



**HAL**  
open science

## Contribution à l'amélioration de la quantification des acides nucléiques par qPCR et RT-qPCR

Pascal Pugnère

► **To cite this version:**

Pascal Pugnère. Contribution à l'amélioration de la quantification des acides nucléiques par qPCR et RT-qPCR. Médecine humaine et pathologie. Université de Grenoble, 2012. Français. NNT : 2012GRENS028 . tel-00870512

**HAL Id: tel-00870512**

**<https://theses.hal.science/tel-00870512>**

Submitted on 7 Oct 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.

## THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biotechnologie, Instrumentation, Signal et Imagerie pour la Biologie, la Médecine et l'Environnement (BIS)**

Arrêté ministériel : 7 août 2006

Présentée par

**Pascal PUGNIÈRE**

Thèse dirigée par le **MCS. , ndrè PEINNEQUIN** et  
codirigée par et le **Pr. Emmanuel DROUET**

préparée au sein du **Service de Biologie Moléculaire de L'Institut de Recherche Biomédicale des , rmées , ntenne de La Tronche**

dans l'**École Doctorale de l'Ingénierie de la Santé, la Cognition et l'Environnement (EDISCE)**

## **Contribution à l', amélioration de la Quantification des , cides Nucléiques par qPCR et RT-qPCR**

Thèse soutenue publiquement le **11 octobre 2012**,  
devant le jury composé de :

**M. Laurent BEZIN**

Maître de Conférences de l'Université Claude BERNARD, LYON I (Rapporteur)

**M. Franck CEPP,**

Professeur Agrégé du Val-de-Grâce, Hôpital d'Instruction des Armées de Bégin, PARIS (Rapporteur)

**M. François TRUËB,**

Médecin en Chef, Hôpital d'Instruction des Armées du Val-de-Grâce, PARIS (Examinateur)

**M. Joël LUN, RDI**

Professeur de l'Université Joseph FOURIER, GRENOBLE (Examinateur)

**M. , ndrè PEINNEQUIN**

Médecin Chef des Services, Institut de Recherche Biomédicale des Armées, LA TRONCHE (Directeur)

**M. Emmanuel DROUET**

Professeur de l'Université Joseph FOURIER, GRENOBLE (Co-Directeur)





*À Cathy,*

*À Nicolas & Stéphane,  
À mes parents & grands-parents,*

*À tous ceux qui ont cru en moi.*



*" La Biologie donne des réponses, la Vie ... "*

*Giovanni Quirici*

*17 Novembre 2010*

*Grimpeur, alpiniste, écrivain et philosophe, diplômé de Biologie - Université de Genève  
(Bidogno, 9 Mai 1978 - Face Nord de l'Eiger, 12 Août 2011)*



## Avant-propos

J'ai effectué mon doctorat sous la direction du Médecin Chef des Services **André PEINNEQUIN** et sous la co-direction du Professeur **Emmanuel DROUET** depuis le mois d'octobre 2008 jusqu'au mois d'octobre 2012 dans le Service de Biologie Moléculaire de l'Institut de Recherche Biomédicale des Armées – antenne de La Tronche (Université Joseph FOURIER - Grenoble).

Mon intérêt pour le domaine de la biologie moléculaire et l'ensemble des techniques que cette discipline est aujourd'hui capable de proposer au monde scientifique date de la fin des années 90. Technicien de laboratoire de formation, j'ai ainsi pu accumuler au fil des ans connaissances et maîtrises de techniques sans jamais relâcher l'intérêt et la curiosité que je leur portais. Ainsi, la voie de la recherche s'est imposée dès le début des années 2000 avec la réalisation d'un Diplôme d'Etude Appliquée en Physiologie Intégrée. Je rencontrais au même moment et pour la toute première fois André et son sens inné des mathématiques... S'en est suivi un long et hasardeux parcours initiatique à travers le monde de l'entreprise et de la recherche en secteur public, fait de rencontres extraordinaires, de succès, mais aussi d'échecs... Il fut à mon sens le plus enrichissant qui soit.

Dès le mois d'octobre 2006, je retrouvais André pour travailler à nouveau avec lui sur ce qui fut le Pôle de Génomique, en tant qu'ingénieur de recherche. Après un important processus d'apprentissage et de maturation pendant deux longues années faites de réflexions et d'échanges, nous finissons par concrétiser cette curiosité scientifique de la technique d'amplification de l'ADN en temps réel sous la forme d'un doctorat. Réalisé dans le cadre de la formation continue, il fut à l'image de toutes les thèses, tantôt décourageant, tantôt exaltant, mais restera à jamais pour moi une véritable expérience de découverte et d'apprentissage de la connaissance.

Ce manuscrit ne traite pas de la technique de *Reverse Transcription – quantitative Polymerase Chain Reaction* (RT-qPCR) de manière exhaustive mais définit plus précisément ses limites, notamment dans certains cadres d'application. Il souligne les différents facteurs capables de l'influencer ainsi que les améliorations pouvant être apportées au cours des différents stades de sa réalisation.

Dans un premier temps, une introduction générale sera consacrée aux différentes caractéristiques de la technique de RT-qPCR ainsi qu'à son utilisation dans la quantification des acides nucléiques. Je développerai un premier chapitre traitant de l'inhibition et la variabilité inter-échantillons, phénomènes souvent rencontrés au cours des analyses de RT-qPCR. Une étude d'évaluation de la spécificité d'amorces modifiées par acides nucléiques bloqués (*LNA*) sera ensuite présentée. Enfin, dans une troisième partie, des considérations relatives à la détermination de la température de fusion et des paramètres thermodynamiques d'un oligonucléotide seront abordées. Je terminerai ce manuscrit par les perspectives pouvant être attendues à la suite de ces différents résultats.



## Remerciements

Tout d'abord, je tiens à remercier mon Directeur de Thèse, le Médecin Chef des Services **André PEINNEQUIN**, qui a su patiemment me guider à travers ces différents Travaux de Recherche. Recevoir la connaissance de la RT-qPCR de la part d'un *Géo-Trouve-Tout* sans pareil aura été pour moi un véritable privilège et la plus enrichissante des expériences scientifiques. Je tiens aussi à le remercier infiniment pour m'avoir donné l'occasion de travailler avec lui pendant six longues années au sein du Pôle de Génomique. Ce fut un réel plaisir de travailler avec lui, dans l'environnement le plus stimulant qu'il m'est été donné de rencontrer.

Je tiens à remercier très sincèrement le Professeur **Emmanuel DROUET** d'avoir accepté d'être mon co-Directeur de Thèse et pour m'avoir ainsi permis de me lancer dans cette incroyable quête de la Connaissance. Je remercie ainsi le *Scientifique*... Quant au *Montagnard*, je lui devrais toujours l'inestimable Trésor des fabuleuses lectures qu'il m'a fait et continue de me faire découvrir, depuis *Leslie Stephen* jusqu'à *Edward Whymper*, *René Desmaison* ou encore *Walter Bonatti*, ... le plus grand, nous sommes d'accord ! En espérant encore vous recroiser, sans doute sur les pentes du *Massif de Belledonne* ou au détour d'un sentier du *Parc National des Écrins* ! ...

Je tiens également à remercier tout particulièrement **Catherine MOURET**, la plus agréable des collègues de travail avec qui il m'est été donné de travailler. Sa réserve, sa discrétion, sa gentillesse et sa disponibilité m'ont chaque jour profondément touché. J'ai eu l'immense privilège de recevoir de sa part le meilleur des enseignements techniques de la PCR dès notre première rencontre en 2001. Pour cet héritage, je lui serai toujours infiniment reconnaissant. Elle n'a jamais cessé de m'étonner par cette exceptionnelle maîtrise technique mais je crois bien qu'après plus de 7 longues années passées à l'observer, j'en suis enfin arrivé à la certitude qu'elle est sans doute aussi un peu magicienne. En fait, je reste convaincu qu'elle sait vraiment tout faire avec une paire de pipettes dans les mains !

Je remercie sincèrement Monsieur **Joël LUNARDI** pour m'avoir fait l'honneur d'accepter de se joindre à ce Jury de Thèse et lui suis reconnaissant de l'intérêt qu'il a su porter à ces différents Travaux de Recherche. Je souhaite aussi remercier sincèrement Monsieur **Laurent BEZIN** et Monsieur **Franck CEPPA** pour avoir accepté d'être les rapporteurs de ce Travail de Thèse, le Médecin en Chef **François TRUÉBA** pour avoir accepté de faire partie de la composition de ce Jury.

Un remerciement tout particulier et très personnel pour **Cathy**, pour tout l'amour qu'elle me témoigne chaque jour, pour son indéfectible soutien mais aussi et surtout pour avoir su me supporter patiemment tout au long de ces derniers mois. J'espère désormais pouvoir lui témoigner toute l'attention qu'elle mérite ...

Merci à toi **Kév**, taciturne frère d'Auvergne. Tu vis comme *l'Homme libre* que chacun rêve de devenir ... Où que l'on soit plus tard, je garderai précieusement cette belle et indéfectible Amitié. Et tu as raison, « *La Verte* » est bien la plus belle Montagne du Monde, je suis d'accord avec toi ! Cependant, n'oublie jamais une seule chose : « *La montagne n'est ni juste, ni injuste. Elle est dangereuse.* » Merci à toi **Sibylle**, pour ton amitié, ta sincérité, ta confiance mais surtout pour le soutien que tu as régulièrement su me témoigner.

Merci à toi **Dan**, frère des Cévennes, charismatique Capitaine de la *Gamba Me Fai Mal*, intrépide explorateur de mondes engloutis, philanthrope patenté à la gentillesse sans limite ! Grâce à toi, le *Mont Aigoual* fait désormais parti de mes prochains « sommets » à gravir ... « Sommet » bien trop méconnu des montagnards il est vrai... Du haut de ses vertigineux 1565 mètres d'altitude, j'ignorais jusque là l'existence même de ce colosse de granite et de roches métamorphiques, sans doute resté dans l'ombre du *Mont Mézenc* (1753 m d'altitude).

Merci à toi **Laurette**, pour ton extraordinaire gentillesse, sans doute la chose la plus rare que l'on puisse rencontrer de nos jours. Je n'oublierai jamais ces délicieux et ô combien exceptionnels petits gâteaux crêpe gavotte-chocolat au lait ... Merci à toi **David**, pour toutes ces nombreuses marques de sympathie et d'estime régulièrement témoignées ... Merci à vous deux pour votre précieuse Amitié.

Merci à toi **Thom**. Infatigable dans le sport tout comme dans la réflexion intellectuelle. Tu ignores toujours toute la motivation que tu as su me procurer durant ces dernières années. De toute évidence, *Force* et *Détermination* sont l'apanage du Breton ... Et te rencontrer sur le chemin de la Science m'aura agréablement obligé à répondre à « *quelques petites questions* », comme tu dis si bien ...

Merci à toi **Diane**, pour ton humanité si touchante ainsi que pour tout le soutien que tu as su me témoigner, souvent dans les moments les plus difficiles. Je crois que je ne te remercierai jamais assez pour toute l'aide que tu as su m'apporter à chaque fois que j'en avais besoin...

Merci à toi **Hervé**, pour ton amitié et ton soutien. Avec **Olivier** et **Xavier**, vous avez su me donner l'envie de travailler dans la Recherche. Pour cela je t'en remercie, tout comme pour tous tes innumérables conseils qui se sont toujours avérés très précieux dans chacun des choix que j'ai pu avoir à faire.

Merci à toi **Johannes**, pour ces magnifiques sorties alpi' que j'ai pu partager avec toi. Tu m'as appris la rigueur nécessaire à la préparation d'une course. Continue toujours d'avancer droit devant toi, tu marches dans les traces de Kurt Albert... « *Parce qu'encore plus beau que la montagne, c'est la montagne ensemble !* »

Je tiens à remercier absolument toutes les personnes rencontrées dans les différents Départements de Recherche du **CRSSA**. En six longues années, la liste est tellement importante que je ne souhaiterai oublier aucune personne : étudiant(e)s, technicien(ne)s, ouvrier(e)s d'état, ingénieurs d'étude, vétérinaires, médecins, pharmaciens ...

Enfin, je souhaiterai personnellement remercier toute ma famille et plus particulièrement ma mère et mon père. Je les remercie sincèrement pour m'avoir toujours donné le meilleur, pour moi-même comme pour mes frères, ainsi que pour tout l'amour qu'ils me témoignent chaque jour si généreusement. Je leur suis reconnaissant de m'avoir sans cesse encouragé dans les moments difficiles tout au long de ces années de Recherche. Je tiens également à remercier Nico & Steph, les meilleurs frères qui soient...

# Liste des Abréviations

18S	ARN ribosomal 18 S
28S	ARN ribosomal 28 S

## A

---

A	Adénine
ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire (à l'ARN)
ARN	Acide ribonucléique
ARNm	ARN messenger
ARNr	ARN ribosomique
ARNt	ARN de transfert
atm	atmosphère

## B

---

BLAST	<i>Basic Local Alignment Search Tool</i> (Outil de recherche d'alignements)
bp	<i>Base Pair</i>
BrEt	Bromure d'éthidium
BSA	<i>Bovine Serum Albumin</i>

## C

---

°C	Degré Celsius
C	Cytosine
CA-RT	<i>Constant Amount - Reverse Transcription</i> , Transcription inverse à quantité constante d'ARN
CD 19	Cluster de Différenciation 19
cDNA	<i>complementary DNA</i>
C <sub>0</sub> t	<i>Concentration over time</i>
C <sub>p</sub>	<i>Crossing-point</i>
C <sub>q</sub>	Cycle de quantification
C <sub>t</sub>	<i>Threshold cycle</i>
CV	Coefficient de Variation
CV-RT	<i>Constant Volume - Reverse Transcription</i> , Transcription inverse à volume constant d'extrait d'ARN

## D

---

<i>DNA</i>	<i>deoxyribonucleic acid, acide désoxyribonucléique</i>
<i>DNase</i>	<i>désoxyribonucléase</i>
<i>DMSO</i>	<i>Diméthylsulfoxyde</i>
<i>dNTP</i>	<i>désoxyribonucléotides triphosphate</i>
<i>ddNTP</i>	<i>didésoxyribonucléotides triphosphate</i>
<i>dsDNA</i>	<i>double stranded DNA</i>

## E

---

<i>EDTA</i>	<i>Ethylènediamine tétraacétique</i>
<i>EGFP</i>	<i>Enhanced Green Fluorescent Protein</i>
<i>et al.</i>	<i>et alii - et les autres</i>
<i>EtBr</i>	<i>Ethidium Bromide</i>
<i>et coll.</i>	<i>et collaborateurs</i>
<i>e.g.</i>	<i>exempli gratia, par exemple</i>

## F

---

<i>FAM</i>	<i>6-carboxyfluorescéine, fluorophore</i>
<i>FFPE</i>	<i>Formalin-Fixed, Paraffin-Embedded, fixé au formol, inclus en paraffine</i>
<i>FRET</i>	<i>Fluorescence Resonance Energy Transfer, Transfert par résonance d'énergie de la fluorescence</i>

## G

---

<i>G</i>	<i>Guanine</i>
<i>GAPDH</i>	<i>Glyceraldehyd-3-Phosphate Deshydrogenase</i>
<i>GFP</i>	<i>Green Fluorescent Protein</i>
<i>GITC</i>	<i>Thiocyanate de guanidinium</i>
<i>GSP</i>	<i>gene-specific primer</i>

## H

---

<i>HRM</i>	<i>High-Resolution DNA Melt curve analysis</i>
<i>HPRT</i>	<i>Hypoxanthine Phosphoribosyltransferase</i>

## I

---

<i>IDT</i>	<i>Integrated DNA Technologies</i>
<i>IL-6</i>	Interleukine 6
<i>IRBA</i>	Institut de Recherche Biomédicale des Armées
<i>ITC</i>	<i>Isothermal Calorimetry</i> , calorimétrie isothermale
<i>IUPAC</i>	<i>International Union of Pure and Applied Chemistry</i>
<i>i.e.</i>	<i>id est</i>

## K

---

<i>Kb</i>	kilo paire de base
-----------	--------------------

## L

---

<i>LOD</i>	Limite de Detection, <i>Limit of Detection</i>
<i>LOL</i>	Limite de Linéarité, <i>Limit of Linearity</i>
<i>LOQ</i>	Limite de Quantification, <i>Limit of Quantification</i>

## M

---

<i>min</i>	minute
<i>MIBBI</i>	<i>Minimum Information for Biological and Biomedical Investigations</i>
<i>MIQE</i>	<i>Minimum Information for Publication of Quantitative Real-Time PCR Experiments</i>
<i>MIT</i>	<i>Massachusetts Institute of Technology</i>
<i>mRNA</i>	messenger RNA

## N

---

<i>n</i>	Nombre de cycles de PCR
<i>N<sub>0</sub></i>	Nombre de molécules cibles au départ de la réaction
<i>N<sub>n</sub></i>	Nombre de molécules cibles au bout du cycle <i>n</i>
<i>Nbre</i>	Nombre
<i>NF</i>	<i>Normalization Factor</i> , facteur de normalisation
<i>nt</i>	Nucléotide
<i>NTC</i>	<i>No Template Control</i> , contrôle négatif

## P

---

<i>p.</i>	Page
<i>pb</i>	paire de bases

PCR	<i>Polymerase Chain Reaction / Réaction de polymérisation en chaîne</i>
PNA	<i>Peptide Nucleic Acids, acides nucléiques peptidiques</i>
PPIA	<i>Peptidylprolyl Isomerase A (cyclophilin A)</i>

## Q

---

qPCR	quantitative PCR / PCR en temps réel
------	--------------------------------------

## R

---

RDML	<i>Real-time PCR Data Markup Language</i>
RT-qPCR	<i>Reverse-Transcription - quantitative Polymerase Chain Reaction</i>
RIN	<i>RNA Integrity Number</i>
RMN	Résonance Magnétique Nucléaire
RNA	<i>ribonucleic acid</i>
RNase	ribonucléase
RPA	<i>Ribonuclease Protection Assay</i>
ROR (vaccin)	vaccin Rougeole Oreillons Rubéole, <i>Measles Mumps Rubella (vaccine)</i>
rRNA	<i>ribosomal RNA</i>
RT	<i>Reverse Transcription, transcription inverse</i>

## S

---

sec	seconde
SD	<i>Standard Deviation, écart-type</i>
sdm	<i>Second Derivative Maximum</i>
SDS	Sodium Dodecyl Sulfate
SE	Erreur Standard, <i>Standard Error</i>
SNP	<i>Single Nucleotide Polymorphism, Polymorphisme nucléotidique</i>
SPR	<i>Surface Plasmon Resonance, Résonance Plasmonique de Surface</i>

## T

---

T	Thymine
T(°C)	Température
TAMRA	tetramethylrhodamine, <i>quencher</i>
tHDA	<i>thermophilic Helicase-Dependent Amplification</i>
$T_H$	Température d'hybridation
$T_m$	<i>Melting Temperature, Température de fusion</i>
Tris	Tris hydroxyméthylaminométhane

## U

---

$U$	Uracile
$UV$	Ultraviolet

## $\Delta$

---

$\Delta C_p$	Changement de capacité thermique
$\Delta G$	Variation d'énergie libre de Gibbs, mesure de la stabilité globale du système
$\Delta H$	Variation d'enthalpie (énergie potentielle du système)
$\Delta S$	Variation d'entropie (désordre moléculaire)

## $\varepsilon$

---

$\varepsilon$	Efficacité de PCR
---------------	-------------------

## $\lambda$

---

$\lambda$	Longueur d'onde
-----------	-----------------

# Table des Matières

Avant-propos  
Remerciements  
Liste des Abréviations

---

## INTRODUCTION PAGE 1

---

<b>1.1. Amplification de l'ADN par PCR.....</b>	<b>page 8</b>
1.1.1. Définitions .....	page 8
1.1.2. Les étapes de la PCR .....	page 9
1.1.3. Caractéristiques de la PCR .....	page 14
1.1.4. Limites de la PCR .....	page 15
1.1.5. Thermodynamique et hybridation moléculaire .....	page 17
1.1.5.1. Paramètres thermodynamiques .....	page 18
1.1.5.2. Constante d'équilibre .....	page 19
1.1.5.3. Estimations de $\Delta S$ et $\Delta H$ pour un oligonucléotide de séquence connue .....	page 20
1.1.5.4. Applications à la PCR .....	page 23
1.1.6. Cinétiques d'amplification .....	page 25
<b>1.2. Quantification de l'ADN par qPCR selon le modèle exponentiel .....</b>	<b>page 31</b>
1.2.1. Notion de Threshold Cycle $C_T$ , Crossing Point $C_p$ et de quantification Cycle $C_q$ .....	page 31
1.2.2. Quantification absolue et courbe de quantification .....	page 33
1.2.2.1. Quantification absolue .....	page 33
1.2.2.1. Courbe de quantification .....	page 34
<b>1.3. Les Directives MIQE .....</b>	<b>page 37</b>

---

## OBJECTIFS SCIENTIFIQUES PAGE 41

---

## MATÉRIELS & MÉTHODES PAGE 45

---

<b>3.1. Échantillons biologiques.....</b>	<b>page 45</b>
<b>3.2. Purification des ARN totaux .....</b>	<b>page 45</b>
3.2.1. Extraction des ARN totaux.....	page 45
3.2.2. Contrôle de la qualité de l'ARN.....	page 46
<b>3.3. Transcription inverse des ARN messagers.....</b>	<b>page 48</b>
3.3.1. Transcription inverse classique.....	page 48
3.3.2. Transcription inverse à volume constant d'extrait d'ARN.....	page 49
3.3.3. Évaluation de l'efficacité relative de RT par l'analyse 5':3'.....	page 50

<b>3.4. Amplification de l'ADN par Polymérisation en Chaîne .....</b>	<b>page 52</b>
3.4.1. Design des amorces de qPCR .....	page 52
3.4.2. Optimisation des amorces .....	page 54
3.4.3. Amplification par qPCR - système SybrGreen I .....	page 56
3.4.4. Intercalants ADN .....	page 57
3.4.5. Mesure du $T_m$ par courbe de fusion sur LightCycler™ .....	page 60
3.4.6. Amplification par qPCR - système Sonde d'hydrolyse .....	page 63
3.4.7. Système de sonde d'hydrolyse (Taqman®) .....	page 64
3.4.8. Modèle d'évaluation de la spécificité des amorces modifiées .....	page 65
3.4.9. Test d'inhibition de la PCR .....	page 66
<b>3.5. Méthodes prédictives de la température de fusion.....</b>	<b>page 68</b>
<b>3.6. Calorimétrie Différentielle à balayage .....</b>	<b>page 72</b>
<b>3.7. Quantification relative d'ARN messagers par la méthode du <math>\Delta\Delta</math> de <math>C_q</math> .....</b>	<b>page 75</b>
<b>3.8. Utilisation du logiciel geNorm .....</b>	<b>page 81</b>
<b>3.9. Analyses statistiques .....</b>	<b>page 82</b>
3.9.1. Méthode de Bland-Altman .....	page 82
3.9.2. Les raisons de son utilisation .....	page 82
<b>3.10. Prélèvements animaux .....</b>	<b>page 84</b>
<b>VARIABILITE INTER-ÉCHANTILLONS &amp; INHIBITION EN RT-qPCR</b>	<b>PAGE 85</b>
<hr/>	
<b>4.1. Introduction .....</b>	<b>page 85</b>
<b>4.2. Résultats expérimentaux .....</b>	<b>page 89</b>
4.2.1. Évaluation de la qualité de l'ARN .....	page 89
4.2.2. Évaluation de la qualité de la RT .....	page 90
4.2.3. Évaluation de l'inhibition de la réaction de qPCR .....	page 96
4.2.4. Conséquences des inhibiteurs volume-dépendants sur la quantification .....	page 99
4.2.5. Agréments entre les différentes techniques d'extraction d'ARN utilisées .....	page 104
<b>4.3. Discussion .....</b>	<b>page 106</b>
<b>4.4. Conclusion .....</b>	<b>page 110</b>

<b>5.1. Introduction</b> .....	<b>page 112</b>
<b>5.2. Résultats expérimentaux</b> .....	<b>page 124</b>
5.2.1. Efficacité PCR & température d'hybridation .....	page 125
5.2.2. Prédiction de la température de fusion des amorces modifiées .....	page 129
5.2.2.1. Comparaisons de trois méthodes de prédiction du $T_m$ .....	page 129
5.2.2.2. Comparaisons des méthodes de prédiction du $T_m$ avec la mesure directe par courbe de fusion.....	page 131
5.2.3. Conception d'un modèle d'évaluation de la spécificité d'amorces modifiées .....	page 133
5.2.3.1. Essais préliminaires .....	page 133
5.2.3.2. Modèle d'évaluation de la spécificité des amorces .....	page 134
5.2.4. Spécificité des amorces modifiées .....	page 136
5.2.4.1. Estimation de l'amplification non spécifique .....	page 136
5.2.4.2. Spécificité des amorces tous couples confondus .....	page 137
5.2.4.3. Spécificité des amorces modifiées par 2 monomères LNA en condition d'hybridation optimale .....	page 139
5.2.5. Sensibilité des amorces modifiées .....	page 140
5.2.6. Mesure du pourcentage d'amorces hybridées en fonction de la température .....	page 142
5.2.7. Définitions de règles de design .....	page 144
5.2.7.1. Nature de la base substituée .....	page 144
5.2.7.2. Positionnement de la base substituée .....	page 144
<b>5.3. Discussion</b> .....	<b>page 145</b>

## DÉTERMINATION DE LA TEMPÉRATURE DE FUSION

### & DES PARAMÈTRES THERMODYNAMIQUES D'UN OLIGONUCLÉOTIDE

<b>6.1. Introduction</b> .....	<b>page 150</b>
6.1.1. Rappel sur la complémentarité des bases de l'ADN .....	page 150
6.1.2. Comportements de l'ADN en fonction de la température .....	page 151
6.1.2.1. Dénaturation de l'ADN .....	page 151
6.1.2.1. Effet hyperchromique de l'ADN .....	page 153
6.1.2.2. Renaturation de l'ADN .....	page 154
6.1.3. Méthodes prédictives de la température de fusion .....	page 157
6.1.3.1. Méthode de calcul du $T_m$ basée sur la séquence de l'oligonucléotide .....	page 158
6.1.3.2. Méthodes de calcul du $T_m$ tenant compte de la concentration en sodium .....	page 158
6.1.3.3. Méthodes de prédiction du $T_m$ selon le modèle du plus proche voisin .....	page 159
6.1.3.4. Variations autour de l'équation thermodynamique du $T_m$ .....	page 161
6.1.4. Problématique de l'étude .....	page 163
<b>6.2. Résultats expérimentaux – Méthode de mesure directe du <math>T_m</math></b> .....	<b>page 165</b>
6.2.1. Caractérisation de la méthode de mesure directe du $T_m$ .....	page 165
6.2.1.1. Effet de la concentration d'amorce .....	page 165
6.2.1.2. Effet du magnésium .....	page 165
6.2.1.3. Effet de la vitesse de la transition thermique .....	page 167

6.2.1.4. Effet des intercalants fluorescents .....	page 169
6.2.1.5. Reproductibilité inter-instrument .....	page 171
6.2.2. Validation de la méthode de mesure directe du $T_m$ par courbe de fusion.....	page 172
6.2.2.1. Comparaison des valeurs de $T_m$ obtenues par les deux techniques avec leurs tampons respectifs.....	page 173
6.2.2.2. Comparaison des valeurs de $T_m$ obtenues en tampon de PCR avec les deux techniques .....	page 175
6.2.2.3. Comparaison des valeurs de $T_m$ obtenues en tampon phosphate 1M Na <sup>+</sup> avec les deux techniques .....	page 177
6.2.3. Détermination du $T_m$ par les méthodes prédictives en tampon phosphate 1M Na <sup>+</sup> .....	page 179
6.2.3.1. Comparaison de différents modèles prédictifs .....	page 179
6.2.3.2. Modèle prédictif dnaMATE .....	page 181
6.2.4. Effet séquence-dépendant des cations sur le $T_m$ des amorces.....	page 183
6.2.5. Application de la méthode aux amorces modifiées par LNA.....	page 185
<b>6.3. Discussion – Méthode de mesure directe du <math>T_m</math>.....</b>	<b>page 188</b>
<b>6.4. Résultats expérimentaux – Paramètres thermodynamiques de l'hybridation .....</b>	<b>page 190</b>
6.4.1. Précision de la mesure des paramètres thermodynamiques.....	page 191
6.4.2. Reproductibilité de la mesure des paramètres thermodynamiques.....	page 192
<b>6.4. Discussion – Paramètres thermodynamiques de l'hybridation .....</b>	<b>page 196</b>

## PERSPECTIVES

PAGE 198

Glossaire

Sites Internet Utilisés

Publications

Références bibliographiques



Chapitre I

# Introduction

---

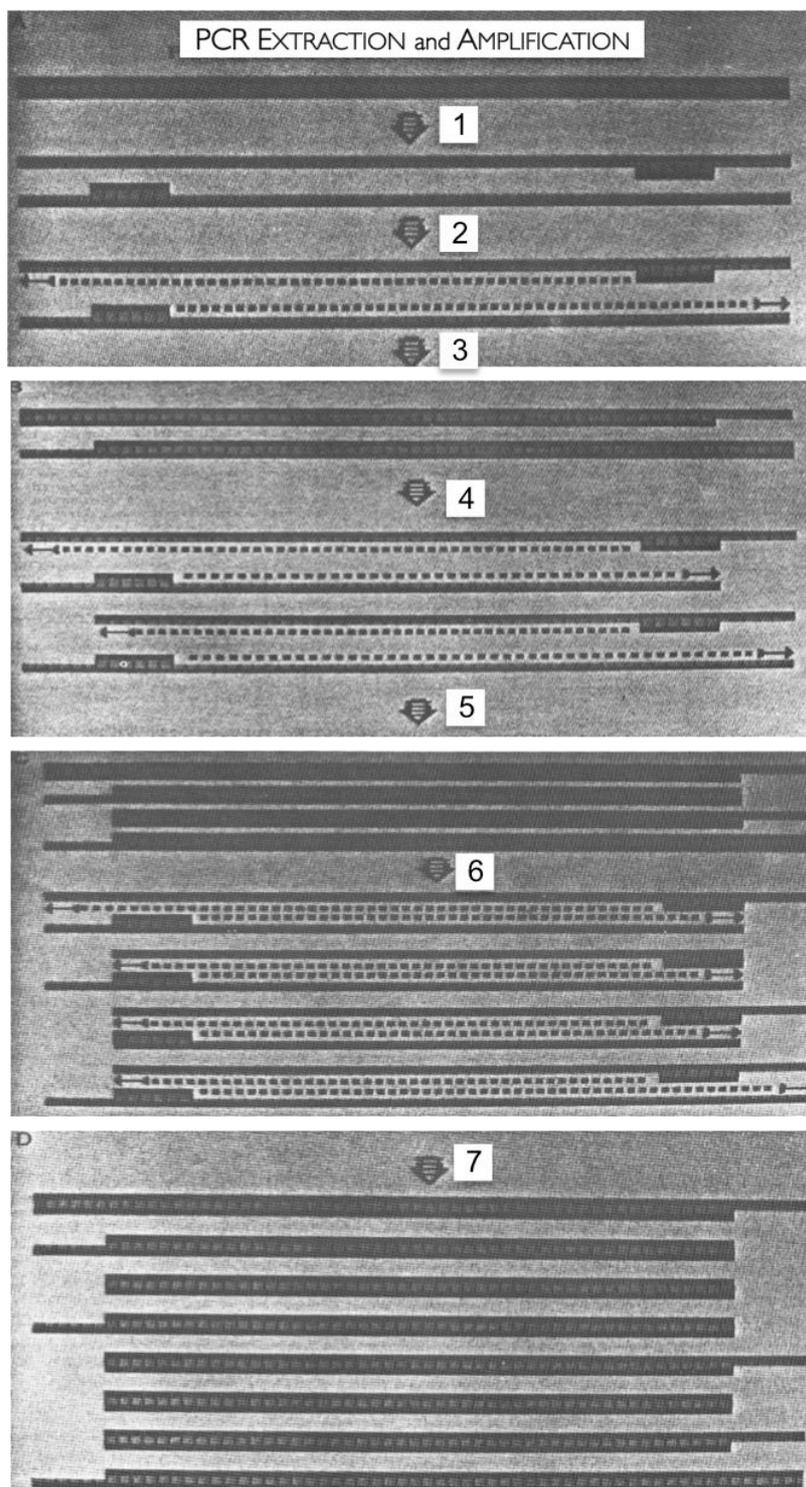
Depuis sa première découverte par Friedrich MIESCHER en 1869 (Miescher 1871 ; Dahm 2008), l'acide désoxyribonucléique (ADN), alors connu sous le nom de *nucléine*, fut très vite reconnu comme le véritable support de l'information génétique (Avery, Macleod et al. 1943). À la suite de sa caractérisation structurelle par James Dewey WATSON et Francis CRICK (Watson and Crick 1953) en 1953, la recherche s'est alors rapidement orientée vers les différentes méthodes capables de le quantifier. Les premières études de quantification par **hybridation** sur membrane de nitrocellulose remontent aux années 60 avec les travaux de David GILLESPIE (1965) (Gillespie and Spiegelman 1965). Les acides nucléiques présentant une bande d'absorption maximale autour de 260 nm, la spectrophotométrie dans l'ultraviolet a tout d'abord été très vite utilisée pour leur quantification (Petrakis 1953). Elle constitue encore à l'heure actuelle la méthode la plus couramment utilisée (Tataurov, You et al. 2008).

Dans les années 1970 apparaissent successivement les premières techniques de détection spécifique de l'ADN avec le *Southern blot* (Southern 1975) puis spécifique de l'acide ribonucléique (ARN) avec le *Northern blot* (Alwine, Kemp et al. 1977 ; Thomas 1980). Toujours basées sur l'**hybridation** d'acides nucléiques à partir de membranes, ces deux techniques similaires sont respectivement capables de détecter l'ADN et l'ARN de manière spécifique afin de les quantifier à l'aide de radio-isotopes radioactifs de courte demi-vie comme le phosphore 32 ( $^{32}\text{P}$ ) par exemple. À la même époque, BERK *et al.* travaillent sur la cartographie d'un **ARN messenger (ARNm)** (Berk and Sharp 1977) au moyen d'une endonucléase issue d'*Aspergillus oryzae* capable de cliver les acides nucléiques lorsqu'ils sont sous forme simple brin (Ando 1966). Cette approche intéressante se développe deux ans plus tard (Weaver and Weissmann 1979) et en 1983, ZINN *et al.* rapportent pour la première fois dans la littérature scientifique la technique de *Ribonuclease Protection Assay* (RPA) (Zinn, DiMaio et al. 1983). Décrite plus en détail par Douglas MELTON en 1984 (Melton, Krieg et al. 1984), la RPA est capable d'identifier une ou plusieurs molécules d'ARN de séquence connue au moyen d'un ARN **antisens** ou d'une **sonde** ADN, **complémentaires** de la séquence d'intérêt. Après exposition à l'endonucléase SI, les séquences d'intérêt se trouvant sous forme double brin, donc protégées de l'activité enzymatique, sont ensuite détectées puis quantifiées. Un intérêt éphémère sera porté à cette méthode puisqu'apparaîtra deux années plus tard la technique la plus sensible et la plus fiable dans la quantification des acides nucléiques (Mullis, Faloona et al. 1986)...

La genèse de la réaction de l'amplification en chaîne de l'ADN, plus communément appelée **PCR** (pour *Polymerase Chain Reaction*), prend très probablement naissance en 1971 lorsque Kjell KLEPPE publie une étude sur les mécanismes de réparation de l'ADN ayant lieu

pendant la réplication cellulaire (Kleppe, Ohtsuka et al. 1971). Les études basées sur l'utilisation d'une **ADN polymérase** - ADN dépendante dans le but de polymériser un ADN synthétique à partir de courtes séquences déjà appelées "*primers*" sont sans doute à l'origine du concept de la **PCR**. Durant le début des années 80, alors que l'**ARN polymérase** est le principal sujet d'études, l'idée de la **PCR** fait tout doucement son chemin. En 1985, c'est avec la publication des travaux de Randall SAIKI (Saiki, Scharf et al. 1985) que l'on entend parler pour la première fois d'amplification enzymatique de séquence cible à partir d'**amorces**. Ce groupe de recherche de *Cetus Corporation* travaille alors sur l'établissement d'un diagnostic prénatal à la fois sensible et rapide de l'anémie falciforme, en détectant les séquences génomiques de la  $\beta$ -globine à partir de **sondes** oligonucléotidiques ADN.

L'idée de la **Réaction de Polymérisation en Chaîne**, encore appelée *Polymérisation par extension d'amorces*, naît dans l'esprit d'un chimiste américain de 39 ans, Kary MULLIS, un soir d'avril 1983. Désireux d'améliorer le séquençage de fragments d'ADN déjà rendu possible par Frederick SANGER (Sanger and Coulson 1975 ; Sanger, Nicklen et al. 1977), c'est au clair de lune et au volant de sa voiture le long de l'*U.S. Highway 101* que le chercheur californien imagine cette remarquable réaction. *« J'ai de nouveau arrêté la voiture et ai commencé à dessiner les lignes des molécules d'ADN s'hybridant et s'allongeant, les produits du premier cycle devenant des matrices pour le suivant dans une sorte de réaction en chaîne ... Sur le point de s'endormir, Jennifer [BARNETT, alors la petite amie de MULLIS] protesta de nouveau. Tu ne vas pas le croire, m'écriai-je, complètement excité. C'est incroyable ! Elle refusa de se réveiller Immédiatement, je repris place au volant de ma voiture. Le fin fond de l'Anderson Valley est l'endroit où les séquoias apparaissent et où les "bons-à-rien" ont toujours vécu. Avec cette découverte, je me sentais sur le point de rompre avec la tradition de cette ancienne vallée. Il était vraiment difficile pour moi de dormir cette nuit-là, avec toutes ces bombes désoxyribo-nucléaires explosant dans mon cerveau »* (Mullis 1990). Ses travaux sont publiés trois ans plus tard (Mullis, Faloona et al. 1986), faisant de lui le Père de la **PCR** (**Figure 1.1**). La polymérisation de l'ADN est dans un premier temps réalisée avec la plus importante sous-unité de l'**ADN polymérase** I d'*Escherichia coli* : le fragment de *Klenow*. Il faudra attendre trois ans avant de voir apparaître l'utilisation d'une **ADN polymérase** - ADN dépendante thermorésistante issue d'une *Deinococci* thermophile vivant dans les sources d'eaux chaudes, *Thermus aquaticus* (Brock and Freeze 1969). Randall SAIKI décrit ainsi cette remarquable amélioration permettant une cyclisation continue de la réaction dans son étude publiée en 1988 (Saiki, Gelfand et al. 1988).



**Figure I.1 :** Première représentation des trois premiers cycles de la Réaction de Polymérisation en Chaîne de l'ADN, publiée en 1986 (Mullis, Faloona et al. 1986). L'hybridation de deux amorces au niveau des deux brins d'une séquence matricielle connue va ici conduire à une amplification de la cible d'un facteur 8. Les détails du premier cycle sont représentés au niveau des réactions 1, 2 et 3. Les détails du deuxième cycle sont représentés au niveau des réactions 4 et 5, et ceux du troisième cycle, au niveau des réactions 6 et 7.

Très rapidement, la PCR devient la technique de laboratoire la plus spécifique et la plus sensible pour la détection de l'ADN (Williams 2009). En 1993, on attribue à Kary MULLIS le Prix Nobel de Chimie pour le développement de cette technique révolutionnaire trouvant son utilité et son application dans un très grand nombre de domaines de la biologie.

Par la suite, des avancées technologiques majeures ne cesseront de se succéder (Erllich, Gelfand et al. 1991). En 1992, à la suite d'une erreur survenue au cours d'une expérimentation, Russell HIGUCHI propose de détecter l'accumulation du produit de PCR, appelé amplicon, à l'aide d'un intercalant fluorescent, le bromure d'éthidium (EtBr), sans avoir à ouvrir le tube où se déroule la réaction biochimique (Higuchi, Dollinger et al. 1992 ; Wittwer, Herrmann et al. 1997). La méthode présente alors l'intérêt de combiner à la fois l'amplification et la détection de la séquence cible, cela en une seule étape. La possibilité d'enregistrer une cinétique de fluorescence au fur et à mesure de la réaction de synthèse de l'ADN (Higuchi, Fockler et al. 1993) marque un tournant décisif dans l'évolution de la PCR que l'on nommera désormais PCR en Temps Réel ou *Real Time PCR*. En 1993, *Molecular Probes* dépose le brevet d'un intercalant capable de se lier de façon non spécifique à l'ADN : le *SybrGreen I* (mentionné pour la première fois dans le *Bioprobes n° 18*). En effet, cette cyanine asymétrique est capable d'émettre un signal fluorescent une fois liée au petit sillon d'un ADN double brin. L'intensité de la fluorescence émise peut alors être directement corrélée à la quantité de séquences polymérisées présentes dans le tube (Marino, Cook et al. 2003). Très vite reconnu pour ses qualités remarquables (Schneeberger, Speiser et al. 1995 ; Becker, Reith et al. 1996), le *SybrGreen I* deviendra le marqueur fluorescent le plus utilisé pour cette technologie, même si de nombreux autres composés chimiques fluorescents voient toujours le jour (Eischeid 2011), comme par exemple l'*EvaGreen* (Mao, Leung et al. 2007). Parallèlement, des thermocycleurs de plus en plus performants sont développés, permettant notamment des vitesses de transition thermique de plus en plus importantes. En 1996, *Idaho Technology* commercialise le premier fluorimètre pour micro-échantillons présentant un contrôle rapide de température, l'*Idaho Technology LC 24*. C'est en 1997 qu'apparaissent les premiers *LightCycler™*, produits par *Roche Molecular Biochemicals* (Wittwer, Ririe et al. 1997) (**Figure 1.2**).



MACHINE PCR PRIMITIVE  
« MACHINE GENE »



« MR. CYCLE » - CETUS (1985)



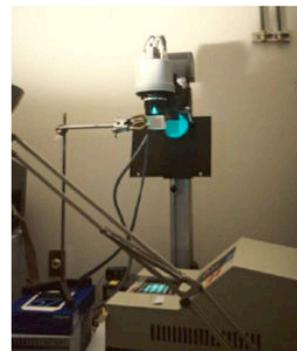
PERKIN ELMER CETUS 480 (1988)



ABI 9700 (1989)



AIR THERMO-CYCLER 1605  
IDAHO (1990)



APPROCHE PAR CAMÉRA CCD  
PERKIN ELMER CETUS 480 (1991-92)



ABI 7700  
(1992)



PERKIN ELMER  
GENEAMP 2400 (1994)



LIGHTCYCLER 24  
IDAHO (1996)



LIGHTCYCLER 1.0 (1997)



LIGHTCYCLER 2.0 (2005)



LIGHTCYCLER 480 II (2008)

Figure 1.2 : Évolution des machines de PCR depuis les premiers bains-marie robotisés (Source : Agesworth) en passant par « Mr. Cycle », premier thermocycleur de Cetus (Source : <http://classic.the-scientist.com/article/display/36690/>), jusqu'à la toute dernière plateforme de PCR en Temps Réel en format multi-puits (Roche Applied Science).

Le suivi en temps réel de l'accumulation de la cible détectée au cours de la réaction d'amplification constitue véritablement une révolution pour la quantification des acides nucléiques. Le traitement et l'analyse de ces nouvelles données de cinétiques vont consécutivement permettre l'établissement des deux principaux modèles de quantification : la quantification absolue et la quantification relative. Les premières études de quantification réalisées à partir de la PCR apparaissent dès 1989 (Becker-Andre and Hahlbrock 1989). La réaction de polymérisation en chaîne est alors utilisée de manière compétitive afin d'obtenir une quantification absolue de l'acide nucléique recherché (Gilliland, Perrin et al. 1990). Il est ainsi précisément déterminé le nombre de copies du transcrit d'un gène cible initialement présent au niveau d'un échantillon donné. Le second modèle de quantification, celui de la quantification relative, ne verra le jour que quelques années plus tard (Livak 1997 & 2001; Livak and Schmittgen 2001), permettant quant à lui d'établir une quantité de gène cible de manière relative à un ou plusieurs gènes dits " de référence ". L'étude des modulations géniques et la quantification des ARN messagers qui en découle vont alors rapidement constituer un important sujet de recherche pour nombre de scientifiques. En raison de ce très fort intérêt et face à la popularité grandissante de la méthode d'amplification de l'ADN, la technique de RT-qPCR va très vite s'imposer comme une méthode de choix dans la quantification de ces différents profils d'expression (Chelly, Kaplan et al. 1988 ; Wang, Doyle et al. 1989 ; Gilliland, Perrin et al. 1990 ; Bustin 2000). Dès la fin des années 90, Stephen BUSTIN, alors Professeur de Science moléculaire de l'Université Queen Mary de Londres, se révèle par la qualité des travaux qu'il réalise sur la détection et la caractérisation de cellules tumorales de patients atteints de cancers colorectaux (Bustin, Gyselman et al. 1999). Auteurs de nombreuses publications et de deux ouvrages de référence (Bustin 2004 ; Bustin 2010), il deviendra au fil des ans un expert reconnu dans le domaine de la Médecine moléculaire. Publiée en 2000, sa première revue est approuvée par la communauté scientifique. Elle décrit les avantages considérables que la biotechnologie peut désormais proposer à ses utilisateurs : simplicité, rapidité, reproductibilité, sensibilité, spécificité et précision de la réaction.

En 2002, Stephen BUSTIN, conscient des principaux problèmes auxquels sont déjà confrontés les différentes équipes de recherche utilisant cette technologie, publie une seconde revue beaucoup plus critique (Bustin 2002), soulignant l'hétérogénéité des réactifs utilisés, les diverses méthodes employées, la normalisation et les traitements statistiques des données obtenues. Deux années plus tard est publiée une revue précisant les dangers et les écueils régulièrement rencontrés au cours de la pratique de cette technique. Stephen BUSTIN y

indique surtout qu'il est d'une importance capitale d'établir rapidement des directives standard concernant la pratique de cette analyse (Bustin and Nolan 2004). Ces propos sont à nouveau renforcés par une nouvelle publication parue l'année suivante (Bustin, Benes et al. 2005). C'est en 2009 que " *The MIQE (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) Guidelines* " est publiée dans le journal *Clinical Chemistry* (Bustin, Benes et al. 2009 ; Bustin, Benes et al. 2011) et apparaît vraiment comme un événement déterminant dans la maturation de la technologie de RT-qPCR. Cet ensemble de recommandations développées par un consortium international de scientifiques académiques a pour principal objectif d'établir un accord universel concernant les questions fondamentales que soulèvent l'utilisation d'une telle méthode en vue de la quantification des acides nucléiques (qualité et contrôle de la quantité de l'ARN, analyse et présentation des résultats, normalisation et standardisation des protocoles). Il fait remarquer la nécessité d'une certaine transparence dans les expérimentations réalisées (Lefever, Hellemans et al. 2009) et publiées dans les revues biomédicales spécialisées considérant tous les aspects de l'ADN et de l'ARN dans un contexte cellulaire, telles que la revue *BMC Molecular Biology* (Bustin, Beaulieu et al. 2010).

---

## I.1. Amplification de l'ADN par PCR

---

### I.1.1. Définitions

L'*Oxford Dictionary of Biochemistry and Molecular Biology* (Datta, Smith et al. 2000) définit plus précisément la PCR comme étant une « méthode par laquelle une séquence spécifique de nucléotides d'un acide désoxyribonucléique double-brin est amplifiée. La séquence est identifiée par l'utilisation de courts oligonucléotides synthétiques complémentaires des deux régions terminales de la séquence d'ADN à amplifier ; ces nucléotides sont rallongés par une ADN polymérase à partir de la lecture de la matrice d'ADN. Les nouvelles chaînes couvrent la région délimitée par les deux régions terminales choisies. Les chaînes naissantes sont dénaturées par la chaleur et le processus est répété un grand nombre de fois. Une amplification d'un facteur 10 000 d'un fragment d'ADN ciblé peut ainsi être obtenue. » Pour ce faire, « l'utilisation d'une ADN polymérase thermostable obtenue à partir de la bactérie thermophile *Thermus aquaticus* (Brock and Freeze 1969) évite la nécessité de rajouter de l'ADN polymérase à chaque cycle car l'enzyme n'est pas inactivée à la température de dénaturation de l'ADN. »

Les capacités et les limites de cette réaction biochimique peuvent aussi être résumées dans une seconde définition beaucoup plus fonctionnelle. En effet, la polymérisation par extension d'amorces peut être considérée comme « une amplification exponentielle, cyclique et ciblée de l'ADN ». Son caractère exponentiel lui confère une grande sensibilité de détection. La reproductibilité de la méthode est quant à elle conférée par le caractère cyclique de la réaction. En effet, une réinitialisation de la réaction se produit à chaque cycle. Cette synchronisation permet alors une efficacité initiale maximum. Enfin, la spécificité est conférée par le caractère ciblé de la réaction. Cette caractéristique est vraie en conditions ordinaires de détection mais disparaît aux limites de la quantification.

### 1.1.2. Les étapes de la PCR

La PCR est constituée d'une succession d'une quarantaine de cycles (on estime entre 34 et 35 environ le nombre théorique de cycles de PCR permettant de détecter 1 copie), comportant chacun 3 principales étapes : la **dénaturation** de l'ADN double brin, l'**hybridation** ou anelage des **amorces sens** et **antisens** et enfin l'extension ou élongation des **amorces**. Tous les éléments nécessaires à la réaction (**amorces**, déoxyribonucléotides (dNTPs) et *Taq* polymérase) sont présents en très large excès dans le mélange réactionnel qui sera soumis aux différentes températures correspondant à chacune des étapes. Le thermocycleur est la machine qui permettra de réaliser automatiquement ces cycles de température.

- **Dénaturation de l'ADN**

- **Dénaturation initiale** : Avant le commencement des différents cycles de la réaction d'amplification de l'ADN, une importante première étape de chauffage à 95°C, d'une durée de 10 à 15 minutes est généralement réalisée. Cette étape permet en effet d'activer les *Taq* polymérases de type " Hot Start ". Dans un premier temps, ces *Taq* polymérases étaient formulées avec des anticorps monoclonaux labiles à la chaleur et capables de neutraliser efficacement l'activité de la protéine à température ambiante. Par la suite, de véritables *Taq* polymérases sont apparues, issues du génie génétique (Birch 1996). Cette première étape thermique permet aussi de déshybrider l'ADN double brin, d'éliminer les éventuelles **structures secondaires**, de dénaturer des enzymes pouvant exister dans la solution comme la **transcriptase inverse**. D'autre part, le chauffage va permettre une augmentation du **mouvement brownien** et donc une homogénéisation du mélange réactionnel.

- **Dénaturations pendant les cycles** : Ces phases sont réalisées à 95°C pendant 1 minute. Elles permettent de synchroniser les réactions (**précision de la PCR**) en réduisant les doubles hélices ADN sous la forme de molécules linéaires simple brin (déshybridation de l'ADN). Au cours de cette **dénaturation**, l'ADN devient alors accessible aux **amorces** intervenant au cours de l'étape suivante. D'autre part, ces étapes permettent aussi de décrocher les **ADN polymérases** qui seraient encore liées à une **matrice**.

- **Annelage / Hybridation des amorces**

Le mélange réactionnel est ensuite rapidement refroidi jusqu'à la **température d'hybridation** ( $T_H$ ) à laquelle les **amorces**, en excès dans le mélange, vont pouvoir se fixer sur la totalité des séquences cibles. Au cours de cette étape cruciale de la PCR, la quantification d'une

cible bien déterminée va imposer l'**hybridation** d'une quantité d'**amorces** proportionnelle à la quantité de la cible, et cela à chaque cycle afin de permettre une amplification de type exponentiel. Il faut donc 100% d'**hybridation** si l'on souhaite s'affranchir des variations aléatoires possibles (*observées à l'aide de **duplicates** d'échantillon*).

Généralement située entre 45 et 60°C, la **température d'hybridation** correspond généralement à la **température de fusion** ( $T_m$ ) du **duplexe** moins 5 à 10°C et dépend essentiellement de trois principaux paramètres :

- **composition en bases de l'amorce** : la liaison hydrogène additionnelle de la **paire de bases** G-C rend les **duplexes** GC-riches beaucoup plus stables que les **duplexes** AT-riches.
- **longueur et concentration de l'amorce** : Le taux de renaturation est en effet proportionnel à ces deux paramètres. De fortes concentrations d'**amorces** favorisent ainsi la formation de **duplexes**.
- **compositions du mélange réactionnel** : Influence de la composition ionique du milieu ( $Mg^{2+}$ ,  $Na^+$ , additifs ...).

Les sites d'**hybridation** des **amorces** sont choisis de telle manière que l'**ADN polymérase** synthétise la région d'intérêt au cours de l'étape d'extension. À cette étape, la cinétique est alors d'ordre 2 : dans un premier temps, les **amorces** cherchent lentement leurs **séquences complémentaires** dans un **mouvement brownien** pour qu'une fois retrouvées, elles puissent s'hybrider de manière plus rapide. Le phénomène qui survient alors est décrit sous le terme de "**zipping de l'ADN**" et débute au niveau de sites bien précis de la séquence, appelés **noyaux de nucléation** (Smith, Britten et al. 1975 ; Murugan 2002 ; Sikorav, Orland et al. 2009).

Cependant, il est important de souligner l'existence d'une compétition entre **amorces** et brins **complémentaires** au niveau moléculaire. En effet, en raison de la longueur des séquences, l'**hybridation** des brins **complémentaires** est thermodynamiquement plus favorable que celle des **amorces** sur le brin matriciel. Cela est compensé au cours de la réaction par une concentration d'**amorces** largement supérieure à celle de la cible et du brin **complémentaire**. D'autre part, lorsqu'une **hybridation** non simultanée des deux **amorces** se produit, une réhybridation du brin **complémentaire** devient alors plus favorable, entraînant une diminution du rendement de la réaction. Très rapidement, la concentration de la cible va augmenter parallèlement à la diminution de la concentration des **amorces**, favorisant de plus en plus l'**hybridation** de la cible sur elle-même. Ce phénomène d'**hybridation** constitue la principale cause de l'arrêt de l'amplification caractérisé par la phase " plateau ". En effet, l'activité enzymatique de la **Taq**

*polymérase* reste largement convenable en fin de réaction (Lawyer, Stoffel et al. 1993 ; Birch 1996)). Ces différents problèmes liés à la difficulté de l'*hybridation* des *amorces* sur leur cible contribuent donc à une vision beaucoup plus dynamique de l'inhibition de *PCR*, malgré le fait que des inhibitions enzymatiques puissent encore être possibles (*inhibiteurs d'ADN polymérase* comme l'*acyclovir* par exemple).

- ***Extension / Élongation des amorces***

Au cours de cette phase, la température remonte rapidement jusqu'à 72°C, température de polymérisation optimale de la *Taq polymérase*. Cette *ADN polymérase* - ADN dépendante va alors synthétiser un brin *complémentaire* à partir des deux *amorces* hybridées en " lisant " la séquence du brin matriciel dans le sens 3'OH vers 5'P (polymérisation par *extension d'amorces*). Elle étend ainsi les *amorces* au niveau de leurs extrémités 3'OH en incorporant correctement les nucléotides présents dans le milieu selon les règles d'appariement établies en 1953 par WATSON & CRICK (Watson and Crick 1953) (***reproductibilité de la PCR***). Cette caractéristique rend alors compte de fidélité enzymatique (Keohavong and Thilly 1989 ; Eckert and Kunkel 1990 ; Eckert and Kunkel 1991) de la *Taq polymérase*. On observe donc à cette étape la polymérisation d'un brin néo-synthétisé dont la séquence est *complémentaire* de celle du brin matriciel défini comme cible. Le temps alloué pour cette étape est dépendant de la taille de l'amplicon et de la *processivité* de l'enzyme (Von Hippel, Fairfield et al. 1994). Il doit être en excès de manière à permettre un rendement de synthèse optimal.

Toutes ces étapes sont réalisées les unes après les autres aussi rapidement que possible (Wittwer, Fillmore et al. 1990) au cours de 30 à 40 cycles d'amplification durant lesquels la quantité d'amplicons va augmenter de manière exponentielle.

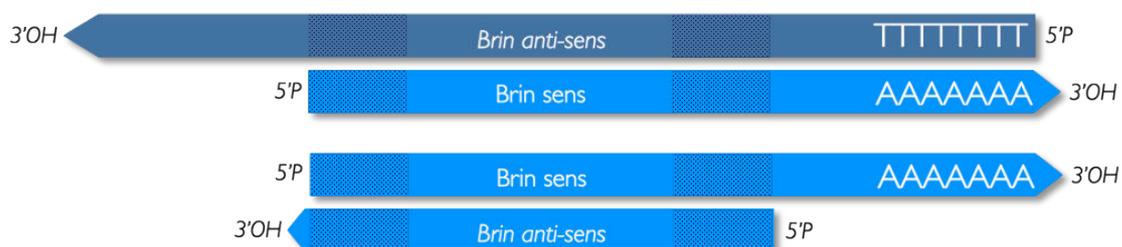
### • TRANSCRIPTION INVERSE (RT)



### • 1<sup>ER</sup> CYCLE DE PCR



### • 2<sup>ÈME</sup> CYCLE DE PCR



### • 3<sup>ÈME</sup> CYCLE DE PCR



ARN ADNC (RT) ADNC (PCR) CIBLES

Figure 1.3 : Espèces ADN natives et néo-synthétisées au cours des 3 premiers cycles de la réaction d'amplification d'une séquence cible en RT-PCR.

Aux cours des premiers cycles d'une réaction de RT-PCR, différentes espèces vont successivement être obtenues (Figure 1.3). Au premier cycle, un brin *sens* est synthétisé à partir du brin d'ADN *complémentaire* obtenu pendant la *transcription inverse*. Ce brin *sens* est précisément défini au niveau de son *extrémité 5'P* mais pas au niveau de son *extrémité 3'OH*.

Au second cycle, il existe alors dans le tube deux types d'ADN matriciel : le brin d'ADN *complémentaire* initialement présent dans l'échantillon d'origine et le brin d'ADN *sens* nouvellement synthétisé d'une longueur variable, composé de la région d'intérêt flanquée de la séquence de l'*amorce* en *5'P*. Généralement, la taille des fragments varie de 100 *paires de bases* (pb) à 2 kilobases (Kb).

Au troisième cycle, les premiers amplicons apparaissent. Ces fragments d'ADN double-brins néo-synthétisés et bornés par les *amorces* vont alors servir à leur tour de *matrice* pour les *amorces* présentes en excès dans le mélange réactionnel. La molécule d'ADN *complémentaire* d'origine est toujours présente et le restera jusqu'à la fin de la réaction mais rapidement, les fragments néo-synthétisés vont apparaître de manière prédominante dans le tube. Pendant ce cycle, toutes les espèces d'ADN qui existeront au cours de la réaction d'amplification sont obtenues, soient :

- un brin d'ADN *antisens* issu de la RT.
- trois brins d'ADN *sens* précisément définis à leur *extrémité 5'P* uniquement.
- un brin d'ADN *sens* correspondant à l'amplicon, c'est-à-dire précisément défini à ses deux *extrémités*.
- trois brins d'ADN *antisens* correspondant à l'amplicon, c'est-à-dire précisément défini à ses deux *extrémités*.

### I.1.3. Caractéristiques de la PCR

On peut distinguer cinq caractéristiques majeures que sont la **spécificité** et la **sensibilité analytique**, la **reproductibilité**, la **précision** et la **répétabilité**.

- **Spécificité analytique**

Elle est conférée par le caractère ciblé de la détection d'une séquence connue existant dans le milieu réactionnel au milieu de séquences non spécifiques également présentes dans le mélange réactionnel. Elle est importante dans des conditions ordinaires de détection, c'est-à-dire lorsque la séquence cible est présente en quantité suffisante dans l'échantillon. Cette spécificité vient cependant à disparaître lorsque l'on se rapproche des limites de quantification (LOQ, définie plus loin). La notion de spécificité, notamment concernant l'évaluation des **amorces**, sera fortement considérée dans les études réalisées au cours de cette thèse.

- **Sensibilité analytique**

Elle est conférée par le caractère exponentiel de la réaction et fait référence au nombre minimum de copies de séquences cibles présentes dans un échantillon pouvant être mesuré avec précision par **PCR** en temps réel. Généralement, la sensibilité est exprimée comme étant la limite de détection (LOD, définie plus loin), c'est-à-dire la concentration pouvant être détectée à l'aide d'une réaction de **PCR** optimisée (Kramer and Coen 2001) avec une certitude raisonnable (la plupart du temps, avec 95% de probabilité). Cette notion de sensibilité sera elle aussi considérée au cours de ces travaux.

- **Reproductibilité**

Elle est conférée par le caractère cyclique de la réaction. En effet, une réinitialisation de ses différents paramètres est observée à chaque cycle avec théoriquement une efficacité initiale maximum et une synchronisation des réactions. Cette précision à long terme (ou variance inter-essai) fait référence à la variation des résultats observés entre les différentes exécutions de programmes thermiques de **PCR** (ou *runs* de **PCR**) entre différents laboratoires ayant réalisé la même analyse. Elle est généralement représentée par l'écart-type (SD) ou le coefficient de variation (CV) du nombre de copies ou de concentrations. Il a cependant été démontré que les valeurs caractéristiques de chaque amplification, notées  $C_q$  (pour *Cycle de quantification*), générées à partir de différents *runs* sont sujettes à une variation inter-*run* inhérente (Hellemans, Mortier et al. 2007). Il n'est donc pas approprié de calculer des ratios de variation de  $C_q$  inter-*run*.

- **Précision**

Elle fait référence à la différence entre les concentrations expérimentalement mesurées et les concentrations réelles, représentées par le calcul du facteur existant entre les valeurs les plus hautes et les valeurs les plus basses (*fold-change*) ou d'estimations du nombre de copies.

- **Répétabilité**

C'est une précision à court terme (ou variance intra-essai) qui fait référence à la précision et à la robustesse de l'analyse lorsque les mêmes échantillons sont analysés à plusieurs reprises. Elle peut être exprimée par l'écart-type (SD) de la variance des  $C_q$ . Le coefficient de variation (CV) du nombre de copies ou la variance de la concentration sont aussi des valeurs statistiques permettant d'apprécier cette notion. Le coefficient de variation ne doit toutefois pas être utilisé lorsque des  $C_q$  sont considérés (Schmittgen and Livak 2008).

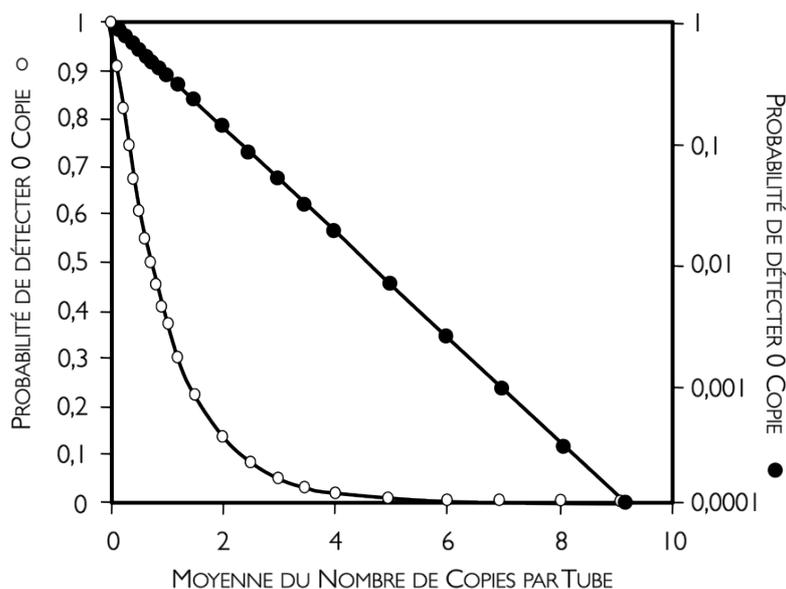
#### 1.1.4. Limites de la PCR

La réaction d'amplification de l'ADN présente deux limites majeures que sont la limite de détection (*Limit of Detection*, LOD) et la limite de quantification (*Limit of Quantification*, LOQ).

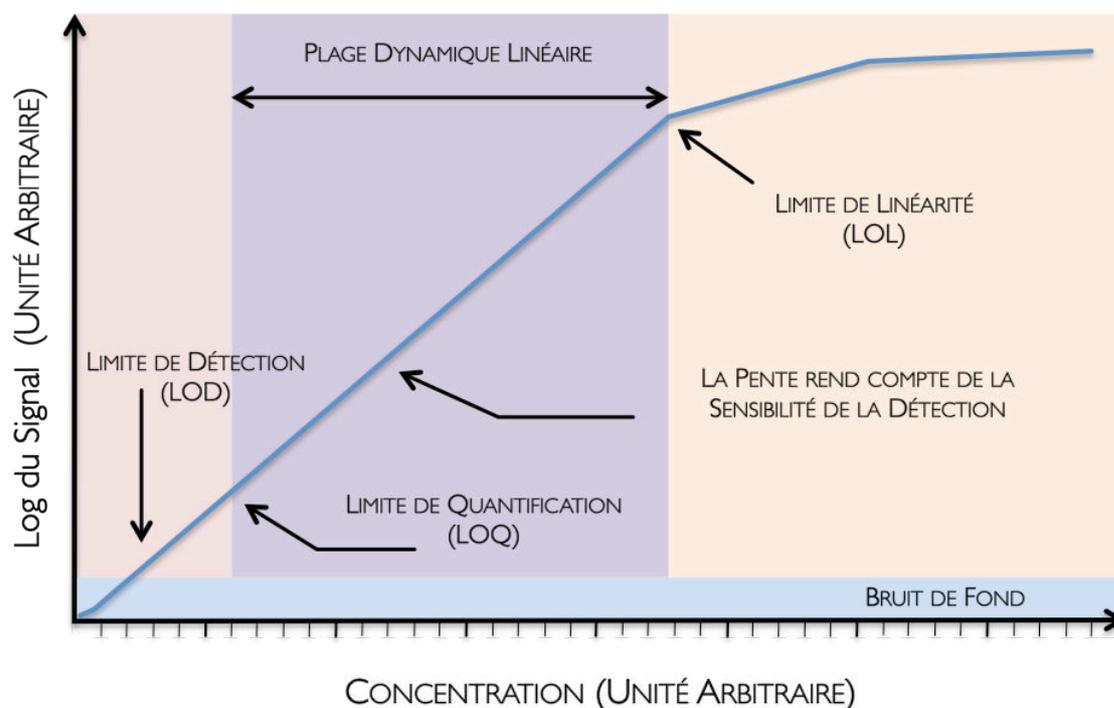
- **Limite de détection (LOD)**

D'après l'*International Union of Pure and Applied Chemistry* (IUPAC), la limite de détection est la plus petite quantité ou concentration d'un composé pouvant être significativement distinguée de manière fiable par rapport à un bruit de fond ou un niveau de base au niveau d'un échantillon à analyser (**Figure 1.5**).

Si l'on cherche simplement à détecter la présence d'une cible (détection qualitative), alors il faut s'interroger sur la signification d'un résultat négatif qui peut être très souvent trompeur. Ainsi, des résultats expérimentaux inférieurs à la limite de détection théoriquement possible ne devraient jamais être présentés. Bien que difficile, un calcul de la LOD peut cependant être réalisé en tenant compte de l'**hétéroscédasticité** des données généralement obtenues en **PCR** en temps réel (Burns and Valdivia 2008). Ainsi, si l'on suppose une distribution de Poisson, on peut déterminer le pourcentage de chance de détecter une seule copie de la cible au cours de la réaction de **PCR**.



**Figure 1.4 :** La probabilité de détecter zéro copie en fonction du nombre réel moyen de copies par réaction. Les données ont été extraites des tableaux de distribution de la Loi de Poisson et sont à la fois présentées sur une échelle linéaire (à gauche, points blancs) et une échelle logarithmique (à droite, points noirs).



**Figure 1.5 :** Courbe de calibration précisant les moments où la limite de détection (LOD), la limite de quantification (LOQ) et la limite de linéarité (LOL) sont atteints. Les limites de quantification et de linéarité définissent la plage dynamique linéaire. Source : *Sample Preparation Techniques in Analytical Chemistry*. Mitra, S. First edition ed2003, Hoboken, New Jersey, USA: Wiley-Interscience. 488, page 5.

Environ 5 copies en moyenne sont nécessaires pour que 99% des réactions de PCR soient capables de détecter au moins 1 copie alors que 10 copies par réaction en moyenne sont nécessaires pour réduire le risque d'obtenir un résultat faux-négatif en dessous de 0,01% (Figure 1.4). En théorie, la valeur de la LOD la plus sensible possible est de 3 copies par PCR (Wittwer and Kusakawa 2004). La modélisation et la détermination appropriée de la limite de détection en qPCR reste toujours l'objet de recherches en cours.

- **La limite de quantification (LOQ)**

La limite de quantification d'une méthode analytique est la plus petite quantité ou concentration d'un composé pouvant être déterminée quantitativement avec un niveau acceptable de précision et d'exactitude au niveau d'un échantillon (Figure 1.5). La LOQ est atteinte lorsque la réaction perd son caractère exponentiel.

### 1.1.5. Thermodynamique et hybridation moléculaire

La phase d'hybridation des amorces sur leur cible constitue le point critique de la quantification par qPCR. Dans une logique d'optimisation de la réaction, la compréhension thermodynamique de ce mécanisme est donc absolument capitale. Au cours de cette étape, un oligonucléotide d'une vingtaine de bases et de séquence connue va s'hybrider avec une ou plusieurs autres séquences partiellement ou totalement complémentaires en milieu aqueux. La spécificité de l'hybridation est alors dictée par la complémentarité des bases établie par WATSON & CRICK : la guanine (G) se lie à la cytosine (C) par l'établissement de trois liaisons hydrogène et l'adénine (A) se lie à la thymine (T) ou l'uracile (U) par l'établissement de seulement deux liaisons hydrogène. Par cette liaison hydrogène supplémentaire, les paires de bases G-C contribuent ainsi davantage à la stabilité d'un duplexe, expliquant ainsi pourquoi le pourcentage GC d'une séquence - mais ne devrait-on pas plutôt parler du nombre de liaisons hydrogène impliquées ? - revêt une importance majeure dans le design des amorces utilisées en PCR (Haynie 2008).

### 1.1.5.1. Paramètres thermodynamiques

L'énergie libre de Gibbs, l'enthalpie et l'entropie sont les trois paramètres thermodynamiques permettant de comprendre l'état dans lequel se trouve un système thermodynamique. Elles sont définies de la manière suivante :

- L'énergie libre (notée  $G$ ) est une fonction d'état introduite par Josiah Willard GIBBS en 1878 indiquant la direction dans laquelle va spontanément évoluer un système. La variation de cette énergie, notée  $\Delta G$ , représente le travail effectué par un système subissant un processus réversible à température et pression constantes. Le système considéré va alors spontanément évoluer dans le sens de la diminution de l'énergie libre de Gibbs selon l'équation :

$$\Delta G = \Delta H - T\Delta S \quad (1.1)$$

où :  $\Delta G$  est la variation d'énergie libre ( $\text{kcal.mol}^{-1}$ )

$\Delta H$  est la variation d'enthalpie ( $\text{kcal.mol}^{-1}$ )

$T$  est la température absolue ( $^{\circ}\text{K}$ )

et  $\Delta S$  est la variation d'entropie ( $\text{kcal.mol}^{-1}.\text{K}^{-1}$ ).

La quantité  $\Delta G$  ( $\text{cal.mol}^{-1}$ ) est négative pour **des processus spontanés libérateurs d'énergie**. Elle est positive pour **des réactions consommatrices d'énergie**.

- La variation d'enthalpie, notée  $\Delta H$ , est une mesure du flux de chaleur au cours d'un processus. Lorsque  $\Delta H$  est négatif, la réaction est dite **exothermique** : le flux de chaleur progresse depuis le système jusqu'à l'environnement. Lorsque  $\Delta H$  est positif, la réaction est dite **endothermique** : le flux de chaleur progresse alors depuis l'environnement jusqu'au système.

- L'entropie est une fonction thermodynamique qui mesure le degré de désordre d'un système. Elle peut parfois être exprimée en unité entropique ( $1 \text{ eu} = 1 \text{ cal.mol}^{-1}.\text{K}^{-1}$ ). Une valeur  $\Delta S$  négative correspond ainsi à une diminution du désordre, une valeur  $\Delta S$  positive correspondant à une augmentation du désordre. La formation de **duplexe** d'ADN au cours de l'**hybridation** entraîne par conséquent une diminution du désordre.

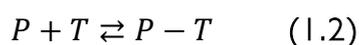
Au cours de la phase d'**hybridation**, l'établissement de liaisons hydrogène va devenir le paramètre déterminant dans le maintien ou la rupture de la stabilité de l'interaction. On sait qu'une molécule d'eau est capable de former jusqu'à quatre liaisons hydrogène avec des

molécules environnantes. La question qui se pose alors est pourquoi ces liaisons établies entre deux séquences sont-elles plus favorables que lorsqu'elles s'établissent au sein même d'une molécule simple brin ou tout simplement avec les molécules d'eau ? Deux molécules simple brin présentent tout simplement des états d'enthalpie et d'entropie plus importants qu'une double hélice ADN à la température d'**hybridation**. En effet, deux molécules qui s'hybrident pour former un **duplexe** constituent de manière évidente une situation beaucoup plus ordonnée que lorsque ces deux mêmes molécules évoluent librement dans un milieu aqueux (Haynie 2008).

Les valeurs  $\Delta G^\circ$ ,  $\Delta H^\circ$  et  $\Delta S^\circ$  représentent les variations d'énergie libre, d'enthalpie et d'entropie dans des conditions standard de température (température d'intérêt) et de pression (1 atm). La valeur  $\Delta G^\circ_T$  représente la variation d'énergie libre standard à la température T. Les valeurs  $\Delta H$  et  $\Delta S$  dépendent à la fois de la température et de la concentration en sels (SantaLucia 1998) et **devraient être déterminées** dans les conditions de **PCR**. Il est important de noter que ces valeurs dépendent également de la température. Cependant, il a été montré qu'au moins pour de courtes séquences, elles peuvent être considérées comme constantes dans une gamme de température allant de 0 à 100°C (SantaLucia and Turner 1997).

### 1.1.5.2. Constante d'équilibre

En premier lieu, définissons ce qui constitue le système thermodynamique considéré en **PCR**. Il correspond au contenu intégral du capillaire en verre du *LightCycler*<sup>TM</sup>, c'est-à-dire les brins d'acides nucléiques, les solvants et tampons, les sels ainsi que toutes les autres molécules susceptibles de le composer (dNTP, *SybrGreen I*, DMSO etc.). L'« environnement » est alors défini comme le reste de l'univers tout entier. Au cours de la **PCR**, différentes molécules vont interagir entre elles : **amorces**, **matrices** spécifiques, **matrices** non spécifiques, **polymérase**... L'**hybridation** peut être définie comme une réaction d'équilibre entre l'**amorce** (P), sa cible (T) et le **duplexe** constitué de l'**amorce** hybridée sur sa cible (P-T) selon l'équation suivante :



À l'équilibre, la proportion entre les différentes espèces est donnée par la constante d'équilibre :

$$k = \frac{[P-T]}{[P].[T]} \quad (1.3)$$

où les concentrations des espèces sont données en **mol.l<sup>-1</sup>**

La relation entre la variation d'énergie libre et la constante d'équilibre à une température donnée est alors donnée par l'équation :

$$\Delta G^{\circ}_T = -R.T.\ln(k) \quad (1.4)$$

où  $R$  est la constante des gaz parfaits ( $1,9872 \text{ cal.mol}^{-1}.\text{K}^{-1}$ ).

En combinant les équations 1.4 et 1.7, on obtient :

$$T = \frac{\Delta H^{\circ}}{\Delta S^{\circ} - R.\ln(k)} \quad (1.5)$$

où  $T$  est la température en **degré Kelvin** ( $^{\circ}\text{K}$ )

Cette équation est utilisable à la température de fusion où l'on observe l'égalité :  $[T] = [P - T]$  (1.6), la constante d'équilibre devenant alors :

$$k = \frac{1}{[P]} \quad (1.7)$$

En supposant que la concentration initiale d'amorces  $[P_i]$  est supérieure à la concentration initiale de cibles  $[T_i]$ , on démontre facilement qu'à l'équilibre, la concentration d'amorces libres, notée  $[P]$ , est donnée par l'équation :

$$[P] = [P_i] - \frac{[T_i]}{2} \quad (1.8)$$

Dans des conditions de PCR où la concentration de la séquence cible est négligeable par rapport à celle des amorces, on peut donc estimer que  $[P] \approx [P_i]$ . Par contre, si  $[P_i] = [T_i]$ , alors :

$$[P] = \frac{[P_i]}{2} \quad (1.9)$$

Ces formules ne sont valables que pour des **oligonucléotides** qui ne sont pas auto-complémentaires.

### *1.1.5.3. Estimations de $\Delta S$ et $\Delta H$ pour un oligonucléotide de séquence connue*

Établi pour la première fois par Philip BORER en 1974 (Borer, Dengler et al. 1974) à partir des dix appariements possibles selon les règles de WATSON & CRICK, le modèle dit « *du plus proche voisin* » constitue le modèle le plus utilisé pour l'estimation des paramètres thermodynamiques  $\Delta S$  et  $\Delta H$  d'une séquence connue.

	$\Delta H^\circ$ (kcal.mol <sup>-1</sup> )	$\Delta S^\circ$ (cal.mol <sup>-1</sup> .K <sup>-1</sup> )	$\Delta G^\circ_{37}$ (kcal.mol <sup>-1</sup> )
AA/TT	-7,9 ± 0,2	-22,2 ± 0,8	-1,00 ± 0,01
AT/TA	-7,2 ± 0,7	-20,4 ± 2,4	-0,88 ± 0,04
TA/AT	-7,2 ± 0,9	-21,3 ± 2,4	-0,58 ± 0,06
CA/GT	-8,5 ± 0,6	-22,7 ± 2,0	-1,45 ± 0,06
GT/CA	-8,4 ± 0,5	-22,4 ± 2,0	-1,44 ± 0,04
CT/GA	-7,8 ± 0,6	-21,0 ± 2,0	-1,28 ± 0,03
GA/CT	-8,2 ± 0,6	-22,2 ± 1,7	-1,30 ± 0,03
CG/GC	-10,6 ± 0,6	-27,2 ± 2,6	-2,17 ± 0,05
GC/CG	-9,8 ± 0,4	-24,4 ± 2,0	-2,24 ± 0,03
GG/CC	-8,0 ± 0,9	-19,9 ± 1,8	-1,84 ± 0,06
Initiation avec un G-C terminal	0,1 ± 1,1	-2,8 ± 0,2	0,98 ± 0,05
Initiation avec un A-T terminal	2,3 ± 1,3	4,1 ± 0,2	1,03 ± 0,05

**Tableau I.1** : Paramètres thermodynamiques unifiés des dix appariements possibles pour l'initiation et la propagation de l'hélice ADN en 1 M NaCl (Allawi and SantaLucia 1997).

Ce modèle suppose que la stabilité d'une **paire de base** spécifique dépend des bases qui lui sont voisines. Lors de la formation d'un **duplexe**, cette stabilité est alors mesurée par la variation d'énergie libre standard  $\Delta G^\circ$ , définie comme étant la somme de trois termes :

- (1) **une diminution de l'entropie** liée à la perte de liberté observée lors de l'établissement des premières liaisons hydrogène (énergie libre d'initiation).
- (2) **la somme des termes** pour les interactions existant entre toutes les **paires de bases**.
- (3) **une pénalité entropique** pour les séquences auto-**complémentaires**.

Après la publication de plusieurs études établissant des valeurs de  $\Delta G^\circ_{37}$  selon différentes approches (Gotoh and Tagashira 1981 ; Vologodskii, Amirikyan et al. 1984 ; Breslauer, Frank et al. 1986 ; Delcourt and Blake 1991 ; Doktycz, Goldstein et al. 1992 ; SantaLucia, Allawi et al. 1996 ; Sugimoto, Nakano et al. 1996 ; Allawi and SantaLucia 1997), John SANTALUCIA Jr. détermine en 1998 les paramètres thermodynamiques unifiés des dix appariements possibles permettant de prédire la stabilité d'un **duplexe** ADN en fonction de la

base la plus proche (**Tableau 1.1**) (SantaLucia 1998). À partir de ce modèle, la tendance observée concernant la stabilité des 4 bases de l'ADN en fonction du plus proche voisin (Borer, Dengler et al. 1974) et à la température de 37°C est alors la suivante :

$$GC > CG > GG > GA \approx GT \approx CA > CT > AA > AT > TA$$

- **Correction pour la concentration en Na<sup>+</sup>**

L'équation 1.5 exprime la température en *Kelvin*. Après conversion en degré *Celsius*, cette équation devient :

$$T = \frac{\Delta H^\circ}{\Delta S^\circ - R \cdot \ln(k)} - 273,15 \quad (1.10)$$

Cette formule exprime la température en degrés *Celsius* et suppose une concentration ionique en Na<sup>+</sup> de 1M. Par conséquent, elle ne montre pas la dépendance de la température par rapport à la concentration ionique du milieu dans lequel se déroule l'*hybridation*. Bien que des formules plus compliquées existent, une *expression* de la température tenant compte de la concentration en sodium est donnée par l'équation suivante (Wetmur 1991) :

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ - R \cdot \ln(k)} + 16,6 \times \log_{10} \frac{[Na^+]}{1,0 + 0,7 [Na^+]} - 269,3 \quad (1.11)$$

- **Correction pour la concentration en Mg<sup>2+</sup>**

On sait que les cations divalents se lient invariablement au niveau des groupements phosphate des acides nucléiques, chargés négativement. On a démontré par différentes techniques que l'effet stabilisant des ions Mg<sup>2+</sup> sur un *duplexe* ADN est environ **140 fois plus important** que celui des ions Na<sup>+</sup> (Nakano, Fujimoto et al. 1999). Bien que présentant un effet non linéaire sur la *température de fusion* d'une séquence, le magnésium permet généralement une augmentation du  $T_m$  (Eichhorn and Shin 1968). La valeur théorique d'une séquence est alors augmentée d'environ 5 à 8°C à une concentration de 1,5 mM de Mg<sup>2+</sup>.

En étudiant les effets du magnésium, du potassium, du sodium, des ions Tris et des désoxynucléotides triphosphates sur les profils de *fusion* de *duplexes* ADN, Richard OWCZARZY confirme l'effet dominant de la concentration de magnésium libre sur la stabilité du *duplexe*, les dNTP lient les ions magnésium stœchiométriquement (1:1) avec une haute affinité (Owczarzy, Moreira et al. 2008).

#### 1.1.5.4. Applications à la PCR

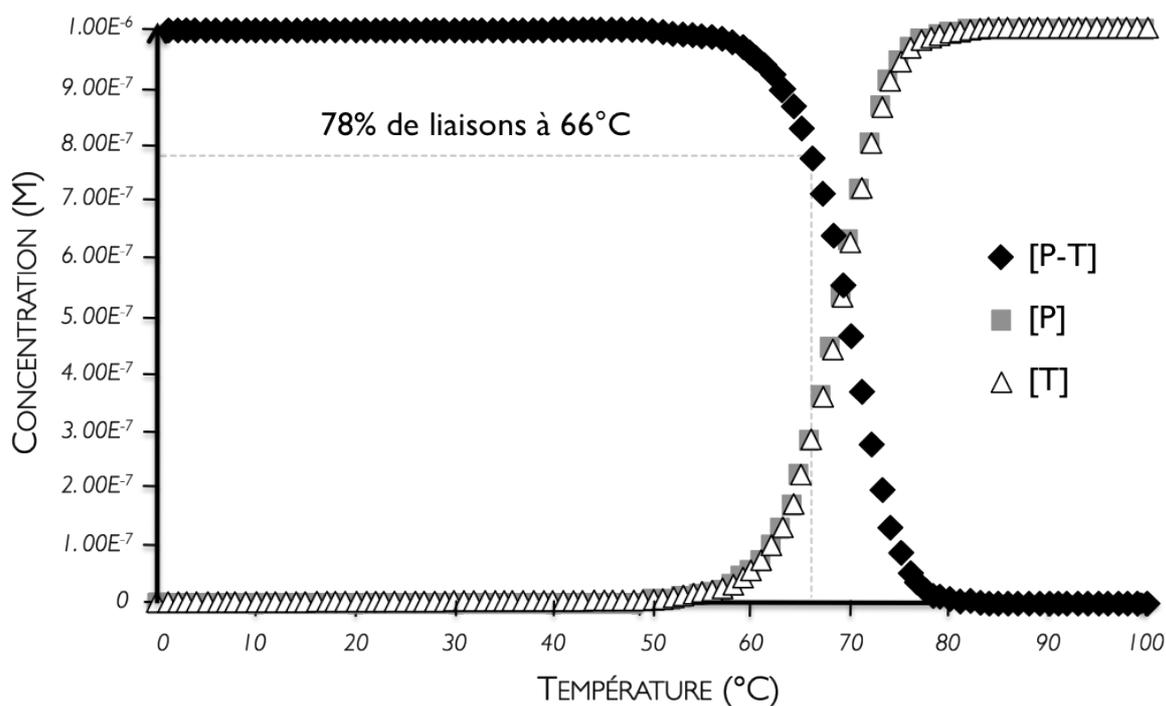
Lorsque l'on met au point une réaction de PCR, la principale difficulté consiste à déterminer la température idéale d'hybridation en fonction des amorces que l'on a dessinées. En théorie, il est possible de déterminer le pourcentage d'amorces hybridées en fonction de la température, en se basant à la fois sur l'équation (1.5) et sur les valeurs de  $\Delta H$  et de  $\Delta G$  (Figure 1.5) (SantaLucia 2007).

Cette méthode est très peu utilisée car elle présente plusieurs inconvénients. Tout d'abord, un inconvénient théorique apparaît clairement puisqu'il est généralement assez difficile de déterminer précisément des paramètres thermodynamiques dans un milieu de PCR. Une prédiction de la valeur du  $T_m$  est traditionnellement la valeur la plus souvent utilisée en PCR afin d'équilibrer les amorces et de déterminer la température d'hybridation idéale. On considère alors souvent que 100% d'hybridations est obtenu lorsque  $T_H = T_m - 5^\circ\text{C}$ . En dehors de la précision de prédiction de cette valeur de  $T_m$ , cette méthode présente donc un dernier inconvénient majeur : au regard de l'équation 1.10, des amorces peuvent avoir le même  $T_m$  mais un pourcentage d'amorces hybridées différent à la température  $T_m - 5^\circ\text{C}$ .

Par ailleurs, le modèle à deux états (équation 1.3) ne traduit que partiellement ce qui se passe réellement dans le milieu réactionnel. En effet, il faudrait tenir compte de la totalité des hybridations qui ont lieu parallèlement à l'hybridation de l'amorce sur sa cible, comme par exemples :

- les hybridations de la cible sur son brin complémentaire
- les hybridations générant des structures secondaires
- les hybridations de l'amorce sur elle-même ou sur les autres amorces (espèces à haute concentration)
- les hybridations d'amorces sur des séquences non ciblées présentes dans le milieu réactionnel (Hybridation non spécifique)

Ces différents évènements d'hybridations modifient très probablement la cinétique d'hybridation et sont au moins en partie responsables (avec les inhibiteurs de PCR) des différences de cinétique d'hybridation existant au niveau d'un échantillon biologique et lors d'une PCR « idéale », c'est-à-dire un fragment recombinant détecté dans un mix de PCR avec impossibilité d'hybridations parasites.



**Figure 1.5 :** Simulation du profil d'hybridation pour une transition simple et non-auto-complémentaire de deux états, en considérant les valeurs  $\Delta H^\circ$  et  $\Delta S^\circ$  à l'aide des équations 1.3 et 1.4. La détermination du pourcentage de liaisons peut être réalisée à n'importe quelle température. Les concentrations d'amorces libre P (carrés) et de cible T (triangles) sont superposées (Source : " *Physical Principles and Visual-OMP Software for Optimal PCR Design* ", *Methods in Molecular Biology: PCR Primer Design*, Anton Yuryev, Ed., Humana Press).

Ces hybridations parallèles expliquent deux importants problèmes en PCR :

- a) les hybridations d'amorces sur des séquences non ciblées présentes dans le milieu réactionnel sont responsables des limites de quantification de la PCR. De plus, dans un contexte de cible rare à détecter, l'hybridation des amorces sur des séquences non spécifiques présente un bénéfice thermodynamique plus important qu'une hybridation spécifique, générant généralement et rapidement des produits non spécifiques par simple recombinaison (Malgoyre, Banzet et al. 2007). Ceci est largement favorisé par le fait qu'à la différence de l'ARN, certains mésappariements rencontrés au niveau de l'ADN, comme au niveau de la guanine par exemple, ont la propriété de stabiliser ces hybridations non spécifiques (SantaLucia and Hicks 2004).

- b) une perte d'efficacité de l'**hybridation** des **amorces** lors de l'accumulation de la **matrice**. En effet, une **hybridation** des **amorces** sur la cible prédomine à faible concentration de **matrice** alors qu'une réhybridation de la **matrice** sur son brin **complémentaire** est plutôt favorisée à forte concentration de **matrice**.

### 1.1.6. Cinétiques d'amplification

Avec l'apparition des premières courbes d'amplification obtenues par **PCR** en temps réel, plusieurs études ont dans un premier temps cherché à modéliser ce phénomène d'amplification de l'ADN. Même si le modèle exponentiel semble être *a priori* le modèle le plus approprié pour la modélisation de la **PCR**, il ne reste limité qu'aux premiers cycles de la réaction. D'autre part, de par sa nature exponentielle, de très faibles erreurs sur la détermination de l'efficacité de la réaction se traduiront par une grande dispersion des valeurs de  $N_0$ .

- **Le modèle exponentiel de la PCR**

Dans ce modèle, le nombre de molécules obtenues au bout de  $n$  cycles ( $N_c$ ) peut être estimé par une première équation théorique, de fonction exponentielle :

$$N_c = N_0 \times 2^n \quad (1.12)$$

- où :
- $N_c$  est le nombre de copies détectées au bout du cycle  $n$
  - $N_0$  est le nombre de copies initialement présentes dans le tube
  - $n$  est le nombre de cycles

Dans l'équation 1.12, on considère une efficacité de réaction **constante et théorique** égale à 2. Au départ de la réaction, on constate que la fonction, linéaire dans un premier temps, perd ce caractère au cours des cycles suivants. En effet, pour des raisons stochastiques de dynamiques de réaction au cours du cycle d'amplification, une molécule peut ou ne peut pas être dupliquée par l'**ADN polymérase**, expliquant clairement qu'aucune quantification absolue du nombre de copies de séquence cible dans le tube ne peut être obtenue à l'aide de cette réaction. Expérimentalement, l'équation 1.12 n'est donc vérifiée que dans les tout premiers cycles de la réaction.

Afin d'établir une équation plus réaliste de la PCR, il est nécessaire de tenir compte de l'efficacité d'amplification  $\epsilon$  de la réaction :

$$N_c = N_0 \times (1 + \epsilon)^c \quad (1.14)$$

où :  $N_c$  est le nombre de copies détectées au bout du cycle  $c$

$N_0$  est le nombre de copies initialement présentes dans le tube

$\epsilon$  est l'efficacité d'amplification de la réaction

Une relation linéaire de type  $y = ax + b$  peut ensuite être définie en dérivant la dernière équation de la manière suivante :

$$\text{Log } N_n = \text{Log } (1 + \epsilon) \times n + \text{Log } N_0 \quad (1.15)$$

où :  $a$  serait la pente de valeur  $\text{Log } (1 + \epsilon)$

et  $b$  l'ordonnée à l'origine de valeur  $\text{Log } N_0$

Cette modélisation permet de déterminer à la fois l'efficacité de PCR et le nombre initial de séquences cibles par simple régression linéaire. Elle est cependant imparfaite car elle suppose à tort qu'il existerait à la fois un processus infini de l'accumulation d'amplicons et une efficacité constante du premier cycle jusqu'au cycle de quantification  $C_q$ . En effet, l'efficacité de la réaction est variable d'un cycle à l'autre avec des cinétiques différentes pour un même gène détecté à partir d'un même échantillon tissulaire (Liu and Saint 2002). D'autre part, il a été démontré de manière expérimentale que la précision des données quantitatives obtenues à l'aide de ce modèle n'existe que sur une plage dynamique relativement restreinte évaluée entre 7 et 8 log (Morrison, Weis et al. 1998). En effet, lorsque l'on arrive aux limites de détection, la réaction devient beaucoup plus aléatoire et peut présenter une forte variabilité, notamment lorsqu'un très faible nombre de molécules cibles doit être détecté. La moindre erreur d'hybridation des amorces conduit alors à une amplification de produits non spécifiques et une stratégie de détection plus efficace devient alors nécessaire (Teo, Choi et al. 2002 ; Malgoyre, Banzet et al. 2007).

Il est important de souligner que, dans ce modèle, la fonction mathématique définit le phénomène d'amplification de l'ADN comme un processus illimité (Figure 1.6A). Bien que cela soit manifestement irréaliste, on considère cette approximation raisonnable, au moins concernant les tout premiers cycles de la réaction. Ce modèle suppose également une efficacité

d'amplification constante (Livak 1997 & 2001; Pfaffl 2001) et sensiblement identique pour tous les échantillons analysés. Cette hypothèse n'est évidemment pas vraie tout au long de la réaction mais encore une fois, cette approximation peut encore être considérée comme raisonnable dans les premiers cycles de la réaction (Liu and Saint 2002).

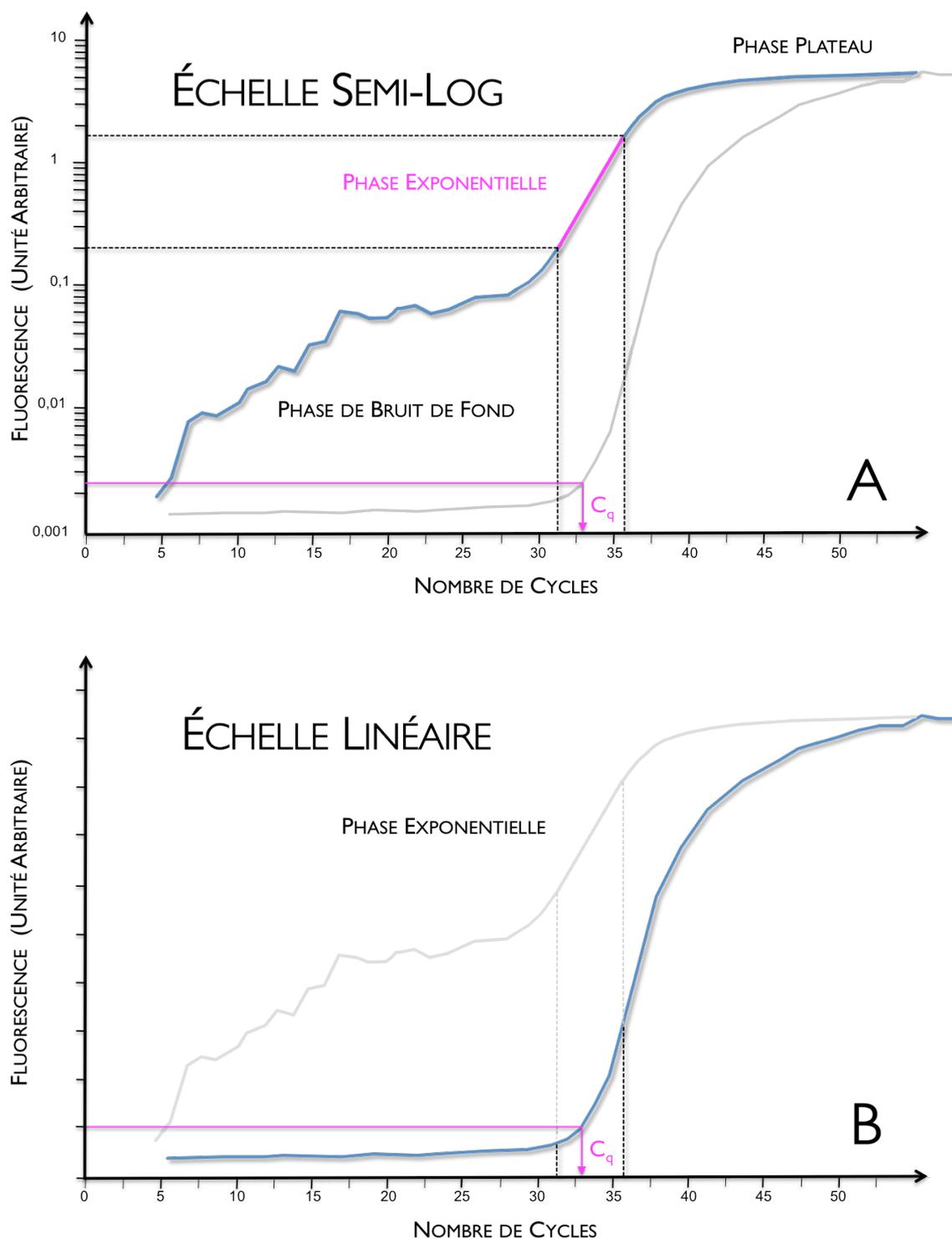


Figure 1.6 : Représentations semi-logarithmique (A) et linéaire (B) d'une cinétique d'amplification d'ADN par mesure de fluorescence.

Sur une représentation semi-logarithmique, trois phases peuvent être distinguées : une phase dite « de bruit de fond », une phase exponentielle et une phase dite de « plateau » (**Figure 1.6A**). Durant toute la phase de bruit de fond, une certaine quantité de séquence cible est dénaturée de manière récurrente jusqu'à ce que les amorces puissent s'y fixer spécifiquement. À ce stade, l'hybridation des amorces au niveau de cette séquence est très favorable en raison de leur présence en très large excès dans le milieu réactionnel. Une fois hybridées, une importante quantité d'amplicons va être générée pendant 4 à 5 cycles de PCR. La fluorescence accumulée et enregistrée va alors augmenter de manière exponentielle. Cette phase dite « exponentielle » correspond à la partie linéaire d'une courbe représentée en échelle semi-logarithmique. À ce moment précis, l'efficacité de la réaction est maximale et peut facilement être déduite de manière précise à partir de la courbe. En effet, celle-ci correspond exactement à la pente de la droite au point d'inflexion. Au même moment et à partir d'une représentation linéaire de cette même amplification, le cycle de quantification de l'échantillon  $C_q$  peut alors être précisément déterminé (**Figure 1.6B**). Au moment de la phase plateau, l'efficacité de la réaction diminue rapidement. Ceci est expliqué par la présence d'une trop grande quantité d'amplicons alors capable d'empêcher une fixation efficace des amorces. À ce moment précis, les séquences cibles tendent à se renaturer préférentiellement sur elles-mêmes et l'hybridation des amorces est alors largement défavorisée.

