



**HAL**  
open science

**Etude structurale par RMN des changements  
conformationnels de la séquence SL1 de l'ARN de  
l'isolat VIH-1 laï**

Fabien Kieken

► **To cite this version:**

Fabien Kieken. Etude structurale par RMN des changements conformationnels de la séquence SL1 de l'ARN de l'isolat VIH-1 laï. Autre. Université d'Orléans, 2004. Français. NNT : . tel-00007782

**HAL Id: tel-00007782**

**<https://theses.hal.science/tel-00007782>**

Submitted on 16 Dec 2004

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

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



THESE  
PRESENTEE  
A L'UNIVERSITE D'ORLEANS  
POUR OBTENIR LE GRADE DE  
DOCTEUR DE L'UNIVERSITE D'ORLEANS

Discipline : **CHIMIE**

PAR

*Fabien KIEKEN*

*Etude structurale par RMN des changements conformationnels de  
la séquence SL1 de l'ARN de l'isolat VIH-1<sub>lat</sub>*

Soutenue le 2 avril 2004

**MEMBRES DU JURY :**

- M. F. DARDEL	Directeur de Recherche, Fac. Pharmacie Paris V	Rapporteur
- M. O. MAUFFRET	Directeur de Recherche, IGR Villejuif	Rapporteur
- M. G. GUILLAUMET	Professeur, Université d'Orléans	Examineur
- M. J. PAOLETTI	Professeur, Université d'Orléans	Examineur
- M <sup>me</sup> . F. KEITA-PAQUET	Chargé de Recherche, CBM	Co-Directeur de Thèse
- M. G. LANCELOT	Directeur de Recherche, CBM	Co-Directeur de Thèse

# PREAMBULE

Bien que les premiers cas de syndrome d'immunodéficience acquise (SIDA) aient été décrits dès 1981, ce n'est qu'en 1983 que la nature virale de l'agent infectieux fut découverte. Le virus de l'immunodéficience humaine (VIH) fut isolé pour la première fois par l'équipe du Pr. Montagnier cette même année, et identifié comme la cause du SIDA. Son développement à l'échelle de la planète est tel qu'on a substitué le terme d'épidémie à celui de pandémie. L'organisation mondiale pour la santé (OMS) estime que 42 millions de personnes étaient touchés par le VIH en 2002 dont près de 30 millions uniquement en Afrique sub-saharienne. Si la progression de la pandémie semble stagner en Europe et en Amérique du nord, des études très récentes révèlent une très forte progression en Asie du Sud-Est. Le SIDA est désormais l'une des principales causes de mortalité dans le monde, l'Afrique étant tout particulièrement sinistré.

Le développement de multi-thérapies contre le VIH permet pour le moment une prolongation significative de la durée de vie des patients. Ces multi-thérapies se heurtent cependant à la capacité de mutation du VIH et à l'apparition rapide de résistances aux différents agents dirigés contre lui. Il apparaît donc essentiel d'identifier de nouvelles cibles virales de manière à lutter plus efficacement contre la maladie.

Notre travail porte sur la compréhension au niveau moléculaire du mécanisme d'interaction de la dimérisation du génome de VIH. La dimérisation est une étape clé du cycle rétroviral. Il est entre autre corrélé aux étapes d'encapsidation de l'ARN génomique et de maturation des particules virales. L'inhibition de la dimérisation représente une nouvelle voie de traitement potentiel du VIH.

# Table des Matières

<b>ABREVIATIONS</b> .....	<b>7</b>
<b>INTRODUCTION</b> .....	<b>9</b>
<b>I GENERALITES SUR LES RETROVIRUS</b> .....	<b>10</b>
<i>I-1 Taxonomie et caractéristiques rétrovirales</i> .....	10
<i>I-2 Aspects structuraux des particules virales</i> .....	11
<i>I-3 Organisation du génome rétroviral</i> .....	13
<i>I-4 Cycle rétroviral de VIH-1</i> .....	15
I-4-1 Entrée du virus dans la cellule .....	16
I-4-2 Transcription inverse de l'ARN .....	16
I-4-3 Intégration de L'ADN proviral au génome de la cellule hôte .....	17
I-4-4 Expression du génome viral .....	17
I-4-5 Traduction des protéines .....	17
I-4-6 Assemblage et encapsidation .....	17
I-4-7 Bourgeonnement et maturation .....	18
<i>I-5 La dimérisation de L'ARN génomique</i> .....	18
I-5-1 Un génome diploïde .....	18
I-5-2 Identification de la séquence responsable de la dimérisation .....	19
I-5-3 Initiation de la dimérisation par la formation d'un complexe boucle-boucle .....	20
I-5-4 Maturation du dimère boucle-boucle en dimère étendu par la protéine NCp7 ...	22
I-5-4-a. Propriétés générales de la protéine de la nucléocapside NC .....	22
I-5-4-b. Rôle de la NCp7 dans la maturation du dimère .....	22
I-5-4-c. Mécanisme de passage du dimère instable au dimère stable .....	25
<i>I-6 La dimérisation et ses implications dans le cycle de VIH-1</i> .....	26
I-6-1 Dimérisation et Encapsidation .....	27
I-6-2 Dimérisation et Recombinaison génétique .....	27
I-6-3 Dimérisation et Traduction .....	28
I-6-4 Dimérisation chez d'autres rétrovirus .....	28
<b>II DETERMINATION STRUCTURALE DES ACIDES NUCLEIQUES PAR RMN</b> .....	<b>30</b>
<i>II-1 Stratégies d'attributions des acides nucléiques</i> .....	31
Complexité de l'étude des ARN par RMN .....	31

