from miniaturized arrays for full comparative DNA analysis. *Nature Med.* 1997;3:1413–1416. [[PubMed](#)]
8. Tost J., Kucharczyk R., Lechner D., Gut I.G. The GOOD assay: a purification-free procedure for genotyping by MALDI mass spectrometry. In: Weissensteiner T., Griffin H.G., Griffin A., editors. *PCR Technology—Current Innovations*. London: CRC Press; 2004. pp. 121–130.
9. Mengel-Jørgensen J., Sanchez J.J., Børsting C., Kirpekar F., Morling N. Typing of multiple single-nucleotide polymorphisms using ribonuclease cleavage of DNA/RNA chimeric single-base extension primers and detection by MALDI-TOF mass spectrometry. *Anal. Chem.* 2005;77:5229–5235. [[PubMed](#)]
10. Astatke M., Grindley N.D., Joyce C.M. How *E.coli* DNA polymerase I (Klenow fragment) distinguishes between deoxy- and dideoxynucleotides. *J. Mol. Biol.* 1998;278:147–165. [[PubMed](#)]
11. Gelfand D.H., Kalman L.V., Reichert F.L. Thermostable DNA polymerases having reduced discrimination against ribo-NTPs. US Patent. 5,939,292. 1999.
12. Sauer S., Gut I.G. Extension of the GOOD assay for genotyping single nucleotide polymorphisms by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 2003;17:1265–1272. [[PubMed](#)]

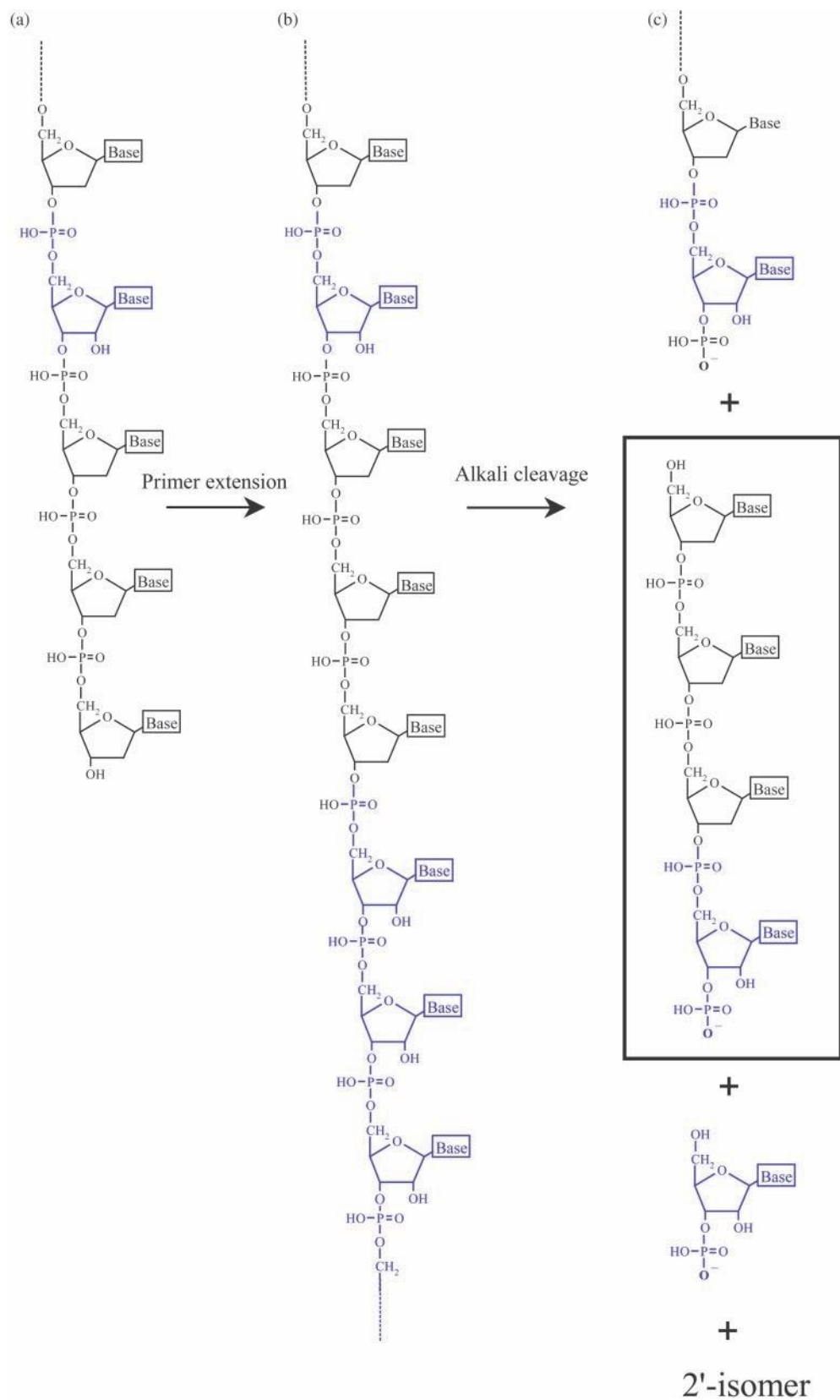
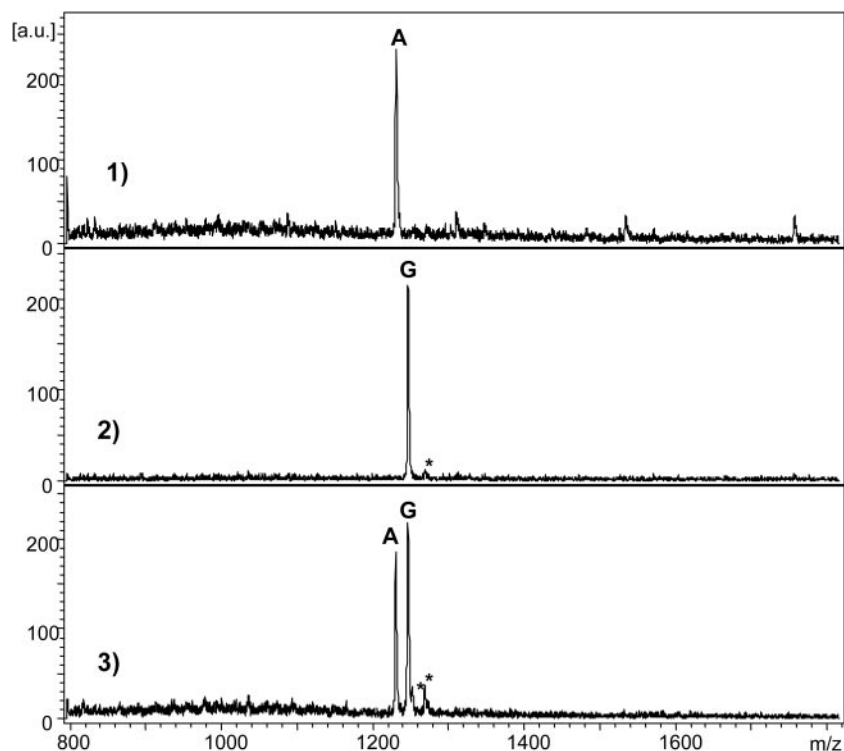


Figure 1. Principle of the ribonucleotide primer extension for SNP genotyping: (a) shows the primer prior to extension with a ribose sugar at the fourth nucleotide from the 3⁰ end; (b) extension of the primer with ribonucleotides; (c) cleavage of the extended product by alkali results in the cleavage inside the primer and after the first added nucleotide generating a mixture of 2⁰-HPO₄-3⁰OH and 2⁰OH-3⁰-HPO₄ products of identical mass.

| Name | | Forward PCR primer | Reverse PCR primer | Extension primer |
|-----------------|---|-------------------------|--------------------------|--|
| Cav3-t+543c | Y | 5'-TAAACCAGCACGCGTTCC | 5'-CGTGTACGGTCCAGCAAACAA | 5'-TTCAGTGGCCAAATGUAAG 5'-TTCAGTGGCCAAATGUA [^] A [^] G |
| FANCA 10162 | W | 5'-GAGGGCTGTACCCTCCTACC | 5'-GAGAGGCAGTCCCATGATA | 5'-CCGCTCTTTTCAGAGGCA 5'-CCGCTCTTTTCAGAGG [^] C [^] A |
| FANCA 10191 | S | 5'-TGCCTGTAATCCCAGCTACC | 5'-AAGGCCTGGAGATAAGCAGC | 5'-GGCCTGAGCATTGGTCCTT 5'-GGCCTGAGCATTGGTCC [^] T [^] T |
| KIAA 0604_18550 | R | 5'-CCTGCTCCCTCCTGAGTATG | 5'-GGTCTTGCTCTGCCACTTA | 5'-CGCAGAGTGGCCTGTC 5'-CGCAGAGTGGCCTGT [^] C [^] C |

Table 1. Sequences of the primers used for the four tested SNPs. The symbol ‘^’ indicates a phosphorothioate bridge and is combined with a final conditioning by alkylation. Bolded nucleotide in extension primer designates position of ribose sugar in primer.

A



B

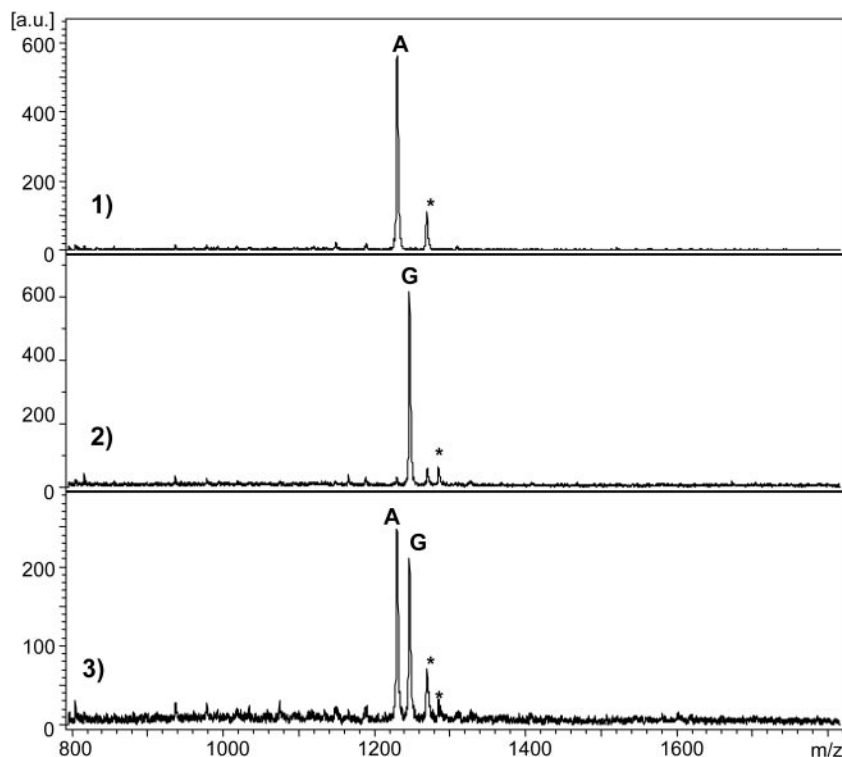
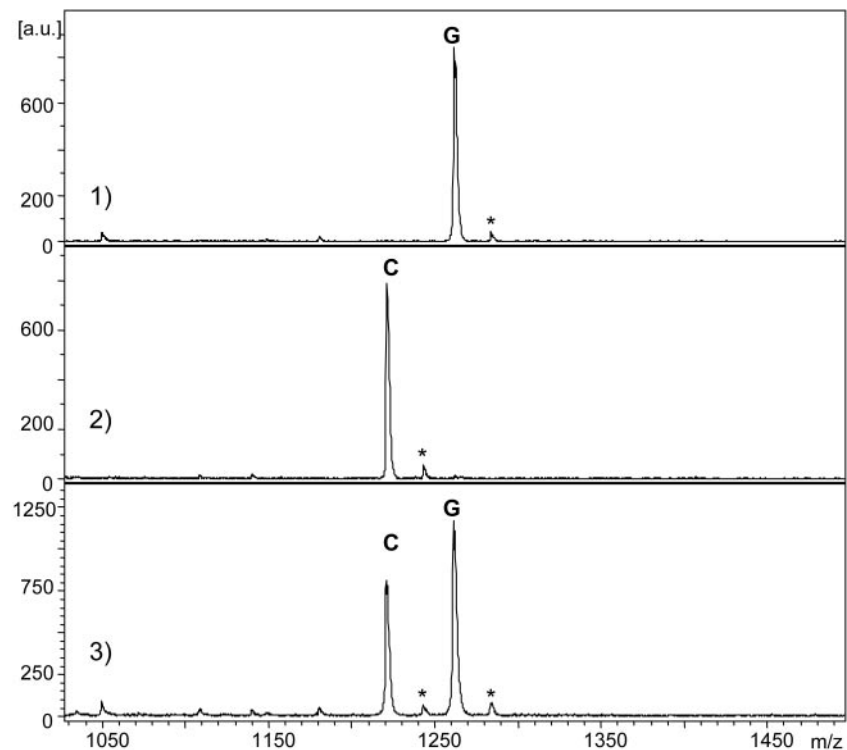


Figure 2. Spectra of the SNP KIAA 0604_18550 (A/G): (a) shows cleavage with NaOH. Na⁺ adduct peaks (asterisk) are observed at +22 Da from the real allele-peaks; (b) shows cleavage with KOH. K⁺ adduct peaks (asterisk) are observed at +38 Da from the real allele-peaks. In addition +22 Da Na⁺ adducts are detected due to ubiquitous Na⁺. (1) Mass spectrum of a homozygous DNA sample for the first allele, (2) mass spectrum of a homozygous DNA sample for the second allele, and (3) mass spectrum of a heterozygous sample.

A



B

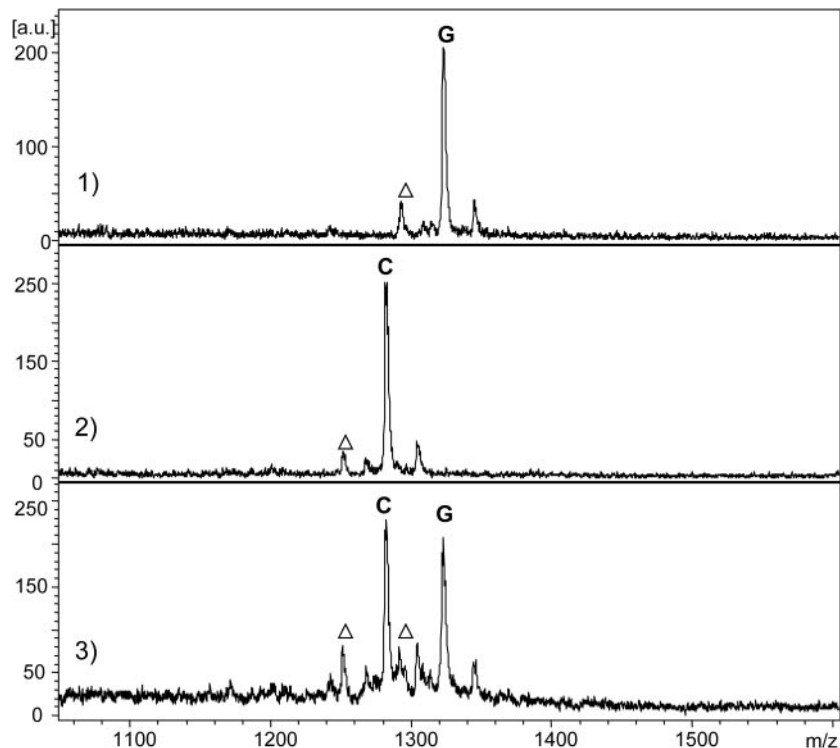


Figure 3. Spectra of the SNP FANCA 10191: (a) shows the protocol using desalting with ion exchange resin, Na⁺ adducts are labeled with (asterisk); (b) shows the protocol with alkylation and no desalting. Peaks associated with exchanges of S by O, which then cannot be alkylated, are labeled with open triangle. Homozygous samples for each of the two alleles and a heterozygous sample are shown (1–3°).

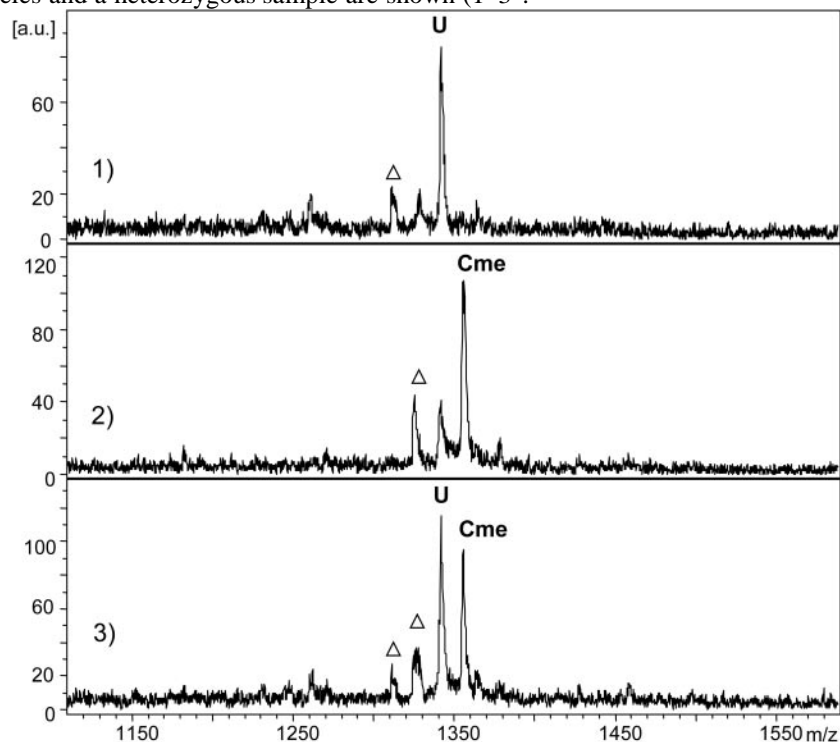


Figure 4. Mass spectra of the SNP Cav3-t+543c: conditioning is carried out according to the GOOD assay where the sample is subjected to alkylation, which neutralizes the negative charges on the phosphate backbone. An

example for U and C^{Me} is shown for DNA samples homozygous for either of the alleles as well as a heterozygous DNA sample (1–3). Peaks associated with exchanges of S by O, which then cannot be alkylated, are labeled with open triangle

| Name | Allele 1 | Mass allele 1/Da | Allele 2 | Mass allele 2/Da |
|-----------------|--|------------------|--|------------------|
| Cav3-t+543c | AAGC | 1277.9 | AAGU ^{Me} | 1292.9 |
| | AAGC ^{Me} | 1291.9 | AAGU | 1278.9 |
| | A [^] A [^] GC | 1338.0 | A [^] A [^] GU ^{Me} | 1353.0 |
| | A [^] A [^] GC ^{Me} | 1352.0 | A [^] A [^] GU | 1339.0 |
| FANCA 10162 | GCAA | 1277.9 | GCAU | 1254.8 |
| | G [^] C [^] AA | 1338.0 | G [^] C [^] AU | 1315.0 |
| FANCA 10191 | CTTC | 1219.8 | CTTG | 1259.8 |
| | C [^] T [^] TC | 1279.9 | C[^]T[^]TG | 1320.0 |
| KIAA 0604_18550 | TCCA | 1228.8 | TCCG | 1244.8 |
| | T [^] C [^] CA | 1289.0 | T[^]C[^]CG | 1305.0 |

Table 2. Sequences and masses of the allele-products. The symbol ‘^’ indicates a phosphorothioate bridge. After alkylation with methyl iodide, each phosphorothioate bridge results in a change of +30 Da relative to a regular phosphate bridge (mass difference between OH and SCH₃). Bolded nucleotide designates added ribonucleotide

Annexe 2: DNA sequencing by MALDI-TOF MS using alkali cleavage of RNA/DNA chimeras

Nucleic Acids Res. 2007 April; 35(8): e62. Published online 2007 April 10. doi: [10.1093/nar/gkm056](https://doi.org/10.1093/nar/gkm056)

Florence Mauger,¹ Keith Bauer,² Cassandra D. Calloway,² Jérémy Semhoun,¹ Tetsuya Nishimoto,² Thomas W. Myers,² David H. Gelfand,² and Ivo G. Gut^{1*}

¹Centre National de Génotypage, Bâtiment G2, 2 rue Gaston Crémieux, 91057 Evry Cedex, France and ²Roche Molecular Systems Inc., 1145 Atlantic Avenue, Alameda, California 94501, USA

*To whom correspondence should be addressed Phone: +33 160 87 84 00, Fax: +33 160 87 83 83, ; Email: ivo.gut@cng.fr

Received November 9, 2006; Revised January 17, 2007; Accepted January 19, 2007

ABSTRACT

Approaches developed for sequencing DNA with detection by mass spectrometry use strategies that deviate from the Sanger-type methods. Procedures demonstrated so far used the sequence specificity of RNA endonucleases, as unfortunately equivalent enzymes for DNA do not exist and therefore require transcription of DNA into RNA prior to fragmentation.

We have developed a novel, rapid and accurate concept for DNA sequencing using mass spectrometry and RNA/DNA chimeras and applied it to sequence mitochondrial DNA. Our method is based on the preparation of a chimeric RNA/DNA with a DNA polymerase that also incorporates ribonucleotides. Sequencing is carried out with one ribonucleotide (ATP, CTP or GTP) and the other three nucleotides in their deoxyribo-form. The product is treated with alkali, which cleaves 3' of all ribonucleotides to form a terminal 3' phosphate. Conditions have been streamlined so that molecular, biological and alkali cleavage conditions are compatible with matrix-assisted laser desorption/ionization time-of-flight (MALDI) mass spectrometric analysis. Fragment analysis by MALDI MS provides a sequence-specific fingerprint, which allows the identification of differences between a reference and another sequence. Due to the mass profile, the position and kind of the mutation can be assigned. These differences between signatures are indicative of known, unidentified, rare and private mutations.

This novel DNA sequencing protocol was applied to sequence the hypervariable region 1 (HV1) of mitochondrial DNA in 22 individuals.

INTRODUCTION

For nucleic acid analysis, mass spectrometers have predominantly been applied for SNP genotyping. Particularly, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI) is the method of choice due to its reliability, accuracy and speed (1). DNA analysis by mass spectrometry was reviewed extensively and for a general overview we refer the reader to publications cited therein (2,3). More recently, novel methods to determine more complex sequence variability in defined loci have been described (4,5).

The most evolved methods for DNA sequencing with MALDI mass spectrometric detection use an initial PCR amplification with primers that contain transcription start sites at their 5' ends. After transcription of the PCR product, sequence-specific RNA endonucleases are used to generate fragments terminating on a specific base. Sizing the fragments by MALDI allows the identification of the sequence differences relative to a reference. The desire for high throughput, parallel processing, simplified handling and low-cost reagents have influenced the development of these methods.

Here we present a novel concept for determining the complete genotype in one of the hypervariable regions (hypervariable region I) of human mitochondrial DNA combined with MALDI mass spectrometry for detection. We chose to sequence this region of the mitochondrial genome because it is well characterized and has a high degree of sequence polymorphism. Also, since mitochondria are maternally inherited, mtDNA is haploid and in most cases we can expect to have only a single sequence. Hypervariable regions I and II (HVI/HVII) of human mitochondrial genome are first amplified in duplex (6–8). We then take advantage of a novel class of thermostable DNA polymerases that efficiently incorporate and rapidly extend NTPs to generate chimeric primer extension products. These products contain three deoxyribonucleotides and the fourth nucleotide in its ribo form (A, C or G). The chimeric products are sequence-specifically fragmented by treatment with alkali, generating oligonucleotide fragments each terminating with the ribonucleotide of the cycled primer extension reaction. Taking advantage of the resolving power, precision and speed of the mass spectrometer, the masses of the fragments are determined. Analysis of differentially substituted (e.g. A, C, G) extension products is able to unambiguously provide the sequence of the original locus. Additional confirmation of sequence is readily obtained using a complementary primer on the opposite strand and one of the NTPs.

MATERIALS AND METHODS

Twenty-two DNA samples were used for this work and results are summarized in [Table 1](#). They were selected based on representing the major variants of the hypervariable region (HV) of human mitochondrial DNA. All samples were also sequenced by fluorescent Sanger sequencing for control.

Primers for PCR were from the LINEAR ARRAY Mitochondrial DNA HVI/HVII Region-Sequence Typing Kit (Roche Diagnostics, Indianapolis, IN, USA). Primers for ribosubstitution cycled extensions were synthesized by Roche Molecular Systems Core Chemistry Department (Alameda, CA, USA). Sequences of the primers used in this study are listed in [Table 2](#). Deoxynucleoside triphosphates, dNTPs, (N = A, C, G, T) and ribonucleoside triphosphates, NTPs, (N = A, C, G,) were purchased from GE

Healthcare (Saclay, France). Qiagen QiaQuick PCR Purification Kit was purchased from Qiagen (Courtaboeuf, France). Ion exchange resin AG 50W-X8 H⁺ was purchased from BioRad (Marmes la Coquette, France). General chemical reagents were purchased from Aldrich (Steinheim, Germany). Thermocycling procedures were carried out in Eppendorf Gradient Thermocyclers (Eppendorf, Germany).

KB17 DNA polymerase was provided by Roche Molecular Systems (Alameda, CA, USA). This designer DNA polymerase is a chimeric DNA polymerase comprising the 5'-nuclease domain of *Thermus* sp. Z05 DNA polymerase and the 3'-nuclease and 5' to 3' DNA polymerase domains of *Thermotoga maritima* (*Tma*) DNA polymerase. The 5'-nuclease activity is eliminated in this enzyme by the introduced G46E mutation (9). The wild-type *Tma* proofreading activity has been modulated in KB17 DNA Polymerase by introduction of an L329A mutation in 'Motif I' of the 3' to 5' exonuclease domain (10,11). The enzyme's polymerase domain contains a mutation that eliminates the selectivity of the wild-type enzyme against incorporation of C'-2-substituted nucleotides and which facilitates efficient incorporation of ribonucleoside triphosphates (9). Finally, the enzyme contains three additional mutations in the polymerase domain, which enhance binding to template and result in faster extension rates, particularly in high ribonucleotide substitution reactions [please contact Thomas Myers (Thomas.Myers@Roche.com) at Roche Molecular Systems for research samples].

PCR amplifications of HVI/HVII regions were carried out in 50 µl with 0.2 pg/µl of mitochondrial DNA, 0.3 µM each primer, 10 mM Tris/HCl, 0.1 mM EDTA at pH 8.0, 1.0 mM MgCl₂, 0.3 mM each dNTP, 1× PCR Buffer II and 0.1 U/µl *AmpliTaq* Gold DNA Polymerase. The thermal cycling profile for the duplex PCR was 14 min at 94°C followed by 34 cycles of 15 s at 92°C, 30 s at 59°C and 30 s at 72°C. This was concluded with 10 min at 72°C. The PCR product was purified with the Qiagen QiaQuick PCR Purification kit and the concentration was measured.

The extension reactions were carried out in a total volume of 20 μl , 5 μl of 0.02 μM purified PCR product, 0.5 μl of KB17 polymerase (10 U/ μl), 0.5 μl of 4 mM NTP/dNTPs mix, 2.5 μl of 200 mM tricine pH 7.75, 2.6 μl of 500 mM potassium acetate, 1 μl of 30 mM magnesium acetate, 2 μl of extension primer (10 pmol/ μl) and 5.9 μl water. The composition of the NTP/dNTPs mix was ATP, dCTP, dGTP and dTTP; CTP, dATP, dGTP and dTTP or GTP, dATP, dCTP and dTTP, respectively. The thermal cycling profile for the extension reaction was 15 s at 89°C followed by 20 cycles of 15 s at 89°C and 4 min at 62°C. Extension primers also contained 5' 6-FAM labels, so that full-length extension products could be controlled by capillary electrophoresis (Figure 1). For capillary electrophoresis, 5 μl of the extension reaction was loaded onto a MegaBace1000 (GE Healthcare, Amersham, UK).

For alkali cleavage, 1.8 μl of 3.3 M sodium hydroxide and 3.2 μl of water were added to the 15 μl of extension reaction for a final concentration of 0.3 M and incubated at 70°C for 1.5 h. Samples were desalted by the addition of cation exchange resin charged with H⁺. One-third of the total reaction volume was added in resin and incubated for 20 min at room temperature under agitation. Thereafter, the sample was centrifuged for 2 min at 134 \times g to sediment the resin. All of the supernatant was removed. Trihydroxyacetophenone (THAP) was used as matrix (12). For preparation, 0.5 μl of 0.2 M of 2,4,6 and 2,3,4 THAP in 50% acetonitrile and 0.3 M of ammonium citrate in water in 6/3/2 (v/v) was deposited on an anchor position of MALDI target plate (AnchorChip™ Target with a spot size of 400 μm , Bruker Daltonik GmbH, Bremen, Germany). Afterwards, 0.5 μl of sample was added and dried at room temperature.

The target was introduced into the MALDI mass spectrometer (Autoflex and Ultraflex II, Bruker Daltonik GmbH, Bremen, Germany) for analysis. Analysis was carried out in negative ion mode, with an acceleration voltage of 20 kV using a pulsed ion extraction delay of 100 ns in linear and reflectron mode and with external calibration. Each spectrum obtained was the sum of 200 laser shots for linear mode and 400 laser shots for reflectron mode.

RESULTS

The principle of mitochondrial DNA sequencing developed here is a cleavage of RNA/DNA chimeras and analysis of fragment fingerprints by MALDI mass spectrometry.

Mitochondrial DNA is unique in that 100–1000s of copies can be found in a single cell, far exceeding the two copies of genomic DNA. It is also haploid, which means that although different sequences exist in the human population, none of the polymorphic positions are found to be heterozygous (Table 1).

A duplex PCR of the hypervariable region of the mtDNA, HVI and HVII, was used to prepare a template and reduce the complexity for the following extension step. The product was purified to remove residual primers and dNTPs from the PCR prior to the extension reaction.

Primer extension is carried out with a reaction mixture containing one ribonucleotide and the three other nucleotides in their deoxyribo-form with a thermostable ribonucleotide incorporating chimeric DNA polymerase that has modulated proofreading activity. Extension primers are prepared with 2'OMe or 2'NH₂ on the minus one terminal base from the 3' end to prevent degradation by the DNA polymerase due to proofreading (Table 2). The polymerase can readily incorporate up to 100% of ATP, 100% of CTP or 100% of GTP and all of the deoxyribonucleotides and full extension is obtained (Figure 1). The product is a chimeric DNA/RNA of about 440 nt. Due to a mobility shift between RNA and DNA in capillary electrophoresis, the different full extension chimeric RNA/DNA products display different mobility for the ATP, CTP and GTP track. We did not find conditions that allowed the incorporation of 100% UTP. However, this was not a reason for concern as, on one hand the three other tracks provide sufficient information to unambiguously identify all sequence differences, and on the other leave the possibility to use the dUTP in conjunction with uracil-*N*-glycosylase (UNG) decontamination.

Fragments for the mass spectrometric analysis are generated by alkali cleavage with sodium hydroxide (Table 4, Supplementary Data). This results in the cleavage 3' after each added ribonucleotide. Under these conditioning and analysis methods, a residual 2'

or 3' phosphate group remains at the 3' terminal base. From the mass we cannot distinguish 2'-H₂PO₄ + 3'-OH and 2'-OH + 3'-H₂PO₄. Quite likely products are a mixture of these two species. Due to the negative charges on the sugar phosphate backbone, adducts of the phosphate groups with Na⁺ are observed when samples are not sufficiently well desalted. The most common method for desalting, the addition of an ion exchange resin H⁺ was used here.

In linear detection mode, the 3-mer to 20-mer interval can be analyzed whereby the instrument resolution is lower throughout and the number of species at the bottom of this interval make calling difficult, while at the top of the range sensitivity drops off. In reflectron mode, the 3-mer to 12-mer interval can be analyzed. Isotopic resolution is achieved beyond the 7-mers and detection sensitivity drops off faster than in linear mode, which might be due to post-source decay of the larger fragments. Calculated masses are listed in Table 4a and c (Supplementary Data). We chose to analyze samples in both linear and reflectron mode and use the combined information for sequence assignment. In the three and four base fragment range many fragments have the same mass and cannot be distinguished.

We used the reverse reactions of HVI with ATP, CTP and GTP to determine the sequences of 22 DNA samples blinded. Combining fragment fingerprints from the different ribonucleotides on the same fragment of the HVI region was used to obtain complete sequence coverage. For confirmation, the forward reaction of HVI with ATP, CTP and GTP was measured. Forward GTP is complementary to reverse CTP (Table 4b, Supplementary Data), which due to the sequence has many fragments around 20 bases. In Table 1, compound results for the 22 DNA samples are listed and differences from the Cambridge reference sequence (CRS) (13) are noted. Samples C018 and C207 carry the CRS. With the three reactions in one orientation we were able to determine polymorphism without ambiguity in the majority of samples. The determined sequences were compared with the reference from Sanger capillary electrophoresis sequencing. The results generated by our method matched the reference perfectly.

For illustration we are detailing the analysis of samples C018 and C004. This example exemplifies

all situations of changes in the fingerprints that can occur. There are two sequence differences relative to the CRS (16224 and 16311). Table 3 lists the differential fragments. In reverse ATP, a decrease of the relative intensity from 3 to 2 of a peak at mass 2272.5 Da is observed due to the loss of the fragment GGGTTGA (Figure 2A) and an increase of relative intensity from 1 to 2 of another peak at 2207.4 Da due to the appearance of the fragment TGTGCTA. TGTGTGA at 2247.4 Da remains unchanged and is used as the reference. A new peak with a mass of 3868.5 Da appears which corresponds to the fragment GTTGGGGTTGA (Figure 2B). In reverse CTP, the appearance of a new fragment at 2815.8 Da (TTTATGTGC) and the disappearance of a fragment at 2799.8 Da (TTTATGTAC, Figure 3) are observed. Effectively the peak shifted 16 Da due to an A to G base change in the fragment. However, as there are two As in this fragment an ambiguity remains. It is removed by examination of the reverse GTP spectrum in which a peak at 2191.4 Da which corresponds to the fragment TACTATG (Figure 4) disappeared and an increase in relative intensity from 2 to 3 of a peak at 1574 Da is observed that corresponds to CTATG. Polymorphism 16224 is not detectable in reverse GTP because it results in a fragment of less than three bases. Also, it cannot be detected in reverse CTP as the indicative fragment is a 31-mer which is too big to resolve. Both polymorphisms of sample C004 can easily be assigned by comparing the spectra of the ATP, CTP and GTP reactions with the reference.

We created software that aids with the interpretation of results. Using the reference sequence a theoretical peak profile is established. This is then overlaid with the actual spectrum. The treatment of the actual spectrum is limited to a baseline correction using an adaptation of the Convex Hull algorithm. No smoothing is applied to prevent any loss of information. The peak recognition is simplified because with the theoretical spectrum, the expected masses of the peaks are known. Differences of the measured and theoretical spectrum are used to call sequences. Software to carry out similar tasks is available from Bruker Daltonik (FlexAnalysis) or Sequenom (www.sequenom.com).

DISCUSSION

Here we describe a novel principle for resequencing DNA using a MALDI mass spectrometer for detection. PCR products are subjected to a cycled primer extension reaction using three deoxyribonucleotides, one ribonucleotide and a ribonucleotide incorporating thermostable DNA polymerase. In contrast to methods using a modified T7 RNA polymerase (14), which require the PCR primers to contain the RNA polymerase promoter sequence and necessitate individual PCRs rather than permitting multiplex PCR, this thermostable DNA polymerase is used with a sequence-specific primer rather than a transcription start site. This allows positioning primers freely. Cycled extension products may be substituted with 100% of any of the three ribonucleotides (ATP, CTP or GTP). After the cycled extension reaction, the product is simply cleaved with alkali to generate a sequence-specific family of oligonucleotide products that each terminates with the ribonucleotide used in the cycled extension reaction. No expensive base-specific enzymes are required. A-terminated, C-terminated or G-terminated products were generated with equivalent efficiency and ease, while the incorporation of UTP is not as efficient and does not work at 100% replacement. This is not a major cause for concern because the U channel should be used for dUTP in conjunction with uracil-*N*-glycosylase for carryover contamination control. Products were analyzed by MALDI mass spectrometry. Assembling and merging the profiles for the individual reactions yielded unambiguous sequence for each of the 22 samples. Cycled primer extension reactions with a complementary primer and one of the NTPs in the opposite sense may be used to confirm the assembled sequence.