Afin de mesurer la quantité de cible présente dans l'échantillon et de déterminer l'efficacité de la réaction d'amplification avec la meilleure précision qui soit, plusieurs autres modèles ont été imaginés. Pour ce faire, l'utilisation d'algorithmes mathématiques s'est avérée nécessaire, permettant ainsi le développement de nombreux modèles. Globalement, deux types de modèles peuvent être distingués :

- les modèles linéaires (de type logarithmique)
- les modèles non-linéaires (de type sigmoïdal, polynomial ou biochimique)

- **Les modèles linéaires**

Ces modèles sont utilisés afin d'estimer la quantité de séquences cibles initialement présentes au départ de la réaction ( $N_0$ ) (Gentle, Anastasopoulos et al. 2001 ; Marino, Cook et al. 2003 ; Peirson, Butler et al. 2003 ; Ramakers, Ruijter et al. 2003) (**Figure 1.7B**).

À l'heure actuelle, la méthode *LinReg*<sup>®</sup> (Ramakers, Ruijter et al. 2003) est l'exemple le plus connu construit à partir de ce modèle, et dont l'équation de base est :

$$\log N_c = \log N_0 + c \times \log \epsilon \quad (1.16)$$

où :  $N_c$  est le nombre de copies détectées au bout du cycle  $n$

$N_0$  est le nombre de copies initialement présentes dans le tube

$n$  est le nombre de cycles

L'équation 1.16 est une fonction linéaire de pente égale à  $\log \epsilon$  lorsque l'on considère la cinétique au moment de la phase exponentielle. La régression linéaire est directement liée à l'efficacité de la réaction et à la quantité de séquences cibles initialement présentes au départ de la réaction ( $N_0$ ). Tout comme pour le modèle exponentiel, on notera qu'en raison de la nature exponentielle de l'équation, de très faibles erreurs sur la détermination de l'efficacité de la réaction se traduiront par une grande dispersion des valeurs de  $N_0$ .

- **Les modèles non-linéaires**

Des études ont tenu à prendre en compte l'ensemble de la cinétique d'amplification pour le développement de nouveaux modèles. Bien que relativement plus complexes que les modèles précédents, ils présentent l'avantage d'utiliser des données plus importantes de fluorescence générées au cours de la PCR et ne nécessitent pas forcément la fixation arbitraire d'une valeur « seuil ».

- Un premier modèle de type sigmoïdal a été proposé par LIU et al. (Liu and Saint 2002) en 2002 (Figure 1.7C). Il peut être défini par l'équation 1.17. Défendu deux ans plus tard dans une nouvelle étude (Rutledge 2004), il semble constituer la meilleure modélisation de la réaction à ce jour (Rutledge and Stewart 2008).

$$N_c = \frac{a}{1 + \exp\left(\frac{x_0 - c}{b}\right)} \quad (1.17)$$

où :  $N_c$  est le nombre de copies détectées au bout du cycle  $c$

$a$ ,  $b$  et  $x_0$  sont des paramètres d'ajustement

$c$  est le nombre de cycles

- Le modèle polynomial à quatre paramètres d'Ales TICHOPAD constitue une autre modélisation particulièrement intéressante (Tichopad, Dilger et al. 2003). À la différence des méthodes basées sur le modèle exponentiel et sur l'utilisation d'une gamme de dilutions (Pfaffl 2000 ; Rasmussen 2001), ce modèle considère uniquement la phase exponentielle précoce de la réaction d'amplification. D'après son auteur, le calcul d'une **efficacité de PCR** sensiblement identique à l'**efficacité** réelle de la réaction serait ainsi obtenu.

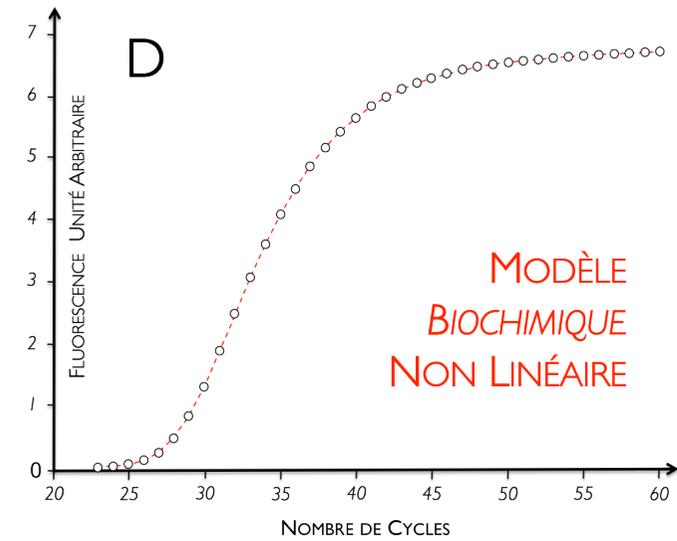
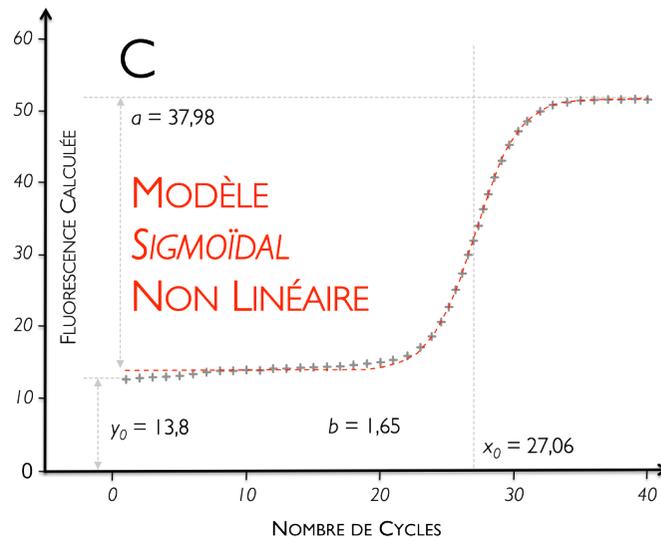
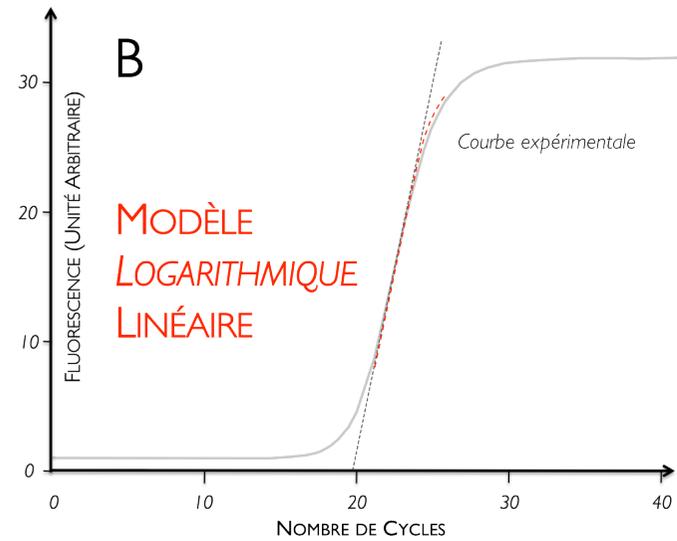
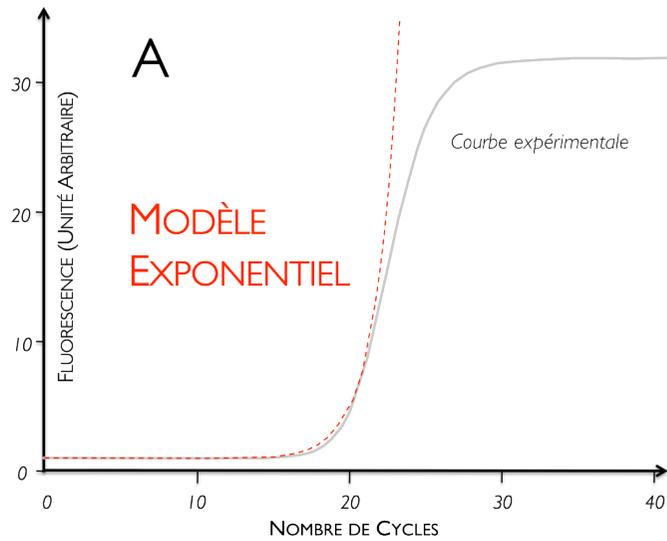
$$N_c = \frac{a}{1 + \left(\frac{c}{x_0}\right)^b} \quad (1.18)$$

- où :
- $N_c$  est le nombre de copies détectées au bout du cycle  $c$
  - $a$  est la différence entre la fluorescence max et le bruit de fond
  - $c$  est le nombre de cycles
  - $x_0$  est le nombre de cycles où la courbe présente le point d'inflexion
  - $b$  est la pente de la courbe nombre de cycles

- La modélisation des processus biochimiques se réalisant au cours des différents cycles de la réaction ont récemment été développés (Mehra and Hu 2005; Smith, Miller et al. 2007), permettant aussi l'estimation des concentrations absolues de séquences cibles de départ (Figure I.7D).

- D'autres modèles peuvent encore être rencontrés, construits à partir d'équations fondamentales comme la fonction de *Gompertz* (Schlereth, Bassukas et al. 1998), l'équation de *Richards* (Yin, Goudriaan et al. 2003), l'équation de *Hill* (Hill 1910 ; Zhao and Fernald 2005) ou bien encore l'équation de *Chapman*.

Concernant tous ces modèles, il est toutefois important de remarquer qu'un problème majeur peut se poser, notamment lorsque la phase plateau de la réaction est considérée dans la quantification de l'échantillon. En effet, une importante variabilité est fréquemment observée à ce niveau, essentiellement due à l'inhibition de la **Taq polymérase** par l'accumulation des amplicons (Kainz 2000) et affectant de manière significative la quantification d'une séquence cible (Morrison and Gannon 1994). La validité d'un modèle construit sur la considération de cette phase passera donc inévitablement par le calcul de l'**efficacité de PCR** correspondant à chacune des réactions.



**Figure 1.7 :** Représentations graphiques de modèles de cinétiques en PCR. **A)** Modèle exponentiel. **B)** Modèle logarithmique. **C)** Modèle sigmoïdal. **D)** Modèle biochimique

---

## I.2. Quantification de l'ADN par qPCR selon le modèle exponentiel

---

### I.2.1. Notion de *Threshold Cycle* $C_T$ , *Crossing Point* $C_p$ et de *quantification Cycle* $C_q$

L'évaluation de la quantité de séquence cible nécessite la définition d'un paramètre de quantification caractéristique de l'échantillon. Recherché au-dessus du bruit de fond réactionnel de fluorescence, ce paramètre devra être déterminé au moment de la phase exponentielle de la réaction (**Figure 1.8**), seul moment où la mesure de quantification est possible sans biais (Wiesner, Beinbrech et al. 1993). Depuis la création du concept de PCR quantitative, quatre différents paramètres ont successivement été définis :

- **Le *Crossing Point*, noté  $C_p$**

Cette mesure est définie après établissement d'un seuil de fluorescence placé juste au-dessus du bruit de fond le plus important de la totalité des échantillons (Wittwer, Ririe et al. 1997 ; Rasmussen 2001). Le  $C_p$  correspond donc au nombre de cycles pour lequel la fluorescence enregistrée au niveau d'un échantillon dépasse un tel seuil. Cette méthode pose cependant deux problèmes : (1) comment déterminer la valeur du seuil de sensibilité et (2) comment prendre en compte la différence de profils entre échantillons.

- **Le *Threshold Cycle*, noté  $C_T$**

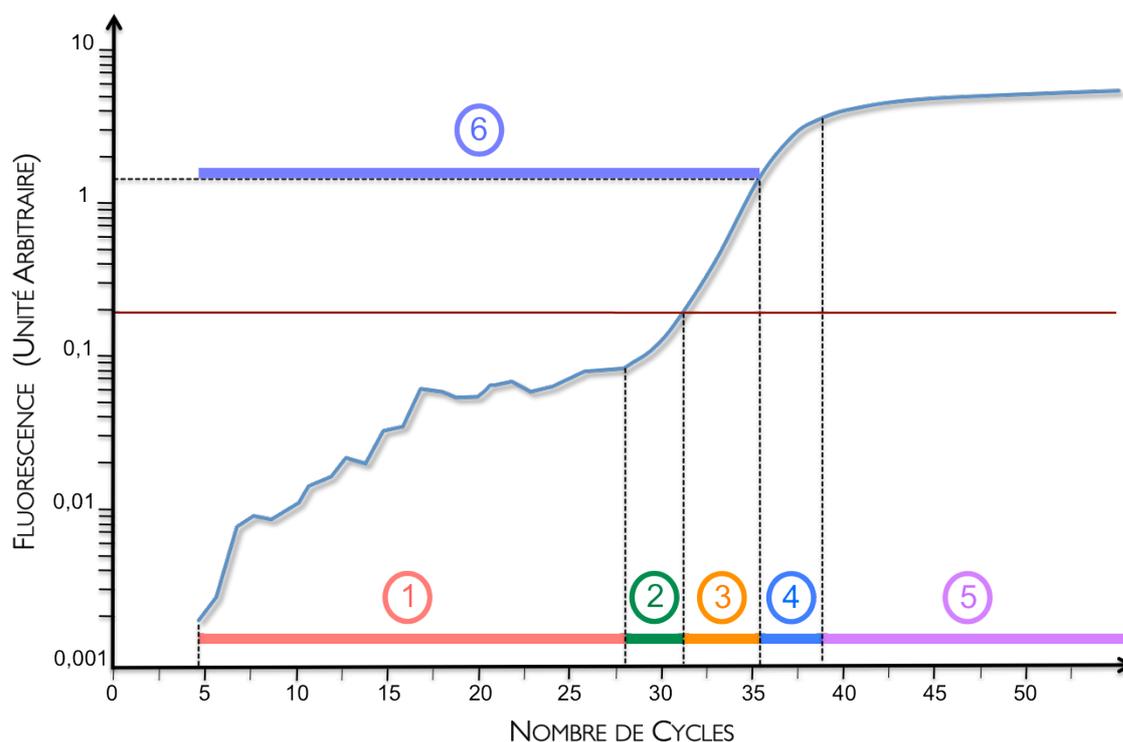
Le  $C_T$  résoud les problèmes posés par la détermination du  $C_p$  en utilisant un seuil de fluorescence pour chaque échantillon, basé sur une estimation statistique du bruit de fond (2011). Ce seuil, *threshold* en anglais, est une valeur numérique assignée et spécifique de chaque échantillon. Il équivaut en général à dix fois l'écart type de la fluorescence des 12 premiers cycles. Le  $C_T$  est donc caractéristique de l'échantillon mais ne tient absolument pas compte de la variation du bruit de fond de fluorescence qui peut être très importante d'un échantillon à l'autre, surtout pendant les douze premiers cycles de la réaction. Il est important de remarquer que cette méthode est efficace mais elle est souvent à l'origine de la majorité des faux positifs en diagnostic. En effet, si le cycle de sortie d'un échantillon positif appartient à l'intervalle de détermination du seuil, l'échantillon est considéré comme négatif. L'utilisation de cette méthode impose donc une vérification systématique des données sources.

- **Le Crossing Point sdm (pour second derivative maximum), noté  $C_{p-sdm}$**

Malgré son nom, cette méthode est plus proche du  $C_T$  que du  $C_p$ . Le seuil est défini pour chaque échantillon. Il correspond à la fluorescence à laquelle on observe un maximum de la dérivée seconde de la courbe d'amplification. Cette valeur est calculée à l'aide d'un algorithme qui peut être mis en défaut quand les niveaux des plateaux d'amplification sont sensiblement différents. Sur le plan théorique, le maximum de la dérivée seconde de la courbe d'amplification de la séquence ciblée en fonction du nombre de cycles correspond au cycle le plus tardif où l'amplification est exponentielle. Cette méthode permet ainsi d'optimiser le rapport signal/bruit. Cette valeur est utilisée dans les logiciels développés par Roche Molecular Biochemicals. C'est la méthode que nous utilisons en routine au laboratoire.

- **Le quantification Cycle, noté  $C_q$**

Le choix du mode de quantification est fixé par le fabricant de l'appareil utilisé. La majorité des appareils utilisent le  $C_T$  à l'exception des appareils Roche qui lui préfèrent le  $C_{p-sdm}$ . En outre, la plupart des appareils peuvent utiliser un  $C_p$  simple. Ces différentes méthodes utilisées correctement sont robustes et, en 2009, il a été proposé de remplacer toutes ces désignations par le terme générique de *quantification Cycle* ( $C_q$ ). Proposé par Stephen BUSTIN (Bustin, Benes et al. 2009), il rentre ainsi en accord avec les données standards de PCR RDML (*Real-Time PCR Data Markup Language* ; [http:// www.rdml.org](http://www.rdml.org)) (Lefever, Hellemans et al. 2009).



**Figure 1.8 :** Représentation semi-logarithmique illustrant les différentes phases d'une cinétique d'amplification par PCR. Le seuil de quantification est représenté par la ligne rouge. 1) Zone de bruit de fond de fluorescence. 2) Zone de mesure de fluorescence biaisée. 3) Phase exponentielle d'amplification. 4) Fin de phase exponentielle. 5) Phase plateau. 6) Phase de cinétique répondant à une loi exponentielle mais où seulement la zone 3 est réellement exploitable.

## 1.2.2. Quantification absolue et courbe de quantification

### 1.2.2.1 Quantification absolue

Les premiers travaux basés sur cette approche ont été réalisés par Michael BECKER-ANDRÉ dès 1989 (Becker-Andre and Hahlbrock 1989) et bien que souvent critiquée pour des raisons de significativité, cette approche quantitative est universellement employée. Cette méthode permet de déterminer la quantité précise d'ADN se trouvant au niveau d'un échantillon donné à partir d'une courbe de quantification (appelée courbe standard externe) obtenue à partir de différentes concentrations connues de séquences cibles (Mitra 2003). Elle rapporte toujours un résultat final de manière relative à une unité définie, comme par exemple un nombre de copies par ng d'ARN total, un nombre de copies par génome, un nombre de copies par cellule, un nombre de copies par gramme de tissu ou un nombre de copies par mL de sang.

Elle est définie par l'équation exponentielle suivante :

$$N_c = N_0 \times (1 + \epsilon)^c \quad (1.14)$$

où :  $N_c$  est le nombre de copies détectées au bout du cycle  $c$

$N_0$  est le nombre de copies initialement présentes dans le tube

$\epsilon$  est l'efficacité d'amplification de la réaction

### 1.2.2.2. Courbe de quantification

La réalisation d'une gamme de quantification reste toujours de nos jours une étape très discutée par les utilisateurs de la RT-qPCR, elle n'en reste cependant pas moins une étape incontournable dans l'optimisation d'une réaction de PCR. Réalisée à partir du *pool* des ADN complémentaires (ADNc) obtenus pour chaque échantillon, elle va globalement prendre en compte la qualité de ces acides nucléiques avec les éventuels inhibiteurs les accompagnant. Ainsi, une réaction d'amplification de bonne qualité présentera une gamme linéaire sur 4 à 5 log avec une pente de -3,32 (100% d'efficacité de PCR) et un coefficient de corrélation élevé ( $r^2 > 0,99$ ). Il est important de remarquer qu'il existe un danger car ces courbes sont dans la pratique relativement peu reproductibles et il est nécessaire de calculer leur intervalle de confiance (Hellemans, Mortier et al. 2007).

Cette méthode est donc incontournable, mais pose des problèmes conceptuels insolubles :

- la dilution d'ADN recombinant dans du tampon de PCR ne « mime » pas de façon pertinente les paramètres de la PCR dans des échantillons biologiques.
- La dilution d'un *pool* d'échantillons est également irréaliste car d'une part, le ratio  $\frac{\text{matrice spécifique}}{\text{matrice non spécifique}}$  est constant sur la totalité de la gamme, d'autre part la concentration d'inhibiteurs diminue avec les dilutions. De plus, cette méthode ne permet pas la considération d'une grande gamme de concentrations. L'idéal serait une dilution de la matrice cible dans une quantité constante d'échantillon biologique dépourvue de cette matrice.

Cette droite présente plusieurs caractéristiques :

- **La pente**

Sa valeur est de  $-(1/(\log I + \epsilon))$ . Pente de la droite de quantification et **efficacité de la réaction PCR** sont donc étroitement liées et un tableau de valeurs peut facilement être établi (**Figure 1.9**). Ainsi, à une **efficacité de réaction** de 100% correspond une pente de valeur égale à -3,32 (**Tableau 1.2**).

- **L'efficacité de PCR  $\epsilon$**

La notion d'efficacité de réaction est le premier élément de divergence entre modèle et réalité. Souvent considérée comme constante et identique entre tous les échantillons au cours d'une analyse dans le modèle exponentiel, ce paramètre présente en réalité une importante variabilité tout au long de l'amplification, en raison de différents facteurs. Il a ainsi été observé des différences d'**efficacités** d'amplification au cours des premiers cycles de la réaction (Nogva and Rudi 2004).

Si l'on considère que la quantité de fluorescence correspondant au cycle de sortie  $S$  est toujours la même car la détection est effective pour une quantité constante de produit, l'**efficacité de PCR** peut être calculée de la manière suivante :

$$S = \text{Constante } I \times (1 + \epsilon)^n \quad (1.18)$$

où :  $\epsilon$  est l'efficacité de PCR

$n$  le nombre de cycles

En transformant par le log :

$$\log(S) = \log(C) + n \log(1 + \epsilon)$$

$$n \log(1 + \epsilon) = -\log(C) + \text{Constante } 2$$

$$n = -(1/(\log(1 + \epsilon)) \times \log(C) + \text{Constante } 3$$

La pente  $p$  de la droite **nombre de cycles =  $f(\log(\text{nombre de cycles}))$**  est elle-même fonction de l'**efficacité de PCR**.

$$p = -(1/(\log I + \epsilon))$$

$$\log(1 + \epsilon) = -(1/p)$$

$$(1 + \epsilon) = 10^{-(1/p)}$$

$$e = 10^{-(1/p)} - 1 \quad (1.19)$$

où :  $e$  est l'efficacité de la réaction

$p$  est la pente de la courbe de quantification

Les calculs précédents utilisent un seuil de fluorescence constant ( $C_p$ ), ce qui n'est pas le cas de la majorité des analyses réalisées en qPCR ( $C_T$ ). Cette approche constitue donc une source d'erreurs. D'autre part, elle n'estime qu'une **efficacité de réaction** moyenne alors qu'elle varie au cours de la réaction (Nogva and Rudi 2004).

Pente	Taux d'amplification	Efficacité
-3,32	2,00	100,0 %
-3,40	1,97	96,8 %
-3,50	1,93	93,0 %
-3,60	1,90	89,6 %
-3,70	1,86	86,3 %
-3,80	1,83	83,3 %
-3,90	1,80	80,4 %
-3,92	1,80	79,9 %

**Tableau 1.2** : Correspondances entre pente de courbe de quantification et efficacité de réaction de PCR.

Il existe une autre méthode d'estimation de l'**efficacité** basée sur une méthode de régression linéaire obtenue par le logiciel *LinReg*<sup>®</sup> (Ramakers, Ruijter et al. 2003). Cette méthode, qui est discutable pour déterminer l'**efficacité** réelle de la PCR au niveau d'un échantillon, peut être utilisée pour déterminer l'**efficacité** globale à partir d'une gamme (Cikos, Bukovska et al. 2007). Les valeurs obtenues ainsi sont sensiblement inférieures à celle déterminées à partir de la pente.

### I.3. Les directives MIQE : Informations Minimum pour la publication d'expérimentations de PCR en temps réel.

---

- *Les raisons de la publication*

Depuis son apparition, la méthode de RT-qPCR est devenue omniprésente en biologie moléculaire. Elle a tellement suscité d'engouement dans le domaine de la recherche et dans celui du diagnostic qu'il s'en est suivi un nombre considérable d'innovations technologiques, notamment en terme d'instrumentations, de kits et de réactifs. D'innombrables variations techniques et méthodologiques capables d'affecter la réalisation d'expériences de l'analyse de l'expression génique sont ainsi apparues dans de multiples domaines. Cette absence de consensus admis sur les meilleurs moyens de réaliser une analyse de qPCR et d'interpréter convenablement les résultats qui en découlent s'est inévitablement traduite par des incohérences, parfois même par certaines contradictions lors de la présentation des résultats expérimentaux. Les premières publications ont donc souvent restitué une description inadéquate des études réalisées et ont présenté des analyses de données non convenables. En effet, le manque d'informations caractérisant la plupart des documents décrivant cette technologie ne permet pas au lecteur d'évaluer de manière critique la qualité des résultats présentés. Plus précisément, des informations détaillées sur l'acquisition et la manipulation des échantillons, la qualité et l'intégrité de l'ARN, les caractéristiques des amorces et sondes utilisées, les détails de la transcription inverse, les efficacités de PCR ainsi que les paramètres d'analyse des différentes données étaient souvent omis. Très fréquemment, la normalisation des échantillons était réalisée en ne considérant qu'un seul gène de référence, sans justification adéquate. D'une manière générale, cela a rapidement conduit à des conclusions inappropriées et l'impossibilité de reproduire les études les plus pertinentes. Par la suite, la validité des résultats s'en est retrouvée très discutable et par conséquent largement critiquée par les reviewers au cours de l'évaluation des manuscrits.

En pratique, ces conséquences néfastes ont abouti à la corruption de la littérature scientifique concernée. La publication d'une étude dans la rubrique " *Révélation de l'année 2005* " du journal *Science* (Huang, Bohlenius et al. 2005) suivie de son retrait deux ans plus tard (Bohlenius, Eriksson et al. 2007) est un très bon exemple de cette situation problématique. Un autre exemple plus grave encore est l'établissement d'une corrélation abusive entre vaccin ROR (Rougeole Oreillons Rubéole) et autisme en raison d'une mauvaise mise en œuvre de la technique de PCR quantitative (Bustin 2008). Les scientifiques les mieux informés et les plus

spécialisés dans cette technologie ont ainsi pris conscience de la faiblesse des normes concernant la conceptualisation, l'exécution et l'interprétation des analyses obtenues en RT-qPCR. Très logiquement, ils ont rapidement établi une liste de recommandations nécessaires à la réalisation et la publication de ce type d'expérimentations (Bustin, Benes et al. 2009).

- **Les recommandations proposées**

Directement inspirées des recommandations établies pour les *microarrays* ainsi que pour d'autres analyses de protéomique sous l'abréviation " MIBBI " (*Minimum Information for Biological and Biomedical Investigations*, <http://www.mibbi.org>, (Taylor, Field et al. 2008)), les directives décrites dans MIQE - *Minimum Information for qPCR Experiments* (Bustin, Benes et al. 2009) constituent la norme standard concernant la présentation des données obtenues en qPCR. Rédigées par un consortium international de scientifiques académiques, elles définissent point par point les prérequis indispensables permettant la réalisation convenable des expérimentations et l'interprétation appropriée des données obtenues.

Ces directives demandent ainsi la précision de toutes les conditions expérimentales pouvant être pertinentes au regard des reviewers. Les différentes caractéristiques de chaque analyse réalisée doivent être clairement décrites, cela afin de permettre une meilleure évaluation de la validité des protocoles utilisés. La divulgation complète de tous les réactifs utilisés est exigée, tout comme les séquences travaillées et les méthodes d'analyse pratiquées, rendant ainsi les expérimentations beaucoup plus reproductibles. Considérées comme « essentielles » ou seulement « souhaitables », ces directives peuvent être retranscrites de manière plus exhaustive dans le tableau récapitulatif suivant :

## Checklist des Informations Essentielles en vue de Publications d'Expérience en PCR en Temps Réel

DESIGN EXPERIMENTAL		OLIGONUCLEOTIDES UTILISES POUR LA qPCR	
Définition des groupes expérimentaux et des groupes contrôles	essentielle	Séquences des amorces	essentielle
Nombre d'individus dans chaque groupe	essentielle	Numéro d'identification RTPrimerDB	souhaitable
Analyse apportée par le centre d'expertise du laboratoire	souhaitable	Séquences des sondes utilisées	souhaitable <sup>b</sup>
Contributions d'autres auteurs	souhaitable	Localisation et identité des éventuelles modifications apportées	essentielle
ECHANTILLON		Fabriquant des oligonucléotides	souhaitable
Description	essentielle	Méthode de purification	souhaitable
Volume / masse de l'échantillon analysé	souhaitable	PROTOCOLE DE qPCR	
Microdissection ou macrodissection	essentielle	Ensemble des conditions de la réaction	essentielle
Procédure de traitement	essentielle	Volume réactionnel et quantités d'ADN complémentaire / ADN	essentielle
Si congélation, comment et avec quelle rapidité ?	essentielle	Concentrations en amorces, sonde, Mg <sup>2+</sup> et dNTP	essentielle
Si fixation, comment et avec quelle rapidité ?	essentielle	Identification de l'ADN polymérase utilisée	essentielle
Conditions de stockage de l'échantillon et période de stockage (spécialement pour les échantillons traités au FFPE)	essentielle	Nom du kit et du tampon utilisés, nom du fabriquant	essentielle
EXTRACTION DES ACIDES NUCLEIQUES		Composition chimique exacte du tampon utilisé	souhaitable
Procédure et / ou instrumentation	essentielle	Additifs (SybrGreen I, DMSO, etc...)	essentielle
Nom du kit utilisé, détails et modifications éventuellement apportées	essentielle	Fabriquant des plaques / tubes et numéros de catalogue	souhaitable
Origine des réactifs additionnels utilisés	souhaitable	Paramètres complets du programme thermique	essentielle
Détails du traitement à la DNase ou la RNase	essentielle	Mise au point de la réaction	souhaitable
Évaluation de la contamination par l'ADN ou l'ARN	essentielle	Fabriquant de l'instrument de qPCR	essentielle
Quantification de l'acide nucléique	essentielle	VALIDATION DE LA qPCR	
Instrumentation et méthode	essentielle	Illustrations de l'optimisation (à partir de gradients)	souhaitable
Pureté (A <sub>260</sub> / A <sub>280</sub> )	souhaitable	Spécificité (gels électrophorétiques, séquence, courbe de fusion ou digestions enzymatiques)	essentielle
Rendement	souhaitable	Pour le SybrGreen I, C <sub>q</sub> du NTC	essentielle
Intégrité de l'ARN : méthode et instrumentation	essentielle	Droite de calibration avec pentes et ordonnée à l'origine	essentielle
RIN / RQI ou C <sub>q</sub> en 3'OH et 5'P des transcripts	essentielle	Efficacité PCR calculée à partir de la pente	essentielle
Investigations électrophorétiques	souhaitable	Intervalles de confiance pour l'efficacité PCR ou SE	souhaitable
Analyse de présence d'inhibiteurs (Dilutions, spike, ou autres)	essentielle	R <sup>2</sup> ou droite de calibration	essentielle
TRANSCRIPTION INVERSE		Plage de linéarité dynamique	essentielle
Conditions de la réaction complète	essentielle	Variation du C <sub>q</sub> à la LOD	essentielle
Quantité d'ARN et volume réactionnel	essentielle	Intervalles de confiance à travers la plage	souhaitable
Type d'amorçage utilisé (si utilisation de GSP) et concentration	essentielle	Mise en évidence de la LOD	essentielle
Type de reverse transcriptase utilisée et concentration	essentielle	Si PCR multiplex, efficacité et LOD pour chaque analyse	essentielle
Temps et température	essentielle	ANALYSE DES DONNEES	
Fabriquant des réactifs utilisés et numéros de catalogue	souhaitable	Programme d'analyse qPCR (provenance et version)	essentielle
C <sub>q</sub> avec et sans transcription inverse	souhaitable <sup>a</sup>	Méthode de détermination du C <sub>q</sub>	essentielle
Condition de stockage de l'ADN complémentaire	souhaitable	Identification d'outlier	essentielle
INFORMATIONS CONCERNANT LA CIBLE DETECTEE PAR qPCR		Résultats des NTCs	essentielle
Symbole du gène cible	essentielle	Justification du nombre et du choix des gènes de référence	essentielle
Numéro d'accession de la séquence travaillée	essentielle	Description de la méthode de normalisation	essentielle
Localisation de l'amplicon	souhaitable	Nombre et concordance des répliques biologiques	souhaitable
Spécificité <i>in silico</i> (BLAST ou autre types d'analyses)	essentielle	Nombre et états (transcription inverse ou qPCR) des répliques techniques	essentielle
Pseudogènes, rétro-pseudogènes ou gènes homologues ?	souhaitable	Répétabilité (variation intra-essai)	essentielle
Alignement de séquences	souhaitable	Reproductibilité (variation inter-essai, CV)	souhaitable
Analyse de la structure secondaire de l'amplicon	souhaitable	Puissance d'analyse	souhaitable
Localisation de chaque amorce au niveau des exons / introns	essentielle	Méthodes d'analyses statistiques utilisées pour la significativité des résultats	essentielle
Quelles isoformes épissées sont ciblées ?	essentielle	Logiciels (provenance, version)	essentielle
		C <sub>q</sub> ou soumission des données brutes en RDML	souhaitable

Toutes les informations essentielles doivent être soumises avec le manuscrit. Les informations souhaitables devraient être soumises si disponibles. Notifier si les amorces proviennent de RTPrimerDB, les informations sur la cible qPCR, les oligonucléotides, les protocoles, et si la validation est disponible à partir de cette source.

<sup>a</sup> L'évaluation de l'absence d'ADN par un contrôle sans transcription inverse est une information essentielle lors de la première extraction de l'ARN. Une fois que l'échantillon a été validé comme étant dépourvu de toute contamination ADN, la réalisation de ce type de contrôle devient souhaitable et n'est plus indispensable.

<sup>b</sup> La divulgation de la séquence de la sonde est fortement souhaitable. Cependant, parce que certains distributeurs commercialisant des détections par sonde déjà prédéfinies ne fournissent pas toujours cette information, il ne peut être exigé que cette information soit essentielle. L'utilisation de telles détections n'est donc pas encouragée.

Ces normes lèvent donc dans une totale transparence toute ambiguïté sur la fiabilité des résultats. Elles sont par conséquent fortement recommandées lors de la soumission initiale de tout manuscrit décrivant cette technique (Taylor, Wakem et al. 2010). La *check-list* proposée est également disponible à l'adresse <http://www.RDML.org/MIQE>. Parallèlement, le consortium a développé un format universel pour l'échange de données de qPCR, indépendant de l'instrumentation utilisée, le format *Real-time PCR Data Markup Language* (RDML) dans le but de ré-analyser plus facilement des données soumises à publication. Ces directives peuvent être retrouvées aux adresses <http://www.RDML.org> et <http://www.RDML.org/MIQE> ainsi que dans les publications respectives de MIQE (Bustin, Benes et al. 2009 ; Bustin, Beaulieu et al. 2010 ; Bustin, Benes et al. 2011) et RDML (Lefever, Hellemans et al. 2009 ; Lefever, Vandesompele et al. 2009).

# Objectifs Scientifiques

---

L'objectif principal de ce travail de thèse est de contribuer à l'amélioration de la sensibilité et de la précision de la quantification des acides nucléiques par la technique de RT-qPCR.

Cette technique est théoriquement capable de détecter et de mesurer des quantités infimes d'acides nucléiques au niveau d'une grande variété d'échantillons. Elle constitue de ce fait la méthode d'excellence aussi bien pour le diagnostic moléculaire que pour les sciences de la vie ou la médecine (Kubista, Andrade et al. 2006). Sa simplicité conceptuelle et sa facilité d'utilisation, combinées avec sa rapidité d'exécution, sa sensibilité et sa spécificité d'analyse en ont fait la pierre angulaire de la quantification des acides nucléiques. Cependant, des auteurs ont très précocement attiré l'attention sur les limites de cette technique, en particulier sur les biais qui pouvaient altérer la qualité des mesures (Bustin 2002).

C'est dans ce contexte que nous avons placé notre travail. Nous avons réalisé 3 études indépendantes. La première avait pour but d'augmenter la précision de la quantification des ARNm en diminuant la variabilité inter-échantillons. La deuxième étude a consisté à augmenter la spécificité des amorces en substituant des analogues de nucléotides (*Locked Nucleic Acids, LNA*) au niveau de la partie 5', de façon à augmenter la sensibilité. Le dernier volet, qui a été initialement développé en réponse à des impératifs de l'étude précédente, a permis de mettre au point un outil capable de déterminer les paramètres thermodynamiques de l'hybridation directement dans le milieu réactionnel. Il devrait permettre de mieux comprendre les bases de la variabilité inter-échantillons de l'hybridation.

## I. Diminution de la variabilité inter-échantillons

La standardisation des différentes étapes de la technique de RT-qPCR est un prérequis indispensable pour une diminution de la variabilité inter-échantillons. En particulier, depuis les travaux du groupe de Mikael KUBISTA (Stahlberg, Hakansson et al. 2004 ; Stahlberg, Kubista et al. 2004) l'utilisation d'une quantité constante d'ARN pour réaliser la RT est un dogme universellement observé. En effet, le rendement de la transcriptase inverse dépend de la quantité absolue d'ARN utilisé. L'utilisation de l'ARNm constitue la seule exception admise. En effet, leur quantification qui n'est réalisable que depuis l'apparition de la micro-fluidique, reste relativement peu précise. D'autre part, comme ils ne constituent qu'une faible fraction des ARN

**totaux** (1 à 3%), d'importantes variations relatives d'**ARNm** correspondent à des variations absolues négligeables lorsque l'**ARN total** est considéré.

L'arrêt de la commercialisation d'un kit d'extraction d'**ARNm** automatisé que nous utilisions nous a obligés à nous servir d'un kit de purification d'**ARN total**, c'est-à-dire de passer d'une **RT** réalisée avec un volume constant d'extrait d'**ARNm**, à une quantité constante d'**ARN total**, i.e. à un volume variable d'extrait d'**ARN** en fonction du rendement d'extraction. Cette modification s'est traduite par une augmentation de la variabilité de nos résultats et une dégradation de la **normalisation** que nous avons objectivée par l'augmentation des indices  $V_{n/n+1}$  du logiciel *geNorm* (Vandesompele, De Preter et al. 2002). Nous avons émis les hypothèses suivantes :

- il existerait des inhibiteurs de **RT-qPCR** dont la quantité ne serait pas corrélée à la quantité d'**ARN** extrait mais au volume d'échantillon utilisé (inhibiteurs volume-dépendants).
- une utilisation d'un volume constant d'extrait d'**ARN** au cours de l'étape de **RT**, en compensant les différences de quantité d'**ARN** inter-échantillons par de l'**ARN de transfert (ARNt)** de levure, devrait diminuer la variabilité inter-échantillons.

Nous avons vérifié ces hypothèses en comparant la variabilité de la **RT-qPCR** réalisée à « quantité constante » ou à « volume constant » sur des répliques techniques ou biologiques provenant de différents tissus et extraits avec différents protocoles.

## II. Optimisation de la spécificité d'hybridation des amorces de PCR

La spécificité de la **PCR** est conférée par le caractère ciblé de l'amplification de la séquence choisie. Si cette spécificité est vérifiée lorsque la cible se trouve en quantité suffisante dans le mélange réactionnel, elle est souvent mise en défaut lorsque les quantités de séquences cibles diminuent. Cette perte de spécificité est indiscutable aux limites de détection et peut être favorisée en présence d'inhibiteurs de **PCR**.

L'augmentation de la **température d'hybridation**, un temps d'**hybridation** plus court (Cha and Thilly 1993) ou une longueur plus importante des séquences d'**amorces** (Wu, Ugozzoli et al. 1991) permettent une amélioration significative de la spécificité obtenue en **PCR**. Cependant, cette augmentation de **stringence** de l'**hybridation** ne permet toujours pas de résoudre ce

problème. En effet, il existe un équilibre entre spécificité et **efficacité** : l'augmentation de la **stringence** de l'**hybridation** est souvent délétère pour l'**efficacité** de la qPCR et par conséquent pour la qualité de la quantification souhaitée.

Pour résoudre ce dilemme, nous proposons une méthode permettant d'augmenter la spécificité sans retentissement sur l'**efficacité** de la qPCR. Notre but n'est pas de diminuer l'**hybridation** non spécifique qui nous semble inéluctable lorsque la séquence cible diminue, mais plutôt d'empêcher que cette **hybridation** ne conduise à une amplification non spécifique. L'incorporation de *Locked Nucleic Acids* (Koshkin, Singh et al. 1998 ; Obika, Nanbu et al. 1998), au niveau des **amorces** de PCR permet d'augmenter la **stringence** de l'**hybridation** sans diminuer l'**efficacité** de la qPCR. Notre laboratoire a montré qu'en substituant des bases situées au niveau de la **partie 5'** de l'**oligonucléotide** par des LNA, on favorisait une **hybridation** non spécifique au niveau de cette zone, tout en diminuant l'amplification non spécifique qui nécessite l'**hybridation** de la **partie 3'** de l'**amorce** (Malgoyre, Banzet et al. 2007). Cette augmentation de spécificité de la PCR, observée au cours de la quantification de **transcrits** faiblement exprimés, a été obtenue pour un couple d'**amorces** présentant 4 substitutions en 5', choisies afin de maximiser l'augmentation du  $T_m$  prédit *in silico*. Cependant, ces modifications n'ont permis qu'un gain de **température d'hybridation** très inférieur à celui attendu au regard des prédictions thermodynamiques.

Le but de mon travail a donc été de déterminer des règles de substitution des bases 5' des **amorces** par des LNA, permettant une amélioration de la quantification des cibles rares. L'hypothèse de départ guidant les substitutions était que, dans l'étude de MALGOYRE *et al.*, un nombre trop important de substitutions par LNA était responsable d'une faible augmentation obtenue de la **température d'hybridation**. Nous avons par conséquent testé des substitutions de seulement 2 nucléotides, choisis en fonction de leur position ou de leur nature. Ce travail a nécessité la mise au point d'un modèle pour quantifier la spécificité d'une part et une méthode de détermination précise du  $T_m$  dans les conditions de PCR d'autre part. La mise au point de cette méthode nous a conduit à développer le dernier volet de ce travail de thèse.

### III. Synchronisation de l'hybridation des amorces en PCR

L'efficacité de la qPCR peut être définie par le pourcentage moyen de cibles amplifiées à chaque cycle. Si les amorces sont équilibrées, elle correspond au pourcentage d'amorces hybridées de façon spécifique à la fin de l'étape d'hybridation. Si le pourcentage d'hybridation des deux amorces est différent, l'efficacité est donnée par le pourcentage le plus faible.

Le paramètre de prédiction le plus utilisé concernant l'hybridation des amorces est celui de la température de fusion ( $T_m$ ), température à laquelle la moitié des amorces est hybridée sur la cible à l'équilibre. Ce paramètre dépend à la fois de l'oligonucléotide (séquence, taille, pourcentage de bases G-C) et du milieu réactionnel (concentrations en oligonucléotides, en  $Mg^{2+}$ , en  $Na^+$ , en additifs etc. ...). Il est généralement déterminé par des méthodes thermodynamiques prédictives basées sur la séquence et la concentration en cations. Cependant, la valeur de ces prédictions est limitée car d'une part, elles sont optimisées pour des conditions salines (1M  $Na^+$ ) qui ne correspondent pas au milieu réactionnel de la PCR, et d'autre part, il existe de nombreuses disparités entre les différentes méthodes (Panjkovich and Melo 2005). Afin d'évaluer le  $T_m$  des amorces substituées par des LNA de façon précise et dans les conditions réelles de la PCR, nous avons utilisé une mesure directe du  $T_m$  par fluorescence en hybridant la séquence de l'amorce contre sa séquence complémentaire. De concept relativement simple, cette méthode a été précédemment proposée (Teo, Choi et al. 2002) sans validation. Nous avons donc validé dans un premier temps cette méthode par comparaison à une technique de référence, la nano Calorimétrie Différentielle à Balayage (nanoDSC).

Dans un second temps, nous avons émis l'hypothèse que la mesure de la modification du  $T_m$  d'une amorce en fonction de sa concentration ou de la concentration de  $Mg^{++}$  pourrait permettre de guider l'optimisation des réactions de PCR. Cependant, les faibles variations de  $T_m$  observées pour des modifications ayant un impact significatif sur la qPCR nous a obligé à rechercher un paramètre plus pertinent. Bien que le  $T_m$  (température à laquelle 50% des amorces sont hybridées) soit le paramètre classiquement utilisé, le paramètre à prendre en compte est le pourcentage d'amorces hybridées à la température d'hybridation (SantaLucia 2007). L'estimation de ce paramètre nécessite de déterminer les variations d'enthalpie ( $\Delta H$ ) et d'entropie ( $\Delta S$ ). Théoriquement, ces paramètres peuvent être calculés à partir des valeurs du  $T_m$  obtenues pour différentes concentrations d'amorces. Nous avons alors recherché les conditions de mesures du  $T_m$  permettant une estimation précise de ces paramètres thermodynamiques.

# Matériels & Méthodes

---

### 3.1. Échantillons biologiques

---

En raison de la variabilité biologique, de nombreux échantillons analysés au cours d'une analyse de **RT-qPCR** présentent fréquemment des concentrations variables en inhibiteurs endogènes, souvent co-purifiés au cours de l'étape d'extraction et donc directement liés au volume d'extrait d'ARN. Afin d'évaluer l'impact de la variabilité inter-échantillons et d'identifier à quel(s) niveau(x) l'inhibition de ce type d'inhibiteurs est la plus marquée, nous avons choisi de travailler à partir de deux types d'échantillons bien différents :

- *les répliques biologiques* : obtenus à partir du prélèvement de tissu splénique provenant de 12 animaux, ces échantillons indépendants présentent des concentrations variables d'inhibiteurs endogènes échantillon-dépendant. Leur utilisation va nous permettre de mettre en évidence une variabilité inter-échantillons récurrente, problématique dans les étapes de **standardisation** et de **normalisation** des données d'**expression** génique (Willems, Leyns et al. 2008).

- *les répliques techniques* : obtenus à partir du broyat d'un seul prélèvement original (1 muscle squelettique ou 1 cervelet), ces 6 répliques présentent tous le même niveau d'inhibiteurs endogènes échantillon-dépendants. Leur utilisation va nous permettre de démontrer qu'une **transcription inverse** à volume constant d'ARN permet dans ce cas de figure particulier, une **normalisation** plus efficace de la variabilité inter-échantillon concernant les concentrations en inhibiteurs endogènes ou exogènes volume-dépendants.

### 3.2. Purification des ARN totaux

---

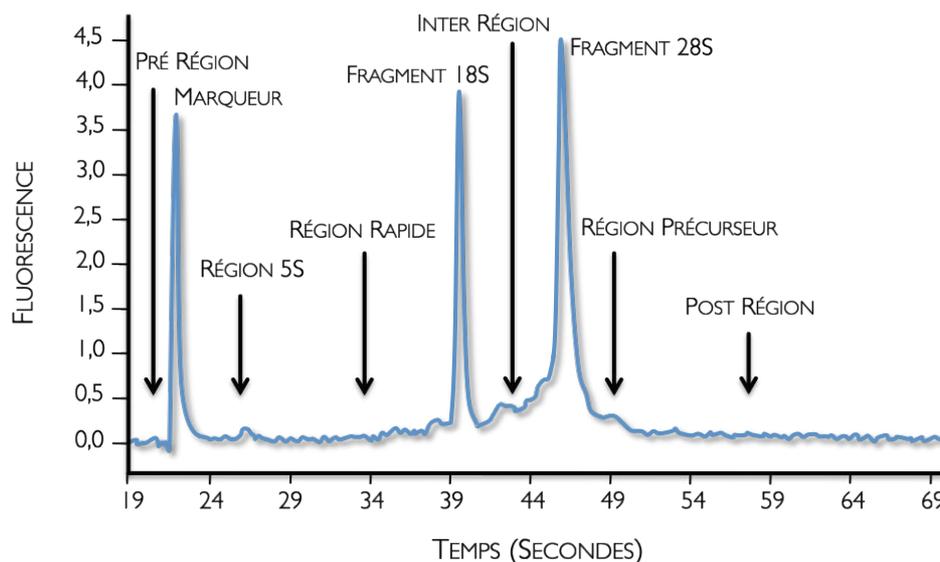
#### 3.2.1. Extraction des ARN totaux

Chaque échantillon tissulaire a été prélevé sur animal anesthésié puis directement conservé dans du **RNAlater**<sup>®</sup> (Ambion, Austin, TX, USA) à -20°C en vue d'analyses de biologie moléculaire. 12 rates (8 mg), 1 cervelet (48 mg) et un muscle soléaire (48 mg) sont ensuite homogénéisés dans 200 µL de tampon de lyse composé d'agents chaotropiques inhibant les ribonucléases (thiocyanate de guanidinium (McGookin 1984), GITC, 50 mg/mL, en tampon RLT, Qiagen, Courtaboeuf, France) afin de préserver l'intégrité des **ARNm** et d'agents

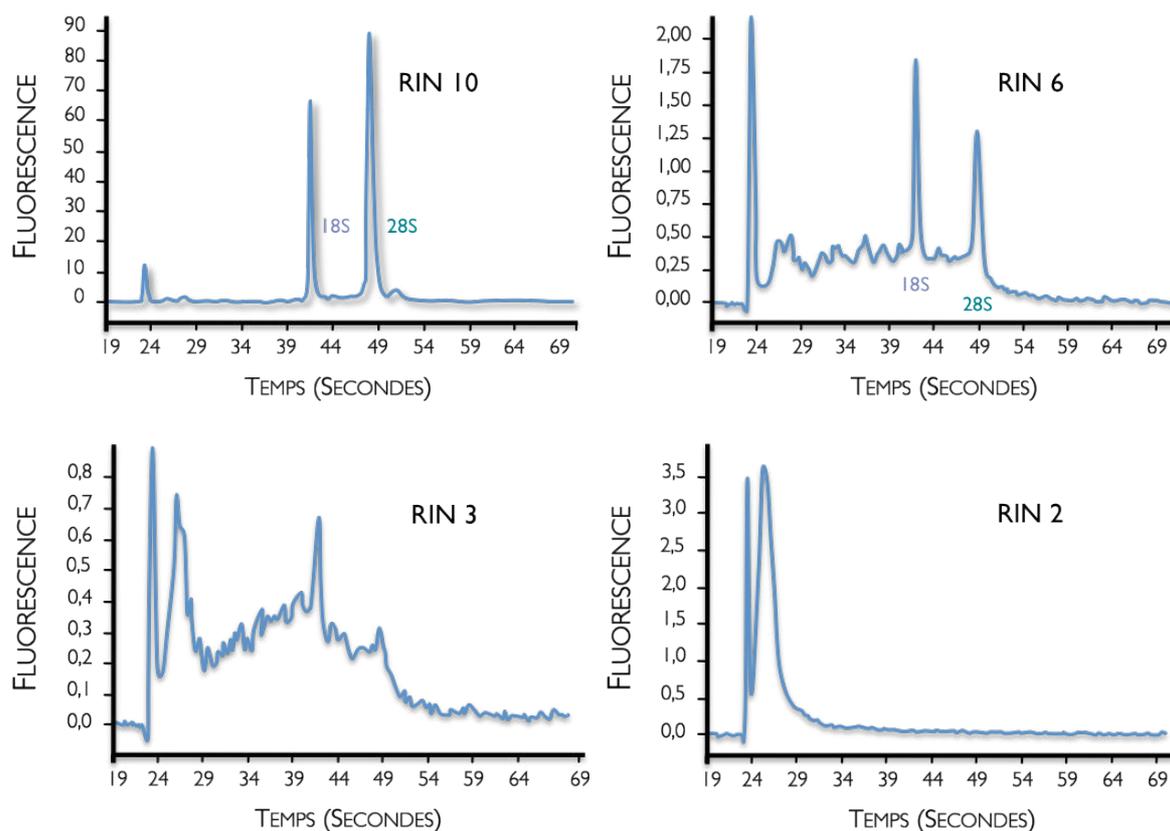
dénaturants ( $\beta$ -mercaptoéthanol, 2-Hydroxy-1-éthanethiol, Sigma-Aldrich, Buchs, SG, Suisse) en utilisant un broyeur à billes MM300 (Fréquence d'utilisation : 30 Hz, 2 x 2 min, 1 bille d'acier de 5 mm par tube, Rescht, Haan, Allemagne). Pour le cervelet et le soléaire, les lysats ont été aliquotés en six tubes et le volume a été ajusté à 600  $\mu$ L par addition de tampon de lyse RLT. L'ARN total a ensuite été isolé en utilisant une extraction au phénol-chloroforme (Chomczynski and Sacchi 1987 ; Haimov-Kochman, Fisher et al. 2006) sur instrument QIAcube (Qiagen, Courtaboeuf, France), selon les instructions du fabricant, en utilisant deux purifications différentes : une extraction basée sur l'utilisation de colonne de silice avec une étape à la DNase (*RNeasy Mini Kit*, Qiagen, Courtaboeuf, France) et une extraction organique suivie d'un *clean-up* sur colonne de silice (*RNeasy Lipid Tissue Mini Kit*, Qiagen, Courtaboeuf, France). Les extraits d'ARN ont été élués par 50  $\mu$ L d'eau exempte de RNase.

### 3.2.2. Contrôle de la qualité de l'ARN

L'intégrité des ARN est évaluée en utilisant le kit RNA 6000 Nano LabChip sur Bioanalyseur 2100 (Agilent Technologies, Santa Clara, CA, USA) par mesure du *RIN* (*RNA Integrity Number*) (Schroeder, Mueller et al. 2006). Comme la dégradation de l'ARN est un processus graduel, il existe en parallèle une diminution du ratio *ARN ribosomal* 18S / *ARN ribosomal* 28S et une augmentation du signal de base entre les deux pics ribosomaux (**Figure 3.1**). Développé afin de standardiser le contrôle de la qualité de l'ARN, le *RIN* prend en compte la totalité d'un tracé électrophorétique d'un échantillon d'ARN. L'algorithme du logiciel de la machine restitue donc une valeur permettant la classification de l'ARN total eucaryotique, basée sur un système de numérotation de 1 à 10, 1 étant le profil le plus dégradé et 10 étant le profil d'un échantillon intact (**Figure 3.2**).



**Figure 3.1 :** Électrophorogramme détaillant les régions indicatives de la qualité de l'ARN. D'après "RNA Integrity Number (RIN) - Standardization of RNA Quality Control", Agilent Application Note, Publication Number 5989-1165EN, 2004.



**Figure 3.2 :** Exemples de différents électrophorogrammes d'échantillon évalué par le RIN. Les échantillons varient d'un état intact (RIN 10) à un état dégradé (RIN 2). D'après "RNA Integrity Number (RIN) - Standardization of RNA Quality Control", Agilent Application Note, Publication Number 5989-1165EN, 2004.

### 3.3. Transcription inverse des ARN messagers

#### 3.3.1. Transcription inverse classique

La synthèse d'ADNc a été réalisée en utilisant le kit de transcription inverse d'Eurogentec (*Reverse Transcriptase Core Kit*, Saraing, Belgique), à partir de 400 ng d'ARN préalablement dénaturés en bain marie pendant 2 min à 70°C et selon la composition suivante :

	Volume (µL)	Concentration finale
Eau dépourvue de RNase	4,1	
Tampon réactionnel 10X	2,0	1X
Solution de MgCl <sub>2</sub> 25 mM	4,0	5 mM
Solution de dNTP 2,5 mM	4,0	500µM pour chaque dNTP
Amorces oligo dT <sub>15</sub>	1,0	2,5 µM
Inhibiteur de RNase	0,4	0,4 U/µL
Transcriptase inverse EuroScript	0,5	1,25 U/µL
Échantillon	4,0	100 ng/µL

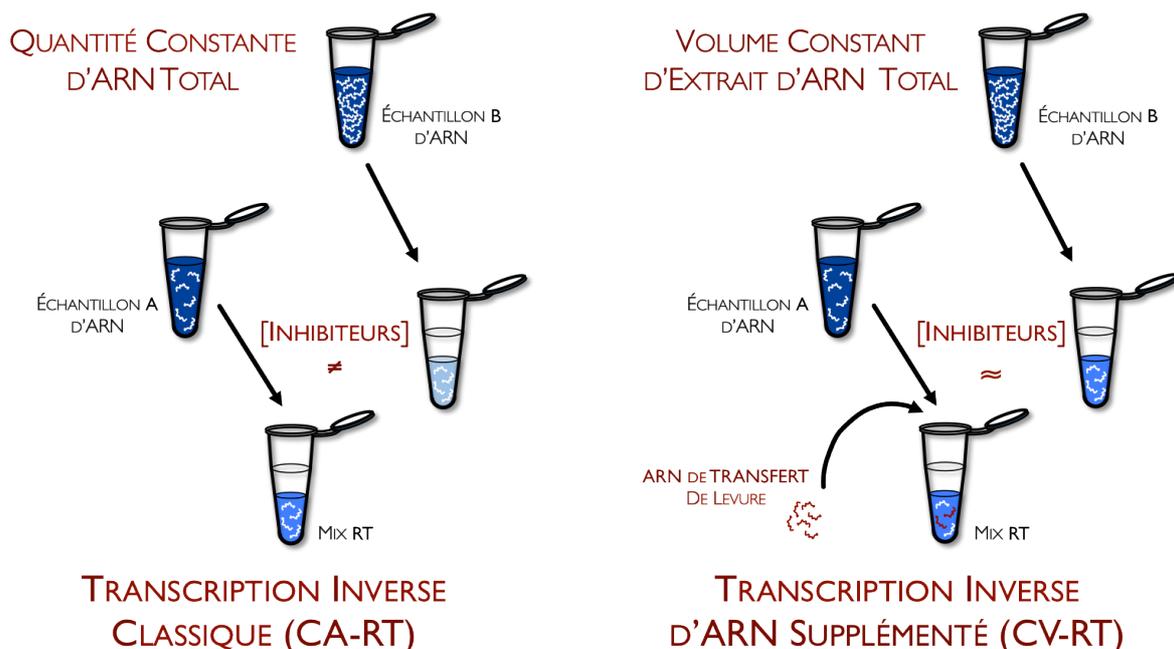
**Tableau 3.1 :** Composition du mix réactionnel pour la synthèse du premier brin d'ADNc d'ARN messagers à partir de 400 ng d'ARN totaux dénaturés, volume final de 20 µL (Eurogentec - Reverse Transcriptase Core Kit).

Le mélange réactionnel est ensuite incubé au bain marie à la température de 25°C pendant 10 min (stabilisation thermique et hybridation des amorces oligo dT<sub>15</sub>) avant d'être placé dans un autre bain marie à la température de 48°C pendant 30 min (synthèse de l'ADNc). Une dénaturation thermique est effectuée à 95°C pendant 5 min afin de détruire l'ARN résiduel. L'ADNc est enfin placé dans la glace pendant 5 min avant d'être stocké à -80°C.

### 3.3.2. Transcription inverse à volume constant d'extrait d'ARN

La synthèse d'ADNc a été réalisée en utilisant le kit de transcription inverse d'Eurogentec (*Reverse Transcriptase Core Kit*, Saraing, Belgique) à partir de 200 ng d'ARN en utilisant 2,5  $\mu$ M d'amorces oligo dT<sub>15</sub> et un inhibiteur de RNase dans un volume final de 10  $\mu$ L, conformément aux instructions du fabricant. La RT a été effectuée de deux manières différentes :

- soit en utilisant 200 ng d'ARN total pour chaque échantillon, on parlera alors de CA-RT.
- soit à partir d'un volume constant d'extrait d'ARN total pour chaque échantillon. Pour cela, un volume d'extrait est déterminé à partir de l'échantillon d'ARN le plus concentré pour obtenir la quantité de 200 ng d'ARN total. Ce volume est ensuite prélevé pour chaque échantillon afin de réaliser la transcription inverse. Afin de conserver une quantité constante d'ARN au cours de la réaction pour chaque échantillon, les moins concentrés sont ajustés à 200 ng par addition d'un volume variable d'ARN de transfert de levure (Ambion, Austin, TX, USA). On parlera alors dans ce cas-là de CV-RT (**Figure 3.3**). À chaque fois que la transcription inverse est réalisée, un mélange unique d'eau, d'amorces, de dNTP, de magnésium et d'enzyme est utilisé pour chaque groupe d'échantillons (rates, muscle ou cervelet) et de RT. Chaque groupe (un tissu + une méthode d'extraction) correspond à un set de RT. Enfin, afin de minimiser la variabilité rencontrée au cours des pipetages de faibles volumes, les extraits d'ARN purs ont été dilués de manière à n'utiliser qu'un volume d'essai de 4  $\mu$ L minimum.



**Figure 3.3 :** Transcription inverse à quantité constante d'ARN (CA-RT) : la quantité constante d'ARN prélevée au niveau de deux échantillons de concentrations différentes implique une variabilité de la concentration en inhibiteurs. La Reverse Transcription à Volume Constant d'extrait d'ARN (CV-RT) permet de normaliser les inhibiteurs, la quantité constante d'ARN étant préservée par supplémentation d'ARNt de levure pour les échantillons les moins concentrés.

### 3.3.3. Évaluation de l'efficacité relative de RT par l'analyse 5':3'

Ce test proposé par Tania NOLAN en 2006 (Nolan, Hands et al. 2006) est initialement conçu pour évaluer l'intégrité des ARNm en amplifiant les extrémités 5'P et 3'OH d'un gène. En prenant pour cible un gène de référence considéré comme représentatif de l'intégrité de tous les ARNm dans un échantillon d'ARN donné, cette analyse fournit une mesure raisonnable de la dégradation des transcrits d'intérêt. De plus, elle permet d'obtenir des résultats indépendants de l'intégrité de l'ARN ribosomal. Les ARNm ne se dégradent pas tous de la même manière et à la même vitesse, il peut être nécessaire de concevoir des tests similaires pour des cibles spécifiques. Au cours de notre étude, nous avons choisi le gène de la protéine ribosomale large P0 (*ribosomal protein, large, P0* - Rplp0, NM\_022402) du rat. Deux couples d'amorces de PCR sont dessinés, l'un s'hybridant au niveau de l'extrémité 5'P de la séquence de l'ARNm, l'autre au niveau de l'extrémité 3'OH (Tableaux 3.9). Les conditions et les performances de ces deux réactions sont décrites plus loin (Tableaux 3.10). La réalisation d'un ratio de quantités détectées à ces deux endroits du gène reflète ainsi le succès relatif que la transcriptase inverse

aura rencontré tout au long du **transcrit** à partir de l'**amorce** oligo dT<sub>15</sub> au cours de sa synthèse d'**ADNc**. De toute évidence, la progression de la **transcriptase inverse** jusqu'à l'**extrémité 5'P** du gène dépend directement de l'intégrité des **ARNm**, l'enzyme étant incapable de l'atteindre si la molécule d'**ARNm** est dégradée (Cheng, Chen et al. 1995). Par conséquent, un ratio 5':3' proche de 1 indique une intégrité élevée et plus ce ratio est proche de zéro, plus la dégradation est importante.

En travaillant sur un même ARN, ce test nous permet d'évaluer l'efficacité de la **transcription inverse**. Cela n'est possible que lorsqu'un seul extrait d'ARN est utilisé, mais dans des conditions de réaction différentes. Ainsi, pour chaque échantillon, le ratio 5':3' CA-RT / 5':3' CV-RT permet d'évaluer les efficacités relatives de **transcription inverse** entre une RT à quantité constante d'ARN et une RT à volume constant d'extrait ARN additionné d'**ARN de transfert** de levure. Ces efficacités relatives sont calculées de la manière suivante :

$$E_{RT \text{ éch.A}} = \frac{\varepsilon_{3'} \cdot \text{Moyenne}(C_q \text{ échantillons}) - C_q \text{ éch.A}}{\varepsilon_{5'}} \quad (3.1)$$

---

## 3.4. Amplification de l'ADN par Polymérisation en Chaîne

---

### 3.4.1. Design des amorces de qPCR

Le principal objectif de notre *design* est d'empêcher le phénomène d'*hybridation* non spécifique qui survient inéluctablement au cours de la *PCR*, plus particulièrement lorsque de faibles concentrations de séquences cibles doivent être détectées. Notre stratégie de dessin d'*amorce* est donc définie à partir du fait qu'au cours de la phase de polymérisation, seules les *parties 3'OH* hybridées des *amorces* vont conduire à l'amplification d'un produit, qu'elles se soient fixées de manière spécifique ou non spécifique.

Ainsi, en forçant l'*hybridation* non spécifique à se réaliser plutôt dans la *partie 5'P* de l'*amorce*, nous espérons diminuer de manière significative l'amplification de produits non spécifiques en laissant à l'*amorce* la possibilité de se déshybrider en cas d'*hybridations* non spécifiques. Les derniers travaux du laboratoire, publiés par MALGOYRE *et al.* (Malgoyre, Banzet *et al.* 2007) et point de départ de cette étude, démontrent bien que l'incorporation de monomères *LNA* au niveau de la *partie 5'P* des *amorces* améliore la spécificité de détection dans la quantification des *ARNm* faiblement exprimés. Cependant, ils mettent aussi en évidence que l'augmentation de la *température d'hybridation* reste inférieure aux prédictions thermodynamiques, malgré l'incorporation de 4 monomères *LNA* au niveau des deux *amorces* utilisées. À la lumière des travaux de David LATORRA (Latorra, Arar *et al.* 2003) qui suggèrent que trop de substitutions de monomères *LNA* dans la séquence de l'*amorce* apparaît préjudiciable d'une part et que leur positionnement représente une importante considération d'autre part, nous choisissons de nous limiter à seulement 2 monomères. Le *design* est réalisé en prenant en considération les résultats thermodynamiques obtenus par Patricia McTIGUE en 2004 (McTigue, Peterson *et al.* 2004). À la manière de Joshua LEVIN (Levin, Fiala *et al.* 2006), nous adoptons ici 3 différentes façons d'incorporer ces monomères : (a) espacés au niveau de l'*extrémité 5'P* ; (b) rapprochés au niveau de l'*extrémité 5'P* et (c) rapprochés en milieu de séquence.

Plus schématiquement, notre *design* peut se décomposer en 3 séries distinctes :

- une première série correspondant au premier couple synthétisé pour l'étude de MALGOYRE *et al.* c'est-à-dire 4 substitutions de monomères LNA réparties uniformément sur l'ensemble de la séquence de l'*amorces*, le couple d'*amorces* ADN servant dans un premier temps de couple référence (RIL6 F0-R0).
- une deuxième série où le critère considéré est le positionnement de la base substituée, comme décrit précédemment.
- une troisième série où le critère considéré est la nature de la base substituée, un bénéfice plus important étant obtenu par la substitution de bases pyrimidiques (d'après des informations rapportées du 4<sup>ème</sup> Congrès International de qPCR " *Diagnostics & Molecular Markers* " ; Freising-Weihenstephan, Allemagne, 2009 et d'après McTIGUE *et al.*, 2004).

La conception des *amorces* est dans un premier temps réalisée à l'aide du logiciel *MacVector* (Accelrys, San Diego, CA, USA). La *température de fusion* de chaque *amorce* est ensuite optimisée avec le logiciel d'*Exiqon* (*Exiqon Tm prediction version 1.1*). La synthèse est effectuée par la société Eurogentec (Saraing, Belgique) après purification par Chromatographie en phase Liquide à Haute Performance, HPLC. Les caractéristiques des séquences des *amorces sens* (*forward*, notées *F*) et *antisens* (*reverse*, notées *R*) utilisées au cours de nos expérimentations sont récapitulées au niveau du **Tableau 3.2.** suivant :

**Tableau 3.2 :** Caractéristiques des oligonucléotides modifiés par LNA spécifiques de l'interleukine 6, IL-6 (GenBank NM\_012589), utilisés pour l'étude de l'influence des LNA sur la sensibilité et la spécificité de la qPCR. Ces oligonucléotides sont synthétisés puis purifiés par Chromatographie en phase Liquide à Haute Performance, HPLC (Eurogentec, Saraing, Belgique). Les substitutions de LNA sont marquées en plus grande lettre. Les séquences : RIL6 FX, 300-322 ; RIL6 RX, 428-406. \*Les séquences des amorces RIL6 F3L et RIL6 F6L sont identiques; \*\* Les séquences des amorces RIL6 F5L et RIL6 F7L sont identiques.

Oligonucléotides	Règle de design	Nombre de LNA	Niveau d'incorporation	Séquence 5'P – 3'OH		
RIL6 F0	Sans LNA	0		GCGATGATGCACTGTCAGAAAAC		
RIL6 R0	Sans LNA			AACGGAACTCCAGAAGACCAGAG		
RIL6 F1L	LNA espacés en extrémité 5'P	4	répartition uniforme en 5'P	GCGATGATGCACTGTCAGAAAAC		
RIL6 R1L	LNA espacés en extrémité 5'P		AACGGAACTCCAGAAGACCAGAG			
RIL6 F2L	Selon la POSITION de la base substituée	2	espacés en extrémité 5'P	GCGATGATGCACTGTCAGAAAAC		
RIL6 R2L			espacés en extrémité en 5'P	AACGGAACTCCAGAAGACCAGAG		
RIL6 F3L*			rapprochés en extrémité 5'P	GCGATGATGCACTGTCAGAAAAC		
RIL6 R3L			rapprochés en extrémité 5'P	AACGGAACTCCAGAAGACCAGAG		
RIL6 F4L			rapprochés en milieu de séquence	GCGATGATGCACTGTCAGAAAAC		
RIL6 R4L			rapprochés en milieu de séquence	AACGGAACTCCAGAAGACCAGAG		
RIL6 F5L**			Selon la NATURE de la base substituée	2	espacés en en extrémité 5'P	GCGATGATGCACTGTCAGAAAAC
RIL6 R5L					espacés en en extrémité 5'P	AACGGAACTCCAGAAGACCAGAG
RIL6 F6L*	rapprochés en extrémité 5'P	GCGATGATGCACTGTCAGAAAAC				
RIL6 R6L	rapprochés en extrémité 5'P	AACGGAACTCCAGAAGACCAGAG				
RIL6 F7L**	rapprochés en milieu de séquence	GCGATGATGCACTGTCAGAAAAC				
RIL6 R7L	rapprochés en milieu de séquence	AACGGAACTCCAGAAGACCAGAG				

### 3.4.2. Optimisation des amorces

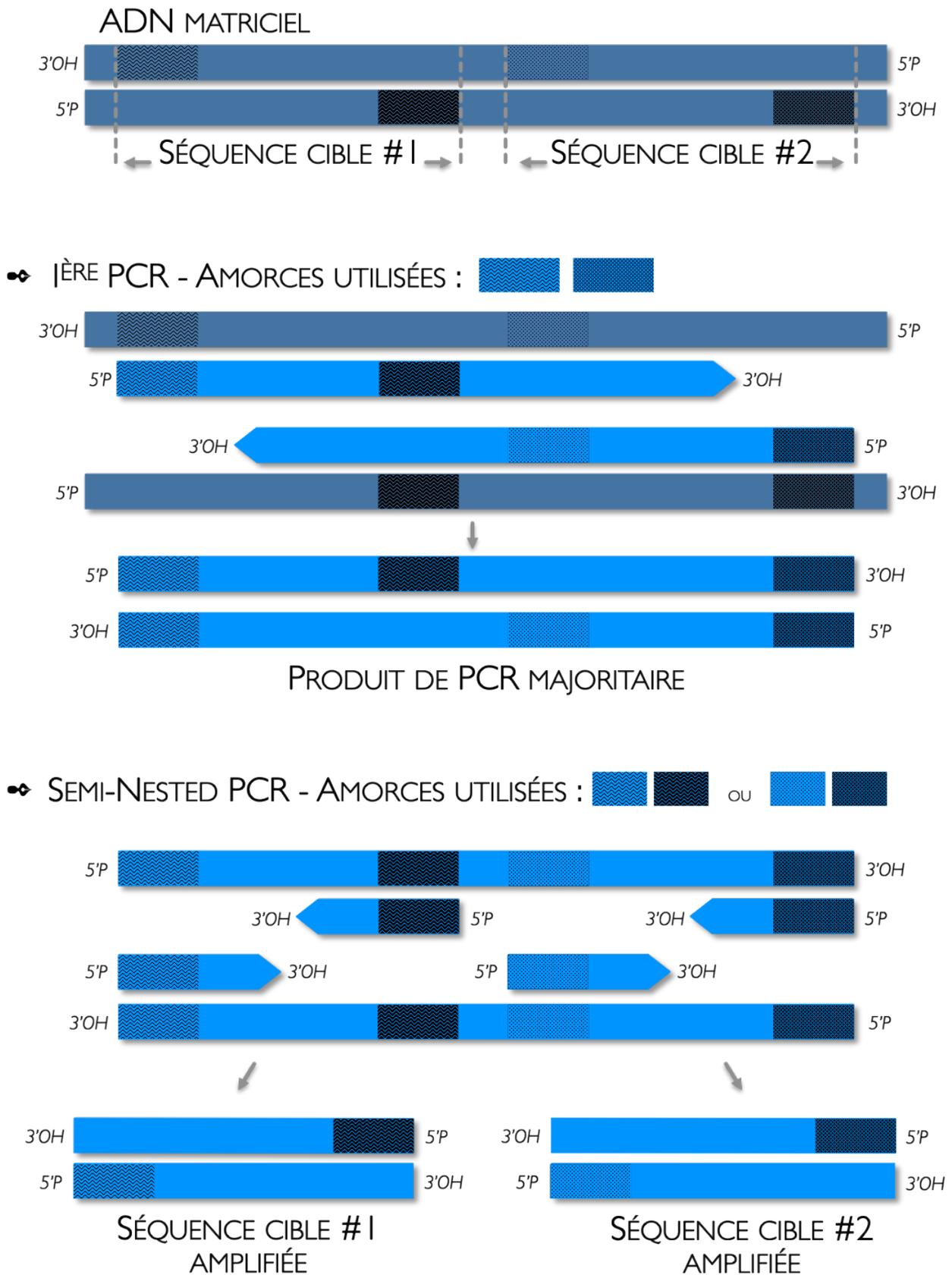
L'optimisation des **amorces** est systématiquement effectuée lorsqu'une détection en système *SybrGreen I* est réalisée. Elle est effectuée pour chaque type d'échantillon et chaque détection de gène. La méthode utilisée est la suivante :

#### 1. Mise au point par *Semi-Nested PCR* sur ADN recombinant (Figure 3.4)

Un premier couple d'**amorces** permet de détecter spécifiquement la cible. Après amplification puis dilutions en cascades sur 5 log, cet ADN recombinant est utilisé afin de permettre la mise au point d'un deuxième couple d'**amorces** qui servira quant à lui au dosage des échantillons. Tous les essais sont réalisés en **duplicate** pour chaque échantillon.

#### 2. Mise au point sur *pool* d'ADN complémentaires :

Une fraction de l'ensemble des échantillons est constituée sous la forme d'un *pool* permettant la réalisation de 5 points de concentrations différentes (1 ; 0,5 ; 0,25 et 0,125  $\mu$ L de RT pure). Les conditions de réaction sont optimisées afin d'obtenir une **efficacité de réaction** optimale (pente entre -3,2 et -3,5). On détermine ainsi la concentration idéale qui sera utilisée pour le dosage des échantillons à l'étape suivante.



**Figure 3.4 :** Diagramme de deux *Semi-Nested PCR*, réactions d'amplification d'ADN réalisées à l'aide de couples d'amorces s'hybridant à l'intérieur de la séquence d'un premier produit de PCR.

### 3.4.3. Amplification par qPCR - système *SybrGreen I*

Chaque réaction de qPCR a été réalisée sur thermocycleur *LightCycler™ 1.2* en capillaire borosilicate de 20  $\mu\text{L}$  (*Roche Applied Science*, Basel, Suisse). Cet appareil mesure à chaque cycle la fluorescence émise par le *SybrGreen I* et enregistrée au niveau des différents capillaires contenant les échantillons. Les cycles de quantification ( $C_q$ ) ont été déterminés en utilisant la méthode du maximum de la dérivée seconde (logiciel *LightCycler*, version 3.5, *Roche Applied Science*, Basel, Suisse). Le mélange est réalisé selon le **Tableau 3.3** suivant :

	Volume ( $\mu\text{L}$ )	Concentration finale
Eau dépourvue de DNase	x	-
Solution de $\text{MgCl}_2$ 25 mM	1,6	3 mM
	2,4	4 mM
	3,2	5 mM
Amorce sens	0,4	0,4 $\mu\text{M}$
Amorce antisens	0,4	0,4 $\mu\text{M}$
<i>Taq</i> polymérase + <i>SybrGreen</i> + dNTP + Tampon 10X	2,0	1X
Échantillon	5,0	0,25 à 0,5 $\mu\text{L}$ d'ADNc pur

**Tableau 3.3** : Composition du mix réactionnel pour une amplification d'un échantillon d'ADNc en système *SybrGreen I*, volume final de 20  $\mu\text{L}$ . La concentration du magnésium peut être ajustée en fonction de la qualité de l'hybridation. La concentration des solutions de travail des amorces est de 20 $\mu\text{M}$ . (*Roche Applied Science* - Kit LC Fast Start DNA SYBR® Master Green I).

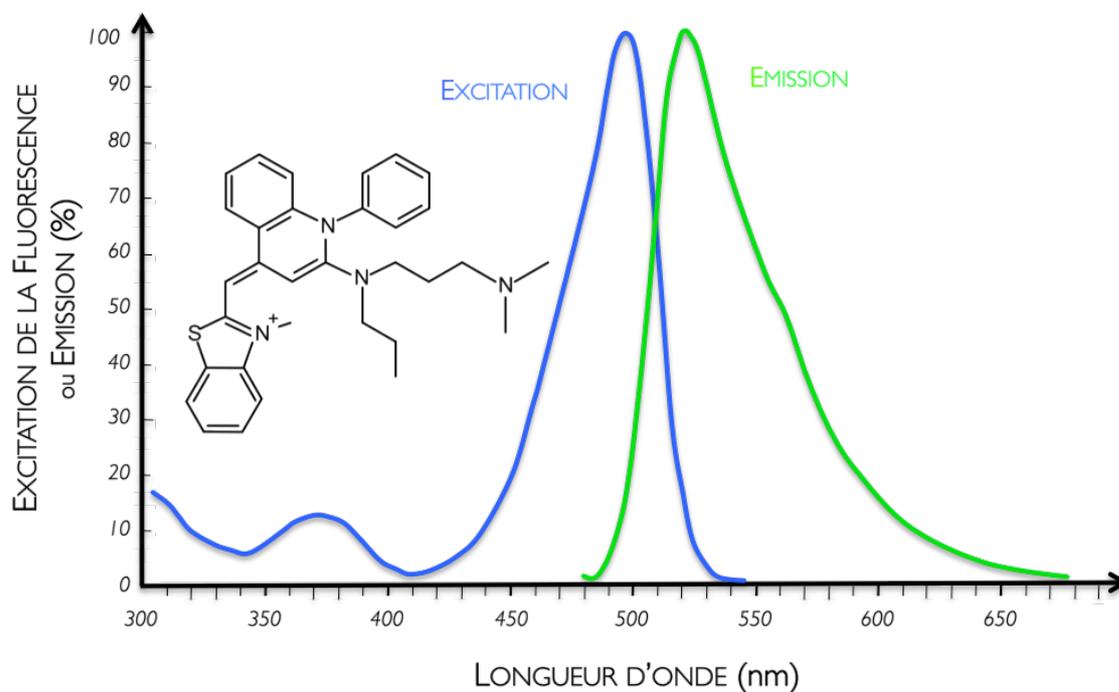
L'ADNc est ensuite amplifié par PCR selon le programme thermique suivant :

1. 1 cycle de **dénaturation** initiale à 95 °C pendant 10 min (20°C/sec), sans acquisition
2. 40 cycles d'amplification constitués :
  - 1<sup>er</sup> segment de **dénaturation** à 95 °C pendant **20 sec** (20°C/sec), sans acquisition
  - 2<sup>ème</sup> segment d'**hybridation** entre 53-60 °C pendant **5 à 10 sec** (20°C/sec), sans acquisition
  - 3<sup>ème</sup> segment de polymérisation à 72 °C pendant un temps **t (sec)** = nombre de bases du produit / 25 (20°C/sec), acquisition en fin de segment
3. 1 cycle de **fusion** constitué :
  - 1<sup>er</sup> segment de **dénaturation** à 95 °C pendant **20 sec** (20°C/sec), sans acquisition
  - 2<sup>ème</sup> segment d'**hybridation** entre 63-70 °C pendant **30 sec** (20°C/sec), sans acquisition
  - 3<sup>ème</sup> segment de **fusion** à 95 °C pendant 0 sec (0,1°C/sec), acquisition en continu
4. 1 cycle de **refroidissement** à 40 °C pendant **30 sec** (20°C/sec), sans acquisition

### 3.4.4. Intercalants ADN utilisés pour la fusion des duplexes

- **Le SYBR<sup>®</sup> Green I**

Le kit *LC Fast Start DNA SYBR<sup>®</sup> master Green I* (Roche Applied Science, Basel, Suisse) est le kit utilisé pour le système *SybrGreen I*. Cet intercalant ADN non spécifique constitue la molécule la plus utilisée au cours des analyses par PCR quantitative (Wittwer, Ririe et al. 1997 ; Zipper, Brunner et al. 2004) (Figure 3.5). C'est une cyanine asymétrique qui se lie préférentiellement à l'ADN double brin, au niveau du petit sillon de la molécule, mais de manière non systématique (Yue, Singer et al. 1997). Il est cependant capable de parfois se lier à l'ADN simple brin. Le complexe ADN- *SybrGreen I* résultant absorbe la lumière bleue ( $\lambda_{\max} = 488 \text{ nm}$ ) et émet de la lumière verte ( $\lambda_{\max} = 522 \text{ nm}$ ). On peut cependant lui attribuer certains inconvénients comme la modification de la précision de la quantification lorsque plusieurs amplicons de différentes tailles sont analysés, la quantité de fluorescence enregistrée étant dépendante de la longueur de la séquence cible détectée, et une inhibition de la réaction de type concentration-dépendante (Ririe, Rasmussen et al. 1997 ; Nath, Sarosy et al. 2000).



**Figure 3.5 :** Caractéristiques spectrophotométriques et structure moléculaire du SYBR<sup>®</sup>Green I (Yue, Singer et al. 1997).

- *L'EvaGreen*<sup>®</sup>

*L'EvaGreen*<sup>®</sup> (Biotium, Hayward, CA, USA (Mao and Leung 2010; Mao, Leung et al. 2010)) est un intercalant ADN fluorescent de dernière génération. Au cours de notre étude, il n'a été utilisé qu'au cours de la réalisation de *courbes de fusion* afin d'étudier l'effet d'un intercalant ADN saturant sur la mesure directe de la *température de fusion*. Cette molécule est en effet capable de s'intercaler au niveau d'une double hélice ADN de façon saturante, ce qui la rend particulièrement intéressante lorsqu'une importante résolution s'avère nécessaire dans certaines applications telles que la *qPCR* (Mao, Leung et al. 2007), l'analyse de *courbe de fusion ADN Haute Résolution (HRM)*, le suivi en temps réel d'amplifications dépendantes d'hélicase thermophile (tHDA), la quantification d'ADN en solution (Wang, Chen et al. 2006) ou encore l'électrophorèse en gel capillaire. De plus, cette molécule présente des spectres d'excitation et d'émission très proches de ceux de la fluorescéine (FAM) ou du *SybrGreen I*. Sans présence d'ADN, ce colorant est faiblement fluorescent avec un pic d'excitation à 495 nm et pic d'émission à 525 nm (**Figure 3.6**). Elle est thermiquement très stable et résistante à l'hydrolyse, ce qui la rend facile à manipuler. Non fluorescente par elle-même, elle le devient fortement lorsqu'elle se lie à l'ADN double brin. Elle présente aussi l'avantage d'être non mutagène et non cytotoxique (membranes cellulaires imperméables à la molécule), contrairement au *SybrGreen I*, qui pénètre dans la cellule rapidement et est connu pour être un puissant exhausteur de mutations (Ohta, Tokishita et al. 2001).

D'autre part, l'effet inhibiteur de cet intercalant au cours de la *PCR* est bien moins important que celui du *SybrGreen I* et il reste moins susceptible de provoquer des amplifications non spécifiques. En conséquence, *L'EvaGreen* peut être utilisé à une concentration beaucoup plus élevée que le *SybrGreen I*, permettant ainsi l'enregistrement d'un signal de fluorescence plus robuste au cours de la réaction.

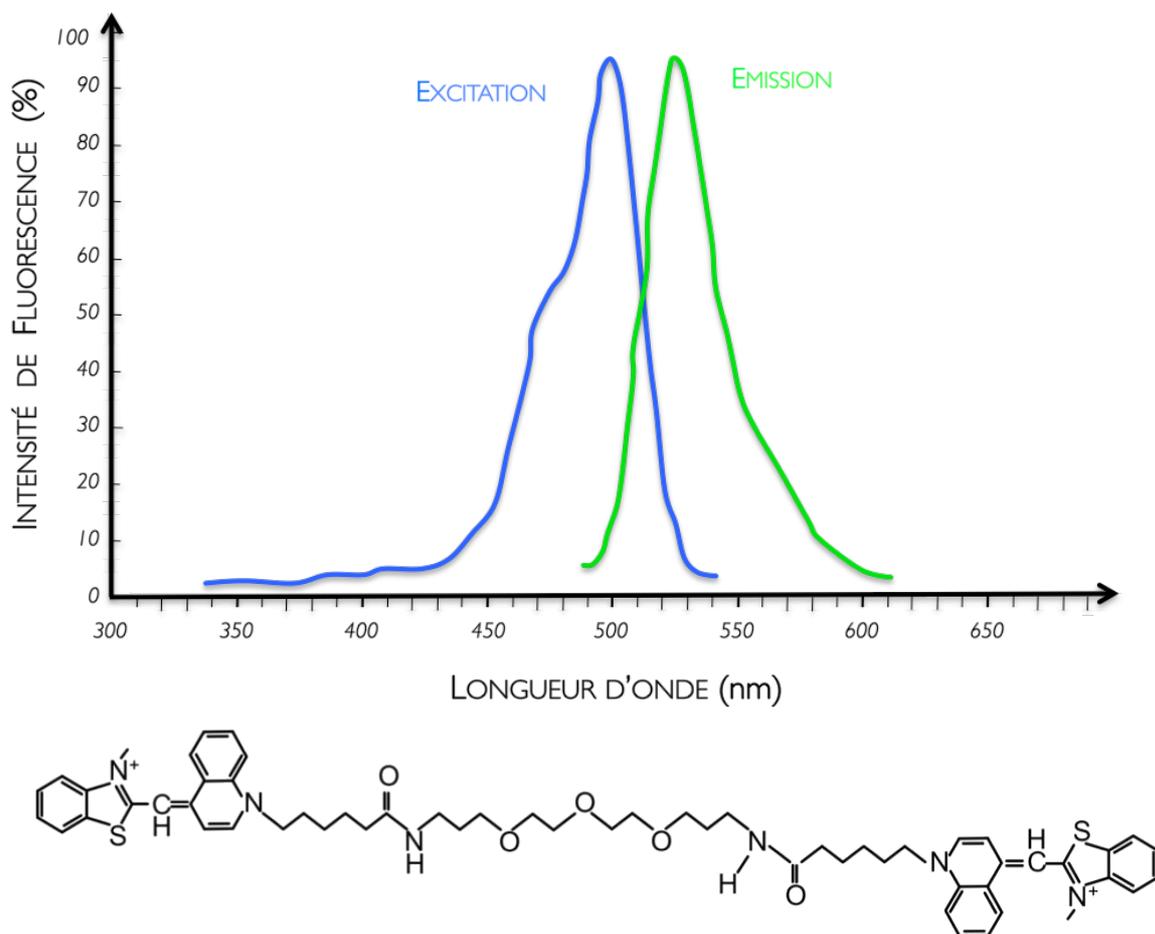


Figure 3.6 : Caractéristiques spectrophotométriques et structure moléculaire de l'EvaGreen® (Mao, Leung et al. 2010) (Mao and Leung 2010)

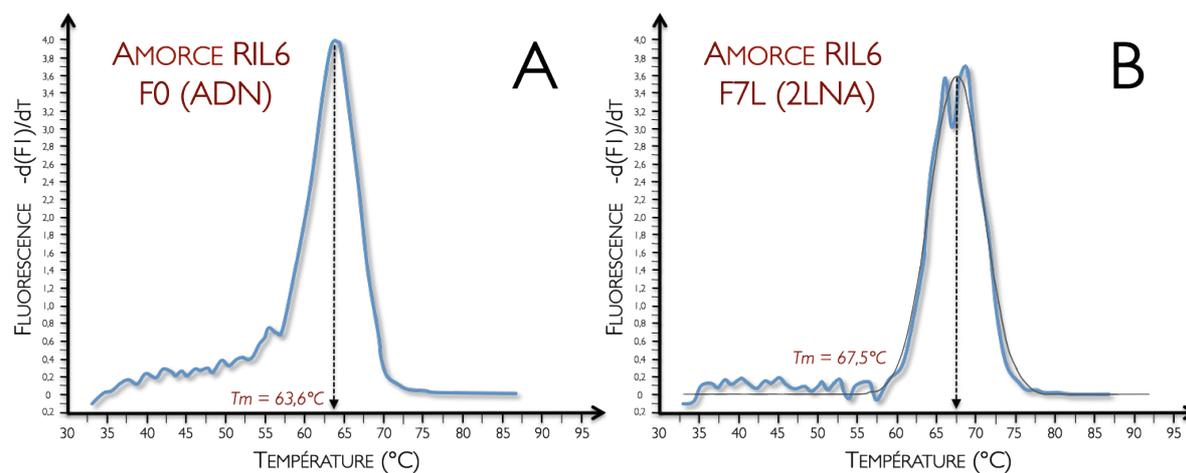


Figure 3.7 : Pics de fusion obtenus en EvaGreen pour l'amorce ADN sens spécifique de l'IL-6 (NM\_012589) (A) et la même amorce modifiée par 2 monomères LNA rapprochés en milieu de séquence, aux mêmes concentrations, vitesse de transition thermique de 0,1°C/sec. L'intercalant saturant permet l'observation des deux incorporations de LNA au niveau de la fusion de l'amorce modifiée (B).

### 3.4.5. Mesure du $T_m$ par courbe de fusion sur *LightCycler*<sup>TM</sup>

La température de fusion de chaque amorce ADN (Tableau 3.4) et chaque amorce modifiée par LNA (Tableau 3.2) est directement mesurée en réalisant une courbe de fusion d'un duplexe composé de l'amorce elle-même hybridée à sa séquence complémentaire, comme décrit précédemment (Teo, Choi et al. 2002). Cette dissociation par la chaleur de l'ADN double brin est réalisée en capillaire en verre sur *LightCycler*<sup>TM</sup> (Roche Applied Science, Basel, Suisse). Chaque fusion est réalisée en duplicate soit en mix de PCR (Kit Fast Start DNA Master SYBRGreen - 32 mM de Na<sup>+</sup>, 5 mM de Mg<sup>2+</sup>, 1,2 mM de dNTP, Roche Applied Science, Basel, Suisse) aux concentrations de 0,6 µM et de 20 µM ; soit en tampon phosphate, utilisé en calorimétrie (1 mM Na<sub>2</sub>EDTA, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1 M NaCl, pH 7) à la concentration de 20 µM, dans un volume final de 20 µL.

La dénaturation de l'ADN est obtenue selon le programme thermique suivant :

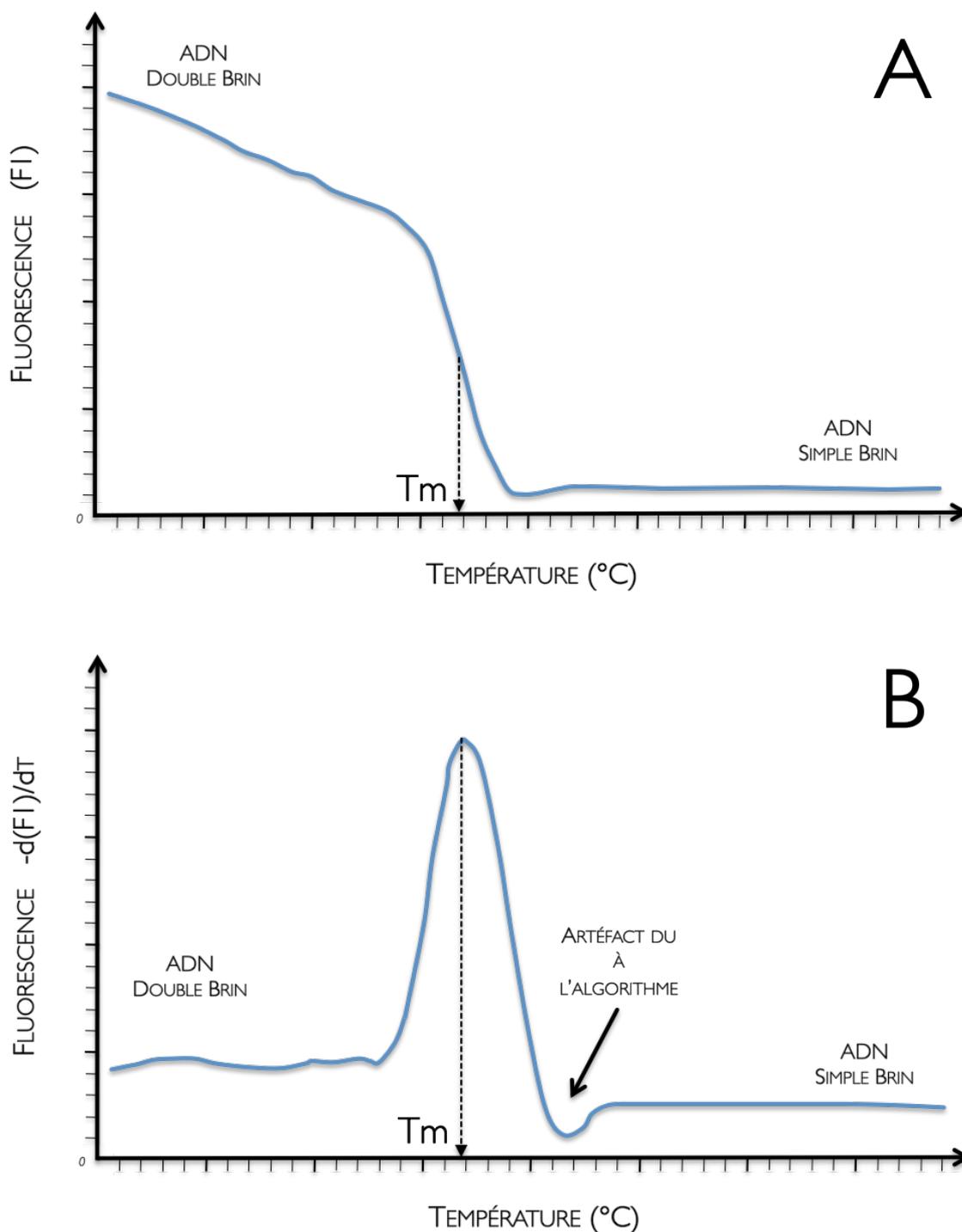
1. 1 cycle de dénaturation constitué :

- 1<sup>er</sup> segment à 95 °C pendant 1 min ; vitesse de transition thermique de 20°C/sec, sans acquisition.
- 2<sup>ème</sup> segment indiquant la température de consigne à laquelle la fusion doit commencer : 30°C. La vitesse de transition thermique est de 20°C/sec. Aucune acquisition.
- 3<sup>ème</sup> segment indiquant la température de consigne à laquelle la fusion doit se terminer : 95°C. La vitesse de transition thermique est de 0,1°C/sec (tests réalisés à 1,5°C/sec et 3°C/sec). Acquisition en continu.

2. 1 cycle de refroidissement constitué d'un segment indiquant la température de consigne à laquelle la dénaturation doit se terminer : 40°C. La vitesse de transition thermique est de 20°C/sec. Durée de refroidissement : 30 secondes. Aucune acquisition.

Les valeurs du  $T_m$  ont été mesurées à l'aide du logiciel *Roche Molecular Biochemicals LightCycler version 3.5* (Roche Applied Science, Basel, Suisse), avec et sans correction automatique de bruit de fond.

Au cours d'une analyse par **courbe de fusion**, la fluorescence est enregistrée en fonction de la température (**Figure 3.8A**). Au début du programme thermique, le mélange réactionnel se trouve à basse température et les séquences contenues dans le capillaire se trouvent alors à l'état double brin. En conséquence, le signal de fluorescence enregistré est élevé. Au fur et à mesure que la température augmente de façon constante, le *SybrGreen I* se libère du petit sillon de la double hélice, dans un premier temps au niveau de séquences AT-riches. La baisse de fluorescence va ainsi passer par une certaine température où 50% de l'ADN se trouve dénaturé. Cette température caractéristique d'une séquence ADN est alors appelée **température de fusion**, notée  $T_m$ . Après **dénaturation** complète d'un **duplexe**, le logiciel **Roche Molecular Biochemicals LightCycler version 3.5** restitue l'analyse sous forme de pic de fusion, obtenu par représentation graphique de la dérivée première négative de la fluorescence enregistrée au niveau de l'échantillon analysé (**Figure 3.8B**).



**Figure 3.8 :** Courbe de fusion d'un duplexe ADN (A) et son pic de fusion correspondant, représenté par la dérivée première négative de la fluorescence enregistrée en fonction de température (B). Le signal fluorescent, enregistré à 530 nm, est obtenu à partir du *SybrGreen I*, intercalant ADN capable d'émettre une fluorescence plus importante lorsqu'il se trouve lié au petit sillon d'une double hélice ADN.

**Tableau 3.4 :** Caractéristiques des oligonucléotides utilisés pour l'étude de la mesure de la température de fusion. Les oligonucléotides sont synthétisés puis purifiés par Chromatographie en phase Liquide à Haute Performance, HPLC (Eurogentec, Saraing, Belgique).