<i>II-2 Informations structurales révélées par RMN</i> .....	32
II-2-1 Mesure des distances .....	32
II-2-2 Mesure des angles dièdres .....	33
II-2-3 Le déplacement chimique du proton .....	34
II-2-4 Le déplacement chimique du phosphore .....	34
II-2-5 Les liaisons hydrogène .....	34
II-2-6 Méthodologie d'attribution séquentielle .....	35
<i>II-3 Les Nouvelles Approches</i> .....	36
II-3-1 Marquage Isotopique .....	36
II-3-2 TROSY .....	37
II-3-3 Les couplages dipolaires résiduels .....	37
<b>III MODELISATION MOLECULAIRE D'ACIDES NUCLEIQUES A PARTIR DE DONNEES RMN</b> .....	38
<i>III-1 La fonction d'énergie potentielle des logiciels XPLOR et CNS</i> .....	38
III-1-1 Les différents termes de l'énergie chimique .....	39
III-1-2 Les termes effectifs .....	39
<i>III-2 Algorithmes de modélisation</i> .....	40
III-2-1 Mécanique Moléculaire .....	40
III-2-1-a. La méthode de plus forte pente .....	41
III-2-1-b. La méthode du gradient conjugué .....	41
III-2-2 Dynamique Moléculaire .....	42
III-2-2-a. Méthode du recuit simulé .....	43
III-2-2-b. Recuit dans l'espace cartésien .....	44
III-2-2-c. Recuit dans l'espace des angles dièdres .....	44
<b>RESULTATS</b> .....	<b>46</b>
PROBLEMATIQUE .....	47
<b>I ETUDE DE LA FORME MONOMERIQUE DE LA TIGE BOUCLE SL1</b> .....	50
<i>I-1 Préparation de l'échantillon</i> .....	50
<i>I-2 Etude RMN</i> .....	51
I-2-1 Etude dans l'eau .....	51
I-2-2 Etude dans l'eau lourde .....	53
I-2-2-a. Mise en évidence des H2 des adénines .....	53
I-2-2-b. Attribution séquentielle ribose-base .....	53
I-2-3 Spectre Phosphore .....	57

<i>I-3 Structure du monomère SL1*</i> .....	58
I-3-1 SL1* est monomérique.....	58
I-3-2 Modélisation moléculaire.....	60
I-3-2-a. Etablissement des contraintes .....	60
Contraintes sur les angles.....	60
Contraintes de distance.....	60
I-3-2-b. Etablissement des incertitudes sur les contraintes .....	61
I-3-2-c. Résolution de la structure tridimensionnelle.....	62
I-3-3 Analyse structurale de la tige boucle SL1* .....	64
<i>I-4 Simulation de dynamique moléculaire</i> .....	68
<i>I-5 Discussion</i> .....	71
<i>I-6 Conclusion</i> .....	72
II ETUDE DU DIMERE INSTABLE DU DOMAINE SL1 DE VIH.....	74
<i>II-1 Bilan des études structurales menées sur le domaine SL1 de VIH</i> .....	74
II-1-1 Etude des dimères stables .....	74
II-1-2 Etude sur les dimères instables .....	77
<i>II-2 Pourquoi s'intéresser de nouveau à la forme instable de VIH-1 Lai ?</i> .....	79
<i>II-3 Préparation de l'échantillon</i> .....	80
II-3-1 Synthèse de l'oligonucléotide .....	80
II-3-2 Préparation de l'échantillon .....	82
II-3-2-a. Sélection de la forme stable et instable.....	83
II-3-2-b. Ajout de sels .....	85
II-3-2-c. Caractérisation spectrale des deux formes de dimères .....	86
<i>II-4 Etude RMN</i> .....	87
II-4-1 Etude dans l'eau légère .....	87
II-4-1-a. Les spectres HSQC <sup>15</sup> N.....	87
II-4-1-b. Spectres HNN-COSY .....	88
II-4-1-c. Attribution des protons échangeables .....	89
II-4-1-d. Attribution des protons non échangeables.....	91
II-4-2 Etude dans l'eau lourde.....	91
II-4-2-a. Spectre HSQC <sup>13</sup> C .....	93
II-4-2-b. Attribution des protons H2' et H3' .....	94
II-4-2-c. Le dimère instable est un dimère boucle-boucle .....	96