In this study, we used MALDI mass spectrometers in linear and reflectron mode. The linear mode allows covering a larger mass range (1–8 kDa) but does not resolve different species with similar masses. The reflectron mode provides isotopic resolution over its entire but limited mass range (1–4 kDa). Particularly in the low mass range, species can be differentiated using the reflectron mode, which are not resolvable in linear mode. We found it advantageous to record spectra in both modes. This gave us choices for analysis. Modern instrumentation can be set up to record spectra automatically in linear and reflectron mode.

The presented procedure constitutes another addition to the ever-growing arsenal of resequencing methods. Its main advantage over other resequencing methods using mass spectrometric detection lies in the ease of operation and probably lower cost than current methods. Any appropriate sequence-specific extension primer may be used. Thus, amplicons could be generated in multiplex PCRs obviating the T7 RNA polymerase promoter sequence in the PCR primer. As the KB17 DNA polymerase efficiently incorporate and rapidly extends the provided NTPs, expensive nucleotide analogs (e.g. α -S-CTP or -UTP) are unnecessary. Further, in contrast to other enzymes used in sequencing or resequencing, KB17 DNA polymerase contains a 3'–5' exonuclease or proofreading activity that assists in maintaining high fidelity in primer extension reactions. The extension out to 440 bases is efficient and no indication of premature termination of extension was detected. Theoretical estimations indicate that 1000 bases of sequence could still be resolved by our approach without marked problems due to congested spectra. Beyond 1000 bases, problems due to ambiguities are anticipated. As fragmentation uses facile, inexpensive and complete alkali cleavage, base-specific nucleases or other enzymes are avoided. Finally, ribonucleotides are orders of magnitude less expensive than the ddNTPs that are commonly used in primer extension assays.

Clearly, the mechanism of primer-directed transcription/resequencing described here could also be of use for applications other than DNA sequencing.

SUPPLEMENTARY DATA

Supplementary Data is available at NAR online.

ACKNOWLEDGEMENTS

This work was supported by the French Ministry of Research (Ministère délégué à la Recherche) and the European Community through the integrated project 'MolTools' under contract LSHG-CT-2003-503155 and through funding from Roche Applied Science. We thank Dr. Roderic Fuerst for encouragement and support of these efforts. Funding to pay the Open Access publication charge was provided by the authors' institutions.

Conflict of interest statement. None declared.

REFERENCES

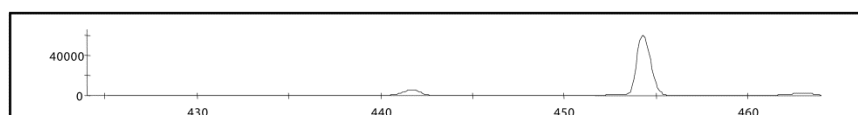
1. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* 1988;60:2299–2301.
2. Nordhoff E, Kirpekar F, Roepstorff P. Mass spectrometry of nucleic acids. *Mass Spectrom. Rev.* 1996;15:67–138.
3. Tost J, Gut IG. Genotyping single nucleotide polymorphisms by mass spectrometry. *Mass Spectrom. Rev.* 2002;21:388–418.
4. Krebs S, Medugorac I, Seichter D, Forster M. RNaseCut: a MALDI mass spectrometry-based method for SNP discovery. *Nucleic Acids Res.* 2003;31:e37.
5. Ehrich M, Bocker S, van den Boom D. Multiplexed discovery of sequence polymorphisms using base-specific cleavage and MALDI-TOF MS. *Nucleic Acids Res.* 2005;33:e38.
6. Hall TA, Budowle B, Jiang Y, Blyn L, Eshoo M, Sannes-Lowery KA, Sampath R, Drader JJ, Hannis JC, et al. Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: a novel tool for the identification and differentiation of humans. *Anal. Biochem.* 2005;344:53–69.
7. Kline MC, Vallone PM, Redman JW, Duewer DL, Calloway CD, Butler JM. Mitochondrial DNA typing screens with control region and coding region SNPs. *J. Forensic Sci.* 2005;50:377–385.
8. Oberacher H, Niederstatter H, Huber CG, Parson W. Accurate determination of allelic frequencies in mitochondrial DNA mixtures by electrospray ionization time-of-flight mass spectrometry. *Anal. Bioanal. Chem.* 2006;384:1155–1163.
9. Mauger F, Jaunay O, Chamblain V, Reichert F, Bauer K, Gut IG, Gelfand DH. SNP genotyping using alkali cleavage of RNA/DNA chimeras and MALDI time-of-flight mass spectrometry. *Nucleic Acids Res.* 2006;34:e18.
10. Schonbrunner NJ, Myers TW, Gelfand DH. US Patent Office Publication; 2004. US patent no. 20040005599(A1)
11. Schonbrunner NJ, Fiss EH, Budker O, Stoffel S, Sigua CL, Gelfand DH, Myers TW. Chimeric thermostable DNA polymerases with reverse transcriptase and attenuated 3'-5' exonuclease activity. *Biochemistry.* 2006;45:12786–12795.
12. Zhu YF, Chung CN, Taranenko NI, Allman SL, Martin SA, Haff L, Chen CH. The study of 2,3,4-trihydroxyacetophenone and 2,4,6-trihydroxyphenone matrices for DNA detection in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Comm. Mass Spectrom.* 1996;10:383–388.
13. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, et al. Sequence and organization of the human mitochondrial genome. *Nature.* 1981;290:457–465.
14. Hahner S, Ludemann HC, Kirpekar F, Nordhoff E, Roepstorff P, Galla HJ, Hillenkamp F. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI) of endonuclease digests of RNA. *Nucleic Acids Res.* 1997;25:1957–1964.

Table 1. Sequences of 22 DNA samples of HVI region of mitochondrial DNA

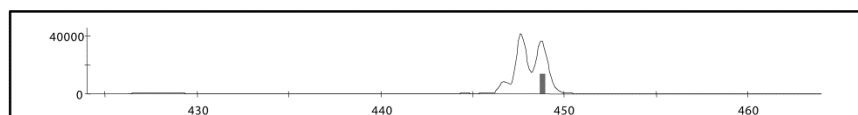
| Region | Name | Duplex PCR primer | Extension primer |
|--------|---------|----------------------|-----------------------------|
| HVI | Forward | CTCCACCATTAGCACCCAA | FAM-CCACCATTAGCACCCAAAG(C)T |
| | Reverse | ATTTACGGAGGATGGT | FAM-CACGGAGGATGGTGGT[C]A |
| HVII | Forward | CACCCTATTAACCACTCACG | |
| | Reverse | CTGTTAAAAGTGCATACCGC | |

Table 2. Sequences of the primers used for HV region for duplex PCR amplification and primer extension reactions

a)



b)



c)

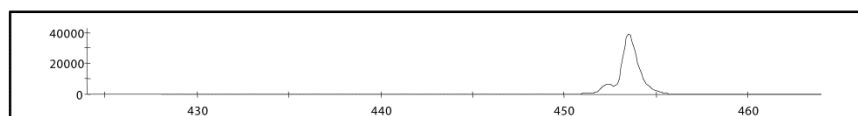
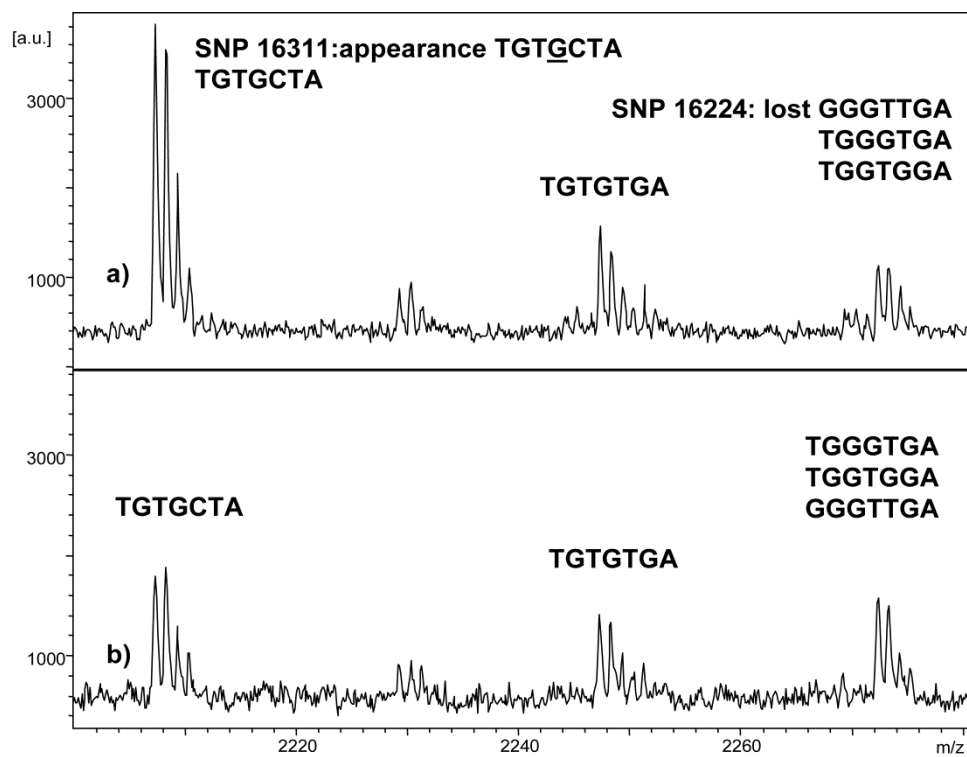


Figure 1. Capillary electropherograms of HVI reverse ribo extension products of sample C207 prior to fragmentation: (A) with ATP; (B) with CTP and (C) with GTP show that extension is full length (440 bases). Mobility is influenced by ribonucleotide content.

| NTP | SNP | C018: Reference | Masses | C004 | Masses |
|-----|-------|---------------------------------|--------|--------------------------------|--------|
| A | 16224 | GTTGA | 1614 | GTTGGGGTTGA | 3868,5 |
| | | GGTTGA | 2272,5 | | |
| | 16311 | TGTA | 1284,8 | TGTGCTA | 2207,4 |
| | | CTA | 940,6 | | |
| C | 16224 | AGTTGATGTGTGATAGTTGAGGGTTGATTGC | 9803,3 | AGTTGATGTGTGATAGTTGGGGTTGATTGC | 9819,3 |
| | 16311 | TTTATGTAC | 2799,8 | TTTATGTGC | 2815,8 |
| G | 16224 | GA | | G | |
| | 16311 | TACTATG | 2191,4 | CTATG | 1574 |

Table 3. Example of the differential peaks of the HVI region with the reverse ribo extensions of sample C018 which is the reference and sample C004

2A



2B

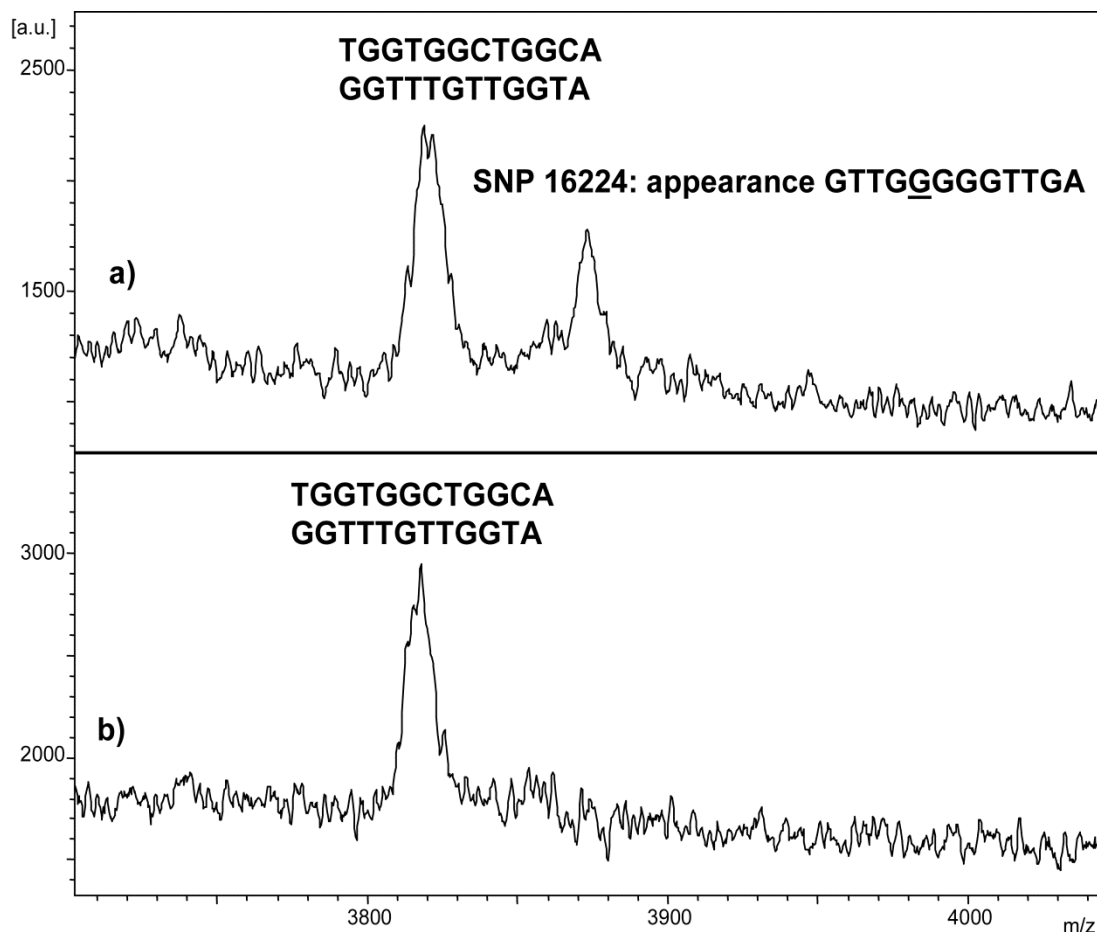


Figure 2. Spectra of the reverse ATP trace of HVI of sample C004 (a) and sample C018 (b) which corresponds to the Cambridge Reference Sequence. **(A)** In the 7-mer region two peaks change intensity. A decrease of the relative intensity from 3 to 2 of the peak at mass 2272.5 Da due to the loss of the fragment GGGTTGA is observed together with an increase of relative intensity of another peak at 2207.4 Da from 1 to 2 due to the appearance of the fragment TGTGCTA. These spectra were recorded on a Bruker Ultraflex II in reflectron mode which achieves isotopic resolution in this mass range. **(B)** In the 12-mer region, a new peak with a mass of 3868.5 Da appears which corresponds to the fragment GTTGGGGGTTGA. These spectra were recorded on a Bruker Autoflex in linear mode.

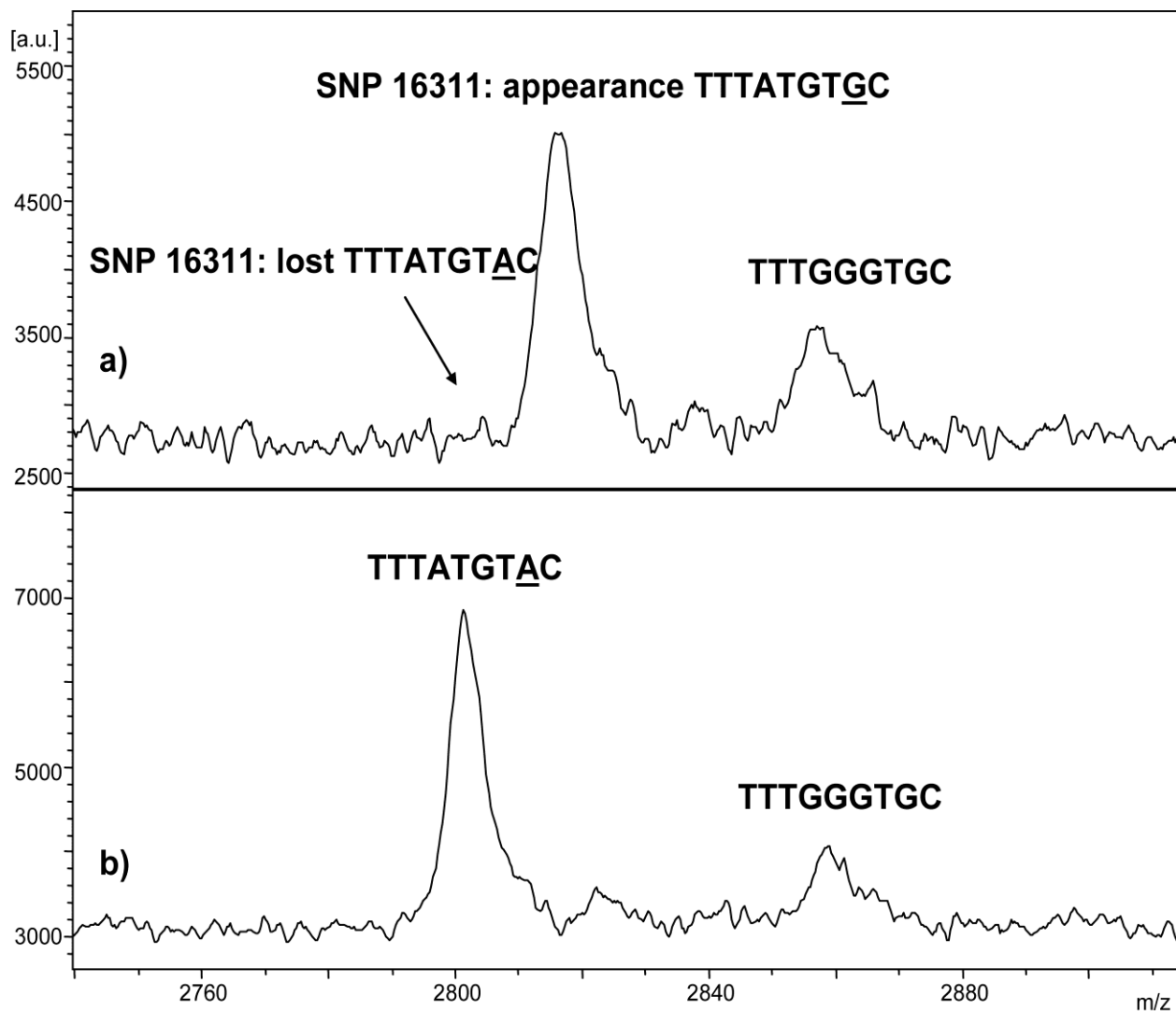


Figure 3. Spectra of the 9-mer region of the reverse CTP trace of HVI of sample C004 (a) and C018 (b) Cambridge Reference Sequence. The polymorphism 16311 of sample C004 results in a shift of the peak at 2799.8 Da (TTTATGTAC) to 2815.8 Da (TTTATGTGC) due to an A to G base change in the fragment. These spectra were recorded on a Bruker Autoflex in linear mode.

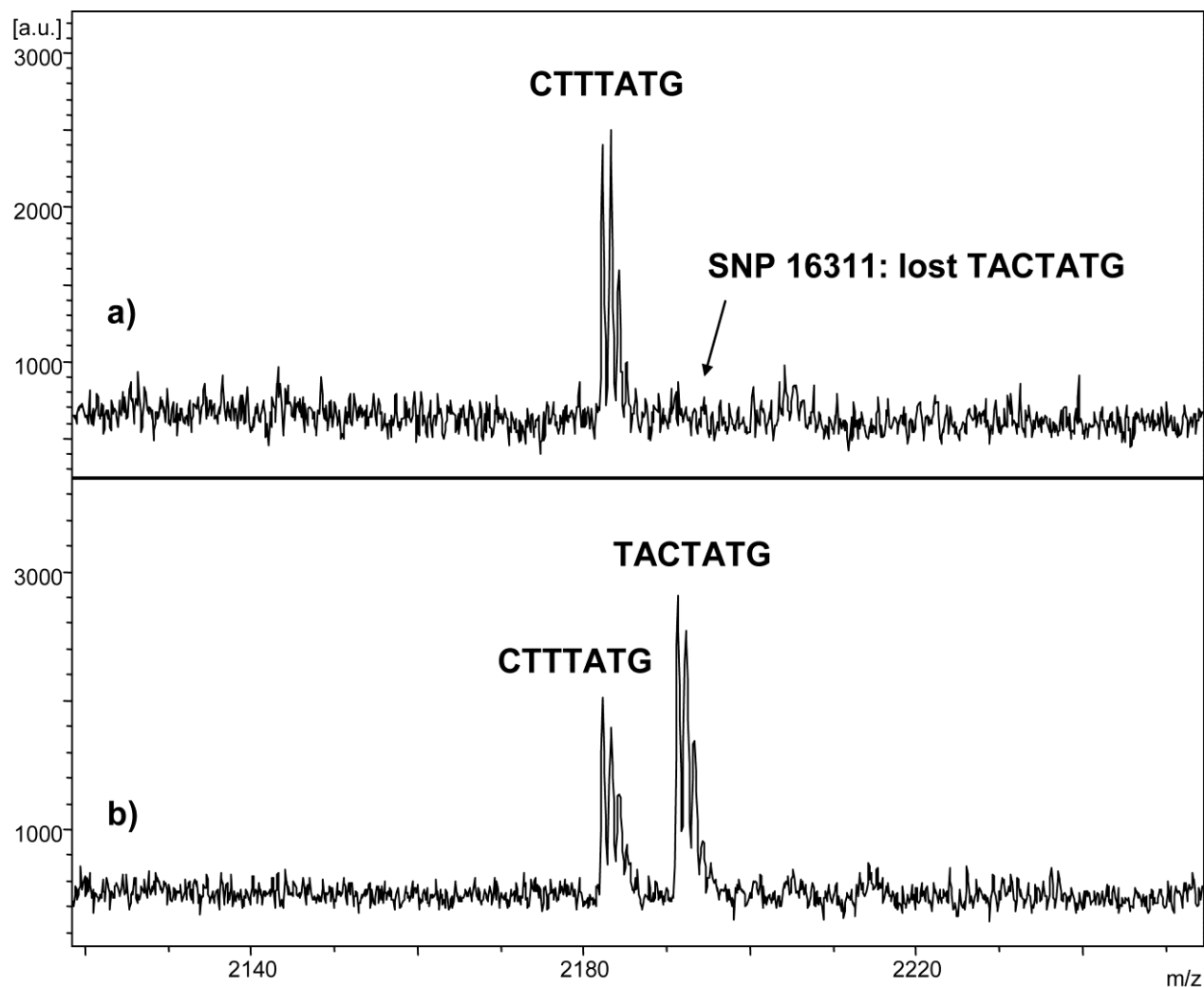


Figure 4. Spectra of the 7-mer region of the reverse GTP trace of HVI of sample C004 (a) and C018 (b) Cambridge Reference Sequence. A fragment at 2191.4 Da which corresponds to TACTATG disappeared. These spectra were recorded on a Bruker Ultraflex II in reflectron mode which achieves isotopic resolution in this mass range.

Annexe 3: Ribo-polymerase chain reaction-A facile method for the preparation of chimeric RNA/DNA applied to DNA sequencing[†]

Florence Mauger¹, Keith Bauer², Jérémy Semhoun¹, Thomas W. Myers², David H. Gelfand¹, Ivo G. Gut^{1,3,*}

Hum Mutat. 2012 Feb 21. doi: 10.1002/humu.22061. Article first published online: 4 APR 2012
DOI: 10.1002/humu.22061

Keywords: polymorphism; ribo-PCR; resequencing; MALDI–TOF mass spectrometry; chimeric RNA/DNA

Abstract

We describe ribo-polymerase chain reaction (PCR), a method for the preparation of chimeric RNA/DNA. The RNA/DNA chimeric nucleic acids are generated directly from genomic DNA starting templates with two locus-specific primers, three nucleotides in their deoxy form and the fourth in its ribo form, a DNA polymerase capable of incorporating ribo bases, a suitable buffer, and thermal cycling. We have applied ribo-PCR to resequence DNA by directly fragmenting the RNA/DNA chimeras with alkali and analyzing the fragments by mass spectrometry (MS). Mass fingerprint is used to identify deviations from the reference sequence. This method readily detects homozygous sequence deviations as well as heterozygous positions directly from genomic DNA samples. With the high-throughput capability of MS, this facile method is well suited for screening DNA sequences of limited regions of the genome in a large number of individuals. It can also be used to sequence multiple distant genomic loci in a single reaction. This novel ribo-PCR resequencing protocol was applied to different genomic loci involving nitric oxide synthase 1 (*NOS1*) and *H19* in 30 individuals and *SLCO1B1* in 95 individuals.

Introduction

Polymerase chain reaction (PCR) has become one of the most powerful and applied methods of molecular biology since its first description 25 years ago [Saiki et al., 1985]. Concurrent with the introduction of *Taq* DNA polymerase for PCR [Saiki et al., 1988], this enzyme was also quickly used for DNA sequencing directly from PCR product [Innis et al., 1988] and then cycle sequencing [Carothers et al., 1989]. The PCR and mutants of *Taq* DNA polymerase became very important for generating templates and ultimately enabling high-throughput, four-color capillary electrophoretic separation platforms in the mid to late 1990s, and the first DNA sequence of a human genome [Lander et al., 2001; Venter et al., 2001].

Mass spectrometers also have been applied for single-nucleotide polymorphism genotyping and DNA sequencing. Particularly, matrix-assisted laser desorption/ionization–time-of-flight mass

spectrometry (MALDI–TOF MS) [Karas and Hillenkamp, 1988] is the method of choice due to its reliability, accuracy, and throughput. The analysis of DNA by MS has been reviewed extensively, and for a general overview, we refer to publications cited therein [Hahner et al., 1997; Nordhoff et al., 1996; Sauer et al., 2000]. More recently, novel methods to determine more complex sequence variability in defined loci have been described [Krebs et al., 2003; Stanssens et al., 2004; Ehrich et al., 2005]. The most evolved methods for DNA sequencing with MALDI–TOF MS detection use an initial PCR amplification with primers that contain transcription start sites at their 5' ends. After transcription of the PCR product, sequence-specific RNA endonucleases are used to generate fragments terminating on a specific base. Sizing the fragments by MALDI–TOF MS allows the identification of the sequence differences relative to a reference. The desire for high throughput, parallel processing, simplified handling, and low cost reagents have influenced the development of these methods.

We had previously described a novel, rapid, and accurate concept for DNA genotyping [Mauger et al., 2006] and sequencing [Mauger et al., 2007] using MS and RNA/DNA chimeras. Using PCR products as templates, these methods are based on preparing and alkali fragmenting chimeric RNA/DNA-cycled extension products generated by a primer with a DNA polymerase that also incorporates ribonucleotides. For sequencing hypervariable region (HVI), both HVI and HVII of the human mitochondrial genome were first amplified in duplex [Kline et al., 2005]. We then took advantage of a novel class of thermostable DNA polymerases that efficiently incorporate ribonucleotides and rapidly extend a primer to generate chimeric primer extension products [Mauger et al., 2007; Mauger et al., 2006]. Due to the formulation of the nucleotide mixture, these products contain three deoxyribonucleotides and the fourth nucleotide in its ribo form (A, C, or G). The chimeric products were sequence-specifically fragmented by treatment with alkali, generating oligonucleotide fragments each terminating with the ribonucleotide of the cycled primer extension reaction. Masses of the fragments were determined by MS. Analyzing the differentially substituted (e.g., A, C, and G) extension products, we were able to unambiguously determine the sequence of the original locus.

Here, we have developed the use of ribo-incorporating DNA polymerases further for ribo-PCR using genomic DNA as template, which is dramatically more complex than the mitochondrial genome and, at the same time, read a forward and reverse trace in one reaction due to the PCR mode. We have applied it to sequence three fragments in different genomic loci (*SLCO1B1**604,843; *H19**103,280; and nitric oxide synthase 1 [*NOS1*]*163,731). Variants in *SLCO1B1* have been associated with a dramatically increased risk to develop myopathies when treated with the cholesterol-lowering drug, Simvastatin [Link et al., 2008]. The *H19* locus is an imprinted locus, which is involved in growth [Edwards and Ferguson-Smith, 2007]. *NOS1* variants have been associated with diseases such as cancer and chronic disorders [Kanwar et al., 2009]. The interest of applying a sequencing method rather than a genotyping method in this context is that sequencing will identify unknown sequence anomalies while at the same time identifying already described variants.

Materials and Methods

Thirty Caucasian, randomly selected DNA samples, and 95 DNA samples with a history of myocardial infarction were used [Link et al., 2008]. Primers were purchased from Eurogentec (Angers, France). Deoxynucleoside triphosphates (N = A, C, G, T) and ribonucleoside triphosphates, NTPs, (N = A, C, G, U) were from GE Healthcare (Saclay, France). Ion-exchange resin AG 50W-X8 was from BioRad (Marnes la Coquette, France), and MassArray Clean Resin Tool Kit was purchased from Sequenom (Hamburg, Germany). General chemical reagents were from Aldrich (Steinheim, Germany). Thermocycling procedures were carried out in Eppendorf Gradient Thermocyclers (Eppendorf, Hamburg, Germany).

PCR primers are prepared with 2'OMe modification on the minus one terminal sugar from the 3' end. This modification inhibits 3'-5' exonuclease digestion of the primer [Mauger et al., 2007; Schönbrunner et al., 2006] by the DNA polymerase due to proofreading activity.

KB17 DNA polymerase [Mauger et al., 2007] was developed by Roche Molecular Systems (Pleasanton, CA). This designer DNA polymerase is a chimeric DNA polymerase comprising the 5'-nuclease domain of *Thermus* sp. Z05 DNA polymerase, and the 3'-nuclease and 5' to 3' DNA polymerase domains of *Thermotoga maritima* (*Tma*) DNA polymerase. The 5'-nuclease activity is eliminated in this enzyme by the introduced G46E mutation [Mauger et al., 2006]. The wild-type *Tma* proofreading activity has been modulated in KB17 DNA polymerase by introduction of an L329A mutation [Schönbrunner et al., 2006] in "motif I" of the 3' to 5' exonuclease domain. The enzyme's polymerase domain contains a mutation that eliminates the selectivity of the wild-type enzyme against incorporation of C'-2-substituted nucleotides and which facilitates efficient incorporation of NTPs. Finally, the enzyme contains three additional mutations in the polymerase domain, which enhance binding to template and result in a faster extension rate, particularly in high ribonucleotide substitution reactions (please contact Thomas Myers [Thomas.Myers@Roche.com] at Roche Molecular Systems for research samples).

Ribo-PCR amplifications were carried out in 20 μ l reactions, with 2 μ l of 10 ng/ μ l human genomic DNA, 0.8 μ l of 10 μ M of each primer (TTGATGGCTCACCATTGAAA and

GGCTAGTGGTGA-GGATGGAA for *NOS1*,
GCCAGTGAGGAGTGTGGAGTA and
TTCTTTCCAGCCCTAGCTCA for *H19* or
CAGGATGATTCTGGCTTCAAA, and
GCCAAATTCTACTGGAAATACAAAA for
SLCO1B1), 4 μ l of 500 mM tricine at pH 8.0
adjusted with potassium hydroxide, 4 μ l of 500 mM
potassium acetate at pH 7.5, 2 μ l of 30 mM
magnesium acetate, 1 μ l of 4 mM each of a mixture
of deoxyguanosine triphosphate (dGTP),
deoxycytidine triphosphate (dCTP), 2'-
deoxythymidine triphosphate (dTTP), and adenosine
triphosphate (ATP); or 2'-deoxyadenosine
triphosphate (dATP), dCTP, dTTP, and guanosine
triphosphate (GTP); or dATP, dGTP, dTTP, and
cytidine triphosphate (CTP); or dATP, dCTP, dGTP,
and uridine-5'-triphosphate (UTP); and 0.5 or 1 μ l for
duplex of 10 U/ μ l KB17 DNA polymerase. The
thermal cycling profile for the PCR was 4 min at
92°C, followed by 46 cycles of 15 sec at 92°C, 5 min
at 62°C. An aliquot (5 μ l) of the PCR was analyzed
on 2% agarose gel to verify that the desired PCR
products were generated.

For alkali cleavage, 5.0 μ l of 1.2 M sodium hydroxide
was added to the 15 μ l amplification reaction for a
final concentration of 0.3 M and incubated at 70°C
for 1.5 hr.

Samples were desalted by the addition of cation-
exchange resin charged with H⁺. Resin (6 mg) was
added to the reaction with the MassArray[®] Clean
Resin Tool Kit (Sequenom) and incubated for 20 min
at room temperature under agitation. Thereafter, the
sample was centrifuged for 2 min at 134 \times g to
sediment the resin.

Trihydroxyacetophenone (THAP) was used as matrix
[Zhu et al., 1996]. For preparation, 0.5 μ l of 0.2 M
each of 2,4,6 and 2,3,4 THAP in 50% acetonitrile,
and 0.3 M of ammonium citrate in water in 6/3/2
(v/v) was deposited on an anchor position of MALDI
target plate (AnchorChip[™] Target with a spot size of
400 μ m; Bruker Daltonik GmbH, Bremen, Germany).
After 0.5 μ l of the supernatant was added, the target
was dried at room temperature.

The target was introduced into a MALDI-TOF mass
spectrometer (Autoflex; Bruker Daltonik GmbH) for
analysis. Analysis was carried out in negative-ion
mode, with an acceleration voltage of 20 kV using
pulsed ion extraction delayed by 150 ns in linear
mode, and with external calibration. Each spectrum
obtained was the sum of 200 laser shots.

We created software that aided in primer design and
with the interpretation of results. The software has
been modified from an earlier version [Mauger et al.,
2007] to account for fragmentation of ribo-PCR
product rather than a cycled single stranded DNA
primer extension product. Using the reference
sequence, a theoretical peak profile is established.
This is then overlaid with the actual spectrum. The
treatment of the actual spectrum is limited to a
baseline correction using an adaptation of the Convex
Hull algorithm. No smoothing is applied to prevent
any loss of information. The peak recognition is
simplified because with the theoretical spectrum, the
expected masses of the peaks are known. Differences
of the measured and theoretical spectrum are used to
call sequences. Software to carry out similar tasks is
available from Bruker Daltonik (FlexAnalysis) or
Sequenom (www.sequenom.com).

Results

In this work, we used a novel thermostable DNA
polymerase that efficiently incorporated and extended
ribonucleotides for ribo-PCR using a genomic
template. A genomic DNA template, two locus
specific primers and a nucleotide mix with one of the
four nucleotide triphosphates of regular PCR
substituted by a ribo-nucleotide triphosphate is used
to generate a chimeric DNA/RNA ribo-PCR product.
Upon simple alkali treatment, a fragment pattern is
generated that allows unambiguous identification of
polymorphisms by MALDI-TOF MS. The
polymorphic position is identified by a
complementary mass fragment change in the ribo-
PCR product. Genotyping results at *NOS1* and *H19*
loci were 100% confirmed by genotyping in an
Amplifluor[®] assay. Genotyping results at the
SLCO1B1 loci were 100% confirmed in the
Illumina[®] Infinium[™] II assay [Link et al., 2008].

As previously reported [Mauger et al., 2007], the
DNA polymerase can readily incorporate up to 100%
ATP, 100% CTP, or 100% GTP in primer extension.
Here, we achieved the preparation of ribo-PCR
product with 100% ATP, 100% CTP, 100% GTP, or
100% UTP on genomic DNA template. The product
is an RNA/DNA chimeric PCR product of about 100
base pairs (bp). For *NOS1* (Supp. Table S1A), ribo-
PCR product was generated with 100% ATP. For
H19, ribo-PCR products with 100% ATP (Supp.
Table S1B) or 100% UTP (Supp. Table S1C) were
generated. For *SLCO1B1*, ribo-PCR products with
100% CTP (Supp. Table S1D) or 100% GTP (Supp.
Table S1E) were generated. The ribo-PCR works well

with homopolymeric stretches of three or four bases of the ribonucleotide utilized. However, longer homopolymeric runs may cause a drop in the efficiency of ribo-PCR.