Gène	Oligonucléotides	Séquence 5'P – 3'OH	Localisation	Longueur du duplex	Numéro d'accèsion GenBank
<i>Chemokine (C-X-C motif) ligand 2, CXCL2</i>	MCXCL2 F1	TCCCTCAACGGAAGAACCAAAGAGA	377-402	25	NM_009140
	MCXCL2 F1 rev	TCTCTTTGGTTCTTCCGTTGAGGGA			
	MCXCL2 R1	CGAGGCACATCAGGTACGATCCA	454-432	23	
	MCXCL2 R1 rev	TGGATCGTACCTGATGTGCCTCG			
<i>Interleukin 6, IL6</i>	RIL6 F10	GCGATGATGCACTGTCAGAAAAC	300-322	23	NM_012589
	RIL6 F10 rev	GTTTTCTGACAGTGCATCATCGC			
	RIL6 R10	AACGGA ACTCCAGAAGACCAGAG	428-406	23	
	RIL6 R10 rev	CTCTGGTCTTCTGGAGTTCCGTT			
<i>Peptidylprolyl isomerase A, PPIA</i>	Mut 3 F	GCGATGATGCACTGGCTGTAAAC	300-322	23	NM_017101
	Mut 3 F Rev	GTTTACAGCCAGTGCATCATCGC			
	RCYCA F2	TATCTGCACTGCCAAGACTGAGTG	381-404	24	
	RCYCA F2 rev	CACTCAGTCTTGGCAGTGCAGATA			
<i>Ribosomal protein, large, P0, RPLP0</i>	RCYCA R1	CTTCTTGCTGGTCTTGCCATTCC	507-485	23	NM_007475
	RCYCA R1 rev	GGAATGGCAAGACCAGCAAGAAG			
	MARBP F1	GAAAATCTCCAGAGGCACCATTG	556-579	23	
	MARBP F1 rev	CAATGGTGCCTCTGGAGATTTTC			
	MARBP R1	TCCCACCTTGCTCCAGTCTTTATC	625-601	25	
	MARBP R1 rev	GATAAAGACTGGAGACAAGGTGGGA			

### 3.4.6. Amplification par qPCR - système Sonde d'hydrolyse

Toutes les réactions ont été réalisées sur thermocycleur *LightCycler™* 1.2 en capillaire de 20  $\mu$ L (*Roche Applied Science*, Basel, Suisse) selon le mélange décrit en **Tableau 3.5**. Là encore, l'appareil va mesurer à chaque cycle la fluorescence émise par la carboxyfluorescéine (FAM) et enregistrée au niveau des différents capillaires correspondant aux échantillons. La spécificité des amorces est préalablement vérifiée en système *SybrGreen I* à une  $T_H$  de 60°C. Les cycles de quantification ( $C_q$ ) ont été déterminés en utilisant la méthode du maximum de la dérivée seconde (logiciel *LightCycler*, version 3.5, *Roche Applied Science*, Basel, Suisse).

	Volume ( $\mu$ L)	Concentration finale
Eau dépourvue de DNase	x	-
	1,6	3 mM
Solution de MgCl <sub>2</sub> 25 mM	2,4	4 mM
	3,2	5 mM
Amorce sens	0,4	0,4 $\mu$ M
Amorce antisens	0,4	0,4 $\mu$ M
Sonde	0,4	0,2 $\mu$ M
Taq polymérase + dNTP + Tampon 10X	2,0	1X
Échantillon	5,0	0,25 à 0,5 $\mu$ L d'ADNc pur

**Tableau 3.5 :** Composition du mix réactionnel pour une amplification d'un échantillon d'ADNc en système *TaqMan*, volume final de 20  $\mu$ L. La concentration du MgCl<sub>2</sub> peut être ajustée en fonction de la qualité de l'hybridation. La concentration des solutions d'amorces est de 20 $\mu$ M, celle de la sonde est de 10 $\mu$ M. La sonde d'hydrolyse est marquée en 5'P par la carboxyfluorescéine (6-FAM) et en 3'OH par un quencheur : la carboxytétraméthylrhodamine (TAMRA). (*Kit LC Fast Start DNA Master HybProbe*, Roche).

L'ADNc est ensuite amplifié par PCR selon le programme thermique suivant :

1. 1 cycle de **dénaturation** initiale à 95 °C pendant 10 min (20°C/sec), sans acquisition
2. 40 cycles d'**amplification** constitués :
  - 1<sup>er</sup> segment de **dénaturation** à 95 °C pendant 10 sec (20°C/sec), sans acquisition
  - 2<sup>ème</sup> segment d'**hybridation** à 60, 62 ou 64 °C pendant 5 sec (20°C/sec), acquisition
  - 3<sup>ème</sup> segment de polymérisation à 72 °C pendant un temps **t (sec)** = nombre de bases du produit / 25 (20°C/sec), acquisition en fin de segment
3. 1 cycle de **refroidissement** à 40 °C pendant 30 sec (20°C/sec), sans acquisition

### 3.4.7. Système de sonde d'hydrolyse (Taqman®)

Ce système, décrit pour la première fois en 1991 (Holland, Abramson et al. 1991), utilise la propriété exonucléase 5'P-3'OH que possède la *Taq polymérase* qui va cliver une sonde marquée, hybridée spécifiquement au niveau de sa séquence cible complémentaire. Très utilisé en qPCR, il apparaît être comme l'un des meilleurs systèmes de détection actuel en termes de spécificité.

La sonde est construite de la manière suivante : un fluorophore, généralement la fluorescéine (FAM), est fixée à l'extrémité 5'P tandis qu'un quencheur qui va éteindre le signal de fluorescence émis par la FAM, traditionnellement la rhodamine (TAMRA), est fixé au niveau de l'extrémité 3'OH de la sonde. Le fluorophore est excité à une certaine longueur d'onde et va transférer son énergie par FRET au niveau du quencheur qui va fluorescer. Cette fluorescence est détectée à une longueur d'onde différente de celle du fluorophore et ainsi le bruit de fond réactionnel est faible. Si aucune hybridation des amorces de PCR ne se réalise en amont de la sonde, cette dernière va rester intacte et aucune fluorescence ne sera détectée. Si au contraire, une hybridation des amorces conduit à la rencontre de la *Taq polymérase* avec l'extrémité 3'OH de la sonde, alors l'activité 5'P-3'OH exonucléasique permettra la libération du fluorophore puis du quencheur dans le milieu réactionnel. La sonde est hydrolysée. Comme l'éloignement entre ces deux composés n'est plus suffisant pour que s'établisse un FRET, une augmentation irréversible de la fluorescence est enregistrée par le système de détection. À une séquence cible détectée correspond donc un signal de fluorescence mesuré. Généralement, une étape combinée d'hybridation et de polymérisation à 60-62°C est réalisée, la plupart des sondes ayant un  $T_m$  autour de 70°C. À cette température est assuré un maximum d'activité exonucléasique 5'P-3'OH de la part de la *Taq polymérase* qui présentera toutefois une réduction de sa processivité.

### 3.4.8. Modèle d'évaluation de la spécificité des amorces modifiées

Le modèle repose sur la détection par **sondes** d'hydrolyse de deux séquences synthétiques, l'une spécifique de l'IL-6 et l'autre, représentative des détections non spécifiques, caractérisée par la séquence de l'EGFP (**Alignements 3.1**). La **matrice** non spécifique a été obtenue au moyen d'**oligonucléotides** mutés à différents degrés dont les séquences figurent au niveau du **Tableau 3.6**.

**Tableau 3.6 :** Séquence 5'P-3'OH des oligonucléotides ayant été utilisés pour la construction du modèle d'évaluation des amorces modifiées par *LNA in vitro*. Ces oligonucléotides sont synthétisés puis purifiés par Chromatographie en phase Liquide à Haute Performance, HPLC (Eurogentec, Saraing, Belgique). Les bases mutées sont notées en minuscules. Les substitutions de *LNA* sont notées en grandes lettres. Les sondes d'hydrolyse sont marquées en 5'P par la carboxyfluorescéine (6-FAM) et en 3'OH par la carboxytétraméthylrhodamine (TAMRA).

Oligonucléotides	Séquence 5'P - 3'OH	Oligonucléotides	Séquence 5'P - 3'OH
RIL6 F0	GCGATGATGCACTGTCAGAAAAC	RIL6 R0	AACGGAACTCCAGAAGACCAGAG
Mut 2F	GCGATGATGCACTGgCAGtAAAC	Mut 2R	AACGGAACTCCAGAtGACgAGAG
Mut 3F	GCGATGATGCACTGgCtGtAAAC	Mut 3R	AACGGAACTCCAGAtGtCgAGAG
Mut 4F	GCGATGATGCACTGgCtGtAtAC	Mut 4R	AACGGAACTCCAGAtGtCgACAG
Sonde spécifique IL6	Séquence 5'P - 3'OH	Sonde spécifique EGFP	Séquence 5'P - 3'OH
RIL6 S10	CCAGTTTGAAGCATCATCATTTC	EGFP S1	CGACACCCTGGTGAACCGCA

**Alignements 3.1 :** Les séquences des matrices sont représentées dans le sens 5'P  $\Rightarrow$  3'OH, les séquences des sondes d'hydrolyse dans le sens 3'OH  $\Rightarrow$  5'P. Ces oligonucléotides sont synthétisés puis purifiés par Chromatographie en phase Liquide à Haute Performance, HPLC (Eurogentec, Saraing, Belgique). Les séquences des amorces de l'IL6 sont identifiées en rouge. Les bases mutées sont notées en minuscules. Les substitutions de *LNA* sont notées en grandes lettres. Les sondes d'hydrolyse sont marquées en 5'P par la carboxyfluorescéine (6-FAM) et en 3'OH par la carboxytétraméthylrhodamine (TAMRA).

- Matrice SPÉCIFIQUE de l'IL6, RIL6LNAMatrix :

GCGATGATGCACTGTCAGAAAACAATCTGAAACTTCCAGAAATACAAA**GAAATGATGGATGCTTCCAACTGG**ATATAACCAGGAAATT  
 CTTTACTAC**CTACGAAGGTTGACC**  
 TGCCTATTGAAAATCTGCTCTGGTCTTCTGGAGTCCGTT  
 Sonde RIL6 S10

- Matrice SPÉCIFIQUE de l'EGFP, RIL6mutEGFP2 :

GCGATGATGCACTGgCAGtAAACGCGCCGAGGTGAAGTTCGAGGG**CGACACCCTGGTGAACCGCA**TCGAGCTGAAGGCTCT**CGTca**TCT  
 GGAGTCCGTT  
 Sonde EGFP S1

### 3.4.9. Test d'inhibition de la PCR

Afin d'évaluer spécifiquement l'effet des inhibiteurs éventuellement co-purifiés au cours de l'étape d'extraction puis retrouvés dans la solution d'ADNc sur la réaction d'amplification de l'ADN cible, nous avons utilisé le test de détection par sonde d'hydrolyse, système *TaqMan* (Holland, Abramson et al. 1991) d'une matrice étrangère d'ADN introduite dans le mix de PCR (Nolan, Hands et al. 2006). Trois tests ont ainsi été mis au point en utilisant le logiciel *Primer3* (<http://frodo.wi.mit.edu/primer3>) :

- une première matrice ADN, publiée dans l'étude précédemment citée issue du gène codant pour le phytochrome B de la pomme de terre (*Solanum tuberosum*, GenBank Y14572) (**Alignement 3.2**), détectée par sonde d'hydrolyse.
- une matrice ADN de l'EGFP (*Enhanced Green Fluorescent Protein* (Thastrup, Tullin et al. 2005), détectée par *SybrGreen I*.
- une matrice ADN issue du gène codant pour la protéine de 14 kDa du virus de la vaccine (GenBank M18173), détectée par *SybrGreen I*.

Les caractéristiques des amorces et sondes utilisées pour ces différentes matrices sont listées au niveau du **Tableau 3.8**. Les caractéristiques des PCR sont résumées au niveau du **Tableau 3.7**. Douze échantillons d'ADNc reverse transcrits de deux manières différentes sont ainsi analysés. Les deux groupes de nombres de copies détectées spécifiquement sont ensuite comparés en utilisant le test statistique de *Wilcoxon*.

Gène Alien	Paramètres de calibration de la courbe			
	$E_{PCR}^a$	penste	Y intercept	$r^{2b}$
Spud-A	100,7	-3,305	42,12	1,00
EGFP	105,3	-3,201	42,22	1,00
14 KD	101,1	-3,295	43,22	1,00

**Tableau 3.7** Caractéristiques des courbes de calibration réalisées à partir d'un *pool* d'échantillons d'ADNc dilué en cascades avec un volume de départ de 5 $\mu$ L de mix RT, dilué jusqu'à 10 fois (4 différentes dilutions analysées en duplicate) pour les détections en système *TaqMan* des matrices *alien* Spud-A, EGFP et 14 KDa. Chaque PCR est effectuée à une  $T_H$  de 60°C, une concentration de sonde de 0,2  $\mu$ M et une concentration de MgCl<sub>2</sub> de 4 mM. <sup>a</sup> pourcentage d'efficacité PCR déterminée à partir de la pente de la courbe de calibration, <sup>b</sup> coefficient de régression linéaire de la courbe de calibration. Spud-A : matrice synthétisée, EGFP : produit de *nested* PCR, 14 KDa : produit de PCR sur plasmide.

**Tableau 3.8 :** Caractéristiques des amorces et des sondes d'hydrolyse utilisées en *TaqMan* pour le test d'inhibition de la qPCR. Les oligonucléotides sont synthétisés puis purifiés par Chromatographie en phase Liquide à Haute Performance HPLC (Eurogentec, Saraing, Belgique). \*Les sondes d'hydrolyse sont marquées en 5'P par la carboxyfluorescéine (6-FAM) et en 3'OH par la carboxytétraméthylrhodamine (TAMRA)

Gène <i>alien</i>	Oligonucléotides	Séquence 5'P – 3'OH	Localisation	Longueur de l'amplicon (pb)	T <sub>m</sub> (°C)	Numéro d'accèsion GenBank
<i>Solanum tuberosum</i> , Spud	Spud FW	AACTTGGCTTTAATGGACCTCCA	449-471	100	66	Y14572
	Spud BW	ACATTCATCCTTACATGGCACCA	527-549			
	Sonde Spud	TGCACAAGCTATGGAACACCACGT*	482-505		72	
<i>Enhanced Green Fluorescent Protein</i> , EGFP	EGFP FW	ACGACGGCAACTACAAGACC	5415-5434	94	62	DQ071886
	EGFP BW	GTCCTCCTTGAAGTCGATGC	5490-5509			
	Sonde EGFP S1	CGACACCCTGGTGAACCGCA*	5458-5477		66	
<i>Vaccinia virus 14kd fusion protein</i> , 14 KD	14 KD FW (3)	CTTGCAATTCCAGCAACTGA	79-98	91	58	M18173
	14 KD BW (3)	CGTCTTCATCGGCTTTAACA	150-169			
	Sonde 14 KD S3	ATTGCTTCGCGTTTAGCCTCTGGCTTT*	123-149		80	

### 3.5. Méthodes prédictives de la température de fusion

La température de fusion des oligonucléotides utilisés au cours de cette étude (Tableaux 3.2, 3.4) est évaluée à l'aide de quatre logiciels différents : *MacVector* (Accelrys, San Diego, CA, USA) ainsi que trois autres logiciels accessibles sur internet :

1. Le logiciel *OligoAnalyzer 3.1*, conçu par Richard OWCZARZY (Owczarzy, You et al. 2004 ; Owczarzy, Tataurov et al. 2008) et disponible à partir du site internet d'*IDT (Integrated DNA Technologies, Coralville, Iowa, USA)*. Il permet de prédire le  $T_m$  de séquences modifiées par différentes molécules en prenant en compte les concentrations de la majeure partie des constituants d'un mélange réactionnel (Figure 3.9).

Lien : <http://eu.idtdna.com/analyzer/applications/oligoanalyzer/>,

#### OligoAnalyzer 3.1

Instructions | Definitions | Feedback

**Sequence** # Bases 23

5'-AACGG+AAC+TCC+AG+AAGACCAGAG-3'

3'

**Target Type** DN2

**Oligo Conc** 0.3  $\mu$ M

**Na<sup>+</sup> Conc** 32 mM

**Mg<sup>++</sup> Conc** 5 mM

**dNTPs Conc** 1.2 mM

---

Results

5' mods

Internal Mods

3' mods

Mixed Bases

**Standard Mixed Base Instructions**

To use a Standard Mixed Base, simply type in the IUB symbol (from the table below) which represents the desired mix.

R	A,G
Y	C,T
M	A,C
K	G,T
S	C,G
W	A,T
H	A,C,T
B	C,G,T
V	A,C,G
D	A,G,T
N	A,C,G,T

**Custom Mixed Base Instructions**

To use Custom Mixed Bases  
Enter the desired percentage of each base (Integers Only, Totaling 100%).  
Press 'Use Mix Base' button to add your custom mixed base.

**Please note:** An additional charge is applied for hand mixing these custom bases.

%A   %C   %G   %T

**Base Notations**  
 DNA = A, C, G, T and U, I  
 Mixed Bases = Please enter bases in UPPERCASE  
 Phosphorothioated DNA = A\*, G\*, C\*, T\*  
 RNA = rA, rG, rC, rU  
 Phosphorothioated RNA = rA\*, rG\*, rC\*, rU\*  
 2'O-Methyl RNA = mA, mG, mC, mU  
 Phosphorothioated 2'O-Methyl RNA = mA\*, mG\*, mC\*, mU\*  
 Locked Nucleic Acid (LNA) = +A, +G, +C, +T

Figure 3.9 : Le logiciel *OligoAnalyzer version 3.1* permet de prendre notamment en compte le type d'acide nucléique, la longueur de la séquence, sa concentration, les concentrations en sodium, magnésium et dNTP. Il permet aussi de considérer le type de base incorporée dans la séquence de l'oligonucléotide (phosphorothioate, méthylation, ou LNA).

2. Le logiciel *Exiqon T<sub>m</sub> prediction 1.1*, conçu par la société *Exiqon* (Vedbaek, Danemark), titulaire du brevet d'utilisation des *LNA*, et disponible à partir du site internet de l'entreprise. Il permet de prédire le  $T_m$  de différentes séquences modifiées par des monomères *LNA* selon le modèle de John SANTALUCIA, Jr. (SantaLucia 1998) (Figure 3.10).

Lien : <http://www.exiqon.com/ls/Pages/ExiqonTMPredictionTool.aspx>

### LNA™ Oligo T<sub>m</sub> Prediction

LNA™-enhanced oligonucleotides have different melting properties from DNA oligonucleotides.

To use this melting temperature prediction (T<sub>m</sub>) tool, just enter your sequence below and press "calculate". DNA nucleotides are entered as A, C, T or G and LNA™ nucleotides are written +A, +C, +T or +G.

This tool is based on a modified nearest-neighbor thermodynamical model and on LNA™ oligonucleotide melting temperature measurements. To learn more about the T<sub>m</sub> prediction press [here](#).

Please use this base code:

- DNA bases: A, C, G, T
- Locked nucleic acid bases (LNA™) are entered as '+\_', e.g., +A

Oligo Sequence:

DNA T<sub>m</sub> :  Use T<sub>m</sub> DNA for applications such as Chromatin ISH-PCR - Southern

RNA T<sub>m</sub>:  Use T<sub>m</sub> RNA for applications such as RNA ISH - microarray - Northern

Figure 3.10 : Le logiciel *Exiqon T<sub>m</sub> prediction version 1.1* permet de prendre notamment en compte le type d'acide nucléique, la longueur de la séquence et le type de base incorporée dans la séquence de l'oligonucléotide (base ordinaire ou *LNA*).

3. Le logiciel *MELTING 4.1f*, conçu par Nicolas LE NOVÈRE et publié dans le Journal *Bioinformatics* (Le Novere 2001). Il permet de prédire le  $T_m$  de différentes séquences ADN selon les paramètres thermodynamiques de 7 études différentes (Breslauer, Frank et al. 1986 ; Freier, Kierzek et al. 1986 ; Sugimoto, Nakano et al. 1995 ; SantaLucia, Allawi et al. 1996 ; Sugimoto, Nakano et al. 1996 ; Allawi and SantaLucia 1997 ; Xia, SantaLucia et al. 1998) ainsi qu'une correction considérée pour les sels selon 3 études différentes (Wetmur 1991; SantaLucia, Allawi et al. 1996 ; SantaLucia 1998 ) (Figure 3.11).

Lien : <http://mobyale.pasteur.fr/cgi-bin/portal.py?#forms::melting>

## MELTING 4.1f

Run

Reset

### enthalpy, entropy and melting temperature

\* Hybridisation type (-H)

\* Nearest Neighbor parameters set (-A)

\* Sequence string (-S)

Complementary sequence (-C)

\* Salt concentration (-N) ?

\* Nucleic acid concentration in excess (-P) ?

Nucleic acid correction factor (-F)

Salt correction (-K)

Force approximative temperature computation (-x)

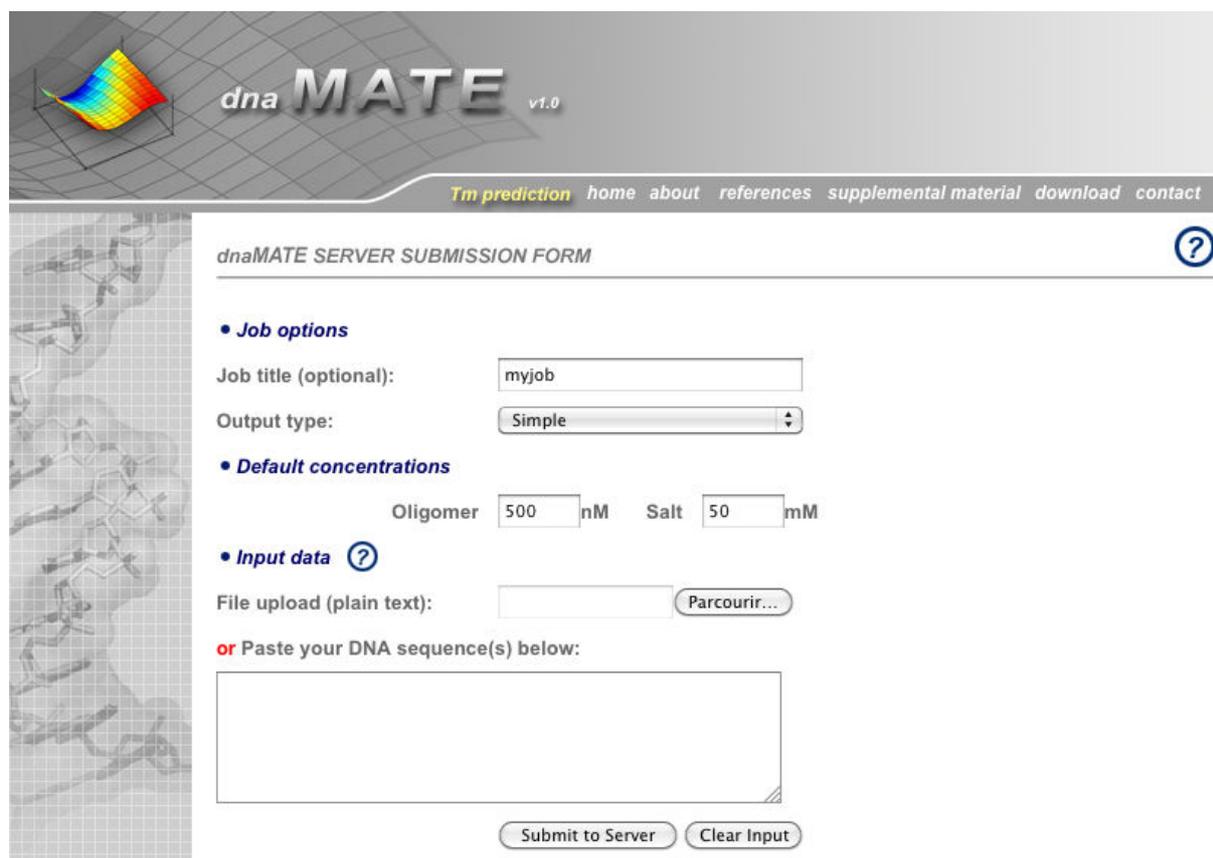
Use parameters for dangling ends (dnadnade.nn) (-D)?

Use parameters for mismatches (dnadnamm.nn) (-M)?

Figure 3.11 : Le logiciel *MELTING version 4.1f* permet de prendre notamment en compte le type d'hybridation, les paramètres thermodynamiques de 7 études différentes, la longueur de la séquence, la concentration en sodium et la correction en sels.

4. Le logiciel *dnaMATE*, conçu par Alejandro PANJKOVICH et publié dans le Journal *Nucleic Acids Research* (Panjkovich, Norambuena et al. 2005). Il permet de calculer une **température de fusion** consensus pour de courtes séquences ADN de 16 à 30 nucléotides selon les paramètres thermodynamiques de 3 études différentes (Breslauer, Frank et al. 1986 ; Sugimoto, Nakano et al. 1996; SantaLucia 1998 ) (Figure 3.12).

Lien : <http://dna.bio.puc.cl/cardex/servers/dnaMATE/tm-pred.html>



The screenshot displays the 'dnaMATE v1.0' web interface. At the top left is a logo with a colorful heatmap. The main header includes the text 'dnaMATE v1.0' and a navigation menu with links: 'Tm prediction', 'home', 'about', 'references', 'supplemental material', 'download', and 'contact'. Below the header is the 'dnaMATE SERVER SUBMISSION FORM' with a help icon. The form is organized into sections: 'Job options' with a text input for 'Job title (optional):' containing 'myjob' and a dropdown for 'Output type:' set to 'Simple'; 'Default concentrations' with inputs for 'Oligomer' (500 nM) and 'Salt' (50 mM); 'Input data' with a 'File upload (plain text):' field and a 'Parcourir...' button, and a 'Paste your DNA sequence(s) below:' text area. At the bottom are 'Submit to Server' and 'Clear Input' buttons.

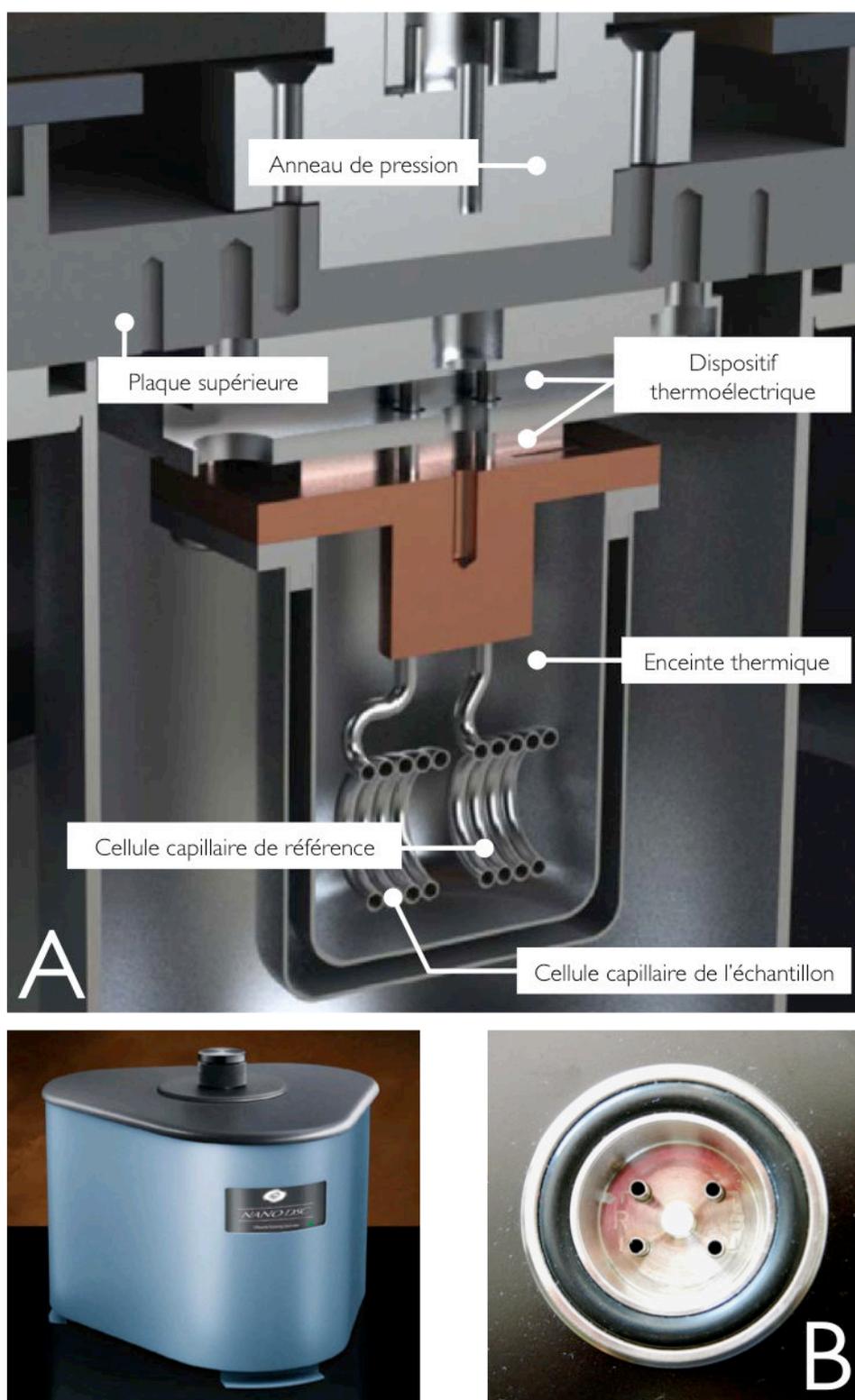
Figure 3.12 : Le logiciel *dnaMATE* permet de prendre en compte les paramètres thermodynamiques de 3 études différentes, la longueur et la concentration de la séquence et la concentration en sodium.

### 3.6. Calorimétrie Différentielle à balayage

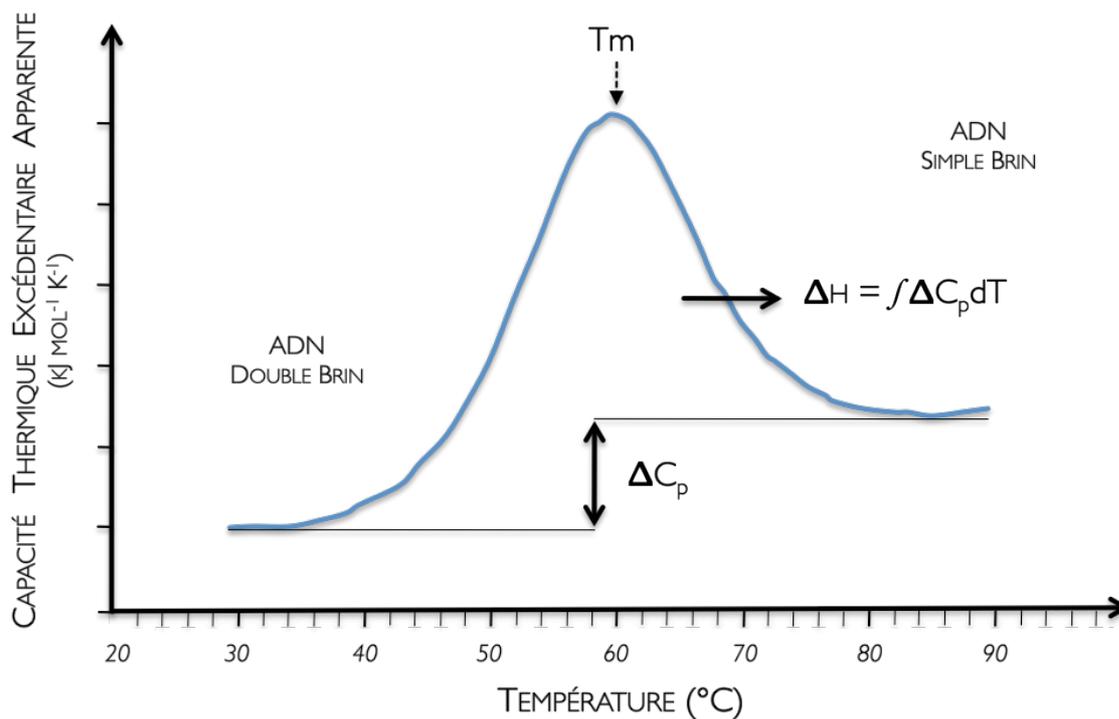
La Calorimétrie Différentielle à balayage (*Differential Scanning Calorimetry*) a été retenue comme méthode de référence afin de mesurer la température de fusion de nos différentes amorces ADN et modifiées par LNA. Ces mesures ont été effectuées sur un nano calorimètre CSC 6300 (TA Instruments, New Castle, Delaware, USA, **Figure 3.13**). L'appareil dispose de deux cellules de mesure en platine : une cellule de référence dans laquelle est uniquement déposé le tampon d'analyse et une deuxième cellule dans laquelle est déposé l'échantillon à analyser, ici un duplexe ADN composé de l'amorce PCR avec sa séquence complémentaire. Chaque cellule possède son propre système de chauffage et de contrôle de la température. Au cours de l'analyse, un programme de température identique est appliqué aux deux cellules. Lorsqu'un phénomène de transition survient dans l'échantillon, celui-ci ne peut plus suivre le programme thermique et une différence de température est ainsi détectée entre les deux cellules. L'appareil apporte alors la quantité de chaleur nécessaire pour maintenir les deux cellules à la même température. Ce sont ces compensations de puissance calorifique qui sont enregistrées en fonction du temps et par rapport à la température. On peut alors en déduire la capacité calorifique de la séquence étudiée.

Pour des raisons financières, la concentration minimale utilisable sur la machine a été évaluée à 20  $\mu\text{M}$  d'amorces. À cette concentration, aucune mesure des paramètres thermodynamiques ( $\Delta H$ ,  $\Delta G$  et  $\Delta S$ ) n'a pu être possible mais une détermination précise du  $T_m$  a été réalisée à partir du thermogramme classiquement obtenu (**Figure 3.14**). Le tampon choisi est un tampon phosphate classiquement utilisé dans ce type d'étude (1 mM  $\text{Na}_2\text{EDTA}$ , 10 mM  $\text{Na}_2\text{HPO}_4$ , 1 M de  $\text{NaCl}$ , pH 7) (Breslauer, Frank et al. 1986 ; Vesnaver and Breslauer 1991 ; Schoppe, Hinz et al. 1996). L'échantillon est préparé dans un volume final de 650  $\mu\text{L}$  qui sera dégazé pendant 4 heures afin de réduire la formation de bulles au cours de l'analyse.

Chaque échantillon a été scanné à 4 reprises sous une pression de 300 kPa et sur une plage thermique allant de 20°C à 100°C (1°C/min). Aucune différence n'a été observée entre les différents thermogrammes d'un même échantillon.



**Figure 3.13 :** Dispositif de mesure en cellule capillaire sur CSC 6300 Nano DSC III. Deux cellules capillaires sont disposées au niveau d'une enceinte thermique sous pression, chauffée par un dispositif thermoélectrique (A). La cellule " R " (pour " Reference ") est dédiée à la solution utilisée pour la solubilisation des molécules pouvant rentrer en interaction tandis que la cellule " S " (pour " Sample ") est dédiée à l'échantillon à analyser, constitué des deux molécules en solution (B).



**Figure 3.14 :** Thermogramme obtenu en Calorimétrie Différentielle à Balayage et représentant l'analyse d'un homoduplexe ADN en solution. Le désappariement des deux brins est reconnu comme un pic aigu endothermique centré à une température caractéristique appelée *Transition midpoint* (noté  $T_m$ ), dans notre cas la température de fusion (notée  $T_m$ ). Les  $T_m$ ,  $\Delta H$  et  $\Delta C_p$  de cette transition sont calculées en ajustant les données à un modèle de transition de deux états et en utilisant une analyse de régression non linéaire des moindres carrés.

### 3.7. Quantification relative d'ARN messagers par la méthode du $\Delta\Delta$ de $C_q$

Afin d'évaluer la précision de la quantification génique, trois gènes cibles sont choisis : la *peptidylprolyl isomérase A* ou *cyclophilin A* (PPIA, NM\_017101), l'*hypoxanthine phosphoribosyltransférase* (HPRT, NM\_012583) et le *Cluster de Différentiation 19* (CD19, NM\_001013237). Les conditions et les performances de ces réactions PCR pour les trois tissus expérimentaux extraits par les deux méthodes décrites précédemment sont résumées au niveau des **Tableaux 3.9 et 3.10**. Le niveau d'**expression** de ces gènes est quantifié de manière relative par la méthode du  $\Delta\Delta$  de  $C_q$ .

Cette méthode de quantification a été proposée en 2001 par Kenneth LIVAK (Livak and Schmittgen 2001) dans le journal *Methods*. Il est important de savoir que la précision du résultat de la quantification va dépendre ici des **efficacités de PCR** obtenues pour le gène cible et le gène de référence. Théoriquement, ces **efficacités** doivent être égales à 100%. Cette méthode est donc robuste si les **efficacités de PCR** sont proches. D'autre part, une correction de ces valeurs peut être effectuée (Pfaffl 2001).

Au départ, considérons une efficacité de 100% au cours de la **réaction de polymérisation en chaîne** (Taux d'amplification = 2). On peut décrire la réaction par l'équation suivante :

Nous avons vu en introduction que  $N_n = N_0 \times 2^n$  (1.12)

$$\text{Soit : } N_{C_q} = N_0 \times \varepsilon^{C_q} \quad (3.2)$$

$C_q$  : cycle de quantification

$N_{C_q}$  : nombre de molécules au cycle de quantification  $C_q$

$N_0$  : nombre initial de molécules

$\varepsilon$  : efficacité de la PCR

Les principes de cette méthode sont les suivants :

- Dans cette méthode de quantification relative, il est nécessaire d'utiliser un calibrateur. Cet échantillon particulier est généralement un échantillon d'ADN provenant d'un d'organisme, d'un tissu ou des cellules n'ayant subi aucune modification au niveau génomique, ni aucun traitement expérimental quel qu'il soit. Il représente donc l'état "normal" au niveau d'un plan

expérimental (lignée sauvage, tissu sain ou un échantillon non traité par différents conditionnements expérimentaux). Au laboratoire, ce calibrateur qui peut être constitué par n'importe quel échantillon que l'on souhaite, est obtenu à partir du *pool* de tous les échantillons de l'expérimentation.

Le calibrateur permet ainsi de normaliser les différences de détection entre la séquence cible et la séquence de référence ou de normaliser des expériences faites à différents moments.

- Dans un premier temps, le rapport de quantité gène cible / gène de référence pour chaque échantillon ainsi que pour le calibrateur est calculé.

- Dans un deuxième temps, le rapport précédemment calculé pour chaque échantillon est divisé par le rapport obtenu pour le calibrateur. On élimine alors à ce moment le biais lié aux différences de sensibilité de détection du gène cible et du gène de référence. Une **normalisation** ainsi réalisée avec l'aide d'un calibrateur permet de comparer de manière cohérente les échantillons entre eux.

Au cycle de détection  $C_q$ , pour un gène cible noté **T** et pour un gène de référence noté **R**, on peut donc poser les équations suivantes :

$$N_T = N_{T0} \times \varepsilon_T^{Cq \text{ de } T} \quad (3.3)$$

$$N_R = N_{R0} \times \varepsilon_R^{Cq \text{ de } R} \quad (3.4)$$

Le rapport de ces deux valeurs est une constante, notée **K** :

$$\frac{N_T}{N_R} = \frac{N_{T0} \times \varepsilon_T^{Cq \text{ de } T}}{N_{R0} \times \varepsilon_R^{Cq \text{ de } R}} = K \quad (3.5)$$

Au cours de cette méthode, on présume les efficacités de PCR pour la détection du gène cible **T** et du gène de référence **R** identiques. On peut alors écrire :

$$\varepsilon_T = \varepsilon_R = \varepsilon \quad (3.6)$$

Comme :

$$\frac{X^A}{X^B} = X^{(A-B)} \quad (2.7) \quad \text{et} \quad \varepsilon_T = \varepsilon_R \quad (3.8)$$

On peut écrire :

$$K = \frac{N_{T0}}{N_{R0}} \times \varepsilon^{(Cq_T - Cq_R)} \quad (3.9)$$

Ainsi, si l'on considère un échantillon **S** à doser et un calibrateur **C**, on peut écrire :

$$K_{\text{échantillon}} = \frac{N_{T0_S}}{N_{R0_S}} \times \varepsilon^{(Cq_T - Cq_R)} \quad (3.10)$$

$$\text{et } K_{\text{calibrateur}} = \frac{N_{T0_C}}{N_{R0_C}} \times \varepsilon^{(Cq_T - Cq_R)} \quad (3.11)$$

Avec :

$$\frac{K_{\text{échantillon}}}{K_{\text{calibrateur}}} = \text{Constante} \quad (3.12)$$

Notons  $X_{\text{échantillon}}$ , la quantité de gène cible normalisée au départ de la réaction  $\left(\frac{N_{T0_S}}{N_{R0_S}}\right)$  et  $\Delta C_q$

la différence de cycles de quantification entre le gène cible et le gène de référence pour cet échantillon :

$$K_{\text{échantillon}} = X_{\text{échantillon}} \times \varepsilon^{(\Delta C_q)} \quad \text{soit} \quad X_{\text{échantillon}} = K_{\text{échantillon}} \times \varepsilon^{-(\Delta C_q)} \quad (3.13)$$

On a donc aussi :

$$K_{\text{calibrateur}} = X_{\text{calibrateur}} \times \varepsilon^{(\Delta C_q)} \quad \text{soit} \quad X_{\text{calibrateur}} = K_{\text{calibrateur}} \times \varepsilon^{-(\Delta C_q)} \quad (3.14)$$

Le rapport des quantités normalisées détectées pour l'échantillon et le calibrateur nous donne la quantité de gène cible normalisée par un gène de référence de manière relative à un calibrateur :

$$\frac{X_{\text{échantillon}}}{X_{\text{calibrateur}}} = \frac{K_{\text{échantillon}} \times \varepsilon^{-(\Delta C_q)}}{K_{\text{calibrateur}} \times \varepsilon^{-(\Delta C_q)}} = \varepsilon^{-\Delta \Delta C_q} \quad (3.15)$$

$$\text{avec : } -\Delta \Delta C_q = -\left(\Delta C_{q_{\text{gène cible échantillon}}} - \Delta C_{q_{\text{gène cible calibrateur}}}\right) \quad (3.16)$$

Soit :

$$X_{\text{gène cible échantillon}} = 2^{-\Delta \Delta C_q} \quad (3.17)$$

Dans la réalité, les **efficacités des PCR** pour le gène cible et le gène de référence sont rarement identiques et il faut corriger la quantification (Pfaffl 2001). On peut le faire de la manière suivante :

$$X_{\text{gène cible échantillon}} = \frac{\left(\varepsilon_{\text{gène cible}}\right)^{\Delta C_{q_{\text{gène cible}}(\text{Calibrateur} - \text{Echantillon})}}{\left(\varepsilon_{\text{gène référence}}\right)^{\Delta C_{q_{\text{gène référence}}(\text{Calibrateur} - \text{Echantillon})}}} \quad (3.18)$$

Un ratio d'expression peut aussi être défini (Stahlberg, Aman et al. 2003) comme suit :

$$\frac{N_T}{N_R} = K_{RS} \times \frac{\eta_R (1 + E_R)^{C_{T_R} - 1}}{\eta_T (1 + E_T)^{C_{T_T} - 1}} \quad (3.19)$$

où  $\eta$  est le rendement de RT,  $E$  est l'efficacité PCR,  $C_T$  est le cycle seuil et  $K_{RS}$  est la sensibilité relative qPCR

**Tableau 3.9 :** Caractéristiques des amorces de qPCR utilisées pour l'étude de la variabilité inter-échantillons. Les oligonucléotides sont synthétisés puis purifiés par Chromatographie en phase Liquide à Haute Performance, HPLC (Eurogentec, Saraing, Belgique).

Gène	Oligonucléotides	Séquence 5'P – 3'OH	Localisation	Amplicon (pb)	T <sub>m</sub> (°C)	Numéro d'accèsion GenBank
<i>Cluster of Differentiation 19, CD19</i>	CD19 FW (2)	TATGCAGCTCCTCAGCTCCAATC	1516-1538	97	70	NM_001013237
	CD 19 BW (2)	GGCTCTGGGTCGTCAGATTTATC	1613-1591		70	
<i>Hypoxanthine phosphoribosyltransferase, HPRT</i>	HPRT FW (1)	CTCATGGACTGATTATGGACAGGAC	113-137	122	74	NM_012583
	HPRT BW (1)	GCAGGTCAGCAAAGAACTTATAGCC	235-211		74	
<i>peptidylprolyl isomerase A PPIA</i>	PPIA FW (3)	GGCAAATGCTGGACCAAACAC	342-362	93	64	NM_017101
	PPIA BW (3)	CTTCCCAAAGACCACATGCTTG	435-414		66	
<i>Ribosomal protein, large, P0, RPLP0</i>	RPLP0 FW (1)	CCTGCACACTCGCTTCCTAGAG	659-680	73	70	NM_022402
	RPLP0 BW (1)	CAACAGTCGGGTAGCCAATCTG	732-711		68	
	RPLP0 FW (3)	GGCGACCTGGAAGTCCAATA	47-67	117	66	
	RPLP0 BW (3)	CATGCGGATCTGCTGCATCT	164-145		62	

A		Conditions de qPCR			Paramètres de calibration de la courbe			
		Gène	T <sub>H</sub> °C	[Primer] $\mu$ M	[MgCl <sub>2</sub> ] mM	E <sub>PCR</sub> <sup>a</sup>	pente	Y intercept
Tissu splénique Extraction organique	ARBP 3'	57	0,5	5	98,8	-3,35	22,74	0,998
	ARBP 5'	55	0,5	3	104,4	-3,22	23,66	0,999
	HPRT	60	0,5	4	98,4	-3,36	27,35	0,983
	PPIA	56	0,5	5	102,6	-3,26	23,26	0,990
	CD 19	54	0,4	4	104,8	-3,21	28,65	0,997

B		Conditions de qPCR			Paramètres de calibration de la courbe			
		Gène	T <sub>H</sub> °C	[Primer] $\mu$ M	[MgCl <sub>2</sub> ] mM	E <sub>PCR</sub> <sup>a</sup>	pente	Y intercept
Tissu splénique Extraction / colonne	ARBP 3'	57	0,5	3	94,9	-3,45	25,29	0,999
	ARBP 5'	55	0,5	4	100,9	-3,30	25,88	0,999
	HPRT	60	0,4	3	91,3	-3,55	31,67	0,997
	PPIA	56	0,5	4	98,0	-3,37	26,15	0,996
	CD 19	56	0,4	3	99,7	-3,33	29,87	0,995

C		Conditions de qPCR			Paramètres de calibration de la courbe			
		Gène	T <sub>H</sub> °C	[Primer] $\mu$ M	[MgCl <sub>2</sub> ] mM	E <sub>PCR</sub> <sup>a</sup>	pente	Y intercept
Cervelet Extraction organique	ARBP 3'	57	0,5	4	97,2	-3,39	28,71	0,997
	ARBP 5'	55	0,4	4	102,6	-3,26	29,60	0,992
	HPRT	60	0,4	4	101,3	-3,29	33,88	0,985
	PPIA	56	0,5	4	89,2	-3,61	27,06	0,997

D		Conditions de qPCR			Paramètres de calibration de la courbe			
		Gène	T <sub>H</sub> °C	[Primer] $\mu$ M	[MgCl <sub>2</sub> ] mM	E <sub>PCR</sub> <sup>a</sup>	pente	Y intercept
Cervelet Extraction / colonne	ARBP 3'	57	0,5	4	88,6	-3,63	29,95	0,993
	ARBP 5'	55	0,4	4	90,2	-3,58	30,97	0,991
	HPRT	60	0,4	4	94,5	-3,46	33,77	0,992
	PPIA	56	0,5	4	109,7	-3,11	26,56	0,998

E	Conditions de qPCR			Paramètres de calibration de la courbe				
	Gène	T <sub>H</sub> °C	[Primer] $\mu$ M	[MgCl <sub>2</sub> ] mM	E <sub>PCR</sub> <sup>a</sup>	pen	Y intercept	r <sup>b</sup>
Tissu musculaire Extraction organique	ARBP 3'	57	0,5	4	91,6	-3,54	27,47	0,995
	ARBP 5'	55	0,4	4	96,1	-3,42	28,04	0,999
	HPRT	60	0,4	4	100,5	-3,31	32,00	0,991
	PPIA	56	0,5	4	96,5	-3,41	28,73	1,000

F	Conditions de qPCR			Paramètres de calibration de la courbe				
	Gène	T <sub>H</sub> °C	[Primer] $\mu$ M	[MgCl <sub>2</sub> ] mM	E <sub>PCR</sub> <sup>a</sup>	pen	Y intercept	r <sup>b</sup>
Tissu musculaire Extraction / colonne	ARBP 3'	57	0,5	4	104,9	-3,21	27,12	0,995
	ARBP 5'	55	0,4	4	95,3	-3,44	27,96	0,998
	HPRT	60	0,4	4	91,6	-3,54	33,14	0,995
	PPIA	56	0,5	4	102,2	-3,27	29,31	0,990

**Tableaux 3.10 :** Conditions de PCR et caractéristiques des courbes de calibration réalisées à partir d'un *pool* d'échantillons d'ADNc dilué en cascades avec un volume de départ de 5 $\mu$ L de mix RT, dilué jusqu'à 10 fois (5 différentes dilutions analysées en duplicate) pour le tissu splénique (A-B) pour le cervelet (C-D) et le tissu musculaire (E-F) en extraction organique et extraction sur colonne de silice. <sup>a</sup> pourcentage d'efficacité PCR déterminée à partir de la pente de la courbe de calibration, <sup>b</sup> coefficient de régression linéaire de la courbe de calibration.

### 3.8. Utilisation du logiciel *geNorm*

La **normalisation** des données est une étape incontournable en **RT-qPCR** car elle permet de corriger les différences techniques inter-échantillons (Bustin, Benes et al. 2009). La **normalisation** interne, qui consiste à utiliser une référence présente dans l'échantillon, est la plus utilisée (Huggett, Dheda et al. 2005). Cette référence est le plus souvent un gène dont l'**expression** est supposée stable dans les échantillons. Le choix de ce gène est problématique. En supposant qu'un tel gène existe, il n'est pas possible de prouver sa stabilité. En effet, l'absence de différence significative pour l'**expression** de ce gène ne peut suffire à démontrer qu'il est stable. Si l'**expression** est pratiquement constante, rien ne permet de démontrer que les rendements des différentes étapes de la **RT-qPCR** sont identiques. Si l'absence de significativité est due à une variabilité importante entre les échantillons, ce qui est le plus souvent le cas, il n'est pas démontré que les rendements sont proportionnels à l'**expression** du gène de référence. Pour compenser cet inconvénient, on utilise comme référence des index multigéniques qui sont basés sur l'**expression** relative des gènes le composant. Si l'**expression** des gènes retenus est proportionnelle au niveau de tous les échantillons, on peut utiliser l'index qui en découle pour corriger les différences de rendements inter-échantillons : *geNorm* est actuellement la méthode de référence multi-gène en **normalisation** interne (Vandesompele, De Preter et al. 2002). Il utilise trois étapes. La première consiste à classer la stabilité relative de chaque gène parmi les gènes testés. La deuxième consiste à évaluer le nombre de gènes nécessaires à la **normalisation** ; on ajoute des gènes dans l'ordre de stabilité décroissante jusqu'à ce que l'addition du dernier gène ne modifie pas la **normalisation**. Enfin, on calcule le facteur de **normalisation** basé sur la moyenne géométrique des gènes retenus.

Dans ce travail, nous avons détourné *geNorm* de son utilisation classique *i.e.* démontrer que certains gènes de références sont stables. En l'utilisant sur des répliques techniques, dans lesquels l'**expression** des gènes est identique, il nous a permis de quantifier les différences de rendements techniques entre échantillons.

### 3.9. Analyses statistiques

Les analyses statistiques ont été réalisées avec le logiciel *Prism version 4.0* (GraphPad Software, La Jolla, CA USA). Les comparaisons ont été réalisées en utilisant le test de *Wilcoxon*. Les corrélations linéaires ont été évaluées en utilisant le test du rang de *Spearman*. La significativité a été acceptée à  $p < 0,05$ . L'agrément entre deux différents dosages a été évalué à l'aide de la méthode de *Bland-Altman*, par comparaison des ratios. Les valeurs sont présentées sous la forme d'une moyenne  $\pm$  écart-type. La variabilité est exprimée par le coefficient de variation qui est la valeur absolue du ratio entre l'écart-type et la moyenne.

#### 3.9.1. Méthode de Bland-Altman

Encore connue sous le nom de représentation de Tukey, cette approche proposée par John Martin BLAND et Douglas ALTMAN pour la première fois dans la revue *The Statistician* en 1983 (Altman and Bland 1983) est une analyse très utilisée en statistiques médicales (Bland and Altman 1986 ; Altman 2002) pour comparer deux mesures cliniques. À la différence de la corrélation linéaire, elle présente l'avantage de comparer deux séries de mesure d'un même paramètre obtenues par deux méthodes différentes en évaluant l'agrément ou l'accord pouvant exister entre ces méthodes. Elle est définie par deux importants paramètres :

- **le biais** : dans l'hypothèse d'une concordance parfaite entre deux séries de valeurs, la somme des différences arithmétiques entre ces valeurs devrait être nulle. Ainsi, on définit le biais comme la moyenne des différences. Il représente la moyenne de l'écart systématique d'une série de valeurs par rapport à une autre.

- **les limites d'agrément** : On détermine un intervalle d'agrément autour du biais. Il est possible de fixer deux seuils de limites de concordance, arbitrairement représentés par le biais -  $1,96 \times SD$  et le biais +  $1,96 \times SD$ . Ils correspondent approximativement à deux écarts-types par rapport au biais et prennent en compte l'intervalle dans lequel sont comprises 95% des différences, dans l'hypothèse que leur distribution suive une loi normale. L'ampleur de l'intervalle d'agrément 95% atteste de la concordance entre les méthodes.

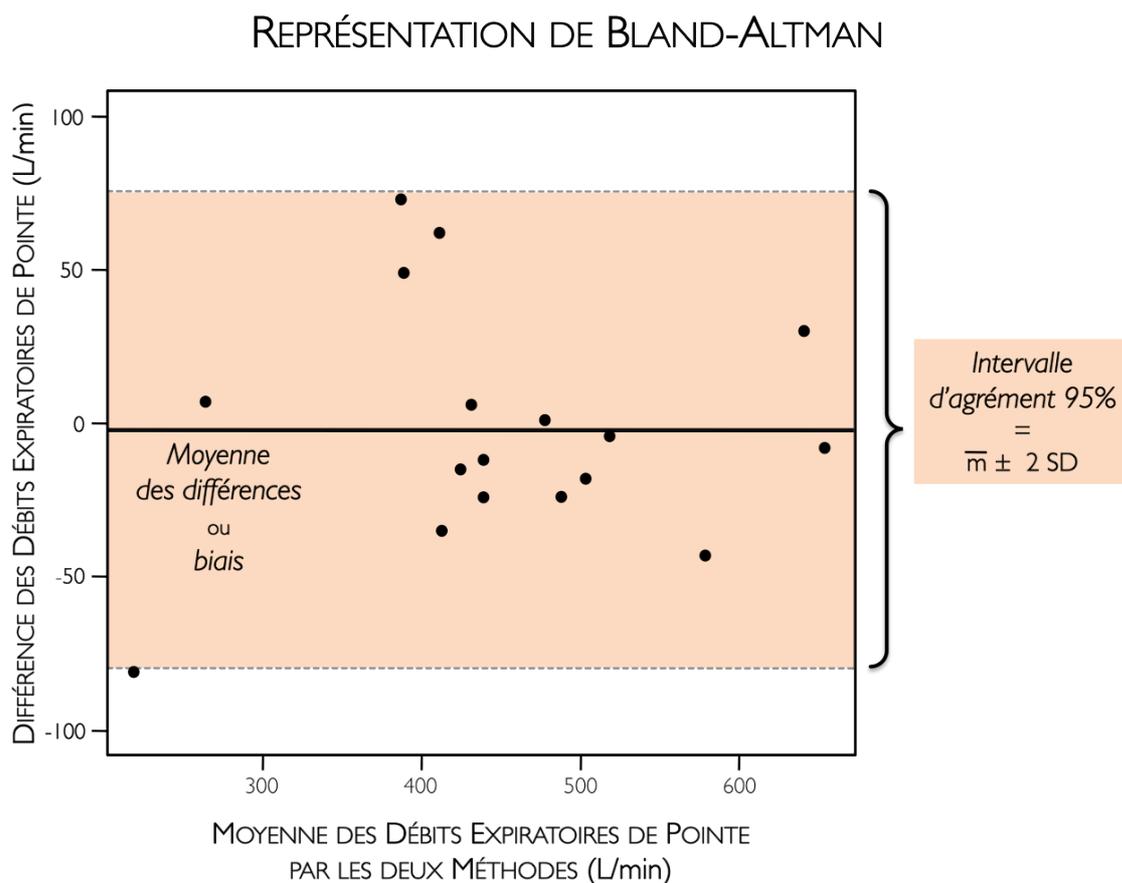


Figure 3.16 : Représentation graphique de la méthode de comparaison statistique de *Bland-Altman*.

Pour un échantillon  $S$  caractérisé par les valeurs  $S_1$  et  $S_2$ , on peut comparer le biais soit par rapport à une différence, soit par rapport à un ratio :

$$S_{(x,y)} = \left( \frac{S_1 + S_2}{2}, (S_1 - S_2) \right) \quad (3.22)$$

$$S_{(x,y)} = \left( \frac{S_1 + S_2}{2}, \left( \frac{S_1}{S_2} \right) \right) \quad (3.23)$$

Les deux types d'analyse ont été rencontrés au cours de ces travaux de thèse. L'étude de la température de fusion des oligonucléotides utilisés en PCR nous a amené à comparer les valeurs obtenues par différentes méthodes en utilisant les différences des valeurs. La

comparaison de quantités détectées en qPCR après différentes méthodes de transcription inverse nous a amené quant à elle à utiliser le ratio des valeurs.

### 3.9.2. Les raisons de son utilisation

La méthode de la régression linéaire caractérisée par le coefficient de corrélation est très souvent utilisée dans le but d'évaluer une concordance. Cette méthode n'est pas utilisable car elle mesure l'existence d'une liaison entre les deux séries de mesures mais absolument pas la concordance des deux techniques pour chaque cas mesuré. Or, cette liaison existe, car il s'agit de deux mesures des mêmes échantillons. La méthode de comparaison de *Bland-Altman* permet d'évaluer la concordance entre les deux méthodes de mesures (Grenier, Dubreuil et al. 2000).

Le cadre de son application est la comparaison de deux méthodes fournissant chacune des erreurs dans leur mesure. Cette méthode représente l'outil le plus adapté pour l'appréciation d'une nouvelle technique de mesure (dans notre cas, la mesure directe du  $T_m$  par courbe de fusion) par rapport à une technique de référence (nano Calorimétrie Différentielle à balayage).

## 3.10. Prélèvements animaux

---

Au cours de ces travaux de doctorat, plusieurs prélèvements tissulaires (rates, muscles squelettiques et cervelets) ont été prélevés chez le rat Sprague Dawley (*Rattus norvegicus*, Centre d'élevage Roger Janvier, Le Genest Saint Isle, France) ayant subi différents conditionnements expérimentaux. Tous ces prélèvements ont été obtenus après anesthésie complète de l'animal et ont ensuite été manipulés en vue des analyses de biologie moléculaire qui vont être décrites dans les lignes suivantes. Toutes ces expérimentations, réalisées par les équipes de recherche en collaboration avec le Service de Biologie Moléculaire, ont été approuvées par le Comité d'éthique de l'Institut Émile Pardé (IRBA - antenne La Tronche, France), conformément aux réglementations communautaires françaises et européennes.

# Variabilité Inter-Échantillons & Inhibition en RT-qPCR

---

## 4.1. Introduction

En raison de sa grande simplicité d'utilisation et de la précision de ses résultats, la technique de **Reverse Transcription - Réaction de Polymérisation en Chaîne en temps réel (RT-qPCR)** est à l'heure actuelle la méthode la plus performante concernant la quantification du niveau d'**expression d'ARN messagers (ARNm)**. En effet, depuis les débuts de la **PCR**, cette technique n'a cessé de se développer jusqu'au statut indéniable de méthode de référence en matière de quantification d'**expression** génique. Inévitablement, cet essor s'est accompagné de plusieurs dérives et des préoccupations essentielles ont récemment été soulevées quant à la fiabilité des données obtenues en **RT-qPCR** (Bustin 2010). Aujourd'hui, une meilleure compréhension des limites de cette méthode constitue l'un des principaux axes de recherche dans le domaine de la biologie moléculaire. Des efforts conséquents ont dernièrement été apportés afin d'améliorer la fiabilité de la quantification par cette méthode (Bustin, Benes et al. 2009). Ainsi, il est désormais établi avec certitude qu'une large part de la variabilité technique observée au cours de ces expérimentations tire son origine d'une étape particulièrement critique au cours de l'analyse : la **transcription inverse (RT)** (Stahlberg, Hakansson et al. 2004 ; Stahlberg, Kubista et al. 2004 ; Bustin 2005). En effet, les variations de rendements de synthèse que cette réaction peut régulièrement présenter pour chaque échantillon diminuent considérablement la précision des résultats de la quantification. La publication régulière de travaux effectués depuis peu sur la capacité de différents facteurs à influencer la synthèse de l'**ADN complémentaire (ADNc)** à partir des ARN purifiés illustre bien l'intérêt porté sur ce sujet. Concentration totale d'ARN utilisée (Curry, McHale et al. 2002; Stahlberg, Hakansson et al. 2004), quantité de **matrice** cible à détecter (Karrer, Lincoln et al. 1995), qualité de l'ARN (Fleige and Pfaffl 2006) ou bien encore composition de la **matrice** utilisée au départ de la réaction (Levesque-Sergerie, Duquette et al. 2007) sont ainsi autant d'exemples retrouvés dans la littérature scientifique.

Par ailleurs, de nombreux facteurs sont susceptibles d'inhiber les réactions de **RT-qPCR** (Wilson 1997), à la fois dépendants des particularités biochimiques de la **matrice** utilisée d'une part, et du processus technique employé afin de l'obtenir d'autre part. Abordons dans un premier temps les singularités pouvant caractériser la **matrice** à utiliser. Il apparaît évident que la qualité de l'extrait d'ARN dépend considérablement de la composition biochimique de l'échantillon à analyser, plus précisément de la proportion en inhibiteurs qu'il peut présenter. On connaît ainsi l'existence de nombreux inhibiteurs retrouvés au niveau de différents organes ou

tissus d'un organisme, pouvant être copurifiés au cours de l'étape d'extraction. Le calcium, un composant inorganique majeur du tissu osseux, est par exemple reconnu comme source majeure d'inhibition de la PCR (Bickley, Short et al. 1996 ; Eilert and Foran 2009 ; Opel, Chung et al. 2010). Le cofacteur hème, notamment retrouvé au niveau de l'hémoglobine et de la myoglobine, est lui aussi considéré comme un important inhibiteur de PCR (Akane, Matsubara et al. 1994). Des niveaux d'inhibition remarquables de la réaction ont donc été identifiés au niveau du muscle squelettique (Belec, Authier et al. 1998), également composé d'un autre important inhibiteur, le collagène (Kim, Labbe et al. 2000). Au niveau du tissu sanguin, d'abondants inhibiteurs de PCR ont été purifiés puis caractérisés (Mercier, Gaucher et al. 1990; Al-Soud and Radstrom 2001 ). La rate constitue donc encore un autre tissu potentiellement riche en inhibiteurs, de par son rôle joué dans la destruction des globules rouges d'une part, mais aussi par son rôle dans l'immunité (Al-Soud, Jonsson et al. 2000) d'autre part. Enfin, le système nerveux, riche en lipides et phospholipides, représente un dernier exemple de source potentielle d'inhibiteurs de RT-qPCR (Rossen, Norskov et al. 1992).

Dans un deuxième temps, l'inhibition de la RT-qPCR peut aussi dépendre de la technique employée lors de la purification des acides nucléiques. Ainsi, les techniques basées sur l'extraction par des solvants organiques présentent un risque potentiel d'inhibition. En effet, bien qu'il soit reconnu pour éliminer efficacement les inhibiteurs hydrophobes endogènes tels que les lipides, le phénol reste un inhibiteur bien caractérisé en PCR (Wiedbrauk, Werner et al. 1995). D'une autre manière, les méthodes de purification sur colonnes de silice, désormais très répandues, se révèlent plus efficaces que les méthodes d'extraction organique dans l'élimination d'un inhibiteur tel que le collagène. Les récents travaux de KEMP *et al.* (2006) illustrent d'ailleurs très bien ces qualités remarquables lorsque sont manipulés d'anciens échantillons d'ADN (Kemp, Monroe et al. 2006). Cependant, elles peuvent se montrer insuffisantes notamment lorsque certains inhibiteurs, provenant des réactifs utilisés au cours de la technique, finissent par se retrouver au niveau de l'extrait d'acides nucléiques. De plus, ces colonnes semblent être beaucoup moins efficaces que les réactifs d'origine organique pour éliminer les inhibiteurs hydrophobes tels que les lipides. Pour cette raison, les fabricants recommandent souvent une première étape de traitement par des solvants organiques avant tout passage de l'extrait à travers la colonne.

La **standardisation** est un facteur clé dans la diminution des variations inter-échantillons. Afin de réduire efficacement ces dernières, les inhibiteurs dépendants de l'étape de purification

des acides nucléiques doivent être retrouvés - dans la mesure du possible - à concentration constante au niveau de tous les échantillons à analyser. Il est néanmoins peu évident d'atteindre un tel objectif, une **normalisation** idéale n'étant pas toujours possible. En raison d'exigences liées au plan expérimental et aux techniques mises en œuvre durant l'expérimentation, des compromis doivent fréquemment être trouvés. D'après le processus de **normalisation** le plus logique, la quantité d'inhibiteurs devrait ainsi être fonction du volume de réactif utilisé. Or, le consensus actuel stipule qu'une quantité constante d'**ARN totaux** doit être utilisée lors de la réalisation de l'étape de **RT** (on nommera cette procédure *Constant Amount - Reverse Transcription*, CA-RT) (Stahlberg, Hakansson et al. 2004 ; Stahlberg, Kubista et al. 2004 ; Bustin 2005 ; Bustin, Benes et al. 2009). Cette règle impose donc l'utilisation de volumes différents d'extrait d'ARN en raison de la variabilité observée au niveau des rendements d'extraction d'ARN. Malheureusement, cela suggère également que des quantités variables d'inhibiteurs provenant à la fois de la préparation de l'échantillon et de la méthode d'extraction utilisée soient retrouvées au niveau des extraits d'ARN. Il est donc plus probable qu'une contamination par le phénol dépende du volume de la solution d'ARN extrait plutôt que de sa concentration. Par ailleurs, il n'existe aucune preuve établie concernant une relation éventuelle entre concentration d'inhibiteurs endogènes copurifiés et rendement d'extraction d'ARN. Par conséquent, une utilisation de volumes variables d'extrait d'ARN (CA-RT) ou de quantités variables d'ARN, en tant que prise d'essai pour l'étape de **RT**, reste entièrement discutable. En effet, un compromis pourrait être la réalisation d'une réaction de **RT** à partir d'un volume constant d'extrait d'ARN (on nommera cette procédure *Constant Volume - Reverse Transcription*, CV-RT) tout en conservant une quantité constante d'ARN entre tous les échantillons par supplémentation en **ARN de transfert (ARNt)** de levure (Stahlberg, Hakansson et al. 2004). Cette méthode devrait nous permettre d'obtenir un rendement constant de la **RT** tout en minimisant les variations inter-échantillons en inhibiteurs-volume dépendants.

Le but de l'étude que nous proposons dans ce chapitre est donc de déterminer dans un premier temps si une CV-RT pourrait sensiblement améliorer la fiabilité des résultats obtenus en **RT-qPCR** en confrontant cette procédure originale à celle classiquement utilisée en laboratoire (CA-RT). Cette comparaison nous oblige d'utiliser de véritables échantillons biologiques présentant des niveaux d'**expression d'ARNm** cibles totalement inconnus et non des **matrices** synthétiques contrôles utilisées précédemment (Levesque-Sergerie, Duquette et al. 2007). L'utilisation de répliques techniques nous a, quant à elle, permis de manipuler des échantillons exprimant tous des quantités identiques de l'**ARNm** cible, bien qu'inconnues. Afin

d'évaluer l'agrément entre les méthodes, nous avons choisi d'utiliser des répliques biologiques présentant une gamme plus importante de niveaux d'expression d'ARNm cibles. Dans un premier temps, nous avons généré différentes matrices d'ARN caractérisées par des types et des concentrations d'inhibiteurs volume-dépendants différents en faisant varier la nature de l'échantillon et la technique de purification employée. Le calcul de l'efficacité de PCR représentant un moyen très sensible et statistiquement robuste pour identifier des matrices contaminées par des inhibiteurs de PCR (Kontanis and Reed 2006), nous avons ensuite étudié les variations inter-échantillons en observant respectivement leurs impacts sur les efficacités de réaction de la RT et de la qPCR.

## 4.2. Résultats expérimentaux

### 4.2.1. Évaluation de la qualité de l'ARN

Au cours de cette étude, deux principaux critères de qualité ont été pris en compte : (1) l'intégrité des molécules d'acides nucléiques, évaluée par le *RNA Integrated Number* (RIN) et (2) la pureté de l'extrait d'ARN, évaluée grâce au rapport  $Abs_{260} / Abs_{280}$  obtenu par spectrophotométrie.

- **Intégrité de l'ARN**

Calculées à partir des électrophorogrammes réalisés sur *Bioanalyser 2100*, les valeurs du RIN obtenues pour la rate, le cervelet et le muscle squelettique sont respectivement de  $7,76 \pm 0,18$  ;  $8,35 \pm 0,15$  et  $7,98 \pm 0,56$ . On observe un écart-type un peu plus élevé pour le tissu musculaire du à une légère différence entre la purification de l'ARN à partir de la méthode d'extraction organique et la méthode d'extraction sur colonne de silice ( $7,55 \pm 0,39$  vs.  $8,42 \pm 0,30$ ). Ces valeurs d'intégrité élevées attestent de la qualité et donc de la validité de tous les échantillons d'ARN utilisés pour les études de RT-qPCR effectuées (Fleige and Pfaffl 2006; Schroeder, Mueller et al. 2006). Les faibles valeurs d'écart-type reflètent quant à elles l'homogénéité de chaque série expérimentale.

- **Pureté de l'ARN**

La pureté de tous les échantillons d'ARN a été évaluée en utilisant le ratio d'absorbances  $A_{260}/A_{280}$ . Ces ratios sont respectivement de  $1,98 \pm 0,07$  et de  $2,05 \pm 0,06$  pour la méthode d'extraction organique et la méthode d'extraction sur colonne de silice. Ces valeurs témoignent de la grande pureté et de l'homogénéité de tous les échantillons (Maniatis, E.F. et al. 1989). Comme il était prévisible, une différence entre les ratios ( $P < 0,001$ ) suggère la présence de traces de composés aromatiques, probablement du phénol, retrouvée dans les extraits d'ARN obtenus par méthode d'extraction organique. Comme l'indique le **Tableau 4.1**, les rendements de purification de l'ARN dépendent à la fois du tissu considéré et de la méthode d'isolement utilisée. Quels que soient les tissus analysés, la méthode d'extraction organique permet des rendements de purification plus élevés que ceux obtenus à partir de la méthode d'extraction sur colonne de silice (respectivement  $P < 0,001$  ;  $P < 0,01$  et  $P < 0,01$  pour la rate, le cervelet et le muscle squelettique) mais conduit à une plus grande variabilité. Cette variabilité est plus élevée dans le muscle squelettique et dans la rate que dans le cervelet. Les gammes de

variation du volume d'extraits d'ARN utilisées pour réaliser la CA-RT nous permettent de comparer de manière pertinente la CA-RT avec la CV-RT.

	Extraction organique avec <i>clean-up</i>			Extraction sur colonne de silice		
	(ng/ $\mu$ L) <sup>a</sup>	RSD <sup>b</sup> (%)	Gamme de volumes <sup>c</sup>	(ng/ $\mu$ L) <sup>a</sup>	RSD <sup>b</sup> (%)	Gamme de volumes <sup>c</sup>
<b>Rate</b>	978 $\pm$ 303	31	1,0 – 2,7	537 $\pm$ 101	19	1,0 – 1,7
<b>Cervelet</b>	220 $\pm$ 53	24	1,0 – 1,9	99 $\pm$ 15	15	1,0 – 1,5
<b>Muscle</b>	123 $\pm$ 45	37	1,0 – 2,8	65 $\pm$ 15	23	1,0 – 1,6

**Tableau 4.1 :** Rendements d'extraction obtenus avec les deux méthodes de purification de l'ARN utilisées. L'ARN total est obtenu à partir de 50 mg de tissu et élué dans 50  $\mu$ L d'eau dépourvue de RNase. Protocole de broyage identique pour les 3 tissus. <sup>a</sup> Les concentrations d'ARN totaux sont mesurées à  $\lambda = 260$  nm. Les résultats sont donnés sous la forme d'une moyenne  $\pm$  écart-type. <sup>b</sup> La variabilité est exprimée par le coefficient défini comme le ratio entre l'écart-type et la moyenne. <sup>c</sup> Gamme de variation du volume d'extraits d'ARN utilisé en  $\mu$ L pour réaliser la RT à quantité constante.

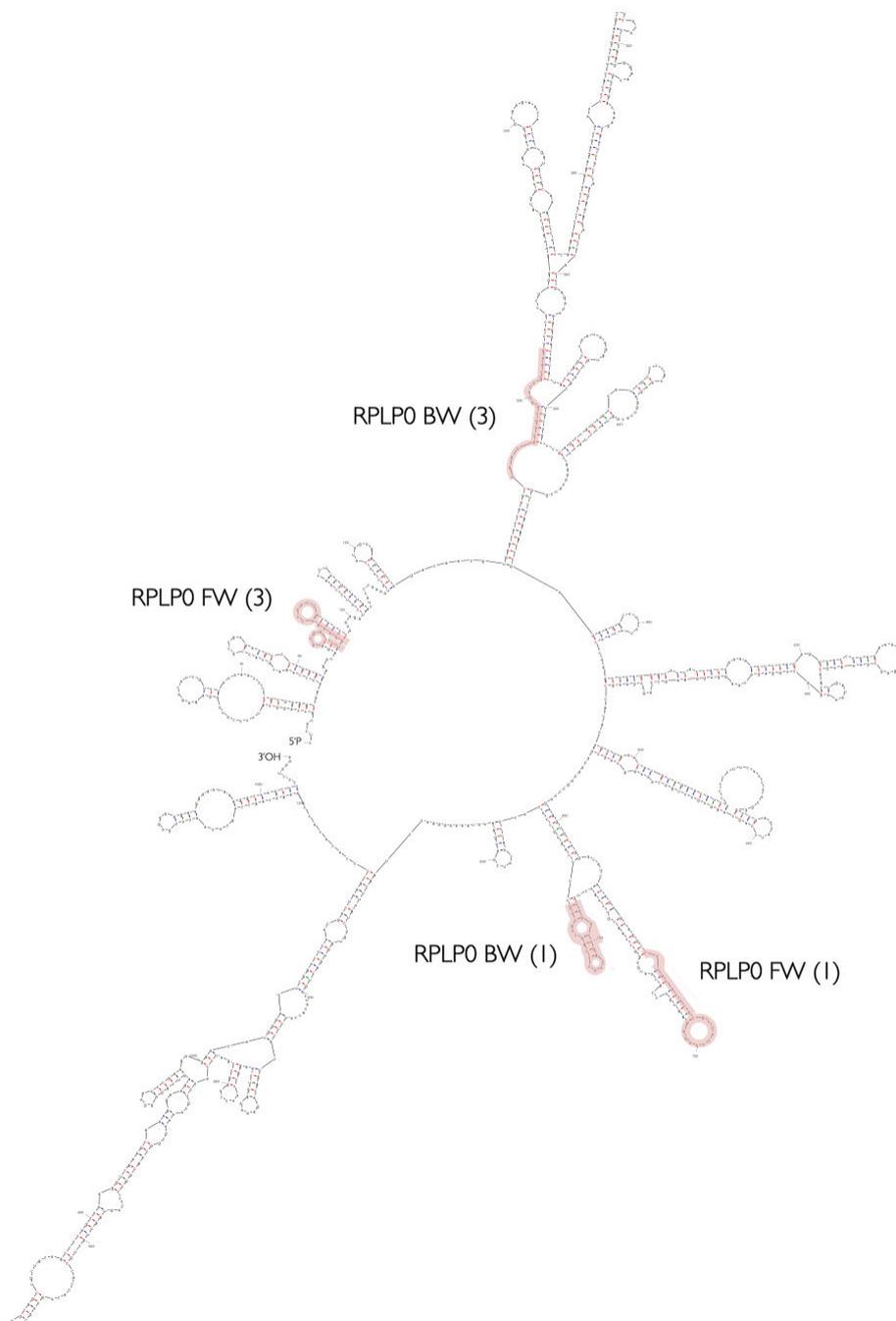
#### 4.2.2. Évaluation de la qualité de la RT

À l'origine, l'analyse 3':5' proposée par NOLAN *et al.* (2006) permet de mesurer l'intégrité de l'ARNm, produit du gène de la *Glyceraldehyde 3-phosphate dehydrogenase* chez l'homme (GAPDH, GenBank NM\_002046). Ce gène est généralement exprimé de manière ubiquitaire par les cellules. L'intégrité de l'ARN est alors évaluée au moyen de deux couples d'amorces PCR placés respectivement au niveau de la région 3'OH et 5'P du transcrit (Nolan, Hands *et al.* 2006).

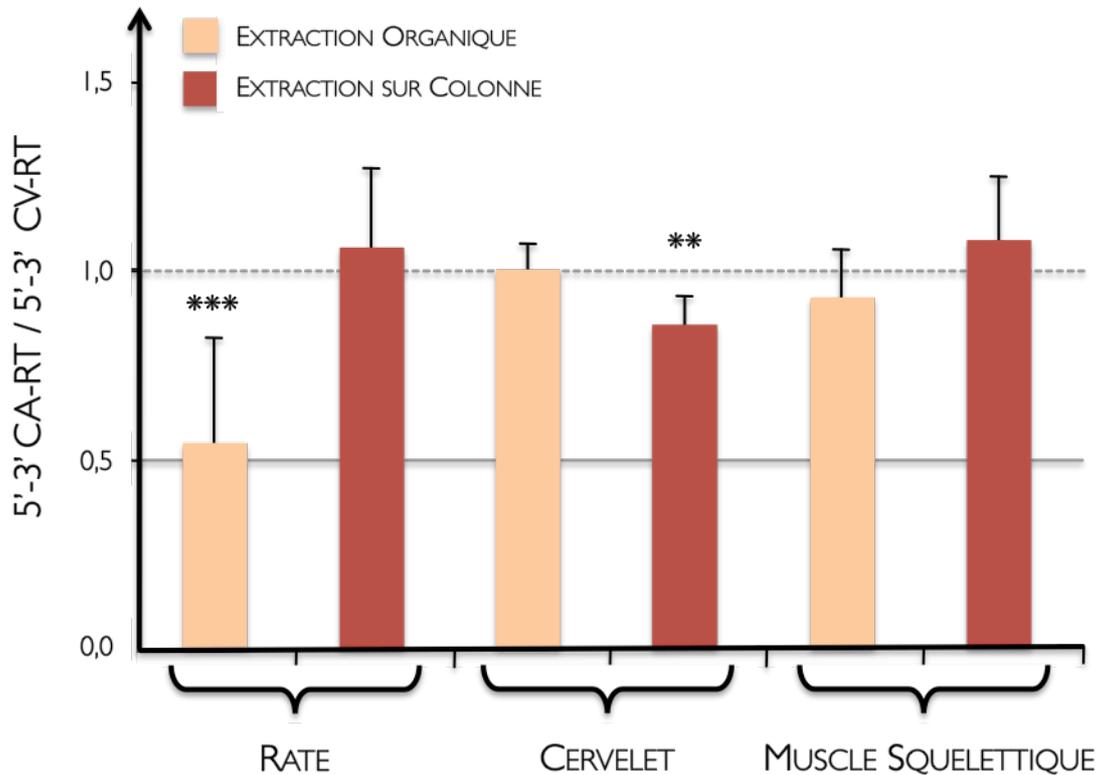
Cette analyse permet d'évaluer l'efficacité de la *transcription inverse* lorsqu'un extrait d'ARN unique non dégradé est utilisé dans des conditions de RT différentes. Dans notre étude, deux couples d'amorces ont donc été dessinés afin de permettre l'amplification des parties 3'OH et 5'P de l'ARNm du gène de la protéine ribosomale P0 : *ribosomal protein, large, P0* (Rplp0, GenBank NM\_022402). La réaction de RT n'étant amorcée que par des oligo-dT spécifiques des ARNm matures, nous avons mesuré les niveaux de quantification de deux amplicons placés respectivement au niveau de la région 3'OH et 5'P de la séquence du transcrit ciblé. Les deux réactions de PCR sont optimisées et effectuées séparément. Le calcul du ratio des amplicons permet alors de rendre compte du rendement de la réaction de synthèse de l'ADNc qui se

réalise à partir de la queue poly-A (**extrémité 3'OH**) et cela, tout au long de la longueur du **transcrit**. Plus clairement, une diminution de ce ratio permet de mesurer les problèmes rencontrés par la **transcriptase inverse** pendant la synthèse du brin d'**ADN complémentaire**.

L'analyse de la séquence de ce gène par l'algorithme *mfold* (Zuker 1999 ; Zuker 2003) met précisément en évidence la présence d'importantes **structures secondaires** au niveau de la molécule d'ARN et situées entre les deux amplicons cibles. La **Figure 4.1** illustre cette simulation qui est réalisée à la température de 48°C, température de synthèse de l'**ADNc** détecté par nos **PCR**.



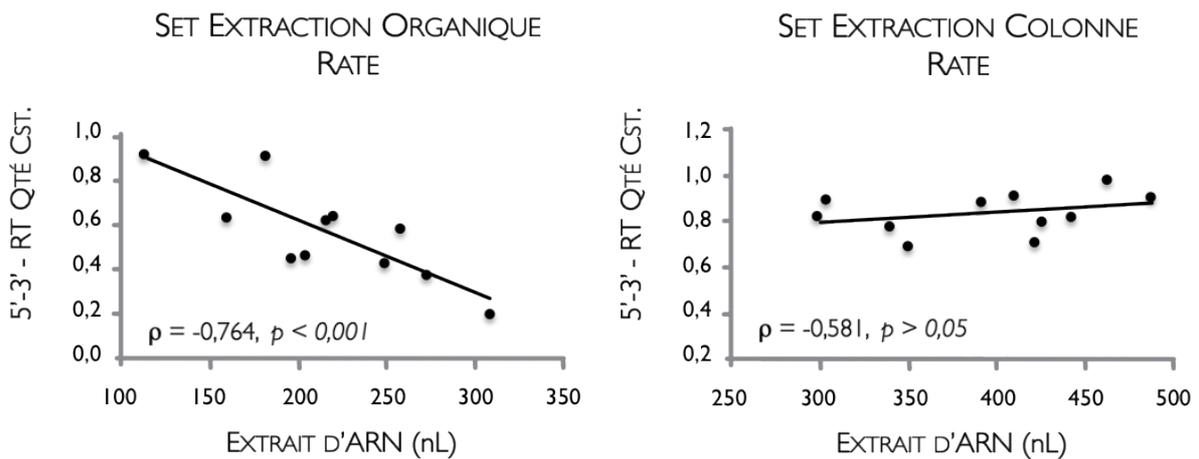
**Figure 4.1 :** Structure secondaire de l'ARN messager du gène *ribosomal protein, large, P0* (Rplp0, GenBank NM\_022402) chez *Rattus norvegicus*, obtenue avec l'algorithme *mfold* util 4.6 (RNA Folding Form version 2.3). Les deux couples d'amorces utilisés en PCR sont identifiés en rose. Les paramètres de réalisation sont : Séquence ARN linéaire, *Folding temperature* de 48°C, Conditions ioniques : 1M NaCl, sans ions divalents,  $\Delta G = -250.92$  kcal/mol,  $\Delta H = -3097.00$  kcal/mol,  $\Delta S = -8862.1$  cal/(K.mol),  $T_m = 76.3^\circ\text{C}$  selon un modèle d'état 2. Les écarts-types sont d'environ  $\pm 5\%$  pour l'énergie libre,  $\pm 10\%$  pour l'enthalpie,  $\pm 11\%$  pour l'entropie et de 2 à 4°C pour la température de fusion.



**Figure 4.2 :** Moyenne des ratios d'efficacités de RT (reverse transcription à quantité constante (CA-RT) / reverse transcription à volume constant (CV-RT)) de chaque échantillon, en utilisant l'analyse qPCR 3':5' et pour les trois tissus analysés. Le ratio des quantités d'amplicons en 5' *versus* quantités d'amplicons en 3' reflète le relatif succès de la reverse transcriptase amorcée à partir d'oligo-dT tout au long de la longueur du transcrit. Le ratio 5'-3' CA-RT / 5'-3' CV-RT évalue l'efficacité relative de la réaction de RT entre des conditions à quantité constante et volume constant d'ARN. Les comparaisons statistiques ont été faites avec le test de *Wilcoxon*. \*\*\*  $P < 0,001$ ; \*\*  $P < 0,01$ .

En observant le ratio de rendements de RT 5'-3' pour un même ARNm mais dans des conditions différentes, on est donc capable d'évaluer une efficacité de RT relative dont la valeur peut être prédictive de la présence d'un problème d'inhibition lorsque celle-ci est inférieure à 1. Les ratios sont calculés pour chaque échantillon à partir de la conversion exponentielle des valeurs de  $C_q$  obtenus à la fois pour la série de RT à quantité constante (5'-3' CA-RT) et pour la série de RT réalisée à volume constant (5'-3' CV-RT). Comme le montre la **Figure 4.2**, il n'existe aucune différence significative entre les différentes efficacités relatives pour quatre des six groupes expérimentaux. En revanche, il existe une différence significative observée pour les ARN d'origine splénique extraits par la méthode organique ainsi que pour les ARN issus du cervelet, extraits à partir d'une méthode sur colonne. Pour ces deux groupes expérimentaux,

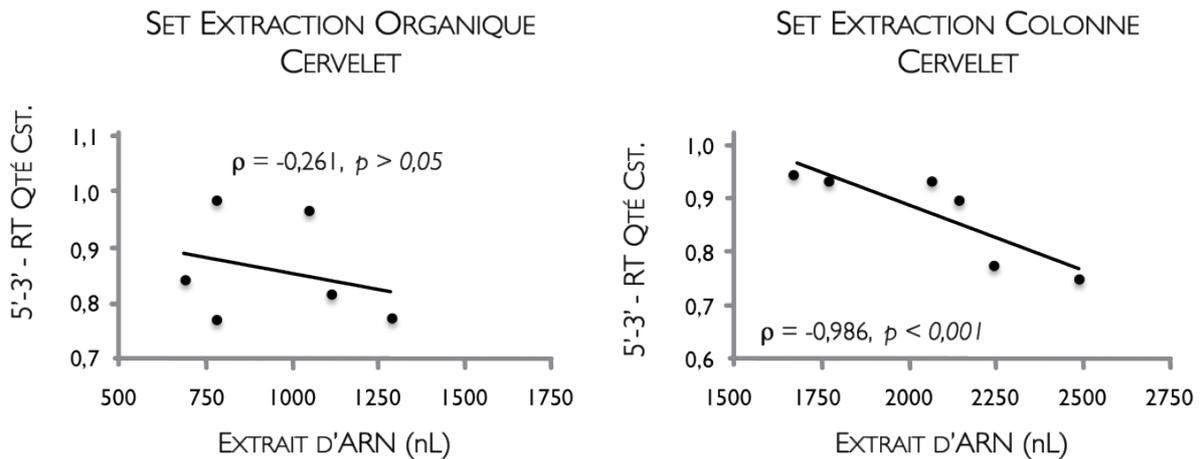
les résultats indiquent ainsi un moins bon rendement de RT lorsque cette réaction est réalisée à quantité constante d'ARN total. Cette diminution d'efficacité est respectivement de 46% pour la rate et de 14% pour le cervelet. D'autre part, il est important de noter que la gamme de variation relative en concentrations d'ARN est plus importante pour le set concernant le tissu splénique extrait par méthode organique (1,0 - 2,7) que pour le set concernant le cervelet extrait par méthode sur colonne (1,0 - 1,5). La variabilité de l'inhibition induite par la CA-RT est également plus importante dans la rate (RSD = 43%) que dans le cervelet (RSD = 10%).



**Figure 4.3a :** Corrélations linéaires non paramétriques de Spearman entre le volume extrait d'ARN utilisé et le ratio 5'-3' pour une transcription inverse à quantité constante d'ARN (CA-RT) issu du tissu splénique.

La Figure 4.3a représente les efficacités de CA-RT en fonction du volume d'extraits utilisés, à la fois pour l'extraction organique et l'extraction sur colonne des ARN d'origine splénique. La signification statistique est évaluée en utilisant le test de corrélation de rangs de Spearman. Les volumes notés représentent la quantité d'extraits bruts d'ARN. Afin de minimiser la variabilité technique observée au niveau du pipetage, les prises d'essai utilisées sont issues de dilutions permettant une utilisation de volumes supérieurs ou égaux à 5  $\mu$ L.

Lorsque les ARN totaux sont extraits par une méthode organique, on observe une corrélation significative entre l'efficacité de CA-RT et le volume d'extrait utilisé (Spearman  $\rho = -0,764$  ;  $p < 0,001$ ). L'existence de cette corrélation démontre ainsi la présence d'inhibiteurs de RT volume-dépendants au niveau des extraits d'ARN obtenus avec ce type d'extraction, contrairement aux extraits d'ARN obtenus à partir d'une méthode sur colonne où aucune corrélation significative n'est observée ( $p > 0,05$ ).



**Figure 4.3b** : Corrélations linéaires non paramétriques de *Spearman* entre le volume extrait d'ARN utilisé et le ratio 5'-3' pour une transcription inverse à quantité constante d'ARN (CA-RT) issu du cervelet.

La **Figure 4.3b** représente les efficacités de CA-RT en fonction du volume d'extraits utilisés, à la fois pour l'extraction organique et l'extraction sur colonne des **ARN totaux** issus du cervelet.

Contrairement au tissu splénique, c'est pour une méthode d'extraction d'ARN sur colonne de silice qu'une corrélation significative est observée entre l'efficacité de CA-RT et le volume d'extrait d'ARN utilisé pour le cervelet (*Spearman*  $\rho = -0,986$  ;  $p < 0,001$ ). En revanche, aucune corrélation significative n'est observée lorsque les ARN sont extraits par la méthode organique ( $p > 0,05$ ).

À partir de ce deuxième exemple et pour une méthode d'extraction différente, on démontre une nouvelle fois l'existence d'inhibiteurs de **RT** volume-dépendants, probablement d'origine lipidique, au niveau des extraits d'**ARN totaux**, obtenus ici par une purification sur colonne de silice.

En résumé, il existe des inhibiteurs volume-dépendants capables d'affecter l'efficacité de la transcription inverse au cours d'une analyse de RT-qPCR. Cette inhibition apparaît dans certaines conditions lorsque la RT réalisée à quantité constante d'ARN. Elle peut être mise en évidence par une analyse 5':3' réalisée sur un même ARN soumis à des conditions différentes de transcription inverse. On démontre enfin que des méthodes d'extraction plus adaptées au tissu analysé permettent d'en limiter sensiblement les effets.

### 4.2.3. Évaluation de l'inhibition de la réaction de qPCR

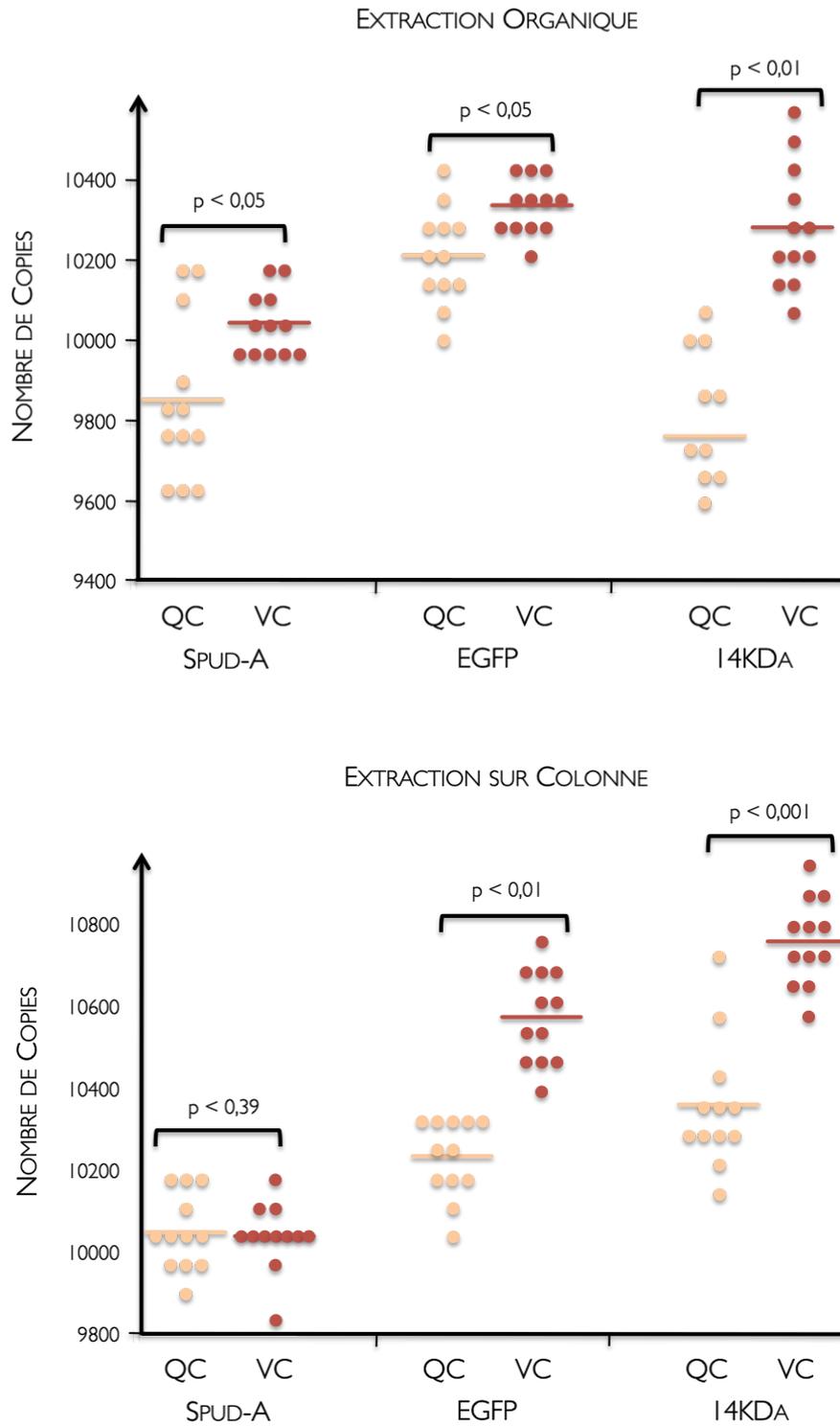
Une évaluation spécifique de cette inhibition est réalisée à l'aide d'un test basé sur l'amplification d'une **matrice** ADN *alien* (Nolan, Hands et al. 2006) préconisé au niveau des recommandations de MIQE (Informations Minimales en vue de la publication d'expérimentations effectuées en PCR quantitative en temps réel) (Bustin, Benes et al. 2009). L'impact de l'inhibition est ainsi évalué à l'aide de trois différentes **matrices** : Spud-A (gène *phyB* de la pomme de terre : *Solanum tuberosum* ; GenBank Y14572), EGFP (*Enhanced Green Fluorescent Protein* (Thastrup, Tullin et al. 2005) ; GenBank DQ071886) et 14 kDa (gène codant pour la protéine de 14 kilo Daltons du virus de la vaccine (GenBank M18173) pour l'ensemble des échantillons de tissu splénique. Les comparaisons statistiques sont réalisées à l'aide du test de Wilcoxon.

La **Figure 4.4** représente le nombre de copies de **matrice** *alien* détectées en fonction du type de RT (CA-RT ou CV-RT) et du type de méthode d'extraction utilisée. On observe que quelle que soit la **matrice** détectée, une inhibition significative est constatée lorsqu'une RT est réalisée à quantité constante d'ARN extrait à partir d'une méthode organique, respectivement  $p < 0,05$  ;  $p < 0,05$  et  $p < 0,01$  pour les **matrices** Spud-A, EGFP et 14 kDa. Comme cela a d'ailleurs pu être démontré au cours d'études précédentes (Huggett, Novak et al. 2008), l'inhibition est ici à la fois **amorce**-dépendante et **matrice**-dépendante. Les diminutions mesurées concernant le nombre de copies de **matrice** *alien* détectées au niveau des échantillons issus des deux différentes RT sont respectivement de 1,9 %, 1,2 % et 5,1 % pour les **matrices** Spud-A, EGFP, et 14 kDa.

En ce qui concerne l'évaluation de l'inhibition au niveau des échantillons obtenus à partir d'une méthode d'extraction sur colonne, on observe des résultats similaires avec une inhibition significative de la réaction de qPCR induite par une CA-RT, respectivement  $p < 0,01$  et  $p < 0,001$  pour les **matrices** EGFP et 14 kDa. On observe également un faible niveau d'inhibition de la réaction de qPCR avec respectivement 3,2 % et 3,7 % de diminution mesurée concernant le nombre de copies de **matrice** *alien* détectées au niveau des échantillons issus des deux différentes RT pour les **matrices** EGFP et 14 kDa. Parmi ces échantillons extraits sur colonne, aucune différence n'est observée lorsque la **matrice** Spud-A est détectée.

Enfin, il est important de noter que les niveaux d'inhibition observés au niveau de la réaction de qPCR sont beaucoup plus faibles que ceux observés au niveau de l'étape de transcription inverse.

En résumé, on constate que les inhibiteurs volume-dépendants générés par CA-RT sont capables d'inhiber la réaction de qPCR. Cette inhibition est à la fois *amorce-* et *gène-dépendante*. Cependant, cet impact sur la PCR reste négligeable par rapport à celui observé sur l'étape de la transcription inverse.



**Figure 4.4 :** Évaluation de l'inhibition de la réaction de qPCR à l'aide du test d'inhibition basé sur l'utilisation des matrices *alien* Spud-A, EGFP et 14 kDa, pour l'ensemble des échantillons d'origine splénique. Les comparaisons statistiques sont réalisées à l'aide du test de *Wilcoxon*.