II-4-2-c-a Spectre NOESY non irradié d'un mélange marqué- non marqué (1/1)	96
II-4-2-c-b Expérience de demi-filtre	97
II-4-2-d. Sommes nous en présence d'un dimère ou d'un tétramère?	102
<i>II-6 Modélisation moléculaire</i>	<i>105</i>
II-6-1 Etablissement des contraintes	105
II-6-2 Résolution de la structure tridimensionnelle	107
<i>II-7 Discussion</i>	<i>122</i>
<i>II-8 Conclusion</i>	<i>125</i>
<b>CONCLUSION ET PERSPECTIVES</b>	<b>126</b>
<b>ANNEXE</b>	<b>130</b>
<b>MATERIELS ET METHODES</b>	<b>132</b>
<b>BIBLIOGRAPHIE</b>	<b>146</b>

# ABREVIATIONS

$\tau_m$	Temps de mélange
1D	Unidimensionnel
2D	Bidimensionnel
ADN	Acide désoxyribonucléique
ALV	Virus de la leucose aviaire
ARN	Acide ribonucléique
ARNm	ARN messenger
ARNt	ARN de transfert
CA	Capside
COSY	CORrelation SpectroscopY
DIS	Dimerization Initiation Site
DLS	Dimer Linkage Structure
env	Enveloppe
FIV	Virus de l'immunodéficience féline
gag	Group specific antigen
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear single quantum coherence
HTLV	Virus de la leucémie des cellules T
IN	Intégrase
LTR	Long Terminal Repeat
MA	Matrice
MMTV	Virus de la tumeur mammaire de la souris
MoMuLV	Virus de la leucémie murine de Moloney
NC	NucléoCapside
Nef	Negative factor
nOe	Effet Overhauser Nucléaire
NOESY	Nuclear Overhauser Effect SpectroscopY
PAGE	PolyAcrylamide Gel Electrophoresis



PBS	Primer Binding Site
PEG	PolyEthylène Glycol
pol	Polymérase
ppm	Partie Par Million
PPT	PolyPurine Tract
PR	Protéase
R	Séquence redondante située aux deux extrémités de l'ARN génomique
Rev	Regulator of virion expression
RMN	Résonance Magnétique Nucléaire
RNase H	Ribonucléase H
RRE	Rev Response Element
RSV	Virus du sarcome de Rous
RT	Rétrotranscriptase ou Transcriptase inverse
SD	Splice Donor
SIV	Virus de l'immunodéficience simienne
SU	Surface
T <sub>2</sub>	Relaxation transversale
TAR	Trans Activating Response element ou élément de trans-activation
Tat	Trans AcTivator
TM	TransMembranaire
TOCSY	TOtal Correlation SpectroscopY
TROSY	TRansverse Optimized SpectroscopY
U3	Séquence unique située en 3' de l'ARN génomique
U5	Séquence unique située en 5' de l'ARN génomique
Vif	Virion Infectivity Factor
VIH-1 et -2	Virus de l'Immunodéficience Humaine de type 1 et 2
Vpr	Viral Protein R
Vpu	Viral Protein U

# INTRODUCTION

# GENERALITES SUR LES RETROVIRUS

## I-1 Taxonomie et caractéristiques rétrovirales

Les rétrovirus appartiennent à la famille des Retroviridae et infectent principalement les vertébrés. La plupart d'entre eux sont exogènes et leur transmission est réalisée par contagion entre individus distincts. D'autres, dits endogènes, sont intégrés au génome de l'hôte et sont transmis héréditairement.