Fragments for the MS analysis are generated by simple alkali cleavage with sodium hydroxide. This results in the cleavage 3' of each incorporated ribonucleotide. Under these conditioning and analysis methods, a residual 2' or 3' phosphate group remains at the 3' terminal base of each fragment. From the mass, we cannot distinguish 2'-H₂PO₄ + 3'-OH and 2'-OH + 3'-H₂PO₄. Quite likely, products are a mixture of these two species. Predicted fragments are listed in Tables S1A–S1E.

Due to the negative charges on the sugar phosphate backbone, adducts of the phosphate groups with Na⁺ are observed when samples are not sufficiently desalted. The most common method for desalting, the addition of an ion-exchange resin H⁺ with the MassArray[®] Clean Resin Tool Kit (Sequenom; www.sequenom.com), was used here.

Due to the mass distributions (Supp. Table S1), samples were analyzed in linear negative-ion mode, and the combined forward and reverse information was used for sequence assignment and resequencing.

This ribo-PCR procedure combined with alkali fragmentation and MS analysis provides a DNA resequencing protocol with a single enzymatic step and three reagent addition steps to the genomic DNA template (ribo-PCR mix, alkali solution, and ion-exchange resin). To our knowledge, this constitutes the simplest DNA sequencing protocol. The ribo-PCR resequencing was applied to different genomic loci in *NOS1* and *H19* in 30 individuals, and *SLCO1B1* in 95 individuals.

In one and two base fragment range, many fragments have identical mass and cannot be distinguished by this analysis. In the high base range (>20 bases), only few fragments have the same masses, but are difficult to detect.

For *SLCO1B1*, two ribo-PCRs are possible: with CTP and GTP. With CTP, a polymorphism (rs4363657) is identified by two fragments (Supp. Table S1D)—11,527.5 Da for allele T and 11,543.5 Da for allele C. In this mass range, mass difference of 16 Da is beyond the resolving power of the instrument, and the polymorphism cannot be detected by MALDI-TOF MS. However, with GTP, the

polymorphism rs4363657 is identified with two readily differentiated fragments (Supp. Table S1E), with masses 4,869.1 and 4,884.2 Da for allele C and for allele T, respectively.

For illustration, we show ribo-PCR resequencing of a portion of *NOS1* that involves rs3741476 (Supp. Fig. S1). Supp. Figure S1 shows resequencing with the interpretation of the spectra. The sequence is completely determined with the forward and reverse fragments in an ATP ribo-PCR.

A few fragments are not found in the mass spectra shown in Supp. Figure S1: two GCA fragments, two GCCA fragments, and one CCGA fragment are missing. These fragments do not have a thymidine base. Furthermore, in ATP ribo-PCR of *H19* (Supp. Table S1B), a GGGCGA fragment is not detected, and for GTP *SLCO1B1* ribo-PCR (Supp. Table S1E), two AAG fragments, one CCAG fragment, and two AAAG fragments are not present either.

It seems that thymidine-containing fragments are easily detected, while the ones without thymidine are more difficult or impossible to detect under these conditions. This may be because of fragmentation of the product. We switched the ion-extraction mode to the positive-ion mode and 3-hydroxypicolinic acid matrix, and still did not detect these species. Due to the coverage of both strands, however, the sequence is completely interpretable for all systems investigated here.

For UTP ribo-PCR of *H19* (Supp. Table S1C), all fragments are detected in the mass spectra, but there are also some additional peaks that stem from depurination of an A or G base [Gut, 1997]. As the number of bases increases, the depurinated peak becomes more intense than the expected peak. All fragments containing U base can be detected, like the thymidine-containing fragments; however, fragments having only one U base at the 3' end position seems to partially prevent the fragmentation of the nearest bases. Accordingly, for ribo-PCR of *H19*, mass spectra of ribo-PCR fragments with UTP are more difficult to analyze than that with ATP.

The combined forward and reverse information of the ribo-PCR resequencing allows all sequence differences to be distinguished (Supp. Table S1) and determined. The *NOS1* and *H19* ribo-PCRs were analyzed with this method and determined for 30 individuals (Table 1). For rs3741476 of *NOS1*, 12

individuals are homozygous A (40%), three individuals are homozygous G (10%), and 15 individuals are heterozygous R (50%). For rs2839698 of *H19*, four individuals are homozygous C (13%), four individuals are homozygous T (13%), and 22 individuals are heterozygous Y (73%). The rs4363657 of the *SLCO1B1* ribo-PCR was analyzed in 95 individuals (Table 2): 53 individuals were homozygous T (56%), 12 individuals were homozygous C (13%), and 30 individuals were heterozygous Y (31%). No previously unknown polymorphisms were detected in the samples sequenced here.

Figure 1 shows mass spectra with a peak at 4,221.7 Da of fragment GCCCCCTTCTTTCA and a peak at 4,236.7 Da of fragment GCCCCCTTTTTTCA that correspond to the G and A alleles of the polymorphism rs3741476 of ATP ribo-PCR of *NOS1*. The homozygote A (Fig. 1A), the homozygote G (Fig. 1B), and the heterozygote R (Fig. 1C) can be determined.

Figure 2 shows mass spectra with a peak at 3,140.0 Da of fragment TGCCTGGGCA and a peak at 4,351.8 Da of fragment TGCCTGGGCGCCTA that correspond to the T and C alleles of the polymorphism rs2839698 of ATP ribo-PCR of *H19*. The homozygote T (Fig. 2A), the heterozygote Y (Fig. 2B), and the homozygote C (Fig. 2C) can be determined.

Figure 3 shows mass spectra of a duplex ATP ribo-PCR of *NOS1* and *H19*. A peak at 4,221.7 Da of fragment GCCCCCTTCTTTCA and a peak at 4,236.7 Da of fragment GCCCCCTTTTTTCA correspond to the G and A alleles of the polymorphism rs3741476 of ATP ribo-PCR of *NOS1*. A peak at 3,140.0 Da of fragment TGCCTGGGCA and a peak at 4,351.8 Da of fragment TGCCTGGGCGCCTA correspond to the T and C alleles of the polymorphism rs2839698 of ATP ribo-PCR of *H19*. A peak at 3,419.2 Da of fragment GTCCGTGTCTA and a peak at 3,509.2 Da of fragment GGGGGCTGCTA correspond to the sequence of *NOS1*. The homozygote G and the heterozygote Y (Fig. 3A), the homozygote A and the homozygote T (Fig. 3B), and the heterozygote R and the homozygote C (Fig. 3C) of *NOS1* and *H19* can be determined.

Supp. Figure S2 shows mass spectra with a peak at 2,492.6 Da of fragment GGGCACCU and a peak at 2,508.6 Da of fragment GGGCGCCU that correspond to the T and C alleles of the polymorphism rs2839698 of UTP ribo-PCR of *H19*. The homozygote C (Supp. Fig. S2A), the heterozygote Y (Supp. Fig. S2B), and the homozygote T (Supp. Fig. S2C) can be determined.

Figure 4 shows mass spectra with a peak at 4,869.1 Da of fragment TACCCACTTTTTTTG and a peak at 4,884.2 Da of fragment TACTCACTTTTTTTG that correspond to the C and T alleles of the polymorphism rs4363657 of GTP ribo-PCR of *SLCO1B1*. The heterozygote Y (Fig. 4A), the homozygote C (Fig. 4B), and the homozygote T (Fig. 4C) can be determined.

Discussion

Here, we describe ribo-PCR for the simple preparation of a chimeric RNA/DNA from a genomic DNA template. We apply ribo-PCR for DNA resequencing by fragmenting the product with alkali and using a MALDI-TOF mass spectrometer for detection by identifying deviations from the reference sequence mass fingerprint.

In contrast to other sequencing approaches with MALDI-TOF MS detection that use a modified T7 RNA polymerase [Hahner et al., 1997], which require the PCR primers to contain an RNA polymerase promoter sequence, this method uses a specially modified, thermostable DNA polymerase with sequence-specific primers that allows positioning of primers freely.

The number of steps is reduced, comparing with our previous MS sequencing protocol [Mauger et al., 2007]. There, a locus-specific PCR product was prepared and purified, followed by three separate, cycled linear extension reactions. In contrast, this is a one-step ribo-PCR with just one ribonucleotide and provides all of the necessary information to correctly resequence a region. The ribo-PCR was successful with 100% substitution of the dATP, dCTP, dGTP, or dTTP by ATP, CTP, GTP, or UTP, respectively. However, we found a sequence-specific preference for amplification. Essentially, GC-rich sequences were amplifiable with the ATP and/or UTP ribo-PCR,

while GC-poor sequences were amplified in GTP or CTP ribo-PCRs.

For the *NOS1* locus, the forward and reverse primer synthesis products are 21–22 and 31% A, respectively, with only a single consecutive run of four A's in the reverse primer synthesis product. In contrast, for *NOS1*, the forward and reverse primer synthesis products are 16–18 and 25% G, respectively, but with a run of five consecutive G's in the forward primer synthesis product. Similarly, for *NOS1*, the forward and reverse primer synthesis products were 32 and 21–22% C, respectively, and the reverse primer synthesis product contains a run of five consecutive Cs.

For the *H19* locus, the forward and reverse primer synthesis products are 24 and 12–14% A, respectively, with no consecutive run of A longer than three. In contrast, the forward and reverse primer synthesis products for *H19* are 41 and 22–24% G, respectively, and 22–23 and 43% C, respectively. For each of the primer synthesis products, there is a consecutive run of four G's or C's in one of the two primer synthesis products.

For the *SLCO1B1* locus, the forward and reverse primer synthesis products are 16 and 17–18% G, respectively, with at most one consecutive run of three G's in each strand. For *SLCO1B1*, the forward and reverse primer synthesis products are 16–17 and 16% C, respectively, with at most one run of three consecutive C's. In contrast, for *SLCO1B1*, the forward and reverse primer synthesis products are 29% and up to 34–35% A, respectively, and the reverse primer synthesis product contains a run of five A's. The percentage calculations do not take account of the primers, but takes account of the reverse complements that are generated in the polymerization reaction.

Thus, the success of ribo-PCR with KB17 DNA polymerase and a particular nucleotide may also depend on the absence of simple sequence repeats of four or more for the particular nucleotide in a primer synthesis product. This enzyme can be used for linear cycled extension, as previously reported [Mauger et al., 2007], and in ribo-PCR.

Products were analyzed by MALDI-TOF MS in linear mode. The linear mode allows covering a large mass range (1–10 kDa), but does not resolve different species with similar masses in the high mass range. In

addition, this procedure takes advantage of information on both DNA strands in the ribo-PCR amplicon. Assembling and merging the profiles for the individual reactions yielded unambiguous results for each of the sequenced samples. We were able to unambiguously identify sequence changes relative to a reference and assign genotypes correctly.

We have demonstrated that ribo-PCR works in the 100 bp region. A chimeric RNA/DNA PCR product is stable with ATP, CTP, GTP, or UTP. For UTP, the mass spectra are more difficult to analyze because of the depurination of fragments containing A or G [Gut, 1997]. As the sequence of both strands is analyzed simultaneously, each ribo-PCR permits sequence verification in excess of 200 bases. It may be possible to extend the range with further optimization; ribo-PCR can be multiplexed (Fig. 3) and thus used to analyze a compound of more than one genomic region in a single reaction (*NOS1* and *H19*).

Fragmentation uses facile, inexpensive and complete alkali cleavage rather than expensive, base-specific nucleases or other enzymes. The KB17 DNA polymerase efficiently and rapidly incorporates the provided NTPs; expensive nucleotide analogues (e.g., α -S-CTP, α -S-UTP, or fluorescently labelled nucleotides) are unnecessary. In contrast to other enzymes used in sequencing or resequencing, KB17 DNA polymerase contains a 3'–5' exonuclease or proofreading activity that assists in maintaining higher fidelity in ribo-PCR. Further, in contrast to most proofreading thermostable DNA polymerases, KB17 DNA polymerase is fully compatible with using dUTP in the PCRs [Schönbrunner et al., 2006], which allows it to be used together for carry-over contamination procedures with dUTP and uracil-N-glycosylase.

This method identifies known polymorphisms such as homozygous or heterozygous positions, and can also be used to detect unknown mutations in the amplified sequence. It constitutes another addition to the ever-growing arsenal of DNA analysis methods. Its main advantage over DNA sequencing methods using MS detection lies in the ease of operation and probably lower cost than the current methods. As such, the method presented here is not more complicated than a genotyping procedure. The application of a sequencing method rather than a genotyping method for loci of diagnostic importance allows the identification of previously unknown polymorphism. Chimeric RNA/DNA molecules might have other uses than the one demonstrated here.

References

- Carothers AM, Urlaub G, Mucha J, Grunberger D, Chasin LA. 1989. Point mutation analysis in a mammalian gene: rapid preparation of total RNA, PCR amplification of cDNA, and Taq sequencing by a novel method. *Biotechniques* 7:494–496, 498, 499.
- Edwards CA, Ferguson-Smith AC. 2007. Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol* 19:281–289.
- Ehrich M, Bocker S, van den Boom D. 2005. Multiplexed discovery of sequence polymorphisms using base-specific cleavage and MALDI–TOF MS. *Nucleic Acids Res* 33:e38.
- Gut I. 1997. Depurination of DNA and matrix-assisted laser desorption/ionization mass spectrometry. *Int J Mass Spectrom* 169:313–322.
- Hahner S, Ludemann HC, Kirpekar F, Nordhoff E, Roepstorff P, Galla HJ, Hillenkamp F. 1997. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI) of endonuclease digests of RNA. *Nucleic Acids Res* 25:1957–1964.
- Innis MA, Myambo KB, Gelfand DH, Brow MA. 1988. DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc Natl Acad Sci USA* 85:9436–9440.
- Kanwar JR, Kanwar RK, Burrow H, Baratchi S. 2009. Recent advances on the roles of NO in cancer and chronic inflammatory disorders. *Curr Med Chem* 16:2373–2394.
- Karas M, Hillenkamp F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60:2299–2301.
- Kline MC, Vallone PM, Redman JW, Duewer DL, Calloway CD, Butler JM. 2005. Mitochondrial DNA typing screens with control region and coding region SNPs. *J Forensic Sci* 50:377–385.
- Krebs S, Medugorac I, Seichter D, Forster M. 2003. RNaseCut: a MALDI mass spectrometry-based method for SNP discovery. *Nucleic Acids Res* 31:e37.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, and others. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
- Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, Gut I, Lathrop M, Collins R. 2008. SLCO1B1 variants and statin-induced myopathy—a genomewide study. *N Engl J Med* 359:789–799.
- Mauger F, Bauer K, Calloway CD, Semhoun J, Nishimoto T, Myers TW, Gelfand DH, Gut IG. 2007. DNA sequencing by MALDI-TOF MS using alkali cleavage of RNA/DNA chimeras. *Nucleic Acids Res* 35:e62.
- Mauger F, Jaunay O, Chamblain V, Reichert F, Bauer K, Gut IG, Gelfand DH. 2006. SNP genotyping using alkali cleavage of RNA/DNA chimeras and MALDI time-of-flight mass spectrometry. *Nucleic Acids Res* 34:e18.
- Nordhoff E, Kirpekar F, Roepstorff P. 1996. Mass spectrometry of nucleic acids. *Mass Spectrom Rev* 15:67–138.
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487–491.
- Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350–1354.
- Sauer S, Lechner D, Berlin K, Lehrach H, Escary JL, Fox N, Gut IG. 2000. A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acids Res* 28:E13.
- Schönbrunner NJ, Fiss EH, Budker O, Stoffel S, Sigua CL, Gelfand DH, Myers TW. 2006. Chimeric thermostable DNA polymerases with reverse transcriptase and attenuated 3'–5' exonuclease activity. *Biochemistry* 45:12786–12795.
- Stanssens P, Zabeau M, Meersseman G, Remes G, Gansemans Y, Storm N, Hartmer R, Honisch C, Rodi

CP, Böcker S, van den Boom D. 2004. High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Res* 14:126-133.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, and others. 2001. The sequence of the human genome. *Science* 291:1304-1351.

Zhu YF, Chung CN, Taranenko NI, Allman SL, Martin SA, Haff L, Chen CH. 1996. The study of 2,3,4-trihydroxyacetophenone and 2,4,6-trihydroxyacetophenone as matrices for DNA detection in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 10:383-388.

:

Table 1. Genotypes of 30 Genomic DNA Samples of *NOS1* rs3741476 and *H19* rs2839698 Determined by Ribo-PCR Resequencing

| Individual | <i>NOS1</i> rs3741476 | <i>H19</i> rs2839698 |
|-------------------|------------------------------|-----------------------------|
| 310003201 | AG | C |
| 310016201 | A | CT |
| 110013201 | A | CT |
| 210015201 | A | CT |
| 710180201 | AG | CT |
| 210131201 | AG | T |
| 710214201 | AG | CT |
| 310038201 | AG | CT |
| 710320201 | A | CT |
| 710315201 | G | CT |
| 310048201 | AG | CT |
| 410017201 | A | T |
| 710322201 | G | CT |
| 710313201 | AG | C |
| 210309201 | AG | CT |
| 210001201 | AG | CT |
| 210010201 | A | CT |
| 210111201 | A | T |
| 710173201 | A | CT |
| 210113201 | AG | CT |
| 210115201 | AG | C |
| 710246201 | G | CT |
| 710248201 | AG | CT |
| 210181201 | A | CT |
| 210175201 | A | CT |
| 210180201 | A | CT |
| 710281201 | A | C |
| 310050201 | AG | CT |
| 210187201 | AG | CT |
| 410127201 | AG | T |

Table 2. Genotypes of 95 Genomic DNA Samples of *SLCO1B1* rs4363657 Determined by Ribo-PCR Resequencing

Annexe 3 : Ribo-polymerase chain reaction-A facile method for the preparation of chimeric RNA/DNA applied to DNA sequencing

| Individual | SCLO1B1 rs4363657 | Individual | SCLO1B1 rs4363657 |
|------------|-------------------|------------|-------------------|
| A002324 | TT | B002S17 | TC |
| A00230Z | CC | B002S1Q | TT |
| A00231K | TC | B002S20 | TT |
| A0023VU | TC | B002S2M | TT |
| B000XFE | TT | B002S2Y | TC |
| B002RZD | TT | B002S38 | CC |
| B002S0T | TC | A0023VQ | TT |
| B002S1H | TT | A00231G | TT |
| B002S1V | TC | A00231S | TC |
| B002S24 | TC | B000USM | TT |
| B002S2T | CC | B002RZ4 | TT |
| B002S32 | TT | B002SDD | TT |
| A002327 | TC | B002S18 | TT |
| A0023SA | TC | B002S1R | TT |
| A00231N | TT | B002S21 | TC |
| A00234T | TC | B002S2Q | CC |
| B000XK5 | CC | B002S2Z | TC |
| B002SDB | TT | B002S39 | TC |
| B002S0U | TT | A00230S | TT |
| B002S1I | TT | A00231H | TT |
| B002S1W | TT | A00231U | CC |
| B002S26 | TT | B000UWG | TT |
| B002S2U | TC | B002RZ5 | TC |
| B002S34 | TT | B002SDE | TT |
| A002329 | TT | B002SDI | TC |
| A002318 | TT | B002S1S | TT |
| A00231O | TC | B002S22 | TT |
| A00234V | TC | B002S2R | TT |
| B002RYZ | TT | B002S30 | TT |
| B002SDC | TT | B002S3E | TT |
| B002SDH | TT | A00230W | TC |
| B002S1M | TT | A00231I | TT |
| B002S1X | TT | A0023VT | TT |
| B002S2J | TT | B000X8P | TC |
| B002S2W | TC | B002RZA | TC |
| B002S35 | TC | B002RZM | TC |
| A00232Z | CC | B002SDJ | CC |
| A00231D | TT | B002S1U | TT |
| B000UI3 | TC | B002S23 | TT |
| B002SD9 | TT | B002S2S | TC |
| B002RZG | CC | B002S31 | TT |
| B002S11 | TT | B002S3F | TC |
| B002S1P | TT | A00231E | CC |
| B002S1Y | TC | A00231Q | TT |
| B002S2K | TT | B000Y2Y | CC |
| B002S2X | TC | B002SDA | TC |
| B002S37 | TT | B002RZH | TT |
| A002331 | CC | | |

Figure 1. Mass spectra of ATP ribo-PCR of *NOS1* with the polymorphism rs3741476. The 14-mer region of the reverse fragment of A: a homozygote A, B: a homozygote G, and C: a heterozygote R.

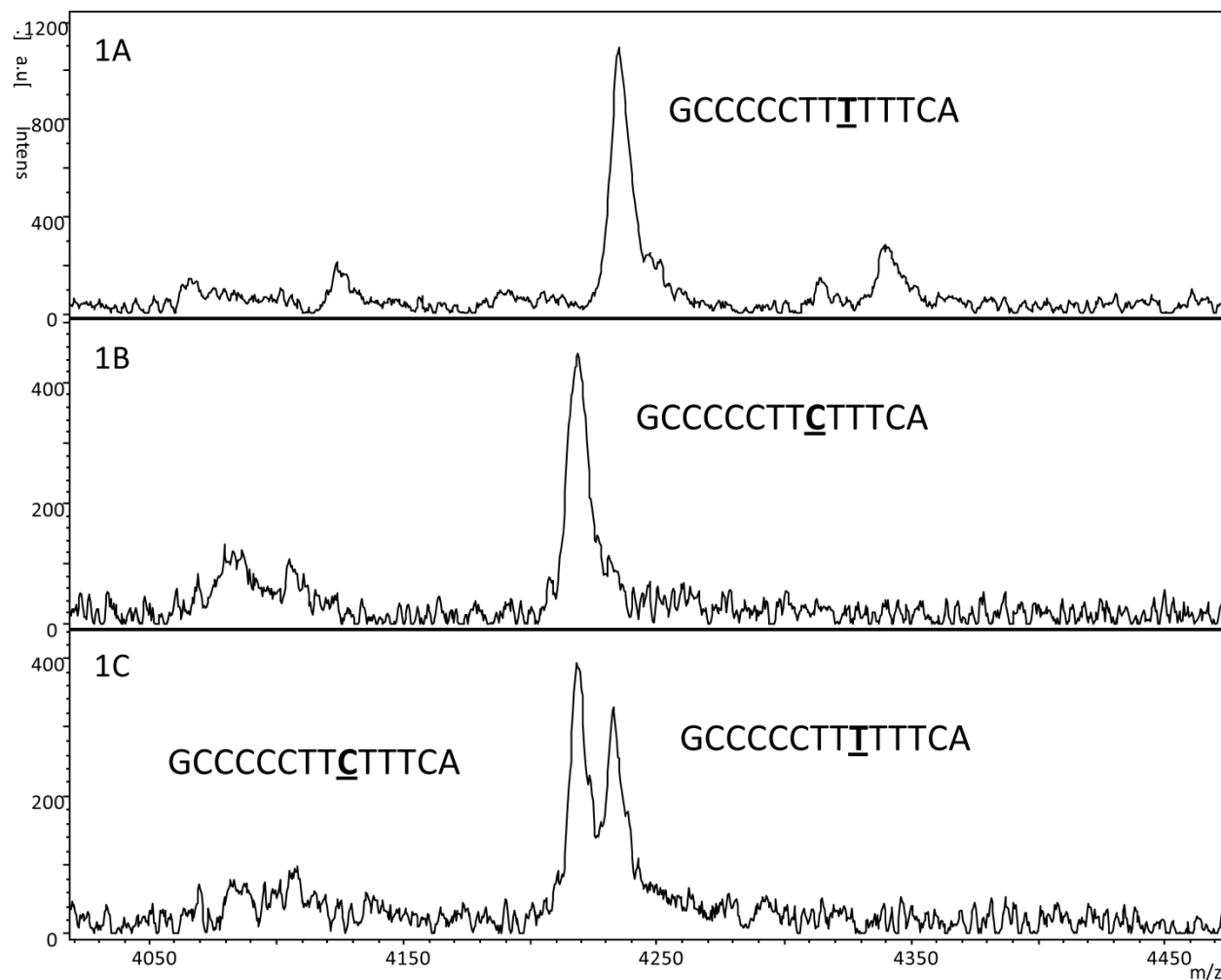
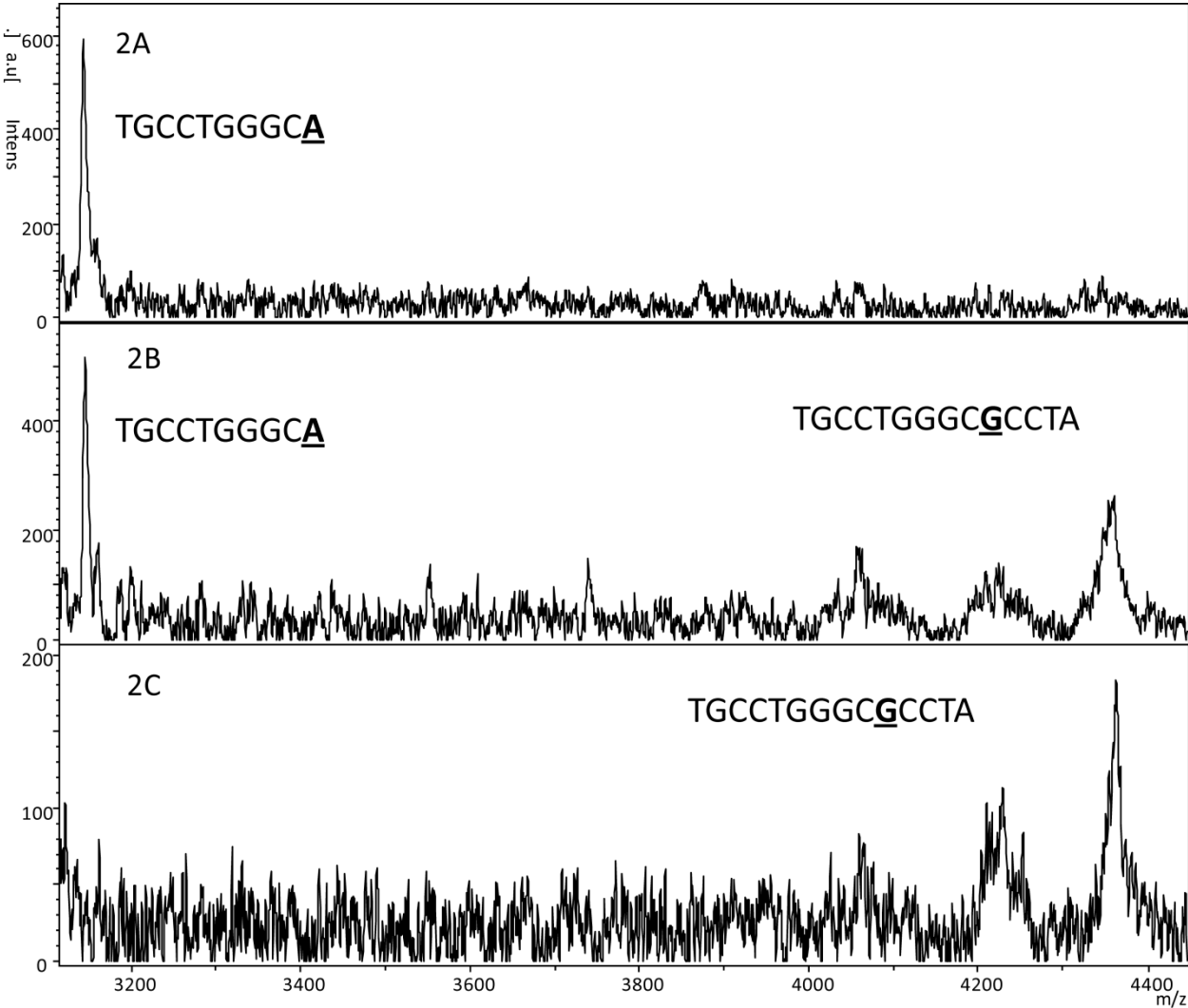


Figure 2. Mass spectra of ATP ribo-PCR of *H19* with the polymorphism rs2839698. The 10- to 14-mer region of the reverse fragment of A: a homozygote T, B: a heterozygote Y, and C: a homozygote C.

Annexe 3 : Ribo-polymerase chain reaction-A facile method for the preparation of chimeric RNA/DNA applied to DNA sequencing



Annexe 3 : Ribo-polymerase chain reaction-A facile method for the preparation of chimeric RNA/DNA applied to DNA sequencing

Figure 3. Mass spectra of ATP ribo-PCR of duplex *NOS1* with the polymorphism rs3741476 and *H19* with the polymorphism rs2839698. The 10- to 14-mer region of the reverse fragment of A: a homozygote G and a heterozygote Y; B: a homozygote A and a homozygote T; and C: a heterozygote R and a homozygote C.

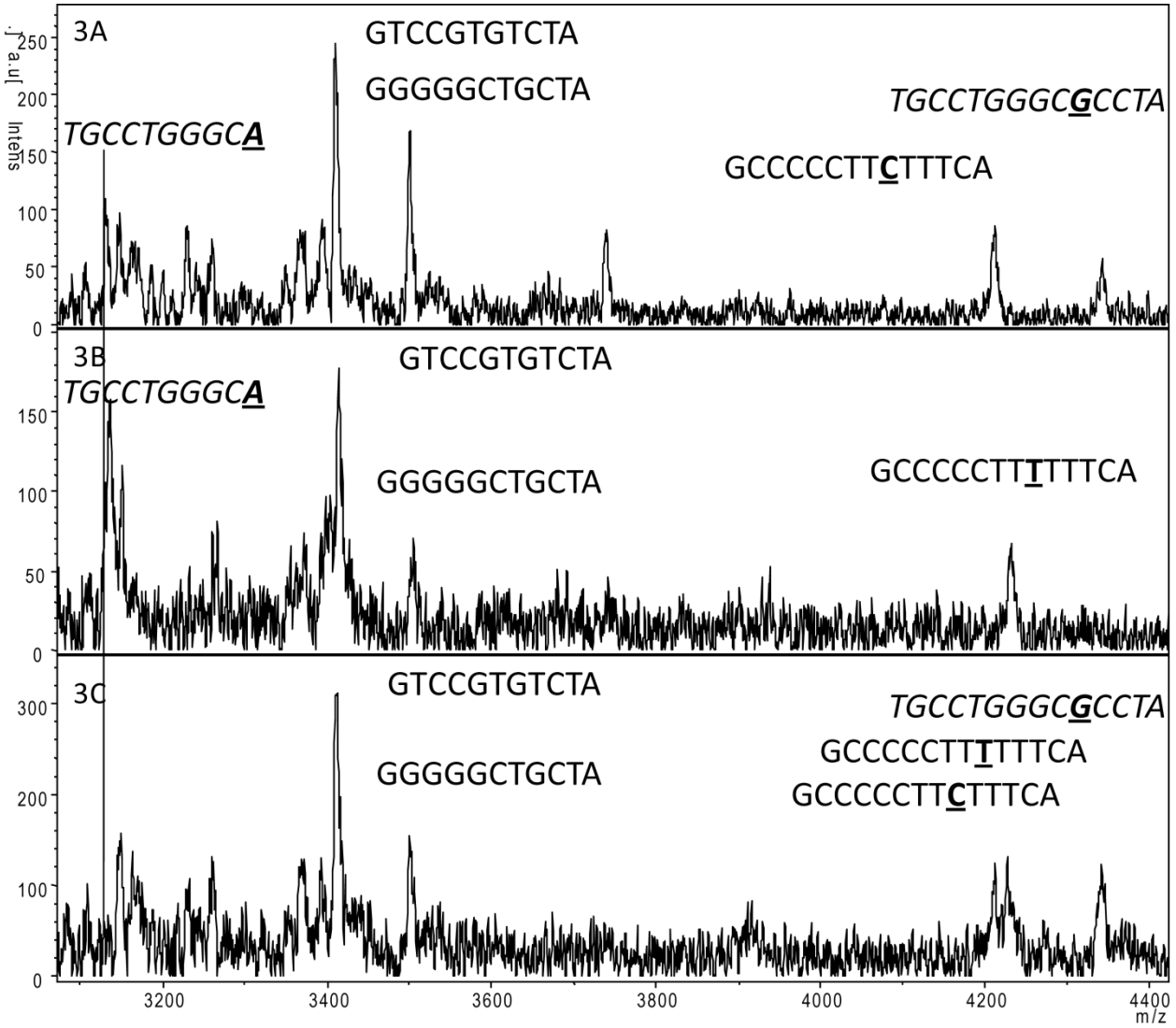
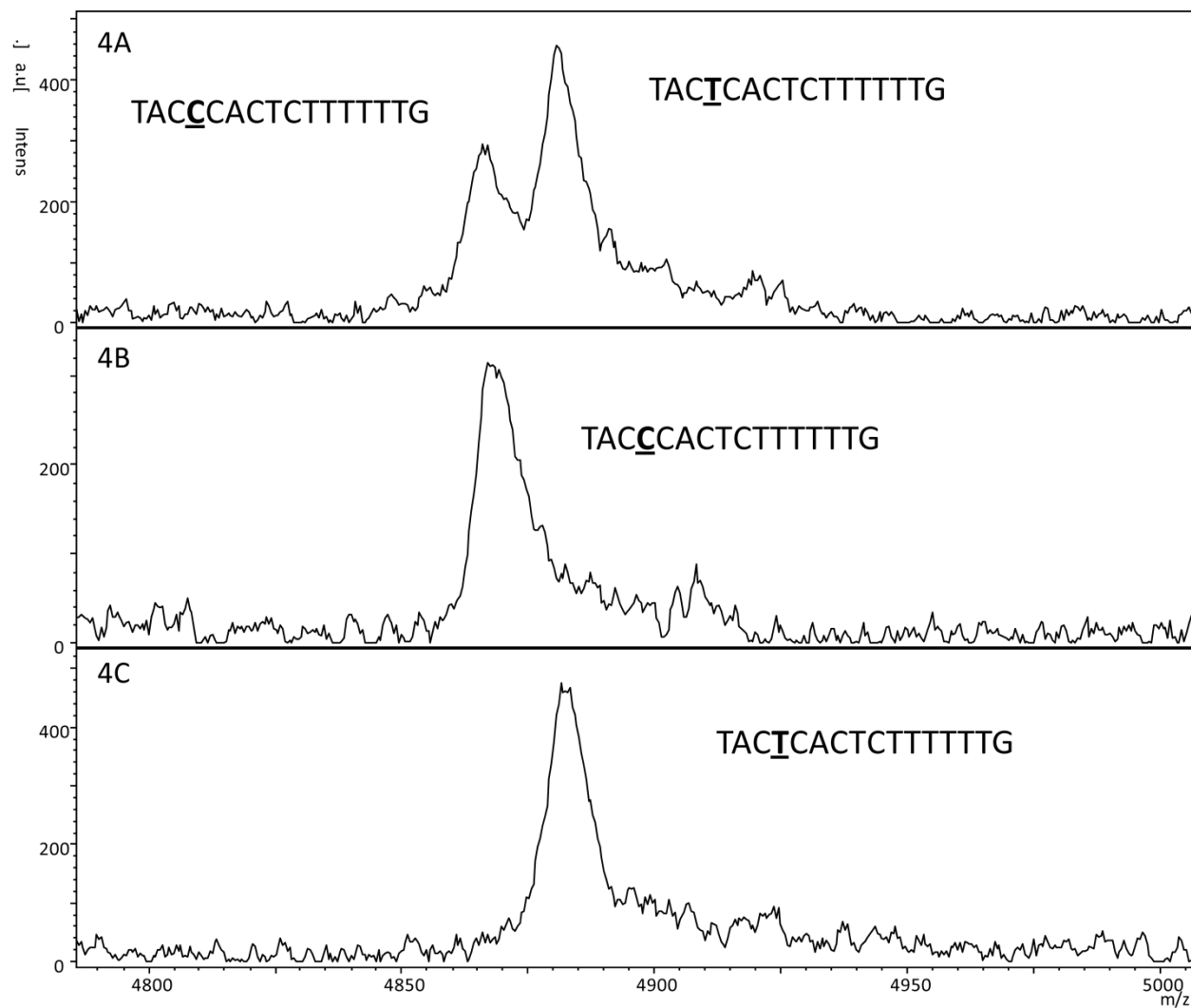


Figure 4. Mass spectra of GTP ribo-PCR of *SLCO1B1* with the polymorphism rs4363657. The 16-mer region of the forward fragment of A: a heterozygote Y, B: a homozygote C, and C: a homozygote T.