#### 4.2.4. Conséquences des inhibiteurs volume-dépendants sur la quantification

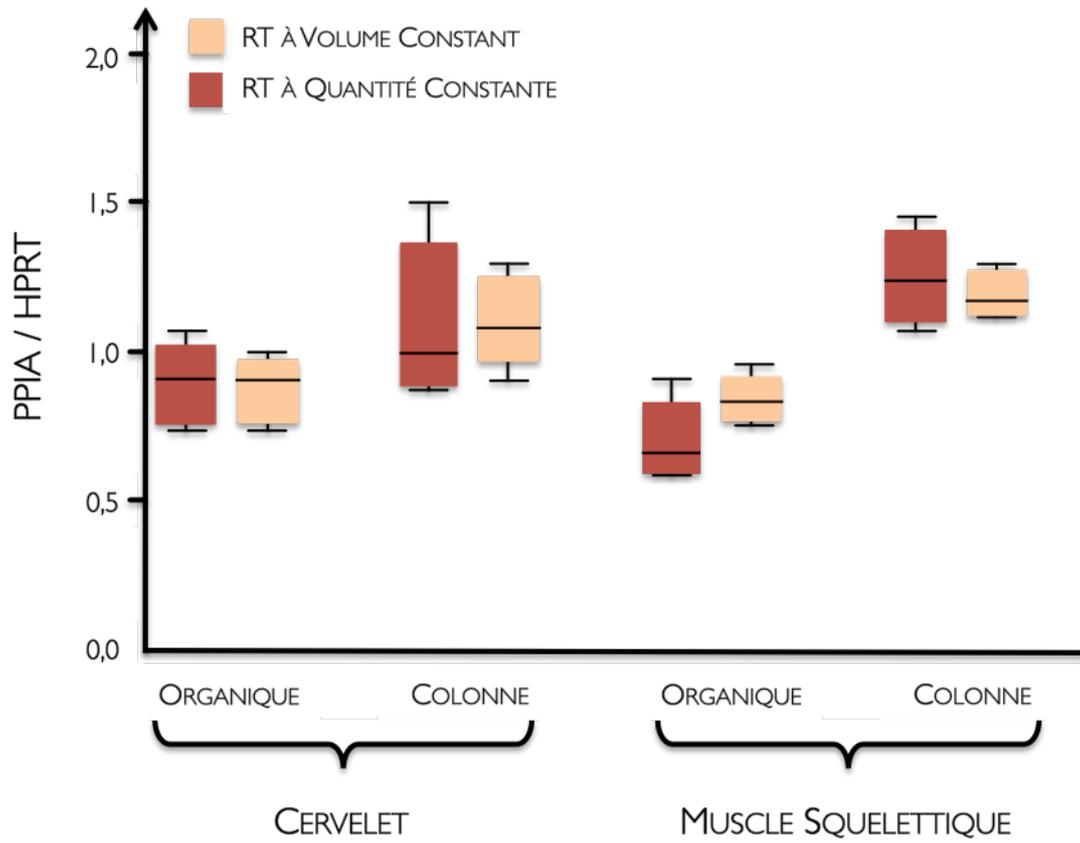
Deux **ARN messagers** : PPIA (*peptidylprolyl isomerase A (cyclophilin A)*) ; GenBank NM\_017101) et HPRT (*hypoxanthine phosphoribosyltransferase* ; GenBank NM\_012583), sont quantifiés au niveau de répliques techniques issus d'un même échantillon originel et pour deux tissus différents (cervelet et muscle squelettique). Les mesures sont donc réalisées sur des échantillons présentant le même taux d'inhibiteurs endogènes tissu-spécifiques.

La **normalisation** des données a été réalisée au moyen (1) d'une quantification relative avec l'établissement d'un ratio PPIA/HPRT et (2) d'un index multiparamétrique avec la valeur  $V_{2/3}$  obtenu avec le logiciel *geNorm*.

##### a) La quantification relative

Un seul gène de référence est ici utilisé en raison de **duplicates** techniques d'échantillons. Une quantification précise est obtenue lorsqu'une faible variabilité du ratio PPIA/HPRT est observée.

Comme le montre la **Figure 4.5** représentant ce ratio en fonction du type de **RT** (CA-RT ou CV-RT) et du type de méthode d'extraction utilisée pour les deux tissus, on observe une variabilité beaucoup plus faible lorsqu'une CV-RT est réalisée. Ce résultat remarquable est notamment observé au niveau des deux sets de **RT** (organique et colonne) réalisées pour des échantillons de tissu musculaire où la variabilité constatée en CA-RT est deux fois plus importante qu'en CV-RT (méthode organique : 16,7 % vs. 8,1 %, méthode sur colonne : 11,4 % vs. 5,8 %). Un résultat similaire est aussi constaté pour les échantillons de cervelet purifiés à partir de la méthode d'extraction sur colonne (22,0 % vs. 12,0 %). Concernant ce même tissu, la différence entre les valeurs observées entre CA-RT et CV-RT, pour le set d'extraction organique, est moins marquée (12,9 % vs. 10,7 %). Aucune différence significative n'est constatée entre CA-RT et CV-RT pour tous les sets de **RT** ( $P > 0,05$  ; test de *Wilcoxon*).



**Figure 4.5 :** Variabilité de l'expression relative du gène PPIA par rapport au gène HPRT au niveau du cervelet et du muscle squelettique. Les boîtes représentent le deuxième et le troisième quartile. Les barres symbolisent les plus fortes et plus faibles valeurs observées.

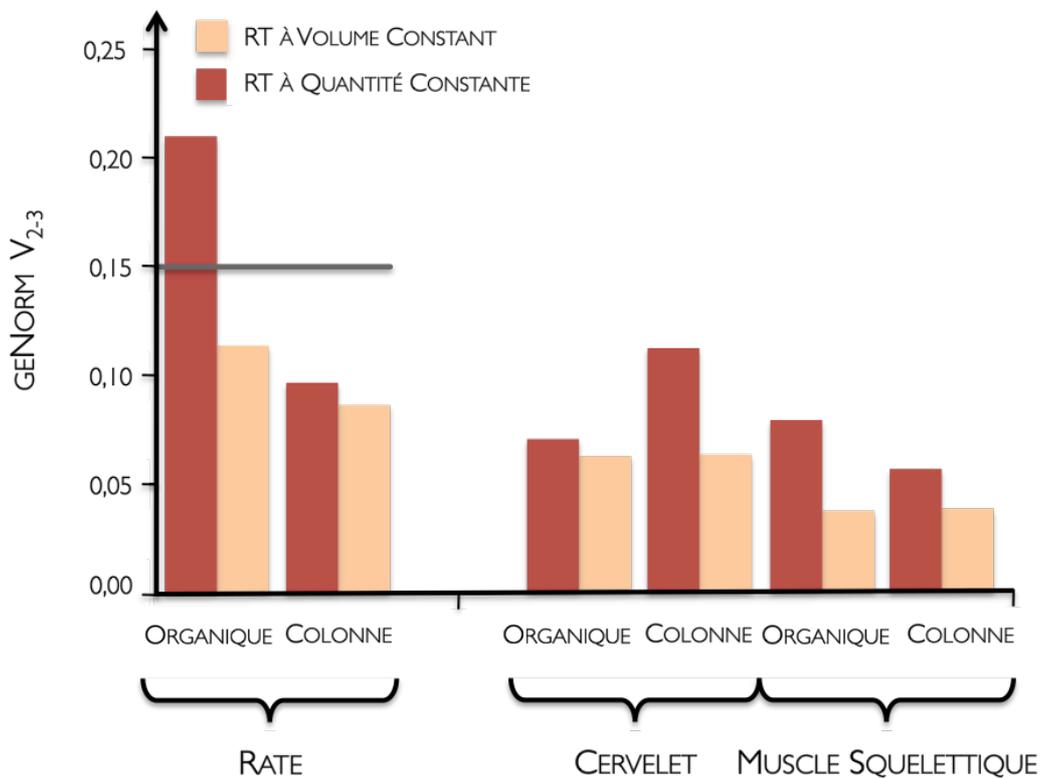
### b) L'index multiparamétrique

De précédentes études ayant démontré que la variabilité technique observée au cours de la technique de RT-qPCR était dépendante de l'ARNm analysé (Stahlberg, Kubista et al. 2004), nous avons évalué la précision de la qPCR en utilisant l'index multigénique  $V_{n/n+1}$  du logiciel de normalisation geNorm. Le seuil limite de 0,15 représente le niveau de variation par paire exigeant l'utilisation d'un gène de référence supplémentaire pour la normalisation de données obtenues à partir d'échantillons indépendants (Vandesompele, De Preter et al. 2002). Nous avons donc utilisé les transcrits de trois gènes (ARBP, HPRT1 et PPIA) afin de normaliser à la fois des répliques biologiques indépendants (tissu splénique) et des répliques techniques (cervelet et muscle squelettique) ayant subi les deux transcriptions inverses. Ainsi, la reproductibilité et la précision de la méthode, depuis la purification de l'ARN jusqu'à la qPCR elle-même, sont évaluées par les valeurs  $V_{2/3}$ . Pour une normalisation précise des résultats, ces valeurs ne doivent pas dépasser le seuil limite de 0,15 (Vandesompele, De Preter et al. 2002). Cet index nous permet ainsi d'apprécier à la fois la stabilité physiologique des gènes de référence utilisés et la qualité de l'analyse effectuée en RT-qPCR.

Comme l'illustre la Figure 4.6 représentant les différentes valeurs  $V_{2/3}$  obtenues pour chaque type de répliques en fonction du type de RT et de la méthode d'extraction utilisée, on observe qu'une CV-RT entraîne une nette diminution de ces valeurs. On soulignera simplement que le type de RT a relativement peu d'impact sur le classement de la stabilité des gènes de référence utilisés (Tableau 4.2).

- Pour des répliques biologiques indépendants (tissu splénique), on observe ainsi une validation de la normalisation pour les deux types de RT lorsque l'extraction est réalisée sur colonne de silice. En revanche, si la CV-RT reste validée pour des échantillons extraits par méthode organique, la CA-RT ne l'est plus ( $V_{2/3} = 0,21$ ). Ce résultat traduit une discrimination relative entre les niveaux d'expression des différents gènes de référence, due à l'inhibition de la RT par les inhibiteurs volume-dépendants. Cette variation est en accord avec les différences d'efficacité relatives de RT observées au niveau de la Figure 4.2.
- Pour des répliques techniques issus d'un même échantillon original (cervelet et muscle squelettique), on observe encore une meilleure normalisation lorsque la CV-RT est utilisée. Plus précisément, l'extraction du cerveaulet sur colonne de silice élimine moins

efficacement les inhibiteurs que l'extraction organique. De ce fait, une meilleure **normalisation** est obtenue avec cette dernière méthode où l'inhibition est plus faible. Pour le tissu musculaire, ces différences entre CA-RT et CV-RT existent pour toutes les méthodes d'extraction utilisées. Bien qu'il n'y ait aucune différence entre les efficacités observées entre CA-RT et CV-RT concernant ce tissu, ces valeurs  $V_{2/3}$  sont plus importantes en CA-RT qu'en CV-RT, que ce soit avec une extraction organique (113%) ou avec une extraction sur colonne de silice (50%).



**Figure 4.6 :** Variations par paire ( $V_{2/3}$ ) calculées par le logiciel *geNorm*. Les index de normalisation multiparamétriques sont calculés en utilisant ARBP, HPRT1, et PPIA comme gènes de référence pour chaque série de RT.

			ARBP	PPIA	HPRT	geNorm
Tissu splénique	Extraction organique avec <i>clean-up</i>	CV+CA-RT	0,46	<u>0,51</u>	<b>0,37</b>	0,165
		CA-RT	0,55	<u>0,64</u>	<b>0,44</b>	0,209
		CV-RT	0,32	<u>0,35</u>	<b>0,26</b>	0,112
	Extraction sur colonne de silice	CV+CA-RT	<u>0,29</u>	0,28	<b>0,25</b>	0,091
		CA-RT	0,28	<u>0,31</u>	<b>0,28</b>	0,096
		CV-RT	<u>0,26</u>	0,18	<b>0,18</b>	0,086
Cervelet	Extraction organique avec <i>clean-up</i>	CV+CA-RT	<u>0,25</u>	<b>0,21</b>	<b>0,21</b>	0,080
		CA-RT	0,22	<b>0,18</b>	<u>0,22</u>	0,070
		CV-RT	<u>0,20</u>	0,19	<b>0,17</b>	0,062
	Extraction sur colonne de silice	CV+CA-RT	0,28	<u>0,28</u>	<b>0,23</b>	0,088
		CA-RT	0,32	<u>0,35</u>	<b>0,26</b>	0,111
		CV-RT	<u>0,21</u>	0,20	<b>0,19</b>	0,063
Muscle squelettique	Extraction organique avec <i>clean-up</i>	CV+CA-RT	0,17	<u>0,22</u>	0,18	0,070
		CA-RT	0,22	<u>0,25</u>	<b>0,19</b>	0,079
		CV-RT	0,09	<u>0,12</u>	<b>0,09</b>	0,037
	Extraction sur colonne de silice	CV+CA-RT	0,13	<b>0,12</b>	<u>0,14</u>	0,042
		CA-RT	0,13	<b>0,12</b>	<u>0,18</u>	0,057
		CV-RT	<u>0,12</u>	0,11	<b>0,09</b>	0,038

**Tableau 4.2 :** Valeurs de stabilité d'expression moyenne M obtenues par le logiciel *geNorm* et déterminées respectivement pour l'ensemble des séries de RT (CA-RT + CV-RT), les séries de CA-RT et les séries de CV-RT. Le gène plus stable figure en **gras**, le gène le moins stable figure en souligné.

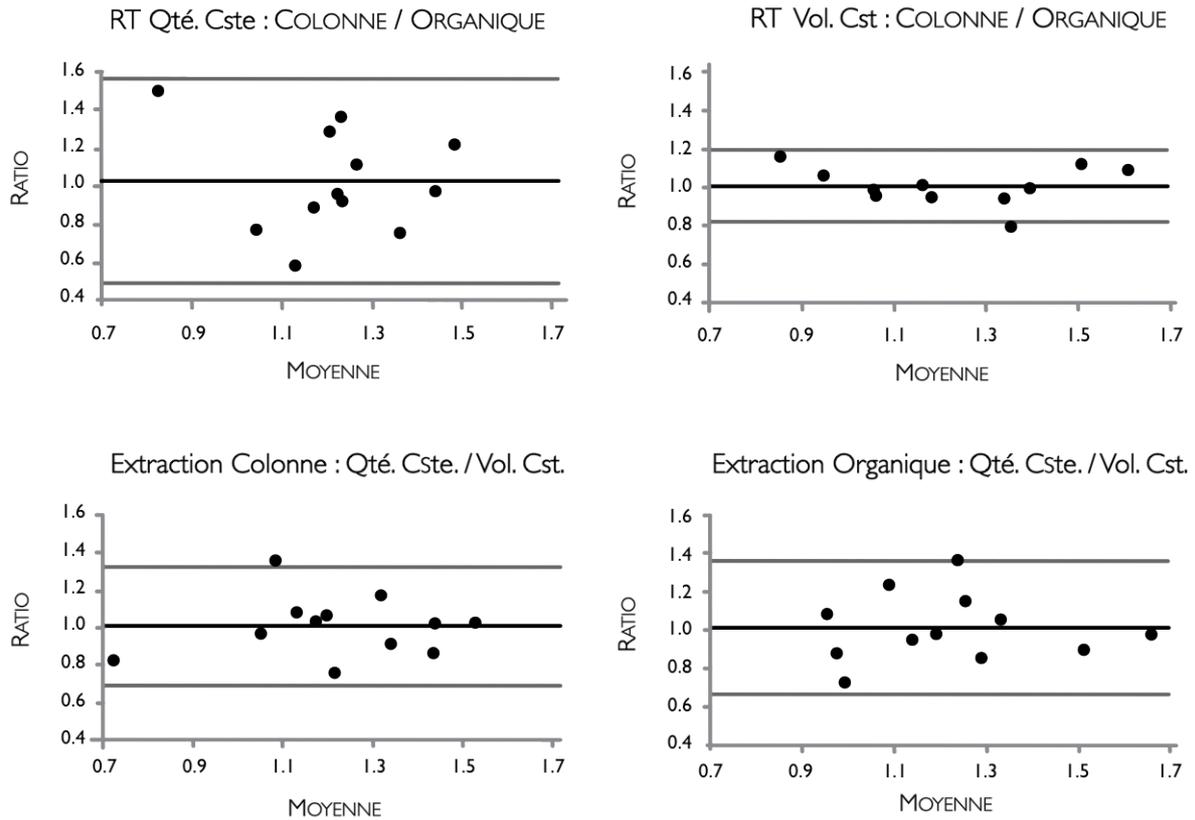
En résumé, au regard des résultats obtenus en CA-RT, une CV-RT est capable de diminuer la variabilité de la qPCR induite par les inhibiteurs volume-dépendants. Cette amélioration est vérifiée par une meilleure normalisation dans toutes les conditions testées. D'autre part, celle-ci est plus importante lorsque les échantillons présentent une inhibition significative de l'étape de transcription inverse.

#### 4.2.5. Agréments entre les différentes techniques d'extraction d'ARN utilisées pour les répliques biologiques

Les niveaux d'expression relatifs du gène CD 19 (GenBank NM\_001013237), préalablement normalisés par le logiciel *geNorm*, sont mesurés au niveau du tissu splénique pour tous les sets de RT. Pour chaque échantillon, un agrément observé entre les différentes mesures du niveau d'expression de CD19, obtenues dans chaque set de RT, est évalué par la méthode de *Bland-Altman* (Altman and Bland 1983 ; Bland and Altman 1986) (Figure 4.7). Cette analyse exprime le rapport des valeurs normalisées mesurées en fonction de la moyenne de ces valeurs.

On sait que la RT est l'étape de la technique qui présente la plus grande variabilité (Stahlberg, Hakansson et al. 2004 ; Stahlberg, Kubista et al. 2004). D'après les résultats obtenus, on constate que le meilleur agrément est obtenu lorsque des niveaux d'expression d'ARNm extraits à partir des deux méthodes (organique et colonne) sont comparés à CV-RT. Ce très bon résultat souligne donc l'efficacité d'une RT à volume constant dans la diminution de la variabilité inter-échantillons. Il est d'autant plus significatif que ni les extractions des ARN, ni les transcriptions inverses ne sont réalisées en même temps (Figure 4.7, en haut, à droite). Des agréments beaucoup moins bons sont observés lorsque des niveaux d'expression d'ARNm obtenus à CA-RT sont comparés avec des niveaux d'expression d'ARNm obtenus à CV-RT à partir d'un même type d'extraction (organique ou colonne, Figure 4.7, les deux graphiques du bas). Ce résultat indique que le choix de la méthode d'extraction est moins important que la méthode (CA-RT ou CV-RT) utilisée pour la transcription inverse des ARN. L'agrément le plus mauvais est observé quant à lui lorsque des volumes variables d'extraits d'ARN (CA-RT) sont comparés à partir de méthodes d'extraction différentes, malgré le fait qu'une seule et même RT soit réalisée (Figure 4.7, en haut, à gauche).

Le meilleur agrément est observé lorsque des niveaux d'expression obtenus à partir d'une CV-RT sont comparés et cela, malgré le fait que différentes extractions et différentes transcriptions inverses soient réalisées. Ce résultat est d'autant plus important que la transcription inverse constitue l'étape la plus variable de l'analyse par RT-qPCR.



**Figure 4.7 :** Quantifications des niveaux d'expression d'ARNm du gène CD19 mesurés et normalisés au niveau de la rate : les agréments effectués entre les différentes mesures sont réalisés en comparant les ratios selon la méthode de *Bland-Altman* parmi les différents sets de RT et d'extraction. L'abscisse et l'ordonnée représentent respectivement la moyenne et le ratio des taux d'ARNm du gène CD19 mesurés au niveau des différents sets testés. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

### 4.3. Discussion

La **standardisation** de la réaction de **RT** constitue une étape déterminante dans la fiabilité de la quantification de l'ARN. En 2004, STAHLBERG *et al.* démontrent que le rendement de **RT** dépend non seulement de la quantité d'ARNm ciblé mais aussi de la concentration d'ARN totaux utilisée (Stahlberg, Hakansson et al. 2004). Le consensus actuel stipule qu'il faut utiliser une quantité constante d'ARN au cours de cette étape (CA-RT). Toutefois, la variabilité observée au niveau des rendements d'extraction d'ARN totaux pour différents échantillons pose souvent un réel problème, obligeant le manipulateur à utiliser des volumes variables d'extraits d'ARN afin de standardiser cette réaction. Au cours de cette étude, nous proposons une nouvelle méthode de **RT** permettant d'améliorer efficacement cette **standardisation** en manipulant un volume constant d'extrait d'ARN totaux pour tous les échantillons, après purification. Cette méthode originale (CV-RT) est rendue possible en maintenant la concentration d'ARN totaux constante au niveau de tous les volumes prélevés par simple supplémentation d'ARNt de levure (Stahlberg, Hakansson et al. 2004).

Les différents résultats de ces travaux soulignent ainsi que certains extraits d'ARN totaux peuvent potentiellement contenir des inhibiteurs capables de perturber la réaction de **RT-qPCR**, notamment en augmentant la variabilité inter-échantillons. Par ailleurs, la présence de ces inhibiteurs au niveau de l'extrait d'acides nucléiques n'est pas uniquement liée à la quantité d'ARN totaux utilisée mais très probablement à la fois à la quantité de tissu traité et aux volumes de réactifs utilisés au cours de l'étape d'éluion des ARN totaux ou de leur remise en solution. Il est important de souligner qu'une telle inhibition est significative pour une faible différence de concentrations relatives d'extraits d'ARN (ici un facteur 3 entre le moins et le plus concentré). Une telle plage de concentrations est significativement plus faible que les gammes 100 fois plus importantes utilisées au cours d'études précédentes réalisées sur les propriétés de la **RT** (Stahlberg, Hakansson et al. 2004 ; Levesque-Sergerie, Duquette et al. 2007). Certains de ces inhibiteurs sont probablement exogènes, liés au réactif utilisé. Le groupe expérimental constitué d'échantillons de tissu splénique et extrait par une méthode organique illustre très probablement cette possibilité. En effet, l'extraction du tissu splénique sur colonne de silice permet d'enlever plus efficacement les inhibiteurs. Les contaminations observées au niveau de l'extraction organique proviennent très probablement du phénol utilisé. Inversement, d'autres inhibiteurs pouvant altérer l'efficacité de la **RT-qPCR** sont d'origine endogène, c'est-à-dire propres au tissu lui-même. Ainsi, c'est au niveau du cervelet que la méthode d'extraction d'ARN

sur colonne de silice a démontré ses limites en n'éliminant pas totalement la présence d'inhibiteurs endogènes volume-dépendants, probablement d'origine lipidique (Rossen, Norskov et al. 1992). L'utilisation d'une CV-RT diminue la différence pouvant exister entre les échantillons d'un groupe expérimental concernant les inhibiteurs endogènes et permet ainsi d'homogénéiser les concentrations relatives en inhibiteurs exogènes.

Il est important de remarquer que la méthode de CV-RT ne permet une diminution de la variabilité inter-échantillons que lorsque des inhibiteurs volume-dépendants sont considérés. En effet, ces inhibiteurs sont caractérisés par une concentration relativement constante au niveau de l'ensemble des échantillons. Cependant, en raison de la variabilité biologique, certains échantillons peuvent présenter des inhibiteurs qui leur sont propres (inhibiteurs échantillon-spécifiques) et dont la concentration peut varier d'un échantillon à l'autre (Tichopad, Didier et al. 2004). Dans cette situation précise, il est important de remarquer qu'aucune RT quelle qu'elle soit, est capable de normaliser la variabilité inter-échantillons pouvant exister au niveau de la concentration de tels inhibiteurs. C'est ici la limite de cette nouvelle procédure.

Cette hypothèse peut être illustrée avec la comparaison réalisée entre les répliques biologiques et les répliques techniques (**Figures 4.3a et 4.3b**). La corrélation existant entre l'inhibition de la RT et le volume d'extrait d'ARN de l'échantillon est plus importante pour le set d'échantillons cervelet extraits par la méthode sur colonne de silice (répliques techniques présentant un taux identique d'inhibiteurs endogènes liés à l'échantillon lui-même) que pour le set d'échantillons de rate extraits par méthode organique (répliques biologiques caractérisés par des taux variables d'inhibiteurs endogènes échantillons-dépendants). Cependant, comme l'indique l'importante valeur du coefficient de Spearman ( $\rho = -0,764$ ), les inhibiteurs volume-dépendants ont probablement plus d'impact sur l'efficacité de RT que les inhibiteurs échantillon-dépendants au niveau du set d'échantillon de tissu splénique extraits par méthode organique. De plus, ces inhibiteurs échantillon-dépendants n'inhibent seulement l'étape de RT que lorsqu'ils ne sont pas éliminés au moment de l'étape d'extraction des ARN. Ainsi, au niveau du set d'échantillons de rate (répliques biologiques) extraits par méthode de purification sur colonne de silice, la plupart des inhibiteurs échantillon-dépendants ont pu être retirés de l'extrait d'ARN, comme l'indique la faible valeur du coefficient de variation de l'inhibition RT (18%, calculée d'après la **Figures 4.3a**). De manière similaire, la même observation est constatée concernant la valeur du coefficient de variation de l'inhibition RT pour le set d'échantillons de cervelet

(réplicas techniques présentant un taux identique d'inhibiteurs endogènes échantillon-dépendants, CV = 24%, calculé d'après la **Figures 4.3b**).

La variabilité induite par une CA-RT est essentiellement due à d'importantes variations inter-échantillons au niveau de l'efficacité de **RT**. Cela l'est beaucoup moins au niveau de l'efficacité de **PCR**. En effet, seulement de très légères variations d'efficacité ont été observées au niveau de cette dernière étape. Cela pourrait s'expliquer par les différences théoriques existant entre les conditions de synthèse d'un **ADN complémentaire** à partir de l'ARN et les conditions de synthèse de l'ADN à partir de l'**ADN complémentaire** matriciel utilisé. La synthèse de l'ADN peut en effet être altérée par un grand nombre de mécanismes (Opel, Chung et al. 2010) tels qu'une perturbation de l'**hybridation** des **amorces**, une inhibition enzymatique potentielle de la **Taq polymérase** ou encore la présence de **structures secondaires** au niveau du simple brin de l'acide nucléique utilisé comme support matriciel. En ce qui concerne la synthèse de l'ADN à partir de l'**ADN complémentaire**, une inhibition enzymatique reste peu probable même si elle a déjà été observée (Wiedbrauk, Werner et al. 1995 ; Tichopad, Polster et al. 2005). La présence de **structures secondaires** éventuelles est également peu probable, à la fois en raison de la courte taille des amplicons générés, du temps de polymérisation relativement court et de l'importante température d'incubation utilisée. Ainsi, une inhibition induite par des troubles liés à l'**hybridation** des **amorces** de **PCR** sur la **matrice** constitue un mécanisme d'inhibition de la **PCR** beaucoup plus probable. En ce qui concerne l'altération du rendement de synthèse de l'**ADNc** à partir de l'ARN en revanche, une inhibition enzymatique ou la présence de **structures secondaires** éventuelles au niveau de la séquence de l'ARN ciblé sont beaucoup plus probables. En effet, la longueur du **transcrit**, la température d'incubation plus faible, sans écarter la possibilité d'une éventuelle dégradation de l'ARN (Cheng, Chen et al. 1995; Golenberg, Bickel et al. 1996) peuvent amener à cette situation problématique. D'autre part, une évaluation précise de l'inhibition de **RT** apparaît relativement compliquée car cette mesure est effectuée après l'étape d'amplification de la cible ADN. En effet, en procédant de cette manière, on convertit un phénomène continu : une longueur plus ou moins importante d'**ADNc** synthétisée par la **transcriptase inverse**, en résultat binaire : la cible qui doit être détectée spécifiquement par les **amorces** de **qPCR** est ou *n'est pas* reverse transcrite par l'enzyme. En général, une température plus élevée utilisée au moment de la **RT** améliore significativement les rendements d'**ADNc** en réduisant significativement la formation de **structures secondaires** tout au long de l'**ARNm** (Brooks, Sheflin et al. 1995 ; Kuo, Leung et al. 1997). De manière logique, ces **structures secondaires**, dépendantes de la séquence de l'ARN à

partir de laquelle elles se forment, sont capables d'influencer significativement le rendement de RT de manière gène-dépendante (Kuo, Leung et al. 1997; Stahlberg, Hakansson et al. 2004) en gênant la processivité de l'enzyme au cours de sa synthèse. Ainsi, la diminution de l'efficacité relative de RT évaluée par l'analyse 5':3' et observée à la fois au niveau du set d'échantillons de tissu splénique extraits par méthode d'extraction organique et au niveau du set d'échantillons de cervelet extraits par méthode d'extraction sur colonne de silice, pourrait être expliquée par la présence de structures secondaires significatives prédites par l'algorithme *mfold* entre la position des amplicons situés aux deux extrémités du transcrit ciblé. Il est important de souligner que ce résultat a été obtenu en utilisant une stratégie d'amorçage par oligo-dT. En effet, les RT d'ARNm devraient également être évaluées en utilisant d'autres stratégies (Zhang and Byrne 1999) telles que l'utilisation de *Random* hexamères ou des amorces spécifiques par exemple.

Les effets de l'inhibition de l'étape de RT sur la fiabilité de l'analyse par la méthode de RT-qPCR sont plus faciles à évaluer. En effet, en utilisant des répliques techniques, le calcul du ratio d'expression relative de deux gènes permet d'évaluer l'impact particulier d'une variabilité induite par une CA-RT sur la quantification des ARNm (d'après la Figure 4.5). Cependant, en raison de l'existence des différentes caractéristiques observées au niveau de la séquence d'un ARNm (structure, longueur, composition), l'inhibition potentielle de la RT pourrait aussi être gène-spécifique (Stahlberg, Kubista et al. 2004). L'utilisation d'un index multiparamétrique basé sur l'évaluation du niveau d'expression de plusieurs gènes constitue donc le moyen le plus efficace pour observer ce phénomène d'inhibition particulier. C'est ce qu'ont pu démontrer les différentes comparaisons réalisées entre les valeurs  $V_{2/3}$  calculées par *geNorm* (Vandesompele, De Preter et al. 2002) à partir de répliques techniques et biologiques, présentant ainsi cet index comme un moyen très sensible de l'évaluation du processus pré-analytique (d'après la Figure 4.6). On a pu constater qu'une CV-RT diminuait significativement la variabilité observée au niveau des échantillons de tissu musculaire au cours d'une analyse de RT-qPCR et cela, quelle que soit la méthode d'extraction utilisée. Certes, ce résultat prouve l'efficacité de la méthode en CV-RT mais il indique surtout qu'aucune méthode d'extraction ne s'est révélée efficace dans l'élimination des inhibiteurs volume-dépendants présents au niveau de ces échantillons. Par conséquent, les CA-RT / CV-RT peuvent être utilisées dans le but de tester différentes méthodes d'extraction employées et d'identifier la méthode la plus efficace pour un type d'échantillon donné.

## 4.4. Conclusion

---

La réaction de **transcription inverse** des ARN est considérée comme l'étape la plus critique en quantification de niveaux d'**expression** génique par **RT-qPCR** (Bustin 2005). La variabilité des rendements obtenus au moment de l'extraction concourt à une augmentation importante de la variabilité inter-échantillons. En raison de l'existence d'inhibiteurs volume-dépendants pouvant à la fois être endogènes et/ou exogènes au tissu analysé, nous démontrons dans cette étude qu'une **RT** à quantité constante d'ARN conduit à une variabilité importante de son rendement de synthèse, aboutissant en définitive, à une altération de la fiabilité du résultat obtenu en quantification.

Bien que l'impact des inhibiteurs soit inéluctable en **RT-qPCR**, il est toutefois possible d'en limiter les conséquences sur la précision d'une quantification. En effet, il existe aujourd'hui un grand nombre de méthodes de purification d'acides nucléiques, particulièrement adaptées aux échantillons connus pour leur composition spécifique en inhibiteurs, comme le sang ou les tissus riches en lipides. Une utilisation appropriée de ces kits et réactifs permet dans un premier temps d'obtenir un rendement d'extraction relativement correct de la **matrice** ainsi purifiée. Dans un deuxième temps, il est possible de mettre en évidence les éventuelles contaminations d'inhibiteurs au niveau de l'extrait par la réalisation d'un simple test **5':3'** (Nolan, Hands et al. 2006). L'analyse, effectuée sur un même acide nucléique extrait dans des conditions différentes, donne ainsi accès au statut de la **matrice** que l'on va utiliser, par comparaison entre les efficacités relatives de **RT** obtenues pour les différentes méthodes d'extraction pratiquées. Cette information est essentielle car elle permet d'éclairer de manière conséquente le manipulateur quant à la démarche à suivre concernant l'étape de **transcription inverse** des ARN. Grâce aux comparaisons par la méthode de *Bland-Altman*, nous avons en effet démontré que le type de procédure choisi (CA-RT ou CV-RT) est plus décisif que celui porté sur la méthode d'extraction à utiliser. Par ailleurs, nous avons aussi démontré à l'aide d'un test d'inhibition (Nolan, Hands et al. 2006) que les inhibiteurs volume-dépendants, qu'ils soient endogènes ou exogènes, présentent un impact beaucoup plus important sur le rendement de la **RT** que sur celui de la **qPCR** et que la nature de cette inhibition était à la fois analyse et gène-dépendante.

En respectant le consensus établi selon lequel la **RT** doit être effectuée à quantité constante d'ARN pour chaque échantillon analysé (Bustin, Benes et al. 2009), nous proposons donc une alternative originale basée sur l'utilisation de volumes constants d'extraits d'ARN

purifiés, alors supplémentés par de l'ARNt de levure (Stahlberg, Hakansson et al. 2004). Les différences inter-échantillons liées au rendement d'extraction sont ainsi diminuées, limitant de manière importante l'impact des inhibiteurs endogènes et/ou exogènes volume-dépendants sur la réaction de RT-qPCR. D'autre part, nous démontrons que l'index de normalisation multiparamétrique de *geNorm* (Vandesompele, De Preter et al. 2002) peut être amélioré lorsque qu'une CV-RT est effectuée, et cela quelque soit la technique de purification pratiquée. Cette prise en compte de la variabilité est dans la ligne droite des directives préconisées par MIQE (Bustin, Benes et al. 2009) et constitue le prérequis absolument indispensable pour l'obtention d'une précision satisfaisante de la quantification de l'ARN messenger par cette méthode. Alors qu'il existe des inquiétudes légitimes quant à la fiabilité des tests effectués en RT-qPCR (Murphy and Bustin 2009), une transcription inverse à volume constant d'extrait d'ARN totaux pourrait fournir une alternative particulièrement pertinente concernant la standardisation de l'étape de RT.

Ce travail a fait l'objet d'une publication dans le journal *Analytical Biochemistry* (Pugniere, Banzet et al. 2011), proposé à la fin du manuscrit.

Chapitre 5

Utilisation  
des *Locked Nucleic Acids* (LNA)  
en qPCR

---

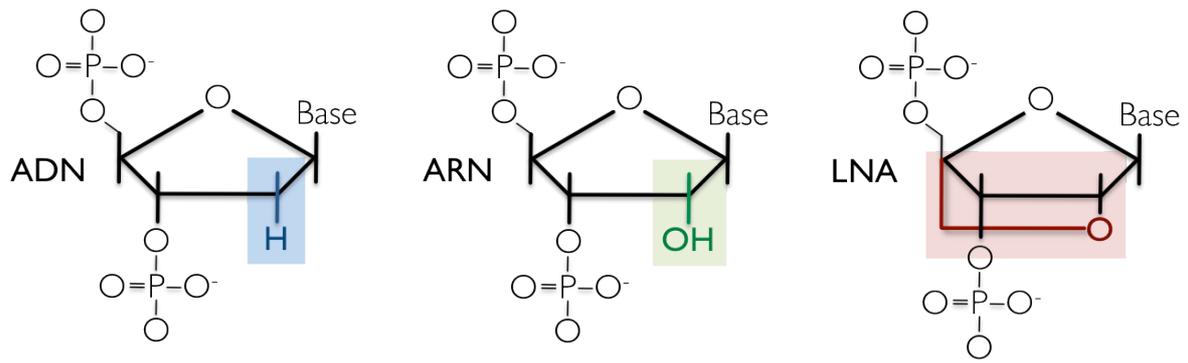
## 5.1. Introduction

Dans le contexte de la détection des acides nucléiques, l'**hybridation** spécifique d'un **oligonucléotide** avec sa **séquence complémentaire** est à la base de la plupart des méthodes de quantification. Cependant, les **oligonucléotides** naturels retrouvés au niveau de la cellule s'avèrent parfois assez mal adaptés pour des raisons d'instabilité ou de faible affinité vis-à-vis de leurs séquences cibles. Dans une optique d'amélioration de l'**hybridation**, l'exploration de nouvelles molécules, analogues aux nucléotides naturels, a suscité un effort permanent depuis la fin des années 70. Cette recherche a eu pour principal objectif d'obtenir des molécules capables de s'hybrider aux nucléotides de manière **complémentaire** de façon plus stable que les **oligonucléotides** naturels, selon les règles d'appariement définies par WATSON & CRICK (Jepsen, Sorensen et al. 2004). Une grande variété de molécules capables de présenter une stabilité thermique accrue a donc vu le jour, telles que les **acides nucléiques peptidiques** ou PNA (Nielsen, Egholm et al. 1991), les oligo-2'-fluoro-3'-aminonucléotide N3'→P5' phosphoramidites (Schultz and Gryaznov 1996) ou les acides nucléiques 5'-anhydro hexitol ou HNA (Van Aerschot, Rozenski et al. 1995), mais sans vraiment présenter de reconnaissance renforcée vis-à-vis de la cible à détecter. Les acides nucléiques bloqués (ou *Locked Nucleic Acids*, LNA) sont ensuite apparus, semblant répondre plus efficacement aux exigences de cette problématique.

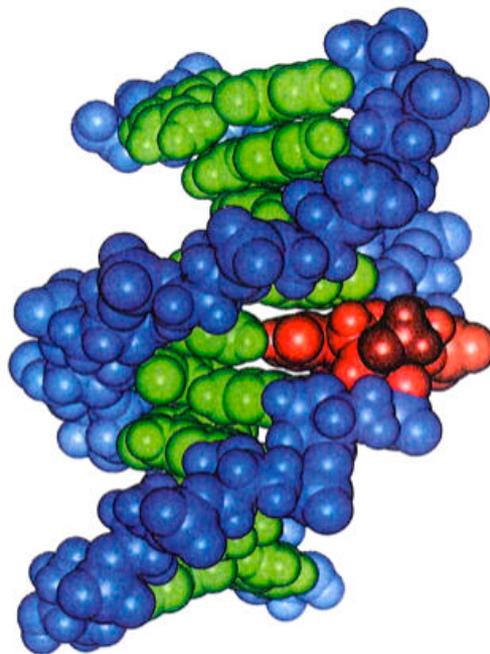
Les LNA sont actuellement les molécules les plus utilisées pour la modification des **oligonucléotides**. Similaires aux nucléobases conventionnelles (adénine, cytosine, guanine, 5-methylcytosine, thymine et uracile), elles ont été synthétisées pour la première fois par OBIKA et al. en 1998 (Obika, Nanbu et al. 1998). Très peu de temps après leur découverte, une deuxième équipe danoise les synthétise de manière totalement indépendante (Koshkin, Singh et al. 1998). En effet, plusieurs réactions biochimiques peuvent conduire à la synthèse de ces 2'-O,4'-C-methylene- $\beta$ -D-ribofuranosyl nucléotides, obtenue par chimie phosphoramidite conventionnelle généralement utilisée dans la production des oligomères (Wengel, Petersen et al. 2003). Consciente de l'important potentiel que peuvent présenter les LNA au niveau de certaines applications biologiques, la société de biotechnologie danoise *Exiqon* va alors très rapidement s'emparer du brevet d'exploitation et financer de nombreuses études de Recherche et Développement pour ces molécules (Kumar, Singh et al. 1998 ; Kvaerno and Wengel 1999).

L'étude structurale des monomères LNA est dans un premier temps réalisée à partir d'études de cristallographie par rayons X (Obika, Nanbu et al. 1997) ou de spectroscopie par Résonance Magnétique Nucléaire (RMN) (Koshkin, Singh et al. 1998 ; Singh, Kumar et al. 1998; Nielsen, Singh et al. 1999 ). Ces dernières montrent que la structure du monomère est " bloquée " en conformation C3'-endo par un pont méthylène 2'O-4'C au niveau du sucre qui les compose (**Figure 5.1**). Malgré cette particularité, les LNA présentent les mêmes caractéristiques que les acides nucléiques naturels, à savoir solubilité en milieu aqueux mais aussi et surtout complémentarité d'appariement selon les règles définies par WATSON & CRICK (Watson and Crick 1953). Il a en effet été démontré une affinité extrêmement importante vis-à-vis de l'ADN et de l'ARN selon ces règles d'appariement (Koshkin, Nielsen et al. 1998 ; Koshkin, Singh et al. 1998), présentant les LNA comme une solution potentielle dans l'amélioration de la discrimination relative des acides nucléiques naturels (Vester and Wengel 2004). D'autre part, ces molécules présentent de remarquables qualités thermiques puisqu'elles permettent une augmentation significative de la **température de fusion** de la séquence dans laquelle elles sont incorporées, favorisant ainsi une plus grande stabilité thermique de l'**oligonucléotide** ainsi modifié (Obika, Nanbu et al. 1998 ; Singh, Kumar et al. 1998).

Les premières études structurales de **duplexes** de séquences oligonucléotidiques modifiées par LNA sont réalisées dans un premier temps (**Figure 5.2**) (Nielsen, Singh et al. 1999 ; Nielsen, Singh et al. 2000 ; Petersen, Nielsen et al. 2000 ; Jensen, Singh et al. 2001 ; Vester and Wengel 2004). Les LNA ont démontré d'excellentes spécificité et affinité pour les acides nucléiques naturels **complémentaires** (Babu and Wengel 2001). En 1999, Christina NIELSEN réalise une étude structurale à partir d'une séquence ADN de 9 nucléotides, modifiée par l'incorporation d'un seul monomère LNA (Nielsen, Singh et al. 1999). Elle observe alors que l'acide nucléique bloqué est capable de réorganiser localement le squelette phosphate de la molécule, réduisant ainsi la flexibilité du duplexe. La conformation résultante est différente des **duplexes** ADN, présentant des nucléobases plus rapprochées.



**Figure 5.1 :** Comparaison des structures de l'ADN, de l'ARN et du LNA, dérivé ribonucléotidique caractérisé par un pont méthylène 2'O-4'C.

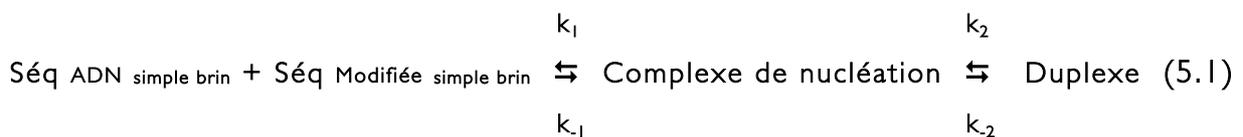


**Figure 5.2 :** Structure d'un duplexe présentant un monomère LNA. Les nucléobases apparaissent en vert, le squelette sucre-phosphate en bleu. Le monomère LNA apparaît en rouge, le pont méthylène 2'O-4'C en rouge foncé. (Nielsen, Singh et al. 1999).

D'après ses résultats, l'incorporation d'un seul LNA ne modifie pas la forme de la double hélice ADN qui reste de forme B (hélice ADN habituellement retrouvée à l'état cellulaire ou en solution aqueuse). Elle constate cependant qu'un changement transitionnel s'est produit à proximité de la substitution du LNA, faisant passer la double hélice de la forme B vers une forme de type A (hélice retrouvée à l'état déshydraté, apparentée à la celle des duplexes ARN). Ces observations seront plus tard confirmées par des études de séquences modifiées par plusieurs LNA (Nielsen, Singh et al. 2000) puis par des études de séquences entièrement composées de LNA (Nielsen, Rasmussen et al. 2004 ; Ivanova and Rosch 2007) (**Figure 5.3**).

D'autre part, il est observé que lorsque les séquences modifiées s'associent avec leurs **séquences complémentaires**, la formation du **duplexe** est accompagnée d'une absorption ionique plus importante et d'une hydratation moins importante du complexe (Kaur, Arora et al. 2008 ; Kaur, Wengel et al. 2008).

L'étude du phénomène d'**hybridation** de séquences modifiées a ensuite constitué une importante thématique dans l'établissement des propriétés des LNA. Elles ont principalement été déterminées par spectrophotométrie à partir de cinétiques de mesures d'absorption à 260 nm. Il est généralement admis que la formation de **duplexe d'oligonucléotides** est un processus comportant deux principaux événements : (1) l'initiation d'un complexe de **nucléation** suivie de (2) l'**hybridation** des séquences l'une sur l'autre, définis selon l'équation :

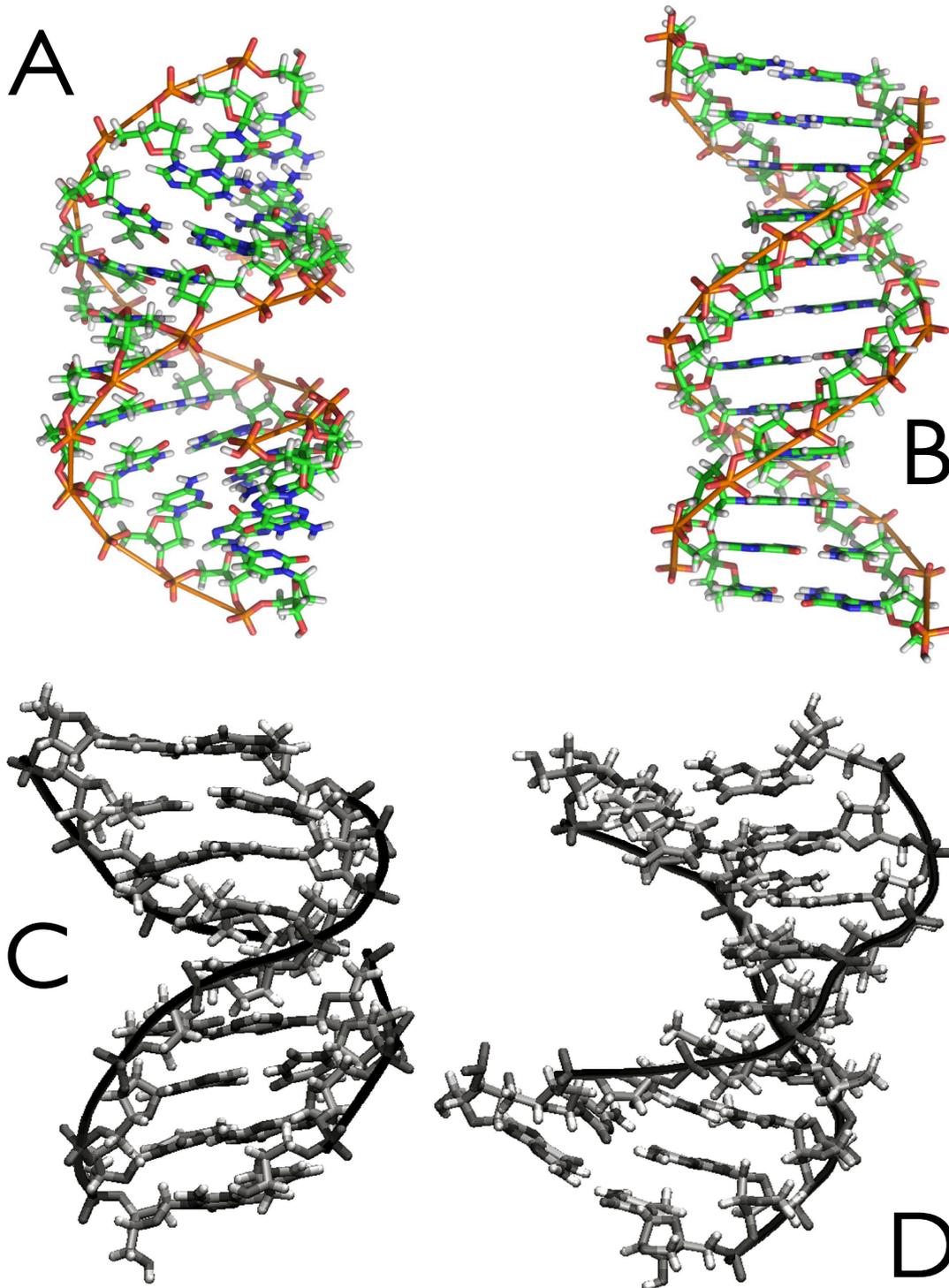


où la constante de dissociation dépendante de la température  $K_D$  est définie par l'équation :

$$K_D = \frac{k_{-1} \times k_{-2}}{k_1 \times k_2} = \frac{[\text{Séq ADN}]_{\text{éq.}} \times [\text{Séq Modifiée}]_{\text{éq.}}}{[\text{Duplexe}]} \quad (5.2)$$

En 2001, CHRISTENSEN *et al.* déterminent les constantes d'association et de dissociation des **duplexes** ADN:ADN et LNA:ADN (Christensen, Jacobsen et al. 2001). Les mesures des constantes d'association calculées pour chacun des **duplexes** sont équivalentes. Cependant, l'étude met en évidence une dissociation plus lente des couples **complémentaires** LNA:ADN, induisant une augmentation significative de la stabilité du **duplexe**. Ces résultats ont été confirmés par **Résonance Plasmonique de Surface (SPR)** (Arora, Kaur et al. 2008). Thermodynamiquement, la stabilisation induite par les monomères LNA résulte donc d'une augmentation favorable de l'enthalpie d'**hybridation** capable de compenser un changement d'entropie défavorable (Kaur, Arora et al. 2008).

Si l'on souhaite efficacement modifier des **oligonucléotides** par des LNA, il est nécessaire de prendre en compte toutes ces modifications. Sur le plan thermodynamique, l'**hybridation** d'un **oligonucléotide** avec sa **séquence complémentaire** peut être considérée comme la résultante de l'**hybridation** de chacune des bases de la séquence (Marky and Breslauer 1982).



**Figure 5.3 :** A) Double hélice droite ARN de forme A. B) Double hélice droite ADN de forme B. Sources : Richard Wheeler - Université d'Oxford. C-D) Comparaison des structures secondaires adoptées par un duplexe ADN:ADN de 9 nucléotides (C) et un duplexe de 9 LNA:ADN (D) (Ivanova and Rosch 2007).

Des modèles dits " *du plus proche voisin* " (Borer, Dengler et al. 1974) ont donc été mis au point afin de permettre la prédiction de ces paramètres thermodynamiques en fonction de la séquence de l'**oligonucléotide**. La stabilité d'un **duplexe** va donc dépendre de la variation de l'énergie libre de Gibbs  $\Delta G^\circ$ . Le nombre de molécules hybridées en fonction de la température dépend quant à lui de la variation d'enthalpie  $\Delta H$  (Stephenson 2010) et dans une moindre mesure, de la variation d'entropie  $\Delta S$ . Ces valeurs peuvent être calculées à toutes les températures à partir de l'équation :

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad (5.3)$$

où :  $G$  est l'énergie libre

$H$  est l'enthalpie

$T$  est la température absolue en **degré Kelvin (°K)**

$S$  est l'entropie

L'appariement de molécules est donc un processus libérateur d'énergie, exothermique ( $\Delta G^\circ$  et  $\Delta H^\circ < 0$ ) : le système tend alors vers un état plus ordonné ( $\Delta S^\circ < 0$ ). Plus la variation de l'énergie libre de Gibbs  $\Delta G^\circ$  est négative, plus la réaction se réalise spontanément et caractérise la stabilité d'un appariement. Inversement, plus la variation de l'énergie libre de Gibbs va tendre vers des valeurs positives, plus un **duplexe** aura tendance à se deshydrider.

Les bases localisées au niveau des **extrémités 5'P** et **3'OH** d'une séquence constituent un point déterminant dans le *design* d'une **amorce PCR**. En 2000, le groupe de John SANTALUCIA Jr. publie d'importants travaux rapportant les paramètres thermodynamiques de séquences ADN présentant des **extrémités** libres non appariées (Bommarito, Peyret et al. 2000), exposés plus tard dans le *Journal Annual Review of Biophysics and Biomolecular Structure* (SantaLucia and Hicks 2004). L'étude rapporte ainsi les incréments de variations d'énergie libre  $\Delta G^\circ_{37}$  observées pour les **extrémités 5'P** et **3'OH** libres non appariées d'un **oligonucléotide**, en 1 M NaCl. (**Tableau 5.1**).

Extrémité libre de la séquence		X = A		X = C		X = G		X = T	
		$\Delta H^\circ$	$\Delta G^\circ_{37}$						
Extrémité 5'P libre	XA/T	0.2	-0.51	0.6	<b>-0.42</b>	-1.1	-0.62	-6.9	-0.71
	XC/G	-6.3	-0.96	-4.4	<b>-0.52</b>	-5.1	-0.72	-4.0	-0.58
	XG/C	-3.7	-0.58	-4.0	<b>-0.34</b>	-3.9	-0.56	-4.9	-0.61
	XT/A	-2.9	-0.50	-4.1	-0.02	-4.2	<b>0.48</b>	-0.2	-0.10
Extrémité 3'OH libre	AX/T	-0.5	-0.12	4.7	<b>0.28</b>	-4.1	-0.01	-3.8	0.13
	CX/G	-5.9	-0.82	-2.6	-0.31	-3.2	<b>-0.01</b>	-5.2	-0.52
	GX/C	-2.1	-0.92	-0.2	<b>-0.23</b>	-3.9	-0.44	-4.4	-0.35
	TX/A	-0.7	-0.48	4.4	<b>-0.19</b>	-1.6	-0.50	2.9	-0.29

**Tableau 5.1** : Variations de l'énergie libre  $\Delta G^\circ_{37}$  (en kcal.mol<sup>-1</sup>) considérées dans le modèle dit " du plus proche voisin " pour les extrémités terminales libres non appariées d'un oligonucléotide, en 1 M NaCl. La barre oblique indique que les séquences sont données dans l'orientation antiparallèle (par exemple, XA/T signifie que l'adénine de la séquence 5'P-XA-3'OH est la base associée à la thymine, selon WATSON & CRICK, et X est non apparié).

Les variations d'enthalpie libre  $\Delta G^\circ_{37}$  rapportées en gras au niveau du **Tableau 5.1** montrent ainsi que la cytosine et la guanine ont plutôt tendance à déstabiliser un **duplexe** ADN lorsqu'elles sont situées au niveau des **extrémités 5'P** et **3'OH** de la séquence d'une **amorce**. Inversement, l'adénine stabilisera plus facilement un **duplexe** lorsqu'elle sera localisée au niveau des **extrémité 5'P** et **3'OH** de cette séquence (Bommarito, Peyret et al. 2000).

Cependant, les premières études thermodynamiques appliquées aux LNA ne sont pas très concluantes. En effet, Niels TOLSTRUP publie *OligoDesign* (Tolstrup, Nielsen et al. 2003) en 2003, un logiciel de conceptualisation d'**oligonucléotides** basé sur le modèle thermodynamique dit " du plus proche voisin " (Borer, Dengler et al. 1974), précédemment amélioré par John SANTALUCIA Jr. (SantaLucia, Allawi et al. 1996; SantaLucia 1998). Créé par la société *Exiqon* dans le cadre des *microarray*, ce logiciel tient compte alors les modifications apportées par les LNA au sein de la séquence par l'analyse d'un score " *flou-logique combiné* ". Ses faibles capacités de prédictions le font alors disparaître très rapidement du site internet de l'entreprise.

En 2004, Patricia McTIGUE réalise la première étude exhaustive sur les paramètres thermodynamiques de l'**hybridation** de LNA en complétant les tables établies auparavant pour la substitution d'un seul monomère LNA. Elle mesure ainsi par spectrophotométrie les variations d'enthalpie  $\Delta H$  et d'entropie  $\Delta S$  observées lors de l'**hybridation** de différents duplexes dont elle détermine le  $T_m$ . Les 32 combinaisons de bases possibles sont analysées ( $5'P-MX^L$  et  $5'P-X^LN$  où **M**, **N** et **X** représentent les quatre bases de l'ADN et  $X^L$  le LNA), les différents paramètres étant calculés à partir de l'équation :

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln(C_T / 4)} \quad (5.4)$$

où :  $R$  est la constante universelle des gaz parfaits (*égale à  $1,9872 \text{ cal.K}^{-1}.\text{mole}^{-1}$* )

$C_T$  est la concentration initiale de simples brins (en égales concentrations)

L'étude fournit alors les règles prédictives de l'**hybridation** d'**oligonucléotides** ADN modifiés par un monomère LNA. Elle montre ainsi le rôle prépondérant des LNA-pyrimidines dans la stabilité d'un **duplexe**. Ce résultat est confirmé par **Calorimétrie Différentielle à balayage** (DSC) (Hughesman, Turner et al. 2008) ainsi que par **Titration Calorimétrique Isotherme** (ITC) (Kaur, Arora et al. 2008 ; Kaur, Wengel et al. 2008). D'autre part, la stabilité de l'**hybridation** est séquence-dépendente : elle va fortement dépendre du contexte de l'incorporation pour chaque base LNA au niveau de la séquence. Les deux voisins du monomère situés en  $5'P$  et  $3'OH$  doivent donc être pris en compte dans la prévision de l'effet d'une incorporation, un voisinage de purines fournissant une plus grande stabilité du **duplexe** (McTigue, Peterson et al. 2004).

La deuxième étude de référence a été réalisée par Richard OWCZARZY en 2011 (Owczarzy, You et al. 2011). Ce travail a été réalisé en collaboration avec le Département de Génétique Moléculaire de la société *Integrated DNA Technologies (IDT)*. Connu pour ses travaux sur les différents paramètres capables d'influencer la **température de fusion** d'une séquence ADN (Owczarzy, You et al. 2004 ; Owczarzy, Moreira et al. 2008; Owczarzy, Tataurov et al. 2008 ), OWCZARZY analyse à son tour l'effet de l'incorporation de plusieurs monomères LNA par mesure de fluorescence (You, Tataurov et al. 2011). Il complète ainsi les paramètres thermodynamiques " *du plus proche voisin* " (Borer, Dengler et al. 1974) qu'il intègre au niveau d'un logiciel disponible en libre accès depuis le site internet d'*IDT* (<http://biophysics.idtdna.com>). On y retrouve de nombreux outils permettant l'amélioration de la conception d'**oligonucléotides** modifiés par LNA, tout en évitant la formation de **structures secondaires** ou d'**hybridations** iso-séquentielles.

Sur la base des travaux de Michael PETERSEN (Petersen and Wengel 2003), on peut résumer les règles d'utilisation des LNA pour des modifications de séquences de la manière suivante :

- L'incorporation d'un ou de plusieurs monomères LNA induit une forte augmentation de stabilité du **duplexe** dans tous les cas avec de l'ARN **complémentaire** et dans la plupart des cas avec de l'ADN **complémentaire**.
- L'amélioration de l'affinité par monomère LNA est généralement plus prononcée pour une interaction avec l'ARN **complémentaire** que pour une interaction avec l'ADN **complémentaire** (augmentation du  $T_m$  de 3 à 7°C pour l'ARN et de 2 à 5°C pour l'ADN).
- L'incorporation de monomères LNA au niveau des **extrémités 3'OH** ou **5'P** induit une augmentation de stabilité thermique moins importante que lorsque cette incorporation est effectuée en milieu de séquence.
- L'augmentation relative de stabilité thermique par monomère LNA pour de courtes séquences est plus importante que celle observée pour des séquences plus longues.
- Les monomères LNA pyrimidiques induisent une augmentation de stabilité thermique plus importante que les monomères LNA puriques.
- Les **duplexes LNA:LNA** sont intrinsèquement très stables. Les possibilités d'**auto-hybridation** iso-séquentielles devraient donc être soigneusement examinées lors de la conception de séquences modifiées par LNA.

L'idée d'associer la PCR et la modification d'**amorces** par LNA apparaît pour la première fois en 2003 avec les travaux de LATORRA *et al.* (Latorra, Campbell *et al.* 2003; Latorra, Hopkins *et al.* 2003). Ces études démontrent que l'incorporation de monomères LNA au niveau des **extrémités 3'OH** des **amorces** de PCR améliore significativement la discrimination allèle-spécifique dans le cadre du génotypage du polymorphisme nucléotidique (SNP). David LATORRA considère alors pour la première fois l'impact de l'incorporation des LNA sur la PCR. Il entreprend alors une étude comparative en faisant varier le positionnement de monomères LNA au niveau de la séquence (Latorra, Arar *et al.* 2003). Il démontre ainsi que les **amorces** modifiées présentent des **températures d'hybridation** maximales plus importantes (en théorie jusqu'à 8°C par nucléotide substitué) et confirme l'amélioration de leurs performances lorsque

des séquences plus courtes sont utilisées. Il souligne également que de trop nombreuses incorporations de LNA au niveau de la séquence de l'amorce sont préjudiciables à l'efficacité de réaction de la PCR (résultat confirmé au cours de la caractérisation cinétique de la formation de duplexe par CHRISTENSSEN *et al.*). L'étude montre ainsi que les monomères LNA doivent être préférentiellement placés au niveau de l'extrémité 3'OH de la séquence et qu'une moindre concentration d'amorces modifiées est nécessaire pour effectuer correctement la réaction.

Les applications PCR basées sur l'utilisation d'oligonucléotides modifiés par LNA se développent rapidement et en 2004, UGOZZOLI *et al.* emploient avec succès des sondes TaqMan modifiées au cours de dosages de mutations de gènes de  $\beta$ -globine et de prothrombine. Les séquences comprennent alors entre 2 et 4 monomères LNA, séparés d'environ 2 à 5 nucléotides. Les résultats montrent que ces sondes présentent une discrimination allélique bien supérieure aux sondes ADN. Les augmentations de sensibilité, de spécificité et de stabilité des duplexes conférées par les LNA rendent ainsi favorable l'utilisation de sondes de plus courte longueur (Ugozzoli, Latorra *et al.* 2004). Ces observations sont confirmées au cours d'études du même type à la fois pour des sondes d'hydrolyse (Costa, Emault *et al.* 2004 ; Johnson, Haupt *et al.* 2004 ; Chou, Meadows *et al.* 2005) et des Molecular Beacon (Sidon, Heimann *et al.* 2006). En 2006, le groupe de Richard OWCZARZY définit les toutes premières règles de *design* à considérer lorsque l'on souhaite conceptualiser des sondes utilisées en PCR en temps réel (You, Moreira *et al.* 2006). Elles peuvent être résumées par les points suivants :

- Placer préférentiellement un triplet de LNA de manière centré au niveau de la base du mésappariement (*exception faite si la sonde contient une guanine au niveau du mésappariement*).
- La substitution des premières bases de la sonde situées au niveau de son extrémité 5'P est à proscrire, sous peine d'augmenter la résistance à l'activité nucléasique de l'enzyme et interfèrerait avec la dégradation souhaitée de la sonde.
- La substitution d'une guanine ou d'une des bases voisines les plus proches doit être évitée au niveau d'un site de mésappariement.
- Des sondes de courte longueur présentent une meilleure discrimination des mésappariements.

- Pour une meilleure qualité de synthèse, il est préférable d'éviter une succession de plus de 4 monomères LNA au niveau de la séquence de la sonde.
- Une meilleure discrimination semble être obtenue lorsque la position du site de mésappariement est à proximité du centre de la sonde. La discrimination est sensiblement diminuée si le mésappariement est localisé à proximité des extrémités du duplexe.
- Les sondes ne doivent pas se replier en structures secondaires iso-séquentielles stables ou pouvant former des auto-dimères, en particulier lorsque ces structures présentent des paires de bases LNA-LNA.
- Si la position du mésappariement est inconnue, la meilleure approche consiste à modifier toutes les adénines jusqu'à ce qu'une température de fusion optimale soit obtenue, les règles énoncées précédemment devant toujours être prises en compte.

L'ensemble de ces recommandations est intégré au niveau du logiciel proposé par IDT depuis son site internet (<http://biophysics.idtdna.com>).

LEVIN *et al.* affirment la même année que l'incorporation de monomères LNA à proximité de l'extrémité 5'P améliore significativement les performances des amorces de PCR, contrairement à une incorporation à proximité de l'extrémité 3'OH (gène la *Taq* polymérase), ou bien lorsque les monomères LNA sont répartis uniformément et de manière espacée sur la longueur de la séquence. Ils suggèrent que l'incorporation de LNA à proximité de l'extrémité 5'P augmente la stabilité de l'amorce lors de son hybridation au niveau du site spécifique tout en interdisant l'extension des autres amorces si celles-ci se sont hybridées au niveau de sites non spécifiques (Levin, Fiala *et al.* 2006 ; Malgoyre, Banzet *et al.* 2007).

Lorsque les transcrits d'un gène sont faiblement exprimés et qu'ils doivent être quantifiés à partir d'échantillons biologiques, des amplifications stochastiques non spécifiques sont fréquemment observées. Dans le but de s'affranchir de ce type d'amplifications problématiques, MALGOYRE *et al.* proposent d'améliorer la détection de l'interleukine 6 (IL-6) en défavorisant les amplifications non spécifiques par l'utilisation d'amorces modifiées par 4 monomères LNA répartis au niveau de la partie 5'P de la séquence. Les  $T_m$  élevées de ces amorces permettent alors la réalisation d'une PCR à une  $T_H$  plus élevée d'une part et une diminution de la probabilité d'amplifications non spécifiques d'autre part, en raison de la dissociation possible de l'amorce au niveau de son extrémité 3'OH en cas d'hybridations non spécifiques. Les résultats de cette étude montrent ainsi qu'une diminution significative d'amplifications de produits non

spécifiques est obtenue pour la quantification de l'IL-6 de rat (Malgoyre, Banzet et al. 2007). L'augmentation de la température d'hybridation obtenue reste cependant inférieure aux prédictions thermodynamiques.

L'ensemble de ces résultats constitue donc le point de départ de cette étude dont l'objectif principal est l'optimisation du bénéfice thermodynamique des LNA au sein de séquences d'amorces de PCR afin de les rendre encore plus spécifiques de leur cible. Sur les critères d'efficacité de réaction, de spécificité et de sensibilité, une analyse de la nature et de la position des bases modifiées sera menée dans le but de définir les règles essentielles à considérer si l'on souhaite modifier efficacement des amorces de PCR par incorporation de monomères LNA.

## 5.2. Résultats expérimentaux

Le **Tableau 5.2** présente les séquences dessinées avec le logiciel d'*Exiqon*. Le but de ce travail de *design* est de maximiser le  $T_m$  par incorporation de LNA. En théorie, l'incorporation de ces monomères permet une augmentation de 2 à 6°C de la température de fusion d'un oligonucléotide ADN. Cette augmentation de température permet donc une utilisation de nos amorces modifiées soit à une  $T_H$  plus élevée, soit à une concentration d'amorces moins importante par rapport aux amorces ADN. Les résultats présentés dans ces paragraphes tiennent compte de ces deux possibilités.

Oligonucléotides	Séquence 5'P – 3'OH	$T_m$ Exiqon (°C)
RIL6 F0	GCGATGATGCACTGTCAGAAAAC	65
RIL6 R0	AACGGA ACTCCAGAAGACCAGAG	66
RIL6 F1L	GCGA <b>T</b> GATG <b>T</b> GC <b>A</b> CTGTCAGAAAAC	72
RIL6 R1L	AACGG <b>A</b> ACT <b>T</b> CC <b>A</b> GAAGACCAGAG	72
RIL6 F2L	G <b>C</b> GATGATG <b>C</b> ACTGTCAGAAAAC	69
RIL6 R2L	A <b>A</b> CGGAACT <b>C</b> CAGAAGACCAGAG	69
RIL6 F3L*	G <b>C</b> GAT <b>G</b> ATGCACTGTCAGAAAAC	69
RIL6 R3L	A <b>A</b> CG <b>G</b> AACTCCAGAAGACCAGAG	70
RIL6 F4L	GCGAT <b>G</b> ATG <b>C</b> ACTGTCAGAAAAC	69
RIL6 R4L	AACGG <b>A</b> ACT <b>C</b> CAGAAGACCAGAG	69
RIL6 F5L**	GCGA <b>T</b> GATG <b>C</b> ACTGTCAGAAAAC	70
RIL6 R5L	AA <b>C</b> GGAACT <b>C</b> CAGAAGACCAGAG	70
RIL6 F6L*	G <b>C</b> GAT <b>G</b> ATGCACTGTCAGAAAAC	69
RIL6 R6L	AA <b>C</b> GGAA <b>C</b> TCCAGAAGACCAGAG	70
RIL6 F7L**	GCGA <b>T</b> GATG <b>C</b> ACTGTCAGAAAAC	70
RIL6 R7L	AACGGAA <b>C</b> T <b>C</b> CAGAAGACCAGAG	69

**Tableau 5.2** : Caractéristiques des oligonucléotides modifiés par LNA spécifiques de l'interleukine 6, IL-6 (GenBank NM\_012589). Les  $T_m$  ont été calculés par le logiciel *Exiqon* (<http://www.exiqon.com/ls/Pages/ExiqonTMPredictionTool.aspx>) pour une concentration de 0,3  $\mu$ M d'amorces ; 32 mM Na<sup>+</sup>. Ces oligonucléotides sont synthétisés puis purifiés par Chromatographie en phase Liquide à Haute Performance, HPLC (Eurogentec, Saraing, Belgique). Les substitutions de LNA sont marquées en plus grande lettre. Localisations : RIL6 FX, 300-322 ; RIL6 RX, 428-406. \*RIL6 F3L = RIL6 F6L ; \*\*RIL6 F5L = RIL6 F7L.

### 5.2.1. Efficacité PCR & température d'hybridation

Afin de comparer objectivement les différentes amorces entre elles, nous avons dans un premier temps validé la  $T_H$  optimale des amorces ADN à partir d'un pool d'échantillons biologiques (Nomenclature : RIL6 F0-R0). Bien que vérifiées préalablement au cours de la précédente étude réalisée par MALGOYRE *et al.*, nous avons confirmé les caractéristiques de la PCR réalisée avec le couple d'amorces ADN à une  $T_H$  de 58°C et 60°C sur les mêmes dilutions. Les caractéristiques des courbes de calibration sont présentées au niveau du **Tableau 5.3**. Ainsi, pour mémoire, on peut utiliser une  $T_H$  de 62°C ou 64°C à condition d'augmenter la concentration d'amorces dans le milieu réactionnel à 0,6 et 0,7  $\mu\text{M}$ , respectivement. Cette augmentation de concentration entraîne cependant une perte de spécificité ((Malgoyre, Banzet *et al.* 2007), résultats non publiés).

	$T_H$ (en °C)	Paramètres de calibration de la courbe			
		$E_{\text{PCR}}^a$	pente	$r^b$	Erreur
RIL6 F0-R0	58	90,0	-3,60	1,00	0,02
	60	80,9	-3,89	1,00	0,04

**Tableau 5.3 :** Caractéristiques des courbes de calibration obtenues pour les amorces ADN RIL6 F0-R0 à partir des duplicates de quatre dilutions en cascade issues d'un pool d'ADNc, en système *SybrGreen I*, à la concentration d'amorces de 0,4  $\mu\text{M}$ , 4 mM  $\text{MgCl}_2$  ; <sup>a</sup> efficacité PCR déterminée à partir de la pente de la courbe de calibration, <sup>b</sup>  $r$  : coefficient de régression linéaire de la courbe de calibration.

Une perte de plus de 9% d'efficacité et une augmentation du taux d'erreur PCR sont observées lors de l'augmentation de la température d'hybridation des amorces ADN de 2°C. Ces résultats confirment une température d'hybridation optimale de 58°C pour les amorces non modifiées.

L'efficacité de PCR des différents couples d'amorces a ensuite été évaluée à partir de dilutions d'ADN recombinant. Trois températures d'hybridation sont dans un premier temps testées (58, 60 et 62°C). Les couples présentant plus de 90% d'efficacité à la  $T_H$  de 62°C sont testés une dernière fois à la température d'hybridation de 64°C (Figure 5.4).

L'amplicon contenant la séquence recherchée est obtenu à l'aide d'une première PCR effectuée sur de l'ADN complémentaire présentant la cible recherchée. À partir du produit récupéré après réaction, une série de dilutions est effectuée en cascades, de manière à constituer les différents points d'une gamme de quantification. Afin d'apprécier l'identité du produit amplifié, une analyse par courbe de fusion en système *SybrGreen I* est effectuée à la fin de chaque PCR. Les différents paramètres obtenus à partir de la gamme de quantification permettent ainsi l'évaluation des différents couples d'amorces. En effet, l'efficacité de réaction est déduite de la pente de la courbe. La sensibilité de détection des amorces est appréciée par les valeurs relatives des  $C_q$  obtenus. Enfin, la spécificité de détection est évaluée par l'analyse de la courbe de fusion et la dispersion des duplicatas pratiqués pour chaque dilution.

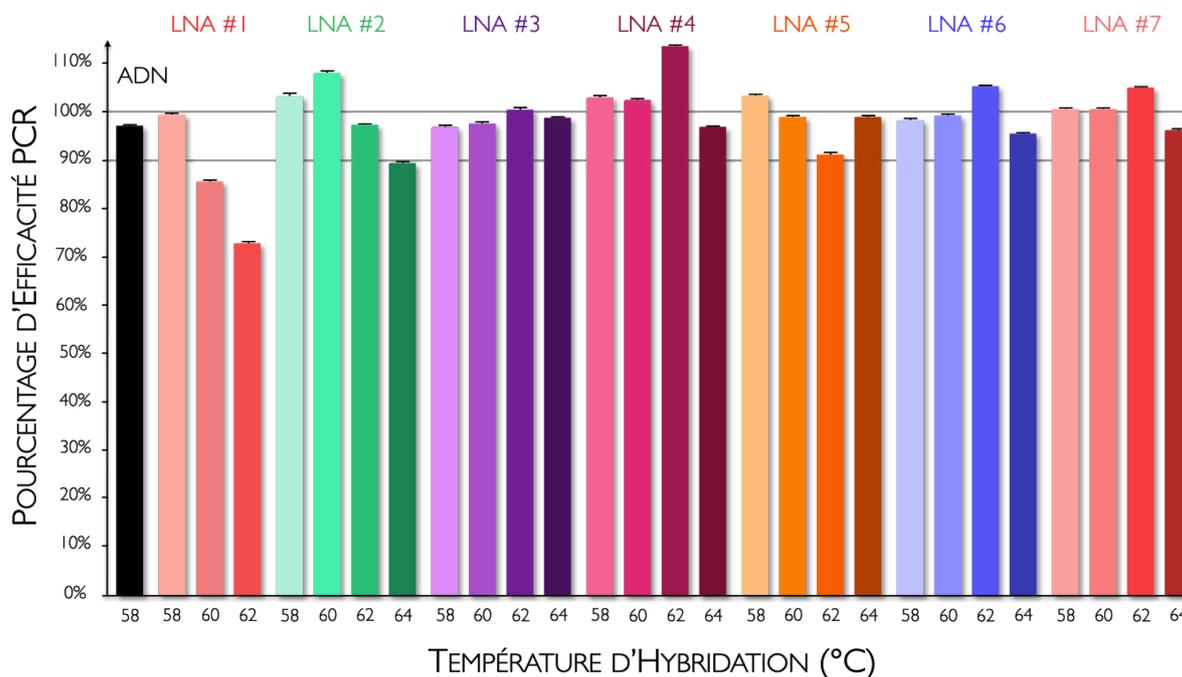
Les résultats montrent qu'une bonne efficacité de PCR est obtenue avec le couple d'amorces ADN à la  $T_H$  de 58°C (97,2%, Tableau 5.3).

Les résultats obtenus pour le couple d'amorces #1, présentant 4 incorporations de monomères LNA, montrent une diminution drastique de l'efficacité de PCR lorsque la  $T_H$  passe de 60°C à 62°C (respectivement de 99,7% à 73,0%, Tableaux 5.4). L'augmentation du taux d'erreur jusqu'à 0,25 indique une augmentation importante de la variabilité de la réaction à la  $T_H$  de 64°C. Une température d'hybridation optimale est donc établie à la  $T_H$  de 60°C pour ce couple.

Concernant les amorces modifiées par 2 incorporations de monomères LNA, on constate des valeurs d'efficacité élevées à toutes les  $T_H$  testées, toutes compatibles pour une quantification de la cible. Sur les trois premières températures testées (58, 60 et 62°C), les couples #3, #4, #6 et #7 présentent une augmentation de l'efficacité de réaction avec des valeurs maximales mesurées à la  $T_H$  de 62°C (respectivement 100,8%, 113,5%, 110,0% et 105,3%). Seuls les couples #2 et #5 présentent une diminution d'efficacité de 60°C à 62°C (respectivement de 10,6% et 7,7%). D'autre part, les taux d'erreur PCR les plus importants sont observés pour le couple #2 (respectivement 0,16 ; 0,20 ; 0,12 aux  $T_H$  58°C, 60°C et 62°C respectivement). Les taux d'erreur les moins élevés sont observés pour les PCR effectuées aux trois  $T_H$  avec le couple d'amorces #5 (0,06 ; 0,10 ; 0,06). Le couple d'amorces #6 montre une augmentation de l'efficacité de PCR aux trois températures étudiées et présente une diminution de la variabilité de la réaction à la  $T_H$  de 62°C (respectivement 0,16 ; 0,11 et 0,06).

Au cours de l'évaluation des couples d'amorces à la  $T_H$  de 64°C, on observe des efficacités de réactions comprises entre 90 et 100% pour les couples #3, #4, #5, #6 et #7. Une température d'hybridation optimale est donc établie à la  $T_H$  de 64°C pour l'ensemble de ces couples. On

détermine une température d'hybridation optimale de 62°C pour le couple #2. Les caractéristiques des différentes gammes sont rapportées au niveau des Tableaux 5.4.



**Figure 5.4 :** Efficacités PCR des différents couples d'amorces en fonction de la température d'hybridation ( $T_H$ ) obtenues à partir de duplicates de dilutions d'ADN recombinant en système *SybrGreen I*. Les valeurs sont présentées avec un intervalle de confiance 95%. La concentration d'amorces est de 0,4  $\mu$ M, la concentration de  $MgCl_2$  est de 4 mM.

La température d'hybridation optimale des amorces ADN est confirmée à 58°C. L'incorporation de 4 monomères LNA au niveau de la partie 5'P de l'amorce (LNA #1) ne permet qu'une très faible augmentation de la température d'hybridation (établie à 60°C pour ce couple), contrairement aux autres couples ne présentant que seulement 2 incorporations.

Les  $T_H$  optimales sont estimées à 62°C pour le couple LNA #2 et à 64°C pour les couples #3, #4, #5, #6 et #7 à la  $T_H$  de 64°C.

	$T_H$ °C	Paramètres de calibration de la courbe			
		pente	$E_{PCR}^a$	$r^b$	Erreur
RIL6 FO-R0	58	-3,39	97,2	1,00	0,12
RIL6 FIL-R1L	58	-3,33	99,7	1,00	0,09
	60	-3,72	85,8	1,00	0,07
	62	-4,20	73,0	0,98	0,25
RIL6 F2L-R2L	58	-3,24	103,7	0,98	0,16
	60	-3,14	108,2	0,98	0,20
	62	-3,38	97,6	1,00	0,12
	64	-3,60	89,7	0,99	0,03
RIL6 F3L-R3L	58	-3,39	97,2	1,00	0,15
	60	-3,38	97,7	0,99	0,03
	62	-3,30	100,8	0,98	0,19
	64	-3,35	99,0	0,99	0,06
RIL6 F4L-R4L	58	-3,25	103,2	0,98	0,23
	60	-3,26	102,6	0,98	0,17
	62	-3,04	113,5	0,99	0,09
	64	-3,39	97,1	0,99	0,06
RIL6 F5L-R5L	58	-3,24	103,5	0,99	0,06
	60	-3,34	99,1	1,00	0,10
	62	-3,55	91,4	0,99	0,06
	64	-3,34	99,2	0,99	0,05
RIL6 F6L-R6L	58	-3,36	98,6	0,98	0,16
	60	-3,34	99,4	1,00	0,11
	62	-3,20	105,4	0,98	0,06
	64	-3,43	95,7	0,99	0,05
RIL6 F7L-R7L	58	-3,31	100,7	1,00	0,10
	60	-3,30	100,8	1,00	0,13
	62	-3,20	105,3	1,00	0,12
	64	-3,41	96,4	0,99	0,04

**Tableaux 5.4 :** Efficacités PCR des différents couples d'amorces en fonction de la température d'hybridation ( $T_H$ ). Les résultats sont obtenus à partir de duplicates de dilutions d'ADN recombinant de concentrations croissantes analysées en duplicates en système *SybrGreen I*. La concentration d'amorces est de 0,4  $\mu$ M, la concentration de  $MgCl_2$  est de 4 mM. <sup>a</sup> efficacité PCR déterminée à partir de la pente de la courbe de calibration, <sup>b</sup>  $r$  : coefficient de régression linéaire de la courbe de calibration.

## 5.2.2. Prédictions de la température de fusion des amorces modifiées

Le pourcentage d'amorces fixées à une température d'hybridation donnée représente le paramètre le plus intéressant pour l'estimation de capacités d'hybridation d'un couple d'amorces (SantaLucia 2007). Cette valeur n'est pas directement liée au  $T_m$  mais dépend du  $\Delta H$  ainsi que du  $\Delta S$ . Elle peut donc être calculée en théorie mais n'est jamais utilisée.

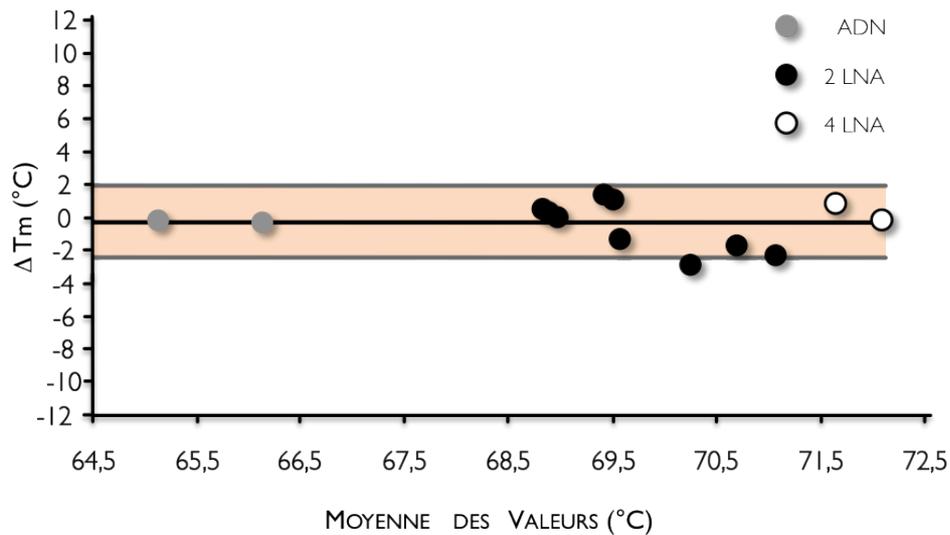
Au cours de cette étude, nous avons plutôt choisi de considérer la température de fusion ( $T_m$ ) comme paramètre de contrôle de l'hybridation, en nous basant sur les valeurs prédites par le logiciel d'Exiqon. On sait que l'incorporation de monomères LNA augmente significativement la température de fusion d'un oligonucléotide (Wengel, Petersen et al. 2003). Une différence minimale existant entre les  $T_m$  de deux amorces est considérée comme un gage de performance pour la PCR, certifiant que les séquences s'hybrideront d'autant plus simultanément que cette valeur sera faible. Au cours de notre design, nous avons donc cherché à diminuer cette différence (homogénéité des  $T_m$  de deux amorces) tout en maximisant le gain de température obtenu par rapport aux  $T_m$  des amorces ADN, à l'aide du logiciel d'Exiqon.

### 5.2.2.1. Comparaisons de trois méthodes de prédiction du $T_m$

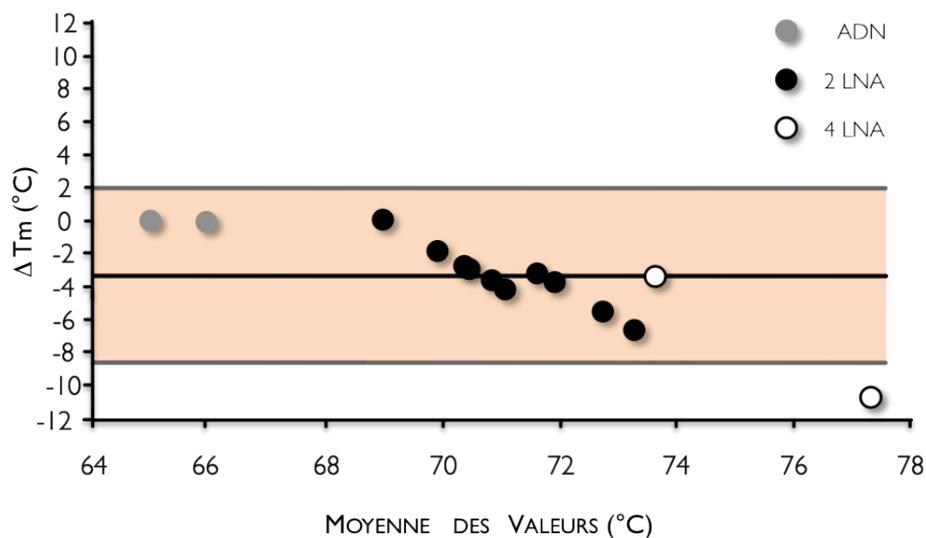
Dans un premier temps, des prédictions de  $T_m$  sont effectuées à partir du logiciel d'Exiqon (<http://lna-tm.com/>), du logiciel *OligoAnalyser 3.1.* d'IDT (Owczarzy, Tataurov et al. 2008) (<http://biophysics.idtdna.com>) et des données obtenues par McTIGUE et al. en 2004 (McTigue, Peterson et al. 2004) (l'ensemble de ces données est restitué au niveau du **Tableau 5.5**).

En comparant les deux premiers logiciels par la méthode de *Bland-Altman* et dans des conditions salines identiques à celles d'un milieu de PCR (32 mM de  $\text{Na}^+$ ), on observe un premier agrément correct avec un intervalle d'agrément de 4,3°C, sans aucun biais de température. Ce résultat montre que les deux logiciels apparaissent sensiblement équivalents dans la prédiction du  $T_m$  des amorces (**Figure 5.5**).

En revanche, un deuxième agrément beaucoup plus dramatique existant entre la méthode prédictive de McTIGUE (développée pour une seule substitution LNA dans un milieu à 1M  $\text{Na}^+$ ) et le logiciel d'Exiqon (en condition saline de PCR) illustre l'impact de la concentration en sodium et la présence d'incorporations multiples sur la valeur prédictive du  $T_m$  (intervalle d'agrément 95% de 10,5°C ; biais de -3,3°C ; **Figure 5.6**).

EXIQON (32 mM Na<sup>+</sup>) – OLIGOANALYSER 3.1 (32 mM Na<sup>+</sup>)

**Figure 5.5 :** Agrément effectué entre les valeurs de  $T_m$  prédites obtenues par le logiciel *Exiqon* et les valeurs de  $T_m$  obtenues par le logiciel d'*IDT*, en comparant les différences par la méthode de *Bland-Altman*. L'abscisse et l'ordonnée représentent respectivement la moyenne et la différence des  $T_m$  mesurés par les deux méthodes. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

EXIQON (32 mM Na<sup>+</sup>) – PRÉDICTION MCTIGUE (1M Na<sup>+</sup>)

**Figure 5.6 :** Agrément effectué entre les valeurs de  $T_m$  prédites obtenues par le logiciel *Exiqon* et les valeurs de  $T_m$  obtenues par la prédiction de *McTIGUE*, en comparant les différences par la méthode de *Bland-Altman*. L'abscisse et l'ordonnée représentent respectivement la moyenne et la différence des  $T_m$  mesurés par les deux méthodes. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

Devant cette incohérence et dans un second temps, nous avons du développer puis valider une méthode de mesure plus précise du  $T_m$  en mesurant directement sa valeur dans le milieu de PCR (cf. Chapitre 6).

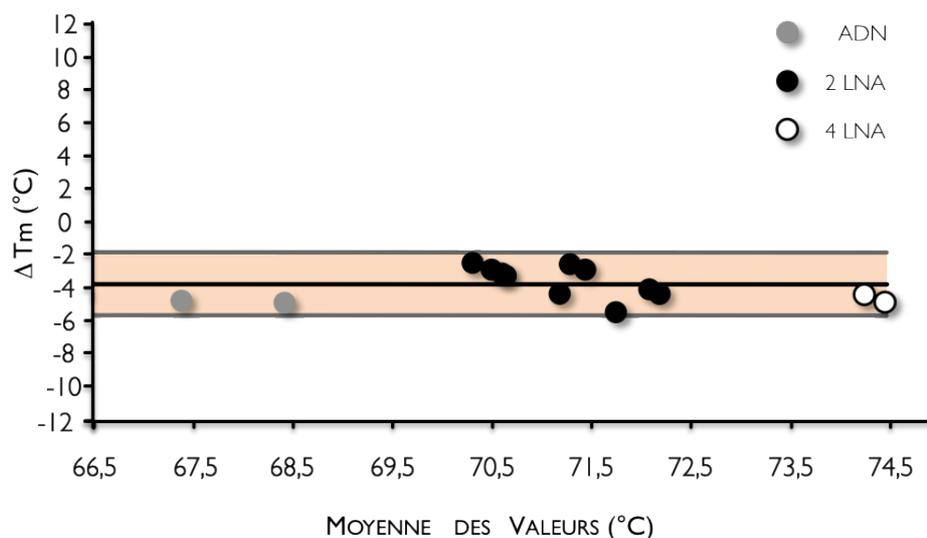
#### 5.2.2.2. Comparaisons des méthodes de prédiction du $T_m$ avec la mesure directe par courbe de fusion

En comparant maintenant les valeurs prédites par le logiciel d'Exiqon et les mesures directes obtenues par *courbe de fusion*, on observe un important biais de  $-3,8^\circ\text{C}$  malgré un intervalle d'agrément convenable ( $3,8^\circ\text{C}$ ). Ce résultat montre que le logiciel ne prend pas en compte la totalité des composants d'un milieu de PCR. De ce fait, il restitue une valeur biaisée qui ne correspond pas à la *température de fusion* réelle mesurée dans le milieu. L'intervalle de  $3,8^\circ\text{C}$  souligne toutefois une cohérence des mesures réalisées par les deux méthodes (**Figure 5.7**). Un résultat similaire a été obtenu en comparant le logiciel *OligoAnalyser 3.1. d'IDT* avec la méthode de mesure directe dans les mêmes conditions ioniques (*résultat non montré*).

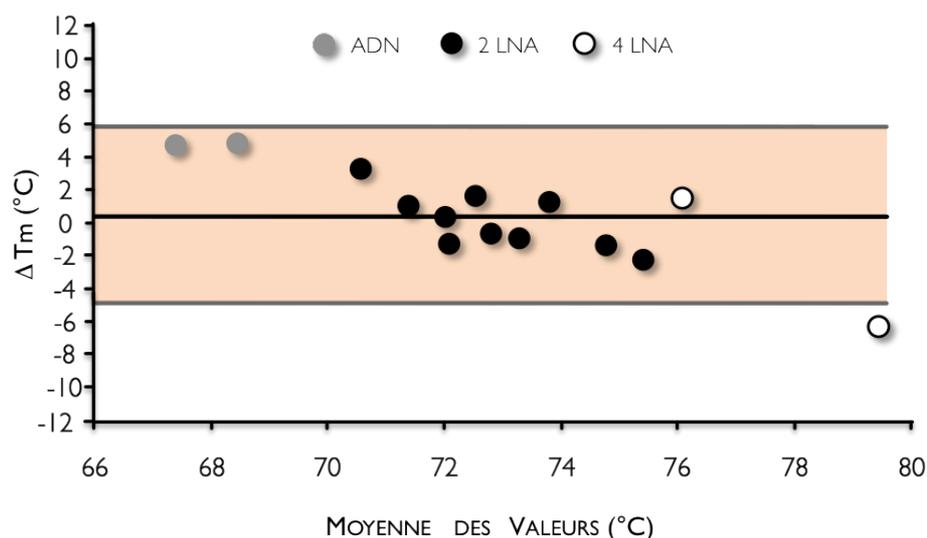
Un agrément tout aussi mauvais est observé entre les valeurs prédites par la méthode de McTIGUE et les mesures directes obtenues par *courbe de fusion* (intervalle d'agrément 95% de  $10,7^\circ\text{C}$  ; biais de  $0,5^\circ\text{C}$  ; **Figure 5.6**). Ce résultat illustre à nouveau la mauvaise efficacité de cette méthode prédictive face à des séquences d'*amorces* présentant 2 à 4 incorporations de LNA.

Les résultats montrent donc que les prédictions de  $T_m$  obtenues à partir des trois méthodes prédictives utilisées (*Exiqon*, *IDT* et *McTIGUE*) varient de manière importante.

En conclusion, les résultats montrent que les prédictions de  $T_m$  obtenues à partir des trois méthodes prédictives utilisées (*Exiqon*, *IDT* et *McTIGUE*) varient de manière importante. D'autre part, la comparaison réalisée avec la mesure par courbe de fusion souligne que les logiciels ne prennent pas en compte tous les constituants du milieu. Il semble donc nécessaire d'utiliser une méthode de mesure directe du  $T_m$  dans le milieu de PCR.

EXIQON (32 mM Na<sup>+</sup>) – COURBE DE FUSION (32 mM Na<sup>+</sup>)

**Figure 5.7 :** Agrément effectué entre les valeurs de  $T_m$  prédites obtenues par le logiciel *Exiqon* et les mesures de  $T_m$  obtenues par courbe de fusion, à la concentration d'amorces de 0,6  $\mu$ M, en comparant les différences par la méthode de *Bland-Altman*. L'abscisse et l'ordonnée représentent respectivement la moyenne et la différence des  $T_m$  mesurés par les deux méthodes. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

PRÉDICTION MCTIGUE (1M Na<sup>+</sup>) – COURBE DE FUSION (32 mM Na<sup>+</sup>)

**Figure 5.8 :** Agrément effectué entre les valeurs de  $T_m$  obtenues par la prédiction de McTigue et les mesures de  $T_m$  obtenues par courbe de fusion, à la concentration d'amorces de 0,6  $\mu$ M, en comparant les différences par la méthode de *Bland-Altman*. L'abscisse et l'ordonnée représentent respectivement la moyenne et la différence des  $T_m$  mesurés par les deux méthodes. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

Oligonucléotides	Séquence 5'P – 3'OH	Prédiction $T_m$ Exiq <sup>n†</sup> (en °C)	Gain <sup>Exiq<sup>n</sup></sup> (en °C)	$\Delta T_m$ Exiq <sup>n</sup> (en °C)	Prédiction $T_m$ IDT <sup>‡</sup> (en °C)	Gain <sup>IDT</sup> (en °C)	$\Delta T_m$ IDT (en °C)	Prédiction $T_m$ McTigue <sup>§</sup> (en °C)	Gain <sup>McTigue</sup> (en °C)	$\Delta T_m$ McTigue (en °C)	Mesure $T_m$ Flu <sup>#</sup> (en °C)	Gain <sup>Flu<sup>#</sup></sup> (en °C)	$\Delta T_m$ Flu <sup>#</sup> (en °C)
RIL6 F0L	GCGATGATGCACTGTCAGAAAAC	65,0		$\Delta = 1,0$	65,3		$\Delta = 1,1$	65,0		$\Delta = 1,4$	69,8		$\Delta = 1,1$
RIL6 R0L	AACGGAAGCTCCAGAAGACCAGAG	66,0			66,4			66,0			70,9		
RIL6 F1L	GCGATGATGCACTGTCAGAAAAC	72,0	7	$\Delta = 0,0$	71,3	6,0	$\Delta = 0,9$	82,6	17,6	$\Delta = 8,4$	76,4	6,6	$\Delta = 0,5$
RIL6 R1L	AACGGAAGCTCCAGAAGACCAGAG	72,0	6		72,2	5,8		75,3	9,3		76,9	6,0	
RIL6 F2L	GCGATGATGCACTGTCAGAAAAC	69,0	4	$\Delta = 0,0$	69,0	3,7	$\Delta = 1,1$	72,6	7,6	$\Delta = 1,9$	71,6	1,8	$\Delta = 0,7$
RIL6 R2L	AACGGAAGCTCCAGAAGACCAGAG	69,0	3		70,1	3,7		71,8	5,8		72,3	1,4	
RIL6 F3L*	GCGATGATGCACTGTCAGAAAAC	69,0	4	$\Delta = 1,0$	68,8	3,5	$\Delta = 0,0$	70,8	5,8	$\Delta = 1,2$	72,0	2,2	$\Delta = 0,5$
RIL6 R3L	AACGGAAGCTCCAGAAGACCAGAG	70,0	4		68,8	2,4		73,1	7,1		72,5	1,6	
RIL6 F4L	GCGATGATGCACTGTCAGAAAAC	69,0	4	$\Delta = 0,0$	68,7	3,4	$\Delta = 1,4$	68,9	3,9	$\Delta = 1,7$	72,2	2,4	$\Delta = 1,2$
RIL6 R4L	AACGGAAGCTCCAGAAGACCAGAG	69,0	3		70,1	3,7		71,7	5,7		73,4	2,5	
RIL6 F5L**	GCGATGATGCACTGTCAGAAAAC	70,0	5	$\Delta = 0,0$	69,0	3,7	$\Delta = 2,4$	73,7	8,7	$\Delta = 0,6$	72,9	3,1	$\Delta = 1,2$
RIL6 R5L	AACGGAAGCTCCAGAAGACCAGAG	70,0	4		71,4	5,0		75,4	9,4		74,1	3,2	
RIL6 F6L*	GCGATGATGCACTGTCAGAAAAC	69,0	4	$\Delta = 1,0$	68,8	3,5	$\Delta = 3,3$	70,8	5,8	$\Delta = 4,6$	72,0	2,2	$\Delta = 2,3$
RIL6 R6L	AACGGAAGCTCCAGAAGACCAGAG	70,0	4		72,1	5,7		76,5	10,5		74,3	3,4	
RIL6 F7L**	GCGATGATGCACTGTCAGAAAAC	70,0	5	$\Delta = 1,0$	69,0	3,7	$\Delta = 2,5$	73,7	8,7	$\Delta = 1,7$	72,9	3,1	$\Delta = 1,6$
RIL6 R7L	AACGGAAGCTCCAGAAGACCAGAG	69,0	3		71,5	5,1		73,1	7,1		74,5	3,6	

**Tableau 5.5 :** Prédiction de la température de fusion des oligonucléotides modifiés par l'incorporation de mères LNA. Les prédictions rapportées sont obtenues † à partir du logiciel Exiqon (0,3 µM d'amères, 32 mM Na<sup>+</sup>), ‡ à partir du logiciel OligoAnalyser 3.1. d'IDT (0,3 µM d'amères, 32 mM Na<sup>+</sup>) § à partir du modèle du plus proche voisin (SantaLucia 1998) et des travaux réalisés par McTIGUE et al. (2004) (McTigue, Petersen et al. 2004) (1 M Na<sup>+</sup>), # à partir de la mesure directe du  $T_m$  à 0,6 µM d'amères, 32 mM Na<sup>+</sup>, 4 mM MgCl<sub>2</sub>. Les meilleurs gains et  $\Delta T_m$  apparaissent sur fond vert, les plus mauvais sur fond rose. Tous les oligonucléotides sont synthétisés puis purifiés par HPLC (Eurogentec, Saraing, Belgique). \*RIL6 F3L = RIL6 F6L, \*\*RIL6 F5L = RIL6 F7L.

### 5.2.3. Conception d'un modèle d'évaluation de la spécificité d'amorces modifiées

#### 5.2.3.1. Essais préliminaires

La mise au point d'un modèle d'évaluation objectif de la spécificité des amorces modifiées par incorporations de LNA a constitué un travail long et laborieux. De nombreux protocoles ont du être utilisés et, bien qu'ils se soient montrés infructueux, sont importants à détailler ici.