Bien qu'ils soient capables d'infecter des cellules hôtes animales différentes et de provoquer des pathologies diverses, tous les rétrovirus possèdent des caractéristiques structurales et fonctionnelles communes qui permettent de les regrouper dans une même famille :

- L'information génétique est portée par un ARN simple brin au lieu d'un ADN. Celui-ci encode pour des polyprotéines constitutives (gag et env) et catalytiques (pol). D'autres gènes codant pour des protéines oncogènes, ou des protéines régulatrices peuvent être présents suivant que le génome du rétrovirus est simple ou complexe. Le terme rétrovirus provient du fait que leur cycle de réplication impose un passage du génome par une forme ADN, *rétro-* se réfère ainsi à la direction inhabituelle ARN-ADN. Ce passage s'effectue par la transcriptase inverse, une ADN polymérase ARN et ADN dépendante. L'ADN est ensuite intégré dans le génome de l'hôte pour servir de matrice à la synthèse de nouveaux ARN viraux, destinés soit à être traduits pour synthétiser des protéines virales, soit à être encapsidés et servir de support à l'information du prochain virion.
- Le génome est toujours diploïde, constitué de deux molécules homologues d'ARN génomiques.
- La morphologie est similaire. La particule virale est d'un diamètre variant entre 80 et 130 nm issue du bourgeonnement de la cellule hôte.

Ces virus peuvent être classés en différents groupes en fonction de leur morphologie et de leurs caractéristiques biologiques et biochimiques. (Coffin 1997) a ainsi décrit sept groupes (Tableau 1).

Groupe	Exemple de virus et d'isolat	Commentaires
ASLV	Virus du sarcome de Rous (RSV)	Exogène et oncogène
Type B	Virus de la tumeur mammaire de la souris (MMTV)	Endogène et exogène, carcinomes mammaires et lymphomes des cellules T
Type C	Virus de la leucémie murine (MoMuLV)	Exogène, lymphome des cellules T
Type D	Virus du singe Mason-Pfizer (MPMV)	Exogène, immunodéficience des singes
HTLV-BLV	Virus de la leucémie des cellules T (HTLV)	Lymphome des cellules T, désordres neurologiques
Lentivirus	Virus de l'immunodéficience humaine de type 1 et 2 (HIV-1 et -2) Virus de l'immunodéficience simienne (SIV) Virus de l'immunodéficience féline (FIV)	Syndrome de l'immunodéficience acquise
Virus spumeux	Virus "spumeux" simien	Exogène, bénin

**Tableau 1.** Classification des rétrovirus.

## I-2 Aspects structuraux des particules virales

La structure du virion de VIH-1, établie par microscopie électronique (Gelderblom et al. 1987), est caractéristique du groupe des lentivirus auquel il appartient. Les particules rétrovirales présentent un diamètre de 110 à 130 nm. Elles sont constituées d'une enveloppe lipidique entourant une capsidie icosaédrique contenant un noyau interne dense : la nucléocapsidie.

Chez VIH, la bicouche lipidique externe contient les glycoprotéines virales de surface SU (gp120) et transmembranaire TM (gp41). Gp 120 contient les déterminants viraux qui se lient aux récepteurs de la cellule hôte. Gp41 comprend un domaine terminal hydrophobe, qui est supposé initier la fusion membranaire entre le virus et la membrane cellulaire.

La membrane interne, ou matrice, est composée de la protéine de matrice MA (p17) et est associée à l'enveloppe.

La capsidie est formée par deux protéines issues du gène gag CA (p24), qui forme la capsidie elle-même, et la NCp pour la nucléocapsidie.

Le noyau nucléoprotéique comprend l'ARN génomique associé à des molécules ARN<sup>t<sub>lys</sub></sup>, servant d'amorce à la transcription inverse, ainsi qu'à de nombreux produits de maturation des polyprotéines Gag (NCp6) et Pol (Nef, Vif et Vpr). La protéine de nucléocapsidie (NC) s'associe sélectivement à l'ARN génomique.

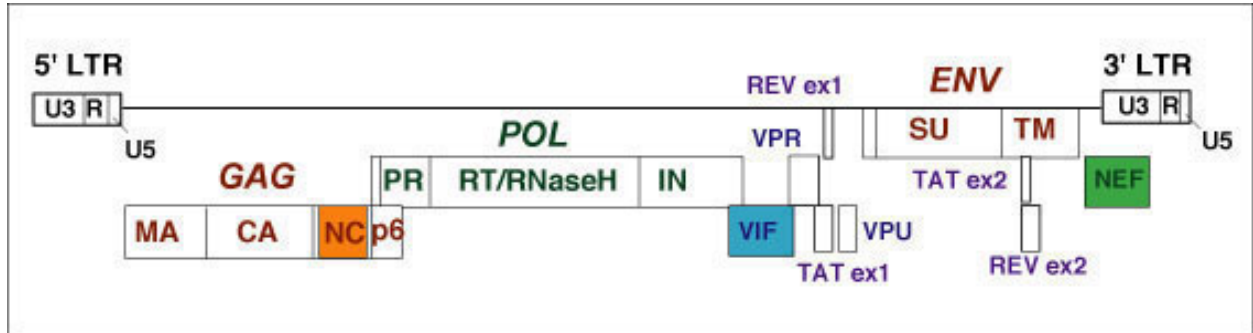
Les fonctions des différentes protéines sont détaillées dans le tableau 2 (Frankel and Young 1998).