Annexe 4: RIBONUCLEOTIDE TAG NUCLEIC ACID DETECTION

Inventors: David H. Gelfand (Oakland, CA, US) Ivo Glynne Gut (Paris, FR) Keith A. Bauer (San Rafael, CA, US) Florence Mauger (Evry, FR)

Assignees: ROCHE MOLECULAR SYSTEMS, INC. Centre National de Genotypage

IPC8 Class: AC12Q168FI

USPC Class: 435 6

Class name: Involving nucleic acid

Publication date: 10/22/2009

Patent application number: 20090263813

Abstract:

The present application provides polynucleotides comprising 5'-tails with sequence segments useful for the detection of target nucleic acid sequences, and methods for their use in detecting target nucleic acids. The polynucleotides are used to amplify a subsequence of a target nucleic acid in the presence of one or more ribonucleotides. The ribonucleotides are incorporated into amplification products at regular intervals complementary to the 5'-tail sequence segments. Cleavage of amplification products at the bond immediately 3' to incorporated ribonucleotides produces detectably distinct fragments indicative of the presence or absence of a target nucleic acid.

Claims:

1. A method of detecting a target nucleic acid, the method comprising:(a) contacting the target nucleic acid with a polynucleotide, a nucleotide set and a nucleotide incorporating biocatalytic component, wherein;(i) the polynucleotide comprises a 5' portion and a 3' portion, the 5' portion comprising one or more contiguous sequence segments, wherein each sequence segment comprises at least three nucleotide bases, and the 5'-end nucleotide base of each sequence segment is unique within one sequence segment and is the same in each sequence segment; and the 3' portion comprising a sequence segment that is substantially complementary sufficient to hybridize to the target nucleic acid and be extended under amplification conditions;(ii) the nucleotide set comprises at least two nucleotide bases in the form of deoxyribonucleotides (dNTPs), and the majority of at least one nucleotide base in the form of a ribonucleotide (rNTP), wherein the base of the ribonucleotide is complementary to the unique 5' nucleotide base of each sequence segment of the polynucleotide; and(iii) the nucleotide incorporating biocatalytic component comprises deoxyribonucleotide and ribonucleotide incorporating activities;(b) amplifying the target nucleic acid under amplification conditions to produce an amplicon comprising a 3' sequence segment complementary to the 5'-portion of the polynucleotide, wherein the 3' sequence segment comprises one or more contiguous

sequence segments, wherein the 3'-end nucleotide base of each sequence segment is a nucleoside monophosphate (NMP) having the same base as the NTP in the nucleotide set;(c) cleaving the amplicon 3' of each NMP into fragments, wherein the cleaving releases the sequence segments as individual fragments; and(d) detecting the amplicon fragments, wherein the detected sequence segment fragments indicate amplification of the target nucleic acid sequence, thereby detecting the target nucleic acid sequence.

2. The method of claim 1, wherein each 5' portion sequence segment is of equal mass.

3. The method of claim 1, wherein at least 80% of at least one nucleotide base is in the form of a ribonucleotide (rNTP).

4. The method of claim 1, wherein the cleaving is performed by subjecting the amplicon to an alkaline solution.

5. The method of claim 4, wherein the alkaline solution comprises at least one of the following: NaOH, KOH or NH₄OH.

6. The method of claim 1, comprising contacting the target nucleic acid sequence with a polynucleotide pair, wherein each polynucleotide in the polynucleotide pair comprise 5' portion sequence

segments of equal mass.

7. The method of claim 1, wherein each sequence segment has the same nucleotide sequence.

8. The method of claim 1, comprising contacting the target nucleic acid with at least one polynucleotide, wherein the detection of the sequence segment fragments from the complement of the polynucleotide indicates the presence of an allele of a polymorphic nucleotide in the target nucleic acid sequence.

9. The method of claim 1, comprising contacting the target nucleic acid with at least two different polynucleotides, wherein the 3'-portion of the polynucleotides differ only at the 3'-end nucleotide base, wherein the 3'-end nucleotide base of each polynucleotide corresponds to a sequence segment of unique mass, and the detection of the sequence segment fragments from the complement of at least one of the polynucleotides indicates the presence of an allele of a polymorphic nucleotide in the target nucleic acid sequence.

10. The method of claim 1, wherein a plurality of different target nucleic acid sequences is contacted with a plurality of different polynucleotides, each of the different polynucleotides comprising sequence segments of a unique mass, wherein the sequence segments of the polynucleotides are identifiers of one of the plurality of different target nucleic acid sequences, whereby detection of the sequence segment fragments from the complement of at least one of the polynucleotides indicates the presence of at least one of the plurality of different target nucleic acid sequences.

11. The method of claim 1, wherein the nucleotide incorporating biocatalytic component comprises a single catalytic domain that comprises deoxyribonucleotide and ribonucleotide incorporating activities.

12. The method of claim 1, wherein the nucleotide incorporating biocatalytic component comprises first and second catalytic domains, wherein the first catalytic domain comprises deoxyribonucleotide incorporating activity and the second catalytic domain comprises ribonucleotide incorporating activity.

13. The method of claim 1, wherein the detecting is performed by mass spectrometry.

14. The method of claim 13, wherein the detection is by gas-phase ion mass spectrometry.

15. The method of claim 13, wherein the detection is by laser desorption-ionization mass spectrometry.

16. The method of claim 1, wherein the polynucleotide comprises a blocking moiety that substantially prevents extension of the polynucleotide, and wherein the method further comprises the step of removing the blocking moiety from the polynucleotide prior to amplifying.

17. A reaction mixture comprising: a) a first polynucleotide comprising a 5' portion and a 3' portion, wherein the 5' portion comprises one or more contiguous sequence segments, wherein each sequence segment comprises at least three nucleotide bases, and the 5'-end nucleotide base of each sequence segment is unique within one sequence segment and is the same in each sequence segment; and the 3' portion comprises a sequence segment that is substantially complementary sufficient to hybridize to a target nucleic acid and be extended under amplification conditions; b) a nucleotide set comprising at least two nucleotide bases in the form of deoxyribonucleotides (dNTPs), and the majority of at least one nucleotide base in the form of a ribonucleotide (rNTP), wherein the nucleotide base provided as a ribonucleotide is complementary to the unique 5' nucleotide base of each sequence segment; and c) a nucleotide incorporating biocatalytic component comprising deoxyribonucleotide and ribonucleotide incorporating activities.

18. The reaction mixture of claim 17, wherein each 5' portion sequence segment is of equal mass.

19. The reaction mixture of claim 17, wherein the 3'-end comprises a removable blocking moiety that substantially prevents extension of the polynucleotide

20. The reaction mixture of claim 19, wherein the 3'-end of the polynucleotide comprises a 2'-terminator moiety.

21. The reaction mixture of claim 17, wherein each sequence segment in the polynucleotide has the same nucleotide sequence.

22. The reaction mixture of claim 17, wherein the nucleotide incorporating biocatalytic component comprises a single catalytic domain that comprises deoxyribonucleotide and ribonucleotide incorporating activities.

23. The reaction mixture of claim 17, wherein the nucleotide incorporating biocatalytic component comprises first and second catalytic domains, wherein

the first catalytic domain comprises deoxyribonucleotide incorporating activity and the second catalytic domain comprises ribonucleotide incorporating activity.

- 24.** The reaction mixture of claim 17, wherein at least 80% of at least one nucleotide base is in the form of a ribonucleotide (rNTP).
- 25.** The reaction mixture of claim 17, further comprising at least one target nucleic acid sequence.
- 26.** The reaction mixture of claim 17, further comprising a second polynucleotide comprising a 5' portion and a 3' portion to form a polynucleotide pair, wherein each polynucleotide in the polynucleotide pair comprise 5' portion sequence segments of equal mass.
- 27.** The reaction mixture of claim 26, comprising a plurality of polynucleotide pairs, wherein each polynucleotide in each polynucleotide pair comprises 5' portion sequence segments of equal mass.
- 28.** The reaction mixture of claim 17, further comprising an amplicon comprising a 5' portion identical to the polynucleotide.
- 29.** A kit comprising: a) a first polynucleotide comprising a 5' portion and a 3' portion, wherein the 5' portion comprises one or more contiguous sequence segments, wherein each sequence segment comprises at least three nucleotide bases, and the 5'-end nucleotide base of each sequence segment is unique within one sequence segment and is the same in each sequence segment; and the 3' portion comprises a sequence segment that is substantially complementary sufficient to hybridize to a target nucleic acid and be extended under amplification conditions; b) a nucleotide set comprising at least two nucleotide bases in the form of deoxyribonucleotides (dNTPs), and the substantial majority of at least one nucleotide base in the form of a ribonucleotide (rNTP), wherein the nucleotide base provided as a ribonucleotide is complementary to the unique 5' nucleotide base of each sequence segment; and c) a nucleotide incorporating biocatalytic component comprising deoxyribonucleotide and ribonucleotide incorporating activities.
- 30.** The kit of claim 29, wherein each 5' portion sequence segment is of equal mass.
- 31.** The kit of claim 29, wherein the 5' portion sequence segment of the polynucleotide comprises at least 3 sequence segments.
- 32.** The kit of claim 31, wherein each sequence segment comprises at least 4 nucleotide bases.
- 33.** The kit of claim 29, wherein each sequence segment comprises at least 7 nucleotide bases.
- 34.** The kit of claim 29, wherein the 3'-end comprises a removable blocking moiety that substantially prevents extension of the polynucleotide.
- 35.** The kit of claim 34, wherein the 3'-end of the polynucleotide comprises a 2'-terminator moiety.
- 36.** The kit of claim 29, wherein each sequence segment in the polynucleotide has the same nucleotide sequence.
- 37.** The kit of claim 29, wherein the nucleotide incorporating biocatalytic component comprises a single catalytic domain that comprises deoxyribonucleotide and ribonucleotide incorporating activities.
- 38.** The kit of claim 29, wherein the nucleotide incorporating biocatalytic component comprises first and second catalytic domains, wherein the first catalytic domain comprises deoxyribonucleotide incorporating activity and the second catalytic domain comprises ribonucleotide incorporating activity.
- 39.** The kit of claim 29, wherein at least 80% of at least one nucleotide base is in the form of a ribonucleotide (rNTP).
- 40.** The kit of claim 29, further comprising a positive control target nucleic acid sequence.
- 41.** The kit of claim 29, further comprising a second polynucleotide comprising a 5' portion and a 3' portion to form a polynucleotide pair, wherein each polynucleotide in the polynucleotide pair comprise 5' portion sequence segments of equal mass.
- 42.** The kit of claim 29, comprising a plurality of polynucleotide pairs, wherein each polynucleotide in each polynucleotide pair comprise 5' portion sequence segments of equal mass.
- 43.** The kit of claim 29, further comprising an alkaline solution.
- 44.** An amplicon comprising a double-stranded polynucleotide comprising: a) a first strand comprising a 5' portion and a 3' portion, wherein the 5' portion comprises two or more contiguous

sequence segments, wherein each sequence segment comprises at least three nucleotide bases, is of equal mass, and the 5'-end nucleotide base of each sequence segment is unique within one sequence segment and is the same in each sequence segment; and the 3' portion comprises at least 5 nucleotides; and b) a second strand complementary to the first strand and comprising a 3' portion comprising a sequence segment comprising one or more contiguous sequence segments, wherein the 3'-end nucleotide base of each sequence segment is an NMP.

45. A system comprising at least one container or support comprising a composition comprising an amplicon of claim 44; b) at least one thermal modulator configured to thermally communicate with the container or the support; c) at least one reagent transfer component that transfers reagents to and/or from the container or the support; and, d) at least one detector configured to detect masses of one or more sequence segments produced in the container or on the support.

46. The system of claim 45, wherein the detector comprises a gas-phase ion spectrometer.

47. The system of claim 45, further comprising at least one controller operably connected to: the thermal modulator to effect modulation of the temperature in the container or on the support, the reagent transfer component to effect transfer of the reagents to and/or from the container or on the support, and/or, the detector to effect detection of the masses of the sequence segments produced in the container or on the support.

48. A polynucleotide comprising a 5' portion and a 3' portion, wherein a) the 5' portion comprises two or more contiguous sequence segments, wherein each sequence segment comprises at least three nucleotide bases, is of equal mass, and the 5'-end nucleotide base of each sequence segment is unique within one sequence segment and is the same in each sequence segment; and b) the 3' portion comprises at least 5 nucleotides, wherein the polynucleotide is less than 100 nucleotide bases in length.

49. The polynucleotide of claim 48, wherein each sequence segment has the same nucleotide sequence.

50. The polynucleotide of claim 48, wherein the 3'-end comprises a removable blocking moiety that substantially prevents extension of the polynucleotide.

51. The polynucleotide of claim 50, wherein the 3'-end comprises a 2'-terminator moiety.

Description:

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 61/046,720, filed on Apr. 21, 2008, the entire disclosure of which is hereby incorporated herein by reference.

FIELD OF THE INVENTION

[0002] The present invention relates to the field of nucleic acid detection. In particular, the present invention provides polynucleotides having multiple contiguous sequence segments, wherein the 5'-end nucleotide base of each sequence segment is unique within the sequence segment and is the same in each sequence segment, and methods for their use in detecting target nucleic acids.

BACKGROUND OF THE INVENTION

[0003] Many methods for detection of target nucleic acids (e.g., SNP genotyping) are known. Currently available homogeneous assays for SNP genotyping include the TAQMAN®, AMPLIFLUOR®, dye binding, allele-selective kinetic PCR, and SCORPION® primer assays. These assays provide one, or at maximum two (if using four different fluorescent dyes) SNPs per reaction well (e.g. dye-binding kinetic PCR requires two wells for one SNP). The available methods for SNP genotyping range from those that allow genotyping of a single SNP in a reaction well to methods that permit genotyping of many thousand SNPs in a single well (e.g., GOLDEN GATE® Assay, Illumina). Present genotyping assay procedures are not readily multiplexed due to the requirement for a different dye for each typed allele, and thus is limited in its potential for improvement. The GOLDEN GATE® assay requires complex analysis devices such as fiber optic array readers. Analyzers for reading of SNP genotyping range from plate readers (optionally, with an included PCR machine), to sequencers (e.g., capillary sequencers), array readers and mass spectrometers. Moreover, to the extent that mass spectrometers have been used for genotyping, currently available procedures do not allow for true multiplexing.

BRIEF SUMMARY OF THE INVENTION

[0004] The present invention provides polynucleotides

for the efficient and highly sensitive concurrent detection of one or more target nucleic acids. The polynucleotides are designed with a 5'-portion comprising one, two or more contiguous sequence segments, such that the mass of a sequence segment, or complement thereof, identifies a target nucleic acid or a target nucleotide within a target nucleic acid. In some embodiments, the sequence segments are of equal mass. Sequence segments of equal mass can, but need not have the same sequence. The sequence segments are cleaved from the complement of the polynucleotide and their mass is measured or detected to identify the presence, absence or level of presence of a target nucleic acid.

[0005]Accordingly, in a first aspect, the present invention provides a polynucleotide comprising one, two or more sequence segments. In some embodiments, the polynucleotide comprises a 5' portion and a 3' portion, wherein

[0006]a) the 5' portion comprises at least one or at least two contiguous sequence segments, wherein each sequence segment comprises at least three nucleotide bases, is of equal mass, and the 5'-end nucleotide base of each sequence segment is unique within one sequence segment and is the same in each sequence segment; and

[0007]b) the 3' portion comprises at least 5 nucleotides, wherein the polynucleotide is less than about 100 nucleotide bases in length. In some embodiments, the polynucleotide is less than about 75, 50 or 25 nucleotides in length.

[0008]In some embodiments, each sequence segment has the same nucleotide sequence.

[0009]In some embodiments, the 5' portion comprises at least 3, 4, 5, 6, 7 or 8 contiguous sequence segments. In some embodiments, each sequence segment comprises at least 3, 4, 5, 6, 7 or 8 nucleotides.

[0010]In some embodiments, the 3'-end comprises a removable blocking moiety that substantially prevents extension of the polynucleotide. In some embodiments, the 3'-end comprises a 2' terminator moiety.

[0011]In a further aspect, the invention provides methods of detecting a target nucleic acid. In some embodiments, the methods comprise:

[0012](a) contacting the target nucleic acid with a polynucleotide, a nucleotide set and a nucleotide

incorporating biocatalytic component, wherein; **[0013]**(i) the polynucleotide comprises a 5' portion and a 3' portion, the 5' portion comprising at least one or at least two contiguous sequence segments, wherein each sequence segment comprises at least three nucleotide bases, and the 5'-end nucleotide base of each sequence segment is unique within one sequence segment and is the same in each sequence segment; and the 3' portion comprising a sequence segment that is substantially complementary sufficient to hybridize to the target nucleic acid and be extended under amplification conditions; **[0014]**(ii) the nucleotide set comprises at least two nucleotide bases in the form of deoxyribonucleotides (dNTPs), and the majority of at least one nucleotide base in the form of a ribonucleotide (rNTP), wherein the base of the ribonucleotide is complementary to the unique 5' nucleotide base of each sequence segment of the polynucleotide; and **[0015]**(iii) the nucleotide incorporating biocatalytic component comprises deoxyribonucleotide and ribonucleotide incorporating activities;

[0016](b) amplifying the target nucleic acid under amplification conditions to produce an amplicon comprising a 3' sequence segment (i.e., 3'-portion) complementary to the 5'-portion of the polynucleotide, wherein the 3' sequence segment comprises at least one or at least two sequence segments, wherein the 3' end nucleotide base of each sequence segment is a nucleoside monophosphate (NMP) having the same base as the NTP in the nucleotide set;

[0017](c) cleaving the amplicon 3' of each NMP into fragments, wherein the cleaving releases the sequence segments as individual fragments; and

[0018](d) detecting the amplicon fragments, wherein the detected sequence segment fragments indicate amplification of the target nucleic acid sequence, thereby detecting the target nucleic acid sequence.

[0019]In some embodiments, the 5' portion sequence segments are of equal mass.

[0020]In some embodiments, at least 80% of at least one nucleotide base is in the form of a ribonucleotide (rNTP).

[0021]In some embodiments, the cleaving is performed by subjecting the amplicon to an alkaline solution.

[0022]In some embodiments, the alkaline solution comprises at least one of the following: NaOH, KOH,

RbOH, Mg(OH)₂, Ca(OH)₂, or NH₄OH. In some embodiments, the alkaline solution comprises at least one of the following: NaOH, KOH or NH₄OH.

[0023]In some embodiments, the methods further comprise the step of contacting the target nucleic acid sequence with a polynucleotide pair, wherein each polynucleotide in the polynucleotide pair comprise 5' portion sequence segments of equal mass.

[0024]In some embodiments, the methods further comprise the step of contacting the target nucleic acid with at least one polynucleotide, wherein the detection of the sequence segment fragments from the polynucleotide, or its complement, indicates the presence of an allele of a polymorphic nucleotide in the target nucleic acid sequence.

[0025]In some embodiments, the methods further comprise the step of contacting the target nucleic acid with at least two different polynucleotides, wherein the polynucleotides differ only at the 3'-end nucleotide base, wherein the 3'-end nucleotide base of each polynucleotide corresponds to a sequence segment of unique mass, and the detection of the sequence segment fragments from at least one of the polynucleotides, or its complement, indicates the presence of an allele of a polymorphic nucleotide in the target nucleic acid sequence.

[0026]In some embodiments, a plurality of different target nucleic acid sequences is contacted with a plurality of different polynucleotides, each of the different polynucleotides comprising sequence segments of a unique mass, wherein the sequence segments of the polynucleotides are identifiers of one of the plurality of different target nucleic acid sequences, whereby detection of the sequence segment fragments from at least one of the polynucleotides, or its complement, indicates the presence of at least one of the plurality of different target nucleic acid sequences.

[0027]In some embodiments, the nucleotide incorporating biocatalytic component comprises a single catalytic domain that comprises deoxyribonucleotide and ribonucleotide incorporating activities. In some embodiments, the nucleotide incorporating biocatalytic component comprises first and second catalytic domains, wherein the first catalytic domain comprises deoxyribonucleotide incorporating activity and the second catalytic domain comprises ribonucleotide incorporating activity.

[0028]In some embodiments, the detecting is performed by mass spectrometry. In some

embodiments, the detection is by gas-phase ion mass spectrometry. In some embodiments, the detection is by laser desorption-ionization mass spectrometry.

[0029]In some embodiments, the polynucleotide comprises a blocking moiety that substantially prevents extension of the polynucleotide, and wherein the method further comprises the step of removing the blocking moiety from the polynucleotide prior to amplifying.

[0030]In another aspect, the invention provides reaction mixtures. In some embodiments, the reaction mixtures comprise:

[0031]a) a first polynucleotide comprising a 5' portion and a 3' portion, wherein the 5' portion comprises at least one or at least two contiguous sequence segments, wherein each sequence segment comprises at least three nucleotide bases, and the 5'-end nucleotide base of each sequence segment is unique within one sequence segment and is the same in each sequence segment; and the 3' portion comprises a sequence segment that is substantially complementary sufficient to hybridize to a target nucleic acid and be extended under amplification conditions;

[0032]b) a nucleotide set comprising at least two nucleotide bases in the form of deoxyribonucleotides (dNTPs), and the majority of at least one nucleotide base in the form of a ribonucleotide (rNTP), wherein the nucleotide base provided as a ribonucleotide is complementary to the unique 5' nucleotide base of each sequence segment; and

[0033]c) a nucleotide incorporating biocatalytic component comprising deoxyribonucleotide and ribonucleotide incorporating activities.

[0034]In a related aspect, the invention provides kits. In some embodiments, the kits comprise:

[0035]a) a first polynucleotide comprising a 5' portion and a 3' portion, wherein the 5' portion comprises at least one or at least two contiguous sequence segments, wherein each sequence segment comprises at least three nucleotide bases, and the 5'-end nucleotide base of each sequence segment is unique within one sequence segment and is the same in each sequence segment; and the 3' portion comprises a sequence segment that is substantially complementary sufficient to hybridize to a target nucleic acid and be extended under amplification conditions;

[0036]b) a nucleotide set comprising at least two nucleotide bases in the form of deoxyribonucleotides

(dNTPs), and the substantial majority of at least one nucleotide base in the form of a ribonucleotide (rNTP), wherein the nucleotide base provided as a ribonucleotide is complementary to the unique 5' nucleotide base of each sequence segment; and

[0037]c) a nucleotide incorporating biocatalytic component comprising deoxyribonucleotide and ribonucleotide incorporating activities.

[0038]Embodiments of the reaction mixtures and kits are as described above for the polynucleotides and the methods, and herein.

[0039]In a further aspect, the invention provides amplicons. In some embodiments, the amplicons comprise a double-stranded polynucleotide comprising:

[0040]a) a first strand comprising a 5' portion and a 3' portion, wherein the 5' portion comprises at least one or at least two contiguous sequence segments, wherein each sequence segment comprises at least three nucleotide bases, is of equal mass, and the 5'-end nucleotide base of each sequence segment is unique within one sequence segment and is the same in each sequence segment; and the 3' portion comprises at least 5 nucleotides; and

[0041]b) a second strand complementary to the first strand and comprising a 3' portion comprising a sequence segment comprising at least one or at least two contiguous sequence segments, wherein the 3'-end nucleotide base of each sequence segment is an NMP.

[0042]In another aspect, the invention provides systems. In some embodiments, the systems comprise at least one container or support comprising:

[0043]a) a composition comprising an amplicon of the invention;

[0044]b) at least one thermal modulator configured to thermally communicate with the container or the support to modulate temperature in the container or on the support;

[0045]c) at least one reagent transfer component that transfers reagents to and/or from the container or the support; and,

[0046]d) at least one detector configured to detect masses of one or more sequence segments produced in the container or on the support.

[0047]In some embodiments, the detector comprises a gas-phase ion spectrometer.

[0048]In some embodiments, the systems comprise at least one controller operably connected to:

[0049]the thermal modulator to effect modulation of the temperature in the container or on the support,

[0050]the reagent transfer component to effect transfer of the reagents to and/or from the container or on the support, and/or,

[0051]the detector to effect detection of the masses of the sequence segments produced in the container or on the support.

DEFINITIONS

[0052]The terms "nucleic acid" or "polynucleotide" apply interchangeably to a polymer that corresponds to a ribose nucleic acid (RNA) or deoxyribose nucleic acid (DNA) polymer, or an analog thereof. This includes polymers of nucleotides such as RNA and DNA, as well as modified forms thereof, peptide nucleic acids (PNAs), locked nucleic acids (LNA®), and the like. In certain embodiments, a nucleic acid can be a polymer that includes multiple monomer types, e.g., both RNA and DNA subunits. A nucleic acid can be or can include, e.g., a chromosome or chromosomal segment, a vector (e.g., an expression vector), an expression cassette, a naked DNA or RNA polymer, the product of a polymerase chain reaction (PCR), an oligonucleotide, a probe, a primer, etc. A nucleic acid can be, e.g., single-stranded, double-stranded, triple-stranded, etc and is not limited to any particular length. Unless otherwise indicated, a particular nucleic acid sequence optionally comprises or encodes complementary sequences, in addition to any sequence explicitly indicated.

[0053]Nucleic acids are not limited to molecules having naturally occurring polynucleotide sequences or structures, naturally occurring backbones, and/or naturally occurring internucleotide linkages. For example, nucleic acids containing one or more carbocyclic sugars are also included within this definition (Jenkins et al. (1995) Chem. Soc. Rev. pp 169-176, which is incorporated by reference). To further illustrate, although a nucleic acid will generally contain phosphodiester bonds, in some cases nucleic acid analogs are included that have alternate backbones. These include, without limitation, phosphoramidate (Beaucage et al. (1993) Tetrahedron 49(10): 1925 and the references therein; Letsinger (1970) J. Org. Chem. 35:3800; Sprinzl et al.

(1977) Eur. J. Biochem. 81:579; Letsinger et al. (1986) Nucl. Acids Res. 14: 3487; Sawai et al. (1984) Chem. Lett. 805; Letsinger et al. (1988) J. Am. Chem. Soc. 110:4470; and Pauwels et al. (1986) *Chemica Scripta* 26:1419), phosphorothioate (Mag et al. (1991) *Nucleic Acids Res.* 19:1437 and U.S. Pat. No. 5,644,048), phosphorodithioate (Briu et al. (1989) *J. Am. Chem. Soc.* 111:2321), O-methylphosphoroamidite linkages (Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press (1992)), and peptide nucleic acid backbones and linkages (Egholm (1992) *J. Am. Chem. Soc.* 114:1895; Meier et al. (1992) *Chem. Int. Ed. Engl.* 31:1008; Nielsen (1993) *Nature* 365:566; and Carlsson et al. (1996) *Nature* 380:207), which references are each incorporated by reference.

[0054]Other analog nucleic acids include those with positively charged backbones (Denpcy et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6097); non-ionic backbones (U.S. Pat. Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; *Angew* (1991) *Chem. Intl. Ed. English* 30: 423; Letsinger et al. (1988) *J. Am. Chem. Soc.* 110:4470; Letsinger et al. (1994) *Nucleoside & Nucleotide* 13:1597; Chapters 2 and 3, *ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research"*, Ed. Y. S. Sanghvi and P. Dan Cook; Mesmaeker et al. (1994) *Bioorganic & Medicinal Chem. Lett.* 4: 395; Jeffs et al. (1994) *J. Biomolecular NMR* 34:17; *Tetrahedron Lett.* 37:743 (1996)) and non-ribose backbones, including those described in U.S. Pat. Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, *ASC Symposium Series 580, Carbohydrate Modifications in Antisense Research*, Ed. Y. S. Sanghvi and P. Dan Cook, which are each incorporated by reference. Several nucleic acid analogs are also described in, e.g., *Rawls, C & E News Jun. 2, 1997 page 35*, which is incorporated by reference. Modifications of the ribose-phosphate backbone may be done to facilitate the addition of additional moieties, such as labeling moieties, or to alter the stability and half-life of such molecules in physiological environments.

[0055]In addition to naturally occurring heterocyclic bases that are typically found in nucleic acids (e.g., adenine, guanine, thymine, cytosine, and uracil), nucleic acid analogs also include those having non-naturally occurring heterocyclic or other modified bases. To illustrate, certain bases used in nucleotides that act as melting temperature (T_m) modifiers are optionally included. For example, some of these include 7-deazapurines (e.g., 7-deazaguanine, 7-deazaadenine, etc.), pyrazolo[3,4-d]pyrimidines, propynyl-dN (e.g., propynyl-dU, propynyl-dC, etc.),

and the like. See, e.g., U.S. Pat. No. 5,990,303, which is incorporated by reference. Other representative heterocyclic bases include, e.g., hypoxanthine, inosine, xanthine; 8-aza derivatives of 2-aminopurine, 2,6-diaminopurine, 2-amino-6-chloropurine, hypoxanthine, inosine and xanthine; 7-deaza-8-aza derivatives of adenine, guanine, 2-aminopurine, 2,6-diaminopurine, 2-amino-6-chloropurine, hypoxanthine, inosine and xanthine; 6-azacytosine; 5-fluorocytosine; 5-chlorocytosine; 5-iodocytosine; 5-bromocytosine; 5-methylcytosine; 5-propynylcytosine; 5-bromovinyluracil; 5-fluorouracil; 5-chlorouracil; 5-iodouracil; 5-bromouracil; 5-trifluoromethyluracil; 5-methoxymethyluracil; 5-ethynyluracil; 5-propynyluracil, and the like. Many non-naturally occurring bases are also described in, e.g., Seela et al. (1991) *Helv. Chim. Acta* 74:1790, Grein et al. (1994) *Bioorg. Med. Chem. Lett.* 4:971-976, and Seela et al. (1999) *Helv. Chim. Acta* 82:1640, which are each incorporated by reference. Additional examples of modified bases and nucleotides are also described in, e.g., U.S. Pat. Nos. 5,484,908; 5,645,985; 5,830,653; 6,639,059; 6,303,315; and U.S. Pat. Application Pub. No. 2003/0092905, which are each incorporated by reference.

[0056]A "nucleotide" refers to an ester of a nucleoside, e.g., a phosphate ester of a nucleoside. To illustrate, a nucleotide can include 1, 2, 3, or more phosphate groups covalently linked to a sugar moiety of the nucleoside (e.g., at a 5' position, 3' position, 2' position, etc.).

[0057]A "nucleotide incorporating biocatalyst" refers to a catalyst that catalyzes the incorporation of nucleotides into a nucleic acid. Nucleotide incorporating biocatalysts are typically enzymes. An "enzyme" is a protein-based catalyst that acts to reduce the activation energy of a chemical reaction involving other compounds or "substrates." A "nucleotide incorporating enzyme" refers to an enzyme that catalyzes the incorporation of nucleotides into a nucleic acid. Exemplary nucleotide incorporating enzymes include, e.g., DNA polymerases, RNA polymerases, terminal transferases, reverse transcriptases, telomerases, polynucleotide phosphorylases, and the like. Other biocatalysts may be DNA-based ("DNAzymes") or RNA-based ("ribozymes"). A "thermostable enzyme" refers to an enzyme that is stable to heat, is heat resistant and retains sufficient catalytic activity when subjected to elevated temperatures for selected periods of time. For example, a thermostable polymerase retains sufficient activity to effect subsequent primer extension reactions when subjected

to elevated temperatures for the time necessary to effect denaturation of double-stranded nucleic acids. Heating conditions necessary for nucleic acid denaturation are well known in the art and are exemplified in U.S. Pat. No. 4,683,202, entitled "PROCESS FOR AMPLIFYING NUCLEIC ACID SEQUENCES," issued Jul. 28, 1987 to Mullis and U.S. Pat. No. 4,683,195, entitled "PROCESS FOR AMPLIFYING, DETECTING, AND/OR-CLONING NUCLEIC ACID SEQUENCES," issued Jul. 28, 1987 to Mullis et al., which are both incorporated by reference. As used herein, a thermostable polymerase is typically suitable for use in a temperature cycling reaction such as a PCR or a 5'-nuclease reaction. For a thermostable polymerase, enzymatic activity refers to the catalysis of the polymerization of the nucleotides in the proper manner to form primer extension products that are complementary to a template nucleic acid.

[0058] Exemplary nucleotide incorporating biocatalysts include, e.g., a G46E E678G CS5 DNA polymerase, a G46E L329A E678G CS5 DNA polymerase, a G46E L329A D640G S671F CS5 DNA polymerase, a G46E L329A D640G S671F E678G CS5 DNA polymerase, a G46E E678G CS6 DNA polymerase, a Δ Z05R polymerase, a E615G Taq DNA polymerase, a *Thermus flavus* polymerase, a TMA-25 polymerase, a E678G TMA-25 polymerase, a TMA-30 polymerase, a E678G TMA-30 polymerase, a Tth DNA polymerase, a *Thermus specie* SPS-17 polymerase, a E615G Taq polymerase, a *Thermus* Z05R polymerase, a T7 DNA polymerase, a Kornberg DNA polymerase I, a Klenow DNA polymerase, a Taq DNA polymerase, a Micrococcal DNA polymerase, an alpha DNA polymerase, a reverse transcriptase, an AMV reverse transcriptase, an M-MuLV reverse transcriptase, a DNA polymerase, an RNA polymerase, an *E. coli* RNA polymerase, an SP6 RNA polymerase, a T3 RNA polymerase, a T4 DNA polymerase, a T7 RNA polymerase, an RNA polymerase II, a terminal transferase, a polynucleotide phosphorylase, a ribonucleotide incorporating DNA polymerase, and the like.

[0059] The term "sequence segment" refers to a polynucleotide subsequence of the 5'-portion (i.e., 5'-tail) of the polynucleotides of the invention. The 5'-portions contain one, two or more sequence segments, each sequence segment having a unique base positioned at the 5'-end that is the same for all sequence segments in a 5'-tail of one polynucleotide. The sequence segments can have from about 3-10 or more bases in length, for example, about 3-8, 3-6, 4-6, 4-8 or 7-8 bases in length, and a 5'-portion can

contain from 2 to about 10 or more sequence segments. The sequence segments can, but need not be of equal mass. Sequence segments of equal mass can, but need not have an identical sequence. Sequence segments in a 5'-tail that share an identical nucleic acid sequence are repeats.

[0060] The terms "5'-end" and "3'-end" interchangeably refer to a nucleotide position on a polynucleotide that is at the 5'- or 3'-terminus (i.e., the 5'- or 3'-terminal base), or 1 or 2 nucleotide base positions from the terminus (i.e., at the (-1) or (-2) position from the 5'- or 3'-terminus) of a nucleic acid or subsequence thereof.

[0061] The term "blocking moiety" or "blocking group" refers to a chemical group or moiety attached to the 3'-end of a polynucleotide that prevents the extension of a nucleic acid, e.g., by at least one nucleotide incorporating biocatalyst. Exemplified blocking moieties include substituents at the 2' position of the 3'-end nucleotide base, i.e., a 2'-terminator nucleotide.