En effet, nous avons initialement cherché à simuler un échantillon biologique au moyen de différentes matrices. Le contexte de l'étude de MALGOYRE *et al.* (Malgoyre, Banzet et al. 2007) était la détection d'une cible rare au niveau d'échantillons biologiques. Nous avons donc amplifié la séquence d'intérêt afin de pouvoir la diluer au niveau d'échantillons biologiques dépourvus d'IL-6 (ADNc de tissu adipeux de porc charcutier, ADNc issus d'ARN dégradés avant RT, etc ...). La spécificité de la détection était alors vérifiée par courbe de fusion. Plusieurs PCR ont ainsi été réalisées à partir des mêmes dilutions et à plusieurs reprises sur un même thermocycleur, puis sur des thermocycleurs différents afin de vérifier la reproductibilité de nos résultats. Ils se sont finalement avérés inutilisables en raison d'une trop importante variabilité au niveau de la détection. Nous avons la nécessité de contrôler le nombre de copies de matrices spécifiques et non spécifiques au niveau de notre système. Pour cela, nous avons d'abord synthétisé des matrices non spécifiques présentant 3 mutations au niveau des extrémités 3'OH et 5'P, en supposant une efficacité d'hybridation plus faible de la part des amorces pour ces matrices. Cette approche s'est avérée inadaptée. En effet, si  $\epsilon$  représente l'efficacité d'hybridation d'une amorce sur une cible non spécifique, et que l'amorce s'hybride sur deux structures identiques, alors l'efficacité réelle de la PCR est  $\epsilon^2$ . Ainsi, s'il n'existe pas suffisamment de mutations au niveau de cette matrice non spécifique, les amorces s'y hybrident avec une efficacité  $\epsilon$  de presque 100%. En revanche, s'il existe beaucoup trop de mutations, l'efficacité d'hybridation diminue nettement, conduisant à une absence d'efficacité de PCR. Nous avons donc créé plusieurs matrices non spécifiques par PCR présentant un nombre de mutations différent, uniquement placées sur l'une des extrémités de la séquence. Cette approche a eu pour avantage le fait de pouvoir tester séparément chaque amorce.

La quantité relative des **matrices** spécifiques et non spécifiques s'est avérée problématique en système *SybrGreen 1*. En effet, cet intercalant non spécifique de l'ADN ne permet pas de quantifier correctement l'amplification non spécifique en raison d'un ratio  $\frac{\text{molécules spécifiques}}{\text{molécules non spécifiques}}$  pouvant être différent entre le moment de la quantification et la fin de la réaction (phase plateau). La révélation de la détection au moyen de **sondes** d'hydrolyse (Holland, Abramson et al. 1991), constitua finalement la solution de notre problème de quantification, l'émission de fluorescence d'un **fluorophore** correspondant à la détection spécifique d'une cible amplifiée.

### 5.2.3.2. Le modèle d'évaluation de la spécificité des amorces

Le modèle d'évaluation repose sur l'utilisation du système de **sonde** d'hydrolyse sur deux types de **matrices** synthétiques (**Tableau 3.3, chapitre Matériels & Méthodes**) :

- une **matrice** spécifique dont la séquence correspond à celle de l'IL-6 (Nomenclature : **RIL6LNAMatrix**). **Amorces** RIL6 à tester ainsi que **sonde** d'hydrolyse sont par conséquent capables de s'y hybrider spécifiquement (**Alignements 3.3, chapitre Matériels & Méthodes**).
- une **matrice** non spécifique dont la séquence comporte à la fois des points de mutation au niveau de ses **extrémités 5'P** ou **3'OH** et une séquence correspondant à l'EGFP (Nomenclature : **Matrice non spécifique 3F** et **Matrice non spécifique 3R**). Sa détection spécifique rendra compte des amplifications non spécifiques.

Le système d'évaluation permet donc la détection spécifique des amplifications de l'IL6 ainsi que la détection spécifique des amplifications non spécifiques pouvant éventuellement se produire au cours de la réaction grâce à l'utilisation de la séquence de l'EGFP (**Alignements 3.3, chapitre Matériels & Méthodes**).

Dans un premier temps, il a été nécessaire d'évaluer le degré de mutation suffisant de la **matrice** non spécifique. En effet, si cette **matrice** présente trop de différences avec les séquences des **amorces** IL-6, les **hybridations** non spécifiques ne sont pas favorisées. Six **amorces** ont donc été dessinées afin de permettre la génération de six différentes **matrices** non spécifiques. La séquence servant de **matrice** de départ pour cette première étape est une matrice synthétique présentant déjà deux mutations en **5'P** et **3'OH** (Nomenclature : **RIL6mutEGFP2, Tableau 3.3, chapitre Matériels & Méthodes**). Des **matrices** uniquement mutées en **5'P** ou en **3'OH** par 2, 3 ou 4 mutations sont ainsi obtenues par PCR,

puis testées avec les amorces IL-6. Les résultats ont alors montré que trois mutations étaient nécessaires au niveau des extrémités 5'P ou 3'OH afin de permettre une observation convenable des amplifications non spécifiques.

Dans un deuxième temps, le degré de spécificité de chaque amorce est évalué par la détection de la matrice spécifique et des deux matrices non spécifiques, celle mutée en 3'OH (Nomenclature : **Matrice non spécifique 3F**) permettant d'évaluer le degré d'aspécificité de l'amorce sens et celle mutée en 5'P (Nomenclature : **Matrice non spécifique 3R**) permettant d'évaluer le degré d'aspécificité de l'amorce antisens. Les différentes détections sont effectuées en *duplicates*, à partir des mêmes solutions d'amorces et au cours d'un même *run* de PCR. D'autre part, afin d'évaluer la spécificité des amorces dans différents contextes de détection, trois mélanges de matrices présentant des probabilités d'amplifications non spécifiques décroissantes, sont utilisés :

1. [Matrice spécifique] = 1X + [Matrice non spécifique] = 100X
2. [Matrice spécifique] = 1X + [Matrice non spécifique] = 10X
3. [Matrice spécifique] = 1X + [Matrice non spécifique] = 1X

Le couple d'amorces ADN est évalué à sa température d'hybridation optimale (58°C) tandis que les amorces modifiées par des monomères LNA sont évaluées à trois températures d'hybridation différentes : 60°C, 62°C et 64°C. Les ratios  $\frac{\text{Détections Non Spécifiques}}{\text{Détections Spécifiques}}$  peuvent ensuite être calculés afin d'évaluer la spécificité de chaque amorce testée. Plus ce ratio est faible, plus l'amorce détecte spécifiquement la cible recherchée. Inversement, plus le ratio est fort, plus l'amorce s'hybride de façon non spécifique. La sensibilité de détection est quant à elle évaluée à partir des différents  $C_q$  obtenus dans une condition donnée et pour chacun des couples.

Les deux principales qualités de ce modèle sont :

1. La détection spécifique de la cible ainsi que la détection spécifique des hybridations non spécifiques des amorces au moyen de 2 différentes sondes d'hydrolyse.
2. Une évaluation objective des performances des différentes amorces sans l'influence d'inhibiteurs de RT-qPCR.

Il peut ainsi s'adapter à l'évaluation de n'importe quel couple d'amorces.

### 5.2.4. Spécificité des amorces modifiées

La comparaison des ratios obtenus pour les différents couples d'amorces est réalisée en prenant la spécificité du couple d'amorces ADN comme référence. L'ensemble des couples est ainsi évalué aux trois concentrations de matrice non spécifique testées.

#### 5.2.4.1. Estimation de l'amplification non spécifique

L'amplification non spécifique de notre référence (amorces ADN à  $T_H = 58^\circ\text{C}$ ) a tout d'abord été évaluée par comparaison du nombre de copies non spécifiques devant être théoriquement détecté avec le nombre de copies réellement mesurées au cours de la réaction. Ce résultat a été établi à la suite de la compensation de fluorescence, en tenant compte du rapport de couplage existant entre les deux sondes utilisées. Afin de déterminer ce dernier, deux PCR ont été réalisées au moyen d'amorces spécifiques aussi bien au niveau de la matrice IL6 (amorces RIL-6 ADN) qu'au niveau de la matrice non spécifique (amorces adaptées). Chaque réaction a été réalisée à une  $T_H$  de  $58^\circ\text{C}$  en système de sonde d'hydrolyse.

Ce rapport de couplage a ensuite été utilisé pour calculer de façon précise les ratios entre amplifications spécifiques et amplifications non spécifiques. Le pourcentage d'amplifications non spécifiques est pratiquement constant quelque soit la quantité de matrice spécifique. Les amorces 5' amplifient en moyenne une copie non spécifique sur 100 et les amorces 3' amplifient en moyenne une copie non spécifique sur 1000. (Figure 5.9). Cette différence est probablement séquence-dépendante, les mutations introduites ayant des conséquences thermodynamiques différentes. Cependant, il semble difficile d'envisager d'en tirer des conclusions sur seulement deux exemples. Lorsque l'on rapporte ces valeurs au nombre de copies spécifiques, on observe comme attendu que l'amplification non spécifique décroît avec la quantité de matrice non spécifique présente dans le milieu réactionnel (Figure 5.10). Ces valeurs nous servent ainsi de référence (100%) pour la mesure de la spécificité de nos différentes amorces modifiées par LNA.

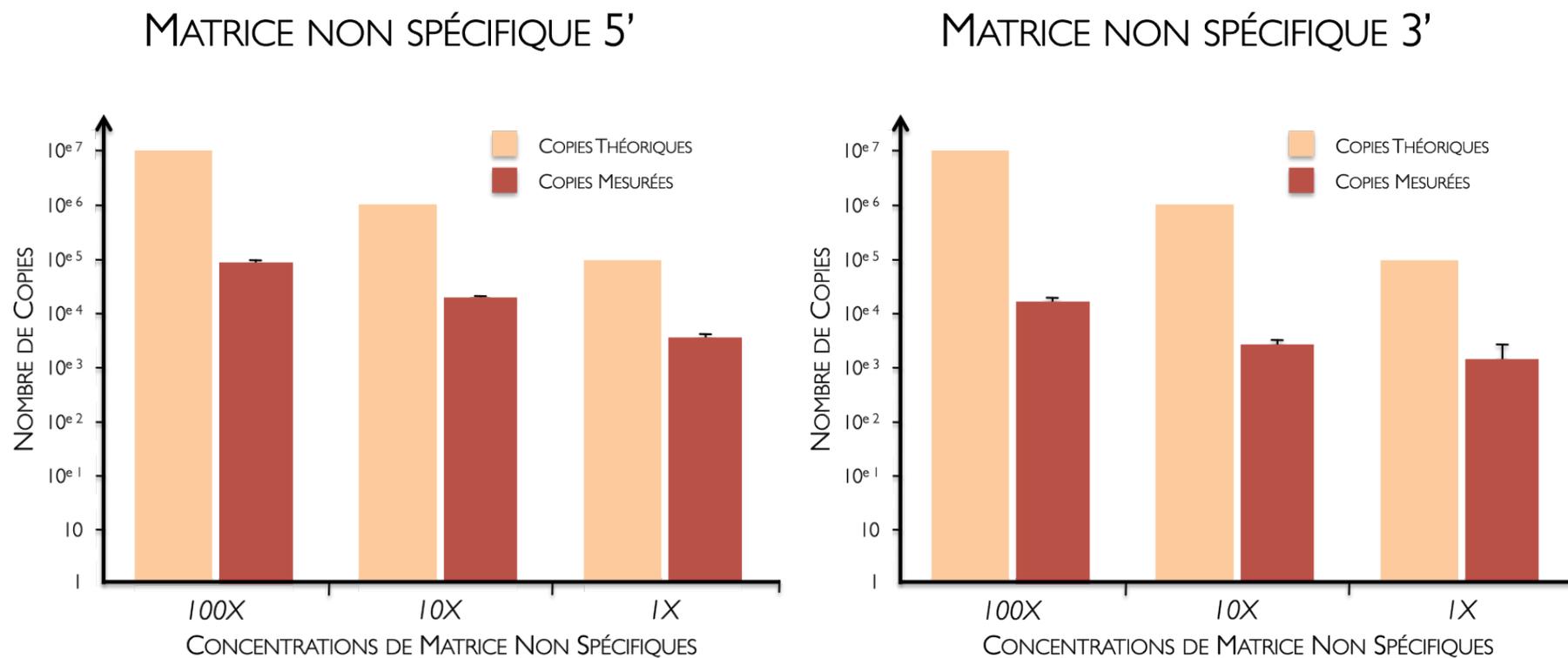


Figure 5.9 : Estimation de l'amplification non spécifique obtenue avec les amorces ADN aux trois concentrations testées (100X, 10X et 1X).

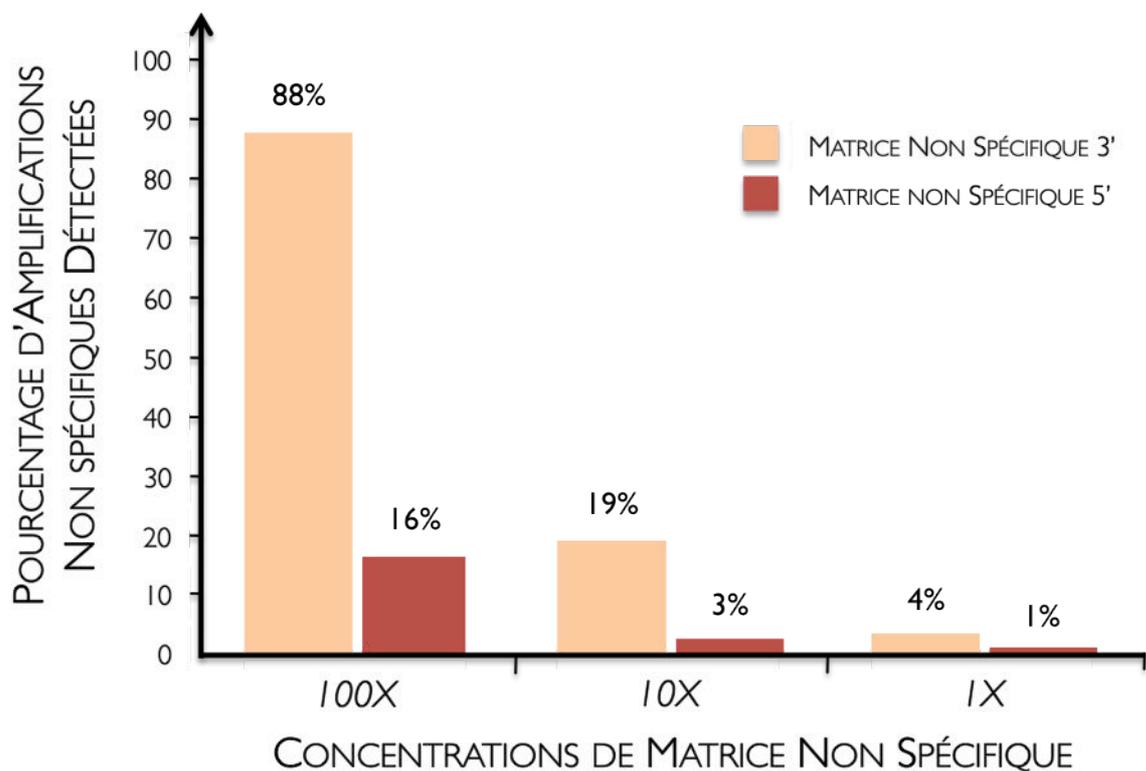


Figure 5.10 : Amplification non spécifique exprimée en pourcentage de l'amplification spécifique.

#### 5.2.4.2. Spécificité des amorces tous couples confondus

Les amorces sens modifiées par deux monomères LNA présentent toutes des degrés de spécificité supérieurs à celui observé pour l'amorce ADN et cela quelque soit le contexte de détection testé. Le degré de spécificité de l'amorce sens #1 est le moins important quelque soit le contexte de la détection testé (respectivement 450%, 211% et 187% pour les mélanges présentant 100X, 10X et 1X de matrice non spécifique).

Lorsque le contexte d'amplifications non spécifiques est favorisé (100X), la spécificité des amorces sens #3, #4, #5, #6 et #7 augmente avec la température d'hybridation à l'exception de l'amorce sens #2. Il convient de souligner ici que des résultats similaires sont obtenus pour les amorces de séquences identiques : RIL6 F3L / RIL6 F6L et RIL6 F5L / RIL6 F7L.

Lorsque le contexte d'amplifications non spécifiques est moins favorable (10X), l'amorce sens #2 devient moins spécifique lorsque la  $T_H$  augmente jusqu'à 64°C (respectivement 27%, 29% et 38% pour  $T_H = 60^\circ\text{C}$ , 62°C et 64°C). Aucune amplification non spécifique n'est détectée à la  $T_H$  de 64°C pour les amorces sens #3, #4, #5 et #7.

Dans le contexte de détection le plus favorable (1X), seule l'amorce sens #5 ne présente plus aucune amplification non spécifique aux trois  $T_H$  analysées. On constate enfin que l'amorce sens #2 continue à s'hybrider de façon non spécifique aux trois  $T_H$ .

Concernant les amorces antisens, toutes les amorces modifiées par LNA présentent des degrés de spécificité plus importants que celui observé pour l'amorce ADN dans un contexte d'amplifications non spécifiques favorisé (100X). On remarque que cette spécificité augmente avec la  $T_H$  pour l'ensemble des couples d'amorces modifiées par 2 LNA.

Ce comportement est encore observé dans un contexte d'amplifications non spécifiques moins favorable (10X). Dans cette condition, aucune amplification non spécifique n'est détectée à la  $T_H$  de 64°C pour les amorces antisens #2, #3, #4, #5 et #7.

Dans le contexte de détection le plus favorable (1X), on constate qu'aucune amplification non spécifique n'est détectée pour les amorces antisens #3, #4, #5 et #6 aux trois  $T_H$  analysées, contrairement aux amorces #1 et #2.

L'amorce antisens #1 est l'amorce la moins spécifique avec la présence d'amplifications non spécifiques dans les trois conditions à  $T_H = 60^\circ\text{C}$  (respectivement 63%, 37% et 105% pour 100X, 10X et 1X, **Figure 5.11**). Il est probable que le gain de spécificité obtenu par Alexandra MALGOYRE n'était dû qu'à l'amorce antisens.

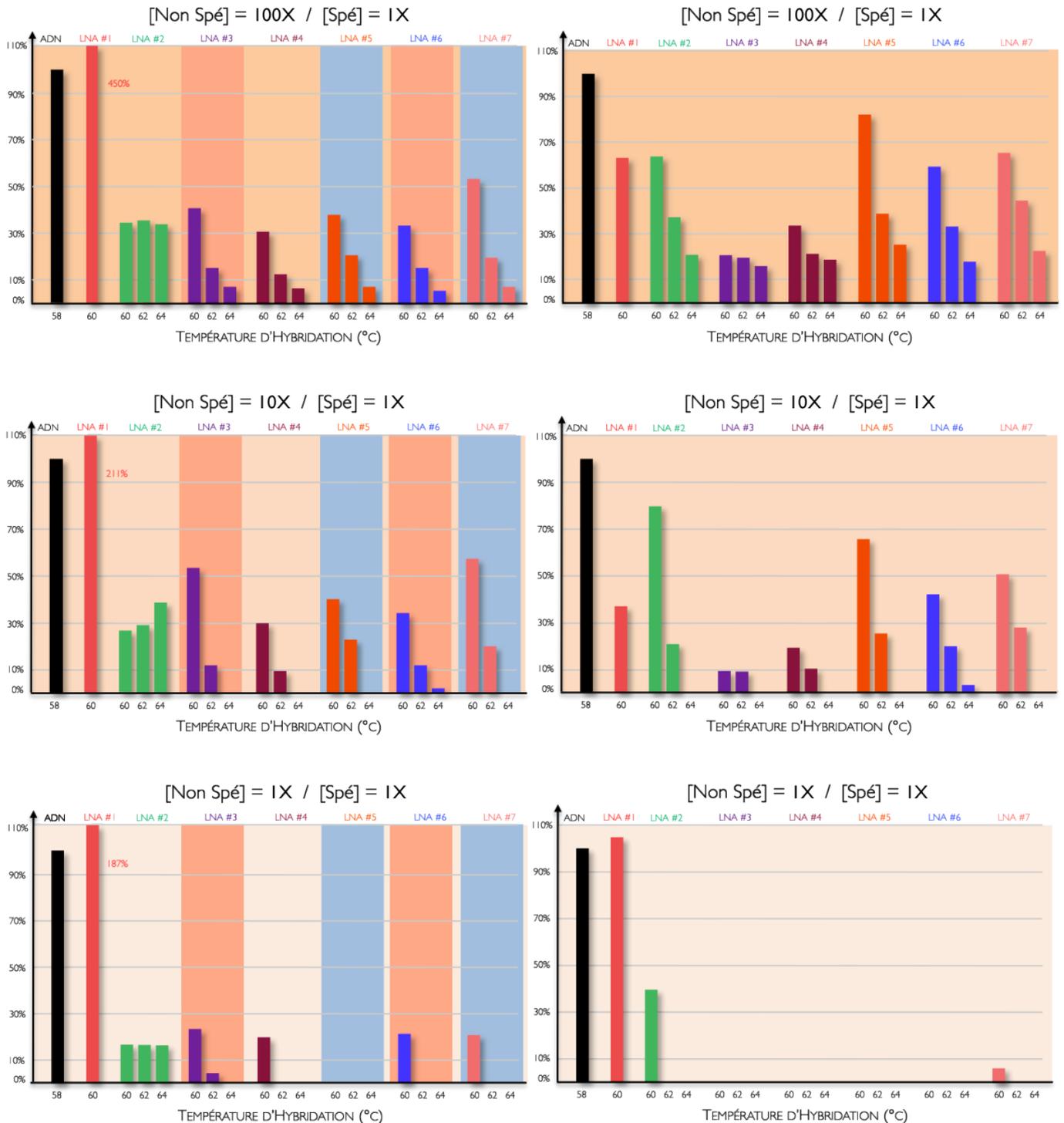
On constate que la spécificité de la majorité des amorces augmente avec la température d'hybridation. La substitution de deux bases par des monomères LNA permet d'obtenir une plus grande spécificité que si l'on substituait 4 bases. Ceci est particulièrement vrai pour l'amorce sens #1 qui est moins spécifique que l'amorce ADN. À la différence des autres amorces, l'amorce sens #2 doublement substituée présente une spécificité indépendante de la température d'hybridation.

## SPÉCIFICITÉ DES AMORCES MODIFIÉES VS. AMORCES ADN

Matrice non spécifique 5' :

5'P - GCGATGATGCACTG**GCTG**TAAAC - 3'OH

Matrice non spécifique 3' :

5'P - CTCT**CgAcA**TCTGGAGTTCGTT - 3' OH

**Figure 5.11** : Spécificité des amorces sens (sur matrice non spécifique 5') et antisens (sur matrice non spécifique 3'), tous couples confondus. L'évaluation du paramètre est réalisée pour une concentration fixe de matrice spécifique (1X) placée dans trois contextes différents (matrice non spécifique présente à 100X, 10X et 1X). Les séquences des amorces RIL6 F3L et RIL6 F6L sont identiques (fond orange) tout comme les séquences des amorces RIL6 F5L et RIL6 F7L (fond bleu).

### 5.2.4.3. Spécificité des amorces modifiées par 2 monomères LNA en condition d'hybridation optimale

Afin de mieux résumer les performances de ces différents couples, les précédents résultats obtenus peuvent être présentés sous une forme beaucoup plus claire, en ne considérant uniquement les températures d'hybridation optimales des amorces (Figure 5.12).

Dans le contexte de détection le moins favorable (100X), on observe une augmentation importante de la spécificité des amorces (respectivement  $64,6\% \pm 0,10$  ;  $93,2\% \pm 0,04$  ;  $93,6\% \pm 0,04$  ;  $92,9\% \pm 0,04$  ;  $94,7\% \pm 0,03$  et  $93,3\% \pm 0,04$  pour les amorces sens #2, #3, #4, #5, #6 et #7 et respectivement  $63,0\% \pm 0,19$  ;  $84,3\% \pm 0,09$  ;  $81,7\% \pm 0,11$  ;  $74,7\% \pm 0,15$  ;  $82,3\% \pm 0,09$  et  $77,8\% \pm 0,13$  pour les amorces antisens #2, #3, #4, #5, #6 et #7).

Ce résultat est comparable lorsque le contexte de détection devient moins favorable (10X) : respectivement  $70,9\% \pm 0,10$  ;  $99,4\% \pm 0,04$  ;  $100,0\% \pm 0,04$  ;  $100,0\% \pm 0,04$  ;  $97,8\% \pm 0,03$  et  $100,0\% \pm 0,04$  pour les amorces sens #2, #3, #4, #5, #6 et #7 et respectivement  $79,1\% \pm 0,19$  ;  $100,0\% \pm 0,09$  ;  $100,0\% \pm 0,11$  ;  $100,0\% \pm 0,15$  ;  $96,3\% \pm 0,09$  et  $100,0\% \pm 0,13$  pour les amorces antisens #2, #3, #4, #5, #6 et #7). Les amorces sens #4, #5 et #7 et les amorces antisens #3, #4, #5 et #7 sont donc les premières amorces totalement spécifiques à la  $T_H$  de  $64^\circ\text{C}$ .

Dans le contexte de détection le plus favorable (1X), les amorces modifiées deviennent toutes totalement spécifiques à l'exception de l'amorce sens #1 (seulement  $83,6\%$  de spécificité).

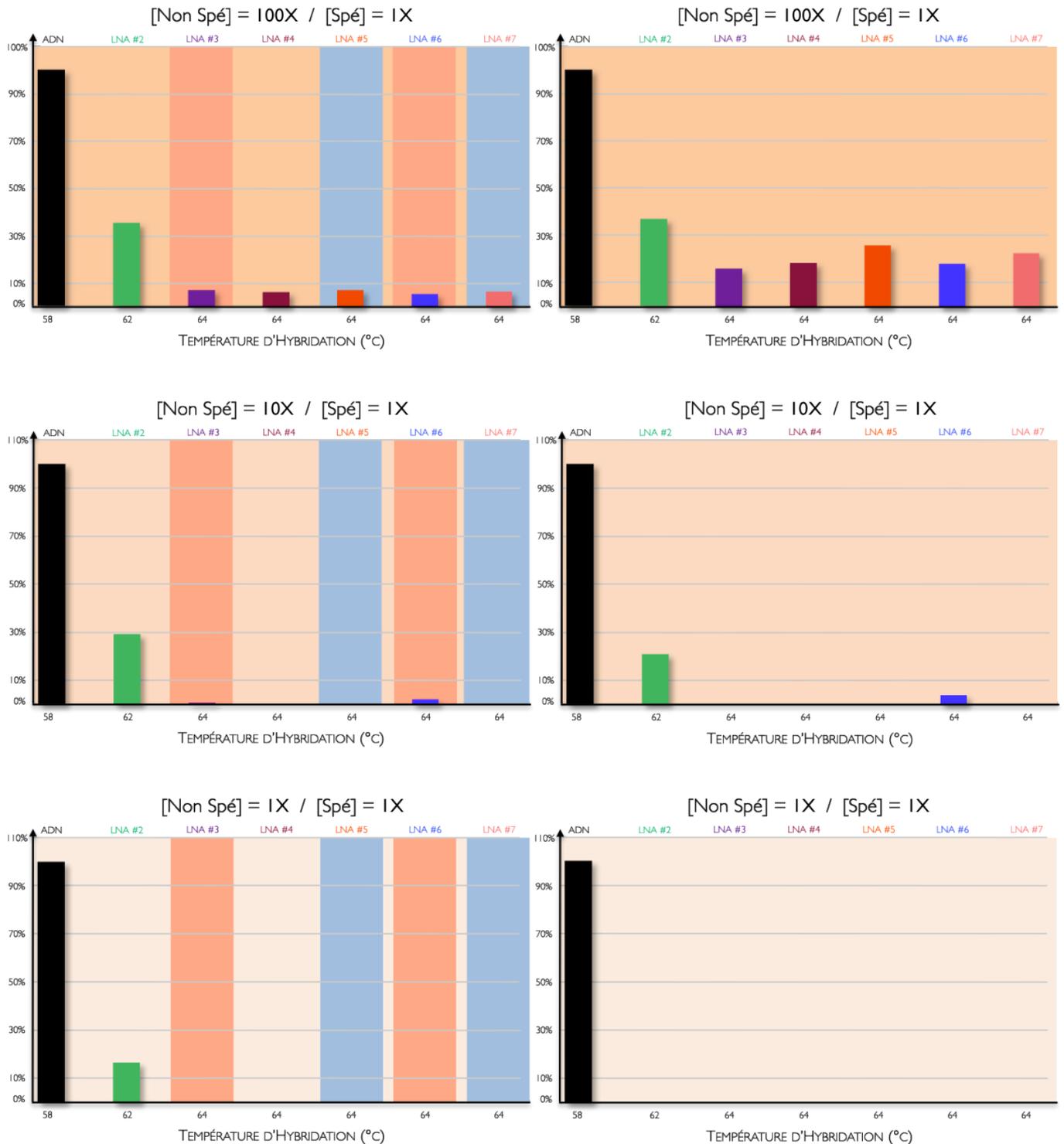
En conclusion, on observe qu'il n'y a pas de lien entre la spécificité et la température de fusion des amorces utilisées.

## SPÉCIFICITÉ CONDITION D'HYBRIDATION OPTIMALE

Matrice non spécifique 5' :

5'P - GCGATGATGCACTG**GCT**GTAAC [ ] - 3'OH

Matrice non spécifique 3' :

5'P - [ ] CTCT**CGACA**TCTGGAGTCCGTT - 3' OH

**Figure 5.12** : Spécificité des amorces sens (sur matrice non spécifique 3F) et antisens (sur matrice non spécifique 3R) en condition d'hybridation optimale. L'évaluation du paramètre est réalisée pour une concentration fixe de matrice spécifique (1X) dans trois contextes de matrice non spécifique différents (100X, 10X et 1X). Les séquences des amorces RIL6 F3L et RIL6 F6L sont identiques (fond orange) tout comme les séquences des amorces RIL6 F5L et RIL6 F7L (fond bleu).

### 5.2.5. Sensibilité des amorces modifiées

La sensibilité des différentes **amorces** a été obtenue par l'observation des valeurs de  $C_q$  obtenues pour la détection de la **matrice** spécifique de l'IL-6 à la concentration de 1X dans différents environnements non spécifiques et différentes températures (concentrations de **matrice** non spécifique à 100X, 10X, 1X, X/10 et X/100). Les deux illustrations de la **Figure 5.13** représentent la moyenne de ces cycles de détection pour toutes les **amorces sens** et **antisens**, obtenue pour la détection spécifique de la **matrice** IL-6 en fonction de la **température d'hybridation** utilisée au cours de la réaction. La sensibilité est observée à la **température d'hybridation** optimale de nos différentes **amorces**.

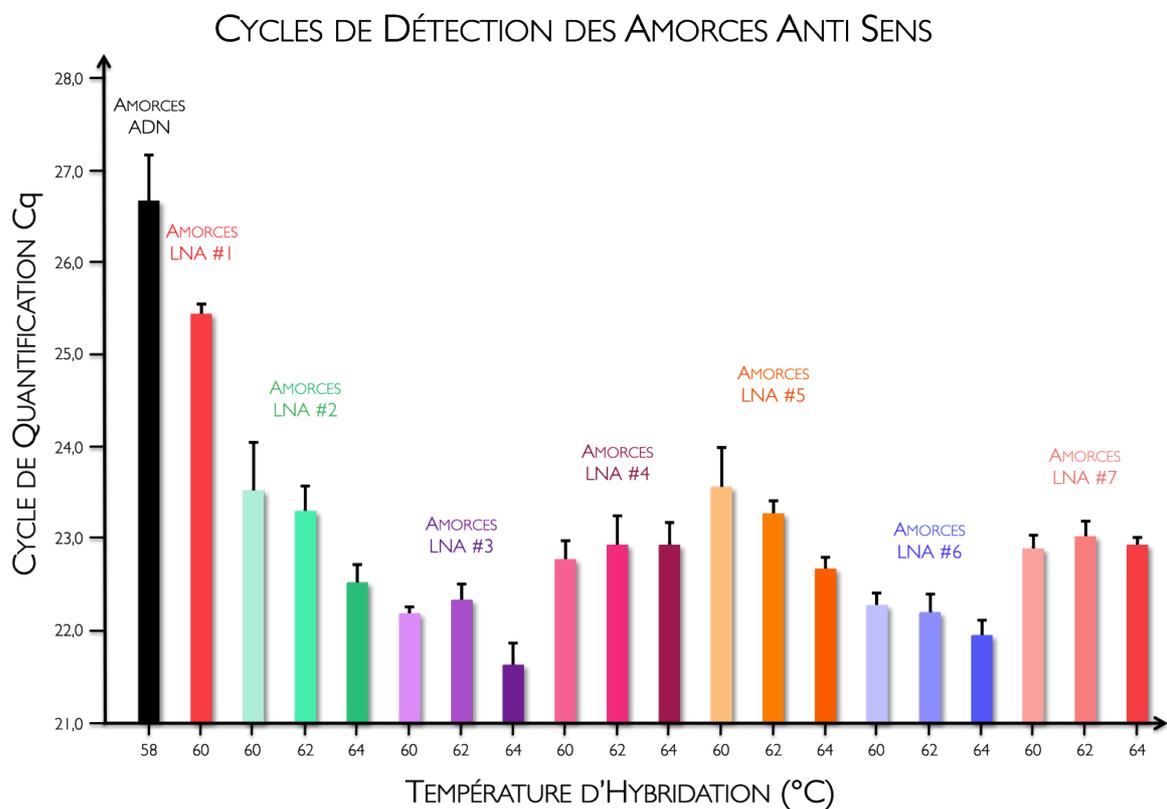
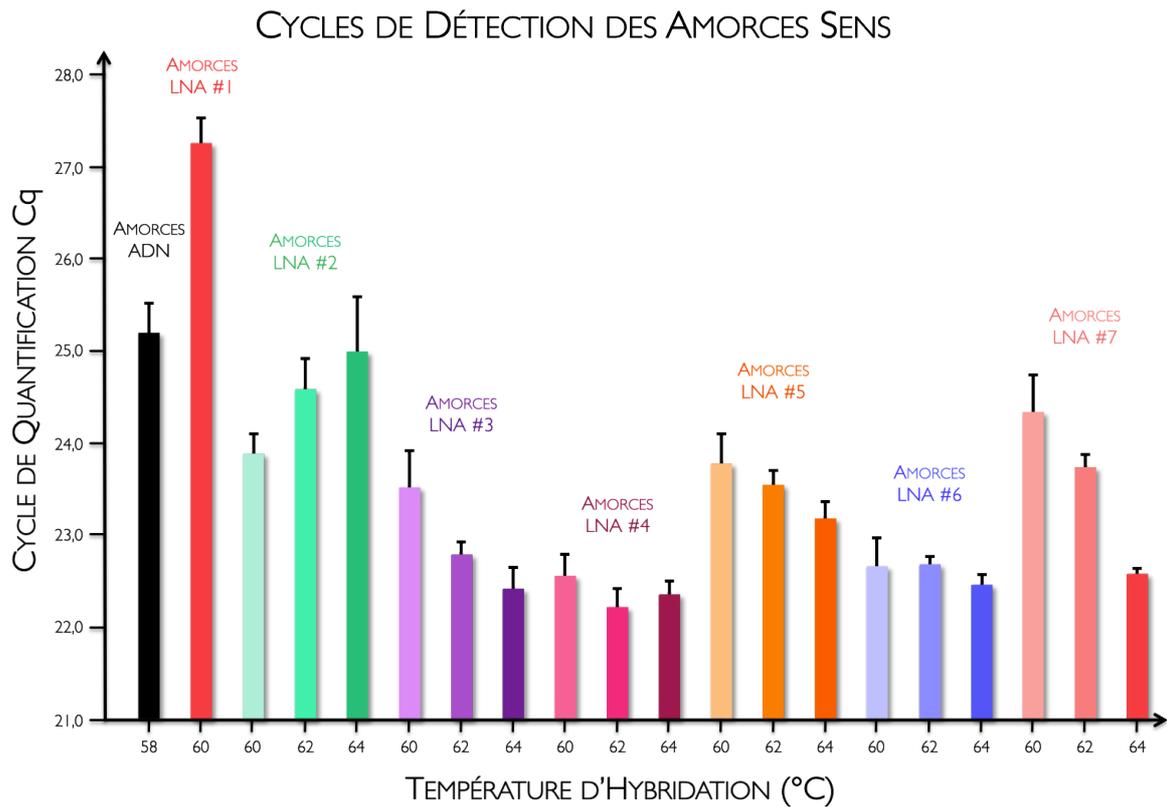
Pour la majorité des **amorces**, la sensibilité augmente avec la **température d'hybridation** à l'exception des **amorces sens #4** et **#7** qui restent constantes. Une interprétation possible de ce résultat est que l'augmentation de la spécificité induit une augmentation du nombre d'**amorces** disponibles pour se lier sur la cible. Cette hypothèse est confortée par l'observation de l'**amorce sens #2**. En effet, la spécificité de cette **amorce** est indépendante de la température et on observe un décrochage de l'amplification spécifique avec la **température d'hybridation**.

D'autre part, à la **température d'hybridation** optimale, les **amorces** présentant 2 substitutions de LNA sont significativement plus sensibles que les **amorces** ADN.

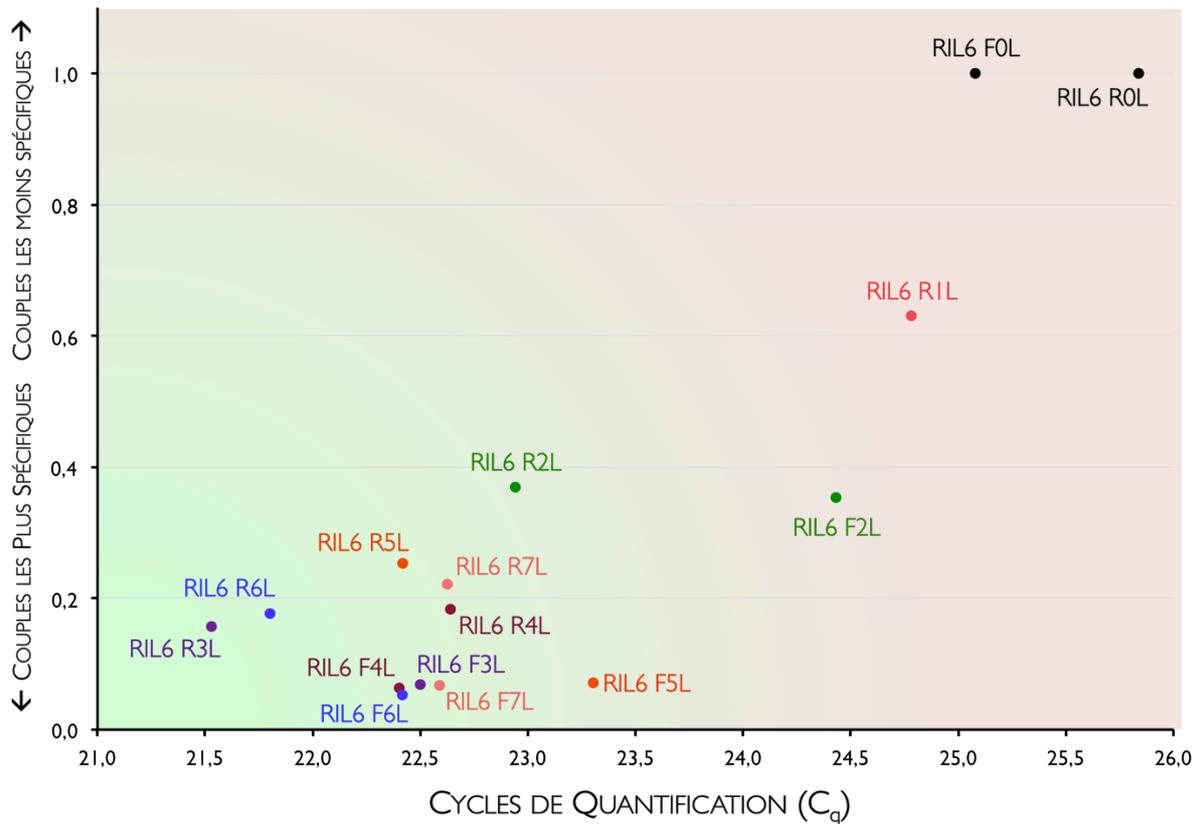
Le couple d'**amorces #1**, caractérisé par 4 substitutions de LNA, présente un comportement différent. L'**amorce sens** est moins sensible que l'**amorce sens** ADN ce qui n'est pas le cas pour les **amorces antisens** où la séquence ADN est moins sensible que l'**amorce** modifiée.

Quelle que soit l'**amorce** considérée des couples **#3**, **#5**, **#6** et **#7**, on observe qu'une augmentation de la  $T_H$  jusqu'à 64°C favorise une augmentation de sensibilité de détection.

Les résultats montrent enfin que les couples d'**amorces** modifiées **#3** et **#6** constituent les couples les plus sensibles pour une détection spécifique de la cible (**Figure 5.14**).



**Figure 5.13 :** Cycles moyens de détection observés pour les différentes amorces sens et anti-sens modifiées par LNA. Les  $C_q$  sont obtenus à partir de la détection de cinq mélanges de matrices spécifique (1X) et non spécifique (100X, 10X, 1X, X/10 et X/100), analysés en duplicates.



**Figure 5.14 :** Représentation graphique illustrant la sensibilité des différentes amorces en fonction de leurs spécificités respectives. Les amorces se trouvant dans la proximité de la zone verte sont plus spécifiques et plus sensibles que celles se trouvant dans la zone rouge (moins spécifiques et moins sensibles). Les séquences des amorces RIL6 F3L et RIL6 F6L sont identiques tout comme celles des amorces RIL6 F5L et RIL6 F7L. L'amorce RIL6 F1L n'apparaît pas sur ce graphique car son aspécificité est trop importante (d'une valeur de 4,5).

On constate qu'une augmentation de la  $T_H$  induit une augmentation de la sensibilité de détection des amorces modifiées pour la matrice spécifique. Les couples d'amorces #3 et #6 permettent les détections la plus sensibles de la séquence ciblée. Le couple d'amorces #2 fait exception au reste du groupe d'amorces dessinées car il présente une amorce sens indépendante de la  $T_H$ . Enfin, le couple d'amorces #1 permet la détection la moins sensible de matrice spécifique.

### 5.2.6. Mesure des paramètres thermodynamiques des amorces modifiées

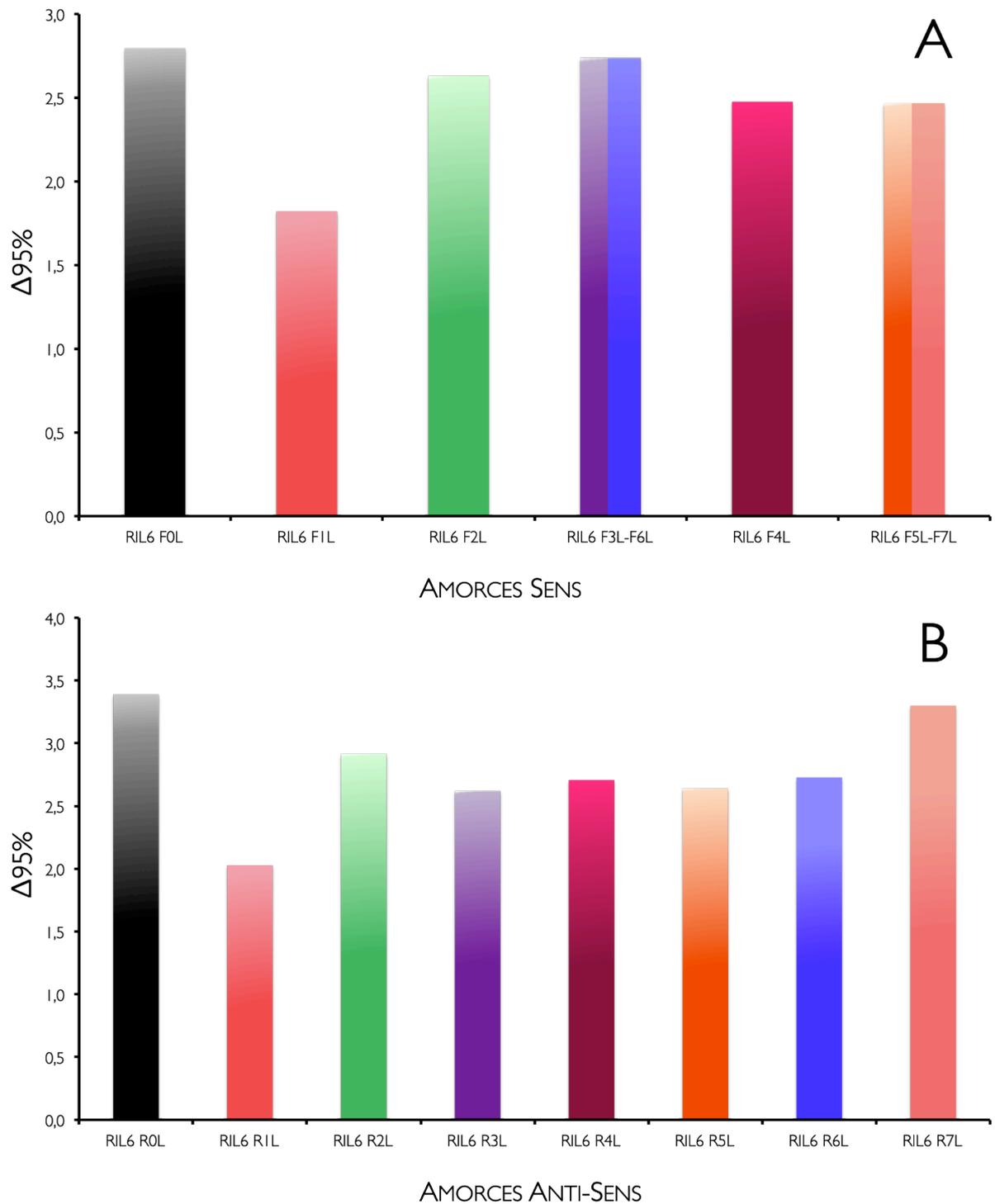
Les paramètres thermodynamiques ( $\Delta H$ ,  $\Delta S$  et  $\Delta G$ ) de nos différentes amorces sont mesurés à partir de la méthode de mesure directe du  $T_m$ , développée dans le **Chapitre 6**. Les résultats obtenus pour les amorces sens et antisens sont résumés au niveau des **Tableaux 5.6.A** et **5.6.B**, respectivement.

Une forte diminution de la valeur de  $\Delta H$  est observée pour l'amorce sens modifiée par 4 LNA par rapport à l'amorce ADN (-393 contre -245 kcal.mol<sup>-1</sup> pour les amorces FIL et ADN, respectivement). En revanche, nous n'observons que de faibles différences entre la valeur de  $\Delta H$  des amorces modifiées par 2 LNA par rapport à l'amorce ADN (-270 kcal.mol<sup>-1</sup> en moyenne contre -245 kcal.mol<sup>-1</sup> pour les amorces modifiées par 2 LNA et les amorces ADN, respectivement). Une observation identique est constatée pour la valeur de  $\Delta S$ .

Ces résultats confirment la différence majeure de comportement des amorces modifiées par 4 LNA et les amorces modifiées par 2 LNA ; il existe probablement un effet coopératif pour les substitutions multiples. Les faibles différences observées entre les amorces modifiées par 2 LNA et les amorces ADN aux limites de la sensibilité de la méthode de mesure ne permettent pas de tirer de conclusions sur les différents types de substitutions (**Figure 5.15**).

A		RIL6 F0L	RIL6 FIL	RIL6 F2L	RIL6 F3L	RIL6 F4L	RIL6 F5L		
$T_H$		58	60	62	64	64	64		
$\Delta H$		-245	-393	-264	-255	-280	-283		
$\Delta S$		-683	-1091	-733	-706	-782	-787		
$\Delta 95\%$		2,8	1,8	2,6	2,7	2,5	2,5		
B		RIL6 R0L	RIL6 R1L	RIL6 R2L	RIL6 R3L	RIL6 R4L	RIL6 R5L	RIL6 R6L	RIL6 R7L
$T_H$		58	60	62	64	64	64	64	64
$\Delta H$		-203	-353	-239	-267	-258	-266	-258	-233
$\Delta S$		-560	-977	-661	-740	-715	-736	-711	-639
$\Delta 95\%$		3,4	2,0	2,9	2,6	2,7	2,6	2,7	3,3

**Tableaux 5.6 :** Paramètres thermodynamiques des amorces sens (A) et anti-sens (B) obtenus à partir de la méthode de mesure directe du  $T_m$  par courbe de fusion. La variation d'enthalpie  $\Delta H$  est donnée en kcal.mol<sup>-1</sup> et la variation d'entropie  $\Delta S$  en cal.mol<sup>-1</sup>.K<sup>-1</sup>. La variation d'énergie libre  $\Delta G$  (en cal.mol<sup>-1</sup>) est calculée à la température d'hybridation optimale des amorces (°C). Le  $\Delta 95\%$  représente la différence de température (°C) entre la température où l'on observe 95% d'amorce hybridée et la température de fusion de la séquence.



**Figure 5.15 :** Différence de température ( $^{\circ}\text{C}$ ) entre la température où l'on observe 95% d'amorce hybridée et la température de fusion de la séquence, observées pour les amorces *sens* (A) et *antisens* (B) à la température d'hybridation optimale des amorces (F0LR0L :  $58^{\circ}\text{C}$ , F1LR1L :  $60^{\circ}\text{C}$ , F2LR2L :  $62^{\circ}\text{C}$ , F3LR3L à F7LR7L :  $64^{\circ}\text{C}$ ).

## 5.2.7. Définitions de règles de design

### 5.2.7.1. Nature de la base substituée

Sur la base de la spécificité précédemment établie, nous avons étudié la nature de la base substituée en classant les différentes amorces modifiées par ordre croissant de spécificité. Les résultats obtenus sont montrés en **Figure 5.16**.

Concernant les amorces *sens*, on constate une plus grande spécificité de l'amorce lorsque celle-ci présente une purine et une pyrimidine LNA au sein de sa séquence.

Cependant, cette observation n'est pas retrouvée au niveau des amorces *antisens* où la présence de purines ou de pyrimidines au sein de la séquence ne semble pas jouer un rôle déterminant dans la spécificité de l'amorce.

En conclusion, il est très difficile de prédire quel type de base purique ou pyrimidique il serait nécessaire de substituer pour apporter le plus de spécificité à l'amorce.

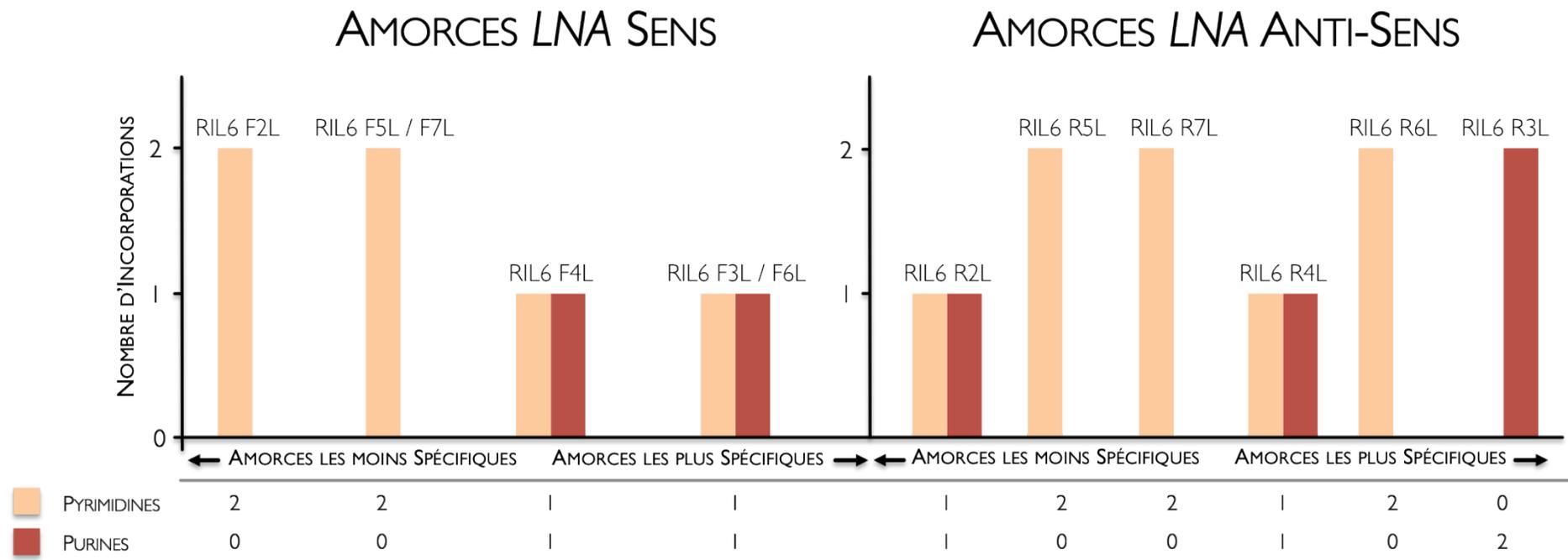
### 4.2.7.2. Positionnement de la base substituée

Nous avons ensuite étudié les 3 types de positionnements de bases substituées (espacées ou rapprochées en *extrémité 5'P*, rapprochées en milieu de séquence) en fonction de la spécificité de l'amorce modifiée. Les résultats sont illustrés en **Figure 5.17**.

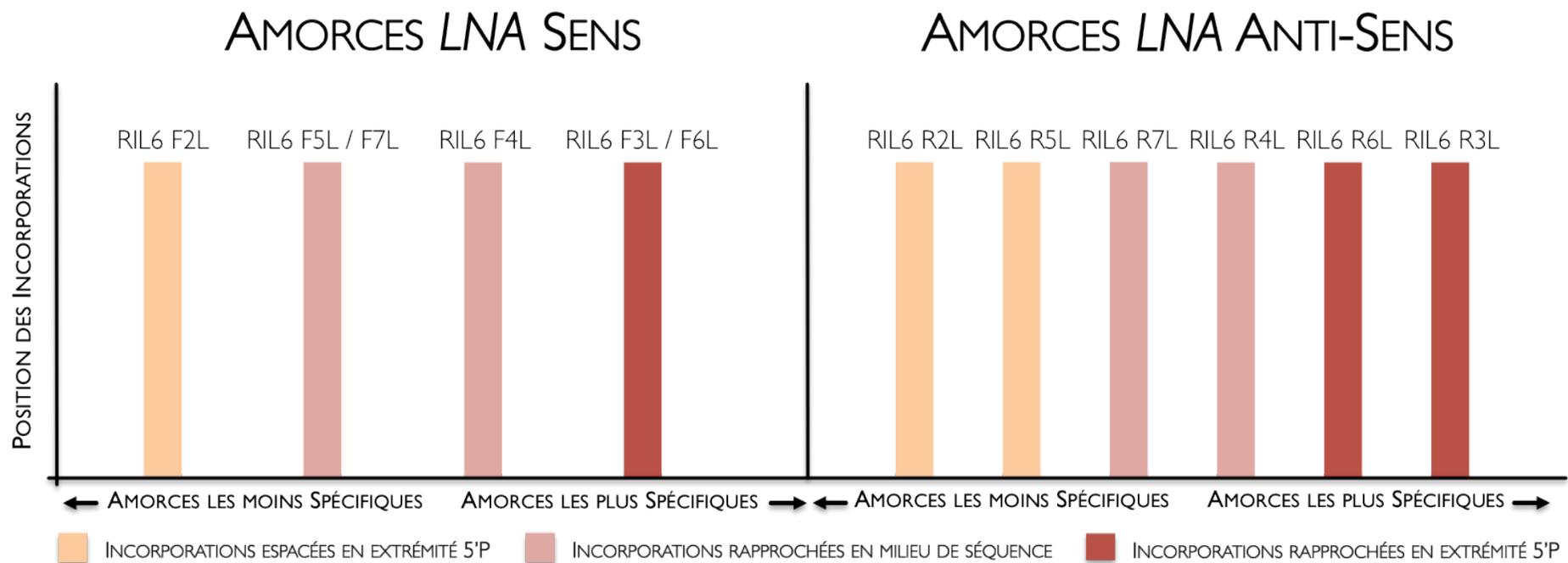
Concernant les amorces *sens*, on observe une plus grande spécificité de l'amorce lorsque les substitutions sont réalisées de façon rapprochée en *extrémité 5'P* de la séquence de l'amorce. Les incorporations réalisées de façon rapprochée en milieu de séquence semblent quant à elles apporter un peu plus de spécificité que les incorporations positionnées de façon écartée.

Une observation similaire est constatée au niveau des amorces *antisens*.

En conclusion, des incorporations de LNA réalisées de façon rapprochées en *extrémité 5'P* de la séquence améliorent beaucoup plus la spécificité de l'amorce que lorsqu'elles sont réalisées en milieu de séquence ou positionnées de façon écartée.



**Figure 5.16 :** Spécificité des amorces *sens* (sur matrice non spécifique 3F) et *antisens* (sur matrice non spécifique 3R) en fonction de la nature de la base substituée, en condition d'hybridation optimale. L'évaluation du paramètre est réalisée pour une concentration fixe de matrice spécifique (1X) dans trois contextes de matrice non spécifique différents (100X, 10X et 1X). La concentration des amorces utilisées est de 0,4  $\mu$ M.



**Figure 5.17 :** Spécificité des amorces *sens* (sur matrice non spécifique 3F) et *antisens* (sur matrice non spécifique 3R) en fonction du positionnement de la base substituée en condition d'hybridation optimale. L'évaluation du paramètre est réalisée pour une concentration fixe de matrice spécifique (1X) dans trois contextes de matrice non spécifique différents (100X, 10X et 1X). La concentration des amorces utilisées est de 0,4  $\mu$ M.

### 5.3. Discussion

Depuis leur découverte en 1998 (Koshkin, Singh et al. 1998; Obika, Nanbu et al. 1998), les *Locked Nucleic Acids* (LNA) ont été utilisés dans de nombreuses applications de biologie moléculaire afin d'améliorer à la fois la force et la spécificité d'hybridation d'un oligonucléotide pour sa séquence complémentaire. De nombreuses tentatives d'optimisation au moyen de ces analogues ont donc été réalisées dans le cadre de la PCR, notamment pour la PCR de détection spécifique d'allèle, pour le design de sondes (d'hydrolyse, *Molecular Beacons* ...) ou bien encore d'amorces de PCR (Petersen and Wengel 2003; Jepsen, Sorensen et al. 2004; Vester and Wengel 2004) pour revues). Il existe cependant très peu de données dans la littérature scientifique concernant la conception d'amorces modifiées par LNA. L'étude de David LATORRA (Latorra, Arar et al. 2003) reste à ce jour la seule étude complète réalisée sur le sujet. Elle conclue que ce type de modification conduit à l'augmentation des températures d'hybridation maximales (jusqu'à 8°C par nucléotide substitué en théorie) et que ces monomères doivent être préférentiellement placés au niveau de l'extrémité 3'OH de la séquence pour une détection spécifique d'allèle en PCR (Latorra, Campbell et al. 2003). D'autre part, l'auteur pressent déjà dans cette étude que l'établissement de règles efficaces sur la conception d'amorces modifiées par l'incorporation de LNA sera probablement très complexe à mettre en œuvre.

Au sein du laboratoire, nous avons proposé la substitution de bases situées en extrémité 5'P de manière à diminuer les amplifications non spécifiques et par conséquent augmenter la spécificité des amorces de PCR (Malgoyre, Banzet et al. 2007). Le but de notre étude était donc de définir des règles de design concernant le nombre et le positionnement des substitutions afin d'améliorer les performances de ces amorces. En particulier, il apparaissait que 4 incorporations de LNA permettaient un gain de température d'hybridation très inférieur à celui attendu au regard des modèles thermodynamiques. Nous avons donc émis l'hypothèse qu'un nombre trop important de monomères LNA était responsable de cette différence. Nous avons alors décidé de tester ces amorces avec seulement 2 substitutions en les positionnant plutôt au niveau de l'extrémité 5'P de la séquence (rapprochées en extrémité 5'P de la séquence, rapprochées en milieu de séquence ou espacées en extrémité 5'P).

Il n'existe pas de modèle de référence pour tester de façon objective la spécificité des amorces utilisées en PCR. Malgré de nombreuses tentatives, l'utilisation d'échantillons

biologiques ne nous a pas permis d'obtenir de résultats suffisamment reproductibles. Nous avons du tester de nombreux modèles construits sur l'utilisation de séquences synthétiques. Nous sommes finalement parvenus à conceptualiser un modèle construit sur l'utilisation de cibles synthétiques non spécifiques mutées au niveau de l'**extrémité 3'OH** du site de fixation des **amorces**. En faisant cela, nous avons pris soin de laisser 2 ou 3 bases spécifiques en **extrémité 3'OH**, de manière à favoriser l'amplification non spécifique. Il est impossible de démontrer que ces configurations correspondent bien aux cibles non spécifiques existant au niveau des échantillons biologiques. Cependant, l'importante homologie entre les séquences spécifique et non spécifiques (93,6%) nous paraît judicieuse pour tester le gain en spécificité de ces substitutions. Une évaluation indépendante de chaque **amorce** est donc réalisée, les mutations ponctuelles étant uniquement placées sur l'une ou l'autre des **extrémités** de la **matrice** non spécifique (**Alignements 3.1**).

D'autre part, la quantification de l'amplification non spécifique a été faite à l'aide d'une **sonde** d'hydrolyse spécifique (Holland, Abramson et al. 1991). Dans l'étude initiale, nous avons utilisé une caractérisation du produit amplifié par **courbe de fusion** afin de détecter les amplifications non spécifiques. Or, au cours d'études préliminaires à la conception de notre modèle d'évaluation de spécificité, nous avons montré que cette méthode ne permettait pas la quantification de l'amplification non spécifique. En effet, le ratio  $\frac{\text{Molécules Spécifiques}}{\text{Molécules Non Spécifiques}}$  existant au moment de la quantification (fin de la phase exponentielle) peut être très sensiblement différent du ratio  $\frac{\text{Molécules Spécifiques}}{\text{Molécules Non Spécifiques}}$  existant au moment de la phase plateau.

L'hypothèse formulée au départ de l'étude est vérifiée. Les **températures d'hybridation** des **amorces** doublement substituées sont significativement supérieures à la **température d'hybridation** de l'**amorce** présentant 4 incorporations de monomères LNA. Il faut souligner que le gain obtenu pour la **température d'hybridation** n'est pas lié au gain observé au niveau de la **température de fusion** des **amorces** (seulement 2°C pour les **amorces** LNA #1 contre 6°C pour les **amorces** doublement substituées). Ce résultat suggère que 4 substitutions de bases par des LNA diminuent le bénéfice thermodynamique (diminution de la **température d'hybridation**) par rapport à des doubles ou des simples substitutions. Malgré un impact plus faible sur le  $T_m$  et un  $\Delta G$  moins important, l'utilisation de seulement 2 monomères LNA permet une augmentation plus importante de la **température d'hybridation**. Ce résultat est compatible avec la littérature (Latorra, Arar et al. 2003 ; Christensen 2007). Il reste néanmoins important de souligner que 4 substitutions augmentent de manière significative la spécificité des **amorces** utilisées sur échantillons biologiques (d'après MALGOYRE *et al.*, 2007). D'autre part, il est démontré que la

substitution par des monomères LNA rigidifie le squelette phosphate de la molécule (Grunweller and Hartmann 2007) et diminue la constante de dissociation ADN / base LNA (Christensen, Jacobsen et al. 2001). Ces modifications pourraient expliquer l'augmentation de la température d'hybridation maximale. Nos résultats suggèrent que cette augmentation présente cependant une limite : ces amorces sont relativement peu spécifiques, essentiellement l'amorce sens #1. Nous n'avons pas réussi à expliquer cette discordance de résultats. En effet, l'augmentation du  $T_m$  permet l'hybridation d'une quantité plus importante d'amorces sur les matrices spécifiques et non spécifiques à une température donnée mais paradoxalement, on observe une faible efficacité de l'amplification ( $C_q$  élevés). Ce déficit de sensibilité pourrait être dû à des cinétiques d'hybridation différentes. La vérification de cette hypothèse nécessite donc des expérimentations complémentaires.

Bien que les températures de fusion obtenues avec les logiciels d'Exiqon et OligoAnalyser 3.1.d'IDT soient relativement proches (d'après la Figure 5.5), on constate que ces valeurs prédites sont biaisées lorsqu'on les compare à une mesure directe du  $T_m$  à partir du mix de PCR (d'après la Figure 5.7). Ces résultats peuvent être expliqués par la différence existant entre un milieu théorique paramétré par le logiciel et le milieu réel où se déroule la PCR. D'autre part, on observe que la méthode prédictive de McTIGUE, pourtant développée pour des amorces modifiées par 1 monomère LNA, n'est pas non plus adaptée à une prédiction réaliste du  $T_m$  de nos amorces modifiées (d'après la Figure 5.8). Ce mauvais agrément peut non seulement s'expliquer par le fait que cette méthode ne considère l'incorporation que d'un seul LNA au niveau de la séquence de l'amorce mais aussi et surtout par le fait que ses paramètres thermodynamiques soient obtenus à la concentration de 1M NaCl. Ainsi, même si de récents travaux réalisés par Richard OWCZARZY (Owczarzy, You et al. 2011) ont permis de compléter les tables des paramètres thermodynamiques pour deux monomères LNA consécutifs au sein d'une même séquence oligonucléotidique, ces méthodes restent inappropriées pour une prédiction précise du  $T_m$  de telles amorces utilisées en PCR.

Le degré de spécificité de chaque amorce est évalué par détecteurs spécifiques au moyen de sondes d'hydrolyse. En considérant l'ensemble des couples, les résultats obtenus montrent que l'amorce sens du couple #1 présente un fort degré d'aspécificité. De même, et bien que d'une moindre importance, un résultat similaire peut être observé pour l'amorce antisens (d'après la Figure 5.11). Or, il est démontré que les LNA sont capables d'induire une certaine rigidité conformationnelle responsable de la stabilité du duplexe (Grunweller and

Hartmann 2007). De plus, ils présentent des constantes de dissociation plus importantes que les bases ADN (Christensen, Jacobsen et al. 2001 ; Arora, Kaur et al. 2008). On peut donc penser qu'un nombre trop important de LNA pourrait devenir préjudiciable à la spécificité de la détection, ne permettant pas un niveau de déshybridation suffisant lorsque l'amorce se fixe de manière non spécifique.

Les amorces les plus spécifiques sont celles ne présentant que 2 monomères LNA au niveau de leur séquence (couples #3, #4, #5 et #7, d'après la Figure 5.12). Ce résultat souligne que le nombre de bases substituées constitue un paramètre important dans le design des amorces modifiées. De toute évidence, la température d'hybridation élevée des amorces modifiées par seulement 2 monomères LNA est responsable de cette spécificité ( $T_H = 64^\circ\text{C}$ ). En raison de l'augmentation de stabilité favorisée par l'incorporation de LNA, une utilisation d'amorces modifiées de plus courtes longueurs tout aussi spécifiques que des amorces ADN classiques pourrait être envisageable en PCR (Costa, Ernault et al. 2004 ; You, Moreira et al. 2006 ; Owczarzy, You et al. 2011). D'une part, la sensibilité des amorces à 2 substitutions de LNA est globalement supérieure à la sensibilité des amorces ADN et des amorces incorporées par 4 LNA, d'autre part, cette sensibilité augmente avec la température d'hybridation. Ce résultat suggère que l'augmentation de la spécificité augmente la disponibilité des amorces pour la cible et donc le rendement de la réaction d'amplification. Si cette hypothèse est vraie, cela signifierait qu'une PCR ayant 100% d'efficacité n'est pas une réaction dans laquelle la totalité des cibles sont hybridées avec des amorces mais une réaction dans laquelle la même proportion de matrices est hybridée à chaque cycle. Des résultats non publiés du laboratoire, obtenus en dehors de ce travail, semblent corroborer cette hypothèse. En particulier, le pourcentage de l'amorce hybridée sur une quantité constante de séquence cible semble dépendre de l'environnement ADN dans lequel se présente cette séquence cible, de façon analogue à ce qui a pu être observé dans le cadre de la RT (Levesque-Sergerie, Duquette et al. 2007).

La sensibilité des différentes amorces est évaluée par la moyenne des cycles de détection de la matrice spécifique mélangée à cinq concentrations différentes de matrices non spécifiques. Les résultats montrent que les amorces présentant 2 incorporations de LNA détectent plus précocement la cible que les amorces présentant 4 incorporations (d'après la Figure 5.13).

Il est important de souligner que le nombre d'amorces testées reste insuffisant pour dégager des règles de design qui soient vraiment indiscutables. Il apparaît cependant que 2 substitutions

rapprochées en **extrémité 5'P** permettent une meilleure spécificité que des incorporations rapprochées en milieu de séquence et que 2 incorporations espacées en **extrémité 5'P** (d'après la **Figure 5.17**). Ce résultat peut s'expliquer par le fait que ces différentes configurations pourraient correspondre à un ancrage de moins en moins important de l'**amorce** au niveau de la **partie 5'P** de sa séquence. En effet, des incorporations rapprochées en **extrémité 5'P** pourraient laisser beaucoup plus de possibilités pour une déshybridation éventuelle du reste de la séquence. Même si ce premier résultat est en accord avec les données de la littérature (Levin, Fiala et al. 2006 ; Malgoyre, Banzet et al. 2007), il nécessite d'être vérifié sur un plus grand nombre de séquences.

Concernant la nature de la base substituée, nous n'avons observé aucun effet sur la spécificité de détection de la cible (d'après la **Figure 5.16**). On sait que les pyrimidines LNA jouent un rôle important dans la stabilité d'un **duplexe** (McTigue, Peterson et al. 2004; Hughesman, Turner et al. 2008 ; Kaur, Arora et al. 2008 ; Kaur, Wengel et al. 2008 ). On pouvait donc imaginer que la substitution des bases puriques nous permettrait une plus grande spécificité. Dans notre étude, si cet effet existe, il est très faible et probablement masqué par la différence de position entre les différentes incorporations.

Le **Tableau 5.7** résume la spécificité des différentes **amorces** en fonction de la nature et de la position des substitutions réalisées par LNA.

**Tableau 5.7 :** Tableau de classement des différents couples d'amorces spécifiques de l'interleukine 6, IL-6 (GenBank NM\_012589) présentant à la fois les séquences, la nature et la position des incorporations de monomères LNA pour chaque amorce testée. Les amorces les plus spécifiques apparaissent sur fond vert, les amorces les moins spécifiques apparaissent sur fond rose. \*Les séquences des amorces RIL6 F3L et RIL6 F6L sont identiques. \*\*Les séquences des amorces RIL6 F5L et RIL6 F7L sont identiques.

Oligonucléotides	Séquence 5'P – 3'OH	Nature de la Base	Nombre & Positionnement en 5'P
RIL6 F0	GCGATGATGCACTGTCAGAAAAC	-	-
RIL6 R0	AACGGAACTCCAGAAGACCAGAG	-	-
RIL6 F1L	GCGATGATGCAC T GTCAGAAAAC	PYRIMIDINES PURINE	4, ESPACÉS
RIL6 R1L	AACGGAACTCCAG AAGACCAGAG	PURINES PYRIMIDINE	
RIL6 F2L	GC GATGATGC ACTGTCAGAAAAC	PYRIMIDINES	2, ÉCARTÉS
RIL6 R2L	AA C GGAACTC CAGAAGACCAGAG	PURINE PYRIMIDINE	
RIL6 F3L*	GC GATGATGCACTGTCAGAAAAC	PYRIMIDINE PURINE	2, EXTRÉMITÉ
RIL6 R3L	AA C GGAACTCCAGAAGACCAGAG	PURINES	
RIL6 F4L	GCGATGATGC ACTGTCAGAAAAC	PURINE PYRIMIDINE	2, MILIEU
RIL6 R4L	AACGGAACTC CAGAAGACCAGAG	PURINE PYRIMIDINE	
RIL6 F5L**	GCGATGATGC ACTGTCAGAAAAC	PYRIMIDINES	2, ÉCARTÉS
RIL6 R5L	AA C GGAACTC CAGAAGACCAGAG	PYRIMIDINES	
RIL6 F6L*	GC GATGATGCACTGTCAGAAAAC	PYRIMIDINE PURINE	2, EXTRÉMITÉ
RIL6 R6L	AA C GGAACTCCAGAAGACCAGAG	PYRIMIDINES	
RIL6 F7L**	GCGATGATGC ACTGTCAGAAAAC	PYRIMIDINES	2, MILIEU
RIL6 R7L	AACGGAACTC CAGAAGACCAGAG	PYRIMIDINES	

Détermination  
de la Température de Fusion  
et des Paramètres Thermodynamiques  
d'un Oligonucléotide

---

## 6.1. Introduction

Les performances de la qPCR dépendent du comportement des amorces au cours de l'étape d'hybridation (SantaLucia 2007). Ainsi, l'une des meilleures façons d'optimiser cette réaction est de permettre une hybridation simultanée des deux amorces au niveau de leurs séquences complémentaires. Le paramètre de prédiction de l'hybridation le plus utilisé est celui de la température de fusion d'un duplexe ADN, notée  $T_m$ . Cette température correspond au point médian de la dénaturation de la double hélice d'ADN induite par une augmentation de la température. À cette température, 50% de l'ADN présent dans le milieu se trouve alors sous forme simple brin.

### 6.1.1. Rappel sur la complémentarité des bases de l'ADN

Les quatre bases de l'ADN s'associent selon les règles définies par WATSON & CRICK (Watson and Crick 1953). Ces règles d'appariement sont les suivantes :

- l'adénine (*purine notée A*) s'apparie avec la thymine (*pyrimidine notée T*)
- la cytosine (*pyrimidine notée C*) s'apparie avec la guanine (*purine notée G*)

Ces appariements purines-pyrimidines se réalisent obligatoirement en raison d'un espace beaucoup trop faible entre deux purines et d'un espace beaucoup trop important entre deux pyrimidines pour une adaptation convenable de la structure de la double hélice ADN (Watson and Crick 1953). Les hybridations spécifiques entre l'adénine avec la thymine d'une part et la guanine avec la cytosine d'autre part, son dues à la proximité de groupements amines et hydroxyles des deux bases. Deux liaisons H sont établies entre une A et une T, et trois liaisons H sont établies entre une C et une G (**Figure 6.1**). Ainsi, les règles de WATSON & CRICK permettent de déduire la séquence complémentaire d'un brin ADN.

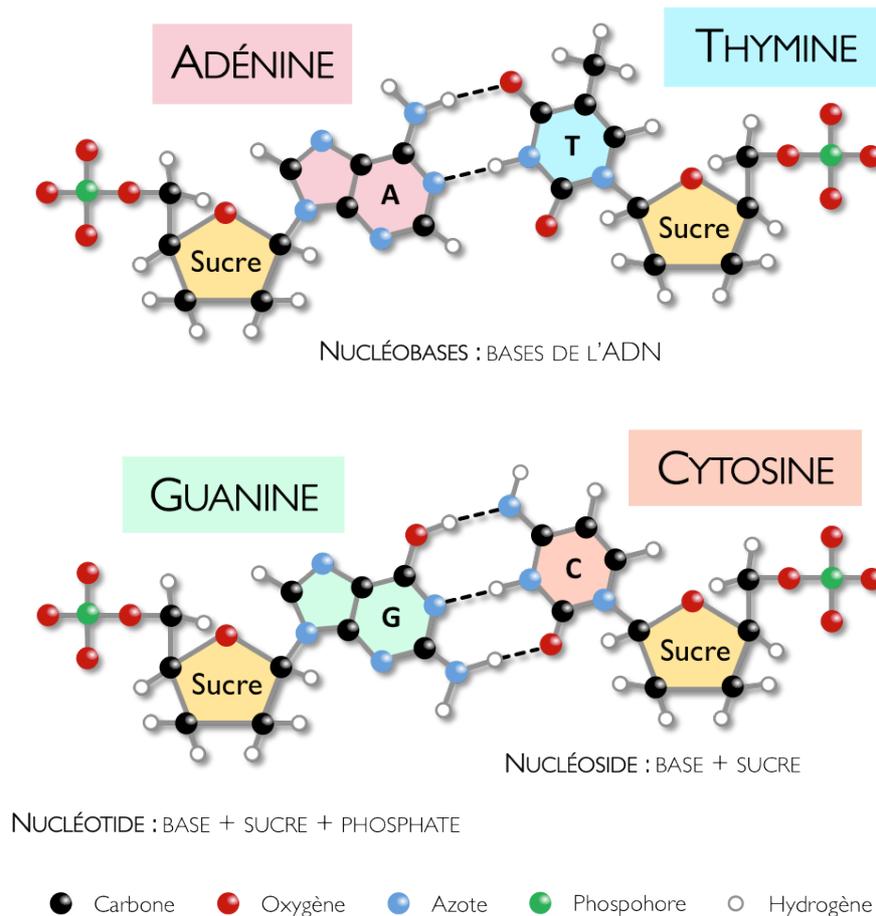


Figure 6.1 : Appariement des quatre nucléobases de l'ADN selon les règles d'appariement définies par WATSON & CRICK en 1953.

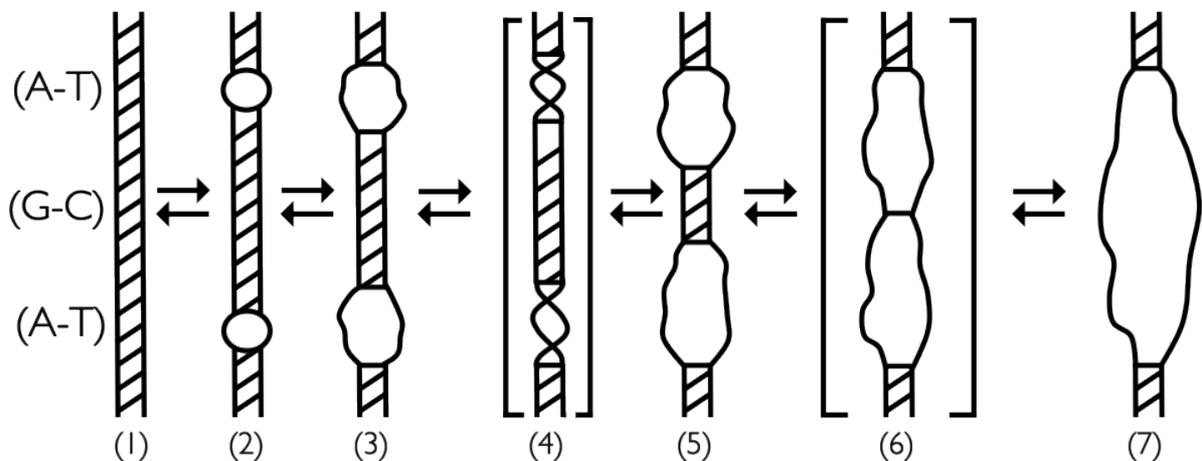
### 6.1.2. Comportements de l'ADN en fonction de la température

Les effets de la température sur les molécules d'acides nucléiques ont très rapidement été mis en évidence. Avant même la découverte de la double hélice d'ADN, René THOMAS décrira l'existence d'une **structure secondaire** due à des liaisons labiles dans les molécules d'acides nucléiques (Thomas 1951), puis leur **dénaturation** trois ans plus tard (Thomas 1954).

#### 6.1.2.1. Dénaturation de l'ADN

Lorsque de l'ADN double brin est chauffé, les liaisons hydrogène établies entre les deux **séquences complémentaires** de la molécule se déstabilisent et finissent par se rompre. Ce phénomène physique s'explique par la présence de zones présentant une énergie libre de liaison plus faible (Marko and Siggia 1995 ; Cocco and Monasson 1999 ; Strick, Allemand et al.

2000). Des ouvertures locales, parfois dénommées « bulles de dénaturation » (Fye and Benham 1999 ; Hwa, Marinari et al. 2003), peuvent alors apparaître sous l'effet d'une augmentation de température en imposant une torsion supplémentaire au reste de la molécule. Il a été identifié que ces zones énergétiquement favorables à un tel phénomène sont caractérisées par des séquences AT-riches (Marmur and Doty 1962 ; Breslauer, Frank et al. 1986 ; Delcourt and Blake 1991 ; Steger 1994). La **Figure 6.2** illustre ce modèle de **dénaturation** de l'ADN, proposé pour la première fois par Hanns-Christof SPATZ en 1969 (Spatz and Crothers 1969).



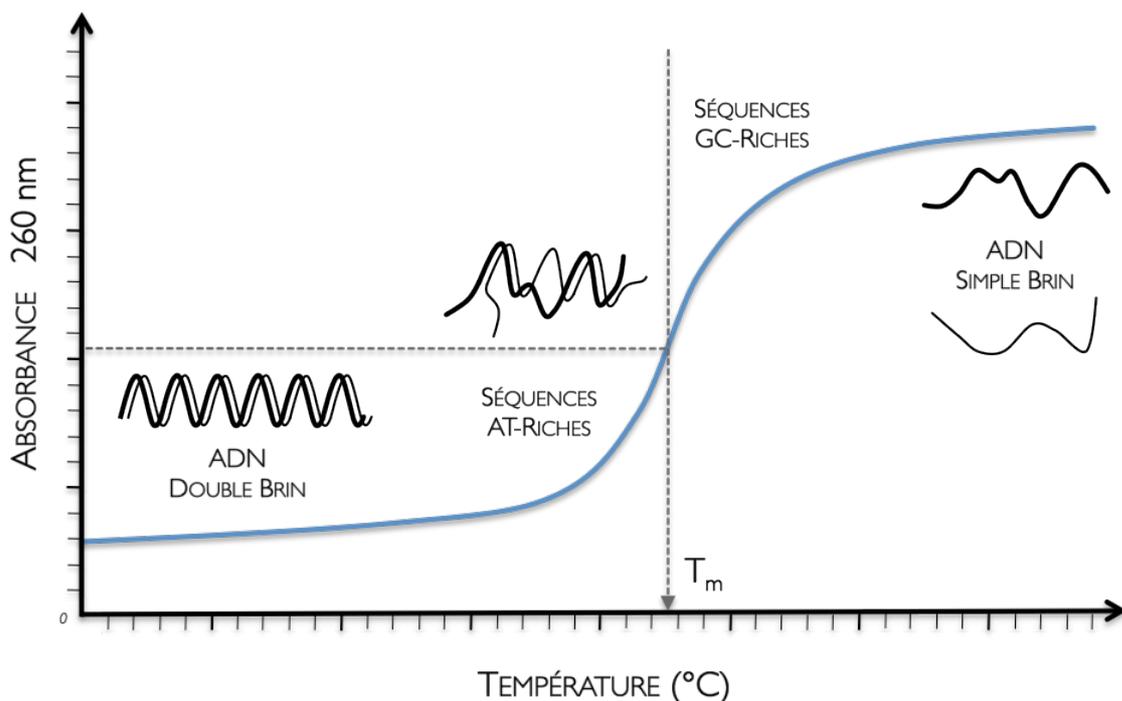
**Figure 6.2** : Dénaturation de l'ADN à partir de la présence de bulles de dénaturation. L'état (1) est une double hélice intacte contenant des régions localement riches en paires de bases A-T et G-C. La nucléation de la fusion se produit au niveau des régions AT-riches, produisant à l'équilibre une répartition des états (2) et (3). Une perturbation supplémentaire par la température doit se réaliser pour que la molécule passe à l'état (3). La réponse immédiate est le désappariement de quelques paires de bases à l'extrémité de sections hélicoïdales, avec une torsion reprise au niveau des régions adjacentes, état (4). Il s'ensuit le déroulement de la molécule, les états (4) et (5) sont à l'équilibre, avec le désappariement de quelques paires de bases au niveau des extrémités des sections hélicoïdales. À la température finale, certaines régions hélicoïdales sont proches de leur température critique de transition et subissent une transition lente de l'état (5) par l'intermédiaire de l'état (6) vers l'état (7). La transition de l'état (5) vers l'état (7) (dans des conditions de relaxation) est proposée comme l'origine de l'effet lent de la dénaturation (d'après CROTHERS et al. 1969).

D'autre part, il a été démontré que les liaisons hydrogène des paires de bases situées aux extrémités de la molécule d'ADN sont beaucoup moins stables que celles situées à l'intérieur de la séquence (Zimm 1960). Le phénomène de fusion ou de **dénaturation** de l'ADN est donc de nature *coopérative* : la déshybridation des **extrémités** d'une molécule double brin

ainsi que de ses régions AT-riches, beaucoup plus dynamiques, finit par déstabiliser les régions GC-riches adjacentes plus fortement hybridées. Une augmentation suffisante de la température conduit alors à la **dénaturation** progressive de la double hélice d'ADN toute entière en deux molécules d'ADN simple brin et antiparallèles.

### 6.1.2.1. Effet hyperchromique de l'ADN

La **dénaturation** de l'ADN double brin peut être observée par spectrophotométrie (**Figure 6.3**). À la longueur d'onde de 260 nm, le coefficient d'extinction molaire d'une solution d'ADN double brin est de  $0,020 \text{ (pg/ml)}^{-1} \cdot \text{cm}^{-1}$ . Il est de  $0,027 \text{ (pg/ml)}^{-1} \cdot \text{cm}^{-1}$  pour une solution d'ADN simple brin. Pour de courts **oligonucléotides** simple brin, il est dépendant de la longueur et de la composition en bases de la séquence (Maniatis, E.F. et al. 1989). On observe alors une augmentation de 40% de l'absorption de la lumière ultraviolette à 260 nm par cette solution. Décrit pour la première fois par Michel MICHELSON (Michelson 1958), cet effet est appelé *effet hyperchromique* de l'ADN. La **température de fusion**, notée  $T_m$ , est alors observée lorsque 50% de l'ADN se retrouve sous une forme simple brin dans le milieu.



**Figure 6.3 :** Mesure de la densité optique d'une solution d'ADN en fonction de la température. La courbe obtenue illustre l'hyperchromicité d'un acide nucléique en fonction de la température à laquelle il est soumis. Durant ce processus, les séquences AT-riches sont les premières régions à se désapparter en raison de leur faible nombre de liaisons hydrogène. La température de fusion, notée  $T_m$ , est la température à laquelle la moitié de l'ADN se trouve à l'état simple brin.

Au moins quatre facteurs sont capables d'influencer la valeur du  $T_m$  :

- **La composition en bases**

L'appariement de deux brins d'ADN est maintenu grâce à la complémentarité des bases. La **dénaturation** de l'ADN nécessite la rupture des liaisons hydrogène. Or, ce nombre est différent selon le couple de bases considéré. Le pourcentage GC est donc un paramètre important à prendre en compte dans l'estimation de la **température de fusion** d'une séquence ADN.

- **La longueur des fragments d'ADN**

Le  $T_m$  d'une molécule d'ADN augmente avec sa taille. En effet, plus une séquence est longue, plus elle contient de liaisons hydrogène à dissocier.

- **Les mésappariements**

Différentes études ont montré que les **mésappariements** étaient capables d'influencer la **température de fusion** d'une séquence ADN (Hall, Gula et al. 1980 ; Baldino, Chesselet et al. 1989). Pour 1% de **mésappariements**, sa valeur est généralement diminuée de 1°C (Maniatis, E.F. et al. 1989). Mais à la différence de l'ARN, certains **mésappariements** observés au niveau de l'ADN sont stabilisant, en particulier les guanines (Allawi and SantaLucia 1997).

- **La nature du milieu de l'ADN**

Les cations monovalents sont capables de diminuer la répulsion électrostatique des groupements phosphate chargés négativement. Aux fortes concentrations (supérieures à 1M), ils n'influencent pas la valeur du  $T_m$ . Par contre, on observe généralement une diminution du  $T_m$  pour des concentrations plus faibles. Les cations divalents influencent de manière beaucoup plus importante la valeur du  $T_m$ . Ainsi, l'effet stabilisant des ions  $Mg^{2+}$  est environ 140 fois plus important que celui des ions  $Na^+$  sur la stabilité d'un **duplexe** ADN (Nakano, Fujimoto et al. 1999). D'autres substances telles que le formamide, le DMSO, les ions ammonium et la bêtaïne diminuent les interactions entre les bases et abaissent par conséquent la valeur du  $T_m$ .

### 6.1.2.2. La renaturation de l'ADN

Un processus réversible caractérise les changements observés au niveau de la **structure secondaire** de l'ADN double brin. En effet, si la chaleur dénature cette molécule, un refroidissement lui restitue sa conformation d'origine *i.e.* la double hélice. Dans ces conditions, l'incidence de la vitesse de refroidissement représente un impact non négligeable sur le résultat

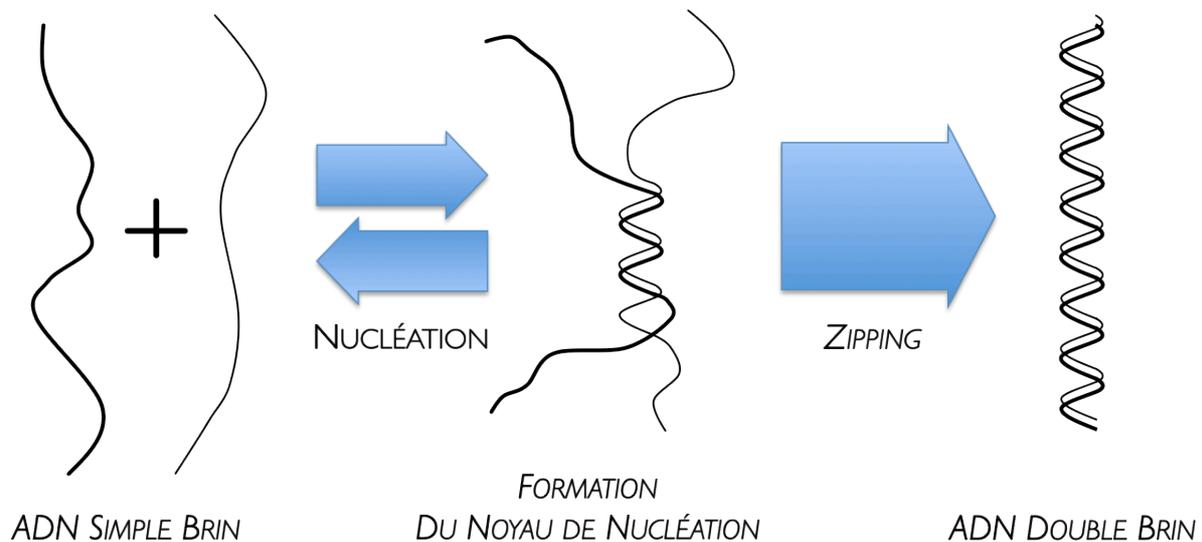
obtenu. Ainsi, lorsqu'il est rapide, il ne permet la formation que de quelques régions double brin formées par l'appariement de bases ou de courtes régions complémentaires intra ou intermoléculaires. Inversement, plus ce refroidissement se réalise lentement, plus la probabilité de rencontre des brins d'ADN complémentaires devient importante. Cette renaturation de régions complémentaires à partir de différents brins d'ADN caractérise ainsi l'hybridation des séquences d'acides nucléiques. Elle dépend de la collision aléatoire entre les brins d'ADN complémentaires. En fin de renaturation, toutes les molécules de l'échantillon ont retrouvé une forme double brin, présentant la même absorbance que l'échantillon d'origine.

Il est important de noter que la complexité d'une séquence ADN peut être mesurée à l'aide d'une courbe de  $C_0t$  représentant le pourcentage d'ADN simple brin observé en fonction de la concentration totale d'ADN et du temps (Britten, Graham et al. 1974) pour revue). Au cours de cette analyse, la probabilité de nucléation est proportionnelle à la racine carrée de la longueur des séquences ADN (Britten and Davidson 1976)

La cinétique d'hybridation de deux brins d'ADN est une réaction d'ordre 2 (Marmur and Doty 1961). Un premier événement de nucléation conduit tout d'abord à l'association correcte de quelques paires de bases entre les deux séquences complémentaires. Ces premiers niveaux d'interaction au niveau de la séquence sont appelés noyaux de nucléation. L'étape qui se produit ensuite est une association beaucoup plus rapide survenant entre les deux molécules, un peu à la manière d'une fermeture éclair. C'est au cours de cette étape bidirectionnelle appelée zipping de l'ADN (Murugan 2002) que le reste de l'appariement des bases se réalise ensuite (Figure 6.4). Cette cinétique de deuxième ordre se compose donc d'une première étape bimoléculaire assez lente et d'une deuxième étape monomoléculaire beaucoup plus rapide, de l'ordre de  $10^6$  à  $10^7$  paires de bases par seconde (Porschke and Eigen 1971).

Si l'on observe attentivement l'enchaînement de ces deux étapes, on comprend facilement que l'étape limitante de la réaction d'hybridation est l'événement de nucléation bimoléculaire. Les travaux réalisés par Gérald MANNING en 1976 ont montré qu'une longueur minimale d'une quinzaine de nucléotides était nécessaire afin d'établir la formation d'hybrides stables et spécifiques entre des acides nucléiques, dans des conditions thermiques favorables, que ce soit au cours de l'hybridation ou de la renaturation. Selon lui, le noyau de nucléation peut être défini par deux séquences non appariées d'une longueur de 13 à 16 bases consécutives et complémentaires. Dans ces conditions, un événement de nucléation peut alors se produire avec l'établissement d'une ou de plusieurs paires de bases consécutives au niveau

de ce même **noyau** (Manning 1976). Ce mécanisme est caractérisé par une réversibilité observée au cours de l'établissement du **noyau de nucléation** (Figure 6.4).



**Figure 6.4** : Cinétique d'hybridation d'ordre 2 de deux séquences d'ADN complémentaires de forme simple brin (d'après (Wetmur 1991)).

Au moins quatre facteurs sont capables d'influencer l'hybridation :

- **La température**

La probabilité de rencontre de deux **séquences complémentaires** et donc la vitesse d'hybridation augmente avec la température. Cependant le pourcentage de molécules hybridées à l'équilibre diminue avec la température. Par conséquent les meilleures vitesses d'association sont observées pour des températures situées 25°C en dessous du  $T_m$  de la molécule considérée (Wetmur and Davidson 1968).

- **La taille du fragment**

Au cours de l'hybridation de deux **séquences complémentaires** strictement identiques, la vitesse de réassociation augmente de manière proportionnelle avec la racine carrée de la longueur des fragments considérés (Britten and Davidson 1976).

- **Nature et concentration des acides nucléiques**

La vitesse de réassociation dépend à la fois de la nature de l'hybride et de sa concentration. Si l'ADN est en excès dans le milieu, la vitesse d'hybridation ADN/ADN est

supérieure à celle de l'ARN/ADN. Si l'ARN est en excès dans le milieu, la vitesse d'**hybridation** ADN/ADN est équivalente à celle de l'ARN/ADN (Casey and Davidson 1977).

- **La force ionique du milieu**

La concentration en cations joue un rôle important au cours de l'**hybridation** de **séquences complémentaires**. À titre d'exemple, la vitesse d'**hybridation** est proportionnelle au cube de la concentration en  $\text{Na}^+$  à faible concentration saline. La dépendance de la vitesse d'**hybridation** à l'égard de la concentration en sodium devient moins forte aux fortes concentrations en sodium, mais augmente encore par un facteur 2 entre 0,4 et 1,0 M (Wetmur 1976). D'autre part, la présence d'ions  $\text{Mg}^{2+}$  dans le milieu accélère le processus d'**hybridation** de **séquences complémentaires** de manière significative (Carrillo-Nava, Mejia-Radillo et al. 2008).

### 6.1.3. Méthodes prédictives de la température de fusion

Bien que ce paramètre puisse être déterminé expérimentalement dans certaines circonstances, il est beaucoup plus courant de calculer mathématiquement sa valeur. Ainsi, une grande variété d'équations considérant la composition en bases, la longueur, la concentration en sels, les valeurs thermodynamiques et/ou celles du *plus proche voisin* ont été développées afin d'estimer la valeur de la **température de fusion** pour de courtes séquences oligonucléotidiques (Panjkovich and Melo 2005). Cependant, il est important de remarquer que toutes ces méthodes prédictives ont été déterminées à partir de séquences différentes et dans des conditions théoriques spécifiques (en général, 1M  $\text{Na}^+$ ), très différentes de celles utilisées en **PCR**. Il n'y a donc pas de véritables consensus sur les corrections à apporter en fonction de la force ionique et des cations présents dans le milieu de **PCR**.

### 6.1.3.1. Méthode de calcul du $T_m$ basée sur la séquence de l'oligonucléotide

La formule la plus simple qui soit connue est la règle de WALLACE (Thein and Wallace 1986) :

$$T_m = 2 \times (wA + xT) + 4 \times (yG + zC) \quad (6.1)$$

où  $w$ ,  $x$ ,  $y$  et  $z$  désignent respectivement le nombre d'adénines, de thymines, de guanines et de cytosines retrouvées dans la séquence de l'oligonucléotide (Marmur and Doty 1962). Cette formule n'est cependant valable que pour la formation de **duplexes** sans aucun **mésappariement** dont la longueur reste **inférieure à 14 bases**, à une concentration d'oligonucléotide de 50 nM, en présence de 50 mM de cations monovalents et à un pH proche de 7,0.

Pour des oligonucléotides plus longs, il est proposé la formule suivante (Marmur and Doty 1962 ; Wallace, Shaffer et al. 1979 ; Maniatis, E.F. et al. 1989) :

$$T_m = 64,9 + 41,0 \times \frac{yG+zC-16,4}{wA+xT+yG+zC} \quad (6.2)$$

Cette équation suppose que l'**hybridation** se produit dans des conditions standard, en solution tampon de 50 mM de  $\text{Na}^+$ , à une concentration d'oligonucléotides de 50 nM et à un pH proche de 7,0.

### 6.1.3.2. Méthodes de calcul du $T_m$ tenant compte de la concentration en sodium

Pour des oligonucléotides de **moins de 14 nucléotides**, la formule (6.1) est modifiée de manière à prendre en compte la concentration en sodium du milieu où se déroule l'**hybridation** (Maniatis, E.F. et al. 1989):

$$T_m = 2 \times (wA + xT) + 4 \times (yG + zC) - 16,6 \times \log_{10} (0,050) + 16,6 \times \log_{10} [\text{Na}^+] \quad (6.3)$$

Le terme  $16,6 \times \log_{10} ([\text{Na}^+])$  permet d'ajuster la **température de fusion** en fonction des changements de concentration en sodium. Le terme  $\log_{10} (0,050)$  permet cet ajustement pour une concentration de 50 mM en  $[\text{Na}^+]$ . D'autres sels monovalents ou bivalents peuvent avoir un effet sur le  $T_m$  d'un oligonucléotide mais d'une manière générale, les ions sodium majoritaires dans les tampons de PCR ont été les premiers étudiés, même si des traces de cations divalents souvent négligées sont importantes à prendre en compte (Nakano, Fujimoto et al. 1999). À partir des travaux d'Ellis BOLTON (Bolton and Mc 1962), Edward FRITSCH

établit une formule plus précise du  $T_m$  pour des oligonucléotides de longueur plus importante, de 15 à 70 nucléotides (Maniatis, E.F. et al. 1989) :

$$T_m = 81,5^\circ\text{C} + 16,6^\circ\text{C} \times (\log_{10} [Na^+]) + 0,41^\circ\text{C} \times (\%GC) - 600/N \quad (6.4)$$

où  $N$  est le nombre de nucléotides présents dans la séquence

Dans cette équation, les troisième et quatrième termes permettent de prendre en compte respectivement la teneur en bases GC et la longueur de la séquence.