<b>protéines virales</b>	<b>Fonctions</b>
<b>Protéines de l'enveloppe</b>	
<b>TM</b>	Protéine transmembranaire
<b>SU</b>	Protéine de surface
<b>Protéines structurales</b>	
<b>MA</b>	Protéines de la matrice myristilées à leur extrémité NH <sub>2</sub>
<b>CA</b>	Protéines de la capsid
<b>NC</b>	Protéines de nucléocapside, protéines chaperonnes
<b>Protéines enzymatiques</b>	
<b>RT</b>	Transcriptase inverse, responsable du passage ARN <sub>génomique</sub> en ADN <sub>proviral</sub>
<b>IN</b>	Intégrase, responsable de l'insertion de L'ADN <sub>proviral</sub> dans le patrimoine génétique de la cellule hôte
<b>PR</b>	Protéase, responsable de la maturation des précurseurs polyprotéiques
<b>Protéines régulatrices</b>	
<b>Nef</b>	(Negative Factor) augmente le taux de réplication et réduit le nombre de récepteurs CD4 à la surface de la cellule infectée
<b>Rev</b>	(Regulator of Virion expression) régule l'épissage des ARNm ainsi que leur transport à partir du noyau
<b>Tat</b>	(TransAcTivator) transactivateur de l'ADN <sub>proviral</sub>
<b>Vif</b>	(Virion Infectivity Factor) permet la formation des particules virales
<b>Vpr</b>	(Viral Protein R) intervient dans la translocation de L'ADN viral vers le noyau
<b>Vpu</b>	(Viral Protein U) augmente le nombre de virions relargués à la surface de la cellule infectée et diminue le niveau d'expression des récepteurs CD4

**Tableau 2.** Classification des différentes protéines impliquées dans le cycle de réplication du VIH.

## I-3 Organisation du génome rétroviral

Le génome rétroviral de VIH est encodé par une molécule d'ARN simple brin d'environ 10 kb (Figure 3).



**Figure 3.** Organisation générale de VIH-1.

Il comprend une large région codante encadrée de deux régions non codantes constituées par une répétition de séquences identiques les LTR (Long Terminal Repeat). Ces séquences LTR contiennent les signaux pour l'intégration de l'ADN proviral dans le génome de la cellule hôte et pour l'expression de protéines virales.

Ces régions non codantes contiennent :

- une coiffe en 5'
- une région R (Redondant) présente aux deux extrémités et qui participe à la rétrotranscription en permettant le saut de brin.
- Une région U5, c'est une région très structurée qui contient des séquences pour le processus d'hybridation entre l'ARN et l'ARN<sub>t</sub><sup>lys</sup><sub>3</sub>.
- Une région PBS (Primer Binding Site) qui est la région où s'effectue la fixation de l'ARN<sub>t</sub><sup>lys</sup><sub>3</sub> sur l'ARN génomique. Celle-ci sert ensuite d'amorce à la transcription inverse.
- Une région DLS (Dimer Linkage Site). Cette région contient le site d'initiation de la dimérisation DIS (Dimer Initiation Site). Elle possède également le site d'épissage majeur SD (Splice Donnor) et la séquence  $\psi$  d'encapsidation de l'ARN génomique (Figure 7).

Les régions codantes contiennent principalement 3 gènes gag, pol et env ainsi que des séquences intervenant dans le cycle rétroviral du virus :

- Le gène gag (group specific AntiGen) encode pour l'ensemble des protéines de structure qui donnera après clivage par la protéase virale les protéines de matrice (MA), de capsid (CA), de nucléocapside (NC).
- Le gène pol (POLymerase) est à l'origine des différentes protéines à activité enzymatique : l'intégrase (IN), la reverse transcriptase (RT) et la protéase (PR)
- Le gène env (ENVeloppe) encode pour les deux glycoprotéines d'enveloppe SU gp120 et TM gp41.
- La séquence RRE (Rev Response Element) localisée au niveau du gène Env interagit avec la protéine Rev pour l'exportation des ARNm non épissés et épissés vers le cytoplasme.
- La séquence PPT (PolyPurine Tract), très riche en purine ; elle est localisée en amont de U3 et sert d'amorce à la synthèse du brin (+) de l'ADN proviral.