[0062] A "2'-terminator nucleotide" refers to a nucleotide analog that comprises a blocking group (BG) at the 2'-position of the sugar moiety of the nucleotide. A 2'-terminator nucleotide can be non-extendible by one or more nucleotide incorporating biocatalysts. That is, once a 2'-terminator nucleotide is incorporated into a nucleic acid (e.g., at a 3'-terminal end of the nucleic acid), the blocking group prevents further extension of a nucleic acid by at least one nucleotide incorporating biocatalyst. An exemplary blocking group is a phosphate group. Exemplary 2'-terminator nucleotides include 2'-monophosphate-3'-hydroxyl-5'-triphosphate nucleosides and 2'-monophosphate-3'-hydroxyl-5'-diphosphate nucleosides. 2'-terminator nucleotides are described in detail, for example, in U.S. Patent Publication Nos. 2007/0219361 and 2007/0154914, the disclosures of each of which are hereby incorporated herein by reference.

[0063] The term "target nucleic acid" refers to any nucleic acid comprising a subsequence to be detected. The target nucleic acid can be DNA or RNA. A target nucleic acid can come from any source, including genomic DNA, mRNA, cDNA. The target nucleic acid can be naturally occurring or synthetic (e.g., an amplicon, a vector, etc). A target nucleic acid, can be, but need not be purified or isolated. Depending on the nature of a detection assay, the target nucleic acid can be from plant or animal tissue, or taken from a reaction mixture. There is no limit on the length of a target nucleic acid, although a target nucleic acid can

be exposed to restriction endonucleases before being subject to detection or identification by the present methods. For the purposes of the present methods, a target nucleic acid is prepared using methods known in the art.

[0064]The term "amplification conditions" refers to conditions in an amplification reaction (e.g., a PCR amplification, a RT-PCR amplification) that allow for hybridization of an extendable polynucleotide (e.g., a primer) with a target nucleotide, and the template-dependent extension of the extendable polynucleotide. As used herein, "amplification conditions" or conditions sufficient for amplifying a target nucleic acid are well known in the art. See, e.g., PCR Primer: A Laboratory Manual, by Dieffenbach and Dveksler, eds., 2003, Cold Spring Harbor Press; and PCR Protocols, Bartlett and Stirling, eds., 2003, Humana Press.

BRIEF DESCRIPTION OF THE DRAWINGS

[0065]FIG. 1 illustrates a schematic representation of the polynucleotides and methods of the invention. In this exemplary embodiment, the polynucleotide contains a 5'-portion or 5'-tail with 3 identical sequence segments (i.e., "repeats") (5'-NTAATAATAAT-3'; SEQ ID NO: 1). The polynucleotide is being used to detect a polymorphic nucleotide in a target nucleic acid. The nucleotide set can contain from about 50% to 100% of a ribonucleotide that is complementary to the nucleotide that is at the 5'-end position of each sequence segment (e.g., rATP=Ā). The ribonucleotides are incorporated into an amplicon strand that is complementary to the polynucleotide. The amplicon is subject to fragmentation, releasing the sequence segments. The mass of the released sequence segments is then detected. Ribonucleotides will also be incorporated into the complementary strand of the target nucleic acid (3'-ĀTTĀTTĀTTĀ-5'; SEQ ID NO:2). Therefore, amplified strands of the target nucleic acid will also be subject to cleavage where ribonucleotides are incorporated.

[0066]FIG. 2 illustrates the successful genotyping of the C allele at a SNP within the H19 gene using methodologies described herein and in Example 1. "GTCTTA" refers to the reverse complement nucleic acid sequence of the repetitive hexamer in the interrogating primer. "Noise" indicates that no discernable signal was detected.

[0067]FIG. 3 illustrates the successful genotyping of the T allele at a SNP within the H19 gene using methodologies described herein and in Example 1.

"GCTCTA" refers to the reverse complement nucleic acid sequence of the repetitive hexamer in the interrogating primer. "Noise" indicates that no discernable signal was detected.

[0068]FIG. 4 illustrates the successful concurrent genotyping of the C and T alleles at a SNP within the H19 gene using methodologies described herein and in Example 1, below. "Noise" indicates that no discernable signal was detected.

[0069]FIG. 5 illustrates the successful identification of HIV transcript using 80% or 90% ribonucleotide (here, rATP) and deoxyribonucleotides dU and dC. No signal was obtained from samples without template. See, Example 2. "CCUA" refers to the reverse complement nucleic acid sequence of the repetitive tetramer in the upstream primer (4×CCUA=SEQ ID NO:6). "GUGA" refers to the reverse complement nucleic acid sequence of the repetitive tetramer in the downstream primer (4×GUGA=SEQ ID NO:7). "Noise" indicates that no discernable signal was detected.

[0070]FIG. 6 illustrates the successful identification of HIV transcript using 80% or 90% ribonucleotide (here, rATP) and deoxyribonucleotides dU and 5-Me-dC (4×CmeCmeUA=SEQ ID NO:8; 4×GUGA=SEQ ID NO:7). No signal was obtained from samples without template. See, Example 2. "Noise" indicates that no discernable signal was detected.

[0071]FIG. 7 summarizes the results from the test samples illustrated in FIGS. 5 and 6 (4×CmeCmeUA=SEQ ID NO:8; 4×CCUA=SEQ ID NO:6).

[0072]FIG. 8A-C illustrates a mass spectra of the 7-mer region of the ATP ribo-PCR of the SNP R of NOS1_361. A. Three fragments GTTTCTA (3×GTTTCTA=SEQ ID NO:9) which corresponds to a homozygote A. B. Three fragments GTTTCTA and GTTTTTTA (3×GTTTTTA=SEQ ID NO: 10) which corresponds to a heterozygote AG. C. Three fragments GTTTTTTA which corresponds to a homozygote G.

DETAILED

DESCRIPTION

1. Introduction

[0073]The present invention provides procedures for multiplex detection that require a minimal number of reaction steps and allow for convenient and true multiplexing. The procedures utilize reactions that incorporate ribonucleotide bases (e.g., an

amplification reaction in which at least one of the four nucleotides (A, C, G, T) is included, completely or partially, as a ribo base (rNTP)). Primers specific to a target nucleic acid are used. One or both of the primers have a 5'-tail with one, two or more sequence segments of about 3-10, or more, bases, and with a base complementary to the ribo base located at the 5'-end of each sequence segment. The sequence segments can, but need not be of equal mass. The amplification results in inclusion of the primer sequence, including its 5'-portion, into an amplicon (i.e., amplification product). The amplicon will also contain incorporated ribo-bases. For analysis by mass spectrometry, the amplification product is subjected to fragmentation, for example, by exposure to alkali. The base selectively cleaves the DNA backbone immediately 3' to the incorporated ribo bases. Thus, incorporation of the primer sequences, and the sequence segments, in the amplification product allows for detection of the presence of the product by detecting the distinct mass of the sequence segments. The usage of a primer translates into the presence of a product of the mass of the copied and amplified sequence segment.

[0074]For multiplexing, primers for multiple target nucleic acids (e.g., different alleles) are tailed with different 5'-portion sequence segments that generate products of distinct mass after cleavage. All neosynthetic DNA contains incorporated ribo base and thus gives rise to fragments when subjected to alkali. To facilitate interpretation, the entire system can be calculated so that generally occurring mass signals from the hybridizing portion of the primer and other nucleic acids are distinct from the masses of the 5'-tail sequence segments, or the complements thereof. The sequence segments in a single tail can be of equal mass and can be identical sequences (i.e., repeats). Reiteration of repeat sequence segments results in clear signals above the signals derived from the sequence(s) of portions of the primers that anneal to or copy a target nucleic acid.

[0075]The procedures allow efficient multiplex genotyping for 10 or more samples in a single well with a minimal number of reaction steps. The steps generally require mixing of DNA template and a mastermix. After thermal cycling, a solution of alkali is added and the mass of the fragmented sequence segments is measured, for example, by mass spectrometry. Any known method for mass measurement can be used. When employing mass spectrometry, any kind of mass spectrometer is suitable for analysis.

[0076]The primers can optionally contain a

removable 3'-end blocking agent (i.e., "hot start") to enable the same thermostable DNA polymerase to effect "hot start," thereby improving specificity and permitting a greater degree of multiplexing.

[0077]In one example for performing the present methods, ribo PCR can be carried out with a DNA sample, for each target nucleic acid (e.g., each SNP, each allele), two forward primers and one reverse primer. The forward primers are designed to be specific for one target nucleic acid (e.g., one SNP, one allele), with the 3'-end or penultimate to 3'-end base specifically complementary to an identifying base on the target nucleic acid. Each forward primer has a specific 5'-tail comprising one, two or more sequence segments of identical mass, wherein the 5'-end base of each sequence segment is complementary to the ribo nucleotide (rNTP) in the PCR amplification mixture. The masses of the sequence segments are designed to be different and distinguishable between different alleles, SNPs, and other detectable target nucleic acids.

[0078]To generate sequence segments of distinguishable mass, a the 5'-tail of a primer could contain, for example, about 2-10 sequence segments that are about 3-10 bases in length; or about 4-7 sequence segments about 4-6 bases in length; or about 2-3 sequence segments about 7-8 bases in length.

[0079]Ribo-PCR or ribo-amplification is a PCR or amplification reaction with a mix of deoxynucleotides and at least one nucleotide (ribonucleotide or rNTP). An example of this is a PCR mixture comprising dGTP, the ribonucleotide rTTP (i.e., 5-methyl-UTP), dCTP and dATP, and a DNA polymerase that efficiently incorporates and extends ribonucleotides in a suitable buffer. The reaction can be carried out by any extension reaction, including thermal cycling reactions. After thermal cycling, the reaction mixture can be treated with alkali. The alkali can be a strong base, for example, NaOH or KOH. The alkali cleaves the DNA backbone immediately 3' to the incorporated ribo bases and results in the generation of fragments. Presence of the target nucleic acid results in the usage of the forward primer specific for the target nucleic acid and thus the presence of the amplified 5'-tail sequence. The copied and cleaved complement of the tail has a unique mass. This mass is used to determine the presence of a target nucleic acid in the tested sample. The presence of a mass signal corresponding to the cleaved complementary tail sequence is used to identify the presence of the target nucleic acid in the tested sample. For multiplexing, a plurality of different tail sequences are used, each with a base composition that gives rise to a specific mass of the

DNA after cleavage.

2. Polynucleotides

[0080]The polynucleotides of the invention comprise a 5'-portion or 5'-tail comprised of one, two or more tandem sequence segments and a 3'-portion designed to be sufficiently complementary to anneal to a target nucleic acid to allow for template-based extension of the nucleotide.

[0081]a. 5'-portion

[0082]The 5'-portion or 5'-tail comprises one, two or more tandem (i.e., contiguous) sequence segments, each sequence segment containing a unique nucleotide base (i.e., used only once in the segment) that is positioned at the 5'-end of each sequence segment. In some embodiments, each sequence segment in a 5'-portion or 5'-tail is of an equal mass. In some embodiments, each sequence segment in a 5'-portion or 5'-tail has an identical nucleotide sequence. Tandem sequence segments with identical nucleotide sequences are referred to as repeats. The unique nucleotide base can follow the last (2nd, or 3rd, or 4th, etc) tandem sequence segment. For example, the sequence of a 5'-portion or 5'-tail with 3 sequence segments of identical mass, each 4 nucleotides in length could be: 5'-TAGCTAGCTAGCT-3' (SEQ ID NO:3), wherein the nucleotide base "T" is complementary to the ribo base in the reaction mix.

[0083]As exemplified above, in some embodiments, the 3'-end nucleotide base of the 5'-portion or 5'-tail is complementary to the ribo base in the reaction mix. Because an alkaline solution cleaves the bond immediately 3' to an incorporated ribo base, including at the 3'-end of the 5'-portion a nucleotide base that is complementary to the ribo base in the reaction mix allows for the release of all sequence segments. What remains at the 3'-end of the cleaved amplicon is an incorporated ribo base with a 2' or 3' phosphate. See, FIG. 1. A nucleotide base that is complementary to the ribo base in the reaction mix positioned at the 3'-end of the 5'-portion is required if the 5'-portion contains only one sequence segment. A nucleotide base that is complementary to the ribo base in the reaction mix positioned at the 3'-end of the 5'-portion also is required if two or more sequence segments are different in mass. However, a nucleotide base that is complementary to the ribo base in the reaction mix positioned at the 3'-end of the 5'-portion is optional if the terminal 5'-nucleotide of target-complementary portion is also the same "unique" nucleotide base or if the 5'-portion contains two or more sequence segments of equal mass.

[0084]In some embodiments, an additional nucleotide base is included at the 5'-end of the 5'-portion. The additional 5'-end nucleotide base can be any base. An exemplified 5'-portion with an additional 5'-end nucleotide base could be: 5'-NTAGCTAGCTAGCT-3' (SEQ ID NO:4). Including the additional 5'-end nucleotide base increases the uniformity and accuracy of mass of the released fragments. This is because when the alkaline solution cleaves the bond immediately 3' to an incorporated ribo base a 2' or 3' phosphate moiety remains attached to the ribo base. See, FIG. 1. An additional nucleotide base included at the 5'-end of the 5'-portion finds use when using a nucleotide incorporating biocatalyst that does not carry out template independent extension.

[0085]In some embodiments, the sequence segments in a 5'-portion or 5'-tail have identical mass but different sequences. For example, the sequence of a 5'-portion or 5'-tail with 3 sequence segments of identical mass, each 4 nucleotides in length could be: 5'-TAGCTCAGTGCAT-3' (SEQ ID NO:5), wherein the nucleotide base "T" is complementary to the ribo base in the reaction mix.

[0086]In some embodiments, the sequence segments in a 5'-portion or 5'-tail can be of different lengths (e.g., 3, 4 and/or 5 nucleotide bases) and different masses, but the whole of the 5'-portion or 5'-tail produces a distinguishable "signature" pattern that can be identified, e.g., by mass spectrometry, when the amplicon is subject to cleavage.

[0087]The 5'-portion or 5'-tail can contain from 2 to about 10 sequence segments, or more, as desired, for example, about 2, 3, 4, 5, 6, 7, 8, 9, 10 or more tandem sequence segments. Each sequence segment can be about 3 to about 10 nucleotide bases in length, or longer, for example, about 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotide bases in length. Generally, the longer the sequence segments, the fewer that need be contained in a 5'-tail. Conversely, the shorter the sequence segments, the more that will be contained in a 5'-tail. For example, a 5'-tail could comprise about 2-4 sequence segments that are about 7-8 nucleotide bases in length, or about 4-6 sequence segments that are about 4-6 nucleotide bases in length, or about 5-8 sequence segments that are about 3-5 nucleotide bases in length.

[0088]The polynucleotides are useful as primers in target nucleic acid extension and amplification reactions. The unique nucleotide base located at the 5'-end of each 5'-tail sequence segment is complementary to the ribo nucleotide base included in

a reaction mixture for the extension or amplification of a target nucleic acid hybridized by the polynucleotides of the invention.

[0089]b. 3'-Portion

[0090]The 3'-portion of the polynucleotides is designed to have a nucleic acid sequence sufficiently complementary to hybridize to a target nucleic acid sequence and be extended. The 3'-portion is at least 5 nucleotide bases in length, and can be longer, as desired, for example, about 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotide bases in length. The 3'-end or (-1) or (-2) to the 3'-end nucleotide base of the 3'-portion can be used to distinguish between specific alleles or SNPs in a target nucleic acid, according to methods known in the art. Exemplary reaction conditions, e.g., for primer extension or thermal cycling include about 50 mM Tricine-KOH, pH 7.5, 100 mM KOAc, 3.0 mM Mg(OAc)₂, and 200 μM of each dNTP; annealing or hybridization temperatures can be between about 50° C. and 70° C., for example from about 60-65° C.

[0091]In some embodiments, the 3'-end of the 3'-portion optionally has a blocking group or a blocking moiety that reversibly prevents extension of the polynucleotide. Exemplified reversible blocking moieties include substituents attached at the 2' position of the 3'-end nucleotide (i.e., 2'-terminators). In some embodiments, the 2'-terminator moiety is a phosphate or phosphate analog, for example, a methyl amino phosphate.

[0092]Generally, the blocking groups (BG) utilized at the 2' position of the sugar moiety can include various embodiments. In some embodiments, for example, BG is a negatively charged group and/or a bulky group. To further illustrate, BG is optionally selected from, e.g., CN, NO₂, N₃, a halo group, an ether group, an aldehyde group, a carboxylic acid group, an ester group, an amino group, OCH₃, OCH₂COOH, an O-silylether group, a keto group, an O-lactone group, an O-alkyl group, an O-cyclic alkyl group, an O-alkenyl group, an O-alkynyl group, a carbamate group, an imide group, an amide group, and combinations thereof. More specifically, BG optionally comprises the formula (I):

##STR00001##

[0093]In other embodiments, BG comprises the formula (II):

##STR00002##

where Q is O, S, or NH; X is H, OH, CH₃, BH₃, F, or SeH; and Z is O, S, or Se. To further illustrate, BG optionally comprises the formula (III):

##STR00003##

where Q is O, S, or NH; X is O, S, or NH; Z is O, S, or Se; and R is an alkyl group, an alkenyl group, or an alkynyl group. In another exemplary embodiment, BG comprises the formula (IV):

##STR00004##

where Q is O, S, or NH; X is O, S, or NH; Z is O, S, or Se; L is --CONH(CH₂)_nNH--, --CO(CH₂)_nNH--, or --CONH(CH₂CH₂O)_nCH₂CH₂NH--; n is an integer greater than 0; and R is NH₂, SH, COOH, a quencher moiety, a reporter moiety, biotin, or a affinity moiety.

[0094]Exemplary 2'-terminator nucleotides that find use include 2'-monophosphate-3'-hydroxyl-5'-triphosphate nucleosides and 2'-monophosphate-3'-hydroxyl-5'-diphosphate nucleosides. 2'-terminator reversible blocking moieties are described in detail, for example in U.S. Patent Publication Nos. 2007/0219361 and 2007/0154914.

[0095]The polynucleotides of the invention can optionally contain non-naturally occurring nucleotide backbone linkages, nucleotide bases and nucleotide base analogs, as described herein. The polynucleotides can be single-stranded or double stranded. In some embodiments, the polynucleotides are single-stranded over their full length. The polynucleotides can, but oftentimes do not, contain any labels, for example, attached fluorophores, radioisotopes, enzymes or other identifiers. The polynucleotides can be isolated, synthesized or recombinant. The polynucleotides, inclusive of the 5'- and 3'-portions, can be any length and are generally about 200 nucleotides or fewer, for example, about 150, 125, 100, 75, 50 or 25, or fewer, nucleotides in length.

3. Methods of Detecting

[0096]a. Contacting a Target Nucleic Acid with a Polynucleotide of the Invention

[0097]The first step in performing the methods of detection involves contacting a target nucleic acid with a polynucleotide of the invention, a nucleotide set, and a nucleotide incorporating biocatalytic component.

[0098]i. Nucleotide Set

[0099]The nucleotide sets used in the present methods contain at least one base in the form of a ribonucleotide base (e.g., A, U, G or C or another ribonucleotide, rNTP). In some embodiments, the nucleotide set contains at least two bases in the form of a ribo nucleotide base. The one or more bases included as a ribonucleotide base can be wholly (e.g., 100% ribonucleotide) or partially (e.g., less than 100% ribonucleotide, e.g., where the remainder of the base is a deoxynucleotide or dNTP) included as a ribonucleotide base in the nucleotide set. In some embodiments, the one or more nucleotide bases are present in a majority portion, i.e., more than 50%, e.g., 60%, 70%, 80%, 85%, 90%, 95%, as a ribonucleotide and in a minority portion (i.e., less than 50%) as a deoxyribonucleotide. The one or more ribonucleotides are included with deoxyribonucleotides to round out a complete nucleotide set (A, T, C and G nucleotides are present).

[0100]ii. Biocatalytic Component

[0101]The nucleotide incorporating biocatalytic components used in the present methods comprise both deoxyribonucleotide and ribonucleotide incorporating activities. In some embodiments, the same catalytic domain can catalyze both deoxyribonucleotide and ribonucleotide incorporating activities. In some embodiments, the biocatalytic components comprise at least two catalytic domains, one that has deoxyribonucleotide incorporating activity, the other that has ribonucleotide incorporating activity.

[0102]Exemplified biocatalytic components suitable for use in the present methods include, without limitation, polymerases that are modified Z05, CS5, CS6 or Taq polymerases. CS5 and CS6 chimeric polymerases are further described in, e.g., U.S. Pat. Application Publication No. 2004/0005599, which is incorporated by reference herein in its entirety. *Thermus* species Z05 has been published in PCT International Patent Publication No. WO 92/06200, which is incorporated herein by reference in its entirety. Exemplary modified enzymes include, e.g., a G46E E678G CS5 DNA polymerase, a G46E L329A E678G CS5 DNA polymerase, a G46E L329A D640G S671F E678G CS5 DNA polymerase, a G46E L329A T606S D640G S671F E678G CS5 DNA polymerase, a G46E E678G CS6 DNA polymerase, an E615G Taq DNA polymerase, and the like. These modified enzymes comprise mutations that enhance that incorporation of ribonucleotides, and enhance incorporation of 2'-modified analogs of

ribonucleotides, and/or that reduce or eliminate 5'-3' exonuclease activity, e.g., relative to an enzyme that lacks one or more of these mutations.

[0103]Additional details relating to useful nucleotide incorporating biocatalysts are also provided in, e.g., U.S. patent application Ser. No. 11/873,896 and U.S. Pat. Nos. 5,939,292; 4,889,818; 5,374,553; 5,420,029; 5,455,170; 5,466,591; 5,618,711; 5,624,833; 5,674,738; 5,789,224; 5,795,762; 7,148,049 and 7,179,590, which are each hereby incorporated herein by reference in their entirety for all purposes.

[0104]The production of modified enzymes with, e.g., enhanced efficiency for incorporating ribonucleotides and 2'-terminator nucleotides or other desired properties can be accomplished by various processes including, e.g., site-directed mutagenesis, chemical modification, etc. More specifically, site-directed mutagenesis is generally accomplished by site-specific primer-directed mutagenesis. This technique can be conducted using a synthetic oligonucleotide primer complementary to a single-stranded phage DNA to be mutagenized except for a limited mismatch representing the desired mutation. Briefly, the synthetic oligonucleotide is used as a primer to direct synthesis of a strand complementary to the plasmid or phage, and the resulting double-stranded DNA is transformed into a phage-supporting host bacterium. The resulting bacteria can be assayed by, for example, DNA sequence analysis or probe hybridization to identify those plaques carrying the desired mutated gene sequence. To further illustrate, many other approaches to modify nucleic acids, such as "recombinant PCR" methods can also be utilized.

[0105]In practicing aspects of the present invention (e.g., producing modified enzymes, performing sequencing reactions, etc.), many conventional techniques in molecular biology and recombinant DNA are optionally utilized. These techniques are well known and are explained in, for example, *Current Protocols in Molecular Biology*, 1997-2007 (F. M. Ausubel ed.), Wiley Interscience; Sambrook et al., 2001, *Molecular Cloning: A Laboratory Manual*, Third Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; Berger and Kimmel, *Guide to Molecular Cloning Techniques Methods in Enzymology* volume 152 Academic Press, Inc., San Diego, Calif. (Berger), *DNA Cloning: A Practical Approach*, Volumes I and II, 1985 (D. N. Glover ed.); *Oligonucleotide Synthesis*, 1984 (M. L. Gait ed.); *Nucleic Acid Hybridization*, 1985, (Hames and Higgins); *Transcription and Translation*, 1984 (Hames and Higgins eds.); *Animal Cell Culture*, 1986

(R. I. Freshney ed.); Immobilized Cells and Enzymes, 1986 (IRL Press); Perbal, 1984, A Practical Guide to Molecular Cloning; the series, Methods in Enzymology (Academic Press, Inc.); Gene Transfer Vectors for Mammalian Cells, 1987 (J. H. Miller and M. P. Calos eds., Cold Spring Harbor Laboratory); Methods in Enzymology Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively).

[0106]iii. Polynucleotides

[0107]The 5'-tailed polynucleotides used in the methods are as described above and herein. In some embodiments, a target nucleic acid is contacted with a plurality (e.g., two or more) of 5'-tailed polynucleotides. In some embodiments, each polynucleotide in a plurality is separately identifiable by having a 5'-portion comprising one, two or more sequence segments of a unique mass. In this case, fragmenting the amplicon or complements of each 5'-tailed polynucleotide produces a different mass signal or signature mass profile. In some embodiments, the polynucleotides are contacted to a target nucleic acid as one or more primer pairs. One or both of the polynucleotides in a primer pair can include a 5'-tail comprising tandem sequence segments. Multiple polynucleotides contacting the target nucleic acid as one or more primer pairs can also be provided wherein each primer in the pair comprises a 5'-portion comprising tandem sequence segments of identical mass. In this case, fragmenting the complements of the 5'-portions of each primer in the pair that is used to generate an amplicon produces a single mass signal of greater (e.g., about double) intensity. Alternatively, each primer in a primer pair can contain one sequence segment, the sequence segment in each primer of the pair being of identical mass.

[0108]In some embodiments of the methods, the 5'-tailed polynucleotides can comprise one sequence segment of distinct mass to produce a detectable cleavage product.

[0109]iv. Target Nucleic Acid

[0110]A target nucleic acid can be from any source, e.g., a tissue sample from an animal, a plant, a bacteria or a virus, or can be synthetic, e.g., from a reaction mixture. The target nucleic acid can be, e.g., genomic DNA, a chromosome or chromosomal segment, mRNA, cDNA, a vector (e.g., an expression vector), an expression cassette, a naked DNA or RNA polymer, the product of a polymerase chain reaction (PCR), an oligonucleotide, a probe, a primer, etc. A target nucleic acid can be, e.g., single-stranded, double-stranded, triple-stranded, etc and is not limited

to any particular length. Target nucleic acids used in the present methods oftentimes will contain one or more single nucleotide polymorphisms (SNPs), insertions, deletions, substitutions or other distinguishing features. Target nucleic acids are generally provided in a sample.

[0111]In some embodiments, the target nucleic acid is contacted (i.e., the methods are performed) with the purpose of distinguishing one or more SNPs. In such an application, at least two different allele-distinguishing 5'-tailed polynucleotides are used that differ in their 3'-portion such that they specifically anneal to a particular allele, and also differ in their 5'-portion such that the one or more sequence segments are of a mass that corresponds to and uniquely identifies the allele bound by the 3'-portion.

[0112]In some embodiments, the target nucleic acid is contacted with the purpose of detecting one or more rare mutations in the target nucleic acid. In such an application, one or more polynucleotides having a 3'-portion that specifically anneals to a mutation at a particular position in the target nucleic acid are present (a polynucleotide having a 3'-portion that specifically anneals to the wild-type at the same position can optionally be present in the same or in a separate reaction). Where one mutation in the target nucleic acid sequence is being evaluated, one, two or more different mutation-distinguishing 5'-tailed polynucleotides can be used that differ in their 3'-portion such that they specifically anneal the wild-type or one or more mutation types, and also differ in their 5'-portion such that the one or more sequence segments are of a mass that corresponds to and uniquely identifies the allele bound by the 3'-portion. In applications where two or more mutations in one target nucleic acid are being evaluated, polynucleotides with 3'-portions that specifically detect a mutation at a first position in the target nucleic acid can have 5'-portion sequence segments all of a first mass; polynucleotides with 3'-portions that specifically detect a mutation at a second position in the target nucleic acid can have 5'-portion sequence segments all of a second mass. Alternatively, the masses of the sequence segments in the 5'-portions of the polynucleotides can each be different, operating as unique identifiers.

[0113]In some embodiments, e.g., those involving multiplex amplification or extension, the polynucleotides are provided as two or more primer pairs. In the first primer pair, the 5'-portion of each primer in the pair contains sequence segments of a first mass; the second primer pair has 5'-portions of each primer in the pair containing sequence segments

of a second mass, detectably distinct from the first mass; the third primer pair has 5'-portions of each primer in the pair containing sequence segments of a third mass, detectably distinct from the mass of the first and second sequence segments; subsequent primer pairs have 5'-portions of each primer in the pair containing sequence segments of a mass detectably distinct from all preceding or other primer pairs. The number of primer pairs and amplification or extension reactions is limited by the availability of free nucleotide bases in the reaction mixture. In some embodiments, a reaction mixture can contain 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or more primer pairs. This strategy allows for the concurrent detection of multiple target nucleic acids or multiple targets in a single target nucleic acid.

[0114]b. Amplifying the Target Nucleic Acid to Produce an Amplicon

[0115]i. Methods of Amplification

[0116]The target nucleic acid can be amplified employing one or more of the polynucleotides of the invention using any polynucleotide extension or amplification method known in the art. Amplification using any known variation of a polymerase chain reaction (PCR), including RT-PCR, quantitative PCR, multiplexed and real-time PCR, finds use in the present methods. Amplification also includes isothermal amplification methods and polynucleotide extension methods known in the art. Protocols for carrying out PCR are well known in the art, and are described, for example, in PCR Primer: A Laboratory Manual, Dieffenbach and Dveksler, eds., 2003, Cold Spring Harbor Laboratory Press; A-Z of Quantitative PCR, Bustin, ed., 2004, International University Line; Edwards, Real-Time PCR: An Essential Guide, 2004, Taylor & Francis; Real Time PCR, Dorak, ed., 2006, Taylor & Francis; PCR Protocols: A Guide to Methods and Applications, Innis, et al., eds., 1990, Academic Press, San Diego; PCR Strategies, Innis, et al., eds, 1995, Academic Press, San Diego; and PCR Applications: Protocols for Functional Genomics, Innis, et al., eds., 1999, Academic Press, San Diego. The target nucleic acid is amplified or extended for a sufficient amount of time, or for a sufficient number of cycles to produce a detectable amount of amplicon.

[0117]As discussed above, one or more target nucleic acids can be amplified or one or more target positions in a single target nucleic acid can be amplified. Because there is essentially no limit on the number of different 5'-portions with sequence segments of a distinguishable mass, the methods are particularly suited to multiplex amplification assays. The one or

more polynucleotides used to amplify a target nucleic acid or a specific location within a target nucleic acid can have different 5'-portions with sequence segments of a uniquely identifiable mass. Pairs of polynucleotides used as primer pairs can have 5'-portions with sequence segments of identical mass. In some embodiments, the sequence segments share identical sequences. The multiplex amplification assays can concurrently determine the presence (or absence) of at least 5, 10, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100 or more different target nucleic acids or positions within a target nucleic acid, as desired.

[0118]The amplification and polynucleotide extension reactions of the present invention employ a biocatalytic unit (e.g., polymerase) that comprises enzymatic activity to incorporate ribonucleotides into the nucleic acid strand being extended (e.g., "ribo-PCR," or "ribo-amplification," or "ribo-extension"). Exemplified biocatalytic units with ribonucleotide incorporating activities include the G46E CS6R DNA polymerase and the KB17 DNA polymerase, available from Roche Molecular Systems, and described in, for example, Mauger, et al., Nucleic Acids Research (2007) 35(8):e62 and Mauger, et al., Nucleic Acids Research (2006) 34(3):e18. Additional biocatalytic components that find use are discussed above and herein. See also, U.S. Pat. No. 5,939,292.

[0119]ii. Amplicon Produced

[0120]Amplifying the target nucleic acid sequence will yield double-stranded amplicons,

[0121]wherein the strand of the amplicon complementary to the 5'-portion of the polynucleotide will comprise a 3'-portion containing one or more sequence segments, wherein the 3'-end nucleotide base of the sequence segments contain an incorporated ribo-nucleoside monophosphate (rNMP).

[0122]c. Cleaving the Amplicon

[0123]The sequence segments of the amplicons can be cleaved using any method known in the art. In some embodiments, the amplicons are cleaved at the bond 3' to an incorporated NMP by subjecting the amplicon to an alkaline (i.e., basic solution). The amplicons are exposed to the alkaline solution for a time period and temperature sufficient to effect cleavage of the bonds 3' to an incorporated NMP. Generally, the higher the temperature, the shorter the time period necessary for alkali treatment. For example, the cleaving or fragmenting step can be carried out for about 1.5 hours at 70° C. or for about 8

hours (i.e., overnight) at about 55° C. Alkali treatment temperatures can be as low as refrigerated temperatures (e.g., 4° C.) and as high as near boiling temperatures (e.g., about 95° C.). In some embodiments, the alkaline solution can be any solution with a pH greater than about 8.5, for example, with a pH of at least about 9.0, 9.5, 10.0, 10.5, 11.0, 11.5, 12.0, 12.5, 13.0 or greater. In some embodiments, the alkaline solution will contain at least about 0.2M, 0.3M, 0.5M, 0.8M, 1.0M or more of a basic compound.

[0124]The alkaline solution will generally contain at least one strong base. In some embodiments, a weak base can be used, for example, if contained in the alkaline solution at a higher concentration. Generally, the alkaline solutions will contain at least one basic compound with a pK_b of about 5.0 or less, for example a pK_b of about 4.5, 4.0, 3.5, 3.0, 2.5, 2.0, 1.5, 1.0, 0.5 or less. Exemplified basic compounds for use in fragmenting the amplicons at the bonds 3' to an incorporated NMP include, for example, NaOH, KOH, RbOH and NH₄OH. Other bases or basic compounds will also find use. See, e.g., Mauger, et al., *Nucleic Acids Research* (2007) 35(8):e62 and Mauger, et al., *Nucleic Acids Research* (2007) 34(3):e18 for procedures of alkaline treatment, the disclosures of both of which are hereby incorporated herein by reference in their entirety for all purposes.

[0125]d. Detecting the Amplicon Fragments

[0126]The amplicon fragments or cleaved and released sequence segments can be detected using any method known in the art. In some embodiments, the cleaved fragments are detected by identification of mass. The sequence segments can also be detectably labeled before or after cleavage, e.g., with a fluorophore or a radioisotope. Fragments can be labeled at the 5'- or 3'-ends, or on any nucleotide base throughout. Exemplified methods of detection include, without limitation, mass spectrometry, electrophoretic separation, time-resolved fluorescence detection, spin resonance detection and hybridization to a solid phase (e.g., an array) with subsequent detection.

[0127]In some embodiments, the cleaved sequence segments released from the amplicon are detected using mass spectrometry. Methods of mass spectrometry are known in the art. Any mass spectrometry technique will find use, including for example, laser desorption-ionization mass spectrometry (e.g., matrix assisted laser desorption-ionization mass spectrometry (MALDI) and surface enhanced laser desorption-ionization mass

spectrometry SELDI), gas-phase ion spectrometry, gas-chromatography mass spectrometry, tandem mass spectrometry, and other mass spectrometry techniques. Mass spectrometry methods are well known in the art, and are described for example, in Gross, *Mass Spectrometry: A Textbook*, 2006, Springer Verlag; and Dass, *Fundamentals of Contemporary Mass Spectrometry*, 2007, Wiley Interscience.

[0128]The amplicon fragments that are detected can be cleaved sequence segments from the amplicon that are complementary to or amplified from the 5'-portion of the polynucleotide. Of course, it will be appreciated that cleavable sequence segments can be created throughout the length of the amplicon. Under alkaline conditions, the bonds immediately 3' to where ribonucleotides have been incorporated are subject to cleavage. In some embodiments, amplicon fragments can be detected from the amplified target nucleic acid (e.g., from the sequence segments complementary to the 3'-portion of the polynucleotide). As desired, cleaved sequence segments from the 5'-portion and/or the 3'-portion of the polynucleotide, or the complements thereof, can be detected. The signal(s) from the detected amplicon fragments produce a signature of peaks indicative of the target nucleic acid to be identified.

4. Reaction Mixtures

[0129]The present invention also provides reaction mixtures involved in the methods of the invention. Any reaction mixtures as described above can be generated. An exemplary reaction mixture comprises, for example, one or more polynucleotides of the invention comprising a 5'-portion or 5'-tail comprised of one, two or more tandem sequence segments and a 3'-portion designed to be sufficiently complementary to anneal to a target nucleic acid; a nucleotide set comprising dNTPs and wherein the majority of at least one base is a ribonucleotide; and a biocatalytic component having ribonucleotide incorporating activity. In some embodiments, the reaction mixtures comprise one or more target nucleic acid sequences. The reaction mixtures of the invention may also comprise amplicons produced by the present methods, the amplicon comprising a 3'-portion having sequence segments complementary to the 5'-portion of the polynucleotides of the invention, wherein the 3'-end nucleotide base of each sequence segment is a nucleoside monophosphate (NMP). A plurality of polynucleotides in the reaction mixtures may be provided in polynucleotide pairs, e.g., as primer pairs. The embodiments of the components in the reaction mixtures are as described above and herein.