### 6.1.3.3. Méthodes de prédiction du $T_m$ selon le modèle du plus proche voisin

À côté de ces formules empiriques basées sur la composition générale en bases fortes et bases faibles, il existe des méthodes capables de tenir compte de la séquence précise de l'oligonucléotide (Breslauer, Frank et al. 1986 ; Rychlik, Spencer et al. 1990 ; Rychlik 1995 ; ; SantaLucia and Hicks 2004). Plus précisément, elles s'appuient sur le modèle du plus proche voisin établi par Philip BORER en 1974 (Borer, Dengler et al. 1974). Pour les acides nucléiques, ce modèle suppose que la stabilité d'une paire de bases donnée dépend de l'identité et de l'orientation des paires de bases voisines. L'équation du  $T_m$  qui découle de ce modèle (SantaLucia, Allawi et al. 1996) est :

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln C_T} \quad (6.5)$$

où  $R$  est la constante universelle des gaz parfaits ( $1,9872 \text{ cal.mol}^{-1}.\text{K}^{-1}$ ).

Pour les molécules non auto-complémentaires, le terme  $C_T$  de l'équation 6.5 est remplacé par  $C_T/4$  si les brins sont en concentrations égales ou par  $(C_A - C_B/2)$  si les brins sont à des concentrations différentes, où  $C_A$  et  $C_B$  sont respectivement la concentration du brin le plus concentré et la concentration du brin le moins concentré.

Le modèle du plus proche voisin permet notamment d'assigner les variations  $\Delta H$  et  $\Delta S$  observées lors de la formation de chacune des dix différentes interactions possibles au niveau d'un duplexe ADN (pour mémoire, se reporter au Tableau I.1 restituant les paramètres d'énergie libre ainsi que les paramètres d'initiation en Introduction de ce manuscrit).

Toutes ces méthodes présentent cependant un problème, elles ont été développées à partir de séquences différentes et ne sont précises que pour les séquences du même type. En 2005, Alejandro PANJKOVICH a évalué les capacités des différentes méthodes de prédiction du  $T_m$

pour les séquences utilisées dans le développement des autres méthodes (Panjkovich and Melo 2005). Il a montré que l'utilisation combinée des différentes méthodes dans des conditions adaptées à la séquence cible permettait une excellente prédiction du  $T_m$  dans des conditions de référence (1M Na<sup>+</sup>). Son logiciel, *dnaMATE* (Panjkovich, Norambuena et al. 2005), est basé sur le modèle *du plus proche voisin* et est capable de déterminer un  $T_m$  consensus en fonction de la séquence (longueur, pourcentage GC, etc.) à partir de trois méthodes prédictives : celle de Kenneth BRESLAUER (Breslauer, Frank et al. 1986) démontrant qu'un **duplexe** ADN peut être thermodynamiquement considéré comme la somme des interactions de ses plus proches voisins, celle de John SANTALUCIA (SantaLucia 1998) restituant les paramètres thermodynamiques unifiés pour un **oligonucléotide** et celle de Naoki SUGIMOTO (Sugimoto, Nakano et al. 1996) démontrant que la variation d'énergie libre du facteur d'initiation pour les **duplexes** ADN:ADN était la même que pour les **duplexes** ADN:ARN (3,4 kcal.mol<sup>-1</sup>). La valeur du  $T_m$  est alors obtenue en appliquant un coefficient de correction sur la séquence. On notera cependant que cette méthode reste toujours basée sur un milieu théorique à 1M Na<sup>+</sup>, qui n'a rien à voir avec les conditions ioniques rencontrées en PCR. En effet, il existe un effet stabilisant plus important de la part des ions magnésium (Owczarzy, Moreira et al. 2008) et leur concentration peut être diminuée par simple chélation stoechiométrique par les désoxynucléotides triphosphates présents dans le milieu (Baumforth, Nelson et al. 1999 ; von Ahsen, Wittwer et al. 2001). Face à ce problème, de nombreux auteurs ont donc proposé plusieurs corrections empiriques afin de permettre une prédiction plus précise du  $T_m$  des **oligonucléotides**. Richard OWCZARZY propose alors une prédiction de la stabilité des **duplexes** ADN en solution contenant à la fois des cations monovalents (Na<sup>+</sup>) et divalents (Mg<sup>2+</sup>), toujours selon le modèle *du plus proche voisin* (pour revue (Owczarzy, You et al. 2004 ; Owczarzy, Moreira et al. 2008)). Il montre que le  $T_m$  peut être prédit à partir de l'équation suivante :

$$\frac{1}{T_m(Mg^{2+})} = \frac{1}{T_m(1MNa^+)} + a + b \ln[Mg^{2+}] + f_{GC} (c + d \ln[Mg^{2+}]) + \frac{1}{2(N_{bp}-1)} [e + f \ln[Mg^{2+}] + g(\ln[Mg^{2+}])^2] \quad (6.6)$$

où  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $e$ ,  $f$ , et  $g$  sont des paramètres d'ajustement de l'équation pour une régression linéaire multivariée

Comme les paramètres thermodynamiques dépendent de la séquence des **oligonucléotides**, la valeur prédictive de ces méthodes est encore plus incertaine lorsque sont incorporés des *Locked Nucleic Acids*. La première étude à s'intéresser aux paramètres thermodynamiques des LNA, établis à partir du modèle *du plus proche voisin*, est celle de Patricia McTIGUE (McTigue, Peterson et al. 2004). Elle complète les paramètres existants pour des séquences modifiées par un seul monomère LNA et sera améliorée sept ans plus tard par celle de Richard OWCZARZY qui mettra à jour de nouveaux paramètres pour des séquences modifiées par un et deux monomères LNA consécutifs au sein de la séquence de l'**oligonucléotide** (Owczarzy, You et al. 2011).

#### 6.1.3.4. Variations autour de l'équation thermodynamique du $T_m$

En se basant sur les équations de prédiction thermodynamique du  $T_m$ , on peut envisager de mesurer ces paramètres en modifiant les conditions expérimentales. On peut alors décrire l'**hybridation** d'une **amorce** de manière assez simple par l'équation suivante :



où  $P$ ,  $T$  et  $P-T$  représentent respectivement la concentration d'**amorce** libre, la concentration de la cible ADN libre et la concentration de l'**amorce** hybridée.

À l'équilibre, la proportion relative entre les concentrations d'**amorce** libre, de cible libre et de produit réactionnel est donnée par la constante d'équilibre :

$$k = \frac{[P-T]_{eq}}{[P]_{eq} \cdot [T]_{eq}} \quad (6.8) \quad \text{et} \quad k = \frac{f}{[P]_{eq} \cdot (1-f)} \quad (6.8')$$

où  $f$  est le taux d'ADN cible hybridé.

En PCR, comme la concentration d'**amorce** libre est en excès par rapport à la concentration de la séquence cible libre, on observe  $[P]_{eq} \approx [P]_{in}$ .

Une telle proportion est dépendante de la température. Une méthode consensuelle pour la prédiction du  $T_m$  est alors :

$$T = \frac{\Delta H^\circ}{\Delta S^\circ - R \ln(k)} \quad (6.9)$$

où  $T$  est la température absolue en K

$R$  est la constante universelle des gaz parfaits (1.987 cal.mol<sup>-1</sup>.K<sup>-1</sup>)

$\Delta H$  est la variation d'enthalpie en cal.mol<sup>-1</sup>

et  $\Delta S$  est la variation d'entropie en cal.mol<sup>-1</sup>.K<sup>-1</sup>

L'équation de la température à laquelle l'équilibre atteint un taux  $f$  de cible ADN hybridée ( $T_f$ ) est :

$$T_f = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln\left(\frac{[P] \cdot (1-f)}{f}\right)} \quad (6.9')$$

Le  $T_m$  est alors défini par l'égalité  $[P-T]_{eq} = [T]_{eq}$ , ce qui entraîne :

$$k = 1/[P]_{eq} \quad (6.10)$$

À partir des équations (6.10) et (6.7), la définition du  $T_m$  devient :

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln([P]_{eq})} \quad (6.11)$$

La concentration d'amorce à l'équilibre peut être calculée en utilisant l'équation :

$$[P]_{eq} = [P]_{in} - [P-T]_{eq} \quad (6.12)$$

où  $[P]_{in}$  représente la concentration initiale de l'amorce

À la valeur du  $T_m$ , pour une molécule non auto-complémentaire, la concentration de l'amorce hybridée est donnée par l'équation :

$$[P-T]_{eq} = \frac{[T]_{in}}{2} \quad (6.13)$$

Ainsi, si  $[P]_{in} \geq [T]_{in}$ , l'équation (6.12) conduit à :

$$P]_{eq} = [P]_{in} - \frac{[T]_{in}}{2} \quad (6.14)$$

Si  $[P]_{in} = [T]_{in}$ , il advient que :

$$k = \frac{2}{[P]_{in}} \quad (6.14')$$

Dans les conditions de la PCR, comme  $[T]_{in}$  est négligeable au regard de  $[P]_{in}$ , on obtient :

$$k = \frac{1}{[P]_{in}} \quad (6.14'')$$

Ces équations ont servi de base à notre étude.

#### 6.1.4. Problématique de l'étude

La prédiction du comportement des amorces au cours de l'hybridation est un facteur clé dans l'amélioration de la technique de qPCR. En effet, l'hybridation simultanée des deux oligonucléotides est un prérequis pour une haute efficacité de réaction, permettant ainsi une quantification précise de la séquence ciblée (Panjkovich and Melo 2005). Le paramètre idéal serait le pourcentage d'hybridation des amorces à la température d'hybridation (SantaLucia 2007). Pour des raisons pratiques, le paramètre de prédiction de cette hybridation qui est utilisé est la température de fusion. Classiquement, la procédure consiste à choisir des paires d'amorces présentant des  $T_m$  prédits relativement proches puis de déterminer expérimentalement les conditions optimales de la PCR. Par ailleurs, une sous-estimation de la température de fusion d'une amorce PCR peut fréquemment conduire à l'amplification de produits non spécifiques (Panjkovich and Melo 2005). Un déséquilibre entre les amorces peut donc obliger l'utilisateur à optimiser séparément la concentration des amorces utilisées. Cette démarche est fastidieuse et, dans la pratique courante, elle apparaît comme assez peu efficace (Nolan, Hands et al. 2006).

Depuis les travaux d'Alejandro PANJKOVICH (Panjkovich, Norambuena et al. 2005), il est possible de prédire le  $T_m$  de courtes séquences d'ADN de façon précise mais dans des conditions d'environnement particulières. Or, la composition de ce milieu de référence (1M

$\text{Na}^+$ ) n'a aucun rapport avec un tampon de PCR classique qui contient en moyenne 50 mM d'ions  $\text{Na}^+$  et de 2 à 5 mM d'ions  $\text{Mg}^{2+}$ . Il est important de retenir que les ions  $\text{Mg}^{2+}$  ont un effet stabilisant environ 140 fois supérieur à celui des ions  $\text{Na}^+$  (Nakano, Fujimoto et al. 1999). Si les premières équations de correction du  $T_m$  définies en fonction de la force ionique du milieu utilisaient un facteur constant dépendant des conditions, les modèles plus élaborés prennent en compte un effet séquence-spécifique joué par ces ions (Zubay 1959 ; Mikulecky and Feig 2006). De plus, les tampons retrouvés dans le commerce contiennent de nombreux additifs capables d'affecter l'hybridation, dont la nature biochimique et la concentration restent souvent protégées par des brevets industriels.

Au cours de l'étude précédente, de manière à pouvoir interpréter objectivement les résultats sur la spécificité des amorces modifiées, nous avons besoin de connaître de façon précise le  $T_m$  de nos différentes amorces. Devant la discordance des résultats observés avec les différentes méthodes prédictives (Panjkovich and Melo 2005) alors disponibles (logiciel *OligoAnalyser 3.1* d'IDT (Owczarzy, Tataurov et al. 2008), modèle *du plus proche voisin* de McTIGUE (McTigue, Peterson et al. 2004), nous avons envisagé de mesurer directement le  $T_m$  dans le milieu de PCR par hybridation de l'oligonucléotide avec sa séquence complémentaire. En effet, Yong YOU démontre en 2011 que l'acquisition de profils de fusion de duplexes ADN sur des systèmes de PCR en temps réel peut permettre l'obtention de valeurs thermodynamiques précises par simple analyse de la fluorescence (You, Tataurov et al. 2011). La méthode de mesure directe du  $T_m$  initialement proposée par TEO *et al.* en 2002 (Teo, Choi et al. 2002), permet de tenir compte de l'ensemble des composants présents dans le milieu réactionnel lors de la mesure du  $T_m$  d'un oligonucléotide mais n'a cependant jamais été validée. Or, cette mesure pose un problème conceptuel réel : la vitesse importante de variation de la température des thermocycleurs est-elle compatible avec une mesure précise du  $T_m$ ? (Zipper, Brunner et al. 2004; Owczarzy, Moreira et al. 2008 )

Pour valider cette méthode, nous avons choisi de comparer la méthode avec une méthode de référence : la nano calorimétrie Différentielle à balayage, nanoDSC (Chalikian, Volker et al. 1999). La précision de la mesure du  $T_m$  que nous avons obtenue nous a alors incité à déduire expérimentalement le paramètre considéré comme le plus pertinent de l'hybridation, c'est-à-dire le pourcentage de l'amorce hybridée en fonction de la température. Pour cela, nous avons dû valider des conditions expérimentales permettant une mesure précise et reproductible de  $\Delta H$  et  $\Delta S$  dans le mix de PCR en utilisant l'équation (6.9).

## 6.2. Résultats expérimentaux – Méthode de mesure directe du $T_m$

### 6.2.1. Caractérisation de la méthode de mesure directe du $T_m$

Dans un premier temps, nous avons voulu caractériser la précision de la méthode de mesure directe du  $T_m$  par *courbe de fusion*. Ainsi, en mesurant la *température de fusion* d'un même *duplexe d'oligonucléotides* ADN ( $n=10$ ), on observe une précision *intra-essai* satisfaisante de la méthode avec un écart maximum entre les répliques de  $0,16^\circ\text{C}$  (*écart-type* =  $0,05^\circ\text{C}$ ). La précision *inter-essai* réalisée sur un seul instrument (4 amorces, 3 mesures indépendantes en *duplicate*) est équivalente. L'écart-type était de  $0,05^\circ\text{C}$  et le plus grand écart entre deux mesures était de  $0,17^\circ\text{C}$ .

Dans un deuxième temps et afin de définir la sensibilité de la méthode, nous avons réalisé une série d'expérimentations visant à observer la qualité de cette précision à différentes concentrations d'*amorce*, différentes concentrations de magnésium, différents intercalants fluorescents au cours de la *dénaturation* tout comme au cours de la *renaturation*.

#### 6.2.1.1 Effet de la concentration d'amorce

Afin de savoir si la méthode est capable de mesurer précisément la *température de fusion* d'un *oligonucléotide* dans une gamme de concentrations utilisées en qPCR, nous avons réalisé la *courbe de fusion* de quatre *duplexes* (composés des *amorces* RIL6 F0, RIL6 R0, RIL6 F2L et RIL6 R2L) à cinq différentes concentrations d'*amorce* ( $0,30$  ;  $0,35$  ;  $0,40$  ;  $0,45$  et  $0,50$   $\mu\text{M}$ ). Les résultats obtenus montrent que la méthode est suffisamment précise pour détecter l'augmentation de la valeur du  $T_m$  en fonction de la concentration d'*amorce* utilisée (**Figure 6.5**).

#### 6.2.1.2 Effet du magnésium sur la mesure de la température de fusion

Nous avons voulu caractériser la précision de la méthode dans des concentrations de magnésium couramment utilisées en qPCR. Afin d'observer la précision de la méthode de mesure dans ces conditions, nous avons réalisé une gamme de concentrations croissantes de 3 à 6 mM. La valeur du  $T_m$  en fonction de la concentration en magnésium apparaît aléatoire. La méthode ne permet pas de distinguer les variations du  $T_m$  dans une gamme de concentrations

de magnésium classiquement utilisées en PCR (Figure 6.6). Cela signifie que les gammes de concentrations utilisées en PCR ont une faible influence sur la réaction.

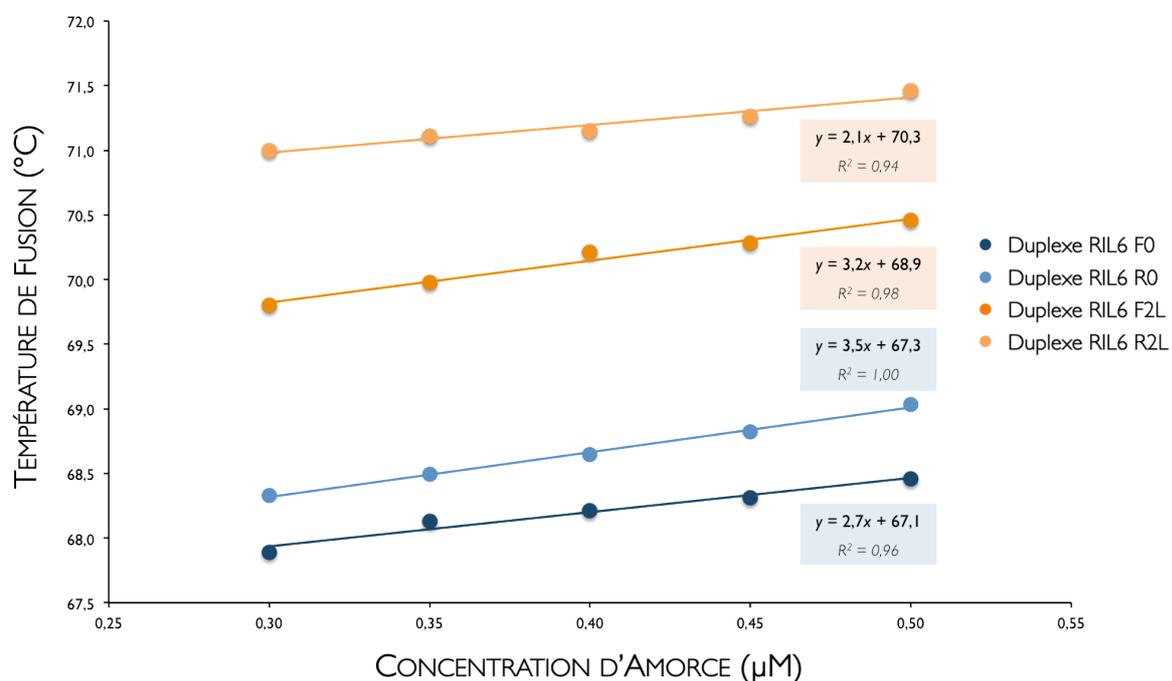


Figure 6.5 : Températures de fusion obtenues à partir d'une gamme de concentrations d'amorce de 0,3 µM à 0,5 µM pour les duplexes RIL6 F0, RIL6 R0, RIL6 F2L et RIL6 R2, analysés en tampon de PCR (32 mM Na<sup>+</sup>, 4 mM Mg<sup>2+</sup>, 1,2 mM dNTP).

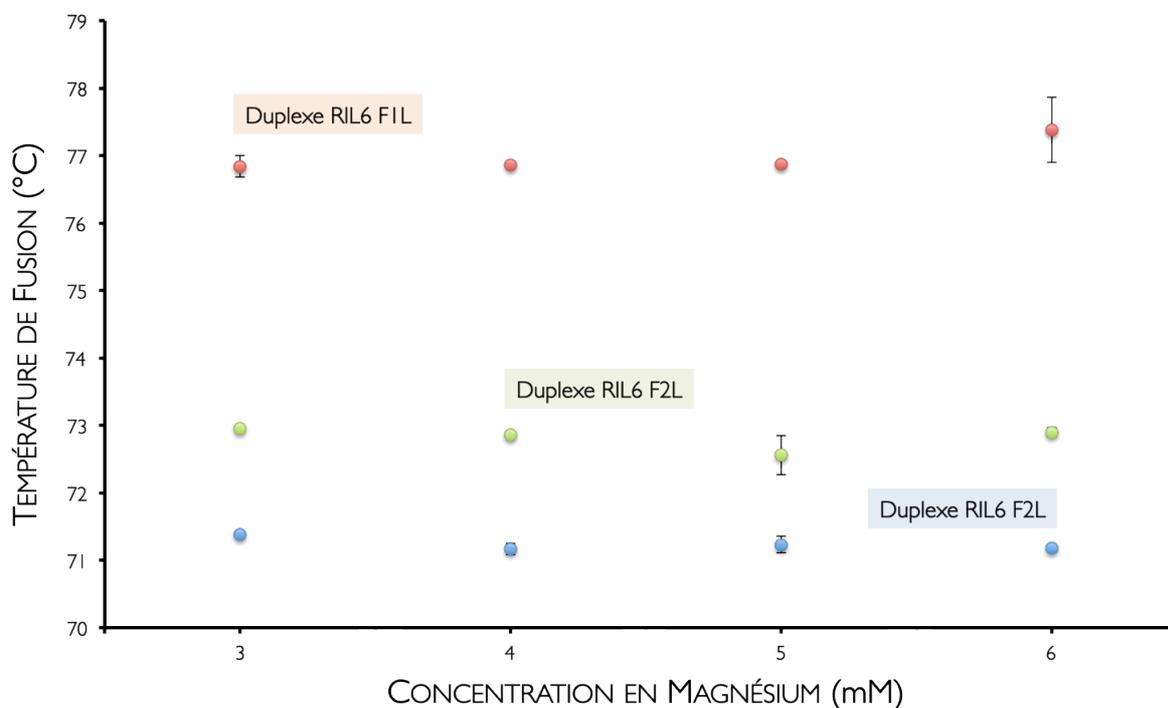
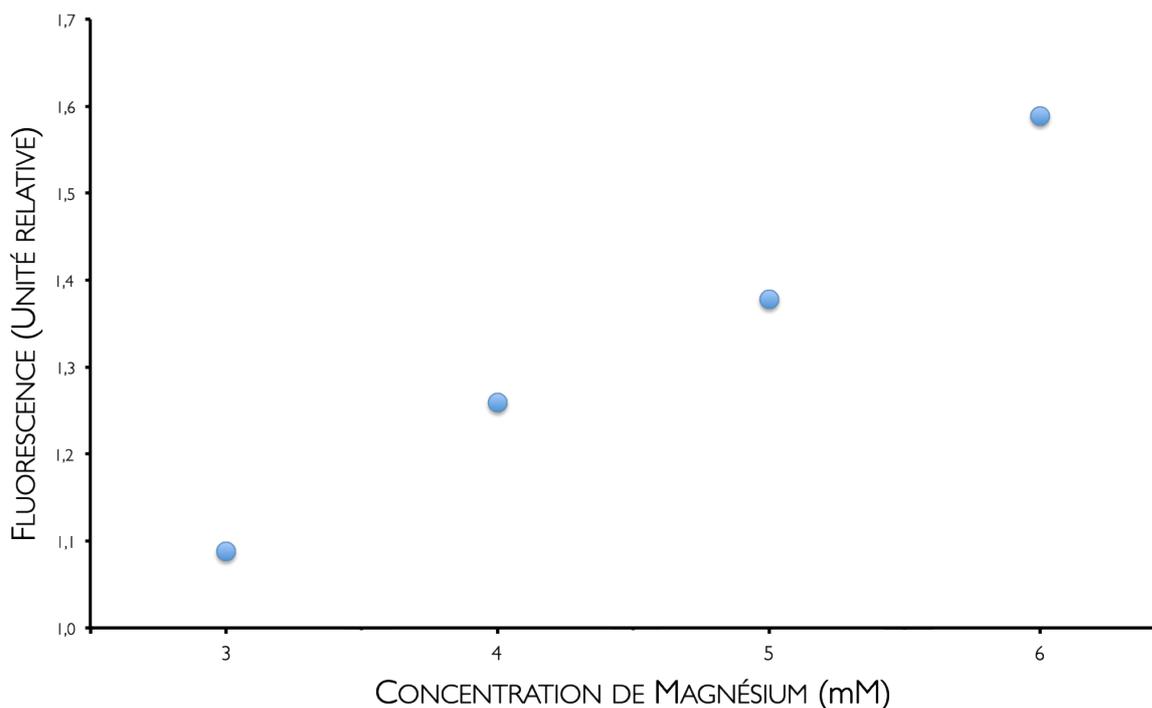


Figure 6.6 : Températures de fusion obtenues pour les duplexes RIL6 F0, RIL6 F1L et RIL6 F2L à 3, 4, 5 et 6 mM de magnésium en tampon de PCR (32 mM Na<sup>+</sup>, 1,2 mM dNTP).



**Figure 6.7 :** Corrélation obtenue entre fluorescence enregistrée à la température de fusion d'un duplexe et concentration en magnésium présent dans le milieu réactionnel.

Cependant, il faut souligner que la fluorescence du *SybrGreen I* est fortement liée au magnésium. En effet, on observe une augmentation de celle-ci d'environ 50% entre 3 et 6 mM (Figure 6.7). Il est à noter que cette différence de hauteur de pic est probablement responsable de la défaillance de l'algorithme. L'outil ne permet donc pas la mesure précise du  $T_m$  aux faibles concentrations de magnésium et nécessiterait le développement d'un algorithme spécifique capable de tenir compte de la hauteur du pic de fusion.

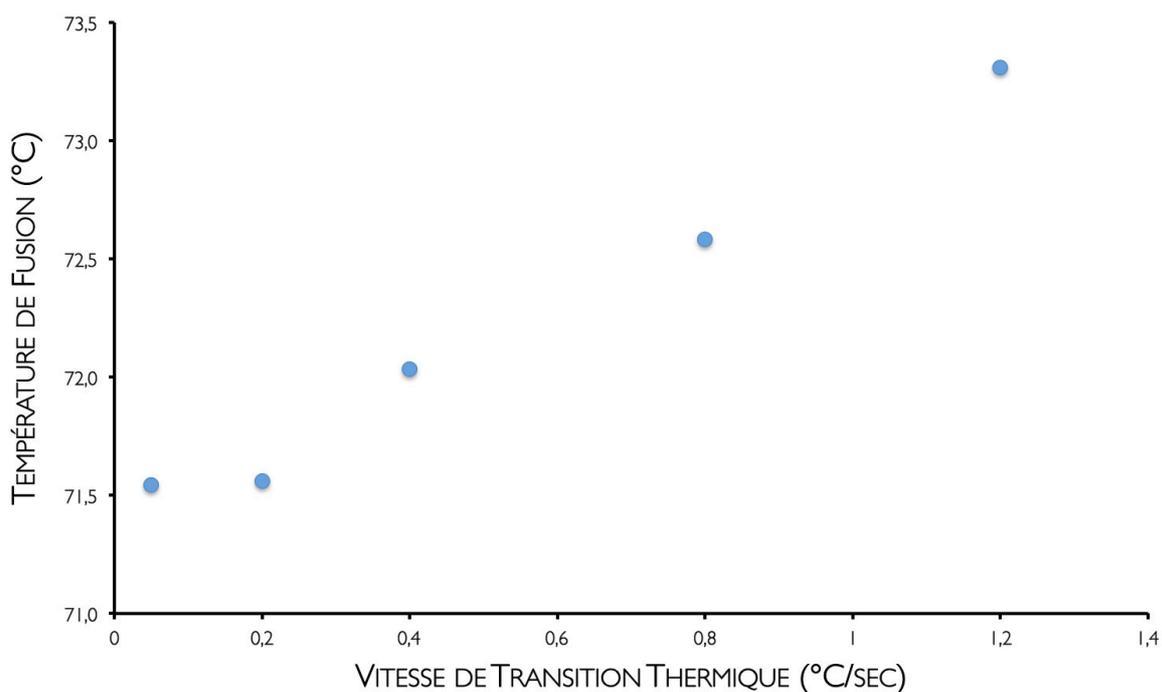
### 6.2.1.3 Effet de la vitesse de la transition thermique

En dénaturant un duplexe d'oligonucléotides (RCYCA R1) à 5 vitesses différentes (0,05 ; 0,20 ; 0,40 ; 0,80 et 1,20°C/sec), les résultats montrent qu'une augmentation de la vitesse de transition thermique induit une augmentation de la valeur de la température de fusion d'une même séquence ADN (Tableau 6.1). Notre méthode semble restituer une mesure relativement stable à partir d'une vitesse de transition de 0,2°C/sec (Figure 6.8). Cela signifie probablement qu'une vitesse élevée ne permet pas d'atteindre l'équilibre observé au  $T_m$ . Il existe donc un décalage du  $T_m$  vers des valeurs supérieures. En revanche, les plus faibles vitesses de transition thermique semblent permettre de mesurer un  $T_m$  suffisamment fiable.

Pour le vérifier, nous avons mesuré un même **duplexe** à la vitesse la plus faible successivement en **dénaturation** et **renaturation** en conditions de qPCR (32 mM Na<sup>+</sup>, 4 mM Mg<sup>2+</sup>, 1,2 mM dNTP). À la vitesse de 0,05°C/sec et 0,2°C/sec, on n'observe aucune différence significative avec une moyenne des valeurs absolues des différences entre  $T_m$  mesurés en **dénaturation** et  $T_m$  mesurés en **renaturation** de 0,15°C ± 0,15°C et 0,12°C ± 0,11°C respectivement ( $n = 8$ ,  $p = 0,3$  et  $p = 0,8$  ; test de *Wilcoxon* apparié). Cette vitesse peut donc être utilisée pour mesurer le  $T_m$ . En revanche, à la vitesse de 0,30°C/sec et 0,50°C/sec, on observe des valeurs de 0,73°C ± 0,11°C et 1,53°C ± 0,19°C ( $n = 8$ , significatif  $p = 0,013$  et  $p = 0,014$  ; test de *Wilcoxon* apparié). Dans ces conditions, l'équilibre n'est pas atteint entre deux points de mesure. Ainsi, une vitesse de 0,05°C/sec permet d'atteindre l'équilibre de l'**hybridation** entre deux points de mesure.

	Vitesse (°C.sec <sup>-1</sup> )				
	0,05	0,2	0,4	0,8	1,2
Moyenne (°C)	71,5	71,6	72,0	72,6	73,3
Écart-type	0,05	0,03	0,06	0,06	0,08

**Tableau 6.1** : Températures de fusion obtenues à partir d'un duplexe d'oligonucléotides ADN en fonction de la vitesse de transition thermique appliquée pendant le programme de fusion. Les mesures ont été obtenues en conditions de qPCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup>, 1,2 mM dNTP) à la concentration de 0,6 µM d'amorce.

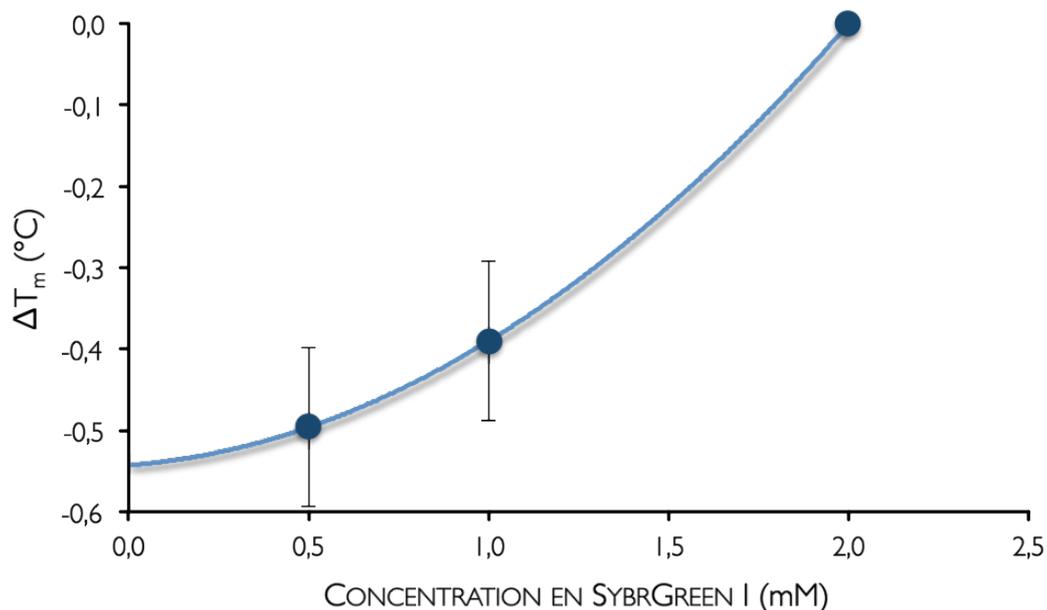


**Figure 6.8** : Mesure de la température de fusion en fonction de la vitesse de transition thermique.

#### 6.2.1.4 Effet des intercalants fluorescents

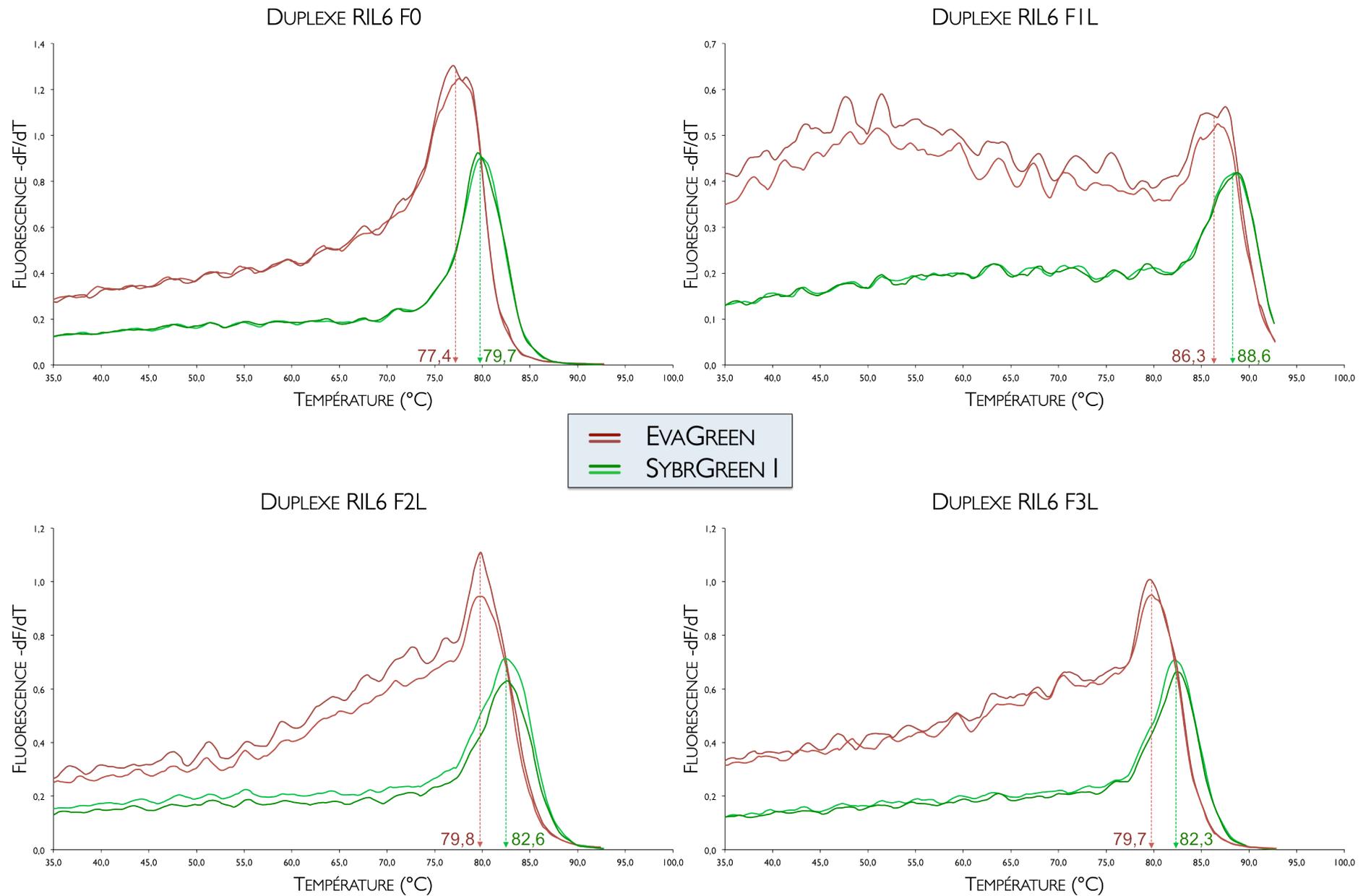
De nombreux éléments présents dans les tampons de PCR sont probablement capables de modifier le  $T_m$ . Il est donc intéressant de mesurer l'effet du *SybrGreen I* sur cette valeur. Cette vérification est importante, notamment pour l'utilisation d'amorces avec des sondes spécifiques utilisées sans intercalant. On sait que le *SybrGreen I* est un intercalant non saturant : le nombre de molécules qui va se fixer dépend de la taille mais aussi de la séquence du duplexe analysé (Gudnason, Dufva et al. 2007).

Nous avons mesuré le  $T_m$  de quatre duplexes par courbe de fusion à trois concentrations de *SyberGreen I* différentes (1X, X/2 et X/4) en tampon phosphate 1M  $\text{Na}^+$ . La concentration 1X correspond à la concentration que nous avons mesurée dans le tampon de PCR en utilisant la méthode décrite par ZIPPER *et al.* (Zipper, Brunner et al. 2004). Cette concentration correspond à la moitié de la concentration utilisée au début de l'utilisation du *SyberGreen I*. Nous observons une faible diminution de la température de fusion d'environ  $0,5^\circ\text{C}$  entre les concentrations 2X et 0,5X. L'extrapolation de cette courbe à l'origine prédit une différence de  $T_m$  négligeable (moins de  $0,2^\circ\text{C}$ ) entre la concentration classiquement utilisée dans les kits commerciaux et l'absence de *SyberGreen I* (Figure 6.9).



**Figure 6.9** : Effet de la concentration en *SybrGreen I* sur la valeur  $T_m$  des oligonucléotides courts, en utilisant une analyse par courbe de fusion. La valeur zéro observée en ordonnées a été attribuée pour une concentration de *SybrGreen I* de 2 mM (ZIPPER *et al.*, 2004) (Zipper, Brunner et al. 2004). Les valeurs de  $T_m$  ont été évaluées dans des conditions de qPCR (32 mM  $\text{Na}^+$ , 5 mM  $\text{Mg}^{2+}$  et 1,2 mM dNTP) à la concentration d'amorce de  $0,6 \mu\text{M}$ . Les valeurs sont représentées par la moyenne  $\pm$  écart-type ;  $n = 6$ .

Nous avons également vérifié les effets d'un intercalant saturant (dont le nombre de molécules fixées ne dépend pas de la séquence) comme l'*EvaGreen* (Eischeid 2011) sur la précision de notre méthode. Nous avons alors effectué des comparaisons de mesures réalisées en *duplicates* avec les deux intercalants fluorescents. À la concentration d'*amorce* de 0,6  $\mu\text{M}$ , le  $T_m$  mesuré en *EvaGreen* est d'environ 2°C inférieur à celui mesuré en *SyberGreen I* (Figure 6.10). D'autre part, les profils de fusion obtenus avec les deux intercalants montrent un bruit de fond supérieur en *EvaGreen* par rapport aux profils de fusion obtenus en *SyberGreen I*. Cette différence de  $T_m$  est probablement due à l'algorithme, plus adapté au *SyberGreen I*, probablement moins adapté à l'*EvaGreen*.



**Figure 6.10** : Courbes de fusion réalisées en *SybrGreen I* et *EvaGreen*, tampon phosphate, ( $1\text{ M Na}^+$ ,  $5\text{ mM Mg}^{2+}$ ) à la concentration d'amorce de  $0,6\text{ }\mu\text{M}$ . Les valeurs reportées sont les moyennes des valeurs mesurées par le logiciel en mode manuel (*LightCycler Software*<sup>®</sup>, Version 3.5 Roche Molecular Biochemicals).

### 6.2.1.5. Reproductibilité inter-instrument

Nous avons utilisé la mesure du  $T_m$  pour évaluer la variabilité de 6 thermocycleurs *LightCyclers*. Pour cela, nous avons réalisé la mesure de la température de fusion de deux duplexes ADN différents (ARBP F1 et RIL6 F0) dans des conditions de qPCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup>, 1,2 mM dNTP) à la concentration de 0,6 µM d'amorce. Chaque mesure a été réalisée deux fois sur chacun des 6 *LightCyclers* disponibles sur la plateforme de génomique. Les résultats obtenus sont présentés au niveau du **Tableau 6.2**. Il existe une corrélation importante entre la mesure des deux duplexes pour les différentes machines. La méthode de mesure directe est donc capable de détecter les différences de température d'étalonnage entre les machines. Globalement, ces températures sont relativement proches, la température de référence de 5 machines est comprise dans un intervalle d'environ 1°C (**Figure 6.11**). La 6<sup>ème</sup> machine obtient des valeurs un peu plus éloignées.

	Duplexe #1		Duplexe #2	
	Moyenne (°C)	Écart-type	Moyenne (°C)	Écart-type
Machine #1	70,05	0,08	70,83	0,05
Machine #2	70,62	0,07	71,45	0,04
Machine #3	70,21	0,01	71,07	0,04
Machine #4	69,70	0,04	70,61	0,02
Machine #5	70,57	0,01	71,38	0,03
Machine #6	69,24	0,08	69,92	0,05

**Tableau 6.2 :** Températures de fusion obtenues à partir de deux duplexes sur six *LightCyclers* différents (*Roche Applied Science*). Les mesures ont été obtenues en conditions de qPCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup> et 1,2 mM dNTP) à la concentration d'amorce de 0,6 µM.

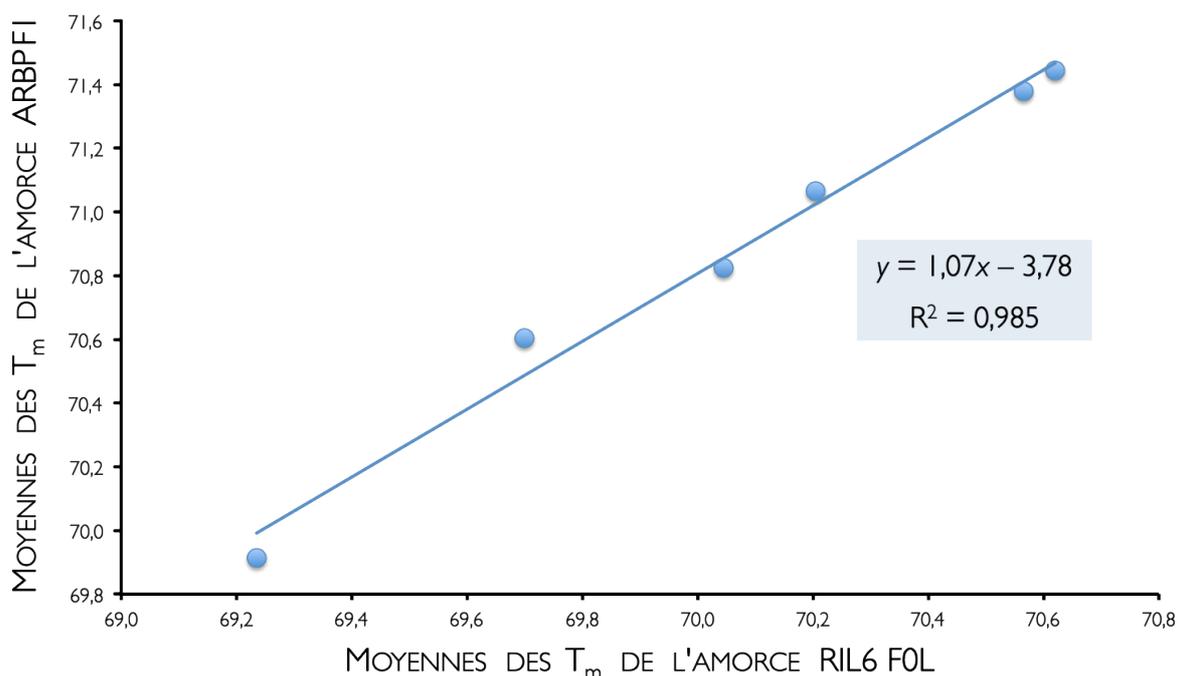


Figure 6.11 : Corrélation obtenue entre les températures de fusion de deux différents duplexes (ARBP FI et RIL6 FOL), sur six *LightCyclers* différents (Roche Applied Science).

### 6.2.2. Validation de la méthode de mesure directe du $T_m$ par courbe de fusion

D'après les résultats précédents, nous pouvons considérer que la mesure directe du  $T_m$  par fluorescence à la vitesse de  $0,05^\circ\text{C}/\text{sec}$  peut être considérée comme fiable et que la reproductibilité *intra-essai* d'environ  $0,1^\circ\text{C}$  est parfaitement adaptée aux conditions de qPCR. Il est nécessaire de valider la méthode de mesure directe du  $T_m$  par rapport à une méthode de référence, la nano Calorimétrie Différentielle à balayage (nano DSC). Pour cela, nous avons choisi d'utiliser un tampon phosphate IM  $\text{Na}^+$  qui est le milieu de référence pour mesurer le  $T_m$ . Cette concentration en sodium est importante car les cations présents dans un mélange réactionnel sont capables de stabiliser l'ADN lorsqu'il se trouve à l'état double brin (Schildkraut 1965). Dans la plupart des approches d'évaluation du  $T_m$  des acides nucléiques, cet effet stabilisant est généralement considéré à la concentration de IM  $\text{Na}^+$ . En effet, un changement de force ionique d'un milieu réactionnel va principalement influencer la constante d'association observée au cours de l'hybridation. En revanche, des constantes de dissociation identiques sont observées au cours de la dénaturation d'un duplexe, que ce soit à la concentration de  $0,05\text{M}$  ou

1M en  $\text{Na}^+$ . De ce fait, de bonnes approximations de valeur de  $T_m$  peuvent être obtenues par tout calcul effectué à 1M  $\text{Na}^+$  (Wetmur 1991).

### 6.2.2.1. Comparaison des valeurs de $T_m$ obtenues par les deux techniques avec leurs tampons respectifs

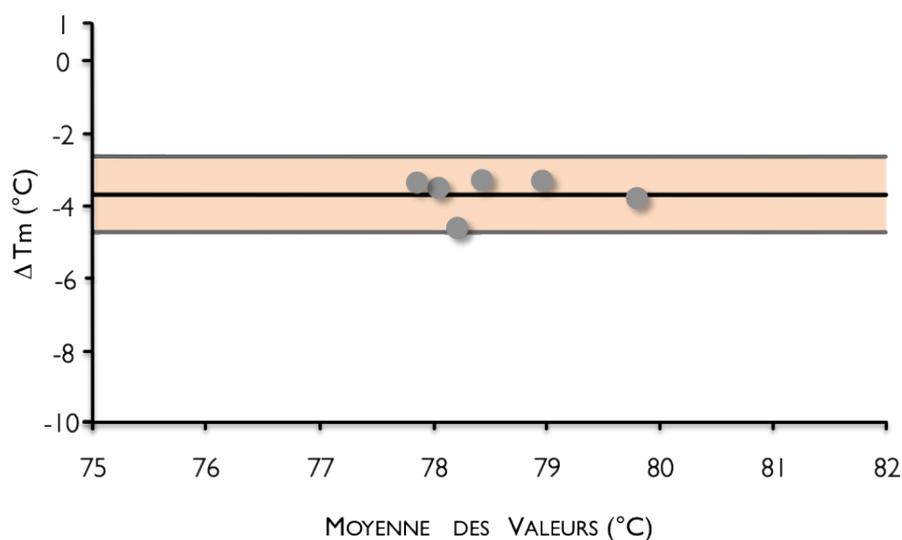
Nous avons comparé les  $T_m$  mesurés pour six oligonucléotides ADN non modifiés. Ces séquences sont respectivement analysées en tampon phosphate 1M  $\text{Na}^+$  par nano DSC et en tampon de PCR (32 mM  $\text{Na}^+$ , 5 mM  $\text{Mg}^{2+}$ , 1,2 mM dNTP) par courbe de fusion sur LightCycler. Les résultats obtenus sont décrits au niveau du **Tableau 6.3**.

On constate qu'il existe un bon agrément entre les deux méthodes (intervalle 95% de 2,3°C) mais avec un biais important de -3,8°C (**Figure 6.12**). Ce biais peut être attribué à l'effet séquence-dépendant des composants des différents tampons sur le  $T_m$  des duplexes. Il faut souligner ici que si la composition du tampon de calorimétrie est entièrement connue, celle du tampon de PCR ne l'est qu'approximativement (en  $\text{Na}^+$ ,  $\text{Mg}^{2+}$  et dNTP). Le tampon et les additifs présents dans ce milieu ne sont pas publics (et la société Roche refuse de les communiquer). Ce biais ne nous permet donc pas de valider notre méthode. Aussi, nous avons choisi de valider notre méthode en réalisant la mesure par nanoDSC en tampon de PCR (32 mM  $\text{Na}^+$ , 5 mM  $\text{Mg}^{2+}$ , 1,2 mM dNTP), tampon représentant les conditions dans lesquelles la mesure nous intéresse.

		$T_m$ nano DSC (°C)	$T_m$ fusion (°C)
Conditions de mesure		DSC-20	MC-20
[Oligonucléotide]		20 $\mu$ M	20 $\mu$ M
[Na <sup>+</sup> ]		1M	32 mM
[Mg <sup>2+</sup> ]		/	5 mM
[dNTP]		/	1.2 mM
Séquence 5'P-3'OH des oligonucléotides			
MARBP R1	TCCCACCTTGTCTCCAGTCTTTATC	79,80	76,31
MCXCL2 F1	TCCCTCAACGGAAGAACCAAAGAGA	80,60	75,86
RCYCA F2	TATCTGCACTGCCAAGACTGAGTG	81,70	77,94
RCYCA R1	CGAGGCACATCAGGTACGATCCA	80,10	76,80
RIL6 R0L	AACGGAACTCCAGAAGACCAGAG	80,70	77,25
RIL6 F0L	GCGATGATGCACTGTCAGAAAAC	79,60	76,15

**Tableau 6.3 :** Mesures de  $T_m$  obtenues par nano DSC en tampon phosphate 1M Na<sup>+</sup> et par courbe de fusion en tampon de PCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup>, 1,2 mM dNTP) pour les oligonucléotides non modifiés à la concentration de 20  $\mu$ M.

### COURBE DE FUSION (TAMPON PCR) – NANO DSC (1M NA<sup>+</sup>)



**Figure 6.12 :** Agrément effectué entre les mesures obtenues par courbe de fusion en tampon de PCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup>, 1,2 mM dNTP) et par nano DSC en tampon phosphate 1M Na<sup>+</sup> pour les oligonucléotides non modifiés à la concentration de 20  $\mu$ M. Les différences sont comparées par la méthode de *Bland-Altman*. L'abscisse et l'ordonnée représentent respectivement la moyenne et la différence des  $T_m$  mesurés par les deux méthodes. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

### 6.2.2.2. Comparaison des valeurs de $T_m$ obtenues en tampon de PCR avec les deux techniques

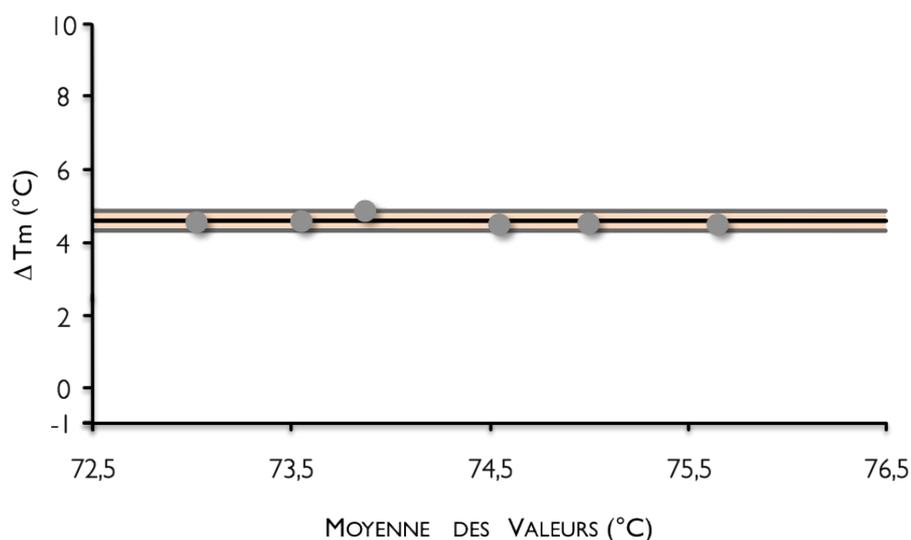
Les résultats de la comparaison des valeurs de  $T_m$  obtenues en tampon de PCR avec les deux techniques sont décrits au niveau du **Tableau 6.4**. Comme attendu, on observe un agrément plus important que précédemment. L'intervalle d'agrément 95% de 0,5°C est presque 5 fois plus faible que celui observé précédemment en tampon phosphate 1M Na<sup>+</sup>. En effet, l'identité de composition du milieu pour les deux mesures permet d'éliminer les effets séquences-spécifiques des composants du tampon. Cependant, on observe un biais inattendu très important de 4,6°C (**Figure 6.13**). Comme nous avons montré que la vitesse de mesure directe était compatible avec une mesure précise du  $T_m$ , cette différence pourrait être due à un additif contenu dans le tampon de PCR, capable de s'adsorber sur les parois du capillaire en verre du *LightCycler* et pas sur celles du serpentin en platine du calorimètre. En particulier, il est connu que la BSA entre dans la composition des tampons de PCR, servant à saturer les parois de verre ou de plastique des contenants utilisés dans les thermocycleurs. Il est à noter qu'en raison du très faible intervalle d'agrément 95% observé ici (0,5°C), le réactif responsable de cette différence n'a probablement pas d'effet séquence-dépendant sur le  $T_m$ .

Afin de vérifier cette hypothèse, nous avons réalisé une dernière analyse en comparant les valeurs de  $T_m$  obtenues par les deux techniques en tampon phosphate 1M Na<sup>+</sup>.

		$T_m$ nano DSC (°C)	$T_m$ fusion (°C)
Conditions de mesure		DSC-20	MC-20
[Oligonucléotide]		20 $\mu$ M	20 $\mu$ M
[Na <sup>+</sup> ]		32 mM	32 mM
[Mg <sup>2+</sup> ]		5 mM	5 mM
[dNTP]		1.2 mM	1.2 mM
Séquence 5'P-3'OH des oligonucléotides			
MARBP F1	GAAAATCTCCAGAGGCACCATTG	70,75	75,36
MARBP R1	TCCCACCTTGTCTCCAGTCTTTATC	71,45	76,35
MCXCL2 F1	TCCCTCAACGGAAGAACCAAAGAGA	71,25	75,83
RCYCA F2	TATCTGCACTGCCAAGACTGAGTG	73,40	77,91
RCYCA R1	CGAGGCACATCAGGTACGATCCA	72,30	76,78
RIL6 R0L	AACGGAACTCCAGAAGACCAGAG	72,75	77,31

**Tableau 6.4 :** Mesures de  $T_m$  obtenues par nano DSC et courbe de fusion en tampon de PCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup>, 1,2 mM dNTP), pour les oligonucléotides non modifiés à la concentration de 20  $\mu$ M.

### COURBE DE FUSION (TAMPON PCR) – NANO DSC (TAMPON PCR)

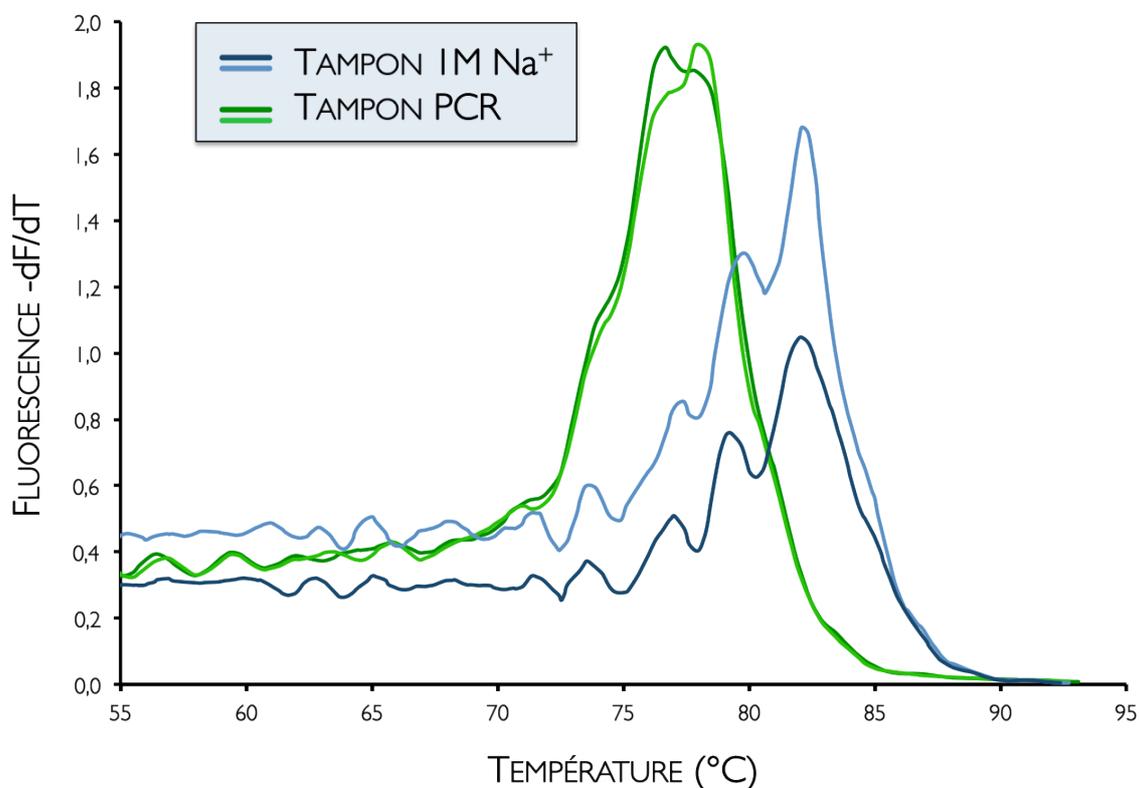


**Figure 6.13 :** Agrément effectué entre les mesures obtenues par courbe de fusion et nano DSC en tampon de PCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup>, 1,2 mM dNTP) pour les oligonucléotides non modifiés à la concentration de 20  $\mu$ M. Les différences sont comparées par la méthode de *Bland-Altman*. L'abscisse et l'ordonnée représentent respectivement la moyenne et la différence des  $T_m$  mesurés par les deux méthodes. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

### 6.2.2.3. Comparaison des valeurs de $T_m$ obtenues en tampon phosphate 1M $\text{Na}^+$ avec les deux techniques

Les résultats de la comparaison des valeurs de  $T_m$  obtenues en tampon phosphate 1M  $\text{Na}^+$  avec les deux techniques sont décrits dans le **Tableau 6.5**. On observe que les deux méthodes de mesure sont équivalentes avec un intervalle d'agrément 95% de 1,4°C et un biais de 0,2°C ; **Figure 6.14**). La composition simple du tampon phosphate 1M  $\text{Na}^+$ , dépourvu d'additifs capables de présenter des comportements différents dans les deux dispositifs de mesure, permet une comparaison non biaisée. L'intervalle d'agrément 95% mesuré en tampon phosphate 1M  $\text{Na}^+$  est environ 3 fois plus élevé que celui mesuré en tampon de PCR. Ce résultat est probablement dû à la différence de précision de la mesure directe du  $T_m$  en tampon de PCR et en tampon phosphate 1M  $\text{Na}^+$ . En effet, l'absence de produits stabilisants du tampon phosphate modifie la forme et la reproductibilité des **courbes de fusion** (**Figure 6.15**). Ces différences affectent l'efficacité de l'algorithme pour la détermination précise du  $T_m$ .

Ce résultat permet cependant de valider notre méthode de mesure directe du  $T_m$  pour des **oligonucléotides** non modifiés.



**Figure 6.15** : Courbes de fusion obtenues pour un même duplexe d'oligonucléotides ADN non modifiés à la concentration d'amorce de 20  $\mu\text{M}$  en tampon phosphate 1M  $\text{Na}^+$  et tampon de PCR (32 mM  $\text{Na}^+$ , 5 mM  $\text{Mg}^{2+}$ , 1,2 mM dNTP).

Conditions de mesure		$T_m$ nano DSC (°C)	$T_m$ fusion (°C)
[Oligonucléotide]		DSC-20	MC-20
[Na <sup>+</sup> ]		20 $\mu$ M	20 $\mu$ M
[Mg <sup>2+</sup> ]		1M	1M
[dNTP]		/	/
Séquence 5'P-3'OH des oligonucléotides			
MARBP R1	TCCCACCTTGTCTCCAGTCTTTATC	79,80	80,18
MCXCL2 F1	TCCCTCAACGGAAGAACCAAAGAGA	80,60	81,10
RCYCA F2	TATCTGCACTGCCAAGACTGAGTG	81,70	81,65
RCYCA R1	CGAGGCACATCAGGTACGATCCA	80,10	79,70
RIL6 R0L	AACGGAActCCAGAAGACCAGAG	80,70	81,01
Mut 3F	GCGATGATGCACTGGCTGTAAAC	81,30	81,70

Tableau 6.5 : Mesures de  $T_m$  obtenues par nano DSC et courbe de fusion en tampon phosphate 1M Na<sup>+</sup>, pour les oligonucléotides non modifiés à la concentration de 20  $\mu$ M.

### COURBE DE FUSION (1M Na<sup>+</sup>) – NANO DSC (1M Na<sup>+</sup>)

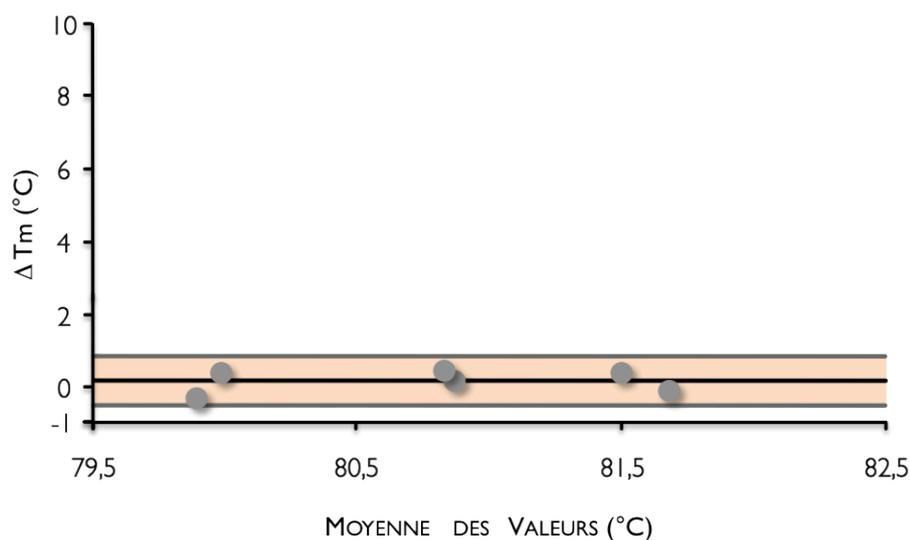


Figure 6.14 : Agrément effectué entre les mesures obtenues par courbe de fusion et nano DSC en tampon phosphate 1M Na<sup>+</sup> pour les oligonucléotides non modifiés à la concentration de 20  $\mu$ M. Les différences sont comparées par la méthode de *Bland-Altman*. L'abscisse et l'ordonnée représentent respectivement la moyenne et la différence des  $T_m$  mesurés par les deux méthodes. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

### 6.2.3. Détermination du $T_m$ par les méthodes prédictives en tampon phosphate 1M $\text{Na}^+$

#### 6.2.3.1. Comparaison de différents modèles prédictifs

À partir des travaux d'Alejandro PANJKOVICH (Panjkovich and Melo 2005) et sur la base de nos précédents résultats, nous avons cherché à évaluer la précision des différentes méthodes de prédiction du  $T_m$ . En comparant les valeurs prédites des différents modèles avec les mesures directes du  $T_m$  obtenues pour plusieurs oligonucléotides ADN par la méthode de référence, nous avons une grande variabilité des prédictions en fonction des méthodes utilisées (résultats présentés au **Tableau 6.6**). Afin d'illustrer au mieux cette observation, le résultat le plus favorable des comparaisons, basé sur la méthode prédictive de SUGIMOTO *et al.* (Sugimoto, Nakano et al. 1996), est présenté au niveau du **Tableau 6.7**. Le biais obtenu de  $13,0^\circ\text{C} \pm 8,7^\circ\text{C}$  montre l'inadaptation de ces méthodes prédictives pour la mesure précise du  $T_m$  (**Figure 6.16**).

	Biais ( $^\circ\text{C}$ )	Agrément 95%
SantaLucia, 1996	10,9	14,8
Sugimoto, 1996	13,0	8,7
Allawi, 1997	8,8	9,5
SantaLucia, 1998	9,0	9,5
Owczarzy, 2008	4,5	9,6

**Tableau 6.6 :** Comparaisons des  $T_m$  mesurés par la méthode de référence (nano DSC) avec les valeurs prédites de 5 différents modèles thermodynamiques. Les conditions utilisées pour réaliser la mesure sont celles qui ont permis d'établir la méthode prédictive (1M  $\text{Na}^+$ ).

En résumé, on constate qu'isolément, la plupart des méthodes prédictives ne permettent pas de déterminer précisément la température de fusion d'un oligonucléotide en 1M  $\text{Na}^+$ .

Conditions de mesure		$T_m$ Sugimoto <i>et al.</i>	$T_m$ nano DSC
		(°C)	(°C)
[Oligonucléotide]		Prédiction-20	DSC-20
[Na <sup>+</sup> ]		I M	I M
[Mg <sup>2+</sup> ]		/	/
[dNTP]		/	/

Séquence 5'P-3'OH des oligonucléotides			
MARBP FI	GAAAATCTCCAGAGGCACCATTG	85,95	78,50
MARBP RI	TCCCACCTTGTCTCCAGTCTTTATC	86,48	79,80
MCXCL2 FI	TCCCTCAACGGAAGAACCAAAGAGA	88,16	80,60
RCYCA RI	CGAGGCACATCAGGTACGATCCA	88,34	80,10
RIL6 F0L	GCGATGATGCACTGTCAGAAAAC	87,12	79,60
RIL6 R0L	AACGGA ACTCCAGAAGACCAGAG	87,66	80,70

Tableau 6.7 : Valeurs de  $T_m$  prédites par le modèle de SUGIMOTO *et al.* (1M Na<sup>+</sup>) et mesures de  $T_m$  obtenues par nano DSC (1M Na<sup>+</sup>) pour des oligonucléotides non modifiés à la concentration de 20  $\mu$ M.

### SUGIMOTO *et al.* (1M Na<sup>+</sup>) – NANO DSC (1M Na<sup>+</sup>)

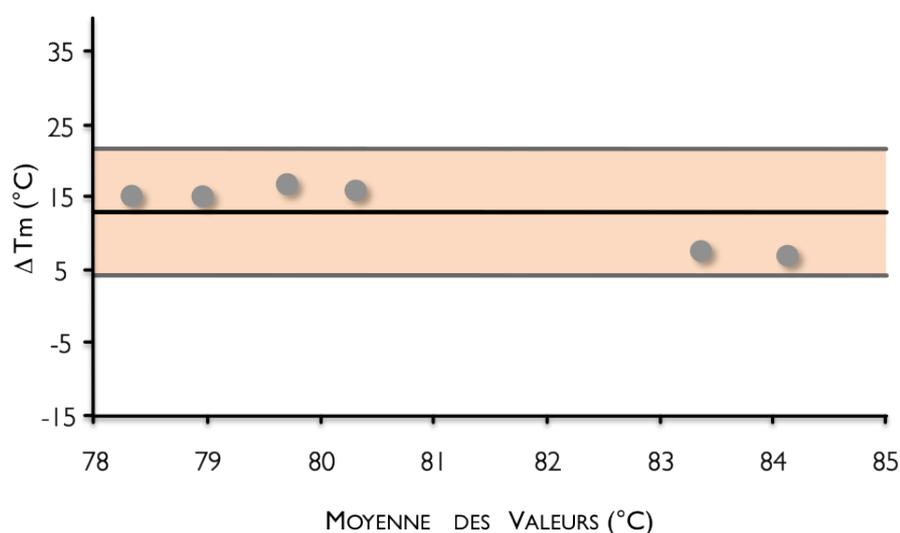


Figure 6.16 : Agrément effectué entre les valeurs de  $T_m$  prédites par la méthode de SUGIMOTO *et al.*, 1996 (1M Na<sup>+</sup>), et les mesures de  $T_m$  obtenues par nano DSC en tampon phosphate 1M Na<sup>+</sup> pour des oligonucléotides ADN à la concentration de 20  $\mu$ M. Les différences sont comparées par la méthode de Bland-Altman. L'abscisse et l'ordonnée représentent respectivement la moyenne et la différence des  $T_m$  mesurés par les deux méthodes. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

### 6.2.3.2. Modèle prédictif *dnaMATE*

L'utilisation isolée de ces méthodes prédictives apparaît donc très limitée. D'après Alejandro PANJKOVICH, elles ont une capacité prédictive du  $T_m$  importante en tampon phosphate, à condition de tenir compte de la séquence globale des amorces : taille et pourcentage GC (Panjkovich and Melo 2005).

Le logiciel *dnaMATE* (<http://protein.bio.puc.cl/cardex/servers/dnaMATE/tm-pred.html>) utilise plusieurs méthodes de prédiction du  $T_m$  (Breslauer, Frank et al. 1986 ; Sugimoto, Nakano et al. 1996; SantaLucia 1998) pondérées par le type de séquence. La comparaison des valeurs prédites par ce logiciel avec les valeurs mesurées par *nano DSC* en tampon phosphate 1M Na<sup>+</sup> est présentée au niveau du **Tableau 6.8**. D'après le biais de  $0,95^\circ\text{C} \pm 1,4^\circ\text{C}$  observé sur la **Figure 6.17**, le logiciel *dnaMATE* présente une excellente capacité à prédire la valeur du  $T_m$  des amorces que nous avons mesurée en tampon phosphate 1M Na<sup>+</sup>. Cependant, ce logiciel ne prend pas en compte les modifications séquence-dépendantes des cations présents dans les tampons de PCR.

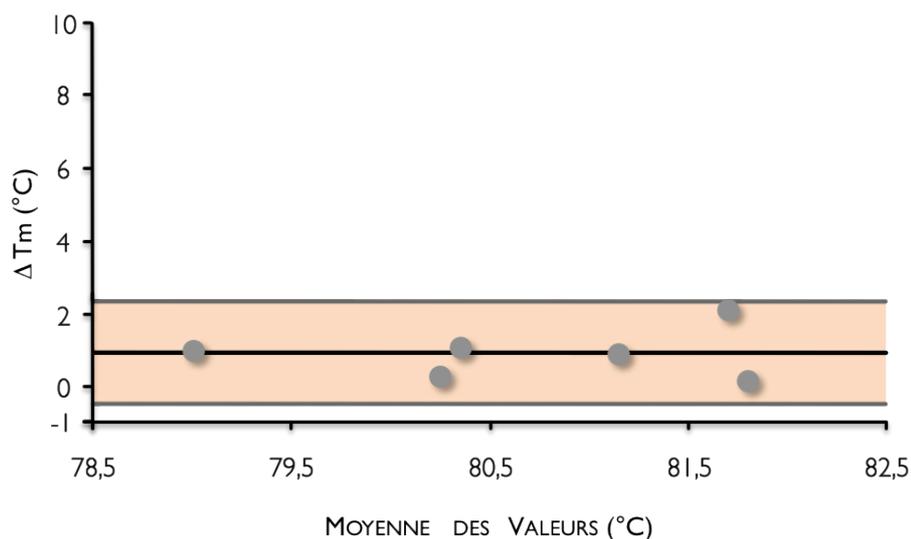
Le logiciel *dnaMATE* développé par Alejandro PANJKOVICH constitue une méthode de prédiction précise de la température de fusion d'un oligonucléotide en tampon phosphate 1M mais elle ne tient pas compte de l'effet des cations présents dans les tampons de PCR.

Conditions de mesure	$T_m$ Panjkovich et al. (°C)	$T_m$ nano DSC (°C)
[Oligonucléotide]	Prédiction-PANJ	DSC-20
[Na <sup>+</sup> ]	20 $\mu$ M	20 $\mu$ M
[Mg <sup>2+</sup> ]	1M	1M
[dNTP]	/	/

## Séquence 5'P-3'OH des oligonucléotides

MARBP F1	GAAAATCTCCAGAGGCACCATTG	79,50	78,50
MARBP R1	TCCCACCTTGTCTCCAGTCTTTATC	80,90	79,80
MCXCL2 F1	TCCCTCAACGGAAGAACCAAAGAGA	82,80	80,60
RCYCA F2	TATCTGCACTGCCAAGACTGAGTG	81,90	81,70
RCYCA R1	CGAGGCACATCAGGTACGATCCA	80,40	80,10
RIL6 R0L	AACGGAACTCCAGAAGACCAGAG	81,60	80,70

**Tableau 6.8 :** Valeurs de  $T_m$  prédites par PANJKOVICH *et al.* (1M Na<sup>+</sup>) et mesures de  $T_m$  obtenues par nano DSC en tampon phosphate (1M Na<sup>+</sup>) pour des oligonucléotides non modifiés à la concentration de 20  $\mu$ M.

NANO DSC (1M Na<sup>+</sup>) – PANJKOVICH *et al.* (1M Na<sup>+</sup>)

**Figure 6.17 :** Agrément effectué entre les valeurs de  $T_m$  prédites par PANJKOVICH *et al.*, 2005 et les mesures de  $T_m$  obtenues par nano DSC en tampon phosphate (1 M Na<sup>+</sup>). Les différences sont comparées par la méthode de *Bland-Altman*. L'abscisse et l'ordonnée représentent respectivement la moyenne et la différence des  $T_m$  mesurés par les deux méthodes. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

#### 6.2.4. Effet séquence-dépendant des cations sur le $T_m$ des amorces

Nous avons utilisé les deux séries de mesure réalisées en *nano DSC* avec les deux tampons (tampon phosphate 1M  $\text{Na}^+$  et tampon de PCR à 32 mM  $\text{Na}^+$ , 5 mM  $\text{Mg}^{2+}$  et 1,2 mM dNTP) pour illustrer l'effet séquence-dépendant des cations. Il est important de souligner que cet effet n'est pas pris en compte par le logiciel *dnaMATE*. Les résultats obtenus pour les sept amorces ADN utilisées sont résumés au niveau du **Tableau 6.9**.

Le biais observé d'environ  $-7^\circ\text{C}$  entre les deux séries de mesure est du à la différence de composition des tampons utilisés. L'intervalle d'agrément 95% de  $4,2^\circ\text{C}$  représente l'effet séquence-dépendant des cations par comparaison avec l'intervalle 8 fois plus faible observé précédemment. Il faut remarquer que ces séquences sont relativement proches avec  $23,7 \pm 0,9$  paires de bases et une richesse en bases GC de  $50\% \pm 3,3$ . Ces différences importantes concernent des oligonucléotides relativement proches au niveau de leurs séquences (**Figure 6.18**).

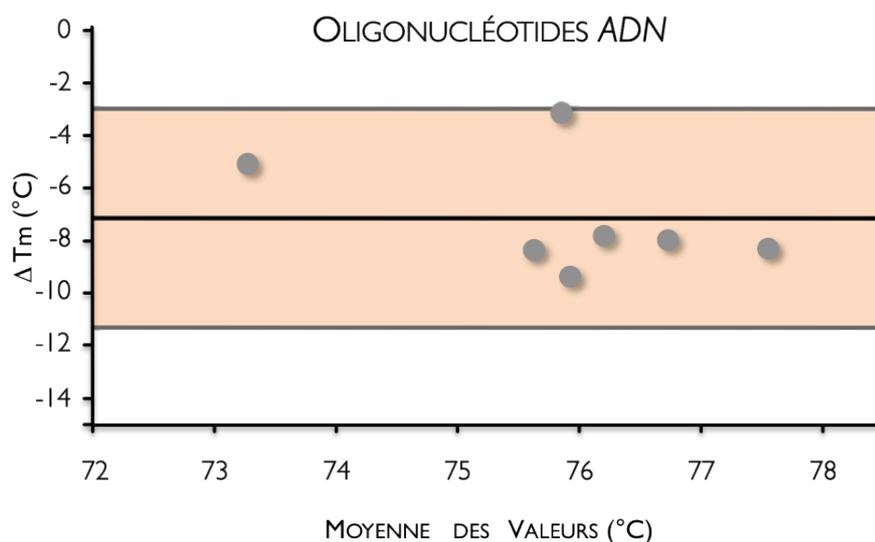
En résumé, la stabilité d'un duplexe dépend à la fois des cations et de certains composants présents dans le milieu réactionnel. Ces effets sont séquence-dépendants et soulèvent le problème de l'utilisation des méthodes prédictives qui souvent, ne tiennent pas compte de cette observation.

Conditions de mesure	$T_m$ nano DSC (°C)	$T_m$ nano DSC (°C)
	DSC-20 PCR	DSC-20 IMNa <sup>+</sup>
[Oligonucléotide]	20 $\mu$ M	20 $\mu$ M
[Na <sup>+</sup> ]	32 mM	1 M
[Mg <sup>2+</sup> ]	5 mM	/
[dNTP]	1,2 mM	/

## Séquence 5'P-3'OH des oligonucléotides

MARBP F1	GAAAATCTCCAGAGGCACCATG	70,75	78,50
MARBP R1	TCCCACCTTGTCTCCAGTCTTTATC	71,45	79,80
MCXCL2 F1	TCCCTCAACGGAAGAACCAAAGAGA	71,25	80,60
MCXCL2 R1	CGAGGCACATCAGGTACGATCCA	74,30	77,40
RIL6 F0L	GCGATGATGCACTGTCAGAAAAC	72,75	80,70
RCYCA F2	TATCTGCACTGCCAAGACTGAGTG	73,40	81,70
RCYCA R1	CGAGGCACATCAGGTACGATCCA	72,30	80,10

**Tableau 6.9** : Mesures de  $T_m$  obtenues par nano DSC en tampon de PCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup>, 1,2 mM dNTP) et tampon phosphate 1M Na<sup>+</sup> pour des oligonucléotides ADN à la concentration de 20  $\mu$ M.

NANO DSC (TAMPON DE PCR) - NANO DSC (1M Na<sup>+</sup>)

**Figure 6.18** : Agrément effectué entre les mesures de  $T_m$  obtenues par nano DSC en tampon PCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup>, 1,2 mM dNTP) et tampon phosphate 1M Na<sup>+</sup> (1 atmosphère), en comparant les différences par la méthode de *Bland-Altman*. L'abscisse et l'ordonnée représentent respectivement la moyenne et la différence des  $T_m$  mesurés par les deux méthodes. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

### 6.2.5. Application de la méthode aux amorces modifiées par LNA

En mesurant le  $T_m$  de nos différentes amorces modifiées et non modifiées par LNA avec la nano DSC en tampon phosphate 1M Na<sup>+</sup> et avec l'analyse par courbe de fusion en tampon de PCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup> et 1,2 mM dNTP), nous obtenons les résultats présentés au niveau des Tableaux 6.10 et 6.11. En raison de la différence entre les tampons, nous avons comparé ces séquences séparément de façon à limiter les effets séquence-dépendants des cations (Figures 6.19A - 6.19B). Les valeurs des biais observés, de -3,9°C et -3,2°C, pour les amorces sens et antisens respectivement, sont proches du biais observé pour les amorces ADN (-3,8°C). De même, les intervalles d'agrément 95% de 1,4°C et 2,3°C pour les amorces sens et antisens respectivement sont proches de l'intervalle d'agrément 95% observé pour les amorces ADN (2,3°C). Ces intervalles sont d'une part dus à la précision des méthodes et d'autre part du à l'effet séquence-dépendant (substitution-dépendant) joué par les cations et les composants du milieu. Comme ces séquences modifiées ont davantage d'homologies entre elles que les séquences ADN testées précédemment, un biais équivalent semble indiquer que les effets séquence-dépendants des cations sont plus importants sur les LNA que sur les bases non modifiées. Il est important de remarquer que les amorces ADN présentent un  $T_m$  supérieur au biais alors que les amorces modifiées par quatre substitutions de LNA présentent un  $T_m$  inférieur au biais. Cela semble indiquer que les LNA stabilisent davantage les duplexes en tampon phosphate qu'en milieu de PCR.

		$T_m$ nano DSC (°C)	$T_m$ fusion (°C)
Conditions de mesure		DSC-20	MC-20
[Oligonucléotide]		20 $\mu$ M	20 $\mu$ M
[Na <sup>+</sup> ]		1M	32 mM
[Mg <sup>2+</sup> ]		/	5 mM
[dNTP]		/	1.2 mM
<b>Set #1 : Séquence 5'P-3'OH des amorces sens</b>			
RIL6 F0L	GCGATGATGCACTGTCAGAAAAC	79,60	76,15
RIL6 F1L	GCGATGATGCACTGTCAGAAAAC	89,00	84,74
RIL6 F2L	GCGATGATGCACTGTCAGAAAAC	82,60	78,39
RIL6 F3L	GCGATGATGCACTGTCAGAAAAC	82,50	78,74
RIL6 F4L	GCGATGATGCACTGTCAGAAAAC	83,30	79,20
RIL6 F7L	GCGATGATGCACTGTCAGAAAAC	83,90	80,36

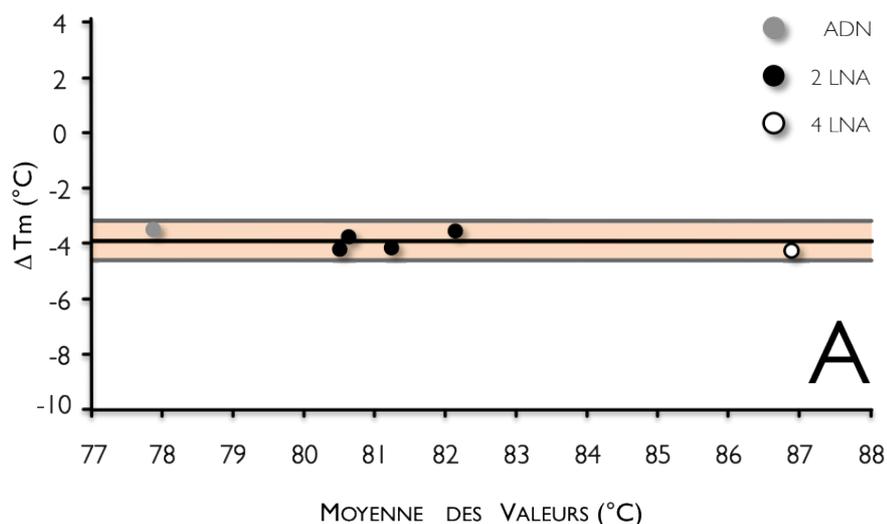
**Tableau 6.10 :** Mesures de  $T_m$  obtenues par nano DSC en tampon phosphate 1M Na<sup>+</sup> et par courbe de fusion en tampon de PCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup>, 1,2 mM dNTP) pour les oligonucléotides sens modifiés et non modifiés à la concentration de 20  $\mu$ M. Les substitutions par LNA sont notées en des grandes lettres.

		$T_m$ nano DSC (°C)	$T_m$ fusion (°C)
Conditions de mesure		DSC-20	MC-20
[Oligonucléotide]		20 $\mu$ M	20 $\mu$ M
[Na <sup>+</sup> ]		1M	32 mM
[Mg <sup>2+</sup> ]		/	5 mM
[dNTP]		/	1.2 mM
<b>Set #2 : Séquence 5'P-3'OH des amorces antisens</b>			
RIL6 R0L	AACGGAACTCCAGAAGACCAGAG	80,60	78,01
RIL6 R1L	AACGGAACTCCAGAAGACCAGAG	88,50	84,48
RIL6 R2L	AACGGAACTCCAGAAGACCAGAG	82,60	80,32
RIL6 R3L	AACGGAACTCCAGAAGACCAGAG	82,80	79,81
RIL6 R4L	AACGGAACTCCAGAAGACCAGAG	84,10	80,70
RIL6 R5L	AACGGAACTCCAGAAGACCAGAG	84,70	81,38
RIL6 R6L	AACGGAACTCCAGAAGACCAGAG	85,30	81,70
RIL6 R7L	AACGGAACTCCAGAAGACCAGAG	86,00	82,29

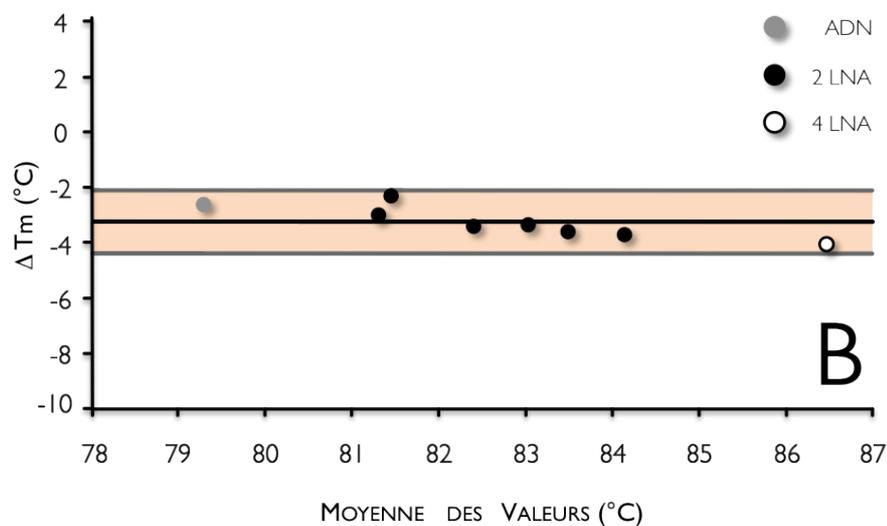
**Tableau 6.11 :** Mesures de  $T_m$  obtenues par nano DSC en tampon phosphate 1M Na<sup>+</sup> et par courbe de fusion en tampon de PCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup>, 1,2 mM dNTP) pour les oligonucléotides antisens modifiés et non modifiés à la concentration de 20  $\mu$ M. Les substitutions par LNA sont notées en des grandes lettres.

COURBE DE FUSION (TAMPON DE PCR) - NANO DSC (1M Na<sup>+</sup>)

## OLIGONUCLÉOTIDES SENS

COURBE DE FUSION (TAMPON DE PCR) - NANO DSC (1M Na<sup>+</sup>)

## OLIGONUCLÉOTIDES ANTI-SENS



**Figure 6.19** : Agréments effectués entre les mesures obtenues par courbe de fusion en tampon de PCR (32 mM Na<sup>+</sup>, 5 mM Mg<sup>2+</sup>, 1,2 mM dNTP) et par nano DSC en tampon phosphate 1M Na<sup>+</sup> pour les oligonucléotides modifiés et non modifiés par LNA sens (A) et antisens (B), à la concentration de 20  $\mu$ M. Les différences sont comparées par la méthode de *Bland-Altman*. L'abscisse et l'ordonnée représentent respectivement la moyenne et la différence des  $T_m$  mesurés par les deux méthodes. La ligne noire représente le biais, les lignes grises représentent l'intervalle d'agrément 95%.

### 6.3. Discussion – Méthode de mesure directe du $T_m$

Le contrôle de l'**hybridation** d'**oligonucléotides** spécifiques sur leurs cibles constitue le point crucial de la majorité des méthodes de quantification des acides nucléiques. Le paramètre de prédiction de l'**hybridation** le plus utilisé est celui de la **température de fusion**, température à laquelle la moitié des **amorces** sont hybridées sur la cible. Sa détermination précise est par conséquent un prérequis indispensable pour une optimisation efficace de la réaction d'amplification d'ADN. La plupart du temps, ce paramètre est estimé à partir de logiciels développés sur la base de différents modèles. Bien que ces derniers prennent désormais en compte de nombreux paramètres tels que la longueur de la séquence, sa composition en bases et bases modifiées, la concentration du milieu en cations ou bien encore les paramètres thermodynamiques *du plus proche voisin* (Borer, Dengler et al. 1974), ils reposent pour la plupart sur des équations établies à partir de séquences différentes. Ainsi, il existe des différences significatives entre les valeurs prédictives (Panjkovich and Melo 2005). De plus, la majorité de ces méthodes ne sont utilisables que dans un tampon phosphate à la concentration de 1M  $\text{Na}^+$ , sensiblement différent des tampons utilisés en **PCR**. Les différences importantes existant entre les méthodes des logiciels proposant des corrections du  $T_m$  en fonction de la composition ionique du milieu n'incitent pas à les utiliser. De plus, elles tiennent uniquement compte de la composition en cations, jamais des nombreux additifs (le plus souvent inconnus en raison de la protection des secrets industriels mais capables de modifier significativement la valeur du  $T_m$ ). D'autre part, peu de logiciels prennent en compte l'effet de substitutions de bases par *Locked Nucleic Acids* (**Chapitre 5**).

En comparant les valeurs de  $T_m$  obtenues par nano **Calorimétrie Différentielle** en tampon phosphate 1M  $\text{Na}^+$ , nous avons vérifié que des méthodes pondérées proposées par Alejandro PANJKOVICH permettaient une prédiction de bonne qualité de la **température de fusion** dans ces conditions-là. Cependant, la comparaison des  $T_m$  mesurés par nano **Calorimétrie Différentielle** en conditions de **PCR** (32 mM  $\text{Na}^+$ , 5 mM  $\text{Mg}^{2+}$ , 1,2 mM dNTP) d'une part et en tampon phosphate 1M  $\text{Na}^+$  d'autre part, montre bien qu'il existe un effet séquence-dépendant non négligeable des cations et de certains composants du tampon. Cet effet, déjà observé auparavant (Mikulecky and Feig 2006), met en évidence l'importance du rôle joué par ces ions durant l'**hybridation** et la **dénaturation** de **séquences complémentaires** (Schildkraut 1965).

Pour résoudre ce problème nous proposons une méthode permettant de mesurer précisément et simplement le  $T_m$  des amorces dans les conditions exactes de la PCR à partir de l'analyse par courbe de fusion sur thermocycleur. Cette méthode consiste à mesurer le  $T_m$  du duplexe constitué de l'amorce hybridée avec sa séquence complémentaire. Dans un premier temps, nous avons caractérisé la précision et la reproductibilité de cette méthode sur thermocycleur *LightCycler*. Les résultats montrent notamment que la méthode est capable de détecter des variations de  $T_m$  de moins de 0,2 °C, ce qui constitue une sensibilité largement suffisante pour les besoins de la qPCR.

L'utilisation de la nano Calorimétrie Différentielle nous a permis de valider cette méthode et de souligner l'effet non négligeable des composants du tampon (4°C pour probablement un seul constituant) qui, à notre connaissance, ne sont pratiquement jamais pris en compte.

Cette méthode peut être utilisée pour optimiser les performances de la qPCR, particulièrement dans le cadre du diagnostic. On peut imaginer de sélectionner des paires d'amorces ayant non seulement des  $T_m$  équivalents mais également des  $T_m$  modifiés de façon analogue par des inhibiteurs de PCR. De telles amorces amélioreraient la qualité de la quantification en diminuant l'impact des variations inter-échantillons. Bien que le  $T_m$  prédit soit généralement la valeur utilisée pour le choix des amorces de PCR, le pourcentage d'amorce hybridée à la température d'hybridation semble un paramètre plus pertinent (SantaLucia 2007). La détermination de ce pourcentage nécessite la connaissance des paramètres thermodynamiques  $\Delta H$  et  $\Delta S$  de cette séquence. En théorie, la connaissance précise du  $T_m$  en fonction de la concentration d'amorce permet de déterminer ces valeurs.

## 6.4. Résultats expérimentaux – Paramètres thermodynamiques de l'hybridation

Au regard de la précision et de la reproductibilité obtenues avec la méthode de mesure directe du  $T_m$ , nous avons pu utiliser l'équation (6.11) afin de déterminer expérimentalement les paramètres thermodynamique  $\Delta H$  et  $\Delta G$  de nos différentes amorces :

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln([P]_{eq})} \quad (6.11)$$

où  $[P]_{eq}$  représente la concentration d'amorce à l'équilibre

En inversant cette équation, on obtient :

$$\frac{1}{T_m} = a \ln([P]_{eq}) + b \quad (6.11')$$

$$\text{avec : } a = \frac{R}{\Delta H^\circ} \text{ et } b = \frac{\Delta S^\circ}{\Delta H^\circ}$$

L'étude des paramètres de cette droite permet de déterminer les  $\Delta H$  et  $\Delta S$  en mesurant la température de fusion de chaque duplexe d'oligonucléotide à plusieurs concentrations d'amorce en modifiant les concentrations  $[P]_{in}$  et  $[T]_{in}$ .

Pour mémoire, si  $[P]_{in} > [T]_{in}$ ,  $[P]_{eq} = [P]_{in} - [T]_{in}/2$  (6.14).

L'écart-type de la pente et de l'ordonnée à l'origine est donné selon les formules :

$$\sqrt{Var(a)} = \frac{s_y \sqrt{1-r^2}}{s_x \sqrt{n-2}} \quad \sqrt{Var(b)} = \frac{s_y \sqrt{(1-r^2) \sum x_i^2}}{s_x \sqrt{n(n-2)}}$$

Dans les conditions de la PCR où  $[T]_{in}$  est négligeable devant  $[P]_{in}$ ,  $[P]_{eq} \approx [P]_{in}$ . Ainsi, la température  $T_f$  à laquelle on observe à l'équilibre un taux  $f$  de cible ADN hybridée, est donnée par l'équation (6.9') :

$$T_f = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln\left(\frac{[P]_{in}(1-f)}{f}\right)}$$

### 6.4.1. Précision de la mesure des paramètres thermodynamiques

Dans un premier temps, nous avons dû déterminer les conditions expérimentales permettant une mesure précise des paramètres. Les premiers essais ont été réalisés avec des dilutions en séries contenant des quantités égales des deux amorces. La précision était insuffisante avec des CV de l'ordre de 15% pour les valeurs de (a) et (b). Afin d'augmenter la précision de la mesure, nous avons essayé de prédire la précision par propagation de l'erreur de mesure du  $T_m$  sur plusieurs types de gamme en se basant sur les principes suivants :

- Avoir une quantité constante d'amorces hybridées à la température de fusion c'est-à-dire travailler avec deux concentrations différentes, la concentration d'une des deux amorces restant constante de façon à optimiser l'efficacité de l'algorithme.
- Une gamme de concentration la plus étendue possible, pour augmenter la précision de la mesure, et comprenant des concentrations les plus proches possibles de celles rencontrées en PCR.

Les conditions optimales que nous avons déterminées correspondent à une concentration initiale d'amorce cible constante de 0,2  $\mu\text{M}$ . Nous avons déterminé une gamme adaptée à ces conditions, elle est définie de la manière suivante :

Concentration de l'amorce à tester ( $\mu\text{M}$ )	Concentration de la séquence complémentaire ( $\mu\text{M}$ )	Concentration réelle de l'amorce ( $\mu\text{M}$ )
0,2	0,2	0,1
0,6	0,2	0,5
1,8	0,2	1,7
5,4	0,2	5,3

**Tableau 6.12 :** Exemple d'une gamme de concentrations d'amorce réalisée pour la détermination des paramètres thermodynamiques  $\Delta H$  et  $\Delta S$  d'une séquence. En pratique, chaque point est réalisé en duplicatas.

L'utilisation de cette gamme permet une précision de mesure qui apparaît suffisante. Sur 9 amorces différentes, on observe des coefficients de variation moyens de  $2,68\% \pm 0,51\%$  pour de la pente (a) et pour l'ordonnée à l'origine (b) de  $0,11\% \pm 0,02\%$ . Une précision supérieure à été observée pour des gammes plus étendues (s'écartant des conditions de PCR) en supprimant le point le moins concentré (0,2  $\mu\text{M}$  / 0,2  $\mu\text{M}$ ).

Afin de vérifier l'exactitude de cette méthode, nous l'avons confrontée à une méthode permettant de déterminer par fluorescence les paramètres thermodynamiques d'amorces ADN (You, Tataurov et al. 2011). Cette méthode consiste à étudier l'hybridation de deux amorces complémentaires modifiées, portant respectivement en 5'P et en 3'OH un fluorophore et un quencheur. Nous avons mesuré les paramètres  $\Delta H$  et  $\Delta S$  de la séquence ADN utilisée par YOU et *al.* dans des conditions similaires (tampon phosphate 1M  $\text{Na}^+$ , 10 mM  $\text{Na}_2\text{HPO}_4$ , 1 mM  $\text{Na}_2\text{EDTA}$ , pH 7). Le **Tableau 6.15** montre une quasi-identité des mesures réalisées avec notre méthode et la mesure réalisée par YOU et *al.* avec la combinaison optimale de fluorophores modifiant le moins la mesure.

Séquence de l'amorce	longueur	Paramètres	YOU et <i>al.</i>	Mesure directe
5'-CGTACATGC-3'	9 nucl.	$\Delta H$ (cal.mol <sup>-1</sup> )	93 300	93 302
		$\Delta S$ (cal.mol <sup>-1</sup> .K <sup>-1</sup> )	-247	-259

**Tableau 6.15** : Paramètres thermodynamiques  $\Delta H$  et  $\Delta S$  obtenus par la méthode de YOU et *al.* et par la méthode de mesure directe.

Pour cette mesure, nous avons obtenu un coefficient de variation moyen de la pente (*a*) de **4,84%** et un coefficient de variation moyen de l'ordonnée à l'origine (*b*) de **0,50%**. Ces valeurs sont supérieures à celles décrites précédemment en raison d'une part d'une diminution de la précision préalablement observée lors des mesures en tampon phosphate (**Figure 6.15**) et probablement d'autre part de la courte longueur de la séquence analysée (9 pb par rapport à 23 pb en moyenne dans notre étude).