## I-4 Cycle rétroviral de VIH-1

Le cycle répliatif (Figure 4) est commun à l'ensemble des rétrovirus. Celui-ci peut-être décomposé en un certain nombre d'étapes successives.

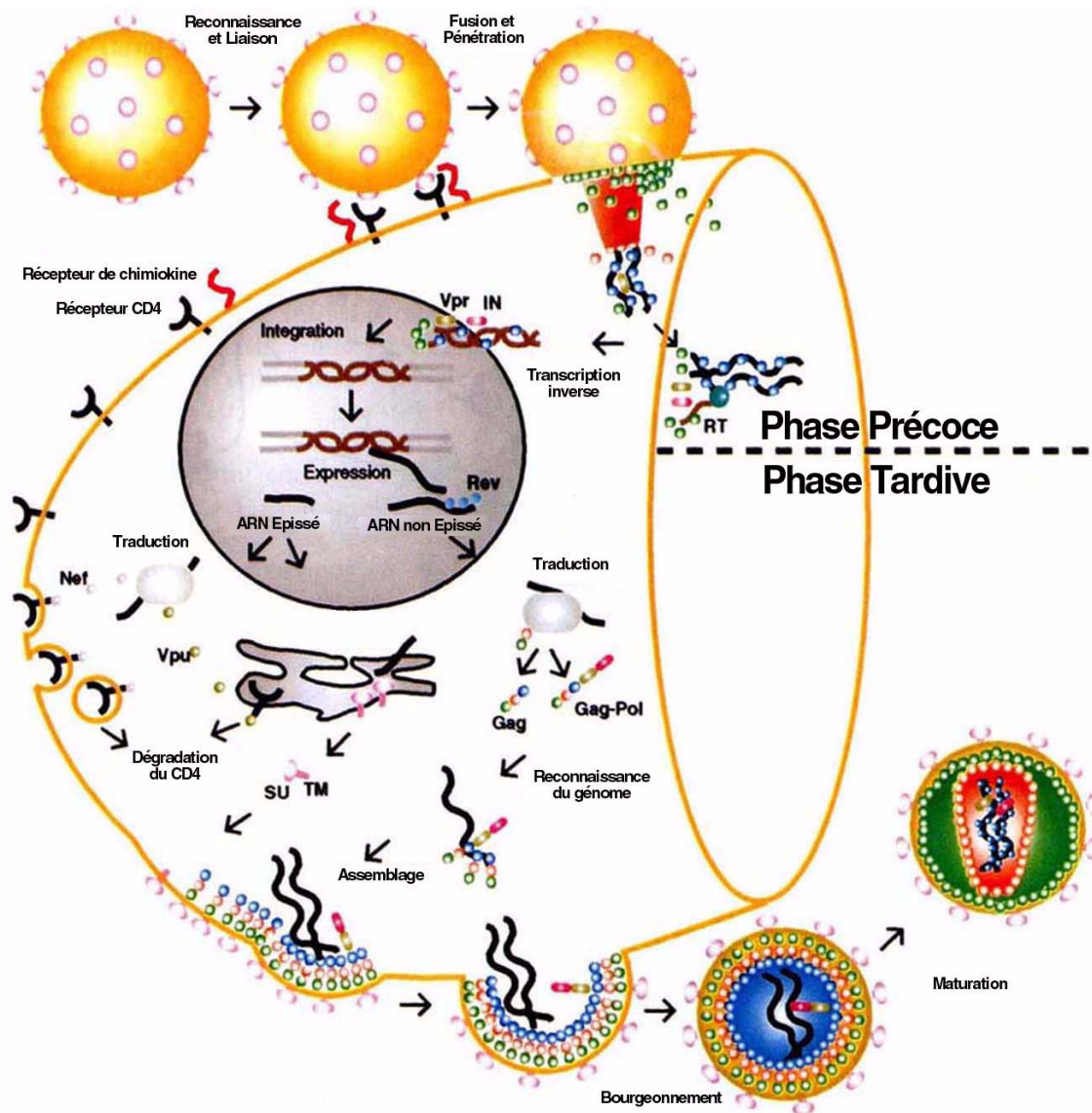


Figure 4. Cycle répliatif de VIH-1 (Turner and Summers 1999).