[0130]In some embodiments, the reaction mixtures comprise two or more primer pairs. In the first primer pair, the 5'-portion of each primer in the pair contains sequence segments of a first mass; the second primer pair has 5'-portions of each primer in the pair containing sequence segments of a second mass, detectably distinct from the first mass; the third primer pair has 5'-portions of each primer in the pair containing sequence segments of a third mass, detectably distinct from the mass of the first and second sequence segments; subsequent primer pairs have 5'-portions of each primer in the pair containing sequence segments of a mass detectably distinct from all preceding or other primer pairs. In some embodiments, a reaction mixture can contain 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or more primer pairs.

[0131]In some embodiments, the reaction mixtures comprise nucleotide sets comprising non-conventional or non-naturally occurring nucleotide bases or labeled nucleotide bases, e.g., with a fluorophore or a radioisotope. Reaction mixtures can also contain buffering components and salts, as needed. In some embodiments, the reaction mixtures comprise an alkaline component, as described above, for example, NaOH or KOH. In accordance with the methods described herein, reaction mixtures containing a high concentration of an alkaline component (e.g., about 0.1 M-0.2 M, or higher) typically have already been subject to amplification.

[0132]In some embodiments, the reaction mixtures comprise amplicons produced by the present methods and an alkaline component, as described herein.

5. Kits

[0133]The present invention also provides kits for use in the methods of the invention. Typically, the kit is compartmentalized for ease of use and contains containers providing components for performing the present methods, for example, one or more polynucleotides of the invention comprising a 5'-portion or 5'-tail comprised of one, two or more tandem sequence segments and a 3'-portion designed to be sufficiently complementary to anneal to a target nucleic acid; a nucleotide set comprising dNTPs and wherein the majority of at least one base is a ribonucleotide; and a biocatalytic component having ribonucleotide incorporating activity. In some embodiments, the kits comprise one or more target nucleic acid sequences, including control nucleic acid sequences (for positive and/or negative controls). The kits may provide reagents for performing a separate control reaction, including control polynucleotides

and control target nucleic acid sequences. A plurality of polynucleotides in the kits may be provided in polynucleotide pairs, e.g., as primer pairs. The embodiments of the components in the kits are as described above and herein.

[0134]In some embodiments, the kits comprise two or more primer pairs. In the first primer pair, the 5'-portion of each primer in the pair contains sequence segments of a first mass; the second primer pair has 5'-portions of each primer in the pair containing sequence segments of a second mass, detectably distinct from the first mass; the third primer pair has 5'-portions of each primer in the pair containing sequence segments of a third mass, detectably distinct from the mass of the first and second sequence segments; subsequent primer pairs have 5'-portions of each primer in the pair containing sequence segments of a mass detectably distinct from all preceding or other primer pairs. In some embodiments, a kit can contain 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or more primer pairs.

[0135]One or more additional containers providing additional reagent(s) can also be included. Such additional containers can include any reagents or other elements recognized by the skilled artisan for use in primer extension or amplification procedures in accordance with the methods described above, including reagents for use in, e.g., nucleic acid amplification procedures (e.g., PCR, RT-PCR), DNA sequencing procedures, or DNA labeling procedures. In other, non-mutually exclusive variations, the kit includes one or more containers providing free nucleotides (conventional and/or unconventional). In specific embodiments, the kit includes alpha-phosphorothioate dNTPs, dUTP, dITP, and/or labeled dNTPs such as, e.g., fluorescein- or cyanin-dye family dNTPs. In still other, non-mutually exclusive embodiments, the kit includes one or more containers providing a buffer suitable for a primer extension reaction. In some embodiments, the kits comprise an alkaline component, as described above, for example, NaOH or KOH.

6. Amplicons

[0136]The invention also provides amplicons produced by the present methods. PCR amplicons, for example, are usually double-stranded and comprise a strand that is complementary to the polynucleotides of the invention. The amplicons comprise a 3' sequence segment complementary to the 5'-portion of the polynucleotides of the invention,

[0137]wherein the 3' sequence segment of the

amplicon comprises at least 1, 2 or more sequence segments, wherein the 3'-end nucleotide base of the sequence segments is a nucleoside monophosphate (NMP) having the same base as the NTP in the nucleotide set used to produce the amplicon.

7. Systems

[0138]The invention further provides systems for automating the present methods. The systems comprise at least one container or support comprising a composition comprising a polynucleotide or an amplicon or a reaction mixture of the invention; a thermal modulator configured to thermally communicate at one or several temperatures with the composition comprising the polynucleotide or amplicon (e.g., a nucleic acid thermocycling machine; an incubator); at least one reagent transfer mechanism that transfers reagents to or from the container or support; and at least one detector configured to detect mass (e.g., a mass spectrometer, a fluorometer) of the fragmented sequence segments. In some embodiments, the systems comprise at least one controller in operable communication with at least one of the thermal modulator, the reagent transfer mechanism and the detector configured to detect mass.

EXAMPLES

[0139]The following examples are provided to illustrate the invention without limitation.

Example 1

Flag-Tag For High Throughput SNP Genotyping

[0140]Amplicon containing a SNP in the human H19 gene was prepared for analysis. The target sequence was

gtgaggagtgtggagtaggyGCCAGGCATCGTGCagaca
ggcgacatcagc (SEQ ID NO: 11) (lower case indicates sequences that anneal to the primers, "y" indicates SNP position, C or T). PCR was performed in a total volume of 20 µl, with 2 µl coming from the genomic DNA samples diluted to 10 ng/µl. The PCR amplifications contained the following components: 50 mM Tricine pH 7.5, 100 mM KOAc, 2.75 mM Mg(OAc)₂, and 1.6% Storage Buffer, which in turn contained 50% v/v glycerol, 100 mM KCl, 0.1 mM EDTA, 20 mM Tris pH 8.0, 1 mM DTT, and 0.5% Tween 20. Also included in the PCR was 0.2 mM each 5-methyl-dCTP and dGTP, 0.4 mM dUTP, 0.18 mM rATP, 0.02 mM dATP, and 0.1 mM pyrophosphate. The nucleotide base mixture contained 90% rATP and 10% dATP. Enzymes used in the PCR amplifications were 0.02 U/µl Uracil-

DNA Glycosylase (UNG) and 20 nM GLTDSE DNA polymerase. See, e.g., PCT Publication Nos. WO 2008/046612 and WO 2009/010251, hereby incorporated herein by reference in their entirety for all purposes. High concentration enzyme stocks (8 U/µl and 5 µM, respectively) were used to minimize glycerol and Tween carry-over. Primers were added to 0.2 µM each. One primer in common to all reactions had the sequence 5'-GCTGATGTCGCCCTGTC-2'-PO₄--U-3' (SEQ ID NO: 12). The SNP-interrogating primer was either 5'-CCTAAGACTAAGACTAAGACTCGTGAGGAGTGTGGAGTAGG-2'-PO₄--C-3' (SEQ ID NO: 13) to detect the H19C allele (see, FIG. 2) or -5'-CCTAGAGCTAGAGCTAGAGCTCGTGAGGAGTGTGGAGTAGG-2'-PO₄-U-3' (SEQ ID NO:14) to detect the H19T allele (see, FIG. 3), or a 50/50 mixture of the two. The portion of the primer corresponding to the target sequence is underlined. These primers include at their 5' ends different repetitive hexamers flanked by T residues: (3× TAAGAC=SEQ ID NO: 15) in the primer to detect the H19C allele or (3× TAGAGC=SEQ ID NO: 16) in the primer to detect the H19T allele (a single hexamer is bolded in the primer sequences). The repetitive hexamers are not complementary to the genomic targets but, when incorporated into amplicon, result in release of a hexamer "flag-tag" after base digestion.

[0141]Thermocycling conditions were: 50° C. for 10 minutes (UNG sterilization); 95° C. for 1 minute (UNG denaturation); then 99 cycles between 92° C. for 15 seconds (denaturation) and 60° C. for 2 minutes (anneal/extension). This was followed by a 5 minute hold at 60° C. Thermocycling was conducted in an ABI 9800.

[0142]Six different genomic DNA samples were tested against all three primer pairs, in duplicate. In addition, 4 "no template" controls were run for each primer pair.

[0143]Following thermocycling, replicate reactions were pooled. Base was added and the amplicons were heat-treated. In order to desalt the sample, a chelating agent was added and the solutions were centrifuged. Supernatant was applied to the matrix and analyzed in the mass spectrometer. By analyzing the data and noting the appearance of the flag sequences, the genotype of each of the 6 samples at the SNP position was determined. The results are depicted in FIGS. 2, 3 and 4.

Example 2

Flag-Tag For High Throughput Screening For Infectious Agents

[0144]The application of flag-tag technology to infectious agent screening was tested using RNA transcripts encoding an HIV-derived sequence. Transcript for these experiments was generated by cloning the gag region from HIV strain HXB2 into an expression vector. After linearizing, transcript was made using T7 RNA polymerase. Transcript was then purified over a poly-dT column. The target sequence was

catgcaggcctattgcaccaGGCCAGATGAGAGAACCAAGGGGaaagtacatagcaggaactactagtagcc ttcagga (SEQ ID NO: 17) (primer sequences are in lower case).

[0145]RT-PCR using this transcript was performed in duplicate, with a total volume of 50 µl per reaction. Reactions were performed with and without 10⁶ copies of transcript per reaction. The reactions contained the following components: 100 mM Tricine pH 7.3, 120 mM KOAc, 1 mM Mn(OAc)₂, 0.2 mM dGTP, 0.4 mM dUTP, a mixture of rATP and dATP such that the total was 0.2 mM with either 80% or 90% being rATP, either 0.2 mM dCTP or 0.2 mM 5-methyl-dCTP, and 0.15 mM pyrophosphate. Enzymes used in the reactions were 0.02 U/µl UNG and 25 nM GLDSE DNA polymerase. See, PCT Publication No. WO 2008/046612. High concentration enzyme stocks (2 U/µl and 2.5 µM, respectively) were used to minimize glycerol and Tween carry-over. Primers were added to 0.2 µM each. The upstream primer had the sequence

TABLE-US-00001 5'-ATAGGTAGGTAGGTAGGTATGCAGGGCCTATTGCACC-2'-PO₄-A-3'

The downstream primer had the sequence 5'-ATCACTCACTCACTCACTCCTGAAGGGTACTA GTAGTTCCTGCTATGTCACT-2'-PO₄--U-3'- . The portion of the primer that anneals to the target sequence is underlined. These primers include at their 5' ends different repetitive tetramers flanked by T residues: TCAC and TAGG (a single tetramer is bolded in the primer sequences). The repetitive tetramers are not complementary to the transcript target but, when incorporated into amplicon, result in release of a tetramer "flag-tag" after base digestion. In this example, both primers contain tags. However, as shown in Example 1, it is not necessary to tag both primers to detect amplicon. The use of 5-methyl-dCTP provides for a larger mass difference between tags containing C and U residues. Because dCTP and dUTP differ by only 1 amu, it is difficult to distinguish between them on a mass basis. In contrast

dUTP and 5-methyl dCTP are readily distinguishable by determining mass (i.e., differ by more than 1 amu).

[0146]Thermocycling conditions were: 50° C. for 2 minutes (UNG sterilization); 60° C. for 60 minutes (reverse transcription step); 93° C. for 1 minute (UNG denaturation); then 60 cycles between 92° C. for 15 seconds (denaturation) and 60° C. for 4 minutes (anneal/extension). Thermocycling was conducted in an ABI 9700.

[0147]Amplicon was generated from reactions containing either 80% or 90% rATP and either dCTP or 5-methyl-dCTP. Negative template reactions for these four conditions were also run. Duplicate reactions for all eight conditions were pooled and prepared for analysis. Base was added and the amplicons were heat-treated. In order to desalt the sample, a chelating agent was added and the solutions were centrifuged. Supernatant was applied to the matrix and analyzed in the mass spectrometer. For all samples containing transcript the expected tag sequence was identified by mass spectrometry, but not in the control samples without template. The results are depicted in FIGS. 5, 6 and 7.

Example 3**Detection of A/G alleles of SNP R in NOS1..361****Allele-Specific PCR:**

[0148]Ribo-PCR amplifications in 201 with 1 ng/µl of human genomic DNA, 0.4 µM each primer (Table 1), 0.15 mM sodium pyrophosphate, 100 mM Tricine/KOH at pH 7.3, 100 mM KCOO at pH 7.5, 3 mM Mg(COO)₂, 0.2 mM each (rATP, dCTP, dGTP and dTTP) and 0.25 U/µl FP-1 DNA Polymerase (i.e., GLTDSE DNA Polymerase). The thermal cycling profile for the PCR was 4 min at 92° C. followed by 60 cycles of 15 s at 92° C., 4 min at 63° C. This was always concluded at 4° C. 5 µl of PCR was put into a 2% of agarose gel to control the PCR.

[0149]Table 1 provides sequences of the primers used in this example. The symbol * indicates a 2'-PO₄ containing residue. The symbol (C) indicates a 2'OME cytidine base. Underline sequence represents the Flag part of the primer and a single heptamer is bolded in the primer sequences.

TABLE-US-00002 TABLE 1 Primer Sequence
Forward 1
CCTAGAAACTAGAACTAGAACTCTG SEQ
ID NO:20 ATGGCTCACCATTGAAAA* Forward 2
CCTAAAACTAAAACTAAAACTCTG SEQ

ID NO:21 ATGGCTCACCATTGAAAG* Reverse
GTCAATGAAGGAAGGTAG(C)A SEQ ID NO:23

Alkali

Cleavage:

[0150]For alkali cleavage (Table 2) 5.0 µl of 1.2 M sodium hydroxide was added to the remaining 15 µl of the amplification reaction for a final concentration of 0.3 M and incubated at 70° C. for 1.5 hours.

[0151]Table 2 shows predicted masses of the fragment of the cleavage of ATP ribo-PCR of the SNP R of NOS1..361. Underline sequence represents the Flag part of the primer and a single heptamer is bolded in the primer sequences.

TABLE-US-00003 TABLE 2 SEQ SNP Sequence ID

| NO: | Mass | Start | End | Sense | tca | 940.6 | 28 | 26 |
|---------|----------------------------------|---------|----------------------------------|---------|-------------|---|---------|--|
| REVERSE | gccca | 1254.8 | 32 | 29 | REVERSE | ttga | 1284.8 | 69 |
| 72 | FORWARD | tggtga | 1943.2 | 38 | 33 | REVERSE | Flag A | -- |
| 23 | 17 | REVERSE | -- | 16 | 10 | REVERSE | gtttcta | 2182.4 |
| 9 | 3 | REVERSE | Flag G | -- | 23 | 17 | REVERSE | -- |
| 16 | 10 | REVERSE | gttttta | 2197.4 | 9 | 3 | REVERSE | ccttcctca |
| 23 | 3009.9 | 59 | 68 | FORWARD | gggggctgcta | 24 | 3509.2 | 48 |
| 58 | FORWARD | G | gtcaatgaaggaaggtagcagcccccttttca | 25 | 10522.8 | 73 | 40 | REVERSE |
| A | gtcaatgaaggaaggtagcagcccccttttca | 26 | 10537.8 | 73 | 40 | REVERSE | A | cctagaactagaaactagaaactctgatggctcaccattgaaaa |
| 27 | 13919 | 1 | 46 | FORWARD | G | cctaaaaactaaaaactctgatggctcaccattgaaaga | 28 | 14230.3 |
| 1 | 46 | FORWARD | | | | | | |

Purification:

[0152]Samples were desalted by the addition of cation exchange resin charged with H.sup.+ . 6 mg of resin was added to the reaction with the MassArray® Clean Resin Tool Kit and incubated for 20 minutes at room temperature under agitation. Thereafter, the sample was centrifuged for 2 min at 134×g to sediment the resin.

MALDI-TOF

MS:

[0153]Trihydroxyacetophenone (THAP) was used as matrix. For preparation 0.5 µl of 0.2 M of 2,4,6 and 2,3,4 THAP in 50% acetonitrile and 0.3 M of ammonium citrate in water in 6/3/2 (v/v) was deposited on an anchor position of MALDI target plate (Anchor Chip® Target with a spot size of 400 µm, Bruker Daltonik GmbH, Bremen, Germany). After 0.5 µl of the supernatant was added and dried at room temperature.

[0154]The target was introduced into the MALDI-

TOF mass spectrometer (Autoflex, Bruker Daltonik GmbH, Bremen, Germany) for analysis. Mass spectrometry analysis (FIG. 8) was carried out in negative ion mode, with an acceleration voltage of 20 kV using a pulsed ion extraction delay of 100 ns in linear and with external calibration. Each spectrum obtained was the sum of 200 laser shots.

[0155]It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

| Sequence | CWU | 1 |
|-----------------------|---|---|
| 28111DNAArtificial | SequenceDescription | of |
| Artificial Sequence5' | portion of polynucleotide target | nucleic acid with 3 sequence segment repeats |
| Intaataataa | t | 11211DNAArtificial |
| SequenceDescription | of Combined DNA/RNA | Molecule complementary strand to 5' portion of polynucleotide target nucleic acid with 3 sequence segment repeats containing incorporated ribonucleotides |
| 2attattatta | n | 11313DNAArtificial |
| SequenceDescription | of Artificial Sequence | polynucleotide 5'-portion or 5'-tail with identical tandem repeat sequence segments |
| 3tagtagtagta | gct | 13414DNAArtificial |
| SequenceDescription | of Artificial Sequence polynucleotide 5'-portion or 5'-tail with identical tandem repeat sequence segments and additional nucleotide base at 5'-end | 4ntagtagct agct |
| 14513DNAArtificial | SequenceDescription | of Artificial Sequence polynucleotide 5'-portion or 5'-tail with different tandem sequence segments |
| 5tagtctagtg | cat | 13616DNAArtificial |
| SequenceDescription | of Combined DNA/RNA Molecule reverse complement of repetitive tetramer in upstream primer | 6ccnaccnaccnaccna |
| 16716DNAArtificial | SequenceDescription | of Combined DNA/RNA Molecule reverse complement of repetitive tetramer in downstream primer (4 GUGA) |
| 7ngagngagn | gagnga | 16816DNAArtificial |
| SequenceDescription | of Combined DNA/RNA Molecule reverse complement of repetitive tetramer in upstream primer (4 CmeCmeUA) | 8ccnaccnaccnaccna |
| 16921DNAArtificial | SequenceDescription | of Combined DNA/RNA Molecule reverse complement of repetitive heptamer in upstream primer (3 GTTTCTA) |
| 9gtttctagtt | tctagttct | a |
| 211021DNAArtificial | SequenceDescription | of Combined DNA/RNA Molecule reverse complement of repetitive heptamer in downstream primer (3 |

GTTTTTA) 10gttttagtt tttagtttt a
 211153DNAArtificial SequenceDescription of
 Artificial Sequencehuman H19 gene SNP target
 sequence 11gtgaggagtg tggagtaggy gccccagcat
 cgtgcagaca gggcgacatc agc 531218DNAArtificial
 SequenceDescription of Combined DNA/RNA
 MoleculePCR amplification primer common to all
 human H19 gene SNP reactions 12gctgatgctg
 ccctgtnu 181342DNAArtificial SequenceDescription
 of Artificial Sequencehuman H19 gene SNP PCR
 amplification SNP-interrogating primer to detect
 H19C allele 13cctaagacta agactaagac tcgtgaggag
 tgtggagtag nc 421442DNAArtificial
 SequenceDescription of Combined DNA/RNA
 Moleculehuman H19 gene SNP PCR amplification
 SNP-interrogating primer to detect H19T allele
 14cctagagcta gagctagagc tcgtgaggag tgtggagtag nu
 421518DNAArtificial SequenceDescription of
 Artificial Sequence5'-end repetitive hexamers in
 primer to detect H19C allele 15taagactaag actaagac
 181618DNAArtificial SequenceDescription of
 Artificial Sequence5'-end repetitive hexamers in
 primer to detect H19T allele 16tagagctaga gctagagc
 181780DNAArtificial SequenceDescription of
 Artificial SequenceHIV strain HXB2 gag region
 target sequence 17catgcagggc ctattgcacc aggccagatg
 agagaaccaa ggggaagtga catagcagga 60actactagta
 cccttcagga 801839DNAArtificial
 SequenceDescription of Artificial SequenceHIV
 strain HXB2 gag region RT-PCR upstream primer
 18atagtaggt aggtaggtca tgcaggcct attgcacna
 391953DNAArtificial SequenceDescription of
 Combined DNA/RNA MoleculeHIV strain HXB2
 gag region RT-PCR downstream primer 19atcactcact
 cactcactcc tgaagggtac tagtagtcc tgctatgtca cnu
 532045DNAArtificial SequenceDescription of
 Artificial SequenceSNP R in NOS1_361 ATP ribo-
 PCR amplification primer Forward 1 20cctagaaact
 agaaactaga aactctgatg gctcaccatt gaaan
 452145DNAArtificial SequenceDescription of
 Artificial SequenceSNP R in NOS1_361 ATP ribo-
 PCR amplification primer Forward 2 21cctaaaaact
 aaaaactaaa aactctgatg gctcaccatt gaaan
 452220DNAArtificial SequenceDescription of
 Artificial SequenceSNP R in NOS1_361 ATP ribo-
 PCR amplification primer Reverse 22gtcaatgaag
 gaaggtagca 202310DNAArtificial
 SequenceDescription of Artificial Sequenceforward
 sense alkali cleavage fragment of ATP ribo-PCR of
 SNP R of NOS1_361 23cctcctca
 102411DNAArtificial SequenceDescription of
 Artificial Sequenceforward sense alkali cleavage
 fragment of ATP ribo-PCR of SNP R of NOS1_361
 24gggggctgct a 112534DNAArtificial
 SequenceDescription of Artificial Sequencereverse
 sense alkali cleavage fragment of ATP ribo-PCR of

SNP R of NOS1_361 25gtcaatgaag gaaggtagca
 gccccctct ttca 342634DNAArtificial
 SequenceDescription of Artificial Sequencereverse
 sense alkali cleavage fragment of ATP ribo-PCR of
 SNP R of NOS1_361 26gtcaatgaag gaaggtagca
 gcccccttt ttca 342745DNAArtificial
 SequenceDescription of Artificial SequenceSNP R in
 NOS1_361 ATP ribo-PCR amplification primer
 forward sense 27cctagaaact agaaactaga aactctgatg
 gctcaccatt gaaaa 452846DNAArtificial
 SequenceDescription of Artificial SequenceSNP R in
 NOS1_361 ATP ribo-PCR amplification primer
 forward sense 28cctaaaaact aaaaactaaa aactctgatg
 gctcaccatt gaaaga 46

Annexe 5: High specificity single-tube multiplex genotyping using ribo-PAP PCR, tag primers, alkali cleavage of RNA/DNA chimeras and MALDI-TOF MS

Soumise Juin 2012 Human Mutation

Florence Mauger¹, David H. Gelfand¹, Amar Gupta², Veeraiah Bodepudi², Stephen G. Will², Keith Bauer², Thomas W. Myers², and Ivo G. Gut^{1,3*}

¹CEA/Institut de Génomique/Centre National de Génotypage, Bâtiment G2, 2 rue Gaston Crémieux, 91057 Evry Cedex, France

²Roche Molecular Systems Inc., 4300 Hacienda Drive, Pleasanton, California 94588, USA

³Centro Nacional de Análisis Genómico, PCB, C/Baldiri Reixac 4, 08028 Barcelona, Spain

Abstract

Here, we describe a high throughput, single-tube, allele-specific ribonucleotide analog pyrophosphorolysis activated polymerization (ribo-PAP) PCR multiplex genotyping and re-sequencing method. An RNA/DNA chimeric PCR product is generated using genomic DNA as starting template, a panel of allele-selective 5'-tagged primers, a reverse primer, one nucleotide in the ribo-form (90-100%), the other nucleotides in the deoxy-form, a DNA polymerase capable of incorporating ribonucleotides, a suitable buffer and thermal cycling. The RNA/DNA chimeric PCR products are fragmented by treatment with alkali and analyzed by mass spectrometry. All allele-selective primers have a 5' repetitive motif where each repeat unit has a unique, distinct mass upon reverse copying and alkali fragmentation. The mass of the complement repeat fragment or flag identifies the primer or primers that were recruited in the ribo-PAP PCR. The method readily identifies homozygous and heterozygous positions in simplex or duplex ribo-PAP PCR. Many different tags can be analyzed simultaneously. The assay can genotype several SNPs in a single tube. It thus constitutes the simplest genotyping protocol with multiplex analysis. This novel genotyping and re-sequencing protocol was applied to different genomic loci: *NOS1* and *H19* in 30 individuals in simplex ribo-PAP PCR and at two *SLCO1B1* loci in 95 individuals in duplex ribo-PAP PCR.

Keywords: Polymorphism, ribo-PAP, ribo-PCR, chimeric RNA/DNA, MALDI-TOF mass spectrometry

Abbreviations: PAP = pyrophosphorolysis-activated polymerization; ribo-PAP = ribonucleotide analog pyrophosphorolysis activated polymerization; ribo-PCR = RNA/DNA chimeric PCR; primer tag sequence = a 6-7 nucleotide sequence with varying A:G ratio and flanked by single CT; specific mass flag = a 6-7 nucleotide sequence with varying C:T ratio and flanked by ribo A or G; MALDI-TOF MS = matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; CPG = controlled pore glass.

Introduction

We have previously described a novel, rapid and accurate concept for DNA genotyping (Mauger et al., 2006) and sequencing (Mauger et al., 2007; Mauger et al., submitted or in press 2011) using mass spectrometry and RNA/DNA chimeras. Our method is based on preparing and alkali fragmenting single primer RNA/DNA chimeric cyclized extension products with ribonucleotides only (1) or cyclized single primer extension products (2) or PCR products (3) containing three dNMPs and one NMP produced with a DNA polymerase that also readily incorporates ribonucleotides.

We applied this method for determining the complete genotype in the hypervariable region I of human mitochondrial DNA combined with MALDI-TOF mass spectrometry for detection (Mauger et al., 2007). We took advantage of a novel class of thermostable DNA polymerases that efficiently utilize NTPs and rapidly extend a primer to generate chimeric primer extension products. Following alkaline hydrolysis and using the resolving power and precision of the mass spectrometer, the masses of the fragments were determined by analysis of differentially substituted (e.g., A, C, G) extension products. We were able to unambiguously determine the sequence of the original locus. Subsequently (Mauger et al., 2011), using only a single reaction we demonstrated that we could generate chimeric RNA/DNA PCR product directly from genomic DNA, simply substituting one of the dNTPs with the corresponding NTP (e.g., A, C, G), fragmenting the resulting amplicon and correctly identifying the sequence polymorphisms among the 3'-NMP-terminated fragments.

Here we sought to develop a single-tube multiplex genotyping method based on the combination of ribo-PCR with an allele selection procedure. Identification of the recruited primer in the allele-specific procedure is achieved by "labeling" each primer with a repeat sequence at its 5' end that upon reverse copy and cleavage gives rise to a product of distinct mass. Allele-specific procedures with tags that report on recruited primers have been described before (Germer and Higuchi, 1999). In this case, one allele-specific primer has a 5'GC-tail that results in differential melting temperature of the two allelic PCR products. Allele-specificity is achieved by complementarity of the most 3' base of the primers to one of the alleles of the polymorphism that is being genotyped. Only the fully complementary primer is extended efficiently and fully copied in the reverse reaction. Another way that has been proposed for allele-distinction is based on the use of primers that are refractive to extension due to the absence of a 3'OH group which is required for extension. By using a DNA

polymerase that incorporates efficiently dideoxynucleotides, the last base of the primer is removed by pyrophosphorolysis, but selectively from the fully complementary primer. Thus, pyrophosphorolysis activates the primer for extension. This procedure is termed Pyrophosphorolysis-Activated Polymerization (PAP, Liu and Sommer, 2000). The main difference between the allele-specific PCR procedure of Germer and Higuchi and PAP is that for PAP an error in primer recruitment does not alter the intended template sequence, while for the Germer and Higuchi procedure mis-extension results in generating a template of the opposite allele. Thus the PAP concept is inherently more resistant to this error. Our approach is based on the PAP concept, but we deviated through the inclusion of primers that are more easily synthesized and are compatible with ribo-PCR and the DNA polymerase used therein. Finally, while it is difficult to multiplex genotype with the prior procedures, we demonstrate the ability to multiplex genotype with our ribo-PAP method.

Thermotoga maritima (*Tma*) ribonucleotide-incorporating thermostable DNA polymerases can utilize 2'-PO₄, 3'-OH, 5'-triphosphate ribonucleotide analogs as substrates. When incorporated, the 2'-PO₄ moiety prevents further extension at the 3'-OH position (Gelfand et al., 2009). In addition, the 2'-PO₄ substitution prevents proofreading by the 3'-5' exonuclease activity of *Tma* DNA polymerase family enzymes (Gelfand, et al 2007; Gelfand et al., 2009). In the presence of pyrophosphate and a template, a *Tma* DNA polymerase with the ability to incorporate ribonucleotides and ribonucleotide analogs is able to remove a complementary 2'-PO₄ 3'-non-extendable nucleotide from a primer by pyrophosphorolysis (Gelfand et al., 2007). We have developed convenient methods to synthesize oligonucleotide primers that contain 2'-PO₄ and 3'-OH groups at the 3'-end (Bodepudi et al., 2011). These non-extendable primers can be used to effect ribonucleotide analog, pyrophosphorolysis-activated polymerization (ribo-PAP) in PCR (Gelfand et al., 2007). In this work, we use this new allele-selective primer "hot start" method, which significantly improves the specificity and selectivity of ribo-PCR assays, to enable a simple means of determining which primers participated in the amplification reaction. The allele-selective primers contain a different, short, repeating sequence in the 5' portion of the primer that serves as a mass tag. Following ribo-PAP PCR in the presence of one ribonucleotide and alkali treatment, a fragment pattern is generated that allows unambiguous identification of which primer(s) participated in the

amplification reaction and the polymorphism(s) that was (were) amplified. Further, we demonstrate that this direct PCR genotyping assay can be readily multiplexed using the presence or absence of the different flag fragments for the identification of different genotypes for the different assays of the multiplex in a single procedure.

This procedure provides a DNA genotyping protocol with a single enzymatic step and three reagent addition steps to the genomic DNA template (ribo-PAP PCR mix, alkali solution and ion exchanged resin). To our knowledge this constitutes the simplest DNA multiplex-capable genotyping protocol. The 5'-tag, 2'-PO₄ primer, ribo-PAP PCR protocol was applied to different genomic loci involving *NOS1* *163731 and *H19* *103280 in 30 individuals in simplex ribo-PAP PCR and at two different loci simultaneously involving *SLCO1B1* *604843 in 95 individuals in duplex ribo-PAP PCR to assist in the determination of individuals at risk from high dose statin treatment. Variants in *SLCO1B1* have been associated with a dramatically increased risk to develop myopathies when treated with the cholesterol-lowering drug Simvastatin (Link et al., 2008).

Materials and Methods

Thirty Caucasian randomly selected DNA samples (3) and 95 DNA samples with a history of myocardial infarction were used (Link et al., 2008).

5'-tag-2'-PO₄ Primers

Oligodeoxyribonucleotide primers were synthesized on an Applied Biosystems 394 synthesizer. The solid support was a 3'-phosphate CPG from Glen Research (Sterling, VA USA) and the first cycle of synthesis used a ribonucleoside-3'TBDMS-2'-O-phosphoramidite from ChemGenes (Wilmington, ME USA). Standard deoxynucleoside 3'-phosphoramidites were used for the second and subsequent cycles of synthesis. Allele-selective primers have a three times repeating, 5'-tag sequence of six or seven nucleotides in length (Table 1). Each allele-selective primer has a distinct mass tag in the 5' portion of the allele-selective primer. The short tags are separated by the complement of the ribonucleotide base that is used in the ribo-PAP PCR. Upon fragmentation with alkali, the ribo-PCR product yields a repeating flag fragment of defined mass that is indicative of the

primer(s) that had (have) been selected for amplification in the ribo-PAP PCR.

As the protected primers have 2'-PO₄ and 3'-OH moieties, care must be taken during deprotection, purification, and storage to prevent an inadvertent isomerisation or loss of the 2'-phosphate group. In order to eliminate the possibility of isomerisation, i.e., a migration of the phosphate group from the 2'- position to the 3'- position on the sugar, the alkaline cleavage and deprotection step of primer synthesis should be performed first using mild basic conditions, followed by the removal of the 3'- TBDMS group with fluoride. Sterile and phosphatase free buffers must be used during purification and storage to prevent loss of the 2'-phosphate blocking group. The primers were purified on a Waters Xterra reverse phase C-18 column using a 24-30% gradient of methanol in 400 mM 1,1,1,3,3,3-hexafluoro-2-propanol solution adjusted to pH 7.9 with triethylamine. Finally, care is taken to avoid phosphatase contamination.

FP-1 DNA Polymerase

Roche Molecular Systems (Pleasanton, CA, USA) provided FP-1 DNA polymerase. This designer DNA polymerase is a chimeric DNA polymerase comprising the 5'-nuclease domain of *Thermus* sp. Z05 DNA polymerase and the 3'-nuclease and 5' to 3' DNA polymerase domains of *Thermotoga maritima* (*Tma*) DNA polymerase. The 5'-nuclease activity is eliminated in this enzyme by the introduced G46E mutation (1). The wild-type *Tma* proofreading activity has been modulated in FP-1 DNA polymerase by introduction of an L329A mutation (Schönbrunner et al., 2006) in "Motif I" of the 3' to 5' exonuclease domain. The enzyme's polymerase domain contains a mutation that eliminates the selectivity of the wild-type enzyme against incorporation of C'-2-substituted nucleotides and which facilitates efficient incorporation of ribonucleoside triphosphates (Mauger et al., 2006, 2007, 2011). Finally, the enzyme contains three additional mutations in the polymerase domain, which enhance binding to template and result in faster extension rates, particularly in high ribonucleotide substitution reactions, and enhance the rate of pyrophosphorolysis (Bauer and Gelfand, 2009). Please contact Thomas Myers (Thomas.Myers@Roche.com) at Roche Molecular Systems for research samples.

The deoxynucleoside triphosphates, dNTPs, (N = A, C, G, T) and ribonucleoside triphosphates, NTPs, (N

= A, C, G) were purchased from GE Healthcare (Saclay, France), ion exchange resin AG 50W-X8 from BioRad (Marmes la Coquette, France). MassArray® Clean Resin Tool Kit was purchased from Sequenom (Hamburg, Germany) and general chemical reagents were from Aldrich (Steinheim, Germany). Thermocycling procedures were carried out in Eppendorf Gradient Thermocyclers (Eppendorf, Germany).

Simplex ribo-PAP PCR amplifications were carried out in 20 µl reactions, with 20 ng human genomic DNA, 0.4 µM of each *NOS1* or *H19* primer (Table 1), 100 mM tricine at pH 7.3, adjusted with KOH, 100 mM potassium acetate at pH 7.5, 3 mM magnesium acetate, 0.15 mM sodium pyrophosphate, 200 µM each dCTP, dGTP, dTTP and ATP for *NOS1* or 200 µM each dCTP, dGTP, dTTP, 20 µM dATP and 180 µM ATP for *H19* and 5 units of 10 U/µl FP-1 DNA polymerase.

Duplex ribo-PAP PCR amplifications were the same as for simplex ribo-PAP PCRs except *SLCO1B1* rs4363657 primers (Table 1) were at 0.4 µM and *SLCO1B1* rs4149056 primers were at 0.2 µM, 200 µM each dATP, dCTP, dTTP, 20 µM dGTP and 180 µM GTP. The thermal cycling profile for the PCR was 4 min at 92°C followed by 46 cycles of 15 s at 92 °C, 4 min at T_{anneal-extend} °C (63 °C for *NOS1*, 66 °C for *H19* and 60 °C for *SLCO1B1*). Five µl of the PCR was analyzed on a 2% agarose gel to verify that the desired PCR products were generated.