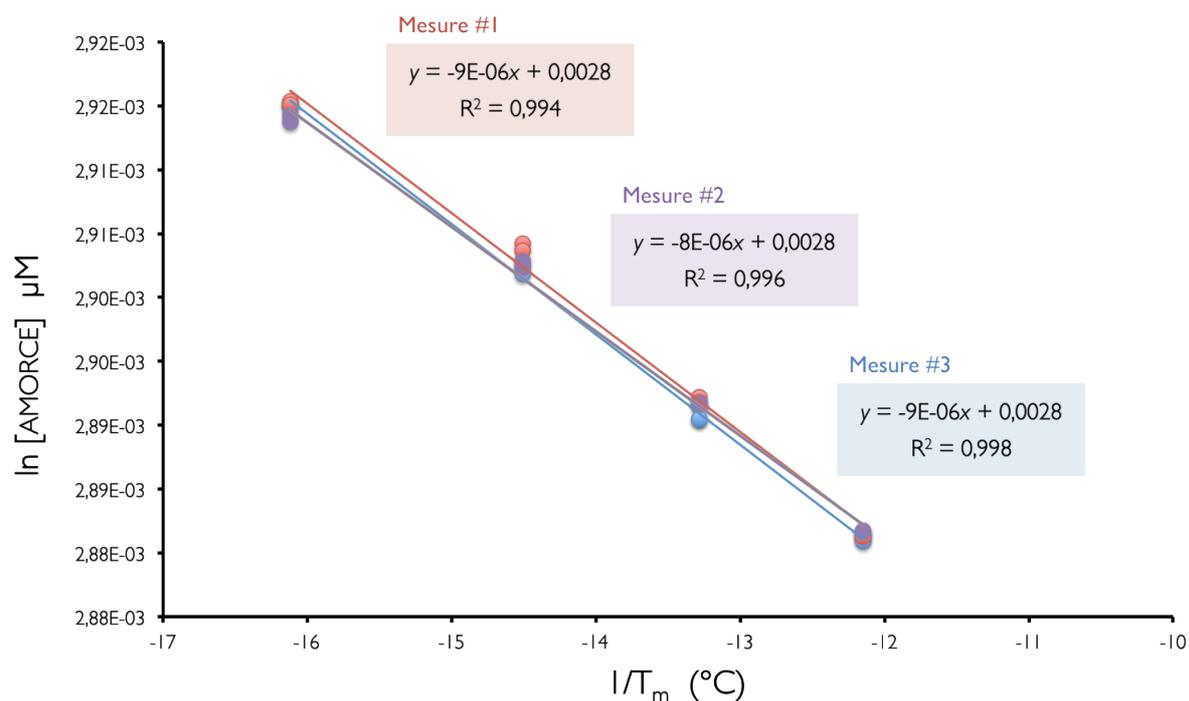
#### 6.4.2. Reproductibilité de la mesure des paramètres thermodynamiques

Dans un deuxième temps, nous avons testé la reproductibilité de la méthode en réalisant trois mesures indépendantes d'une même séquence sur un même thermocycleur. La reproductibilité de cette méthode ( $CV < 3\%$ ) est excellente en comparaison à l'imprécision classiquement rencontrée dans ce type de mesure, de l'ordre de  $\approx 11\%$  (You, Tataurov et al. 2011), pour des mesures cependant réalisées en tampon phosphate 1M  $\text{Na}^+$  stabilisant beaucoup moins l'hybridation. Les résultats obtenus peuvent être résumés sous la forme du **Tableau 6.14** suivant :

	$\Delta H^\circ$	$\Delta S^\circ$	CV pente (a)	CV ordonnée à l'origine (b)
Mesure #1	-231	-642	-2,07%	0,09%
Mesure #2	-239	-666	-3,36%	0,14%
Mesure #3	-243	-676	-2,47%	0,11%
Moyenne	-238	-662		
CV	2,52%	2,63%		

**Tableau 6.14** : Mesures des variations d'enthalpie  $\Delta H$  (en kcal.mol<sup>-1</sup>) et d'entropie  $\Delta S$  (cal.mol<sup>-1</sup>.K<sup>-1</sup>) au cours de trois mesures réalisées de manière indépendante sur un même thermocycleur.

La reproductibilité de la méthode peut être illustrée par les corrélations linéaires qui ont permis de calculer ces paramètres (représentées en **Figure 6.20**) et présentant toutes des coefficients supérieurs à **0,990**.

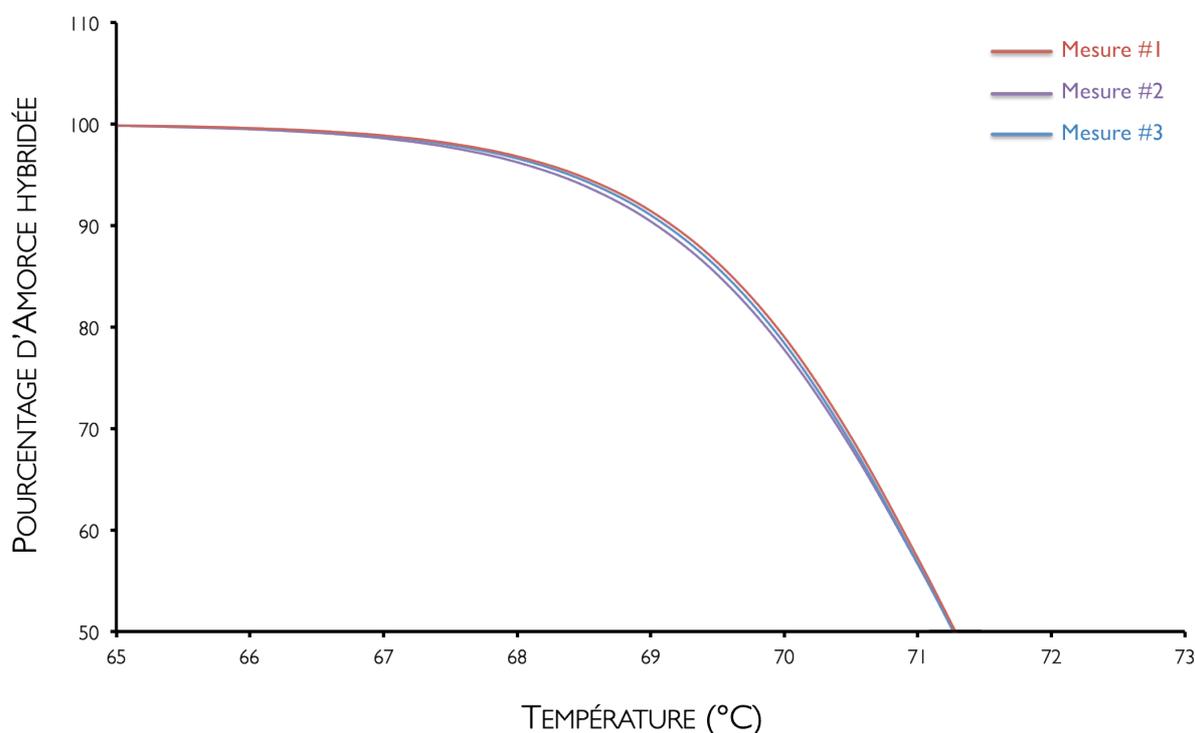


**Figure 6.20** : Corrélations obtenues pour trois mesures indépendantes réalisées en duplicates sur un même thermocycleur.

Les paramètres thermodynamiques ainsi mesurés permettent de prédire le pourcentage de l'amorce hybridée en fonction de la température dans des conditions de PCR (concentration de la cible négligeable par rapport à la concentration d'amorce utilisée ; Figure 6.18) en utilisant l'équation (6.9') :

$$T_f = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln\left(\frac{[P] \cdot (1-f)}{f}\right)}$$

où :  $f$  représente la fraction d'amorce hybridée à la température  $T_f$  et  $[P]$  représente la concentration initiale d'amorces.



**Figure 6.18** : Estimation du pourcentage d'amorce hybridée pour une concentration initiale d'amorce de 0,4  $\mu\text{M}$  dans des conditions de PCR pour trois mesures réalisées de manière indépendante sur un même thermocycleur.

Cette représentation graphique est peu utilisable en pratique : il est plus simple d'indiquer le nombre de degrés séparant le  $T_m$  d'un pourcentage donné d'amorce hybridée. Les résultats montrés au niveau du **Tableau 6.15** montrent l'intérêt de la prédiction thermodynamique par rapport au  $T_m$  : deux amorces présentant un  $T_m$  proche s'hybrident à des vitesses sensiblement différentes.

	Paramètres thermodynamiques	$T_m$	$T_{90\%} - T_m$	$T_{95\%} - T_m$	$T_{99\%} - T_m$
MCXCL2 F1	$\Delta H = -154$ $\Delta S = -417$	72,3	3,3	4,5	6,9
RCYCA F2	$\Delta H = -206$ $\Delta S = -569$	71,9	2,5	3,3	5,2

**Tableau 6.15** : Mesures des paramètres thermodynamiques  $\Delta H$  (en kcal.mol<sup>-1</sup>) et  $\Delta S$  (cal.mol<sup>-1</sup>.K<sup>-1</sup>) observées pour deux différentes amorces dans des conditions de PCR : tampon Roche avec une concentration initiale d'amorce de 0,4  $\mu$ M.

## 6.4. Discussion – Paramètres thermodynamiques de l'hybridation

Le *design* d'*amorces* de PCR est le plus souvent basé sur la nécessité d'utiliser des *amorces* de  $T_m$  identiques, sinon les plus proches possibles, afin d'obtenir une réaction d'amplification d'ADN efficace. Ce principe, qui est dicté par les données le plus souvent disponibles, doit être nuancé. En effet, le  $T_m$  est la température à laquelle la moitié des molécules de même séquence nucléotidique se retrouve spécifiquement liée à leur *séquence complémentaire*. Or, l'*hybridation* des *amorces* au cours de la PCR est réalisée à une *température d'hybridation* inférieure au  $T_m$  des *amorces* (traditionnellement 5 à 10°C). Le paramètre pertinent est donc en dessous du  $T_m$  des *oligonucléotides* utilisés. Le pourcentage de l'*amorce* hybridée à la *température d'hybridation* semble plus intéressant. Bien que ce paramètre dépende du  $T_m$ , des *amorces* ayant un  $T_m$  identiques peuvent avoir un pourcentage d'*amorce* hybridée différent à l'équilibre, 5°C en dessous du  $T_m$ .

Une démarche pertinente est donc la détermination du pourcentage de l'*amorce* hybridée, obtenu à partir des paramètres thermodynamiques  $\Delta H$  et  $\Delta S$  de la séquence (SantaLucia 2007). Le modèle *du plus proche voisin* constitue la méthode la plus utilisée pour la prédiction de ces deux valeurs (SantaLucia 1998; SantaLucia and Hicks 2004) mais elles peuvent aussi être calculées à partir de la *courbe de fusion* de l'*oligonucléotide* (Mergny and Lacroix 2003 ; Schroeder and Turner 2009). Cependant, ces déterminations qui nécessitent le plus souvent des méthodes chères et couteuses en temps (*Differential Scanning Calorimetry* (DSC), *Isothermal Titration Calorimetry* (ITC), *Surface Plasmon Resonance* (SPR-Biacore), ...) sont le plus souvent réalisées dans des milieux différents de ceux utilisés en PCR alors que les valeurs mesurées dépendent du tampon. YOU *et al.* ont récemment proposé une méthode plus simple utilisant un thermocycleur de PCR en temps réel : l'*hybridation* du *duplexe* est quantifiée par l'extinction de la fluorescence d'un *fluorophore* lié en 5'P de l'*oligonucléotide* par un *quencheur* fixé en 3'OH de sa *séquence complémentaire* (You, Tataurov et al. 2011).

Dans cette étude, nous proposons une méthode de mesure précise et rapide des paramètres thermodynamiques de l'*amorce*, directement dans le milieu de PCR en utilisant l'*amorce* de PCR non modifiée. Cette méthode consiste à déterminer le  $T_m$  d'un *duplexe* constitué de l'*amorce* hybridée à sa *séquence complémentaire* à différentes concentrations avec un appareil de PCR en temps réel, directement dans le milieu de PCR. En mesurant la

température de fusion d'un oligonucléotide en fonction des concentrations d'amorces initiales  $[P]_{in}$  et  $[T]_{in}$ , nous avons démontré qu'il est possible de déterminer les variations d'enthalpie ( $\Delta H$ ) et d'entropie ( $\Delta S$ ). Il faut souligner que cette méthode, comme les autres méthodes existantes, ne tient compte que de l'hybridation spécifique de l'amorce sur la cible et ne permet pas de prendre en compte la modification de la cinétique d'hybridation spécifique par l'hybridation non spécifique.

Mesurés dans le milieu de PCR, la connaissance de ces paramètres permet de déterminer le pourcentage de l'amorce hybridée à la température d'hybridation. Il devrait permettre d'optimiser des réactions de PCR réalisées en routine, plus particulièrement dans le cadre du diagnostic. Mais surtout, on peut envisager de comparer le comportement que peut présenter une paire d'amorces vis-à-vis des inhibiteurs présents dans le milieu. La sélection d'amorces, dont l'hybridation est affectée de façon similaire par des inhibiteurs, devrait permettre de diminuer la variabilité inter-échantillon, particulièrement en diagnostic. Elle devrait également permettre de mesurer l'impact d'une mutation ponctuelle sur l'hybridation.

Notre méthode se distingue des autres techniques de détermination des paramètres thermodynamiques par sa simplicité d'utilisation, la possibilité de contrôler *a posteriori* la précision de la mesure et surtout la capacité d'utiliser les amorces non modifiées, exactement dans les conditions de PCR. L'utilisation d'algorithmes plus performants devrait améliorer le domaine d'utilisation de cette méthode (concentration de magnésium, solutions salines concentrées, ...).

# Perspectives

---

## ❖ Diminution de la variabilité inter-échantillons

Au cours de cette étude, nous avons montré l'intérêt de l'utilisation d'une **transcription inverse** à volume constant d'extrait d'ARN dans le cadre de la **standardisation** de la RT-qPCR. Il nous reste désormais à définir l'utilisation pratique de cette méthode qui peut soit être utilisée directement, soit servir à optimiser des protocoles d'extraction, minimisant les différences inter-échantillons.

Nous avons aujourd'hui un recul d'une dizaine d'expérimentations ayant utilisé la **transcription inverse** à volume constant d'extrait d'ARN. Dans la majorité des cas, la méthode de CV-RT donne d'excellents résultats que nous avons pu estimer par les faibles valeurs des facteurs de **normalisation**. Cependant, lorsqu'on utilise des échantillons peu homogènes (peau, tissu adipeux, colon ...), on peut parfois observer un ratio entre les ARN les plus concentrés et les ARN les moins concentrés de l'ordre de 10. Nous avons vérifié que la réalisation d'une **transcription inverse** à quantité constante d'ARN pour ces échantillons biaise significativement le résultat. Pour une expérimentation réalisée à partir d'échantillons de colon de rat, pour lesquels la qualité de l'ARN a été contrôlée par micro-fluidique, puis reverse transcrits à quantité constante, il n'a pas été possible de normaliser les échantillons les moins concentrés avec les mêmes gènes que les échantillons les plus concentrés. Cependant, la réalisation d'une **transcription inverse** à volume constant d'extrait d'ARN entraîne également une perte de sensibilité importante pour les échantillons les moins concentrés. Un compromis difficile doit donc être trouvé pour traiter ces cas particuliers. Comme l'hétérogénéité des rendements d'extraction dépend probablement de l'hétérogénéité des échantillons (pourcentage relatif de cellules et de constituants non cellulaires dans ces échantillons), un travail sur la **standardisation** de l'extraction ne semble pas être la solution. Par contre, la recherche de protocoles d'extraction minimisant les différences inter-échantillons peut être réalisée pour ce type de tissu. Ainsi, nous avons utilisé cette démarche avec succès pour des échantillons de muscle de rat qui différaient par l'âge et surtout par l'absence ou la présence de régénération du tissu musculaire. L'optimisation de l'extraction (en particulier en testant plusieurs enzymes protéolytiques ou des étapes de pré-extraction organique) a permis d'homogénéiser ces échantillons.

## ❖ Optimisation de la spécificité d'hybridation des amorces de PCR par LNA

Le but de cette étude était, dans un premier temps de démontrer que la modification d'un **oligonucléotide** par incorporations de bases modifiées dans la **partie 5'P** de la séquence permettait d'augmenter la spécificité des ces **amorces** et, dans un deuxième temps, de définir des règles optimales de *design* pour ce type de modification. Les expérimentations réalisées au cours de cette étude ont permis de valider cette hypothèse et ont permis d'établir une ébauche de règles de *design*. Ces résultats doivent cependant être validés par des analyses supplémentaires sur un nombre de séquences plus important.

Cependant, nous avons obtenu deux résultats qui vont pouvoir servir de base à de futures expérimentations. Premièrement, il semble y avoir une contradiction entre l'augmentation avérée de la spécificité des **amorces** modifiées par 4 LNA, observée à partir d'échantillons biologiques, et leur faible spécificité au niveau de notre modèle d'évaluation de spécificité. Deuxièmement, nous avons observé qu'il existe une différence significative de spécificité entre les deux **amorces** ADN non modifiées. Ce point est particulièrement intéressant et la compréhension de cette différence pourrait avantageusement remplacer l'étude des LNA. Cette différence est probablement due à une combinaison entre la séquence non spécifique et la séquence de l'**amorce**. Il faudrait donc vérifier cette hypothèse sur d'autres séquences mutées. On sait qu'à la différence de l'ARN, certains **mésappariements** sont capables de stabiliser des **duplexes** ADN (Allawi and SantaLucia 1997). Il pourrait donc être intéressant de mettre au point une méthode de *design* d'**amorces** minimisant le bénéfice thermodynamique de l'**hybridation** non spécifique au regard des tables fournies par Hatim ALLAWI, en privilégiant par exemple les bases C dans la **partie 3'OH** de l'**amorce** ...

## ❖ Détermination de la température de fusion et des paramètres thermodynamiques d'un oligonucléotide

Nous avons montré que la mesure directe du  $T_m$  par **courbe de fusion** réalisée directement dans le milieu de **PCR** représente une méthode fiable et reproductible. Cependant, nous avons observé que dans certaines conditions (variations de la concentration en magnésium, intercalant saturant ou utilisation de tampon phosphate) et pour certaines séquences, l'algorithme du logiciel de nos machines était inadapté. L'utilisation d'algorithmes spécialement conçus pour l'analyse de **courbes de fusion à haute résolution (HRM)** ne permet pas de résoudre ce problème car ils ne sont pas adaptés à une mesure simple du  $T_m$  (fonctionnent par

comparaison). Il pourrait donc être intéressant de développer des algorithmes plus performants. Cela semble possible car nous avons observé, en particulier lors de la mesure des paramètres thermodynamiques, que le pic de fusion était formé en général de deux composants de proportions variables selon les échantillons. Cette situation induit l'algorithme en erreur qui ne va pas toujours choisir le même composant en fonction de la concentration des échantillons. La détermination manuelle au moyen d'une simple règle permet d'obtenir d'excellents résultats. Il semble donc possible d'utiliser un algorithme guidé par l'utilisateur.

Maintenant que l'on dispose d'un outil capable de prédire le pourcentage d'**amorce** hybridée à une température donnée en fonction des conditions de l'étude, il semble intéressant d'étudier l'effet des inhibiteurs de **PCR** sur les **amorces** qui, pour la plupart, interagissent au niveau de l'**hybridation**. Cette méthode devrait permettre de choisir des **amorces** ayant le même comportement face aux inhibiteurs. Utilisées en diagnostic, de telles **amorces** permettraient de diminuer la variabilité inter-échantillons tout en conservant une efficacité de **PCR** convenable.

# Glossaire

---

**ACIDES NUCLÉIQUES PEPTIDIQUES** - Molécules aux bases similaires à l'ADN ou à l'ARN et synthétisées artificiellement. Les acides nucléiques peptidiques présentent cependant un squelette dépourvu de groupements phosphate, composé d'une répétition d'unités N-(2-aminoethyl)-glycine reliées par une liaison peptidique. Ils se notent donc comme des peptides, avec l'extrémité N-terminale en première position (gauche) et l'extrémité C-terminale en dernière position (droite). Non chargés, ils sont capables de former des duplexes PNA/ADN de stabilité plus importante que des duplexes ADN/ADN classiques en raison de l'absence de répulsion électrostatique.

**Anglais : Peptide Nucleic Acid, PNA.**

**ADN POLYMÉRISE** - Polymérase qui synthétise l'ADN. Voir POLYMÉRISE.

**Anglais : DNA polymerase, DNA Dependent DNA polymerase.**

**ADN complémentaire (ADNc)** - Molécule ADN qui a été initialement copiée à partir d'une molécule ARN par reverse transcription. Le terme "ADNc" est couramment utilisé pour décrire l'ADN double brin qui est issu d'une molécule d'ARN simple brin, même si un seul brin de l'ADN est vraiment complémentaire de l'ARN.

**Anglais : cDNA.**

**AMORCE (PRIMER) SENS** - Oligonucléotide complémentaire d'une région spécifique au niveau d'une molécule d'ADN ou d'ARN, et qui est utilisé pour amorcer (initier) la synthèse d'un nouveau brin d'ADN complémentaire à partir de ce site en particulier, au cours d'une réaction ou une série de réactions catalysées par une ADN polymérase. Le brin d'ADN nouvellement synthétisé contient l'amorce à son extrémité 5'P. Généralement, ces oligonucléotides de 15 à 50 nucléotides de long sont synthétisés chimiquement, caractérisés par une séquence spécifique et connue. L'amorce *sens* ou *forward* va s'hybrider sur le brin antisens tandis que l'amorce *antisens* ou *reverse* va s'hybrider sur le brin sens. Voir BRIN ANTISENS.

Cependant, " des amorces aléatoires " (oligonucléotides plus courts, d'environ 6 nucléotides de long et pouvant présenter toutes les combinaisons de séquences possibles) peuvent aussi être utilisées pour la synthèse d'ADN à partir d'une séquence ARN inconnue.

**Anglais : primer.**

**ANNEALLING** - Annelage. Voir HYBRIDATION.

**Anglais : annealing.**

**ARN POLYMÉRISE** - Polymérase qui synthétise l'ARN. Voir POLYMÉRISE.

**Anglais : RNA polymerase, DNA Dependent RNA polymerase.**

**ARNm** - ARN cytoplasmique à l'origine de la synthèse des protéines. Voir TRANSCRIPTION.

**Anglais : mRNA.**

**ARN RIBOSOMAL (ARNr)** - Composants ARN du ribosome capables de jouer un rôle catalytique au cours de la traduction. De quatre tailles différentes chez les eucaryotes (5S, 5.8S, 18S et 28S), ce sont les ARN les plus abondants (80%) au niveau cellulaire, les cellules contenant de multiples copies de gènes ARNr codant pour ces ARN ribosomiaux.

**Anglais : rRNA.**

**ARNt** - Petit ARN se présentant sous la forme de petites molécules repliées dont le rôle est d'apporter des acides aminés spécifiques au niveau des ribosomes actifs, de manière dépendante de la séquence de l'ARNm. Une extrémité de la molécule d'ARN de transfert reconnaît le triplet de nucléotides qui correspond au codon de l'ARNm, tandis que l'autre extrémité (lorsqu'il est activé) est liée de façon covalente à l'acide aminé adéquat.

**Anglais : tRNA.**

**bp** - Paire de bases.

**Anglais : bp.**

**BRIN ANTISENS (ou AMORCE)** - Désigne le brin d'ARN ou d'ADN d'une molécule duplex qui est complémentaire de celui codant pour un polypeptide. Plus précisément, le brin d'ADN qui sert de matrice pour la synthèse de l'ARN et qui lui est complémentaire. Les oligonucléotides " antisens " s'hybrident à l'ARN messager et sont utilisés pour la synthèse de l'ADN complémentaire.

**Anglais : antisens strand (or primer).**

**BROMURE d'ÉTHIDIUM** - Colorant capable de s'insérer dans la structure des acides nucléiques de telle manière qu'ils fluorescent sous lumière UV ( $\lambda = 254$  nm). Formule chimique :  $C_{21}H_{20}BrN_3$ .

**Anglais : ethidium bromide.**

**CALORIMÉTRIE DIFFÉRENTIELLE A BALAYAGE** - Technique d'analyse thermique permettant de mesurer les différences d'échanges de chaleur entre un échantillon à analyser et une référence. Elle permet notamment de déterminer les températures de fusion, transition de phase endothermique absorbant de la chaleur.

**Anglais : Differential Scanning Calorimetry, DSC.**

**CLEAN-UP** - Étape de purification d'ADN ou d'ARN en solution à l'aide d'une membrane de silice. Les acides nucléiques sont tout d'abord retenus à travers la membrane par affinité de charge. Ils sont ensuite élués par l'application d'une forte marche ionique au niveau de la membrane, conférant ainsi un éluât de bonne qualité dépourvu des principaux contaminants du mélange initial.

**Anglais : clean-up.**

**COMPLÉMENTAIRE** - Voir PAIRE DE BASES.

**Anglais : complementary.**

**COURBE DE FUSION** – Méthode qui permet de déduire le  $T_m$  d'un duplexe d'ADN en réalisant la cinétique de deshybridation fluorescence =  $f(T)$ .

**Anglais : Melting curve.**

**COURBE DE FUSION HAUTE RÉOLUTION (HRM)** – Courbe de fusion réalisée avec un intercalant ADN saturant (LCGreen, SYTO 9, EvaGreen, Chromofy, Bebo ou ResoLight) et une vitesse de transition thermique très faible permettant de mettre en évidence de très faibles différences de séquences (1 nucléotide sur 200).

**Anglais : High Resolution Melting curve.**

**DÉNATURATION** - Se réfère à la conversion de l'état double brin des acides nucléiques à l'état simple brin, souvent obtenues par chauffage ou dans des conditions alcalines. Elle est également appelée " fusion " de l'ADN.

**Anglais : denaturation.**

**DIMÈRE D'AMORCES** - Sous-produit potentiel de la PCR constitué de deux séquences d'amorces hybridées l'une à l'autre en raison de succession de bases complémentaires au niveau de leurs séquences. En conséquence, l'ADN polymérase amplifie le dimère d'amorces, conduisant potentiellement à une inhibition de l'amplification de la séquence d'ADN ciblée par PCR et venant ainsi altérer la précision de la quantification.

**Anglais : primer dimer.**

**ds** - Double brin.

**Anglais : ds.**

**DUPLEXE** - Molécule d'acide nucléique au niveau de laquelle deux brins sont bases-appariés l'un avec l'autre.

**Anglais : duplex.**

**DUPLICATE** - Vient d'un terme anglais signifiant " faire un copie à l'identique ". Dans les études que j'ai développées, ce terme désigne le fait de réaliser deux fois la même analyse en même temps (on parle alors de *duplicate technique*) ou le fait de réaliser l'analyse sur deux échantillons provenant d'un même tissu, prélevés au même moment (on parle alors de *duplicate biologique*).

**Anglais : duplicate.**

**EFFICACITÉ PCR** - Notée  $\epsilon$ , élément fondamental à prendre en compte pour considérer la mesure quantitative de l'expression d'un gène. Elle est définie par l'équation  $\epsilon = 10^{-(1/p)}$  - 1 où p est la pente de la droite : Nombre de cycles =  $f(\log([C]))$  dans le range de concentrations données.

**Anglais : PCR efficiency.**

**ÉPISSAGE DE L'ARN** - Série de réactions complexes et comprises de manière incomplète se produisant au niveau du noyau des cellules eucaryotes où les transcrits pré-ARNm issus de l'ADN chromosomique sont modifiés de telle sorte que les régions non codantes du pré-ARNm (introns) sont excisées, et les régions codantes (exons) sont liées de façon covalente, de manière à produire une molécule d'ARNm prête à être transportée vers le cytoplasme. En raison de cette réaction d'épissage, l'ADN eucaryote représentant un gène codant pour une protéine donnée est habituellement beaucoup plus grand que l'ARNm à partir duquel la protéine est effectivement traduite.

**Anglais :** RNA splicing.

**EVAGREEN** - Intercalant ADN stable et fluorescent de dernière génération capable de s'insérer au niveau d'une double hélice ADN de façon systématique. Il est souvent utilisé au cours de réaction d'HRM.

**Anglais :** EvaGreen® dye.

**EXPRESSION** - Se réfère habituellement à l'ensemble du processus de production d'une protéine à partir d'un gène, comprenant ainsi la transcription, la traduction, les modifications post-traductionnelles et éventuellement les réactions d'adressage.

**Anglais :** expression.

**EXTENSION D'AMORCE** - Réaction au cours de laquelle l'ADN est reverse transcrit à partir d'une matrice d'ARN au niveau de laquelle une amorce oligonucléotidique spécifique a été hybridée. Le nouveau produit d'ADNc est une extension de cette amorce qui a été synthétisée à partir de l'extrémité 3'OH du primer en direction de l'extrémité 5'P de l'ARN. Cette réaction est utile pour identifier les séquences 5'P terminales des molécules d'ARN.

**Anglais :** primer extension.

**EXTRÉMITÉ 3'OH** - Désigne l'extrémité de la séquence polynucléotidique terminée par un ose (ribose ou désoxyribose), et notamment par un groupement hydroxyle OH porté par le carbone 3' du sucre.

**Anglais :** 3' end.

**EXTRÉMITÉ 5'P** - Désigne l'extrémité de la séquence polynucléotidique terminée par un phosphate.

**Anglais :** 5' end.

**FLUOROPHORES ou FLUOROCHROMES** - Molécules chimiques composées de plusieurs noyaux aromatiques conjugués ou molécules planes et cycliques possédant une ou plusieurs liaisons  $\pi$ , capables d'émettre de la fluorescence après excitation à une certaine longueur d'onde.

**Anglais :** fluorophore / fluorochrome.

**FLANKING (DESIGN EN)** - Méthode de dessin d'amorces consistant à placer la séquence de l'amorce *forward* un peu avant un site d'épissage d'un intron et la séquence de l'amorce *reverse* juste un peu après. Cette manière de dessiner permet ainsi de ne pas « parasiter » la quantification par PCR par des contaminations d'ADN génomique.

**Anglais :** flanking.

**FOLDING** - Capacité d'un acide nucléique à se replier sur lui-même en formant des appariements Watson-Crick intramoléculaires conduisant localement à la formation de régions en hélice et de régions en boucle au niveau desquelles les bases ne sont pas appariées. Cette topologie est ce qu'on appelle la structure secondaire de la molécule.

**Anglais :** folging.

**FRET (TRANSFERT D'ÉNERGIE PAR RÉSONANCE)** - Transfert d'énergie non radiatif (sans émission de lumière) résultant d'une interaction dipôle – dipôle entre deux molécules (donneur et accepteur d'énergie). Ce phénomène physique nécessite une compatibilité énergétique entre ces molécules. Cela signifie que le spectre d'émission du donneur doit recouvrir, au moins partiellement, le spectre d'absorption de l'accepteur.

**Anglais :** Förster resonance energy transfer.

**HÉTÉRODUPLEXE** - Résultat de l'appariement entre les bases situées sur des simples brins complémentaires issus de différents duplexes parentaux ; ces hétéroduplexes se rencontrent lors de la recombinaison génétique *in vivo* et lors de l'hybridation *in vitro* entre différents brins d'ADN d'origine commune. Ainsi, les séquences des deux brins d'un hétéroduplexe diffèrent. La molécule ne présente donc pas un appariement parfait. La température de fusion d'un hétéroduplexe dépend du nombre de mismatches.

**Anglais :** heteroduplex.

**HÉTÉROSCÉDASTICITÉ** - Provenant du grec *hétéro* - " autre " et *skedasé* - " dissipation ", notion observée lorsque les variances des variables examinées sont différentes. Elle s'oppose à l'*homoscédasticité*, qui correspond au cas où la variance de l'erreur des variables est constante. Les tests statistiques permettant d'apprécier l'hétéroscédasticité sont le test de *White* et le test de *Goldfeld et Quandt*.

**Anglais :** heteroscedasticity.

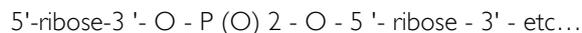
**HYBRIDATION** - Processus d'appariement de bases conduisant à la formation de duplex d'ARN ou d'ADN ou à des duplexes ARN : ADN.

**Anglais :** hybridization.

**KILOBASE** - Unité de 1000 bases nucléotidiques, ARN ou ADN, notée kb.

**Anglais :** kilobase.

**LIAISON PHOSPHODIESTER** - Liaison covalente reliant un groupement hydroxyle située au niveau de l'extrémité 3' du sucre d'un nucléotide avec un groupement phosphate du sucre du résidu nucléotidique adjacent, situé au niveau 5'P, au niveau d'un acide nucléique :



**Anglais : phosphodiester bond.**

**MATRICE** - Brin d'acide nucléique sur lequel une amorce est capable de s'hybrider afin de donner naissance à un autre brin néo synthétisé par une polymérase.

**Anglais : template.**

**MÉSAPPARIEMENT** - Désigne un appariement de deux brins d'ADN non-conforme aux principes biochimiques définis par WATSON & CRICK.

**Anglais : mismatch.**

**MOLECULAR BEACON** - Sonde formée d'un brin d'ADN en épingle à cheveu, chacune des extrémités portant un fluorophore. L'un est dit reporter et l'autre quencher (extincteur). Quand la sonde est refermée, le quencher supprime la fluorescence du reporter (par transfert d'énergie de type FRET). Quand la sonde s'hybride à la source, elle change de conformation, s'ouvre, et le quencher s'éloigne du reporter. La fluorescence n'est plus inhibée et peut alors être détectée.

**Anglais : molecular beacon.**

**MOUVEMENT BROWNIEN** - Décrit pour la première fois en 1827 par le botaniste Robert Brown, encore appelé processus de Wiener, description mathématique du mouvement aléatoire et très irrégulier d'une grosse particule immergée dans un fluide et qui n'est soumise à aucune autre interaction que des chocs avec les petites molécules du fluide environnant.

**Anglais : Brownian motion.**

**NESTED PCR ou PCR NICHÉE** - Méthode très spécifique au cours de laquelle une fraction d'un produit de PCR obtenu au cours d'une première réaction d'amplification (après 30-35 cycles) est soumise à une nouvelle réaction PCR utilisant des amorces différentes s'hybridant à l'intérieur de la région délimitée par la première paire d'amorces utilisée. Voir POLYMERASE CHAIN REACTION.

**Anglais : nested PCR.**

**NORMALISATION** - Processus permettant de corriger les variations observées entre différents échantillons. Elle peut se faire par rapport à une référence externe (*spike*) ou par rapport à une référence interne (gène de référence).

**Anglais : normalization.**

**NOYAU DE NUCLÉATION** - Deux séquences non appariées d'une longueur de 13 à 16 bases consécutives et entièrement complémentaires.

**Anglais : nucleation.**

**OLIGODÉSOXYRIBONUCLÉOTIDE** - Courte molécule d'ADN simple brin, généralement d'une longueur de 5 à 50 nucléotides et qui peut être utilisée comme amorce ou sonde d'hybridation. Les oligodésoxyribonucléotides sont synthétisés chimiquement dans les conditions automatisées.

**Anglais : oligodeoxyribonucleotide.**

**OLIGONUCLÉOTIDE** - Voir OLIGODÉSOXYRIBONUCLÉOTIDE.

**Anglais : oligonucleotide.**

**POINT ABERRANT** - Point de donnée statistique qui semble s'écarter sensiblement des autres points de donnée et apparaît représenter une valeur aberrante. Ces valeurs aberrantes peuvent se produire par hasard dans une distribution, mais elles sont souvent indicatives d'erreurs de mesure. Dans ce cas, il est souhaitable de les rejeter de l'analyse statistique. Une cause fréquente de valeurs aberrantes est un mélange de deux distributions.

**Anglais : outlier.**

**PAIRE de BASES** - Paire de nucléotides complémentaires dans un duplex d'un acide nucléique. En vertu des règles de Watson et Crick, ces paires se composent d'une pyrimidine et d'une purine : c'est-à-dire C-G ou A-T dans le cas de l'ADN ou U-A dans le cas de l'ARN. Cependant, les paires de bases " non canoniques " comme par exemple G-U sont communes dans la structure secondaire de l'ARN.

**Anglais : base pair.**

**PCR** - Voir POLYMERASE CHAIN REACTION

**Anglais : PCR.**

**PCR multiplex** - Amplification simultanée de deux ou plusieurs cibles ADN dans un seul récipient.

**Anglais : multiplex PCR.**

**PCR digitale** – Technique de PCR consistant à diluer l'ADN à analyser en plaques multi-puits de telle sorte qu'une molécule cible soit déposée dans au minimum deux puits. La réaction est effectuée dans des conditions optimales permettant d'amplifier une seule copie de séquence cible. Les amplicons sont hybridés avec des sondes fluorescentes de type *Molecular Beacons*, permettant la détection spécifique de séquences à l'aide de fluorophores différents. La PCR digitale utilise donc le ratio de résultats positifs / résultats négatifs obtenus par chaque réaction de PCR pour comptabiliser le nombre de molécules cibles.

**Anglais : digital PCR.**

**POLYMÉRASE** - Enzyme capable de catalyser l'addition d'un nucléotide à une molécule d'acide nucléique en cours d'élongation à partir d'un brin matriciel. Il existe une grande variété d'ARN et d'ADN polymérases présentant un large éventail d'activités spécifiques et fonctionnant de façon optimale dans des conditions différentes. En général, toutes les polymérases nécessitent la présence d'un brin matriciel ou matrice sur lequel elles vont pouvoir réaliser la synthèse d'un nouveau brin d'ADN ou d'ARN. Les ADN polymérases ont besoin de l'existence d'amorces afin de pouvoir initier la synthèse d'un nouveau brin tandis que les ARN polymérases commencent la synthèse au niveau d'une séquence promotrice spécifique.

**Anglais : polymerase.**

**POLYMERASE CHAIN REACTION (PCR)** - Réaction d'amplification spécifique de l'ADN impliquant plusieurs cycles (30 ou plus) d'hybridation, d'extension d'amorces ainsi que de dénaturation, généralement basée sur l'utilisation d'une ADN polymérase thermostable telles que la *Taq* polymérase. Les amorces sont utilisées par paires, complémentaires de chaque brin de l'ADN et délimitent ainsi une région à amplifier. Dans des conditions optimales, la séquence cible peut être amplifiée un million de fois.

**Anglais : Polymerase Chain Reaction.**

**PROCESSIVITÉ** - Mesure de la vitesse d'une ARN ou une ADN polymérase lorsque le brin matriciel est lu, avant que la protéine ne se dissocie. Elle est déterminée par une longueur moyenne (en kilobases) des brins d'acides nucléiques nouvellement synthétisés et s'applique également aux exonucléases lorsque ces enzymes digèrent un acide nucléique à partir de ses extrémités.

**Anglais : processivity.**

**QUENCHEUR** - Substance capable d'absorber l'énergie d'excitation d'un fluorophore et de la dissiper sous forme de chaleur. Elle est utilisée en biologie moléculaire en conjonction avec des fluorophores. Quand les deux sont rapprochés, comme au niveau d'une sonde oligonucléotidique, l'émission du fluorophore est alors supprimée.

**Anglais : dark quencher.**

**RÉACTION de RT-qPCR** - Série de réactions aboutissant à l'amplification et la quantification indirecte de l'ARN à partir de la synthèse d'un ADN complémentaire. Trois différents types d'amorces (spécifiques de la séquence, oligo dT ou random hexamères) peuvent être utilisés au cours de la synthèse de l'ADNc simple brin, réalisée par la transcriptase inverse à partir de la matrice ARN. Une seconde amorce complémentaire de l'ADNc nouvellement synthétisé ainsi que la *Taq* polymérase sont ensuite ajoutées au mélange réactionnel. L'ADN est alors amplifié après plusieurs cycles de dénaturation, d'hybridation d'amorces et de polymérisation afin d'être quantifié.

**Anglais : one step RT/PCR.**

**RENATURATION (D'ACIDE NUCLÉIQUE)** - Réassociation, après dénaturation, de simples brins d'ADN selon les règles d'appariement définies par WATSON & CRICK.

**Anglais : renaturation.**

**RÉSONANCE PLASMONIQUE DE SURFACE** - Phénomène physique permettant de mesurer la liaison d'un " ligand " sur un " récepteur " adsorbé à la surface d'une couche métallique. Le système de détection par résonance plasmonique de surface mesure alors la variation de l'indice de réfraction au voisinage de l'interface lorsque le ligand se fixe aux récepteurs.

**Anglais : Surface plasmon resonance, SPR.**

**REVERSE TRANSCRIPTION** - Copie d'une molécule d'ARN en molécule d'ADN complémentaire.

**Anglais : reverse transcription.**

**SÉQUENCES COMPLÉMENTAIRES** - Deux séquences d'acides nucléiques appariées l'une par ce que les bases A, T, G et C correspondent respectivement aux bases T, A, C et G de l'autre séquence.

**Anglais : complementary sequences.**

**SÉQUENCE REVERSE COMPLÉMENTAIRE** - Encore nommée « complément inverse », séquence d'acides nucléiques dont l'enchaînement de nucléotides se trouve inversé et complémentaire par rapport à une première séquence. Par exemple, pour une séquence 5'P-ATGGGCCT-3'OH, le complément inverse sera 5'P-AGGCCCAT-3'OH.

**Anglais : reverse complement.**

**SONDE** - Se réfère généralement à une molécule d'ADN ou d'ARN qui a été marquée au <sup>32</sup>P ou avec de la biotine, afin de faciliter sa détection après qu'elle se soit spécifiquement hybridée à sa séquence cible ADN ou ARN.

**Anglais : probe.**

**ss** - " *Single stranded* ", simple brin.

**Anglais : ss, single stranded.**

**STANDARDISATION** - Désigne l'ensemble des mesures permettant de traiter de manière strictement identique l'ensemble des échantillons à analyser. Elles ont lieu depuis le prélèvement de l'échantillon jusqu'à la réaction d'amplification de l'ADN en passant par les étapes d'extraction et de transcription inverse.

**Anglais : standardization.**

**STRINGENCE** - Conditions d'appariement des acides nucléiques déterminant la spécificité de la réaction d'hybridation entre deux molécules simple brin. Des conditions de plus en plus strictes peuvent être obtenues en augmentant la température ou en diminuant la force ionique, permettant ainsi une plus grande spécificité (mais une plus faible sensibilité) de la réaction d'hybridation (stringence élevée). Inversement, une baisse de la température aura plutôt tendance à diminuer la stringence de l'hybridation.

**Anglais : stringency.**

**STRUCTURE EN ÉPINGLE À CHEVEU** - Structure hélicoïdale (duplexe) formée par appariement de bases voisines (inversées) situées sur des séquences complémentaires au sein d'un même brin d'ARN ou d'ADN.

**Anglais :** hairpin.

**STRUCTURE en TIGE - BOUCLE** - Structure secondaire caractéristique de l'ARN au niveau de laquelle deux séquences inversées et complémentaires séparées par une courte séquence, situées sur un seul brin d'ARN, s'apparient pour former une " tige " avec une " boucle " à son extrémité. Elles sont semblables à une épingle à cheveux.

**Anglais :** stem-loop.

**STRUCTURE PRIMAIRE** - Fait respectivement référence à la séquence en nucléotides ou en résidus d'acides aminés au sein de molécules d'acide nucléique ou de protéines. Voir aussi STRUCTURE SECONDAIRE et TERTIAIRE.

**Anglais :** primary structure.

**STRUCTURE SECONDAIRE** - Structure locale retrouvée au sein d'une molécule d'ARN conférée par les appariements de bases entre nucléotides, relativement rapprochés dans la séquence (par exemple, structures en épingles à cheveux, structures en tige-boucle, ... ). Voir aussi STRUCTURE PRIMAIRE et TERTIAIRE.

**Anglais :** secondary structure.

**SYBRGREEN I** - Composé organique aromatique de formule chimique  $C_{32}H_{37}N_4S$  faisant partie des cyanines asymétriques et utilisé pour le marquage de l'ADN double brin au cours de la réaction d'amplification en chaîne de l'ADN.

**Anglais :** SybrGreen I

**Taq POLYMÉRISE** - ADN polymérase qui est très stable à haute température, isolée à partir de la bactérie thermophile *Thermus aquaticus*. Très utile dans les réactions de PCR, cette protéine peut supporter des températures élevées à plusieurs reprises durant les cycles de dénaturation. Certaines Taq sont dites " Hot Start ", elles présentent un inhibiteur, généralement de nature protéique ou chimique et sont activées par le biais d'un chauffage de l'enzyme à 95°C pendant 10 ou 15 minutes.

**Anglais :** Taq polymerase.

**TEMPÉRATURE DE FUSION ( $T_m$ )** - Température médiane du range de températures au niveau duquel l'ADN est dénaturé par la chaleur. C'est donc une température de demi-dénaturation à laquelle une molécule d'acide nucléique est à 50% sous forme de duplexe, 50% sous forme simple brin. Elle est dépendante du nombre et de la proportion de paires de bases GC ainsi que des conditions ioniques dans lesquelles elle est mesurée. Souvent considérée comme une mesure de la stabilité thermique pour une hybridation sonde oligonucléotidique / séquence cible.

**Anglais :** melting temperature.

**TEMPÉRATURE D'HYBRIDATION** - Température dépendant de la composition en bases des amorces oligonucléotidiques utilisées en PCR. Elle est légèrement inférieure (environ de 5°C) à la température de fusion ( $T_m$ ).

**Anglais : annealing temperature.**

**TITRATION CALORIMÉTRIQUE ISOTHERME** - Méthode extrêmement sensible capable de mesurer les données thermodynamiques d'une grande variété de systèmes. Elle permet la mesure de puissances thermiques extrêmement faibles, de l'ordre de 0.1  $\mu$ W et l'obtention d'informations cinétiques (constantes de vitesse d'association/dissociation). Grâce à sa grande sensibilité, la titration calorimétrique isotherme est fréquemment utilisée en biologie, notamment pour l'étude de processus d'association divers (lipides membranaires, protéines, acides nucléiques, assemblages de macromolécules).

**Anglais : Isothermal titration calorimetry, ITC.**

**TRANSCRIT** - Nouvelle molécule d'ARN synthétisée à partir d'un ADN par une ARN polymérase.

**Anglais : transcript.**

**TRANSCRIPTASE INVERSE ou REVERSE TRANSCRIPTASE** - ADN polymérase ARN dépendante, capable de copier une molécule d'ARN en ADN complémentaire simple brin. Ces enzymes sont habituellement purifiées à partir de rétrovirus.

**Anglais : reverse transcriptase.**

**ZIPPING** - Phénomène d'hybridation très rapide de deux séquences complémentaires d'acides nucléiques ayant lieu au niveau des noyaux de nucléation.

**Anglais : zipping.**

# Sites Internet Utilisés

## I. Sites de référence de PCR et RT-qPCR

---

- ❖ Aspects techniques de la quantification de l'expression génique en qPCR & RT-qPCR - Michael PFAFFL  
<http://www.gene-quantification.info/>
- ❖ Site de données de RT-qPCR en format d'échange standard (*Real-time PCR Data Markup Language*)  
<http://www.rdml.org/>  
<http://www.rdml.org/miqe.php>
- ❖ Site internet de Stephen BUSTIN  
<http://www.sabustin.org/>
- ❖ Site internet de TATAA BIOCENTER - Mikael KUBISTA  
<http://www.tataa.com/>
- ❖ Site internet de l'Université du Texas - Gregory SHIPLEY  
<http://www.uth.tmc.edu/research/corelabs/qgcl/index.htm>
- ❖ Site internet de BIOGAZELLE  
<http://www.biogazelle.com/news>
- ❖ Site internet de *geNorm*  
<http://medgen.ugent.be/~jvdesomp/genorm/>
- ❖ Site internet de CEPHEID  
<http://www.cephheid.com/systems-and-software/systems-overview>
- ❖ Site de RT-PCR  
<http://www.rt-pcr.com/>
- ❖ PCR digitale  
<http://digital-PCR.gene-quantification.info/>
- ❖ Site de *PCR Station*  
<http://www.pcrstation.com/>
- ❖ Site d'informations sur la PCR  
<http://www.highveld.com/pages/pcr.html>

## 2. Structures secondaires

---

- ❖ *Folding de l'ARN*  
<http://mfold.bioinfo.rpi.edu/cgi-bin/ma-form1.cgi>  
<http://www.tbi.univie.ac.at/~ivo/RNA/>  
[http://www.genebee.msu.su/services/ma2\\_full.html](http://www.genebee.msu.su/services/ma2_full.html)
- ❖ *Folding de l'ADN*  
<http://mfold.bioinfo.rpi.edu/cgi-bin/dna-form1.cgi>  
<http://motif.genome.ad.jp/>

## 3. Design d'amorces et sondes

---

- ❖ *Logiciels*
  - Mac Vector V.12.6 (MacVector, Inc.)*  
<http://www.macvector.com/>
  - Visual OMP™ Nucleic Acid Software (DNA Software, Inc.)*  
<http://dnasoftware.com/VisualOMP/tabid/108/Default.aspx>
  - Primer Premier V. 6.11 (PREMIER Biosoft)*  
<http://www.premierbiosoft.com/primerdesign/index.html>
  - Primer Explorer V.4 (Fujitsu System Solutions Ltd)*  
<http://primerexplorer.jp/elamp4.0.0/index.html>
  - Primer3 V. 0.4.0 (S. ROZEN)*  
<http://frodo.wi.mit.edu/primer3/>
  - Primer Quest (Integrated DNA Technologies, Inc.)*  
<http://biotools.idtdna.com/PrimerQuest/Home/Index>
- ❖ *Portail de sites de design d'amorces et sondes*  
[http://www.humgen.nl/primer\\_design.html](http://www.humgen.nl/primer_design.html)
- ❖ *BLAST*  
<http://blast.ncbi.nlm.nih.gov/Blast.cgi>  
<http://www.ch.embnet.org/>  
<http://mobyte.pasteur.fr/cgi-bin/portal.py?#forms::blast2>
- ❖ *Librairie de sondes ROCHE*  
<https://www.roche-applied-science.com/sis/rtPCR/upl/index.jsp>

## 4. Calculateurs de $T_m$

---

- ❖ Server de Calculateurs de  $T_m$   
[http://protein.bio.puc.cl/cardex/servers/melting/sup\\_mat/servers\\_list.html](http://protein.bio.puc.cl/cardex/servers/melting/sup_mat/servers_list.html)
- ❖ Calculateurs de  $T_m$  pour oligonucléotides modifiés par LNA  
<http://eu.idtdna.com/analyzer/applications/oligoanalyzer/>  
<http://www.exiqon.com/Is/Pages/ExiqonTMPredictionTool.aspx>
- ❖ Calculateurs de  $T_m$   
<http://www.basic.northwestern.edu/biotools/oligocalc.html>  
<http://mbcf.dfc.harvard.edu/docs/oligocalc.html>  
<http://dna.bio.puc.cl/cardex/servers/dnaMATE/tm-pred.html>  
<http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::melting>  
[http://www6.appliedbiosystems.com/support/techttools/tm\\_calculator.cfm](http://www6.appliedbiosystems.com/support/techttools/tm_calculator.cfm)  
[http://www.neb.com/nebecomm/tech\\_reference/TmCalc/Default.asp?#.UBzbjEQ\\_uuQ](http://www.neb.com/nebecomm/tech_reference/TmCalc/Default.asp?#.UBzbjEQ_uuQ)  
[http://www.finnzymes.fi/tm\\_determination.html](http://www.finnzymes.fi/tm_determination.html)  
<http://www.promega.com/techserv/tools/biomath/calc11.htm>

## 5. ARN

---

- ❖ Site du *Leibniz Institute for Age Research - Fritz Lipmann Institute*  
<http://www.ma.uni-jena.de/ma.php>
- ❖ Site de RNABase  
<http://www.mabase.org/primer/>
- ❖ Identification des régions codantes  
<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>
- ❖ Recherche de sites de fixation de facteurs de transcription  
[http://www.genomatix.de/online\\_help/help\\_matinspector/matinspector\\_help.html](http://www.genomatix.de/online_help/help_matinspector/matinspector_help.html)
- ❖ TESS : *Transcription Element Search System*  
<http://www.cbil.upenn.edu/cgi-bin/tess/tess>
- ❖ *Genevestigator*  
<https://www.genevestigator.com/gv/index.jsp>
- ❖ Site d' *AMBION*  
<http://www.ambion.com/>

## 6. Sites d'intérêts divers

---

- ❖ Moteurs de recherche scientifique  
<http://scholar.google.fr/>  
<http://www.scirus.com>
- ❖ Site de l'*International Union of Pure and Applied Chemistry (IUPAC)*  
<http://www.iupac.org/>
- ❖ PubMed  
<http://www.ncbi.nlm.nih.gov/pmc/>
- ❖ Site de l'*U.S. National Library of Medicine National Institutes of Health*  
<http://www.nlm.nih.gov/>
- ❖ Site de la base de données *Genbank*  
<http://www.ncbi.nlm.nih.gov/genbank>
- ❖ Site d'outils d'analyse  
<http://smart.embl-heidelberg.de>  
<http://molbiol-tools.ca/>  
[http://www.sacs.ucsf.edu/Links/BioInfo\\_Meta.html](http://www.sacs.ucsf.edu/Links/BioInfo_Meta.html)  
<http://sites.univ-provence.fr/~wabim/english/logligne.html>  
<http://biology.fullerton.edu/biol402/molbio.html>
- ❖ Biologie moléculaire et Génétique de l'Université de JUSSIEU  
<http://www.snv.jussieu.fr/bmedia/sommaires/gbm.html>
- ❖ OMIM (*Human genes and genetic disorders*)  
<http://www.ncbi.nlm.nih.gov/omim/>
- ❖ Site de Justbio  
<http://www.justbio.com/>
- ❖ Liens Bioscience  
<http://www.bioexplorer.net>
- ❖ BIOSCI/Bionet  
<http://www.bio.net/>

## 7. Statistiques

---

- ❖ Tests statistiques  
<http://mame.u707.jussieu.fr/biostatgv/?module=tests>

## 8. Glossaires

---

- <http://www.pcrlinks.com/glossary/N.htm>
- [http://www.genomicglossaries.com/content/gene\\_amplification\\_pcr\\_gloss.asp](http://www.genomicglossaries.com/content/gene_amplification_pcr_gloss.asp)

# Publications

---



## Pitfalls of reverse transcription quantitative polymerase chain reaction standardization: Volume-related inhibitors of reverse transcription

Pascal Pugniere<sup>a</sup>, Sebastien Banzet<sup>b</sup>, Thomas Chaillou<sup>b</sup>, Catherine Mouret<sup>a</sup>, Andre Peinnequin<sup>a,\*</sup>

<sup>a</sup> Genomic Core Facility, IRBA La Tronche, 38702 La Tronche, France

<sup>b</sup> Operational Environments, IRBA La Tronche, 38702 La Tronche, France

### ARTICLE INFO

#### Article history:

Received 2 February 2011

Received in revised form 1 April 2011

Accepted 4 April 2011

Available online 15 April 2011

#### Keywords:

Reverse transcription

RT-qPCR

mRNA quantification

Standardization

Inhibitor

MIQE guidelines

### ABSTRACT

A large part of the reliability of reverse transcription quantitative polymerase chain reaction (RT-qPCR) data depends on technical variations. Such variations are mainly attributable to the reverse transcription step. Standardization is a key factor in decreasing the intersample variability. However, an ideal standardization is not always possible, and compromises must be found. Due to technical requirements, the current consensus is that a constant amount of total RNA should be used for the RT step (CA-RT). Because RNA isolation yields are variable, such a practice requires the use of variable volumes of nucleic acid extracts in RT reaction. We demonstrate that some RNA extracts contain both exogenous and endogenous inhibitors. These inhibitors induce a decrease in RT efficiency that significantly impairs the reliability of RT-qPCR data. Conversely, these inhibitors have a slight effect on the qPCR step. To overcome such drawbacks, we proposed to carry out the RT reaction with a constant volume of RNA extract by preserving a constant RNA amount through the supplementation of yeast transfer RNA (CV-RT). We show that CV-RT, compared with the usual CA-RT, allows us to decrease the RT-qPCR variability induced by intersample differences. Such a decrease is a prerequisite for the reliability of messenger RNA quantification.

© 2011 Elsevier Inc. All rights reserved.

Due to an unrivaled combination of simplicity, cost efficiency, accuracy, and availability, reverse transcription quantitative polymerase chain reaction (RT-qPCR)<sup>1</sup> increasingly appears to be the benchmark in the field of messenger RNA (mRNA) quantification. However, serious concerns have been raised regarding the reliability of RT-qPCR data [1]. Significant efforts were recently made to improve the reliability of the quantification [2]. It is now established that a significant part of the technical variability observed in RT-qPCR experiments is determined by variation in RT yields [3]. Thus, it was reported that total RNA concentration [4,5], target template quantity [6], and background RNA [7] influence the RT yield.

Numerous factors are currently described to inhibit the overall RT-qPCR assay [8] that are both template and technical process dependent. Myoglobin and collagen are clearly identified as endogenous RT-qPCR inhibitors in skeletal muscle [9]. Moreover, hemoglobin [10] and lipids [11,12], known to be RT-qPCR inhibitors, are present in the spleen and central nervous system, respectively.

Furthermore, inhibition of RT-qPCR also depends on the RNA isolation method. Phenol, which discards hydrophobic endogenous inhibitors as lipids, is a well-characterized PCR inhibitor [13]. Silica-based columns that are more effective than the organic isolation method for removing collagen [10] carry over fewer inhibitors from the reagent. However, such columns appear to be less effective than organic reagent in discarding hydrophobic inhibitors as lipids. Thus, manufacturers recommend an organic step before silica-based RNA isolation.

Inhibitors related to the isolation step should be found in a constant concentration between samples to lessen intersample variations. However, an ideal standardization is not always possible, and compromises must be found owing to experimental and/or technical requirements. Due to the standardization process, the content of these inhibitors should be dependent on the reagent volume. The current consensus is that a constant amount of total RNA should be used for the RT step (CA-RT) [2]. Due to the variability of RNA isolation yields, using a constant total RNA amount requires the use of variable volumes of nucleic acid extract in RT reaction. This suggests that the amount of inhibitors derived from both sample preparation and RNA extraction may vary between samples. For example, the phenol contamination resulting from organic RNA extraction should correlate with the volume of RNA extract rather than with the RNA amount. Moreover, to our knowledge, there is no established evidence concerning an eventual

\* Corresponding author. Fax: +33 4 76 63 69 40.

E-mail address: [biomol.irba@orange.fr](mailto:biomol.irba@orange.fr) (A. Peinnequin).

<sup>1</sup> Abbreviations used: RT-qPCR, reverse transcription quantitative polymerase chain reaction; mRNA, messenger RNA; CA-RT, RT step with a constant amount of total RNA; CV-RT, RT reaction with a constant volume of RNA extract; tRNA, transfer RNA; RIN, RNA integrity number; cDNA, complementary DNA; RSD, relative standard deviation; MIQE, minimum information for publication of quantitative real-time PCR experiments.

relationship between the concentration of the copurified endogenous inhibitors and the RNA yield. Consequently, the use as RT template, of either variable volumes of RNA extract (CA-RT) or variable amounts of total RNA, is questionable. A compromise could be to carry out the RT reaction with a constant volume of RNA extract (CV-RT) by preserving a constant RNA amount between samples through the supplementation of yeast transfer RNA (tRNA) [4]. Such a method possibly provides a relevant approach, allowing a constant RT yield while minimizing the intersample variations in volume-related inhibitors.

The aim of this study was to determine whether CV-RT could improve the reliability of RT-qPCR assessment compared with the classical use of a constant amount of total RNA (CA-RT). Comparing CA-RT with CV-RT requires us to use realistic biological samples. In contrast to artificially modified templates [7], true biological samples display unknown target mRNA levels. Therefore, for analysis of the reproducibility, we used technical replicates containing identical amounts of target mRNA, albeit unknown. For agreement assessment between methods, we used biological replicates displaying higher ranges of target mRNA levels. First, we generated templates that differ in both quality and concentration of inhibitors by varying both the nature of the sample and the RNA isolation method. These templates, containing potential volume-related inhibitors, were used to study intersample variations in RT and qPCR efficiencies. Then consequences of these potential volume-related inhibitors on intersample differences in both RT efficiency and qPCR efficiency were investigated. We especially focused on accuracy of RNA measurement using a multiparametric normalization index.

## Materials and methods

### Experimental design

The influence of volume-related inhibitors was assessed using three rat tissues. Twelve independent biological spleen samples were processed, whereas a single sample was processed in six technical replicates for the cerebellum and skeletal muscle. Total RNA was extracted for each sample using either an organic isolation followed by a cleanup or a silica-based column isolation. Then RNA extracts were divided into six groups (three tissues times two isolation methods). A unique RT mix was used for each group in two conditions: constant RNA amount (CA) and constant RNA extract volume (CV). Each group (one tissue and one isolation method) was referred to one RT set. All experiments were approved by the ethical committee on animal research care of the French Army Medical Research Center in accordance with applicable French and European Community regulations. After collection, samples were immediately placed in 1 ml of RNeasy lysis reagent (Qiagen), kept at 4 °C for 24 h, and then stored at –20 °C until use.

### RNA isolation

Samples (8 mg spleen, 48 mg cerebellum and skeletal muscle) were then homogenized in guanidinium-based lysis buffer (50 mg/ml, RLT buffer, Qiagen) using a Mixer Mill MM300 device (30 Hz, 2 × 2 min, one 5-mm steel bead, Rescht). For the cerebellum and the skeletal muscle tissues, the lysates were aliquoted in six parts and the volume was adjusted to 600 µl using RLT lysis buffer. Total RNA was then isolated using a QIAcube instrument (Qiagen) according to the manufacturer's instructions using both a silica-based column isolation with an additional DNase step (RNeasy Mini Kit, Qiagen) and an organic isolation followed by a silica-based cleanup (Lipid Tissue Mini Kit, Qiagen). RNA extracts were eluted with 50 µl of RNase-free water.

### Control of RNA quality

RNA quantification and purity were determined in buffer at neutral pH by spectrophotometric analysis (Nanodrop 1000, ThermoFisher Scientific). RNA integrity was assessed using the RNA integrity number (RIN) with a 2100 Bioanalyzer instrument (RNA Nano 6000 LabChip Kit, Agilent Technologies).

### cDNA synthesis

Reverse transcription was performed using a Reverse Transcription Core Kit (Eurogentec) according to the manufacturer's instructions from 200 ng of RNA with 2.5 µM oligo-dT<sub>15</sub> primers and RNase inhibitor in a final volume of 10 µl. Complementary DNA (cDNA) synthesis was realized using either 200 ng of total RNA (CA-RT) or a constant volume of total RNA extract (≤200 ng total RNA) supplemented with yeast tRNA (Ambion) to obtain 200 ng of RNA (CV-RT). In CV-RT, the volume of RNA extract processed matched the volume of the most concentrated sample containing 200 ng of total RNA. To minimize pipetting variability, crude RNA extracts were diluted to use minimal test volumes of 4 µl.

### Primer optimization

Primer design was done with MacVector software (MacVector). The specificity of amplification was checked comparing recombinant DNA with unknown products by melting curve analysis as described previously [14]. The sequences and characteristics of selected primers are listed in part 1 of the [Supplementary material](#). Primer optimization was carried out separately for each RT set as shown in part 2 of the [Supplementary material](#). Briefly, a first primer optimization assay was completed using a 4-log range serial dilution of recombinant DNA and adjusted by using serial dilutions (1-log range) of a pooled sample (calibrator) of the experiment. All assays were carried out using two duplicates per point.

### qPCR

qPCR was carried out with an LC Fast Start DNA Master SYBR Green I Kit in a final volume of 20 µl on a capillary LightCycler instrument (Roche) using 0.25 µl of undiluted cDNA with settings described in part 2 of the [Supplementary material](#). Quantification cycles ( $C_q$ ) [2] were determined using the second derivative maximum method (LightCycler software, version 3.5, Roche). Quantification was achieved by exponential conversion of the  $C_q$  using the PCR efficiency estimated from standard curves for each RT set. To minimize pipetting variability, cDNAs were diluted to use minimal test volumes of 5 µl.

### PCR inhibition

PCR inhibition was specifically assessed using three inhibition assays based on alien DNA template amplification, Spud-A [15], and two laboratory-conceived assays, EGFP and 14 kDa, designed using Primer3 software (<http://frodo.wi.mit.edu/primer3>). Primers and probes are listed in part 3 of the [Supplementary material](#). Recombinant templates (~10<sup>4</sup> copies) were added to 0.25 µl of cDNA and amplified for 40 cycles at 95 °C for 20 s and 60 °C for 30 s (LC FastStart DNA Master HybProbe Kit, Roche, 4 mM MgCl<sub>2</sub>, 0.4 µM primer, and 0.2 µM probe) in a capillary LightCycler instrument.

### Statistical analysis

Statistical analysis was performed with Prism version 4.0 (GraphPad Software). Comparisons were made employing the Wilcoxon test. Linear correlation was evaluated using the Spearman's

rank test. Significance was accepted at  $P < 0.05$ . The agreement between two different assays was assessed using Bland–Altman ratio comparisons. Values are presented as means  $\pm$  standard deviations. Variability is expressed as relative standard deviation (RSD), that is, the absolute value of the ratio of the standard deviation to the mean.

## Results

### RNA quality assessment

RIN values were  $7.76 \pm 0.18$ ,  $8.35 \pm 0.15$ , and  $7.98 \pm 0.56$  for the spleen, cerebellum, and skeletal muscle, respectively. The higher standard deviation observed for skeletal muscle was due to slight differences between organic and silica-based RNA isolation ( $7.55 \pm 0.39$  vs.  $8.42 \pm 0.30$ ). These high integrity values attested that all of the RNA samples were suitable for RT-qPCR studies [16]. The low standard deviation values reflected the homogeneity of each experimental set. RNA purity was assessed using the absorbance  $A_{260}/A_{280}$  ratio (R ratio). The R ratios were  $1.98 \pm 0.07$  and  $2.05 \pm 0.06$  for organic and silica-based RNA isolation methods, respectively. The level of these values attested to the high purity and homogeneity of all the samples. As expected, the difference between R ratios ( $P < 0.001$ ) pointed to traces of aromatic compounds in RNA extracts obtained using the organic isolation method. As shown in Table 1, RNA isolation yields depended on both the tissue and isolation method. For all tissues, the organic isolation method allowed higher isolation yields than the silica-based isolation method ( $P < 0.001$ ,  $P < 0.01$ , and  $P < 0.01$  for the spleen, cerebellum, and skeletal muscle, respectively) but led to higher variability. Variability was higher in skeletal muscle and the spleen than in the cerebellum. The ranges of the volume of RNA extracts used to carry over CA-RT between samples allowed useful comparisons between CA-RT and CV-RT methods (Table 1).

### Assessment of RT quality

We divert from original use the previously described 5'–3' integrity assay to assess RT efficiency [17]. In this assay, the RT reaction is primed using oligo-dT and separate PCR assays are used to quantify the levels of two target amplicons that are spatially separated, with one toward the 5' end and the second toward the 3' end of a single mRNA sequence. Thus, the ratio of amplicons reflects the relative success of the oligo-dT primed RT to proceed along the entire length of the transcript. Classically, a decrease of this ratio provides a measure of degradation of the mRNAs. However, it assesses the RT efficiency only when using a single RNA extract in different reaction conditions. Two primer sets were designed in the respective 3' and 5' parts of the ARBP mRNA. A significant RNA secondary structure was predicted between the two amplicons using the Mfold algorithm [18].

For each sample, the 5'–3' ratios were calculated on exponential conversion of the  $C_q$  values for both the CA-RT set (5'–3'CA) and the

CV-RT set (5'–3'CV). Thus, for each sample, the ratio 5'–3'CA/5'–3'CV assessed the relative RT efficiencies between CA and CV RT reactions. As shown in Fig. 1A, there was no significant difference between CA-RT and CV-RT efficiencies in four of six experimental groups. However, the CA-RT efficiency was lower than the CV-RT efficiency when using the organic RNA isolation method in the spleen and the column RNA isolation method in the cerebellum (46% and 14% decrease of RT efficiencies, respectively). It is noteworthy that the relative range of RNA concentrations was higher in the spleen organic set (1.0–2.7) than in the cerebellum column set (1.0–1.5). The variability of the CA-RT-induced inhibition was also higher in the spleen (RSD = 43%) than in the cerebellum (RSD = 10%). Moreover, the CA-RT inhibition was significantly correlated with the volume of the RT template, as shown in Fig. 1B and C (spleen: Spearman  $\rho = -0.764$ ,  $P < 0.001$ ; cerebellum: Spearman  $\rho = -0.986$ ,  $P < 0.001$ ). Such correlations demonstrate that some RNA extracts contain volume-related inhibitors of RT.

### qPCR inhibition assessment

qPCR inhibition was specifically assessed using three universal inhibition assays based on alien DNA template amplification according to MIQE (minimum information for publication of quantitative real-time PCR experiments) guidelines [2]. All spleen samples were analyzed. CA-RT led to significant inhibition in comparison with CV-RT in five of six cases (Fig. 2). As stated previously, qPCR inhibition was both assay and template dependent [19]. In organic isolated samples, qPCR inhibition was significantly higher in CA samples than in CV samples through the three assays ( $P < 0.05$ ,  $P < 0.05$ , and  $P < 0.01$  for Spud-A, EGFP, and 14 kDa, respectively). However, the levels of qPCR inhibition were low compared with the aforementioned levels of RT inhibition; the decreases of measured alien copy number between CV and CA samples were 1.9%, 1.2%, and 5.1% for Spud-A, EGFP, and 14 kDa, respectively. Similar results were observed in column-extracted samples with significant CA-RT-induced qPCR inhibition ( $P < 0.01$  and  $P < 0.001$  for EGFP and 14 kDa, respectively) with low levels of qPCR inhibition (3.2% and 3.7% decrease of measured alien copy number between CV and CA for EGFP and 14 kDa, respectively). In these samples, no differences were observed using Spud-A assay.

### Gene quantification accuracy

PPIA/HPRT mRNA ratios were assessed in the cerebellum and skeletal muscle using technical replicates. In such samples, the most accurate quantification corresponded to the lower ratio variability. As shown in Fig. 3A, the variability was lower for CV-RT than for CA-RT. In skeletal muscle sets and for the cerebellum column set, the RSD for CV-RT represented approximately half that of the CA-RT (muscle: organic 16.7% vs. 8.1%, column 11.4% vs. 5.8%; cerebellum: column 22.0% vs. 12.0%). For the cerebellum organic set, RSDs were close between CA-RT (12.9%) and CV-RT (10.7%).

**Table 1**  
RNA isolation yield.

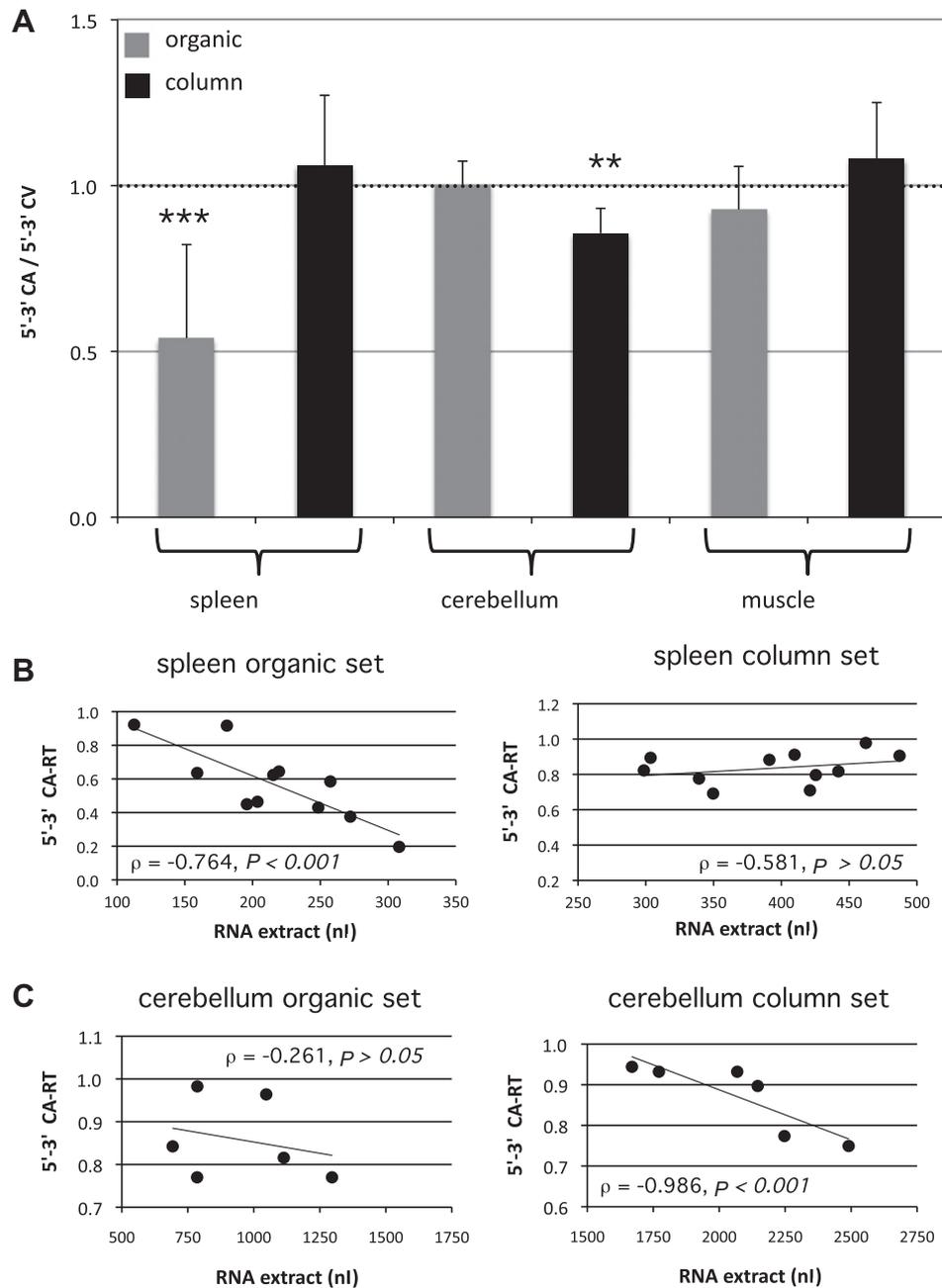
Tissue	Organic isolation yield			Column isolation yield		
	(ng/ $\mu$ l) <sup>a</sup>	RSD <sup>b</sup> (%)	Volume range <sup>c</sup>	(ng/ $\mu$ l) <sup>a</sup>	RSD <sup>b</sup> (%)	Volume range <sup>c</sup>
Spleen	978 $\pm$ 303	31	1.0–2.7	537 $\pm$ 101	19	1.0–1.7
Cerebellum	220 $\pm$ 53	24	1.0–1.9	99 $\pm$ 15	15	1.0–1.5
Skeletal muscle	123 $\pm$ 45	37	1.0–2.8	65 $\pm$ 15	23	1.0–1.6

Note: Total RNA was isolated from 50 mg of tissue and eluted in 50  $\mu$ l of RNase-free water.

<sup>a</sup> Total RNA concentration measured using  $A_{260}$  absorbance. Results are means  $\pm$  standard deviations.

<sup>b</sup> Variability is expressed as relative standard deviation (RSD), that is, the absolute value of the ratio of the standard deviation to the mean.

<sup>c</sup> Range of the volume of RNA extracts between samples used to carry out a constant amount of RT. For each tissue, the RNA yield is significantly greater for organic isolation than for column isolation ( $P < 0.01$  for each tissue).

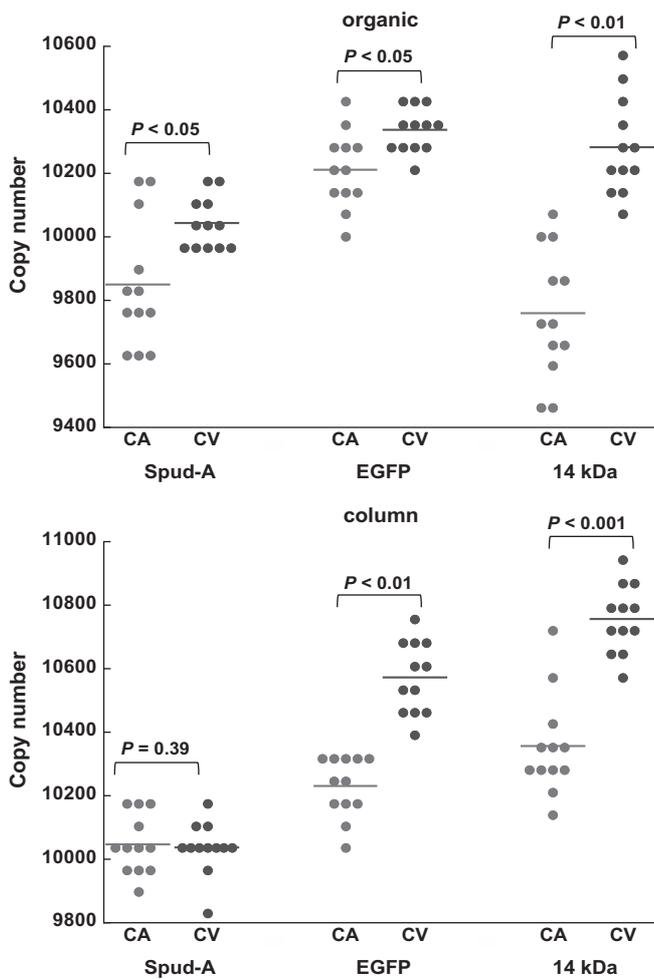


**Fig. 1.** RT inhibition. (A) 5'-3'CA/5'-3'CV: average of the ratios of RT efficiencies between CA-RT and CV-RT of each sample using 5'-3' qPCR assay for the whole tissues. The ratio of the 5' amplicon versus the 3' amplicons reflects the relative success of the oligo-dT primed RT to proceed along the entire length of the transcript. The 5'-3'CA/5'-3'CV ratio assessed the relative RT efficiencies between CA and CV RT reactions. Statistical comparisons were made employing the Wilcoxon test. \*\*\*\* $P < 0.001$ ; \*\*\* $P < 0.01$ . (B and C) Linear correlation between the RNA extract volume and the 5'-3' ratio for CA-RT: (B) spleen extracts; (C) cerebellum extracts. Statistical significance is assessed using the Spearman's rank correlation test. The volumes represent the amount of crude RNA extracts; the templates were diluted to use minimal test volumes of 5  $\mu$ l to minimize pipetting variability.

Because technical variability in RT-qPCR depends on the mRNA species [20], we use a multigenic index to assess the reliability of the quantification. Such an index should be provided using the normalization software geNorm [21]. This normalization method is based on the geometric mean of selected reference genes. The accuracy of geNorm normalization is estimated from the pairwise variation  $V_{n/n+1}$  that must be below a 0.15 threshold. The  $V_{n/n+1}$  value depends on both the physiological stability of reference genes and the RT-qPCR quality. ARBP, HPRT1, and PPIA mRNAs were used to normalize identical samples with different technical processing. In such a case, the  $V_{2/3}$  values assessed the reproducibility and

accuracy of the multistage processing from RNA isolation to qPCR analysis. For the whole RT sets, CV-RT led to lower  $V_{2/3}$  values than CA-RT (Fig. 3B). However, the RT processing had only slight effects on the stability ranking of the reference genes (see part 4 of Supplementary material).

For the spleen and cerebellum, variations were consistent with differences in RT efficiencies described above (Fig. 1). In comparison with the CV-RT  $V_{2/3}$  values, the CA-RT  $V_{2/3}$  values were 86% and 76% higher for spleen organic and cerebellum column, respectively. There was only an 11% increase for each of the other RT sets. Although there were no differences between CA-RT and CV-RT



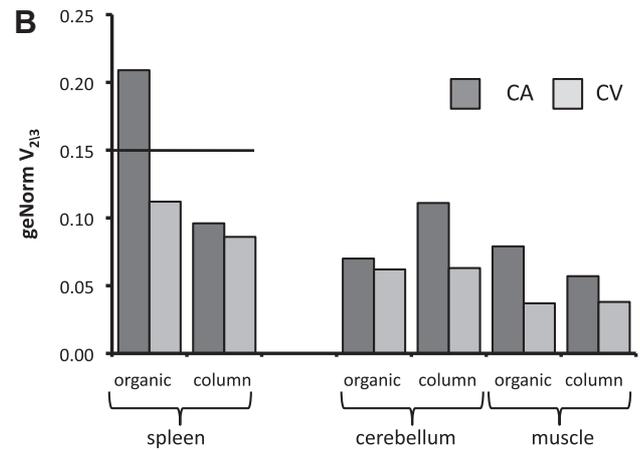
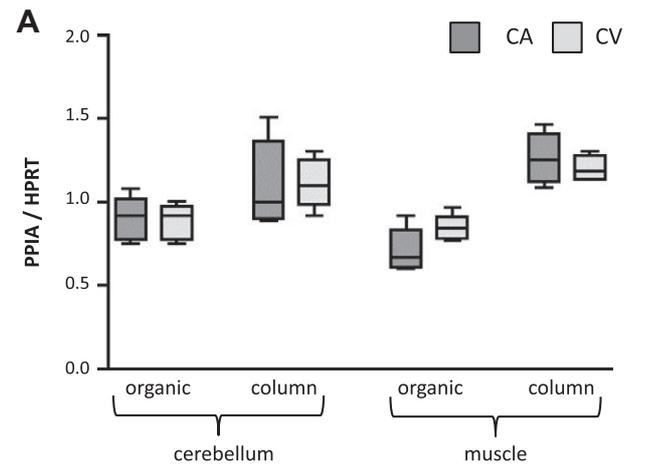
**Fig. 2.** qPCR inhibition assessment using Spud-A, EGFP, and 14 kDa alien inhibition assays for each spleen sample RT set. Approximately  $10^4$  copies of alien recombinant templates were added to 0.25  $\mu$ l of cDNA and amplified in a capillary LightCycler instrument. Statistical comparisons were made employing the Wilcoxon test.

efficiencies for skeletal muscle, the  $V_{2/3}$  values were higher in CA-RT than in CV-RT for both organic and silica-based RNA isolations (113% and 50%, respectively). It is noteworthy that the CV-RT  $V_{2/3}$  values were nearly identical between organic and column RNA isolations when using sample replicates (cerebellum and muscle). However, for independent samples (spleen), these values were slightly different (30% variation).

For the spleen (independent samples), the selected normalization was validated except for the organic CA-RT ( $V_{2/3} = 0.21$ ). Very low  $V_{2/3}$  values were observed for the cerebellum and skeletal muscle due to the use of technical replicates of a single sample.

#### Agreement between technical processes using biological replicates

CD19 mRNA levels were assessed in all spleen RT sets and normalized using geNorm as mentioned earlier. For each sample, the agreement between CD19 measurements in different RT sets was assessed using the Bland–Altman ratio comparison (Fig. 4). These comparisons were carried out using either the same RT mix (column and organic) or two different RT mixes (CA-RT and CV-RT). This point must be highlighted given that RT is the most variable step in RT-qPCR [4]. The best agreement was observed between samples generated using constant RNA extract volumes (CV-RT sets), although two different RT mixes were used. It is noteworthy

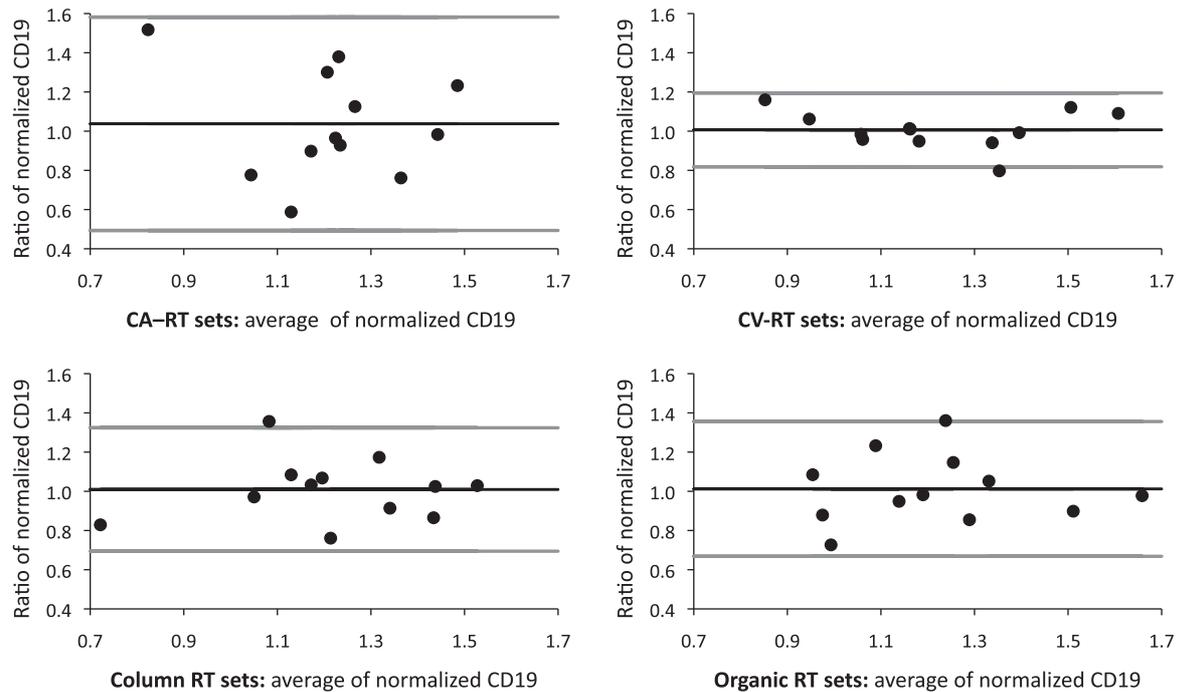


**Fig. 3.** (A) Variability of the relative expression of PPIA versus HPRT in the cerebellum and skeletal muscle. There were no significant differences between CA-RT and CV-RT for all of the RT sets ( $P > 0.05$ , Wilcoxon test). The boxes represent the second and third quartiles, and the whiskers symbolize the lower and the higher values. (B) geNorm pairwise variation ( $V_{2/3}$ ): multiparametric normalization index calculated using ARBP, HPRT1, and PPIA as reference genes for each RT set. For the spleen (independent samples), the 0.15 threshold represents the level of the pairwise variation that requires the inclusion of an additional normalization gene. For the cerebellum and skeletal muscle (technical replicates), the lowest  $V_{2/3}$  value corresponds to the most accurate RT-qPCR quantification.

that the comparisons between samples generated using the same RT mix, but using different volumes of RNA extract, showed worse agreements. The worst agreement was observed between samples generated using both different RT mixes and variable RNA extract volumes (CA-RT sets). Such a result may be explained by both differences in RT efficiencies between organic CV-RT and CA-RT sets (Fig. 1) and the nonvalidated geNorm normalization in the organic group.

#### Discussion

The standardization of the RT reaction is a key factor in the reliability of RNA quantification. Because it was reported that RT efficiency depends on the RNA concentration [4], the current consensus is that a constant amount of RNA (CA-RT) must be used in the RT step. However, due to the variability of the RNA isolation yields, such an approach entails the use of variable volumes of RNA extract. Thus, we propose to improve the standardization of the RT step using a constant volume of RNA extract (CV-RT) while counteracting the differences in the RNA concentration using yeast tRNA as described previously [4].



**Fig. 4.** Quantification of CD19 mRNA levels in spleen: agreement between CD19 measurements using Bland–Altman ratio comparison among CA-RT sets, CV-RT sets, RT sets, and organic RT sets. The abscissa and ordinate represent the average and ratio of CD19 mRNA levels between the tested sets, respectively. Black line: bias; gray lines: 95% limits of agreement.

Our finding points out that some RNA extracts contain inhibitors that increase the intersample variability of RT-qPCR. Moreover, the content of such inhibitors is not related to the total RNA amount but certainly to both the amount of processed tissue and the elution volume of the RNA isolation step. It is noteworthy that such an inhibition is significant within a low 3-fold range in volume of RNA extract. Such ranges are significantly lower than the 100-fold template ranges used in previous studies on the properties of RT [4,7]. Some of these inhibitors are related to the technical process. In the spleen, the prior organic isolation step was a source of exogenous contaminants with significant volume-related inhibitory effects. Conversely, some inhibitors are endogenous. In the cerebellum, the column isolation did not permit the removal of endogenous volume-related inhibitors, probably being fat. The use of CV-RT decreases the intersample difference in endogenous inhibitors while homogenizing the intersample concentration of exogenous inhibitors.

It is noteworthy that the CV-RT practice allows us to decrease the intersample variation only in volume-related inhibitors. Such inhibitors are characterized by a nearly constant concentration in all of the samples. Due to biological variability, some samples may contain sample-related inhibitors whose concentrations vary between samples. The intersample variation of the concentration of such inhibitors is not abolished using either CV-RT or CA-RT. This hypothesis can be illustrated by the comparison between biological and technical replicates (Fig. 1B and C). The correlation between the RT inhibition and the sample volume was higher for the cerebellum column set (technical replicates with identical sample-related inhibitors) than for the spleen organic set (biological replicates). However, as shown by the high Spearman coefficient value ( $\rho = -0.764$ ), the volume-related inhibitors certainly had more impact on RT efficiency than the sample-related inhibitors in the spleen organic set. Moreover, sample-related inhibitors inhibit the RT step only if they are not removed by the isolation step. In the spleen column set (biological replicates), most sample-related inhibitors have been removed, as shown by the low RSD value of RT inhibition (18% as calculated from Fig. 1B). This is similar to

the RSD value of RT inhibition for the cerebellum organic RT set (technical replicates with identical sample-related inhibitors, RSD = 24% as calculated from Fig. 1C).

The CA-RT-induced variability was essentially due to intersample variations in RT efficiency. Conversely, only slight variations in PCR efficiencies were observed. Such differences could be explained by the theoretical differences between RNA to cDNA synthesis conditions and cDNA to DNA synthesis conditions. DNA synthesis could at least be impaired by primer hybridization disorders, enzymatic inhibition, or single-strand nucleic acid secondary structures. For cDNA to DNA synthesis, the upshot of either enzymatic inhibition or secondary structures might be unlikely due to the short size of the amplicons regarding the polymerization time or to the high incubation temperature, respectively. Thus, primer hybridization disorder-induced inhibition might be a significant mechanism of PCR inhibition. In contrast, in RNA to cDNA synthesis, enzymatic inhibition and secondary structures upshot are more likely due to both transcript length and low incubation temperature. An accurate assessment of the RT inhibition is challenging because this measurement is made after the qPCR step. Such an approach converts a continuous phenomenon (the length of synthesized cDNA) to a binary result (the target of the primers either is or is not reverse-transcribed). Moreover, RNA secondary structures are known to influence the RT yield [22,23] and, thus, cause RT-related differences among RNAs [4]. Thus, the 5'–3' decrease in RT efficiency observed in both the spleen organic set and the cerebellum column set could be facilitated by the significant secondary structure predicted between the 3' and 5' amplicons. Such a result was obtained using an oligo-dT priming strategy. It should be equally assessed using another priming strategy of the RT, especially random priming.

The effects of RT inhibition on the RT-qPCR reliability are easier to evaluate. In particular, the relative expression of two genes using technical replicates enables one to assess the impact variability induced by CA-RT on mRNA quantification (Fig. 3A). However, because features (e.g., structure, length, composition) can vary extensively between mRNAs, the potential RT inhibition

could be gene specific [20]. Thus, we used a multiparametric index based on the assessment of multiple genes. Indeed, comparisons of the geNorm  $V_{n/n+1}$  values [21] using both technical and biological replicates appear to be a very sensitive way of evaluating the pre-analytical process (Fig. 4). Thus, in the muscle samples, CV-RT significantly decreased the RT-qPCR variability, for both column and organic isolations, compared with CA-RT. Such a difference indicates that these two isolation methods fail to remove the volume-related inhibitors present in the muscle samples. Therefore, another protocol allowing us to decrease the difference in RT-qPCR variability between CA-RT and CV-RT should be used. Conversely, the similarity of the geNorm  $V_{n/n+1}$  values among CA-RT and CV-RT sets demonstrates the effectiveness of the protocol to eliminate volume-related inhibitors, as shown in both spleen column and cerebellum organic sets. Despite this similarity, in spleen sets the best agreement was observed between CV-RT sets, although RNA isolations and reverse transcription mixes were different. Such a result demonstrates the negative impact of the practice of the usual CA-RT on the reliability of RT-qPCR data.

In summary, the RT reaction is expected to be the most uncertain step in RNA quantification. Due to the variability of RNA isolation yields, the achievement of RT using constant amounts of total RNA increases intersample RT-qPCR variability. Such drawbacks are caused by the presence of volume-related RT inhibitors in the RNA extracts. However, optimization of the RNA isolation protocol to suppress the volume-related inhibitors of the samples allows us to avoid this procedure, although the intersample variability remains higher. Indeed, the use of a constant volume of RNA extracts as RT template allows one to decrease the intersample differences in both endogenous inhibitors and exogenous inhibitors. Such a decrease, which is in the spirit of the MIQE guidelines [2], is a prerequisite for the reliability of RT-qPCR assays. Thus, CV-RT provides a relevant approach for the standardization of the RT step when there are valid concerns about the reliability of RT-qPCR assays [24].

## Acknowledgments

This work was supported by grants from “Delegation Générale pour l’Armement.” The authors thank Gabrielle Cormier for careful reading of the manuscript.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ab.2011.04.008.

## References

- [1] S.A. Bustin, Why the need for qPCR publication guidelines? The case for MIQE, *Methods* 50 (2010) 217–226.
- [2] S.A. Bustin, V. Benes, J.A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M.W. Pfaffl, G.L. Shipley, J. Vandesompele, C.T. Wittwer, The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments, *Clin. Chem.* 55 (2009) 609–620.
- [3] S.A. Bustin, Real-time, fluorescence-based quantitative PCR: a snapshot of current procedures and preferences, *Expert Rev. Mol. Diagn.* 5 (2005) 493–498.
- [4] A. Stahlberg, J. Hakansson, X. Xian, H. Semb, M. Kubista, Properties of the reverse transcription reaction in mRNA quantification, *Clin. Chem.* 50 (2004) 509–515.
- [5] J. Curry, C. McHale, M.T. Smith, Low efficiency of the Moloney murine leukemia virus reverse transcriptase during reverse transcription of rare t(8;21) fusion gene transcripts, *BioTechniques* 32 (2002) 754–772.
- [6] E.E. Karrer, J.E. Lincoln, S. Hogenhout, A.B. Bennett, R.M. Bostock, B. Martineau, W.J. Lucas, D.G. Gilchrist, D. Alexander, In situ isolation of mRNA from individual plant cells: creation of cell-specific cDNA libraries, *Proc. Natl. Acad. Sci. USA* 92 (1995) 3814–3818.
- [7] J.P. Levesque-Sergerie, M. Duquette, C. Thibault, L. Delbecchi, N. Bissonnette, Detection limits of several commercial reverse transcriptase enzymes: impact on the low- and high-abundance transcript levels assessed by quantitative RT-PCR, *BMC Mol. Biol.* 8 (2007) 93.
- [8] I.G. Wilson, Inhibition and facilitation of nucleic acid amplification, *Appl. Environ. Microbiol.* 63 (1997) 3741–3751.
- [9] L. Belec, J. Authier, M.C. Eliezer-Vanerot, C. Piedouillet, A.S. Mohamed, R.K. Gherardi, Myoglobin as a polymerase chain reaction (PCR) inhibitor: a limitation for PCR from skeletal muscle tissue avoided by the use of *Thermus thermophilus* polymerase, *Muscle Nerve* 21 (1998) 1064–1067.
- [10] B.M. Kemp, C. Monroe, D.G. Smith, Repeat silica extraction: a simple technique for the removal of PCR inhibitors from DNA extracts, *J. Archaeol. Sci.* 33 (2006) 1680–1689.
- [11] B. Mercier, C. Gaucher, O. Feugeas, C. Mazurier, Direct PCR from whole blood, without DNA extraction, *Nucleic Acids Res.* 18 (1990) 5908.
- [12] L. Rossen, P. Norskov, K. Holmstrom, O.F. Rasmussen, Inhibition of PCR by components of food samples, microbial diagnostic assays, and DNA-extraction solutions, *Intl. J. Food Microbiol.* 17 (1992) 37–45.
- [13] D.L. Wiedbrauk, J.C. Werner, A.M. Drevon, Inhibition of PCR by aqueous and vitreous fluids, *J. Clin. Microbiol.* 33 (1995) 2643–2646.
- [14] A. Peinnequin, C. Mouret, O. Birot, A. Alonso, J. Mathieu, D. Clarencon, D. Agay, Y. Chancerelle, E. Multon, Rat pro-inflammatory cytokine and cytokine related mRNA quantification by real-time polymerase chain reaction using SYBR green, *BMC Immunol.* 5 (2004) 3.
- [15] T. Nolan, R.E. Hands, W. Ogunkolade, S.A. Bustin, SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations, *Anal. Biochem.* 351 (2006) 308–310.
- [16] S. Fleige, M.W. Pfaffl, RNA integrity and the effect on the real-time qRT-PCR performance, *Mol. Aspects Med.* 27 (2006) 126–139.
- [17] T. Nolan, R.E. Hands, S.A. Bustin, Quantification of mRNA using real-time RT-PCR, *Nat. Protoc.* 1 (2006) 1559–1582.
- [18] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res.* 31 (2003) 3406–3415.
- [19] J.F. Huggett, T. Novak, J.A. Garson, C. Green, S.D. Morris-Jones, R.F. Miller, A. Zumla, Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon, *BMC Res. Notes* 1 (2008) 70.
- [20] A. Stahlberg, M. Kubista, M. Pfaffl, Comparison of reverse transcriptases in gene expression analysis, *Clin. Chem.* 50 (2004) 1678–1680.
- [21] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, *Genome Biol.* 3 (2002) RESEARCH0034.
- [22] E.M. Brooks, L.G. Sheflin, S.W. Spaulding, Secondary structure in the 3' UTR of EGF and the choice of reverse transcriptases affect the detection of message diversity by RT-PCR, *BioTechniques* 19 (1995) 806–812. 814–815.
- [23] K.W. Kuo, M.F. Leung, W.C. Leung, Intrinsic secondary structure of human TNFR-I mRNA influences the determination of gene expression by RT-PCR, *Mol. Cell. Biochem.* 177 (1997) 1–6.
- [24] J. Murphy, S.A. Bustin, Reliability of real-time reverse-transcription PCR in clinical diagnostics: gold standard or substandard?, *Expert Rev. Mol. Diagn.* 9 (2009) 187–197.