### I-4-1 Entrée du virus dans la cellule

L'infection débute par l'adhésion de la particule virale à des récepteurs membranaires de la cellule, les récepteurs CD4. Ces récepteurs sont des récepteurs glycoprotéiques extra-membranaires de la famille des immunoglobulines. Ceux-ci sont principalement exprimés à la surface des lymphocytes T4, des monocytes et des macrophages. L'adhésion du virion est assurée par l'interaction entre la glycoprotéine virale d'enveloppe, la gp120, et les récepteurs CD4 de la cellule hôte. Il existe également un certain nombre de co-récepteurs qui interviennent dans la reconnaissance du virion avec la cellule cible. A ce jour, deux types de co-récepteurs ont été identifiés la fusine (Feng et al. 1996a), qui est un récepteur transmembranaire couplé aux protéines G et les récepteurs de la  $\beta$ -chimiokine de type CCR5 (Deng et al. 1996; Dragic et al. 1996).

Il s'en suit une fusion des membranes et pénétration de la capsid virale dans le cytoplasme de la cellule.

### I-4-2 Transcription inverse de l'ARN

Après pénétration du virion, la transcription inverse de l'ARN génomique est engagé. A l'issu de ce processus, l'ARN simple brin dimérique est converti en ADN double brin possédant à chaque extrémité des séquences LTR, nécessaires à l'amorçage de la rétro transcription par l'ARN<sub>t</sub><sup>lys</sup>. Cet ADN, associé aux protéines virales, sera transporté dans le noyau pour y être intégré au génome de la cellule hôte.

La transcriptase inverse ne possède pas d'activité correctrice. Il existe donc un taux particulièrement élevé de mutations au cours de la rétro transcription. Conjointement aux phénomènes de recombinaison, ce phénomène confère au virus une grande variabilité génétique ce qui lui permet d'échapper aux différentes thérapies dirigées contre lui. A l'heure actuelle, la transcriptase est une des cibles principales des agents thérapeutiques utilisées contre le VIH.

### I-4-3 Intégration de L'ADN proviral au génome de la cellule hôte

Lors de son transport, l'ADN proviral, issu de la transcription inverse de l'ARN du virus, est associé à un certain nombre de protéines virales comme la protéine de matrice MA. Cette protéine contient un signal de localisation nucléaire (Bukrinsky et al. 1993a; Bukrinsky et al. 1993b) qui permet le transport de l'ADN vers le noyau. Une enzyme virale, l'intégrase, permet l'intégration de l'ADN proviral au génome de la cellule infectée.

### I-4-4 Expression du génome viral

L'expression du génome viral se déroule en deux étapes:

- Une première étape où sont exprimées les protéines régulatrices comme Tat, Nef et Rev
- Une deuxième étape où seront traduites les différentes protéines structurales et accessoires

### I-4-5 Traduction des protéines

Les ARN<sub>m</sub> provenant de la transcription de l'ADN proviral vont servir à la traduction des protéines virales. Suivant le type d'ARN<sub>m</sub>, les protéines traduites ne seront pas les mêmes.

Les ARN<sub>m</sub> peuvent encoder pour des précurseurs poly-protéiques, qui après une étape de maturation donnent les différentes protéines de matrice et de capsid, et pour les trois enzymes virales qui sont la protéase, la transcriptase inverse et l'intégrase.

### I-4-6 Assemblage et encapsidation

L'assemblage est réalisé au niveau de la membrane plasmique, où se fixe la partie N-terminale du précurseur Pr55. Les domaines NC de Pr55 interviennent dans l'encapsidation de l'ARN génomique en se fixant à sa région  $\psi$  (Figure 7). Le virion encapsidé contient également un certain nombre de protéines nécessaires à son développement ainsi que l'ARN<sub>t</sub><sup>lys</sup>, amorce de la transcription inverse.

## I-4-7 Bourgeonnement et maturation

Lors du bourgeonnement, on a mise en place de la protéine Env (constituée de deux sous-unités TM et SU). L'acquisition de l'infectiosité est obtenue après une phase de maturation où les précurseurs Pr55 et Pr160 sont clivés par la protéase virale (PR). Ce n'est qu'après cette étape dite de maturation que le virion pourra infecter d'autres cellules.