Alkali cleavage, sample desalting, preparation of matrix, conditioning and MALDI-TOF analysis were as previously described (Mauger et al., 2007; Mauger et al., 2011; Zhu et al., 996). We used software similar to that described previously (Mauger et al., 2007; Mauger et al., 2011) to predict all expected fragments resulting from alkali hydrolysis of ribo-PCR products.

Results

The principle of ribo-PAP PCR multiplex genotyping and re-sequencing developed here is cleavage of RNA/DNA chimeric PCR product and analysis of fragment fingerprints by MALDI-TOF mass spectrometry (Figure 1). The ribo-PAP PCR amplifications are carried out with a reaction mixture containing up to 100% of one ribonucleotide and the three other nucleotides in their deoxyribo-form with a thermostable, ribonucleotide-incorporating DNA polymerase with

modulated proof-reading activity, on genomic input DNA with two blocked allele-specific primers and a reverse primer for each SNP that is genotyped.

As previously reported, the polymerase KB17 can readily incorporate up to 100% ATP, 100% CTP or 100% GTP in primer extension in HVI mitochondrial DNA (2) and in ribo-PCR on genomic DNA template (*NOS1* and *H19* with ATP and *SLCO1B1* rs4363657 with GTP or CTP) (Mauger et al., 2011). Here, we achieved the preparation of ribo-PAP PCR of about 80 bp with the polymerase FP-1 in simplex or duplex PCR. For *NOS1*, the ribo-PAP PCR product is generated with 100% ATP, and there are eight bases between forward and reverse primers. For *H19*, the ribo-PAP PCR product is generated with 90 % ATP and 10% dATP, and there are fifteen bases between forward and reverse primers. For the duplex *SLCO1B1*, the ribo-PAP PCR product is generated with 90% GTP and 10% dGTP, and there are eight and fourteen bases between the forward and reverse primers respectively for rs4363657 and rs4149056. In duplex, there is twice the concentration of rs4363657 primers as rs4149056 primers.

We use a different sequence-specific repeating tag on each of the allele-specific PCR primers. All tags are chosen to have a mass specific RNA/DNA fragmentation flag and be distinct from any other mass fragments from the re-sequenced loci. All tags are designed (1) to have a different ratio of A to G and (2) to contain only one T, C at the end of a repeat to work with ATP or GTP in ribo-PAP PCR (Table1). All tags are repeated three times and flags are in the 6 or 7-mer region to provide facile detection in the MALDI-TOF MS.

Fragments for the mass spectrometric analysis are generated by alkali cleavage with sodium hydroxide. This results in the cleavage 3' of each incorporated ribonucleotide. Under these conditioning and analysis methods, a residual 2' or 3' phosphate group remains at the 3'terminal base. From the mass, we cannot distinguish 2'-H₂PO₄ + 3'-OH from 2'-OH + 3'-H₂PO₄. Quite likely, products are a mixture of these two species. Predicted fragments are listed in Table 2.

Due to the negative charges on the sugar phosphate backbone, adducts of the phosphate groups with Na⁺ are observed when samples are not sufficiently desalted. The most common method for desalting, the addition of a protonated ion exchange resin such as the MassArray® Clean

Resin Tool Kit (www.sequenom.com) was used here.

Due to the mass distributions (Table 2), samples were analyzed in linear negative ion mode in the 6-7-mer region to analyze the alkali-released flags for genotyping and in a larger mass range (1-10 kDa) for re-sequencing.

Figure 2 shows the mass spectra, in 7-mer region, of *NOS1* rs3741476 with a peak at 2182.4 Da for three units of GTTCTA and a peak at 2197.4 Da for three units GTTTTA which correspond to the flags of A and G allele, respectively, of the SNP R. The homozygote A (Figure 2a), the heterozygote R (Figure 2b) and the homozygote G (Figure 2c) can be analyzed and identified.

In mass spectra, in 14-mer region, of *NOS1*, there are also peaks from the re-sequencing that identify the rs3741476 at 4222.7 Da for the fragment GCCCCTTCTTCA that identifies allele G and a peak at 4236.7 Da for the fragment GCCCCTTTTCA that identifies allele A.

Figure 3 shows mass spectra, in 6-mer region, of *H19* rs2839698 with a peak at 1863.2 for three units GCTCTA and a peak at 1878.2 Da for three units GTCTTA, which correspond to the flags of T and C allele, respectively, of the SNP Y. The heterozygote Y (Figure 3a), the homozygote C (Figure 3b) and the homozygote T (Figure 3c) can be analyzed and identified

In mass spectra of *H19*, in the 10 and 14-mer regions, there are also peaks from the re-sequencing that identify the rs2839698 at 3140.0 Da for the fragment TGCCTGGGCA that identifies allele T and a peak at 4351.8 Da for the fragment TGCCTGGGCCTA that identifies allele C.

Figure 4 shows mass spectra, in 6 and 7-mer regions, of the duplex of *SLCO1B1* rs4363657 and rs4149056. In 6-mer region, spectra show peaks at 1848.2 Da for three units of CTCCAG and a peak at 1863.2 Da for three units of CTCTAG, which correspond to the flags of the C and T alleles, respectively, for the SNP Y of rs4363657. In 7-mer region, spectra show peaks at 2152.4 Da for three units TCCTCAG and a peak at 2167.4 Da for three units TCTCTAG, which correspond to the flags of T and C allele, respectively, for the SNP Y of rs4149056. Samples that are homozygote T/T (Figure 4a), the heterozygote (Figure 4b) and the homozygote T/T (Figure 4c) can be analyzed and identified.

In 6 and 7-mer regions, spectra also show peaks that come from the re-sequencing of rs4149056: a peak at 1878.2 Da for the fragment TTCATG, a peak at 2215.4 Da for fragments ATATATG and TAATATG (Figure 4).

In mass spectra, in 16-mer region, of *SLCO1B1*, there are also peaks from the re-sequencing that identify the rs4363657 at 4869.1 Da for the fragment TACCCACTCTTTTTTG that identifies Allele C and a peak at 4884.2 Da for the fragment TACTACTCTTTTTTG that identifies Allele T. In 12-mer region, *SLCO1B1*, there are also peaks from the re-sequencing that identify the rs4149056 at 4302.8 Da for the fragment CATATCCACATG that indicates the presence of allele C.

None of the samples were homozygous at one *SLCO1B1* position and heterozygous at the other *SLCO1B1* position. Accordingly, we infer that these two genotypes are in complete linkage disequilibrium.

In addition to the indicated tag sequences, we have evaluated the effect of switching the tag sequences for the *SLCO1B1* locus multiplex reactions. In all homozygous reaction the flag genotyping results were clear and unambiguous. Occasionally, there was preference for one flag over the other as in the genotyping for *NOS1*. We also saw comparable flag intensities when there were a similar number of T's in the respective flags.

The genotype of the polymorphism rs3741476 of *NOS1* and polymorphism rs2839698 of *H19* was analyzed with this method and was determined for 30 individuals. The identified polymorphisms were 100% validated by the Amplifluor[®] assay and with a ribo-PCR re-sequencing assay (Mauger et al., 2011). The genotype of the polymorphisms rs4363657 and rs4149056 of *SLCO1B1* were analyzed with this method and was determined for 95 individuals. The identified polymorphisms were 100% validated by the Illumina[®] Infinium[™] II assay and capillary electrophoresis (Link et al., 2008).

Discussion

Here we describe a novel one step multiplex DNA genotyping and re-sequencing method using a MALDI-TOF mass spectrometer for detection. Ribonucleotide-containing PCR products are generated using a genomic DNA template, two novel allele-selective primers, a common primer per SNP, one nucleotide in the ribo-form (90-100%), the other nucleotides in the deoxy-form, a ribonucleotide incorporating thermostable DNA

polymerase and a suitable buffer. The ribo-PCR products are fragmented by treatment with alkali, desalted and analyzed by mass spectrometry. Chromatographic, electrophoretic or other separation techniques prior to MS analysis are unnecessary. The mass fingerprint is used to identify the allele selective primer(s) that participated in the amplification reaction.

Our procedure takes advantage of a form of PAP (Liu and Sommer, 2000; Liu and Sommer, 2002; Qin et al., 2007) that confers very high specificity to PCR amplification assays and remarkable increases in allele-selective discrimination (Liu and Sommer, 2002; Qin et al., 2007). With PAP, all primers are “blocked” and cannot be extended by a DNA polymerase. Upon binding to a complementary target region and in the presence of pyrophosphate, the primer 3-terminal, non-extendable nucleotide can be removed by pyrophosphorolysis as a 5'-triphosphate. The DNA polymerase is then able to add a “normal” nucleotide to initiate primer extension. The primer pyrophosphorolysis activation step is repeated in every cycle, imposing the high specificity requirement for each cycle.

The early descriptions for PAP utilized 3'-dideoxy terminated oligonucleotide primers and a mutant thermostable DNA polymerase that efficiently incorporated dideoxynucleotides, but in the presence of pyrophosphate efficiently removed a primer 3'-terminal ddNMP (Liu and Sommer, 2002; Qin et al., 2007). Efficient pyrophosphorolysis has a very strong requirement for 3'-matched primer-template (Liu and Sommer, 2002; Qin et al., 2007). The 3'-dideoxy terminated primers are synthesized enzymatically with terminal transferase (Liu and Sommer, 2000) or usually require synthesis the 5' to 3' direction with 5'-phosphoramidites (Liu and Sommer, 2002) since not all dideoxy bases for conventional 3' – 5' synthesis are commercially available.

After creating thermostable DNA polymerases that could efficiently incorporate ribonucleotides and C'2-substituted ribonucleotide analogs, we found certain ribonucleotide analogs, e.g., 2'-PO₄, 3'-OH, 5'-nucleoside triphosphates, having intact ribose sugar rings arrested all further extension when incorporated, even though a 3'OH moiety was present on the sugar (Gelfand et al., 2009). This led us to synthesize oligonucleotide primers having a 3'-terminal 2'-PO₄, 3'-OH nucleotide (Bodepudi et al., 2011) and evaluate the ability of ribonucleotide-incorporating thermostable DNA polymerases for ribo-PAP primer hot start (Gelfand et al., 2007). The ribo-PAP primers are easily synthesized on automated synthesizers in the conventional 3'-to-5' direction with commercially available CPGs and 3'-phosphoramidites. This ensures that all primers have a 3'-blocking group and eliminates the requirement for acrylamide gel electrophoresis and

purification. In contrast to “dideoxy-PAP” where the dNTP pool size is fairly low, thus increasing the likelihood of reincorporation of a released ddNTP and prematurely terminating extension, dNTP or NTP pool size is much higher (200 μM each or more) minimizing the probability of reincorporating a chain terminating tetraphosphate nucleoside. For *H19*, 90 % ATP and 10 % dATP is used, for *SLCO1B1* 90 % GTP and 10 % dGTP is used, but neither partial substitution with the corresponding dNTP interferes with the interpretation of spectra. Further, whereas the ddNMP-incorporating thermostable enzymes lack a 3'-5' exonuclease (“proofreading”) activity, FP-1 DNA polymerase contains a modulated (Schönbrunner et al., 2006) proofreading activity and is able to remove misincorporated nucleotides. These features of ribo-PAP PCR significantly increase the efficiency of each cycle and contribute to the fewer cycles that are required in a ribo-PAP PCR. Finally, it was unnecessary to design each of the ribo-PAP primers to have allele-selectivity as in “bi-PAP” (Qin et al., 2007; Liu et al., 2004), which permits interrogation of additional genomic sequences and may generate additional internal fragments that serve to confirm the genotyping assignment of the flag.

As previously reported, products were analyzed by MALDI-TOF mass spectrometry in linear mode. The linear mode allows covering a larger mass range (1-10 kDa) but does not resolve different species with similar masses in the high mass range. In ribo-PAP PCR, there is a flag for multiplex genotyping in 6 and 7-mer region. Accordingly, this method enables re-sequencing and multiplex genotyping in a single reaction. Moreover, with the use of different mass tags in the primers, this procedure is amenable to multiplex PCR for genotyping. Thus, in a single reaction, we conveniently genotyped both of the *SLCO1B1* SNPs that have been associated with adverse events following high dose statin treatment (Link et al., 2008).

As fragmentation uses facile, inexpensive and complete alkali cleavage, base-specific nucleases or other enzymes are avoided. The FP-1 DNA polymerase efficiently and rapidly incorporates the provided NTPs; expensive nucleotide analogues (e.g., α-S-CTP, α-S-UTP or fluorescently labelled nucleotides) are unnecessary. In contrast to other enzymes used in genotyping, sequencing or re-sequencing, FP-1 DNA polymerase contains a 3'-5' exonuclease or proofreading activity that assists in maintaining higher fidelity in ribo-PAP PCR. Finally, in contrast to many proofreading thermostable DNA polymerases, FP-1 DNA polymerase, like other *Tma* active site DNA polymerase enzymes (Bodepudi et al., 2011), is fully compatible with using dUTP in the PCRs, which allows it to be used together for carry over

contamination procedures with dUTP and uracil-N-glycosylase (UNG).

The presented procedure constitutes another addition to the ever-growing arsenal of DNA analysis methods. Its main advantage over DNA multiplex genotyping and re-sequencing methods using mass spectrometric detection lays in the ease of operation, ability to genotype multiple independent loci simultaneously in a single reaction and probably lower cost than current methods.

Acknowledgements

This work was supported by the French Ministry of Education Research (Ministère Délégué à l'Enseignement et à la Recherche) and the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n° Health-f4-2008-201418 - READNA.

References

- Bauer KA, Gelfand DH. 2009. Mutant DNA polymerases with improved pyrophosphorolysis (PAP) ability. US Patent Office Publication 20090155802.
- Bodepudi V, Gupta A, Will S. 2011. Synthesis and composition of nucleic acids comprising 2' terminator nucleotides. US Patent No. 7,928,207.
- Gelfand DH, Bauer KA, Gupta AP, Bodepudi V, Niemiec J. 2007. 2'-Terminator related pyrophosphorolysis activated polymerization. US Patent No. 7,745,125.
- Gelfand DH, Reichert FL, Bodepudi V, Gupta A, Will S, Myers T. 2009. 2'-Terminator Nucleotide-Related Methods and Systems. US Patent No. 7,572,581.
- Germer S, Higuchi R. 1999. Single-tube genotyping without oligonucleotide probes. *Genome Res* **9**:72-78.
- Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, Gut IG, Lathrop M, Collins R. 2008. *SLCO1B1* variants and statin-induced myopathy--a genomewide study. *N Engl J Med* **359**:789-799.
- Liu Q, Sommer SS. 2000. Pyrophosphorolysis activated polymerization (PAP): Application to allele-specific amplification. *Biotechniques* **29**:1072-1076.
- Liu Q, Sommer SS. 2002. Pyrophosphorolysis-activatable oligonucleotides may facilitate detection of rare alleles, mutation scanning and analysis of chromatin structures. *Nucleic Acids Res* **30**:598-604.
- Liu Q, Sommer SS. 2004. Detection of extremely rare alleles by bidirectional pyrophosphorolysis-activated polymerization allele-specific amplification (Bi-PAP-A): measurement of mutation load in mammalian tissues. *Biotechniques* **36**:156-166.
- Mauger F, Jaunay O, Chamblain V, Reichert F, Bauer K, Gut IG, Gelfand DH. 2006. SNP genotyping using alkali cleavage of RNA/DNA chimeras and MALDI time-of-flight mass spectrometry. *Nucleic Acids Res* **34**:e18.
- Mauger F, Bauer K, Calloway CD, Semhoun J, Nishimoto T, Myers TW, Gelfand DH, Gut IG. 2007. DNA sequencing by MALDI-TOF MS using alkali cleavage of RNA/DNA chimeras. *Nucleic Acids Res* **35**:e62.
- Mauger F, Bauer K, Semhoun J, Myers TW, Gelfand DH, Gut IG. 2011. Ribo-PCR - a facile method for the preparation of chimeric RNA/DNA applied to DNA sequencing. *Human Mutation Submitted (or In Press)*.
- Qin J, Calabrese P, Tiemann-Boege I, Shinde DN, Yoon S-R, Gelfand D, Bauer K, Arnheim N. 2007. The Molecular Anatomy of Spontaneous Germline Mutations in Human Testis. *PLoS Biology*. **5**:e224.
- Schönbrunner NJ, Fiss EH, Budker O, Stoffel S, Sigua CL, Gelfand DH, Myers TW. 2006. Chimeric thermostable DNA polymerases with reverse transcriptase and attenuated 3'-5' exonuclease activity. *Biochemistry* **45**:12786-12795.
- Zhu YF, Chung CN, Taranenko NI, Allman SL, Martin SA, Haff L, Chen CH. 1996. The study of 2,3,4-trihydroxyacetophenone and 2,4,6-trihydroxyacetophenone as matrices for DNA detection in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **10**:383-388.

Annexe 5: High specificity single tube multiplex genotyping using ribo-PAP PCR, tag primers, alkali cleavage of RNA/DNA chimeras and MALDI-TOF MS

Table 1 Sequences of primers of the ribo-PAP PCRs of *NOS1* rs3741476, *H19* rs2839698, *SLCO1B1* rs4149056 and *SLCO1B1* rs4363657. The sequence of the “tag” is underlined. The sequence of the alkali-hydrolyzed flag complement is indicated. * Denotes a 3’-C’2-PO₄ nucleotide.

| Name | PCR Primer | Tag | Flag |
|-----------------------------|---|----------------|---------|
| <i>NOS1</i> rs3741476 | <u>CCTAGAAACTAGAAACTAGAAACTCTGATGGCTCACCATTGAAA*</u> | CCT(AGAAACT)3C | GTTTCTA |
| | <u>CCTAAAAACTAAAAACTAAAAACTCTGATGGCTCACCATTGAAA*</u> | CCT(AAAAACT)3C | GTTTTTA |
| | GTCAATGAAGGAAGGTAGCA* | | |
| <i>H19</i> rs2839698 | <u>CCTAAGACTAAGACTAAGACTCTGTGAGGAGTGTGGAGTAGGC*</u> | CCT(AAGACT)3C | GTCTTA |
| | <u>CCTAGAGCTAGAGCTAGAGCTCTGTGAGGAGTGTGGAGTAGGT*</u> | CCT(AGAGCT)3C | GCTCTA |
| | GCTGATGTCGCCCTGTCT* | | |
| <i>SLCO1B1</i> rs4149056 | TTTAAAGGAATCTGGGTCATA* | | |
| | <u>CCTAGAGACTAGAGACTAGAGACTCAAGCATATTACCCATGAACA*</u> | CCT(AGAGACT)3C | TCTCTAG |
| | <u>CCTGAGGACTGAGGACTGAGGACTCAAGCATATTACCCATGAAC*</u> | CCT(GAGGACT)3C | TCCTCAG |
| <i>SLCO1B1</i> rs4363657 | CAATTTTTCAGAATAATTTAGTACA* | | |
| | <u>CCTAGAGCTAGAGCTAGAGCTCTGGAAAATACAAAAAAGAGTGA*</u> | CCT(AGAGCT)3C | CTCTAG |
| | <u>CCTGGAGCTGGAGCTGGAGCTCTGGAAAATACAAAAAAGAGTGG*</u> | CCT(GGAGCT)3C | CTCCAG |

Table 2 Sequences and masses of the ribo-PAP PCR fragmentation products: a. *NOS1* rs3741476 with ATP; b. *H19* rs2839698 with ATP; c. *SLCO1B1* rs4363657 and rs4149056 with GTP. Primer nucleotides are in *italics*. The polymorphic nucleotide is in bold and underlined

2a

| <i>NOS1</i> rs3741476 | Sequence | Mass | Start | End | Primer Strand |
|-----------------------|--|---------|-------|-----|---------------|
| | tca | 940.6 | 28 | 26 | REVERSE |
| | gccca | 1254.8 | 32 | 29 | REVERSE |
| | ttga | 1284.8 | 69 | 72 | FORWARD |
| | tggtga | 1943.2 | 38 | 33 | REVERSE |
| Flag A | 3 gtttcta | 2182.4 | 23 | 17 | REVERSE |
| | | | 9 | 10 | REVERSE |
| | | | 9 | 3 | REVERSE |
| Flag G | 3 gttttta | 2197.4 | 23 | 17 | REVERSE |
| | | | 9 | 10 | REVERSE |
| | | | 9 | 3 | REVERSE |
| G | ccttccttca | 3009.9 | 59 | 68 | FORWARD |
| | gggggctgcta | 3509.2 | 48 | 58 | FORWARD |
| A | gccccctt ttt ca | 4221.7 | 53 | 40 | REVERSE |
| | gccccctt ttt ca | 4236.7 | 53 | 40 | REVERSE |
| | gtcaatgaaggaaggtagc a | 6335.1 | 73 | 54 | REVERSE |
| A | cctagaactagaaactagaaactctgatggctcaccattgaaa a | 13919.1 | 1 | 45 | FORWARD |
| G | cctaaaaactaaaaactaaaaactctgatggctcaccattgaaa ga | 14200.3 | 1 | 46 | FORWARD |

2b

Annexe 5: High specificity single tube multiplex genotyping using ribo-PAP PCR, tag primers, alkali cleavage of RNA/DNA chimeras and MALDI-TOF MS

| H19 rs2839698 | Sequence | Mass | Start | End | Primer Strand |
|----------------------|--|-------------|--------------|------------|----------------------|
| | tca | 940.6 | 71 | 73 | FORWARD |
| | cga | 965.6 | 54 | 52 | REVERSE |
| | cga | 965.6 | 23 | 21 | REVERSE |
| T | ccta | 1229.8 | 41 | 38 | REVERSE |
| | ggca | 1294.8 | 48 | 51 | FORWARD |
| | ctcca | 1519.0 | 37 | 33 | REVERSE |
| | | | 20 | 15 | REVERSE |
| Flag T | 3 gctcta | 1863.2 | 14 | 9 | REVERSE |
| | | | 8 | 3 | REVERSE |
| | | | 20 | 15 | REVERSE |
| Flag C | 3 gtctta | 1878.2 | 14 | 9 | REVERSE |
| | | | 8 | 3 | REVERSE |
| | ggcga | 1953.2 | 63 | 68 | FORWARD |
| | ctccta | 2112.4 | 30 | 24 | REVERSE |
| | tcgtgca | 2192.4 | 52 | 58 | FORWARD |
| T | tgctgggc a | 3140.0 | 51 | 42 | REVERSE |
| C | tgctgggc g ccta | 4351.8 | 51 | 38 | REVERSE |
| | gctgatg tc ccctg tc tgca | 6485.2 | 75 | 55 | REVERSE |
| C | cctaagactaagactaagactcgtgaggagtgtggagtagg cg ccca | 14682.5 | 1 | 47 | FORWARD |
| T | cctagagctagagctagagctcgtgaggagtgtggagtagg tg ccca | 14745.5 | 1 | 47 | FORWARD |

2c

| SLCO1B1 | Sequence | Mass | Start | End | Primer Strand |
|------------------|--|-------------|--------------|------------|----------------------|
| rs4149056 | cttg | 1260.8 | 53 | 56 | FORWARD |
| rs4149056 | tatg | 1284.8 | 21 | 18 | REVERSE |
| | | | 59 | 64 | FORWARD |
| | | | 65 | 70 | FORWARD |
| rs4363657 Flag C | 3 ctccag | 1848.2 | 71 | 76 | FORWARD |
| rs4149056 | accag | 1857.2 | 17 | 12 | REVERSE |
| | | | 59 | 64 | FORWARD |
| | | | 65 | 70 | FORWARD |
| rs4363657 Flag T | 3 ctctag | 1863.2 | 71 | 76 | FORWARD |
| rs4149056 | ttcatg | 1878.2 | 38 | 43 | FORWARD |
| | | | 59 | 65 | FORWARD |
| | | | 66 | 72 | FORWARD |
| rs4149056 Flag T | 3 tcctcag | 2152.4 | 73 | 79 | FORWARD |
| | | | 59 | 65 | FORWARD |
| | | | 66 | 72 | FORWARD |
| rs4149056 Flag C | 3 tcctag | 2167.4 | 73 | 79 | FORWARD |
| rs4149056 | atatatg | 2215.4 | 29 | 35 | FORWARD |
| rs4149056 | taatatg | 2215.4 | 46 | 52 | FORWARD |
| rs4363657 | aaaaattg | 2441.7 | 8 | 1 | REVERSE |
| rs4363657 | taccactg | 2754.8 | 32 | 24 | REVERSE |
| rs4363657 | tattccag | 2784.8 | 47 | 55 | FORWARD |
| rs4149056 | attccttaaa | 3290.2 | 11 | 1 | REVERSE |
| rs4149056 C | catatatccacatg | 4302.8 | 35 | 22 | REVERSE |
| rs4363657 | tactaaattattctg | 4637.0 | 23 | 9 | REVERSE |
| rs4363657 C | taccactctttttg | 4869.1 | 31 | 46 | FORWARD |
| rs4363657 T | tactcactctttttg | 4884.2 | 31 | 46 | FORWARD |
| rs4149056 | tttaaaggaaatctgggcat acatg | 7846.1 | 1 | 25 | FORWARD |
| rs4363657 | caattttcagaataatttagtac ag | 8093.3 | 1 | 26 | FORWARD |
| rs4363657 C | cctggagctggagctggagctcctggaatacaaaaaagagtg g | 13800.0 | 77 | 34 | REVERSE |
| rs4149056 C | cctagagactagagactagagactcaagcatattacccatgaac g | 13936.1 | 80 | 36 | REVERSE |
| rs4363657 T | cctagagctagagctagagctcctggaatacaaaaaagagtg ag | 14065.2 | 77 | 33 | REVERSE |
| rs4149056 T | cctgaggactgaggactgaggactcaagcatattacccatgaac ac atatatccacatg | 18236.9 | 80 | 22 | REVERSE |

Figure 1. Principle of the method: ribo-PAP PCR with allele-specific tag sequence primer and ATP, dCTP, dGTP and dTTP, cleavage of ribobases of the ribo-PCR and analysis of the allele-specific flag sequence with the MALDI-TOF MS.

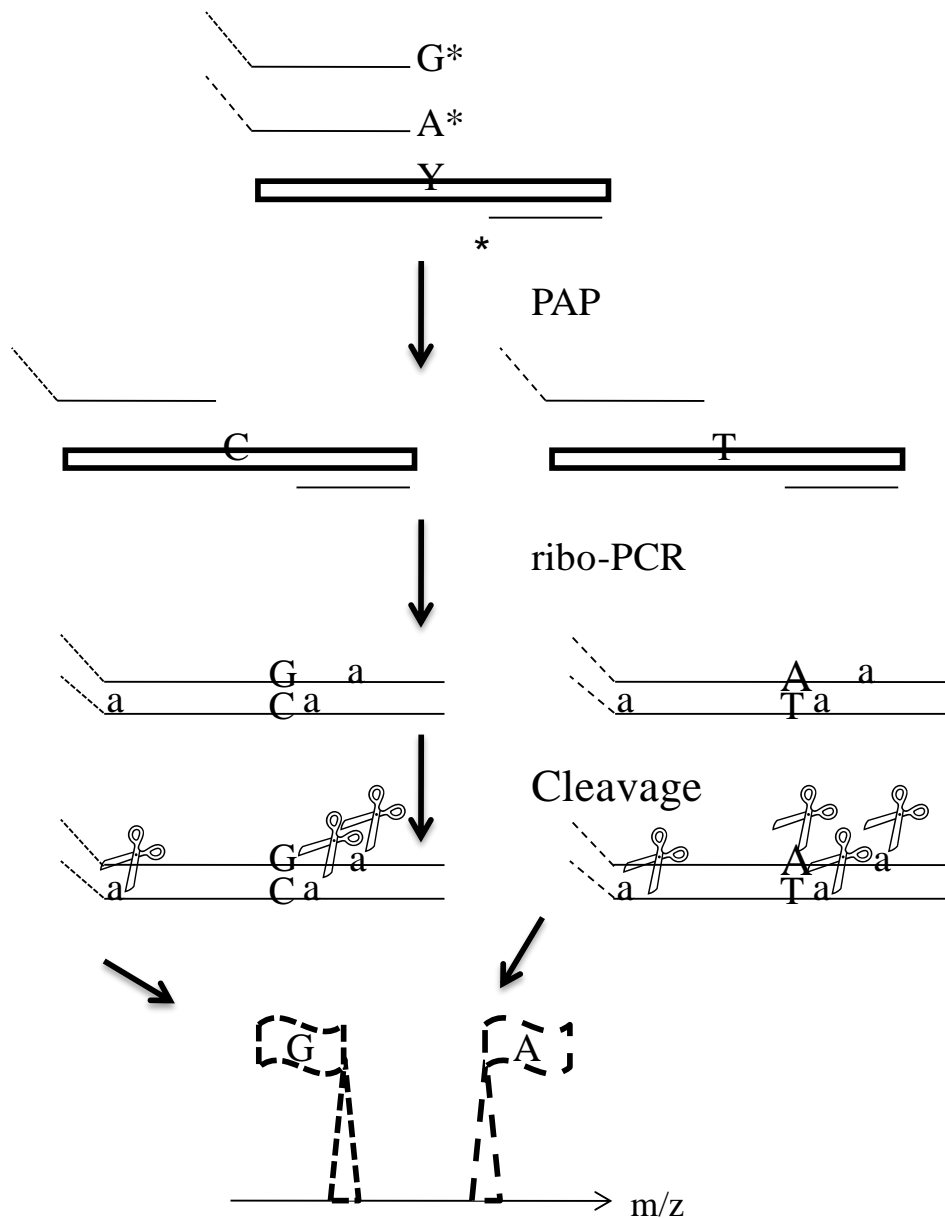


Figure 2. Mass spectra of the 7-mer region of the flag of the simplex ribo-PAP PCR of *NOS1* rs3741476 of: a) homozygote A; b) heterozygote R; c) homozygote G.

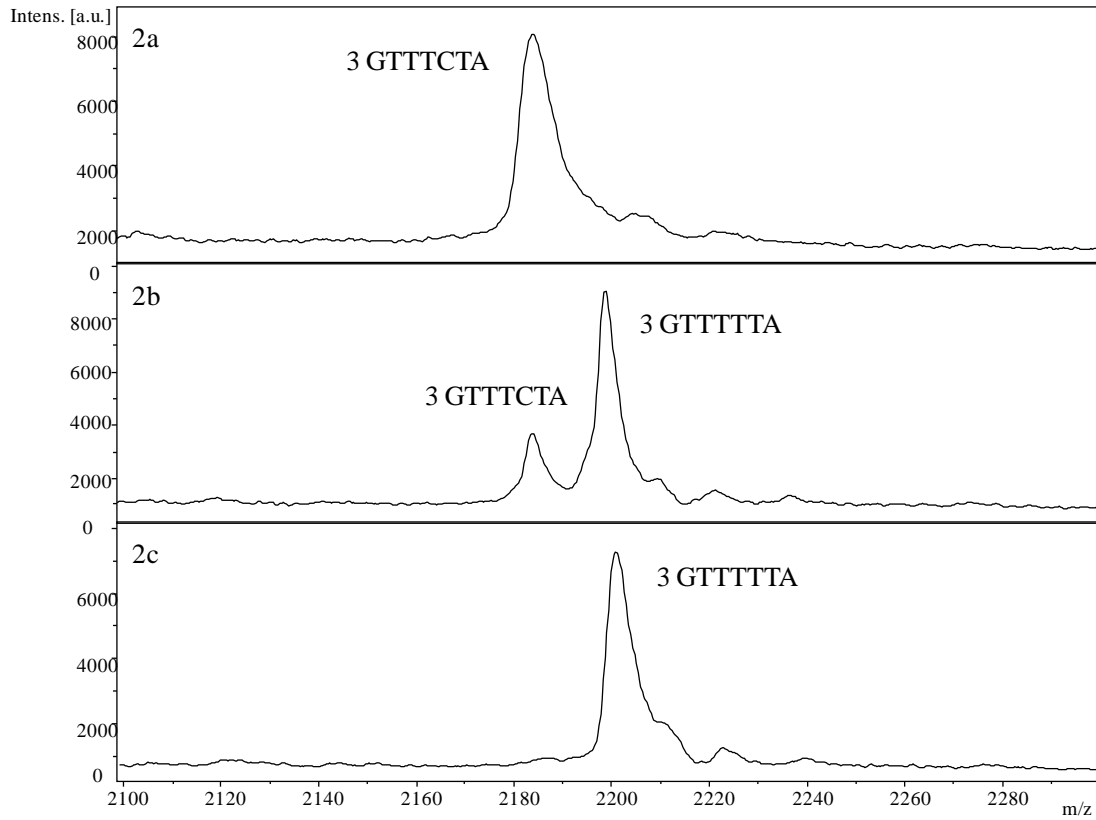


Figure 3. Mass spectra of the 6-mer region of the flag of the simplex ribo-PAP PCR of *H19* rs2839698 of: a) heterozygote Y; b) homozygote T; c) homozygote C.