# Pitfalls in target mRNA quantification for real-time quantitative RT-PCR in overload-induced skeletal muscle hypertrophy

T. Chaillou,<sup>1</sup> A. Malgoire,<sup>1</sup> S. Banzet,<sup>1</sup> R. Chapot,<sup>1</sup> N. Koulmann,<sup>1</sup> P. Pugnère,<sup>2</sup> M. Beaudry,<sup>3</sup> X. Bigard,<sup>1</sup> and A. Peinnequin<sup>2</sup>

<sup>1</sup>Operational environments, <sup>2</sup>Genomic core facility, IRBA La Tronche, BP87, La Tronche; and <sup>3</sup>Laboratoire Réponses cellulaires et fonctionnelles à l'hypoxie, Université Paris13, EA2363, Bobigny, France

Submitted 30 May 2010; accepted in final form 12 December 2010

**Chaillou T, Malgoire A, Banzet S, Chapot R, Koulmann N, Pugnère P, Beaudry M, Bigard X, Peinnequin A.** Pitfalls in target mRNA quantification for real-time quantitative RT-PCR in overload-induced skeletal muscle hypertrophy. *Physiol Genomics* 43: 228–235, 2011. First published December 14, 2010; doi:10.1152/physiolgenomics.00109.2010.—Quantifying target mRNA using real-time quantitative reverse transcription-polymerase chain reaction requires an accurate normalization method. Determination of normalization factors (NFs) based on validated reference genes according to their relative stability is currently the best standard method in most usual situations. This method controls for technical errors, but its physiological relevance requires constant NF values for a fixed weight of tissue. In the functional overload model, the increase in the total RNA concentration must be considered in determining the NF values. Here, we pointed out a limitation of the classical geNorm-derived normalization. geNorm software selected reference genes despite that the NF values extensively varied under experiment. Only the NF values calculated from four intentionally selected genes were constant between groups. However, a normalization based on these genes is questionable. Indeed, three out of four genes belong to the same functional class (negative regulator of muscle mass), and their use is physiological nonsense in a hypertrophic model. Thus, we proposed guidelines for optimizing target mRNA normalization and quantification, useful in models of muscle mass modulation. In our study, the normalization method by multiple reference genes was not appropriate to compare target mRNA levels between overloaded and control muscles. A solution should be to use an absolute quantification of target mRNAs per unit weight of tissue, without any internal normalization. Even if the technical variations will stay present as a part of the intergroup variations, leading to less statistical power, we consider this method acceptable because it will not generate misleading results.

reference genes; normalization factors; total RNA; muscle mass

REAL-TIME reverse transcription-polymerase chain reaction (RT-qPCR) is the gold standard technique used to assess variations in mRNA expression. Specific guidelines have been recently published to evaluate RT-qPCR experiments (4), and it is now clear that a validated normalization strategy is imperative to obtain reliable results.

Normalization using internal reference genes is the most frequently used strategy to detect changes in target mRNA levels (11). It aims to correct variations related to mRNA multistage processes (i.e., extraction yield, reverse-transcription yield, and sample volume). A number of reports recently demonstrated that classical reference genes can vary extensively in cells or tissues and may be influenced by physiological adaptations or experimental conditions (7, 13). Therefore,

using a single internal standard is not suitable to obtain a reliable mRNA quantification. The use of multiple internal standards is commonly accepted and appears to be the most robust normalization method to date (4). Various methods determine the most reliable reference genes, based on their relative stability, using their geometric mean (2, 22, 28). Normalization factors (NFs), based on the selected genes, correct for technical variations that should, on average, be similar for all samples, irrespective of the experimental group. Therefore, the NF values should be constant for all experimental groups if the reference mRNAs content per unit weight of tissue (or per unit cell volume) is not influenced by experimental conditions. If this is not the case, the geometric mean of the selected reference genes will vary between experimental groups. Thus, the quantification of target mRNAs (i.e., absolute target mRNAs values per unit weight of tissue) will generate inconsistent results. Since reverse transcription is performed using a constant amount of total RNA due to technical reasons (4), it seems physiologically relevant to consider changes in the total RNA concentration to determine the NF values for a fixed weight of tissue.

The control of skeletal muscle growth is a physiological domain that now receives important scientific attention. Various *in vivo* models of altered skeletal muscle mass exist (i.e., muscle regeneration, functional overload, electrical stimulation, hormonal administration, unweighting). Functional overload is a commonly used model of drastic muscle hypertrophy (3, 17). Muscle hypertrophy induced by mechanical overload involves adaptive responses of target genes that were recently studied by RT-qPCR analysis (6, 12, 15, 16, 25, 27, 29). In these studies, only a single reference gene was used for internal normalization, and overload-induced changes in the total RNA concentration were not considered in the target mRNAs quantification (1, 14).

The aim of the present study was to propose an appropriate target mRNAs quantification method from RT-qPCR assay, usable in models of drastic muscle hypertrophy. Here, we pointed out a limitation of the classical geNorm-derived normalization for target mRNAs quantification. Indeed, our findings allowed us to select stable reference genes using geNorm software, although the NF values based on these genes were not constant between experimental groups. Therefore, we proposed guidelines to avoid inappropriate quantification in models of muscle mass modulation.

## MATERIALS AND METHODS

### *Experimental Protocol and Surgical Procedures*

Twenty-six female rats of Wistar Han strain, initially weighing 180–200 g, were purchased from Charles River Laboratories (L'Arbresle, France). The animals were housed one per cage in a

Address for reprint requests and other correspondence: A. Peinnequin, IRBA La Tronche, BP 87, 38702 La Tronche Cedex, France (e-mail: andrepeinnequin@crssa.net).

thermoneutral environment (22 ± 2°C) on a 12:12-h photoperiod and were provided with food and water ad libitum. This investigation was carried out in accordance with the Helsinki Accords for Human Treatment of Animals During Experimentation and was approved by the local Animal Ethics Committee. They were assigned to either a control (C, n = 12) or an overload group (OV, n = 14).

After 7 days of acclimatization to the animal room, rats were submitted to a bilateral functional overload of plantaris muscle. Animals were anesthetized with pentobarbital sodium (50 mg/kg), and compensatory overload surgery was accomplished by the bilateral removal of plantaris synergists, as previously described (3). In brief, a small incision was made to the posterior lower limb, and the entire soleus, as well as the major portion of gastrocnemius, was carefully removed. Particular attention was made to ensure that the plantaris neural and the vascular supplies were not damaged, and then the aponeurosis and the skin were sutured.

*Tissue Processing*

Five days (d5) (C<sub>d5</sub>, n = 6 and OV<sub>d5</sub>, n = 6) and 12 days (d12) after surgery (C<sub>d12</sub>, n = 6 and OV<sub>d12</sub>, n = 8), the rats were anesthetized with pentobarbital sodium (80 mg/kg), and the plantaris muscles were carefully dissected and weighed. The proximal half portion of right plantaris was immediately placed in 0.5 ml of RNALater (Ambion, Austin, TX), maintained at +4°C for 24 h, and then stored at -20°C until use. The animals were subsequently euthanized by a lethal injection of pentobarbital sodium.

*RNA Isolation and cDNA Synthesis*

Muscle samples were disrupted (100 mg/ml) with a Mixer Mill MM300 (Rescht, Haan, Germany) in RLT lysis buffer, 1% 2-mercaptoethanol (Qiagen, Courtabouef, France), following the manufacturer's instructions [30 Hz, 2 min; 3 mm ceramic beads (x2)]. Total RNA was isolated using RNeasy Lipid Tissue Mini-kit (Qiagen) with an additional DNase step according to the manufacturer's

instructions using a Qiacube system (Qiagen). The total RNA concentration and purity were assessed by measuring the optical density (230, 260, and 280 nm) with the Nanodrop 1000 Spectrophotometer (ThermoFisher Scientific, Wilmington, DE).

Reverse transcription was carried out in a 20 µl reaction volume from 800 ng total RNA, using the Reverse Transcriptase Core Kit (Eurogentec, Seraing, Belgium), with 50 µM oligo (dT) 15 primer and RNase inhibitor (4 UI), according to the manufacturer's instructions. To minimize intergroup variations (extraction and reverse transcription steps), the same number of samples within each experimental group was simultaneously processed. All samples were stored at -80°C until further analysis.

*Primer Optimization*

Primer design and optimization, concerning dimerization, self-priming, and melting temperature, were carried out using MacVector software (Accelrys, San Diego, CA). Primers were designed in intron flanking to exclude amplification of genomic DNA. Primers were then assessed for specificity, using Blastn algorithm (<http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi>), with search parameters adjusted for a short input sequence. Primer optimization was carried out as described previously (20). Briefly, a first primer optimization assay was performed, using a 4-log range serial dilution of recombinant DNA. Then, final optimization was achieved, using serial dilutions (1-log range) of a pooled sample (calibrator) of the experiment. All assays were carried out using two duplicates per point. Specificity was checked comparing recombinant DNA with unknown products, by melting curve analysis as described previously (20). Selected forward and reverse primers sequences and characteristics are listed in Table 1.

*Real-time qPCR and Quantification*

qPCR was carried out with LightCycler Fast Start DNA Master SYBR Green kit (Roche Applied Science, Mannheim, Germany),

Table 1. Characteristics of primers used in this experiment

Gene	5'-3' Sequence	Reference	Amplicon Length, bp	Annealing Temperature, °C	PCR Efficiency, %
β-actin	TCAGGTCATCACTATCGGCAATG TTTCATGGATGCCACAGGATT	NM_031144	95	58	93
Hprt1 (a)	CTCATGGACTGATTATGGACAGGAC GCAGGTCAGCAAGAAGACTTATAGCC	NM_012583	123	58	93
Hprt1 (b)	ATGA <sup>L</sup> GCACCT <sup>L</sup> CAGGGATTGAATC CGCT <sup>L</sup> GTCTTT <sup>L</sup> AGGCTTTGTA <sup>L</sup> CTTG	NM_012583	83	56	95
Ywhaz	AGACGGAGCTGAGGGACATCTG TTTCTGGCTGCGAAGCATTG	NM_013011	81	56	93
Arbp	CCTGCACACTCGCTTCC <sup>L</sup> TAGAG CAACAGTCGGGTAGCCAATCTG	NM_022402	74	57	96
Pol2ra	TGCTCGGGGAGATAGAGTCCAAGTT TTCAGGGTCATCTGAGTGGCAGGTT	XM_343922	112	57	96
Cyca	TATCTGCACTGCCAAGACTGAGTG CTTCTTGCTGGTCTTGCCATTCC	NM_017101	127	58	94
Tbp	GCCACGAACA <sup>L</sup> ACTGCGTTGAT AGCCCAGCTTCTGCACA <sup>L</sup> ACTCTA	NM_001004198	119	56	96
PGC-1α	ACGCAGGTGCAATGAAACTGAC TGGTGAAGCAGGGTCAAAATC	NM_031347	116	53	86
MAFbx	GCTTGTCGATGTTACCCAAGAA TGAAAGTGAGACGGGAGCAGCTCT	NM_133521	139	58	99
MURF1	GGGAACGACCGAGTT <sup>L</sup> CAGACTATC CCTTACCTGGTGGCTGTTTT <sup>L</sup> C	NM_080903	82	57	85
GDF8	GGGCATGATCTTGCTGTAACCTTC CGTGGAGTGTTCATCACAGTCAAG	NM_019151	132	59	91

β-Actin, beta-actin; Hprt1, hypoxanthine phosphoribosyltransferase; Ywhaz, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide; Arbp, acidic ribosomal phosphoprotein P0; Pol2ra, RNA polymerase II polypeptide A; Cyca, cyclophilin A; Tbp, TATA box binding protein; PGC-1α, peroxisome proliferator-activated receptor-γ coactivator 1 α; MAFbx, muscle atrophy F box; MURF1, muscle ring finger 1; GDF8, growth differentiation factor 8. Hprt1(a) 5' pair is used for normalization. Hprt1(b) 3' pair is used for 5'/3' ratio calculation (see MATERIALS AND METHODS). bp, Base pair. <sup>L</sup>, LNA substitution.

Table 2. Normalized plantaris mass

	C <sub>d5</sub>	OV <sub>d5</sub>	C <sub>d12</sub>	OV <sub>d12</sub>
Normalized plantaris mass, mg/g	0.89 ± 0.03	1.37 ± 0.04*	0.92 ± 0.03	1.67 ± 0.06*§

Data are presented as means ± SE. Normalized plantaris mass refers to the ratio between plantaris mass and body mass. OV, overload; d, day. \*Significantly different from the same time control (C) group, *P* < 0.001; §significantly different from d5 for the same group, *P* < 0.001.

using 0.25–0.5 µl of cDNA in a 20 µl final volume, 4 mM MgCl<sub>2</sub>, and 0.4 µM of each primer (except for β-actin, Arbp, GDF8, and reverse Tbp primer: 0.5 µM). qPCRs were performed, using Light-Cycler (Roche Applied Science) for 50 cycles at 95°C for 20 s, 53–59°C for 5 s (except β-actin, 10 s; annealing temperatures, see Table 1), and 8 s at 72°C. RNA integrity and reverse transcription yield were assessed by a 3'/5' integrity assay (18) with Hprt1(a) and Hprt1(b) selected primers (Table 1). Quantification cycles (Cq) were ascertained, using the second-derivative maximum method (LightCycler Software v 3.5, Roche Applied Sciences). All Cq values are within the standard curve. Relative quantification of samples was performed, using the comparative threshold method, corrected for amplification efficiency variability between genes (21), and improved for use of multiple reference genes (10, 11).

Selection of Reference Genes for Internal Normalization

Potential reference genes were assessed using geNorm software (28). The classical geNorm-derived normalization includes three steps:

1) The first step ranks the potential reference genes according to their relative expression stability. The relative expression stability of two genes *A* and *B* is defined as the variability (standard deviation after logarithmic transformation) of mRNA levels ratio within the samples (*V*<sub>AB</sub>). Logarithmic transformation is required to obtain identical *V*<sub>AB</sub> and *V*<sub>BA</sub> values. The relative expression stability of a gene *A* (*M*<sub>A</sub>) within a group of potential reference genes is defined as the arithmetic mean of *V*<sub>AX</sub> values, where *X* represents each of the other group of genes [for *n* genes (*A*–*N*): *M*<sub>A</sub> = (*V*<sub>AB</sub> + *V*<sub>AC</sub> + . . . *V*<sub>AN</sub>)/(*n* – 1)]. Thus, the average expression stability (*M*) of a group of reference genes is defined as the arithmetic mean of the *M*<sub>X</sub> values [calculation with *n* genes (*A*–*N*): *M* = (*M*<sub>A</sub> + *M*<sub>B</sub> + *M*<sub>C</sub> + . . . *M*<sub>N</sub>)/*n*].

2) The second step sets the number of genes required to validate the normalization. The number of selected genes is assumed as sufficient when the addition of the most stable remaining gene has no significant effect. The number of required genes is assessed using the pair-wise variation between normalization using the *n* and the *n*+1 most stable genes (*V*<sub>*n/n*+1</sub>). The *V*<sub>*n/n*+1</sub> values are calculated using the same formula than the stability expression of two genes (*V*<sub>AB</sub>). The threshold of the *V*<sub>*n/n*+1</sub> value, below which the inclusion of an additional control gene is not required, is 0.15 (28). It is noteworthy that geNorm software selects reference genes from their relative expression stability. It means that it considers their relative variation but not their absolute quantity.

3) The third step consists in calculating the NF for each sample. The NF value is the ratio between the geometric mean of the considered stable reference genes for a sample and the geometric mean of the considered stable reference genes for all samples (28).

Statistical Analysis

Morphologic data and RT-qPCR results are presented as means ± SE. These data were analyzed by two-way ANOVA, to determine the main statistical effects of functional overload, time, and interaction between these two factors. When appropriate, a Newman-Keuls post hoc test was used and a *P* value < 0.05 was considered significant.

RESULTS

Physiological Responses to Functional Overload

Plantaris hypertrophy in functional overload. The normalized muscle mass (expressed by the ratio muscle mass/body mass) was 54 and 82% higher in OV than in C rats at d5 and d12, respectively (*P* < 0.001) (Table 2). Moreover, normalized muscle mass increased by 22% in OV plantaris muscles between d5 and d12 (*P* < 0.001), and did not change in C muscles (overload × time interaction, *P* < 0.01). This finding confirmed that the surgical ablation of synergist muscles induced plantaris muscle hypertrophy.

Total RNA concentration. The total RNA concentration was significantly higher in OV than in C groups at both d5 and d12 (75%, *P* < 0.001 and 34%, *P* < 0.05, respectively) (Fig. 1). No time effect was shown on the total RNA concentration.

Utilization of geNorm Normalization in the Overload-induced Hypertrophy Model

Limitation of classical reference genes selection in this model: necessity to use constant normalization factors. The first normalization strategy consisted in testing genes commonly used as reference genes. Fig. 2A1 shows the *M* value of seven potential genes in all experimental samples. This analysis revealed that four reference genes [Tbp, Hprt1(a), Arbp and Pol2ra], constituted the group of the most stable genes (0.5 < *M* < 0.6), whereas the addition of the three others genes (Cyca, Ywhaz, and β-actin) slightly increased the *M* value (0.65 < *M* < 0.75).

*V*<sub>2/3</sub> (0.165) was > 0.15, but the inclusion of a fourth gene was sufficient to lower the stability factor below the acceptance

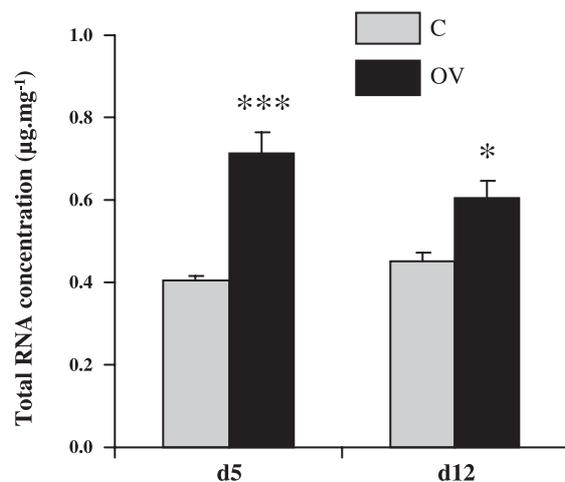


Fig. 1. Total RNA concentration (µg/mg) obtained by Nanodrop analysis. Values are means ± SE. C, control; OV, overload; d, day. Significantly different from the same time C group: \**P* < 0.05, \*\*\**P* < 0.001.

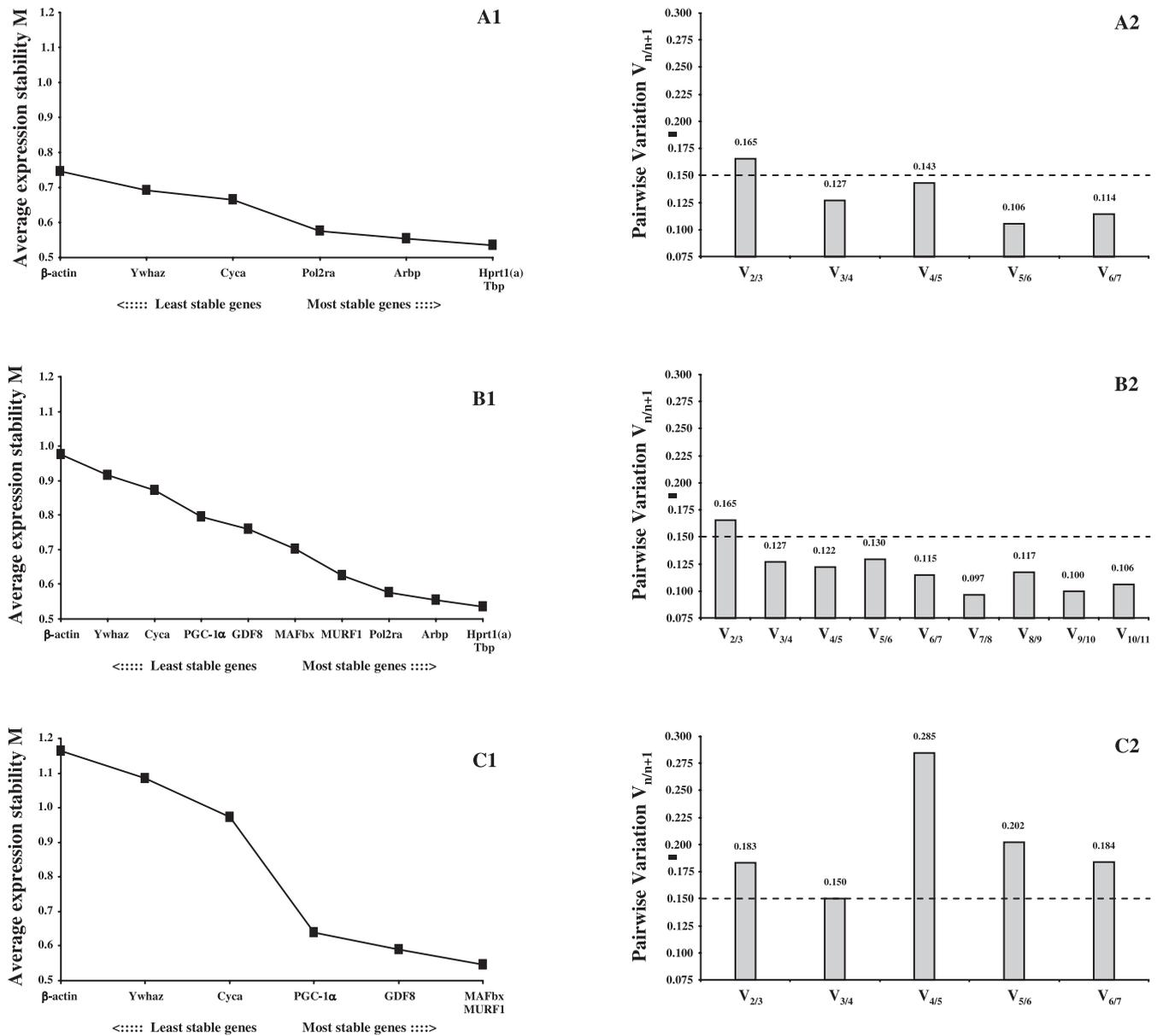


Fig. 2. A1, B1, C1: average expression stability values (M) of 3 combinations of potential reference genes. A2, B2, C2: pair-wise variation ( $V_{n/n+1}$ ) analysis between normalization factors  $NF_n$  and  $NF_{n+1}$ , based on 3 combinations of reference genes.  $V_{n/n+1}$  is assessed to determine the number of control genes required for an accurate normalization. Average stability M and pair-wise variation ( $V_{n/n+1}$ ) are determined using geNorm analysis software. The first geNorm analysis (A1, A2) assessed 7 potential reference genes [Hprt1(a), Tbp, Arbp, Pol2ra, Cyca, Ywhaz, and  $\beta$ -actin]. In the 2nd geNorm analysis (B1, B2), 4 other reference genes were added (MURF1, MAFbx, GDF8, and PGC-1 $\alpha$ ). The 3rd geNorm analysis (C1, C2) evaluated another combination of 7 reference genes (MURF1, MAFbx, GDF8 and PGC-1 $\alpha$ , Cyca, Ywhaz, and  $\beta$ -actin). The  $NF_{6A}$  and  $NF_{4A}$  values exposed in the text were based on the first geNorm analysis (A1, A2). The  $NF_{9B}$  and  $NF_{4C}$  values exposed in the text were respectively based on the 2nd (B1, B2) and 3rd (C1, C2) geNorm analysis.

threshold (Fig. 2A2). Thus, the four most stable genes [Tbp, Hprt1(a), Arbp, and Pol2ra] presented in Fig. 2A1 were at least required for normalization ( $V_{3/4} = 0.127$ ). The optimal geNorm-derived normalization, corresponding to the lowest  $V_{n/n+1}$  value, was obtained with the six most stable genes [Tbp, Hprt1(a), Arbp and Pol2ra, Cyca, Ywhaz] ( $V_{5/6} = 0.106$ ) (Fig. 2A2).

Accurate normalization requires the elimination of systematic variations in levels of reference genes across the investigated sample set (2, 28). Therefore, it is essential to verify the lack of changes in NFs between experimental groups. Due to technical considerations related to reverse transcription step

(4), geNorm-derived NFs are calculated for a fixed weight of total RNA. To be physiologically relevant, the NF values (i.e., reference mRNAs content) should be expressed for a fixed weight of tissue. Because the total RNA concentration does not generally vary under investigation, the NF values are classically considered for a fixed weight of total RNA. In this muscle hypertrophic model, we observed a marked increase in the total RNA concentration, especially at d5 (Fig. 1). Therefore, the NF values presented below were first determined for a fixed weight of total RNA ( $NF_{RNA}$ ) and then converted for a fixed weight of tissue ( $NF_{tissue}$ ), multiplying the NF values for a fixed weight of total RNA by the total RNA concentration.

NFs based on the six most stable genes of the first geNorm analysis (Fig. 2A1) corresponded to the optimal geNorm-derived normalization (Fig. 2A2). These NF values, determined for a fixed weight of tissue (NF6A<sub>tissue</sub>), were significantly higher in OV than in C groups (263%,  $P < 0.001$ , and 95%,  $P < 0.01$ , at d5 and d12, respectively) (Fig. 3A2). Moreover, the NF6A<sub>tissue</sub> values were affected by time in OV groups (overload  $\times$  time interaction,  $P < 0.01$ ) and expanded the between-groups variations observed when considered NF6A for a fixed weight of total RNA (NF6A<sub>RNA</sub>) (Fig. 3A1). The lowest variations in the NF values from the first geNorm analysis were obtained using the four most stable genes [Tbp, Hprt1(a), Arbp, and Pol2ra] (NF4A), whereas NF variations increased when additional genes were included (data not shown). These NF values, determined for a fixed weight of tissue (NF4A<sub>tissue</sub>), still remained higher in OV than in C rats (166%,  $P < 0.001$  and 54%,  $P < 0.05$ , at d5 and d12, respectively) and decreased with time in OV groups (overload  $\times$

time interaction,  $P < 0.01$ ) (Fig. 3B2). NF4A<sub>RNA</sub>, determined for a fixed weight of total RNA, were also affected by overload, especially at d5 (global effect of overload,  $P < 0.001$ ; Fig. 3B1).

*Determination of constant normalization factors in this model.* Each of the seven classical potential reference genes used above (first geNorm analysis, Fig. 2A) was overexpressed in OV groups. Therefore, it was impossible to obtain constant NFs from stable genes between experimental groups in this configuration. Then, we intentionally added four genes into the initial geNorm analysis (Fig. 2B). MURF1, MAFbx, and GDF8, which are involved in the muscle atrophic response, have been shown to be downregulated in functional overloaded muscles (6, 29). PGC-1 $\alpha$ , a gene associated with mitochondrial biogenesis, has already been observed as negatively affected by overload in our laboratory (personal data).

Despite the inclusion of genes presumably downregulated by overload into the initial geNorm analysis, all combinations of

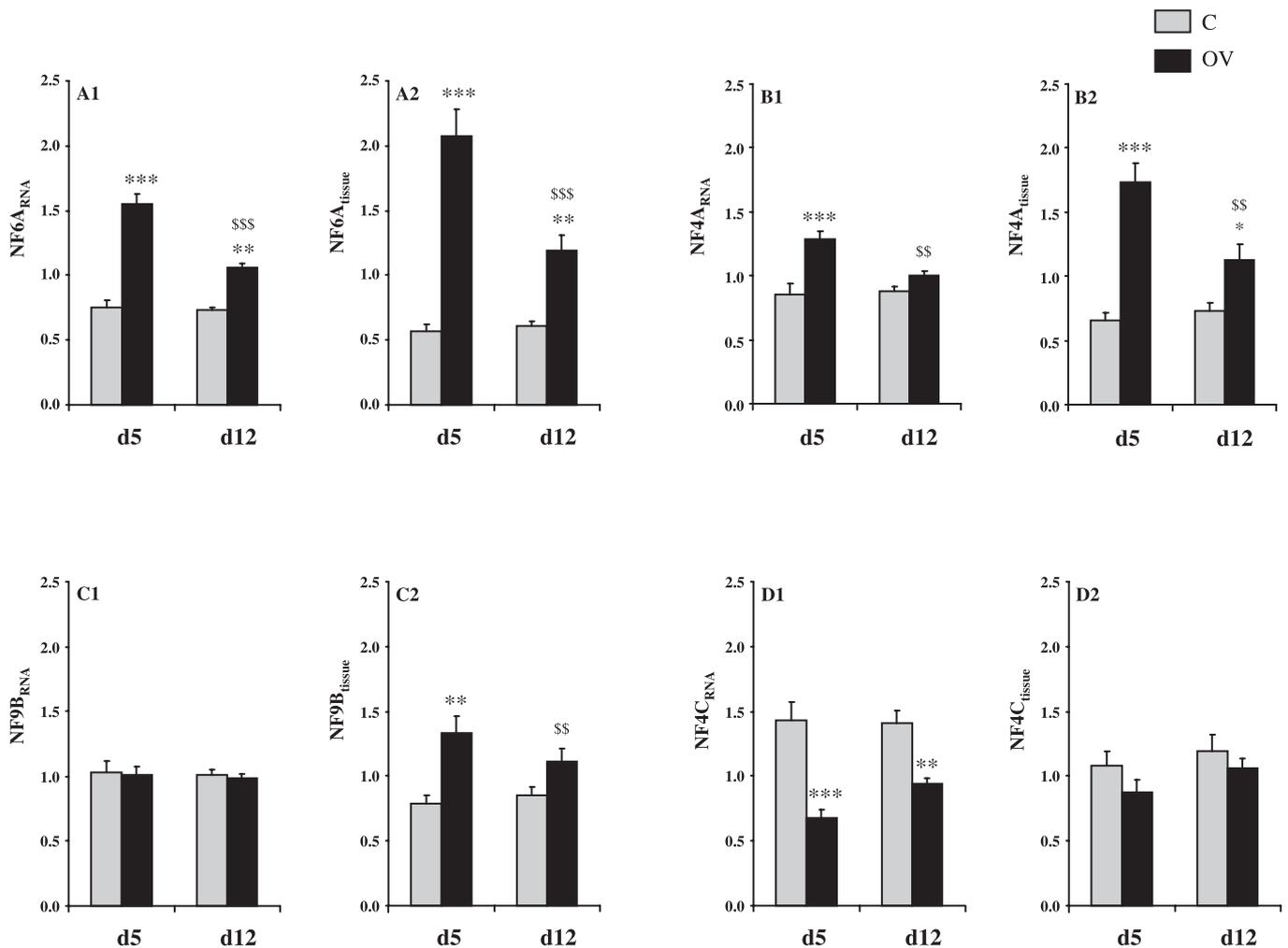


Fig. 3. A1, B1: normalization factors determined for a fixed weight of total RNA, based on the 1st geNorm analysis, obtained from the 6 (NF6A<sub>RNA</sub>) and the 4 (NF4A<sub>RNA</sub>) most stable reference genes, respectively. C1, D1: normalization factors determined for a fixed weight of total RNA, obtained from the nine (NF9B<sub>RNA</sub>, 2nd geNorm analysis) and the 4 (NF4C<sub>RNA</sub>, 3rd geNorm analysis) most stable reference genes, respectively. A2, B2: normalization factors determined for a fixed weight of tissue, based on the 1st geNorm analysis, obtained from the 6 (NF6A<sub>tissue</sub>) and the 4 (NF4A<sub>tissue</sub>) most stable reference genes, respectively. C2, D2: normalization factors determined for a fixed weight of tissue, obtained from the 9 (NF9B<sub>tissue</sub>, 2nd geNorm analysis) and the 4 (NF4C<sub>tissue</sub>, 3rd geNorm analysis) most stable reference genes, respectively. Normalization factors for a fixed weight of tissue were calculated by multiplying normalization factors determined for a fixed weight of total RNA by the total RNA concentration. Values are means  $\pm$  SE. Significantly different from the same time C group: \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . Significantly different from d5 for the same group: \$\$ $P < 0.01$ , \$\$\$ $P < 0.001$ .

stable genes led to higher  $NF_{tissue}$  values in OV groups, especially at d5 (data not shown). The NF values obtained from the nine most stable genes [Hprt1(a), Tbp, Arbp, Pol2ra, MURF1, MAFbx, GDF8, PGC-1 $\alpha$ , Cyca] ( $NF9B_{RNA}$ ) (obtained from the second geNorm analysis, Fig. 2B) were similar between experimental groups when determined for a fixed weight of total RNA (Fig. 3C1), but the physiologically relevant NF9B values ( $NF9B_{tissue}$ ) increased with overload (global effect,  $P < 0.001$ ) (Fig. 3C2).

Since the four genes included above highly diminished variations in the NF values between OV and C groups compared with other genes, we have subsequently checked if these genes alone could be stable enough to be used for the NF calculation. The elimination of the previous most stable genes [Tbp, Hprt1(a), Arbp, and Pol2ra] allowed to define these four genes as the most stable, compared with others (see in the third geNorm analysis, Fig. 2C1). This gene configuration markedly increased the  $V_{n/n+1}$  values (Fig. 2C2), but the combination of these four genes was acceptable ( $V_{3/4} = 0.15$ ). The NF values based on these four genes were lower in OV than in C groups, when determined for a fixed weight of total RNA ( $NF4C_{RNA}$ ) (53%,  $P < 0.001$  and 33%,  $P < 0.01$  at d5 and d12, respectively; Fig. 3D1). However, the  $NF4C_{tissue}$  values were not significantly affected by overload (global effect,  $P = 0.12$ ; Fig. 3D2).

## DISCUSSION

Internal normalization is a commonly used method to accurately quantify target mRNAs from RT-qPCR (4). In the overload model, normalization is generally performed using a single reference gene (6, 12, 15, 16, 25, 27, 29). It is now established that the internal normalization based on multiple selected reference genes is the best standard method to quantify target mRNAs (2, 22, 28). The present study was designed to propose an appropriate quantification method from RT-qPCR assay, usable under conditions of marked increases in muscle mass such as overload-induced hypertrophy.

In this study, we point out a limitation of the classical geNorm-derived normalization (28) in the target mRNA quantification. This software assesses variations in expression of one gene compared with the other ones, in all muscle samples. NFs calculated from the geometric mean of stable genes allow the investigator to correct for technical variations that should, on average, be similar for all samples, irrespective of the experimental group. However, significant intergroup variations in the NF values could not be explained by random intersamples technical variations. The limitation of geNorm is that it does not verify whether the NF values are similar between experimental groups. Here, we reported that reference genes were selected by geNorm software, although their NF values were significantly higher (NF6A, NF4A) or lower ( $NF4C_{RNA}$ ) in overloaded muscles than in control muscles. Consequently, the use of these selected NFs is not acceptable to normalize the target mRNAs levels. Indeed, it would either underestimate (with higher NF values) or overestimate (with lower NF values) the target mRNAs levels in some experimental groups, which would provide false information on the biological response.

It is required to select reference genes that do not vary in tissue across the investigated sample set (2, 28). This means

that the NF values based on the selected genes should be constant within groups when determined for a fixed weight of tissue. Here, we showed as others did (1, 14), that functional overload induced a marked increase in the total RNA concentration. To be physiologically relevant, we decided to convert the NF values classically determined for a fixed weight of total RNA, in NF values expressed for a fixed weight of tissue. This NFs conversion expanded variations observed between OV and C groups for NF6A and NF4A. Moreover, it reveals significant physiological variations in NF9B between these groups at d5. Such variations invalidate this set of nine reference genes, which could have been initially selected if the  $NF9B_{RNA}$  values were considered. These NF variations cannot be explained by intergroup differences in technical variations. Our results show that most of classical reference genes, likely positively affected by overload when expressed per unit weight of total RNA, are severely overexpressed per unit weight of tissue in this hypertrophic model. This is consistent with the fact that functional overload induced an early and drastic muscle hypertrophy. This muscle growth is associated with a cellular remodeling, such as the incorporation of satellite cells, fibroblasts, and immune cells (5). These cell types participate in modifying the hypertrophied muscle composition and certainly influence the expression of reference genes.

The question that arises is whether it is possible to normalize from multiple reference genes in overload. To address this issue, we intentionally added four genes into the initial geNorm analysis, recognized to be downregulated in overload. Only the NF values based on these four genes ( $NF4C_{tissue}$ ) were roughly similar between experimental groups. Nevertheless, the use of these constant  $NF4C_{tissue}$  values is highly questionable. Indeed, three out of four genes (MAFbx, MURF1, GDF8) are involved in the muscle atrophic response. However, reference genes must be selected within different functional classes to avoid that co-regulated genes are selected by geNorm because of their similar expression profile (28). The choice of these three genes is therefore not compatible with the fundamental principles of normalization by multiple reference genes. Moreover, it is physiological nonsense to select genes related to the atrophic response in a muscle hypertrophic situation since several studies showed that these genes were downregulated during muscle growth (23, 24). For these reasons, the use of  $NF4C_{tissue}$  is not acceptable in this model. Moreover, the selection of NFs is only relevant whether they are validated during all the experimental period. We logically consider that the expression of these three genes, which mainly contribute to the  $NF4C_{tissue}$  constancy, could vary extensively over time in this model, especially during the initial phase of muscle growth. Another solution to validate relevant NFs would be the assessment of other genes. According to our results, we support the idea that most of the housekeeping genes, which are involved in the maintenance of cellular homeostasis (cytoskeletal structure, metabolism regulation, gene expression), may be affected by functional overload. A solution would be the identification of some genes that are neither involved in these fundamental biological processes nor involved in the control of muscle mass. Such genes, probably expressed at low levels in this model, could represent only a limited fraction of the mRNAs.

Variations in the total RNA concentration have already been shown in situations of muscle mass modulation. Indeed, an increase was reported during strength training (9) or during

reloading and recovery from hindlimb unweighting (19). In contrast, a decrease was observed during hindlimb unweighting in rats (8). In these situations, the authors chose to express the mRNA levels per unit weight of tissue, multiplying the exponential conversion of the Cq values by the total RNA concentration. It is important to know whether this method is more relevant to compare target mRNA levels between OV and C animals than the classical normalization method by multiple reference genes. As explained above, the use of multiple classical genes would lead to underestimation of the target mRNA levels in OV groups and thus generate inconsistent results. However, this method enables to control for variations related to mRNA multistage processes. In contrast, expressing target mRNAs per unit weight of tissue, without any internal normalization, using the total RNA concentration as a correction factor is not optimal because it 1) is subject to the accuracy of the total RNA measurement and 2) does not control for technical errors. However, it is possible to provide a reasonable assessment of both mRNA integrity and reverse transcription yield by the use of a 3'/5' assay (18). It is noteworthy that such a method, less statistically powerful than any validated internal normalization, necessitates the increase of the sample set. Although this method only allows to assess marked intergroup variations in gene expression, we consider that it is more appropriate because it will not generate misleading results.

Through the overload model, we point out a limitation of the classical geNorm normalization and propose to consider NF as a dimensionless value for a fixed weight of tissue to be physiologically relevant. Thus, the quantification of target mRNAs requires specific attention to avoid generating misleading results. We propose guidelines for optimizing target mRNA normalization and quantification, useful in models of muscle mass modulation.

1) The total RNA concentration should be reliably measured and reverse transcription should be performed using a constant amount of total RNA (4).

2) geNorm software should be used to select stable genes within different functional classes, to determine the number of required genes for normalization, and to calculate the NF values.

3) The NF values should be systematically calculated. If the total RNA concentration does not vary under investigation (*case 1*), as observed in most of experiments, the NF values are determined for a fixed weight of total RNA. If variations in the total RNA concentration are found (*case 2*), the NF values should be determined for a fixed weight of tissue to be physiologically relevant. Then, the constancy of the NF values within experimental groups should be specifically verified in *cases 1* and *2*.

4) If constant NF values are obtained between experimental groups, target mRNAs can be accurately quantified as following:

*Case 1:* Target mRNA level = exponential conversion of the Cq values of target mRNA (for a fixed weight of total RNA)/NF<sub>RNA</sub> (Eq. 1).

*Case 2:* Target mRNA level = exponential conversion of the Cq values of target mRNA (for a fixed weight of tissue)/NF<sub>tissue</sub> (Eq. 2). As the total RNA concentration is a common factor of the numerator and the denominator in Eq. 2, Eq. 2 is equivalent to Eq. 1 and can be alternatively used. It is noteworthy that in this case, the NF values, determined for a fixed

weight of total RNA, are not constant within experimental groups.

5) If the NF values are not constant, it is necessary to test other potential reference genes to find constant and stable reference genes. If this fails, as it did in this study, internal normalization should be invalidated and target mRNAs should be expressed as absolute values per unit weight of tissue. The technical variations will consequently remain present as a part of the intergroup variations, leading to less statistical power.

6) The lack of constant NFs between experimental groups does not mean that the use of normalization by multiple reference genes must be fully excluded. It is conceivable to first assess the effects of muscle mass modulation by an absolute quantification of target mRNAs per unit weight of tissue. However, a classical normalization by multiple reference genes is most probably a relevant method to determine the specific effects of an additional stress on skeletal muscles submitted to variations in muscle mass. Indeed, we observed constant geNorm-derived NF values (NF4A) when OV rats were exposed to hypoxia compared with normoxic conditions (personal data).

In conclusion, we demonstrated that the commonly used geNorm-derived normalization method is questionable for accurately comparing target mRNAs between control and hypertrophied muscles in response to overload, a situation of drastic cellular remodeling. Indeed, we failed to identify stable and constant reference genes that belong to different functional gene classes, a fundamental principle of normalization by multiple genes. Therefore, we proposed to quantify target mRNAs per unit weight of tissue, without any internal normalization, to assess intergroup variations in this model, an in extension in situations where constant NFs were not found.

#### ACKNOWLEDGMENTS

The authors thank Gabrielle Cormier and Laura Murphy for careful reading of the manuscript.

#### DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

#### REFERENCES

- Adams GR, Haddad F, Baldwin KM. Time course of changes in markers of myogenesis in overloaded rat skeletal muscles. *J Appl Physiol* 87: 1705–1712, 1999.
- Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* 64: 5245–5250, 2004.
- Bigard AX, Zoll J, Ribera F, Mateo P, Sanchez H, Serrurier B, Ventura-Clapier R. Influence of overload on phenotypic remodeling in regenerated skeletal muscle. *Am J Physiol Cell Physiol* 281: C1686–C1694, 2001.
- Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 55: 611–622, 2009.
- Carson JA, Nettleton D, Reecy JM. Differential gene expression in the rat soleus muscle during early work overload-induced hypertrophy. *FASEB J* 16: 207–209, 2002.
- Choi H, Selpides PJ, Nowell MM, Rourke BC. Functional overload in ground squirrel plantaris muscle fails to induce myosin isoform shifts. *Am J Physiol Regul Integr Comp Physiol* 297: R578–R586, 2009.
- Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, Zumla A. Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques* 37: 112–114, 116, 118–119, 2004.

8. **Giger JM, Bodell PW, Zeng M, Baldwin KM, Haddad F.** Rapid muscle atrophy response to unloading: pretranslational processes involving MHC and actin. *J Appl Physiol* 107: 1204–1212, 2009.
9. **Heinemeier KM, Olesen JL, Schjerling P, Haddad F, Langberg H, Baldwin KM, Kjaer M.** Short-term strength training and the expression of myostatin and IGF-I isoforms in rat muscle and tendon: differential effects of specific contraction types. *J Appl Physiol* 102: 573–581, 2007.
10. **Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J.** qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8: R19, 2007.
11. **Huggett J, Dheda K, Bustin S, Zumla A.** Real-time RT-PCR normalization: strategies and considerations. *Genes Immun* 6: 279–284, 2005.
12. **Hyatt JP, McCall GE, Kander EM, Zhong H, Roy RR, Huey KA.** PAX3/7 expression coincides with MyoD during chronic skeletal muscle overload. *Muscle Nerve* 38: 861–866, 2008.
13. **Koch I, Weil R, Wolbold R, Brockmoller J, Hustert E, Burk O, Nuessler A, Neuhaus P, Eichelbaum M, Zanger U, Wojnowski L.** Interindividual variability and tissue-specificity in the expression of cytochrome P450 3A mRNA. *Drug Metab Dispos* 30: 1108–1114, 2002.
14. **Lee WJ, McClung J, Hand GA, Carson JA.** Overload-induced androgen receptor expression in the aged rat hindlimb receiving nandrolone decanoate. *J Appl Physiol* 94: 1153–1161, 2003.
15. **McCarthy JJ, Esser KA.** MicroRNA-1 and microRNA-133a expression are decreased during skeletal muscle hypertrophy. *J Appl Physiol* 102: 306–313, 2007.
16. **McClung JM, Mehl KA, Thompson RW, Lowe LL, Carson JA.** Nandrolone decanoate modulates cell cycle regulation in functionally overloaded rat soleus muscle. *Am J Physiol Regul Integr Comp Physiol* 288: R1543–R1552, 2005.
17. **Mounier R, Lantier L, Leclerc J, Sotiropoulos A, Pende M, Daegelen D, Sakamoto K, Foretz M, Viollet B.** Important role for AMPK $\alpha$ 1 in limiting skeletal muscle cell hypertrophy. *FASEB J* 23: 2264–2273, 2009.
18. **Nolan T, Hands RE, Bustin SA.** Quantification of mRNA using real-time RT-PCR. *Nat Protoc* 1: 1559–1582, 2006.
19. **Pandorf CE, Jiang WH, Qin AX, Bodell PW, Baldwin KM, Haddad F.** Calcineurin plays a modulatory role in loading-induced regulation of type I myosin heavy chain gene expression in slow skeletal muscle. *Am J Physiol Regul Integr Comp Physiol* 297: R1037–R1048, 2009.
20. **Peinnequin A, Mouret C, Birot O, Alonso A, Mathieu J, Clarencon D, Agay D, Chancerelle Y, Multon E.** Rat pro-inflammatory cytokine and cytokine related mRNA quantification by real-time polymerase chain reaction using SYBR green. *BMC Immunol* 5: 3, 2004.
21. **Pfaffl MW.** A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29: e45, 2001.
22. **Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP.** Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnol Lett* 26: 509–515, 2004.
23. **Richard-Bulteau H, Serrurier B, Crassous B, Banzet S, Peinnequin A, Bigard X, Koulmann N.** Recovery of skeletal muscle mass after extensive injury: positive effects of increased contractile activity. *Am J Physiol Cell Physiol* 294: C467–C476, 2008.
24. **Roth SM, Martel GF, Ferrell RE, Metter EJ, Hurley BF, Rogers MA.** Myostatin gene expression is reduced in humans with heavy-resistance strength training: a brief communication. *Exp Biol Med (Maywood)* 228: 706–709, 2003.
25. **Sakuma K, Akiho M, Nakashima H, Nakao R, Hirata M, Inashima S, Yamaguchi A, Yasuhara M.** Cyclosporin A modulates cellular localization of MEF2C protein and blocks fiber hypertrophy in the overloaded soleus muscle of mice. *Acta Neuropathol (Berl)* 115: 663–674, 2008.
26. **Tanaka Y, Yamaguchi A, Fujikawa T, Sakuma K, Morita I, Ishii K.** Expression of mRNA for specific fibroblast growth factors associates with that of the myogenic markers MyoD and proliferating cell nuclear antigen in regenerating and overloaded rat plantaris muscle. *Acta Physiol (Oxf)* 194: 149–159, 2008.
27. **Thompson RW, McClung JM, Baltgalvis KA, Davis JM, Carson JA.** Modulation of overload-induced inflammation by aging and anabolic steroid administration. *Exp Gerontol* 41: 1136–1148, 2006.
28. **Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F.** Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3: RESEARCH0034, 2002.
29. **Yamaguchi A, Fujikawa T, Shimada S, Kanbayashi I, Tateoka M, Soya H, Takeda H, Morita I, Matsubara K, Hirai T.** Muscle IGF-I Ea, MGF, and myostatin mRNA expressions after compensatory overload in hypophysectomized rats. *Pflügers Arch* 453: 203–210, 2006.



# Références Bibliographiques

---

1. (2011). Real-time PCR: Understanding Ct. Applied Biosystems Application Note, Applied Biosystems.
2. Akane, A., K. Matsubara, et al. (1994). "Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification." J Forensic Sci **39**(2): 362-372.
3. Al-Soud, W. A., L. J. Jonsson, et al. (2000). "Identification and characterization of immunoglobulin G in blood as a major inhibitor of diagnostic PCR." J Clin Microbiol **38**(1): 345-350.
4. Al-Soud, W. A. and P. Radstrom (2001). "Purification and characterization of PCR-inhibitory components in blood cells." J Clin Microbiol **39**(2): 485-493.
5. Allawi, H. T. and J. SantaLucia, Jr. (1997). "Thermodynamics and NMR of internal G.T mismatches in DNA." Biochemistry **36**(34): 10581-10594.
6. Altman, D. G. (2002). "Commentary on quantifying agreement between two methods of measurement." Clinical chemistry **48**: 801-802.
7. Altman, D. G. and J. M. Bland (1983). "Measurement in Medicine: the Analysis of Method Comparison Studies." The Statistician **32**: 307-317.
8. Alwine, J. C., D. J. Kemp, et al. (1977). "Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes." Proc Natl Acad Sci U S A **74**(12): 5350-5354.
9. Ando, T. (1966). "A nuclease specific for heat-denatured DNA is isolated from a product of *Aspergillus oryzae*." Biochim Biophys Acta **114**(1): 158-168.
10. Arora, A., H. Kaur, et al. (2008). "Effect of locked nucleic acid (LNA) modification on hybridization kinetics of DNA duplex." Nucleic Acids Symp Ser (Oxf)(52): 417-418.
11. Avery, O. T., C. M. Macleod, et al. (1943). "Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. Oswald Theodore Avery (1877-1955)." Clin Orthop Relat Res(379 Suppl): S3-8.
12. Babu, B. R. and J. Wengel (2001). "Universal hybridization using LNA (locked nucleic acid) containing a novel pyrene LNA nucleotide monomer." Chem Commun (Camb)(20): 2114-2115.
13. Baldino, F., Jr., M. F. Chesselet, et al. (1989). "High-resolution in situ hybridization histochemistry." Methods Enzymol **168**: 761-777.
14. Baumforth, K. R., P. N. Nelson, et al. (1999). "Demystified ... the polymerase chain reaction." Mol Pathol **52**(1): 1-10.
15. Becker, A., A. Reith, et al. (1996). "A quantitative method of determining initial amounts of DNA by polymerase chain reaction cycle titration using digital imaging and a novel DNA stain." Anal Biochem **237**(2): 204-207.

16. Becker-Andre, M. and K. Hahlbrock (1989). "Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY)." Nucleic Acids Res **17**(22): 9437-9446.
17. Belec, L., J. Authier, et al. (1998). "Myoglobin as a polymerase chain reaction (PCR) inhibitor: a limitation for PCR from skeletal muscle tissue avoided by the use of *Thermus thermophilus* polymerase." Muscle Nerve **21** (8): 1064-1067.
18. Berk, A. J. and P. A. Sharp (1977). "Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids." Cell **12**(3): 721-732.
19. Bickley, J., J. K. Short, et al. (1996). "Polymerase chain reaction (PCR) detection of *Listeria monocytogenes* in diluted milk and reversal of PCR inhibition caused by calcium ions." Lett Appl Microbiol **22**(2): 153-158.
20. Birch, D. E. (1996). "Simplified hot start PCR." Nature **381** (6581): 445-446.
21. Bland, J. M. and D. G. Altman (1986). "Statistical methods for assessing agreement between two methods of clinical measurement." Lancet **1** (8476): 307-310.
22. Bohlenius, H., S. Eriksson, et al. (2007). "Retraction." Science **316**(5823): 367.
23. Bolton, E. T. and C. B. Mc (1962). "A general method for the isolation of RNA complementary to DNA." Proc Natl Acad Sci U S A **48**: 1390-1397.
24. Bommarito, S., N. Peyret, et al. (2000). "Thermodynamic parameters for DNA sequences with dangling ends." Nucleic Acids Res **28**(9): 1929-1934.
25. Borer, P. N., B. Dengler, et al. (1974). "Stability of ribonucleic acid double-stranded helices." J Mol Biol **86**(4): 843-853.
26. Breslauer, K. J., R. Frank, et al. (1986). "Predicting DNA duplex stability from the base sequence." Proc Natl Acad Sci U S A **83**(11): 3746-3750.
27. Britten, R. J. and E. H. Davidson (1976). "Studies on nucleic acid reassociation kinetics: empirical equations describing DNA reassociation." Proc Natl Acad Sci U S A **73**(2): 415-419.
28. Britten, R. J., D. E. Graham, et al. (1974). "Analysis of repeating DNA sequences by reassociation." Methods Enzymol **29**(0): 363-418.
29. Brock, T. D. and H. Freeze (1969). "*Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile." J Bacteriol **98**(1): 289-297.
30. Brooks, E. M., L. G. Sheflin, et al. (1995). "Secondary structure in the 3' UTR of EGF and the choice of reverse transcriptases affect the detection of message diversity by RT-PCR." Biotechniques **19**(5): 806-812, 814-805.
31. Burns, M. J. and H. Valdivia (2008). "Modelling the limit of detection in real-time quantitative PCR." Eur Food Res Technol (226): 1513-1524.

32. Bustin, S. A. (2000). "Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays." J Mol Endocrinol **25**(2): 169-193.
33. Bustin, S. A. (2002). "Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems." J Mol Endocrinol **29**(1): 23-39.
34. Bustin, S. A. (2004). A-Z of Quantitative PCR Portland, OR, USA, International University Line
35. Bustin, S. A. (2005). "Real-time, fluorescence-based quantitative PCR: a snapshot of current procedures and preferences." Expert Rev Mol Diagn **5**(4): 493-498.
36. Bustin, S. A. (2008). "RT-qPCR and molecular diagnostics: no evidence for measles virus in the GI tract of autistic children " Eur Pharm Rev Dig: 11-16.
37. Bustin, S. A. (2010). The PCR Revolution: Basic Technologies and Applications, Cambridge University Press.
38. Bustin, S. A. (2010). "Why the need for qPCR publication guidelines?--The case for MIQE." Methods **50**(4): 217-226.
39. Bustin, S. A., J. F. Beaulieu, et al. (2010). "MIQE precis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments." BMC Mol Biol **11**: 74.
40. Bustin, S. A., V. Benes, et al. (2009). "The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments." Clin Chem **55**(4): 611-622.
41. Bustin, S. A., V. Benes, et al. (2011). "Primer sequence disclosure: a clarification of the MIQE guidelines." Clin Chem **57**(6): 919-921.
42. Bustin, S. A., V. Benes, et al. (2005). "Quantitative real-time RT-PCR--a perspective." J Mol Endocrinol **34**(3): 597-601.
43. Bustin, S. A., V. G. Gyselman, et al. (1999). "Detection of cytokeratins 19/20 and guanylyl cyclase C in peripheral blood of colorectal cancer patients." Br J Cancer **79**(11-12): 1813-1820.
44. Bustin, S. A. and T. Nolan (2004). "Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction." J Biomol Tech **15**(3): 155-166.
45. Carrillo-Nava, E., Y. Mejia-Radillo, et al. (2008). "Dodecamer DNA duplex formation is characterized by second-order kinetics, positive activation energies, and a dependence on sequence and Mg<sup>2+</sup> ion concentration." Biochemistry **47**(50): 13153-13157.
46. Casey, J. and N. Davidson (1977). "Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide." Nucleic Acids Res **4**(5): 1539-1552.
47. Cha, R. S. and W. G. Thilly (1993). "Specificity, efficiency, and fidelity of PCR." PCR Methods Appl **3**(3): S18-29.

48. Chalikian, T. V., J. Volker, et al. (1999). "A more unified picture for the thermodynamics of nucleic acid duplex melting: a characterization by calorimetric and volumetric techniques." *Proc Natl Acad Sci U S A* **96**(14): 7853-7858.
49. Chelly, J., J. C. Kaplan, et al. (1988). "Transcription of the dystrophin gene in human muscle and non-muscle tissue." *Nature* **333**(6176): 858-860.
50. Cheng, S., Y. Chen, et al. (1995). "Template integrity is essential for PCR amplification of 20- to 30-kb sequences from genomic DNA." *PCR Methods Appl* **4**(5): 294-298.
51. Chomczynski, P. and N. Sacchi (1987). "Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction." *Anal Biochem* **162**(1): 156-159.
52. Chou, L. S., C. Meadows, et al. (2005). "Unlabeled oligonucleotide probes modified with locked nucleic acids for improved mismatch discrimination in genotyping by melting analysis." *Biotechniques* **39**(5): 644, 646, 648 passim.
53. Christensen, U. (2007). "Thermodynamic and kinetic characterization of duplex formation between 2'-O, 4'-C-methylene-modified oligoribonucleotides, DNA and RNA." *Biosci Rep* **27**(6): 327-333.
54. Christensen, U., N. Jacobsen, et al. (2001). "Stopped-flow kinetics of locked nucleic acid (LNA)-oligonucleotide duplex formation: studies of LNA-DNA and DNA-DNA interactions." *Biochem J* **354**(Pt 3): 481-484.
55. Cikos, S., A. Bukovska, et al. (2007). "Relative quantification of mRNA: comparison of methods currently used for real-time PCR data analysis." *BMC Mol Biol* **8**: 113.
56. Cocco, S. and R. Monasson (1999). "Statistical Mechanics of Torque Induced Denaturation of DNA." *Physical Review Letters* **83**(24): 5178-5181.
57. Costa, J. M., P. Emault, et al. (2004). "Chimeric LNA/DNA probes as a detection system for real-time PCR." *Clin Biochem* **37**(10): 930-932.
58. Curry, J., C. McHale, et al. (2002). "Low efficiency of the Moloney murine leukemia virus reverse transcriptase during reverse transcription of rare t(8;21) fusion gene transcripts." *Biotechniques* **32**(4): 768, 770, 772, 754-765.
59. Dahm, R. (2008). "Discovering DNA: Friedrich Miescher and the early years of nucleic acid research." *Hum Genet* **122**(6): 565-581.
60. Datta, S. P., G. H. Smith, et al. (2000). Oxford Dictionary of Biochemistry and Molecular Biology. *Oxford Dictionary of Biochemistry and Molecular Biology revised Edition*. O. U. P. Inc. NewYork, Oxford University Press: 740.
61. Delcourt, S. G. and R. D. Blake (1991). "Stacking energies in DNA." *J Biol Chem* **266**(23): 15160-15169.
62. Doktycz, M. J., R. F. Goldstein, et al. (1992). "Studies of DNA dumbbells. I. Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 endloops: evaluation of the nearest-neighbor stacking interactions in DNA." *Biopolymers* **32**(7): 849-864.

63. Eckert, K. A. and T. A. Kunkel (1990). "High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase." Nucleic Acids Res **18**(13): 3739-3744.
64. Eckert, K. A. and T. A. Kunkel (1991). "DNA polymerase fidelity and the polymerase chain reaction." PCR Methods Appl **1**(1): 17-24.
65. Eichhorn, G. L. and Y. A. Shin (1968). "Interaction of metal ions with polynucleotides and related compounds. XII. The relative effect of various metal ions on DNA helicity." J Am Chem Soc **90**(26): 7323-7328.
66. Eilert, K. D. and D. R. Foran (2009). "Polymerase resistance to polymerase chain reaction inhibitors in bone\*." J Forensic Sci **54**(5): 1001-1007.
67. Eischeid, A. C. (2011). "SYTO dyes and EvaGreen outperform SYBR Green in real-time PCR." BMC Res Notes **4**: 263.
68. Erlich, H. A., D. Gelfand, et al. (1991). "Recent advances in the polymerase chain reaction." Science **252**(5013): 1643-1651.
69. Fleige, S. and M. W. Pfaffl (2006). "RNA integrity and the effect on the real-time qRT-PCR performance." Mol Aspects Med **27**(2-3): 126-139.
70. Freier, S. M., R. Kierzek, et al. (1986). "Improved free-energy parameters for predictions of RNA duplex stability." Proc Natl Acad Sci U S A **83**(24): 9373-9377.
71. Fye, R. M. and C. J. Benham (1999). "Exact method for numerically analyzing a model of local denaturation in superhelically stressed DNA." The American Physical Society **59**(3): 3408-3426.
72. Gentle, A., F. Anastasopoulos, et al. (2001). "High-resolution semi-quantitative real-time PCR without the use of a standard curve." Biotechniques **31**(3): 502, 504-506, 508.
73. Gillespie, D. and S. Spiegelman (1965). "A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane." J Mol Biol **12**(3): 829-842.
74. Gilliland, G., S. Perrin, et al. (1990). "Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction." Proc Natl Acad Sci U S A **87**(7): 2725-2729.
75. Golenberg, E. M., A. Bickel, et al. (1996). "Effect of highly fragmented DNA on PCR." Nucleic Acids Res **24**(24): 5026-5033.
76. Gotoh, O. and Y. Tagashira (1981). "Locations of frequently opening regions on natural DNAs and their relation to functional loci." Biopolymers **20**(5): 1043-1058.
77. Grenier, B., M. Dubreuil, et al. (2000). "Comparaison de deux méthodes de mesure d'une même grandeur : méthode de Bland et Altman." Ann Fr Anesth Réanim **19**: 128-135.
78. Grunweller, A. and R. K. Hartmann (2007). "Locked nucleic acid oligonucleotides: the next generation of antisense agents?" BioDrugs **21**(4): 235-243.

79. Gudnason, H., M. Dufva, et al. (2007). "Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature." Nucleic Acids Res **35**(19): e127.
80. Haimov-Kochman, R., S. J. Fisher, et al. (2006). "Modification of the standard Trizol-based technique improves the integrity of RNA isolated from RNase-rich placental tissue." Clin Chem **52**(1): 159-160.
81. Hall, T. J., J. W. Grula, et al. (1980). "Evolution of sea urchin non-repetitive DNA." J Mol Evol **16**(2): 95-110.
82. Haynie, D. T. (2008). Biologicals Thermodynamics, Second Edition. Cambridge CB2 8RU, UK, Cambridge University Press.
83. Hellemans, J., G. Mortier, et al. (2007). "qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data." Genome Biol **8**(2): R19.
84. Higuchi, R., G. Dollinger, et al. (1992). "Simultaneous amplification and detection of specific DNA sequences." Biotechnology (N Y) **10**(4): 413-417.
85. Higuchi, R., C. Fockler, et al. (1993). "Kinetic PCR analysis: real-time monitoring of DNA amplification reactions." Biotechnology (N Y) **11**(9): 1026-1030.
86. Hill, A. V. (1910). "The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves." J. Physiol. **40**: 4-7.
87. Holland, P. M., R. D. Abramson, et al. (1991). "Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of *Thermus aquaticus* DNA polymerase." Proc Natl Acad Sci U S A **88**(16): 7276-7280.
88. Huang, T., H. Bohlenius, et al. (2005). "The mRNA of the Arabidopsis gene FT moves from leaf to shoot apex and induces flowering." Science **309**(5741): 1694-1696.
89. Huggett, J., K. Dheda, et al. (2005). "Real-time RT-PCR normalisation; strategies and considerations." Genes Immun **6**(4): 279-284.
90. Huggett, J. F., T. Novak, et al. (2008). "Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon." BMC Res Notes **1**: 70.
91. Hughesman, C. B., R. F. Turner, et al. (2008). "Stability and mismatch discrimination of DNA duplexes containing 2,6-diaminopurine and 2-thiothymidine locked nucleic acid bases." Nucleic Acids Symp Ser (Oxf)(52): 245-246.
92. Hwa, T., E. Marinari, et al. (2003). "Localization of denaturation bubbles in random DNA sequences." Proc Natl Acad Sci U S A **100**(8): 4411-4416.
93. Ivanova, A. and N. Rosch (2007). "The structure of LNA:DNA hybrids from molecular dynamics simulations: the effect of locked nucleotides." J Phys Chem A **111**(38): 9307-9319.

94. Jensen, G. A., S. K. Singh, et al. (2001). "A comparison of the solution structures of an LNA:DNA duplex and the unmodified DNA:DNA duplex." J. Chem. Soc., Perkin Trans. 2: 1224-1232.
95. Jepsen, J. S., M. D. Sorensen, et al. (2004). "Locked nucleic acid: a potent nucleic acid analog in therapeutics and biotechnology." Oligonucleotides **14**(2): 130-146.
96. Johnson, M. P., L. M. Haupt, et al. (2004). "Locked nucleic acid (LNA) single nucleotide polymorphism (SNP) genotype analysis and validation using real-time PCR." Nucleic Acids Res **32**(6): e55.
97. Kainz, P. (2000). "The PCR plateau phase - towards an understanding of its limitations." Biochim Biophys Acta **1494**(1-2): 23-27.
98. Karrer, E. E., J. E. Lincoln, et al. (1995). "In situ isolation of mRNA from individual plant cells: creation of cell-specific cDNA libraries." Proc Natl Acad Sci U S A **92**(9): 3814-3818.
99. Kaur, H., A. Arora, et al. (2008). "Thermodynamic, counterion and hydration effects for the incorporation of locked nucleic acid (LNA) nucleotides in duplex." Nucleic Acids Symp Ser (Oxf)(52): 425-426.
100. Kaur, H., J. Wengel, et al. (2008). "Thermodynamics of DNA-RNA heteroduplex formation: effects of locked nucleic acid nucleotides incorporated into the DNA strand." Biochemistry **47**(4): 1218-1227.
101. Kemp, B. M., C. Monroe, et al. (2006). "Repeat silica extraction : a simple technique for the removal of PCR inhibitors from DNA extracts." Journal of archaeological science **33**: 1680-1689.
102. Keohavong, P. and W. G. Thilly (1989). "Fidelity of DNA polymerases in DNA amplification." Proc Natl Acad Sci U S A **86**(23): 9253-9257.
103. Kim, S., R. G. Labbe, et al. (2000). "Inhibitory effects of collagen on the PCR for detection of *Clostridium perfringens*." Appl Environ Microbiol **66**(3): 1213-1215.
104. Kleppe, K., E. Ohtsuka, et al. (1971). "Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases." J Mol Biol **56**(2): 341-361.
105. Kontanis, E. J. and F. A. Reed (2006). "Evaluation of real-time PCR amplification efficiencies to detect PCR inhibitors." J Forensic Sci **51** (4): 795-804.
106. Koshkin, A. A., P. Nielsen, et al. (1998). "LNA (Locked Nucleic Acid): An RNA Mimic Forming Exceedingly Stable LNA:LNA Duplexes." Journal of the American Chemical Society **120**(50): 13252-13253.
107. Koshkin, A. A., S. K. Singh, et al. (1998). "LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition." Tetrahedron **54**(14): 3607-3630.

108. Kramer, M. F. and D. M. Coen (2001). "Enzymatic amplification of DNA by PCR: standard procedures and optimization." Curr Protoc Immunol **Chapter 10**: Unit 10 20.
109. Kubista, M., J. M. Andrade, et al. (2006). "The real-time polymerase chain reaction." Mol Aspects Med **27**(2-3): 95-125.
110. Kumar, R., S. K. Singh, et al. (1998). "The first analogues of LNA (locked nucleic acids): phosphorothioate-LNA and 2'-thio-LNA." Bioorg Med Chem Lett **8**(16): 2219-2222.
111. Kuo, K. W., M. F. Leung, et al. (1997). "Intrinsic secondary structure of human TNFR-I mRNA influences the determination of gene expression by RT-PCR." Mol Cell Biochem **177**(1-2): 1-6.
112. Kvaerno, L. and J. Wengel (1999). "Investigation of restricted backbone conformations as an explanation for the exceptional thermal stabilities of duplexes involving LNA (Locked Nucleic Acids) : Synthesis and evaluation of abasic LNA." Chemical communications (n°7): 657-658.
113. Latorra, D., K. Arar, et al. (2003). "Design considerations and effects of LNA in PCR primers." Mol Cell Probes **17**(5): 253-259.
114. Latorra, D., K. Campbell, et al. (2003). "Enhanced allele-specific PCR discrimination in SNP genotyping using 3' locked nucleic acid (LNA) primers." Hum Mutat **22**(1): 79-85.
115. Latorra, D., D. Hopkins, et al. (2003). "Multiplex allele-specific PCR with optimized locked nucleic acid primers." Biotechniques **34**(6): 1150-1152, 1154, 1158.
116. Lawyer, F. C., S. Stoffel, et al. (1993). "High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity." PCR Methods Appl **2**(4): 275-287.
117. Le Novere, N. (2001). "MELTING, computing the melting temperature of nucleic acid duplex." Bioinformatics **17**(12): 1226-1227.
118. Lefever, S., J. Hellemans, et al. (2009). "RDML: structured language and reporting guidelines for real-time quantitative PCR data." Nucleic Acids Res **37**(7): 2065-2069.
119. Lefever, S., J. Vandesompele, et al. (2009). "RTPrimerDB: the portal for real-time PCR primers and probes." Nucleic Acids Res **37**(Database issue): D942-945.
120. Levesque-Sergerie, J. P., M. Duquette, et al. (2007). "Detection limits of several commercial reverse transcriptase enzymes: impact on the low- and high-abundance transcript levels assessed by quantitative RT-PCR." BMC Mol Biol **8**: 93.
121. Levin, J. D., D. Fiala, et al. (2006). "Position-dependent effects of locked nucleic acid (LNA) on DNA sequencing and PCR primers." Nucleic Acids Res **34**(20): e142.
122. Liu, W. and D. A. Saint (2002). "Validation of a quantitative method for real time PCR kinetics." Biochem Biophys Res Commun **294**(2): 347-353.
123. Livak, K. J. (1997 & 2001). Relative quantification of gene expression. ABI Prism 7700 Sequence detection System User Bulletin #2, Applied Biosystems: 36.

124. Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  Method." Methods **25**(4): 402-408.
125. Malgoyre, A., S. Banzet, et al. (2007). "Quantification of low-expressed mRNA using 5' LNA-containing real-time PCR primers." Biochem Biophys Res Commun **354**(1): 246-252.
126. Maniatis, T., F. E.F., et al. (1989). Molecular Cloning A Laboratory Manual. New York, Cold Spring Harbor Laboratory.
127. Manning, G. S. (1976). "On the application of polyelectrolyte limiting laws to the helix-coil transition of DNA. V. Ionic effects on renaturation kinetics." Biopolymers **15**(7): 1333-1343.
128. Mao, F. and W. Y. Leung (2010). Methods of using with dyes in association with nucleic acid staining or detection and associated technology. United States, Biotium, Inc. . **7,803,943 B2**: 63.
129. Mao, F., W. Y. Leung, et al. (2007). "Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications." BMC Biotechnol **7**: 76.
130. Mao, F., W. Y. Leung, et al. (2010). Dimeric and trimeric nucleic acid dyes, and associated systems and methods. United States, Biotium, Inc. **7,776,567 B2**: 75.
131. Marino, J. H., P. Cook, et al. (2003). "Accurate and statistically verified quantification of relative mRNA abundances using SYBR Green I and real-time RT-PCR." J Immunol Methods **283**(1-2): 291-306.
132. Marko, J. F. and E. D. Siggia (1995). "Statistical mechanics of supercoiled DNA." The American Physical Society **52**(3): 2912-2938.
133. Marky, L. A. and K. J. Breslauer (1982). "Calorimetric determination of base-stacking enthalpies in double-helical DNA molecules." Biopolymers **21**(11): 2185-2194.
134. Marmur, J. and P. Doty (1961). "Thermal renaturation of deoxyribonucleic acids." J Mol Biol **3**: 585-594.
135. Marmur, J. and P. Doty (1962). "Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature." J Mol Biol **5**: 109-118.
136. McGookin, R. (1984). RNA Extraction by the Guanidine Thiocyanate Procedure. Nucleic Acids. J. M. Walker. Clifton, Humana Press. **volume 2**: 113-116.
137. McTigue, P. M., R. J. Peterson, et al. (2004). "Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)-DNA duplex formation." Biochemistry **43**(18): 5388-5405.
138. Mehra, S. and W. S. Hu (2005). "A kinetic model of quantitative real-time polymerase chain reaction." Biotechnol Bioeng **91**(7): 848-860.

139. Melton, D. A., P. A. Krieg, et al. (1984). "Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter." Nucleic Acids Res **12**(18): 7035-7056.
140. Mercier, B., C. Gaucher, et al. (1990). "Direct PCR from whole blood, without DNA extraction." Nucleic Acids Res **18**(19): 5908.
141. Mergny, J. L. and L. Lacroix (2003). "Analysis of thermal melting curves." Oligonucleotides **13**(6): 515-537.
142. Michelson, A. M. (1958). "Hyperchromicity and nucleic acids." Nature **182**(4648): 1502-1503.
143. Miescher, F. (1871). "Ueber die chemische Zusammensetzung der Eiterzellen." Medicinisch-chemische Untersuchungen **4**: 441-460.
144. Mikulecky, P. J. and A. L. Feig (2006). "Heat capacity changes associated with DNA duplex formation: salt- and sequence-dependent effects." Biochemistry **45**(2): 604-616.
145. Mikulecky, P. J. and A. L. Feig (2006). "Heat capacity changes associated with nucleic acid folding." Biopolymers **82**(1): 38-58.
146. Mitra, S. (2003). Sample Preparation Techniques in Analytical Chemistry. Hoboken, New Jersey, USA, Wiley-Interscience.
147. Morrison, C. and F. Gannon (1994). "The impact of the PCR plateau phase on quantitative PCR." Biochim Biophys Acta **1219**(2): 493-498.
148. Morrison, T. B., J. J. Weis, et al. (1998). "Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification." Biotechniques **24**(6): 954-958, 960, 962.
149. Mullis, K., F. Faloona, et al. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." Cold Spring Harb Symp Quant Biol **51 Pt 1**: 263-273.
150. Mullis, K. B. (1990). "The unusual origin of the polymerase chain reaction." Sci Am **262**(4): 56-61, 64-55.
151. Murphy, J. and S. A. Bustin (2009). "Reliability of real-time reverse-transcription PCR in clinical diagnostics: gold standard or substandard?" Expert Rev Mol Diagn **9**(2): 187-197.
152. Murugan, R. (2002). "Revised theory on DNA renaturation kinetics and its experimental verification." Biochem Biophys Res Commun **293**(2): 870-873.
153. Nakano, S., M. Fujimoto, et al. (1999). "Nucleic acid duplex stability: influence of base composition on cation effects." Nucleic Acids Res **27**(14): 2957-2965.
154. Nath, K., J. W. Sarosy, et al. (2000). "Effects of ethidium bromide and SYBR Green I on different polymerase chain reaction systems." J Biochem Biophys Methods **42**(1-2): 15-29.
155. Nielsen, C. B., S. K. Singh, et al. (1999). "The solution structure of a locked nucleic acid (LNA) hybridized to DNA." J Biomol Struct Dyn **17**(2): 175-191.

156. Nielsen, K. E., J. Rasmussen, et al. (2004). "NMR studies of fully modified locked nucleic acid (LNA) hybrids: solution structure of an LNA:RNA hybrid and characterization of an LNA:DNA hybrid." *Bioconjug Chem* **15**(3): 449-457.
157. Nielsen, K. E., S. K. Singh, et al. (2000). "Solution structure of an LNA hybridized to DNA: NMR study of the d(CT(L)GCT(L)T(L)CT(L)GC):d(GCAGAAGCAG) duplex containing four locked nucleotides." *Bioconjug Chem* **11**(2): 228-238.
158. Nielsen, P. E., M. Egholm, et al. (1991). "Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide." *Science* **254**(5037): 1497-1500.
159. Nogva, H. K. and K. Rudi (2004). "Potential influence of the first PCR cycles in real-time comparative gene quantifications." *Biotechniques* **37**(2): 246-248, 250-243.
160. Nolan, T., R. E. Hands, et al. (2006). "Quantification of mRNA using real-time RT-PCR." *Nat Protoc* **1**(3): 1559-1582.
161. Nolan, T., R. E. Hands, et al. (2006). "SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations." *Anal Biochem* **351**(2): 308-310.
162. Obika, S., D. Nanbu, et al. (1998). "Stability and structural features of the duplexes containing nucleoside analogues with a fixed N-type conformation, 2'-O,4'-C-methylenribonucleosides." *Tetrahedron Letters* **39**(30): 5401-5404.
163. Obika, S., D. Nanbu, et al. (1997). "Synthesis of 2'-O,4'-C-methyleneuridine and -cytidine novel bicyclic nucleosides having a fixed C3'-endo sugar pucker." *Tetrahedron Letters* **38**: 8735-8738.
164. Ohta, T., S. Tokishita, et al. (2001). "Ethidium bromide and SYBR Green I enhance the genotoxicity of UV-irradiation and chemical mutagens in *E. coli*." *Mutat Res* **492**(1-2): 91-97.
165. Opel, K. L., D. Chung, et al. (2010). "A study of PCR inhibition mechanisms using real time PCR." *J Forensic Sci* **55**(1): 25-33.
166. Owczarzy, R., B. G. Moreira, et al. (2008). "Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations." *Biochemistry* **47**(19): 5336-5353.
167. Owczarzy, R., A. V. Tataurov, et al. (2008). "IDT SciTools: a suite for analysis and design of nucleic acid oligomers." *Nucleic Acids Res* **36**(Web Server issue): W163-169.
168. Owczarzy, R., Y. You, et al. (2011). "Stability and mismatch discrimination of locked nucleic acid-DNA duplexes." *Biochemistry* **50**(43): 9352-9367.
169. Owczarzy, R., Y. You, et al. (2004). "Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures." *Biochemistry* **43**(12): 3537-3554.
170. Panjkovich, A. and F. Melo (2005). "Comparison of different melting temperature calculation methods for short DNA sequences." *Bioinformatics* **21**(6): 711-722.

171. Panjkovich, A., T. Norambuena, et al. (2005). "dnaMATE: a consensus melting temperature prediction server for short DNA sequences." Nucleic Acids Res **33**(Web Server issue): W570-572.
172. Peirson, S. N., J. N. Butler, et al. (2003). "Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis." Nucleic Acids Res **31**(14): e73.
173. Petersen, M., C. B. Nielsen, et al. (2000). "The conformations of locked nucleic acids (LNA)." J Mol Recognit **13**(1): 44-53.
174. Petersen, M. and J. Wengel (2003). "LNA: a versatile tool for therapeutics and genomics." Trends Biotechnol **21**(2): 74-81.
175. Petrakis, N. L. (1953). "Microspectrophotometric estimation of the desoxyribonucleic acid (DNA) content of individual normal and leukemic human lymphocytes." Blood **8**(10): 905-915.
176. Pfaffl, M. (2000). "Development and Validation of an Externally Standardised Quantitative Insulin-like Growth Factor-I RT-PCR Using LightCycler SYBR Green I Technology." BIOCHEMICA.
177. Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time RT-PCR." Nucleic Acids Res **29**(9): e45.
178. Porschke, D. and M. Eigen (1971). "Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic--oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH." J Mol Biol **62**(2): 361-381.
179. Pugniere, P., S. Banzet, et al. (2011). "Pitfalls of reverse transcription quantitative polymerase chain reaction standardization: Volume-related inhibitors of reverse transcription." Anal Biochem.
180. Ramakers, C., J. M. Ruijter, et al. (2003). "Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data." Neurosci Lett **339**(1): 62-66.
181. Rasmussen, R. P. (2001). Quantification on the LightCycler. Rapid Cycle Real-time PCR, Methods and Applications. S. Meuer, C. T. Wittwer and K. Nakagawara. Heidelberg, Springer Press: 21-34.
182. Ririe, K. M., R. P. Rasmussen, et al. (1997). "Product differentiation by analysis of DNA melting curves during the polymerase chain reaction." Anal Biochem **245**(2): 154-160.
183. Rossen, L., P. Norskov, et al. (1992). "Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solutions." Int J Food Microbiol **17**(1): 37-45.
184. Rutledge, R. G. (2004). "Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications." Nucleic Acids Res **32**(22): e178.

185. Rutledge, R. G. and D. Stewart (2008). "Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time PCR." BMC Mol Biol **9**: 96.
186. Rychlik, W. (1995). "Selection of primers for polymerase chain reaction." Molecular Biotechnology **3**(2): 129-134.
187. Rychlik, W., W. J. Spencer, et al. (1990). "Optimization of the annealing temperature for DNA amplification in vitro." Nucleic Acids Res **18**(21): 6409-6412.
188. Saiki, R. K., D. H. Gelfand, et al. (1988). "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase." Science **239**(4839): 487-491.
189. Saiki, R. K., S. Scharf, et al. (1985). "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia." Science **230**(4732): 1350-1354.
190. Sanger, F. and A. R. Coulson (1975). "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase." J Mol Biol **94**(3): 441-448.
191. Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-5467.
192. SantaLucia, J., Jr. (1998). "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics." Proc Natl Acad Sci U S A **95**(4): 1460-1465.
193. SantaLucia, J., Jr. (2007). "Physical principles and visual-OMP software for optimal PCR design." Methods Mol Biol **402**: 3-34.
194. SantaLucia, J., Jr., H. T. Allawi, et al. (1996). "Improved nearest-neighbor parameters for predicting DNA duplex stability." Biochemistry **35**(11): 3555-3562.
195. SantaLucia, J., Jr. and D. Hicks (2004). "The thermodynamics of DNA structural motifs." Annu Rev Biophys Biomol Struct **33**: 415-440.
196. SantaLucia, J., Jr. and D. H. Turner (1997). "Measuring the thermodynamics of RNA secondary structure formation." Biopolymers **44**(3): 309-319.
197. Schildkraut, C. (1965). "Dependence of the melting temperature of DNA on salt concentration." Biopolymers **3**(2): 195-208.
198. Schlereth, W., I. D. Bassukas, et al. (1998). "Use of the recursion formula of the Gompertz function for the quantitation of PCR-amplified templates." Int J Mol Med **1**(2): 463-467.
199. Schmittgen, T. D. and K. J. Livak (2008). "Analyzing real-time PCR data by the comparative C(T) method." Nat Protoc **3**(6): 1101-1108.
200. Schneeberger, C., P. Speiser, et al. (1995). "Quantitative detection of reverse transcriptase-PCR products by means of a novel and sensitive DNA stain." PCR Methods Appl **4**(4): 234-238.

201. Schoppe, A., H. J. Hinz, et al. (1996). "Xylose-DNA: comparison of the thermodynamic stability of oligo(2'-deoxyxylonucleotide) and oligo(2'-deoxyribonucleotide) duplexes." Eur J Biochem **239**(1): 33-41.
202. Schroeder, A., O. Mueller, et al. (2006). "The RIN: an RNA integrity number for assigning integrity values to RNA measurements." BMC Mol Biol **7**: 3.
203. Schroeder, S. J. and D. H. Turner (2009). "Optical melting measurements of nucleic acid thermodynamics." Methods Enzymol **468**: 371-387.
204. Schultz, R. G. and S. M. Gryaznov (1996). "Oligo-2'-fluoro-2'-deoxynucleotide N3'-->P5' phosphoramidates: synthesis and properties." Nucleic Acids Res **24**(15): 2966-2973.
205. Sidon, P., P. Heimann, et al. (2006). "Combined locked nucleic acid and molecular beacon technologies for sensitive detection of the JAK2V617F somatic single-base sequence variant." Clin Chem **52**(7): 1436-1438.
206. Sikorav, J. L., H. Orland, et al. (2009). "Mechanism of Thermal Renaturation and Hybridization of Nucleic Acids: Kramers' Process and Universality in Watson-Crick Base Pairing." Journal of Physical Chemistry B **113**(12): 3715-3725.
207. Singh, S. K., R. Kumar, et al. (1998). "Synthesis of Novel Bicyclo[2.2.1] Ribonucleosides: 2'-Amino- and 2'-Thio-LNA Monomeric Nucleosides." J Org Chem **63**(18): 6078-6079.
208. Smith, M. J., R. J. Britten, et al. (1975). "Studies on nucleic acid reassociation kinetics: reactivity of single-stranded tails in DNA-DNA renaturation." Proc Natl Acad Sci U S A **72**(12): 4805-4809.
209. Smith, M. V., C. R. Miller, et al. (2007). "Absolute estimation of initial concentrations of amplicon in a real-time RT-PCR process." BMC Bioinformatics **8**: 409.
210. Southern, E. M. (1975). "Detection of specific sequences among DNA fragments separated by gel electrophoresis." J Mol Biol **98**(3): 503-517.
211. Spatz, H. C. and D. M. Crothers (1969). "The rate of DNA unwinding." J Mol Biol **42**(2): 191-219.
212. Stahlberg, A., P. Aman, et al. (2003). "Quantitative real-time PCR method for detection of B-lymphocyte monoclonality by comparison of kappa and lambda immunoglobulin light chain expression." Clin Chem **49**(1): 51-59.
213. Stahlberg, A., J. Hakansson, et al. (2004). "Properties of the reverse transcription reaction in mRNA quantification." Clin Chem **50**(3): 509-515.
214. Stahlberg, A., M. Kubista, et al. (2004). "Comparison of reverse transcriptases in gene expression analysis." Clin Chem **50**(9): 1678-1680.
215. Steger, G. (1994). "Thermal denaturation of double-stranded nucleic acids: prediction of temperatures critical for gradient gel electrophoresis and polymerase chain reaction." Nucleic Acids Res **22**(14): 2760-2768.

216. Stephenson, F. H. (2010). Calculations for Molecular Biology and Biotechnology: A Guide to Mathematics in the Laboratory. London, UK, Academic Press Inc.
217. Strick, T., J. F. Allemand, et al. (2000). "Twisting and stretching single DNA molecules." Progress in Biophysics & Molecular Biology: 115-140.
218. Sugimoto, N., S. Nakano, et al. (1995). "Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes." Biochemistry **34**(35): 11211-11216.
219. Sugimoto, N., S. Nakano, et al. (1996). "Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes." Nucleic Acids Res **24**(22): 4501-4505.
220. Tataurov, A. V., Y. You, et al. (2008). "Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids." Biophys Chem **133**(1-3): 66-70.
221. Taylor, C. F., D. Field, et al. (2008). "Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project." Nat Biotechnol **26**(8): 889-896.
222. Taylor, S., M. Wakem, et al. (2010). "A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines." Methods **50**(4): S1-5.
223. Teo, I. A., J. W. Choi, et al. (2002). "LightCycler qPCR optimisation for low copy number target DNA." J Immunol Methods **270**(1): 119-133.
224. Thastrup, O., S. Tullin, et al. (2005). Novel fluorescent proteins. united States, Biologie A/S. **US 2005/0054050 A1**: 30.
225. Thein, S. L. and R. B. Wallace (1986). The use of synthetic oligonucleotides as specific hybridization probes in the diagnosis of genetic disorders. Genetic Analysis of Human Diseases: A Practical Approach, . Herndon, Virginia, IRL Press: 160 pp. 133-150.
226. Thomas, P. S. (1980). "Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose." Proc Natl Acad Sci U S A **77**(9): 5201-5205.
227. Thomas, R. (1951). "[On the existence of labile bonds in & secondary structure of the nucleic acids' molecule]." Experientia **7**(7): 261-262.
228. Thomas, R. (1954). "[Research on the denaturation of desoxyribonucleic acids]." Biochim Biophys Acta **14**(2): 231-240.
229. Tichopad, A., A. Didier, et al. (2004). "Inhibition of real-time RT-PCR quantification due to tissue-specific contaminants." Mol Cell Probes **18**(1): 45-50.
230. Tichopad, A., M. Dilger, et al. (2003). "Standardized determination of real-time PCR efficiency from a single reaction set-up." Nucleic Acids Res **31**(20): e122.
231. Tichopad, A., J. Polster, et al. (2005). "Model of inhibition of *Thermus aquaticus* polymerase and Moloney murine leukemia virus reverse transcriptase by tea polyphenols (+)-catechin and (-)-epigallocatechin-3-gallate." J Ethnopharmacol **99**(2): 221-227.

232. Tolstrup, N., P. S. Nielsen, et al. (2003). "OligoDesign: Optimal design of LNA (locked nucleic acid) oligonucleotide capture probes for gene expression profiling." Nucleic Acids Res **31**(13): 3758-3762.
233. Ugozzoli, L. A., D. Latorra, et al. (2004). "Real-time genotyping with oligonucleotide probes containing locked nucleic acids." Anal Biochem **324**(1): 143-152.
234. Van Aerschot, A., J. Rozenski, et al. (1995). "An acyclic 5-nitroindazole nucleoside analogue as ambiguous nucleoside." Nucleic Acids Res **23**(21): 4363-4370.
235. Vandesompele, J., K. De Preter, et al. (2002). "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes." Genome Biol **3**(7): RESEARCH0034.
236. Vesnaver, G. and K. J. Breslauer (1991). "The contribution of DNA single-stranded order to the thermodynamics of duplex formation." Proc Natl Acad Sci U S A **88**(9): 3569-3573.
237. Vester, B. and J. Wengel (2004). "LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA." Biochemistry **43**(42): 13233-13241.
238. Vologodskii, A. V., B. R. Amirikyan, et al. (1984). "Allowance for heterogeneous stacking in the DNA helix-coil transition theory." J Biomol Struct Dyn **2**(1): 131-148.
239. von Ahsen, N., C. T. Wittwer, et al. (2001). "Oligonucleotide melting temperatures under PCR conditions: nearest-neighbor corrections for Mg(2+), deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas." Clin Chem **47**(11): 1956-1961.
240. Von Hippel, P. H., F. R. Fairfield, et al. (1994). "On the processivity of polymerases." Ann N Y Acad Sci **726**: 118-131.
241. Wallace, R. B., J. Shaffer, et al. (1979). "Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch." Nucleic Acids Res **6**(11): 3543-3557.
242. Wang, A. M., M. V. Doyle, et al. (1989). "Quantitation of mRNA by the polymerase chain reaction." Proc Natl Acad Sci U S A **86**(24): 9717-9721.
243. Wang, W., K. Chen, et al. (2006). "DNA quantification using EvaGreen and a real-time PCR instrument." Anal Biochem **356**(2): 303-305.
244. Watson, J. D. and F. H. Crick (1953). "The structure of DNA." Cold Spring Harb Symp Quant Biol **18**: 123-131.
245. Weaver, R. F. and C. Weissmann (1979). "Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15 S beta-globin mRNA precursor and mature 10 s beta-globin mRNA have identical map coordinates." Nucleic Acids Res **7**(5): 1175-1193.
246. Wengel, J., M. Petersen, et al. (2003). "Chemistry of locked nucleic acids (LNA): Design, synthesis, and bio-physical properties." Letters in Peptide Science **10**: 237-253.

247. Wetmur, J. G. (1976). "Hybridization and renaturation kinetics of nucleic acids." Annu Rev Biophys Bioeng **5**: 337-361.
248. Wetmur, J. G. (1991). "DNA probes: applications of the principles of nucleic acid hybridization." Crit Rev Biochem Mol Biol **26**(3-4): 227-259.
249. Wetmur, J. G. and N. Davidson (1968). "Kinetics of renaturation of DNA." J Mol Biol **31**(3): 349-370.
250. Wiedbrauk, D. L., J. C. Werner, et al. (1995). "Inhibition of PCR by aqueous and vitreous fluids." J Clin Microbiol **33**(10): 2643-2646.
251. Wiesner, R. J., B. Beinbrech, et al. (1993). "Quantitative PCR." Nature **366**(6454): 416.
252. Willems, E., L. Leyns, et al. (2008). "Standardization of real-time PCR gene expression data from independent biological replicates." Anal Biochem **379**(1): 127-129.
253. Williams, P. M. (2009). "The beginnings of real-time PCR." Clin Chem **55**(4): 833-834.
254. Wilson, I. G. (1997). "Inhibition and facilitation of nucleic acid amplification." Appl Environ Microbiol **63**(10): 3741-3751.
255. Wittwer, C. T., G. C. Fillmore, et al. (1990). "Minimizing the time required for DNA amplification by efficient heat transfer to small samples." Anal Biochem **186**(2): 328-331.
256. Wittwer, C. T., M. G. Hermann, et al. (1997). "Continuous fluorescence monitoring of rapid cycle DNA amplification." Biotechniques **22**(1): 130-131, 134-138.
257. Wittwer, C. T. and N. Kusakawa (2004). Real-time PCR. Molecular microbiology: diagnostic principles and practice. D. H. Persing, F. C. Tenover, J. Versalovic et al. Washington, ASM Press: 71-84.
258. Wittwer, C. T., K. M. Ririe, et al. (1997). "The LightCycler: a microvolume multisample fluorimeter with rapid temperature control." Biotechniques **22**(1): 176-181.
259. Wu, D. Y., L. Ugozzoli, et al. (1991). "The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction." DNA Cell Biol **10**(3): 233-238.
260. Xia, T., J. SantaLucia, Jr., et al. (1998). "Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs." Biochemistry **37**(42): 14719-14735.
261. Yin, X., J. Goudriaan, et al. (2003). "A flexible sigmoid function of determinate growth." Ann Bot **91**(3): 361-371.
262. You, Y., B. G. Moreira, et al. (2006). "Design of LNA probes that improve mismatch discrimination." Nucleic Acids Res **34**(8): e60.
263. You, Y., A. V. Tataurov, et al. (2011). "Measuring thermodynamic details of DNA hybridization using fluorescence." Biopolymers **95**(7): 472-486.

264. Yue, S. T., V. L. Singer, et al. (1997). Substituted unsymmetrical cyanine dyes with selected permeability. United States, Molecular Probes, Inc. **5,658,751**: 63.
265. Zhang, J. and C. D. Byrne (1999). "Differential priming of RNA templates during cDNA synthesis markedly affects both accuracy and reproducibility of quantitative competitive reverse-transcriptase PCR." Biochem J **337** ( Pt 2): 231-241.
266. Zhao, S. and R. D. Fernald (2005). "Comprehensive algorithm for quantitative real-time polymerase chain reaction." J Comput Biol **12**(8): 1047-1064.
267. Zimm, B. H. (1960). "Theory of "Melting" of the Helical Form in Double Chains of the DNA Type " Journal of Chemical Physics **33**(5): 1349-1356.
268. Zinn, K., D. DiMaio, et al. (1983). "Identification of two distinct regulatory regions adjacent to the human beta-interferon gene." Cell **34**(3): 865-879.
269. Zipper, H., H. Brunner, et al. (2004). "Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications." Nucleic Acids Res **32**(12): e103.
270. Zubay, G. (1959). "The interaction of nucleic acid with Mg-ions." Biochim Biophys Acta **32**(1): 233-236.
271. Zuker, M. (1999). Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide. St Louis.
272. Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." Nucleic Acids Res **31**(13): 3406-3415.



# Titre : Contribution à l'Amélioration de la Quantification des Acides Nucléiques par qPCR et RT-qPCR

## Résumé

La qPCR est actuellement la technique de référence en matière de quantification d'ADN. Elle peut être définie comme une amplification exponentielle, cyclique et ciblée de la séquence d'ADN cible. Le caractère exponentiel de la qPCR est à la fois à l'origine de la sensibilité de la méthode mais aussi d'une potentielle variabilité inter-échantillons. Cette variabilité est compensée par le caractère cyclique de la méthode qui entraîne une synchronisation de la réaction pour tous les échantillons à chaque cycle. L'amplification ciblée de la séquence choisie traduit quand à elle la spécificité de la méthode. Néanmoins, cette dernière propriété de la PCR reste la moins vérifiée. La spécificité de la qPCR est indiscutable lorsque la cible à détecter se trouve en quantité suffisante (> 100 copies environ). En revanche, pour des quantités plus faibles ou en présence d'inhibiteurs, la spécificité diminue ou disparaît (limites de détection et de quantification). Cette perte de spécificité retentit de façon significative sur la précision et la reproductibilité du dosage à réaliser. Cependant, si l'hybridation non spécifique est inéluctable dans ces conditions, l'amplification non spécifique consécutive peut être limitée, voire supprimée au moyen d'amorces ancrées dans leur partie en 5' et non étendues par la polymérase. La RT-qPCR, grâce à une étape initiale de transcription inverse (conversion d'un ARN en un ADN complémentaire) permet la quantification des ARN. Cependant, la transcription inverse reste moins reproductible que la PCR, générant des différences inter-échantillons délétères en diagnostic comme en recherche.

Au cours de ces travaux, je propose des méthodes originales améliorant de façon significative différentes étapes de ces techniques. Premièrement, je propose une amélioration de la standardisation de l'étape de transcription inverse capable de diminuer de façon significative la variabilité inter-échantillons ; l'utilisation d'un volume constant d'extrait d'ARN pour chaque échantillon améliore considérablement la précision de la quantification d'ARN messagers. Deuxièmement, je propose une modification des amorces par incorporation de monomères d'acides nucléiques bloqués (*LNA*) ; cette modification permet d'augmenter la spécificité des amorces aux limites de la détection. Enfin, je propose une méthode simple et économique permettant la mesure directe de la température de fusion ( $T_m$ ) dans les conditions réelles de la PCR. Ce paramètre, considéré comme capital pour la réalisation des dosages par qPCR, est généralement obtenu par des méthodes prédictives peu précises. De plus, cette méthode doit permettre de déterminer précisément les paramètres thermodynamiques des amorces ( $\Delta G$ ,  $\Delta H$  et  $\Delta S$ ) et ainsi avoir accès d'une part au pourcentage réel d'amorces hybridées en fonction de la température et d'autre part d'identifier la susceptibilité des amorces aux inhibiteurs.

## Mots-clés

*RT-qPCR, recommandations MIQE, transcription inverse, acides nucléiques bloqués LNA, température de fusion, amorce oligonucléotidique, quantification d'ARNm, standardisation, inhibition*

## Title: Contribution to the Nucleic Acids Quantification Improvement Using qPCR and RT-qPCR

## Abstract

qPCR is currently the gold standard for DNA quantification of DNA. It can be defined as an exponential, cyclic and targeted amplification of a known DNA sequence. The exponential character of qPCR is both the cause of the sensitivity of the method but also a potential variability between samples. This variability is offset by the cyclical nature of the method that causes a synchronization of the reaction for all samples during each cycle. The targeted amplification of the selected sequence reflects the specificity of the method. Nevertheless, this last property of PCR remains the least tested. The specificity of qPCR is unquestionable when the target to detect is sufficient (> 100 copies). However, for smaller quantities or in the presence of inhibitors, the specificity decreases or disappears (limits of detection and quantification). This loss of specificity will have a significant effect upon the accuracy and reproducibility of the assay to be performed. However, if non-specific hybridizations are unavoidable in these conditions, consecutive nonspecific amplification may be limited or eliminated using primers anchored in their 5' part and not extended by the polymerase. RT-qPCR allows RNA quantification because of an initial step of Reverse Transcription (RNA conversion into complementary DNA). However, Reverse Transcription is less reproducible than PCR, generating harmful inter-sample differences both in diagnosis and research fields.

In this work, I propose innovative methods improving significantly different steps of these techniques. First, I propose an improved standardization of the Reverse Transcription step that can significantly reduce the variability between samples; using a constant volume of RNA extract for each sample substantially improves mRNA quantification accuracy. Second, I propose a primers modification by incorporating *Locked Nucleic Acid* residues (*LNA*); this modification increases the specificity of the primers to the limits of detection. Finally, I propose a simple and economical method for direct measurement of the melting temperature ( $T_m$ ) in real PCR conditions. This essential parameter in the achievement of qPCR assays is generally obtained by predictive methods with low accuracy. Furthermore, this method should determine the thermodynamic parameters of the oligonucleotide sequence ( $\Delta G$ ,  $\Delta H$  and  $\Delta S$ ), on the one hand allowing access to the actual percentage of hybridized primers according to the temperature and on the other hand, to identify the primers susceptibility in the presence of inhibitors.

## Keywords

*RT-qPCR, MIQE guidelines, reverse transcription, Locked Nucleic Acid, melting temperature, oligonucleotide primer, mRNA quantification, standardization, inhibition*