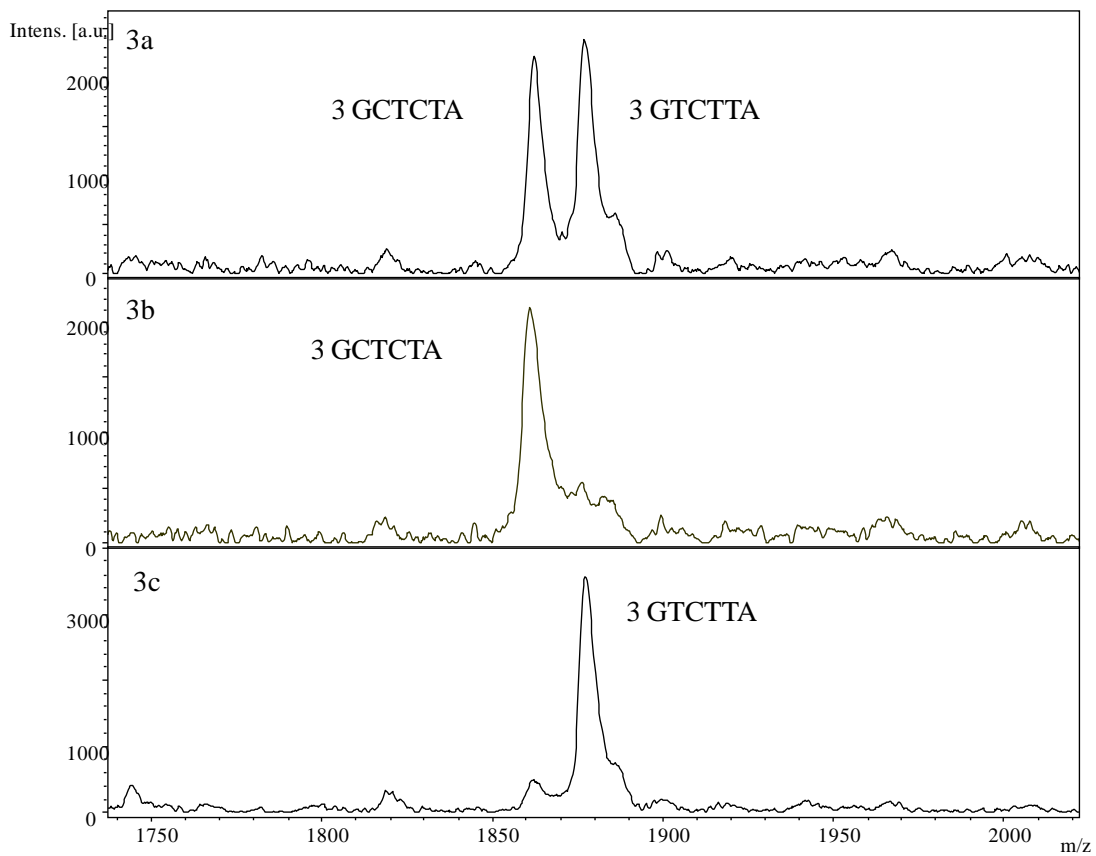
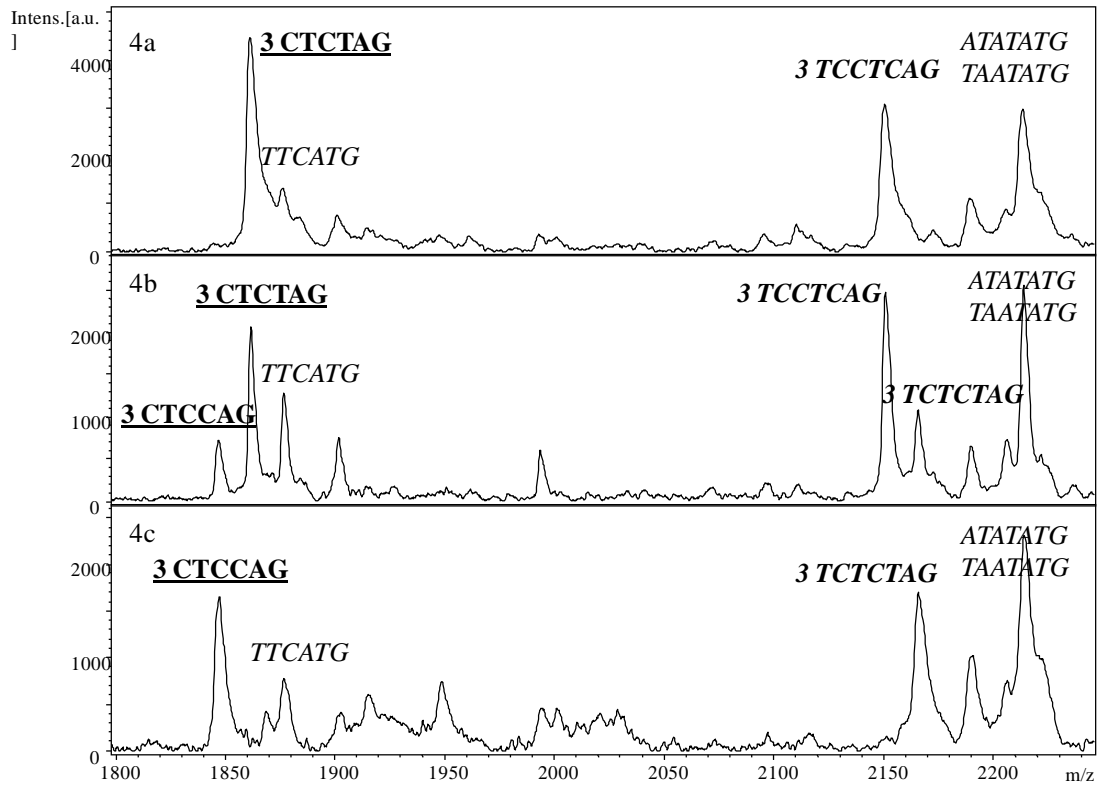


Figure 4. Mass spectra of the 6 and 7-mer region of the flag of the duplex ribo-PAP PCR of *SLCO1B1* rs4363657 and rs4149056 and fragments from the re-sequencing of rs4149056 of: a) homozygote TT; b) heterozygote TC; c) homozygote C. SNP rs4363657 is underlined and SNP rs4149056 is italic. The flag is bolded.



Références bibliographiques

Références bibliographiques

- Albert, T.J. et al., 2005.** Mutation discovery in bacterial genomes: metronidazole resistance in *Helicobacter pylori*. *Nature Methods*, 2(12), p.951–953.
- Albert, E., Baur, M. & Mayr, W., 1984.** *Histocompatibility Testings 1984*,
- Amado, F.M.L. et al., 1997.** Discrimination effects and sensitivity variations in matrix assisted laser desorption/ionization. *Rapid Communications in Mass Spectrometry*, 11(12), p.1347–1352.
- Antequera, F. & Bird, A., 1993.** Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A*, 90, p.11995–9.
- Armstrong, D.W. et al., 2001.** Ionic Liquids as Matrixes for Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal. Chem.*, 73(15), p.3679–3686.
- Barry, J.P. et al., 1995.** Mass and sequence verification of modified oligonucleotides using electrospray tandem mass spectrometry. *Journal of Mass Spectrometry*, 30(7), p.993–1006.
- Bauer, K.A. and G., 2009.** Mutant DNA polymerases with improved pyrophosphorolysis (PAP) ability. US8071536.
- Bellido, M.L. et al., 2010.** MALDI-TOF mass array analysis of RASSF1A and SERPINB5 methylation patterns in human placenta and plasma. *Biology of Reproduction*, 82(4), p.745–750.
- Bodepudi, V., Will, S.G. & Gelfand, D.H., 2011.** Synthesis and compositions of 2'-terminator nucleotides. US7947817.
- Van Berkel, G.J. & Zhou, F., 1995.** Characterization of an Electrospray Ion Source as a Controlled-Current Electrolytic Cell. *Anal. Chem.*, 67(17), p.2916–2923.
- Blazej, R.G., Kumaresan, P. & Mathies, R.A., 2006.** Microfabricated Bioprocessor for Integrated Nanoliter-Scale Sanger DNA Sequencing. *Proceedings of the National Academy of Sciences*, 103(19), p.7240–7245.
- Beavis, R.C. & Chait, B.T., 1989.** Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Communications in Mass Spectrometry: RCM*, 3(12), p.432–435.
- Böcker, S., 2004.** Sequencing from compomers: using mass spectrometry for DNA de novo sequencing of 200+ nt. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 11(6), p.1110–1134.

- Bordoli, R.S. et al., 1994.** Matrix-assisted laser desorption mass spectrometry on a magnetic sector instrument fitted with an array detector. *Rapid Communications in Mass Spectrometry*, 8(8), p.585–589.
- Boschenok, J. & Sheil, M.M., 1996.** Electrospray tandem mass spectrometry of nucleotides. *Rapid Commun Mass Spectrom*, 10, p.144–9.
- Breuker, K. et al., 2003.** Thermodynamic control of final ion distributions in MALDI: in-plume proton transfer reactions. *International Journal of Mass Spectrometry*, 226(1), p.211–222.
- Brown, R.S. & Lennon, J.J., 1995.** Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer. *Analytical Chemistry*, 67(13), p.1998–2003.
- Buetow, K.H. et al., 2001.** High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), p.581–584.
- Carothers, A.M. et al., 1989.** Point mutation analysis in a mammalian gene: rapid preparation of total RNA, PCR amplification of cDNA, and Taq sequencing by a novel method. *Biotechniques*, 7, p.494–6, 498–9.
- Coolen, M.W. et al., 2007.** Genomic Profiling of CpG Methylation and Allelic Specificity Using Quantitative High-Throughput Mass Spectrometry: Critical Evaluation and Improvements. *Nucleic Acids Research*, 35(18), p.e119.
- Cotter, R., 1989.** Time-of-flight mass spectrometry: an increasing role in the life sciences. *Biomedical & Environmental Mass Spectrometry*, 18(8), p.513–532.
- Cotter, R., 1992.** Time-of-flight mass spectrometry for the structural analysis of biological molecules. *Anal. Chem.*, 64(21), p.1027A–1039A.
- Cotter, R., 1999.** The new time-of-flight mass spectrometry. *Analytical Chemistry*, 71(13), p.445A–451A.
- Dai, Y., Whittal, Randy M. & Li, Liang, 1996.** Confocal Fluorescence Microscopic Imaging for Investigating the Analyte Distribution in MALDI Matrices. *Anal. Chem.*, 68(15), p.2494–2500.
- Deamer, D.W. & Akeson, M., 2000.** Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends in Biotechnology*, 18(4), p.147–151.
- Derrington, I.M. et al., 2010.** Nanopore DNA Sequencing with MspA. *Proceedings of the National Academy of Sciences*. Available at:
- Distler, A.M. & Allison, J., 2001.** 5-Methoxysalicylic acid and spermine: a new matrix for the matrix-assisted laser desorption/ionization mass spectrometry analysis of

- oligonucleotides. *J Am Soc Mass Spectrom*, 12, p.456–62.
- Douglas, D.J., Frank, A.J. & Mao, D., 2005.** Linear ion traps in mass spectrometry. *Mass Spectrometry Reviews*, 24(1), p.1–29.
- Dreisewerd, K. et al., 1995.** Influence of the laser intensity and spot size on the desorption of molecules and ions in matrix-assisted laser desorption/ionization with a uniform beam profile. *International Journal of Mass Spectrometry and Ion Processes*, 141(2), p.127–148.
- Dupont, J.M. et al., 2004.** De novo quantitative bisulfite sequencing using the pyrosequencing technology. *Anal Biochem*, 333, p.119–27.
- Edwards, C.A. & Ferguson-Smith, A.C., 2007.** Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol*, 19, p.281–9.
- Edwards, J.R., Ruparel, H. & Ju, J., 2005.** Mass-spectrometry DNA sequencing. *Mutation Research*, 573(1-2), p.3–12.
- Ehrich, M., Nelson, M.R., et al., 2005.** Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A*, 102, p.15785–90.
- Ehrich, M., Bocker, S. & van den Boom, D., 2005.** Multiplexed discovery of sequence polymorphisms using base-specific cleavage and MALDI-TOF MS. *Nucleic Acids Res*, 33, p.e38.
- Emrich, C.A. et al., 2002.** Microfabricated 384-Lane Capillary Array Electrophoresis Bioanalyzer for Ultrahigh-Throughput Genetic Analysis. *Anal. Chem.*, 74(19), p.5076–5083.
- Erlich, H.A. et al., 1991.** HLA-DR, DQ and DP typing using PCR amplification and immobilized probes. *European Journal of Immunogenetics: Official Journal of the British Society for Histocompatibility and Immunogenetics*, 18(1-2), p.33–55.
- Erlich, H.A., Opelz, G. & Hansen, J., 2001.** HLA DNA typing and transplantation. *Immunity*, 14(4), p.347–356.
- Fenn, J. B. et al., 1989.** Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*, 246(4926), p.64–71.
- Fologea, D. et al., 2005.** Detecting Single Stranded DNA with a Solid State Nanopore. *Nano Lett.*, 5(10), p.1905–1909.
- Forster, R.E. et al., 2008** Microchip-Based Sanger Sequencing of DNA. Dans *Next Generation Genome Sequencing*. Wiley-VCH Verlag GmbH & Co. KGaA, p. 153–163.
- Fournier, I. et al., 2002.** Delayed extraction experiments using a repulsive potential before ion extraction: evidence of clusters as ion precursors in UV-MALDI. Part I: dynamical effects with the matrix 2,5-dihydroxybenzoic acid. *International Journal of*

- Mass Spectrometry*, 213(2–3), p.203–215.
- Fournier, I. et al., 2003.** Irradiation effects in MALDI, ablation, ion production, and surface modifications. Part II. 2,5-dihydroxybenzoic acid monocrystals. *Journal of the American Society for Mass Spectrometry*, 14(8), p.893–899.
- Fu, Y. et al., 2006.** A matrix of 3,4-diaminobenzophenone for the analysis of oligonucleotides by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Nucleic Acids Res*, 34, p.e94.
- Gabelica, V. & Pauw, E.D., 2005.** Internal energy and fragmentation of ions produced in electrospray sources. *Mass Spectrometry Reviews*, 24(4), p.566–587.
- Gardiner-Garden, M. & Frommer, M., 1987.** CpG islands in vertebrate genomes. *Journal of Molecular Biology*, 196(2), p.261–282.
- Gelfand, D., 1999.** Thermostable DNA polymerases having reduced discrimination against ribo-NTPs. US5939292
- Gelfand, D. & Gupta, A., 2008.** Sequencing and Genotyping Using Reversibly Terminating Nucleotides.US20080293071.
- Gelfand, D. et al., 2009.** Ribonucleotide Tag Nucleic acid detection.US20090263813
- Germer, S. & Higuchi, R., 1999.** Single-tube genotyping without oligonucleotide probes. *Genome Res*, 9, p.72–8.
- Grotjahn L, 1986.** in *Mass Spectrometry in Biomedical Research*. Dans New York: Gaskell S.J, p. 215–234.
- Glückmann, Matthias et al., 2001.** Mechanisms in MALDI analysis: surface interaction or incorporation of analytes? *International Journal of Mass Spectrometry*, 210–211(0), p.121–132.
- Gresham, D. et al., 2006.** Genome-Wide Detection of Polymorphisms at Nucleotide Resolution with a Single DNA Microarray. *Science*, 311(5769), p.1932–1936.
- Gut, I., 1997.** Depurination of DNA and matrix-assisted laser desorption/ionization mass spectrometry. *INTERNATIONAL JOURNAL OF MASS SPECTROMETRY*, 169, p.313–322.
- Hahner, S. et al., 1997.** Matrix-assisted laser desorption/ionization mass spectrometry (MALDI) of endonuclease digests of RNA. *Nucleic Acids Res*, 25, p.1957–64.
- Gut, I.G. et al., 1997.** Analysis of DNA by ‘Charge Tagging’ and Matrix-assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Communications in Mass Spectrometry*, 11(1), p.43–50.
- Gut, I.G. & Beck, S., 1995.** A Procedure for Selective DNA Alkylation and Detection by Mass Spectrometry. *Nucleic Acids Research*, 23(8), p.1367–1373.

- Gygi, S.P. et al., 1999.** Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*, 17(10), p.994–999.
- Hansen, K.D. et al., 2011.** Increased methylation variation in epigenetic domains across cancer types. *Nature Genetics*, 43(8), p.768–775.
- He, Y.-F. et al., 2011.** Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science*, 333(6047), p.1303–1307.
- Herring, C.D. et al., 2006.** Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nature Genetics*, 38(12), p.1406–1412.
- Hettich, R.L. & Stemmler, E.A., 1996.** Investigation of oligonucleotide fragmentation with matrix-assisted laser desorption/ionization Fourier-transform mass spectrometry and sustained off-resonance irradiation. *Rapid Commun Mass Spectrom*, 10, p.321–7.
- Hunter, E.P.L. & Lias, S.G., 1998.** Evaluated Gas Phase Basicities and Proton Affinities of Molecules: An Update. *Journal of Physical and Chemical Reference Data*, 27(3), p.413–656.
- Hornblower, B. et al., 2007.** Single-molecule analysis of DNA-protein complexes using nanopores. *Nature Methods*, 4(4), p.315–317.
- Horneffer, V. et al., 2006.** Matrix–analyte-interaction in MALDI-MS: Pellet and nano-electrospray preparations. *International Journal of Mass Spectrometry*, 249–250(0), p.426–432.
- Hultman, T. et al., 1989.** Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. *Nucleic Acids Research*, 17(13), p.4937–4946.
- Ikonomou, M.G., Blades, A.T. & Kebarle, P., 1990.** Investigations of the electrospray interface for liquid chromatography/mass spectrometry. *Anal. Chem.*, 62(9), p.957–967.
- Innis, M.A. et al., 1988.** DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc Natl Acad Sci U S A*, 85, p.9436–40.
- Irizarry, R.A. et al., 2009.** The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics*, 41(2), p.178–186.
- Jones, D.S., Schofield, J.P. & Vaudin, M., 1991.** Fluorescent and radioactive solid phase dideoxy sequencing of pcr products in microtitre plates. *DNA Sequence: The Journal of DNA Sequencing and Mapping*, 1(4), p.279–283.
- Jonscher, K.R. & Yates, J.R., 3rd, 1997.** Matrix-assisted laser desorption ionization/quadrupole ion trap mass spectrometry of peptides. Application to the localization of phosphorylation sites on the P protein from Sendai virus. *The Journal*

- of Biological Chemistry*, 272(3), p.1735–1741.
- Ju, J., 1999.** Nucleic acid sequencing with solid phase capturable terminators. US5876936.
- Kaneoka, H. et al., 1991.** Solid-phase direct DNA sequencing of allele-specific polymerase chain reaction-amplified HLA-DR genes. *BioTechniques*, 10(1), p.30, 32, 34.
- Kanwar, J.R. et al., 2009.** Recent advances on the roles of NO in cancer and chronic inflammatory disorders. *Curr Med Chem*, 16, p.2373–94.
- Karas, Michael, Bachmann, D. & Hillenkamp, Franz, 1985.** Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal. Chem.*, 57(14), p.2935–2939.
- Karas, M. et al., 1987.** Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion Processes*, 78(0), p.53–68.
- Karas, M & Hillenkamp, F, 1988.** Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*, 60(20), p.2299–2301.
- Karas, M. & Bahr, U., 1990** Laser desorption ionization mass spectrometry of large biomolecules. *TrAC. Trends in analytical chemistry*, 9(10), p.321–325.
- Karas, Michael. et al., 1995.** Matrix Dependence of Metastable Fragmentation of Glycoproteins in MALDI TOF Mass Spectrometry. *Anal. Chem.*, 67(3), p.675–679.
- Karas, M, Glückmann, M & Schäfer, J., 2000.** Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *Journal of Mass Spectrometry: JMS*, 35(1), p.1–12.
- Karas, Michael et al., 2003.** The initial-ion velocity as a marker for different desorption-ionization mechanisms in MALDI. *International Journal of Mass Spectrometry*, 226(1), p.239–248.
- Karas, Michael & Krüger, R., 2003.** Ion formation in MALDI: the cluster ionization mechanism. *Chemical Reviews*, 103(2), p.427–440.
- Kasianowicz, J.J. et al., 1996.** Characterization of Individual Polynucleotide Molecules Using a Membrane Channel. *Proceedings of the National Academy of Sciences*, 93(24), p.13770–13773.
- Kebarle, P. & Tang, L., 1993.** From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry. *Analytical Chemistry*, 65(22), p.972A–986A.
- Klassen, J.S., Schnier, P.D. & Williams, E.R., 1998.** Blackbody infrared radiative dissociation of oligonucleotide anions. *Journal of the American Society for Mass Spectrometry*, 9(11), p.1117–1124.
- Knochenmuss, R. et al., 2000.** Secondary ion-molecule reactions in matrix-assisted laser

- desorption/ionization. *Journal of mass spectrometry*, 35(11), p.1237–1245.
- Knochenmuss, R., 2002.** A quantitative model of ultraviolet matrix-assisted laser desorption/ionization. *Journal of Mass Spectrometry: JMS*, 37(8), p.867–877.
- Knochenmuss, R & Zenobi, R., 2003.** MALDI ionization: the role of in-plume processes. *Chemical Reviews*, 103(2), p.441–452.
- Knochenmuss, Richard, 2003.** A quantitative model of ultraviolet matrix-assisted laser desorption/ionization including analyte ion generation. *Analytical Chemistry*, 75(10), p.2199–2207.
- Krause, J., Stoeckli, M. & Schlunegger, U.P., 1996.** Studies on the selection of new matrices for ultraviolet matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry: RCM*, 10(15), p.1927–1933.
- Krebs, S. et al., 2003.** RNaseCut: a MALDI mass spectrometry-based method for SNP discovery. *Nucleic Acids Res*, 31, p.e37.
- Kriaucionis, S. & Heintz, N., 2009.** The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science (New York, N.Y.)*, 324(5929), p.929–930.
- Krüger, R. & Karas, Michael, 2002.** Formation and fate of ion pairs during MALDI analysis: anion adduct generation as an indicative tool to determine ionization processes. *Journal of the American Society for Mass Spectrometry*, 13(10), p.1218–1226.
- Kucharzak, R., 2006.** *HLA Typing by SNP Genotyping*. Available at: http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS_derivate_000000003125/0_Title-kucharzak.pdf?hosts=.
- Lander, E.S. et al., 2001.** Initial sequencing and analysis of the human genome. *Nature*, 409, p.860–921.
- Li, L., Golding, R.E. & Whittal, R. M, 1996.** ANALYSIS OF SINGLE MAMMALIAN CELL LYSATES BY MASS SPECTROMETRY. *Journal of the American Chemical Society*, 118(46), p.11662–11663.
- Li, Y., Hunter, R.L. & McIver, R.T., 1994.** High-resolution mass spectrometer for protein chemistry. , *Published online: 04 August 1994; | doi:10.1038/370393a0*, 370(6488), p.393–395.
- Li, Y.L. & Gross, M.L., 2004.** Ionic-liquid matrices for quantitative analysis by MALDI-TOF mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 15(12), p.1833–1837.
- Lieberman, K.R. et al., 2010.** Processive Replication of Single DNA Molecules in a Nanopore Catalyzed by phi29 DNA Polymerase. *J. Am. Chem. Soc.*, 132(50),

- p.17961–17972.
- Link, E. et al., 2008.** SLCO1B1 variants and statin-induced myopathy--a genomewide study. *N Engl J Med*, 359, p.789–99.
- Little, D.P. et al., 1996.** Sequence information from 42-108-mer DNAs (complete for a 50-mer) by tandem mass spectrometry. *Journal of the American Chemical Society*, 118, p.9352–9359.
- Livadaris, V., 2002.** *Une approche thermo-cinétique de la production d'ions en mode MALDI à partir de l'étude d'associations peptidiques non-covalentes formées par désorption/ionisation.*
- Livadaris, V., Blais, J. & Tabet, J., 2000.** Formation of non-specific protein cluster ions in matrix-assisted laser desorption/ionization: abundances and dynamical aspects. *European J. Mass. Spectrom.*, 6(5), p.409–413.
- Liu, Q. & Sommer, S.S., 2000.** Pyrophosphorolysis-activated polymerization (PAP): application to allele-specific amplification. *Biotechniques*, 29, p.1072–6, 1078, 1080 passim.
- Liu, Q. & Sommer, S.S., 2002.** Pyrophosphorolysis-activatable oligonucleotides may facilitate detection of rare alleles, mutation scanning and analysis of chromatin structures. *Nucleic Acids Res*, 30, p.598–604.
- Liu, Q. & Sommer, S.S., 2004.** PAP: detection of ultra rare mutations depends on P* oligonucleotides: « sleeping beauties » awakened by the kiss of pyrophosphorolysis. *Hum Mutat*, 23, p.426–36.
- Loo, J.A., Udseth, H.R. & Smith, R.D., 1989.** Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry. *Analytical Biochemistry*, 179(2), p.404–412.
- Luo, G., Marginean, I. & Vertes, A., 2002.** Internal Energy of Ions Generated by Matrix-Assisted Laser Desorption/Ionization. *Anal. Chem.*, 74(24), p.6185–6190.
- Maiti, A. & Drohat, A.C., 2011.** Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential Implications for Active Demethylation of CpG Sites. *Journal of Biological Chemistry*.
- Makarov, 2000.** Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical Chemistry*, 72(6), p.1156–1162.
- Mamyrin, B.A., 1994.** Laser assisted reflectron time-of-flight mass spectrometry. *International Journal of Mass Spectrometry and Ion Processes*, 131(0), p.1–19.
- Mamyrin et al., 1973.** The mass-reflectron, a new non-magnetic time-of-flight mass spectrometer with high resolution. *Soviet Physics - JETP*, 37(1), p.45–4848.
- Mauger, F. et al., 2006.** SNP genotyping using alkali cleavage of RNA/DNA chimeras and

- MALDI time-of-flight mass spectrometry. *Nucleic Acids Res*, 34, p.e18.
- Mauger, F. et al., 2007.** DNA sequencing by MALDI-TOF MS using alkali cleavage of RNA/DNA chimeras. *Nucleic Acids Res*, 35, p.e62.
- Mauger, F. et al., 2012.** Ribo-PCR - A facile method for the preparation of chimeric RNA/DNA applied to DNA sequencing. *Human Mutation*.
- Maxam, A.M. & Gilbert, W., 1977.** A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), p.560–564.
- McLuckey, Scott A., Van Berker, G.J. & Glish, G.L., 1992.** Tandem mass spectrometry of small, multiply charged oligonucleotides. *Journal of the American Society for Mass Spectrometry*, 3(1), p.60–70.
- McLuckey, S. A. & Habibigoudarzi, S., 1993.** Decompositions of Multiply-Charged Oligonucleotide Anions. *Journal of the American Chemical Society*, 115, p.12085–12095.
- Metzker, M.L., 2009.** Sequencing technologies the next generation. *Nature Reviews Genetics*, 11(1), p.31–46.
- Middleton, D., 1999.** History of DNA typing for the human MHC. *Reviews in Immunogenetics*, 1(2), p.135–156.
- Moskalev, E.A. et al., 2011.** Correction of PCR-Bias in Quantitative DNA Methylation Studies by Means of Cubic Polynomial Regression.. *Nucl. Acids Res.* doi: 10.1093/nar/gkr213 A
- Nelson, C.A. & Fremont, D.H., 1999.** Structural principles of MHC class II antigen presentation. *Reviews in Immunogenetics*, 1(1), p.47–59.
- Nordhoff, E. et al., 1995.** Direct mass spectrometric sequencing of low-picomole amounts of oligodeoxynucleotides with up to 21 bases by matrix-assisted laser desorption/ionization mass spectrometry. *Journal of Mass Spectrometry*, 30(1), p.99–112.
- Nordhoff, E., 1996.** Mass spectrometry of nucleic acids. *Mass Spectrom. Rev*, 15, p.67–138.
- Ong, S.-E. et al., 2002.** Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics: MCP*, 1(5), p.376–386.
- Onnerfjord, P. et al., 1999.** Homogeneous sample preparation for automated high throughput analysis with matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry: RCM*, 13(5), p.315–322.
- Paegel, B.M., Blazej, R.G. & Mathies, R.A., 2003.** Microfluidic devices for DNA sequencing: sample preparation and electrophoretic analysis. *Current Opinion in Biotechnology*, 14(1), p.42–50.

- Pan, S. et al., 1998.** HLA-DR4 (DRB1*0401) transgenic mice expressing an altered CD4-binding site: specificity and magnitude of DR4-restricted T cell response. *Journal of Immunology (Baltimore, Md.: 1950)*, 161(6), p.2925–2929.
- Patil, N. et al., 2001.** Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science (New York, N.Y.)*, 294(5547), p.1719–1723.
- Pfeifer, G.P. & Dammann, R., 2005.** Methylation of the tumor suppressor gene RASSF1A in human tumors. *Biochemistry. Biokhimiia*, 70(5), p.576–583.
- Qin, J. et al., 2007.** The molecular anatomy of spontaneous germline mutations in human testes. *PLoS Biology*, 5(9), p.e224.
- Ragoussis, J. et al., 2006.** Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. *PLoS Genet*, 2, p.e100.
- Robertson, K.D., 2005.** DNA methylation and human disease. *Nature Reviews. Genetics*, 6(8), p.597–610.
- Rodgers, M.T. et al., 1994.** Low-energy collision-induced dissociation of deprotonated dinucleotides: determination of the energetically favored dissociation pathways and the relative acidities of the nucleic acid bases. *International Journal of Mass Spectrometry and Ion Processes*, 137(0), p.121–149.
- Ross, P.L. et al., 2004.** Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics: MCP*, 3(12), p.1154–1169.
- Russell, D.H. & Edmondson, R.D., 1997.** High-resolution Mass Spectrometry and Accurate Mass Measurements with Emphasis on the Characterization of Peptides and Proteins by Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry. *Journal of Mass Spectrometry*, 32(3), p.263–276.
- Ruzov, A. et al., 2011.** Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Research*, 21(9), p.1332–1342.
- Saiki, R.K. et al., 1985.** Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230, p.1350–4.
- Saiki, R.K. et al., 1988.** Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239, p.487–91.
- Sanger, F., Air, G.M., et al., 1977.** Nucleotide sequence of bacteriophage ϕ X174 DNA. , *Published online: 24 February 1977; | doi:10.1038/265687a0*, 265(5596), p.687–695.
- Sanger, F., Nicklen, S. & Coulson, A.R., 1977.** DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), p.5463–5467.

- Sanger, F., 1988.** Sequences, sequences, and sequences. *Annual Review of Biochemistry*, 57, p.1–28.
- Sauer, S. et al., 2000.** A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acids Res*, 28, p.E13.
- Sauer, S., Lehrach, H. & Reinhardt, R., 2003.** MALDI mass spectrometry analysis of single nucleotide polymorphisms by photocleavage and charge-tagging. *Nucleic Acids Res*, 31, p.e63.
- Schonbrunner, N.J. et al., 2006.** Chimeric thermostable DNA polymerases with reverse transcriptase and attenuated 3'-5' exonuclease activity. *Biochemistry*, 45, p.12786–95.
- Schmidt, A., Kellermann, J. & Lottspeich, F., 2005.** A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics*, 5(1), p.4–15.
- Shabarova, Z.A. & Bogdanov, A.A., 1994.** *Advanced Organic Chemistry of Nucleic Acids* 1^{er} éd., Wiley-VCH.
- Shendure, J. & Ji, H., 2008.** Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), p.1135–1145.
- Shendure, J.A. et al., 2011.** Overview of DNA sequencing strategies. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]*, Chapter 7, p.Unit7.1.
- Shima, K. et al., 2011.** Prognostic significance of CDKN2A (p16) promoter methylation and loss of expression in 902 colorectal cancers: Cohort study and literature review. *International Journal of Cancer. Journal International Du Cancer*, 128(5), p.1080–1094.
- Smith, E., Jones, M.E. & Drew, P.A., 2009.** Quantitation of DNA methylation by melt curve analysis. *BMC Cancer*, 9, p.123.
- Smith, R.D. et al., 1990.** New developments in biochemical mass spectrometry: electrospray ionization. *Anal. Chem.*, 62(9), p.882–899.
- Smith, S.S. et al., 1992.** Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation. *Proceedings of the National Academy of Sciences of the United States of America*, 89(10), p.4744–4748.
- Smith-Sørensen, B. & Hovig, E., 1996.** CDKN2A (p16INK4A) somatic and germline mutations. *Human Mutation*, 7(4), p.294–303.
- Song, F., 2003.** Quinaldic acid as a new matrix for matrix-assisted laser desorption/ionization of nucleic acids. *Rapid Commun Mass Spectrom*, 17, p.1802–7.
- Stafford Jr., G.C. et al., 1984.** Recent improvements in and analytical applications of advanced ion trap technology. *International Journal of Mass Spectrometry and Ion Processes*, 60(1), p.85–98.

- Stanssens, P. et al., 2004.** High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Res*, 14, p.126–33.
- Stemmler, E.A. et al., 1995.** Analysis of modified oligonucleotides by matrix-assisted laser desorption/ionization Fourier transform mass spectrometry. *Anal Chem*, 67, p.2924–30.
- Stephens, W., 1946.** A PULSED MASS SPECTROMETER WITH TIME DISPERSION. *Physical Review*, 69(11-1), p.691–691.
- Suckau, D. et al., 2003.** A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Analytical and Bioanalytical Chemistry*, 376(7), p.952–965.
- Svedruzić, Z.M., 2008.** Mammalian cytosine DNA methyltransferase Dnmt1: enzymatic mechanism, novel mechanism-based inhibitors, and RNA-directed DNA methylation. *Current Medicinal Chemistry*, 15(1), p.92–106.
- Szájli, E., Fehér, T. & Medzihradszky, K.F., 2008.** Investigating the quantitative nature of MALDI-TOF MS. *Molecular & Cellular Proteomics: MCP*, 7(12), p.2410–2418.
- Tabet, J. & Cotter, R., 1983.** Time-resolved laser desorption mass spectrometry. 2. measurement of the energy spread of laser desorbed. *International Journal of Mass Spectrometry and Ion Processes*, 54(1-2), p.151–158.
- Tabor, S., Huber, H.E. & Richardson, C.C., 1987.** Escherichia coli thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *The Journal of Biological Chemistry*, 262(33), p.16212–16223.
- Tahiliani, M. et al., 2009.** Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, N.Y.)*, 324(5929), p.930–935.
- Tanaka, H. & Kawai, T., 2009.** Partial sequencing of a single DNA molecule with a scanning tunnelling microscope. *Nature Nanotechnology*, 4(8), p.518–522.
- Tanaka, K. et al., 1988.** Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 2(8), p.151–153.
- Tang, K., Allman, S.L. & Chen, C.H., 1993.** Matrix-assisted laser desorption ionization of oligonucleotides with various matrices. *Rapid Commun Mass Spectrom*, 7, p.943–8.
- Tao, W.A. & Aebersold, R., 2003.** Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. *Current Opinion in Biotechnology*, 14(1), p.110–118.
- Tholey, A. & Heinze, E., 2006.** Ionic (liquid) matrices for matrix-assisted laser desorption/ionization mass spectrometry-applications and perspectives. *Analytical and Bioanalytical Chemistry*, 386(1), p.24–37.
- Thomas, W.K. et al., 2008.** Direct Sequencing by TEM of Z-Substituted DNA Molecules.

- Next Generation Genome Sequencing*. Wiley-VCH Verlag GmbH & Co. KGaA, p. 103–116. .
- Tost, J. et al., 2003.** Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. *Nucleic Acids Res*, 31, p.e50.
- Tost, J., El Abdalaoui, H. & Gut, I.G., 2006.** Serial pyrosequencing for quantitative DNA methylation analysis. *Biotechniques*, 40, p.721–2, 724, 726.
- Venter, J.C. et al., 2001.** The sequence of the human genome. *Science*, 291, p.1304–51.
- Vestal, M.L. & Juhasz, P., 1998.** Resolution and mass accuracy in matrix-assisted laser desorption ionization- time-of-flight. *Journal of the American Society for Mass Spectrometry*, 9(9), p.892–911.
- Vestal, M.L., Juhasz, P. & Martin, S.A., 1995.** Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 9(11), p.1044–1050.
- Vestal, M.L. & Campbell, J.M., 2005.** Tandem time-of-flight mass spectrometry. *Methods in Enzymology*, 402, p.79–108.
- Vorm, O., Roepstorff, P. & Mann, M., 1994.** Improved Resolution and Very High Sensitivity in MALDI TOF of Matrix Surfaces Made by Fast Evaporation. *Anal. Chem.*, 66(19), p.3281–3287.
- Viari, A. et al., 1988.** Characterization and sequencing of normal and modified oligonucleotides by ²⁵²Cf plasma desorption mass spectrometry. *Biomedical & Environmental Mass Spectrometry*, 16(1-12), p.225–228
- Warnecke, P.M. et al., 1997.** Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res*, 25, p.4422–6.
- WATSON, J.D. & CRICK, F.H., 1953.** Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356), p.737–738.
- Wenzel, T. et al., 2003.** Genosnip: SNP genotyping by MALDI-TOF MS using photocleavable oligonucleotides. *Nucleosides, Nucleotides & Nucleic Acids*, 22(5-8), p.1579–1581.
- von Wintzingerode, F. et al., 2002.** Base-specific fragmentation of amplified 16S rRNA genes analyzed by mass spectrometry: a tool for rapid bacterial identification. *Proceedings of the National Academy of Sciences of the United States of America*, 99(10), p.7039–7044.
- Wu, J. & McLuckey, S.A., 2004.** Gas-phase fragmentation of oligonucleotide ions. *International Journal of Mass Spectrometry*, 237(2-3), p.197–241.
- Xiao, M. et al., 2008.** A Single DNA Molecule Barcoding Method with Applications in DNA Mapping and Molecular Haplotyping. Dans *Next Generation Genome Sequencing*. Wiley-VCH Verlag GmbH & Co. KGaA, p. 117–132.

- Yamashita, M. & Fenn, John B., 1984.** Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.*, 88(20), p.4451–4459.
- Zenobi, Renato & Knochenmuss, Richard, 1998.** Ion formation in MALDI mass spectrometry. *Mass Spectrometry Reviews*, 17(5), p.337–366.
- Zhang, Z. et al., 2006.** 3-Hydroxycoumarin as a New Matrix for Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry of DNA. *J Am Soc Mass Spectrom.*
- Zhou, S. et al., 2008.** Optical Sequencing: Acquisition from Mapped Single-Molecule Templates. Dans *Next Generation Genome Sequencing*. Wiley-VCH Verlag GmbH & Co. KGaA, p. 133–151.
- Zhu, Y.F. et al., 1996.** The study of 2,3,4-trihydroxyacetophenone and 2,4,6-trihydroxyacetophenone as matrices for DNA detection in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom.*, 10, p.383–8.
- Zimmerman, J. et al., 1992.** Fully automated Sanger sequencing protocol for double stranded DNA. *Methods in molecular and cellular biology*, 3(1), p.39–42.