## I-5 La dimérisation de L'ARN génomique

### I-5-1 Un génome diploïde

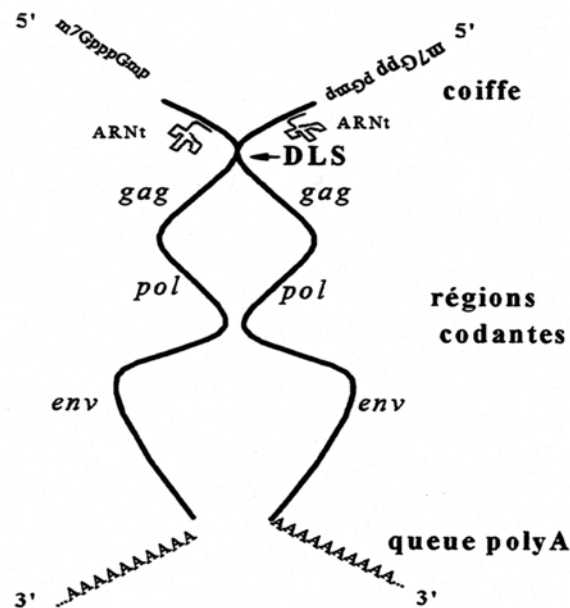
Une des caractéristiques communes à tous les rétrovirus est de posséder un génome sous forme diploïde. En effet, chaque particule virale contient deux copies identiques de l'ARN génomique (+). Il a été également mis en évidence que ces deux brins d'ARN s'associent de façon non covalente pour former des dimères (Canaani et al. 1973; Mangel et al. 1974). La microscopie électronique a pu montrer l'existence d'un site d'interaction ARN-ARN situé à l'extrémité 5' terminale du génome rétroviral (Bender and Davidson 1976; Kung et al. 1975; Murti et al. 1981) (Figure 5).



**Figure 5.** Image obtenue par cryo-microscopie de la dimérisation de l'ARN génomique des rétrovirus.

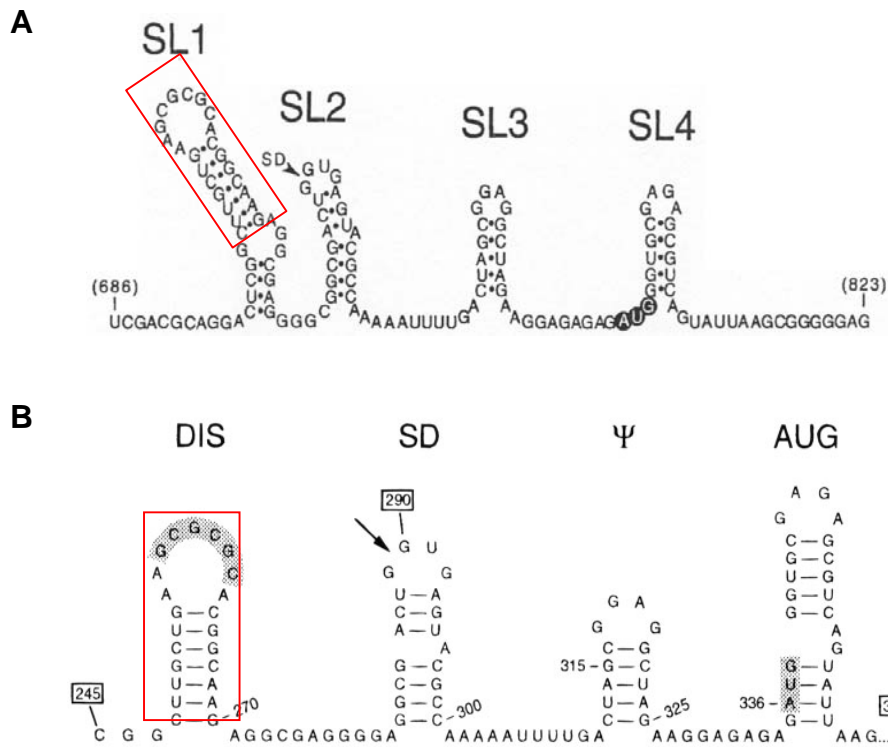
## I-5-2 Identification de la séquence responsable de la dimérisation

Il a été également démontré que cette région est capable de dimériser spontanément en solution au niveau 5' terminal de l'ARN génomique (Awang and Sen 1993; Marquet et al. 1991; Sundquist and Heaphy 1993; Weiss et al. 1993). La région prit alors la dénomination de DLS (Dimer Linkage Structure) (Figure 6).



**Figure 6.** Le génome rétroviral est diploïde (Bender and Davidson 1976; Coffin 1997). Le site de liaison, situé en position 5'-terminale, a été dénommé la DLS (Dimer Linkage Structure).

Plusieurs équipes ont travaillé sur les séquences responsables de l'initiation (DIS : Dimerization Initiation Site) de la dimérisation du génome rétroviral. Ces études ont été effectuées sur les deux sérotypes les plus répandus chez VIH-1, les isolats Mal et Lai. Les résultats obtenus sur l'isolat Mal indiquent que le DIS est une courte séquence auto-complémentaire GUGCAC située en amont du site donneur d'épissage au niveau du domaine  $\psi$  d'encapsulation (Marquet et al. 1994; Skripkin et al. 1994).



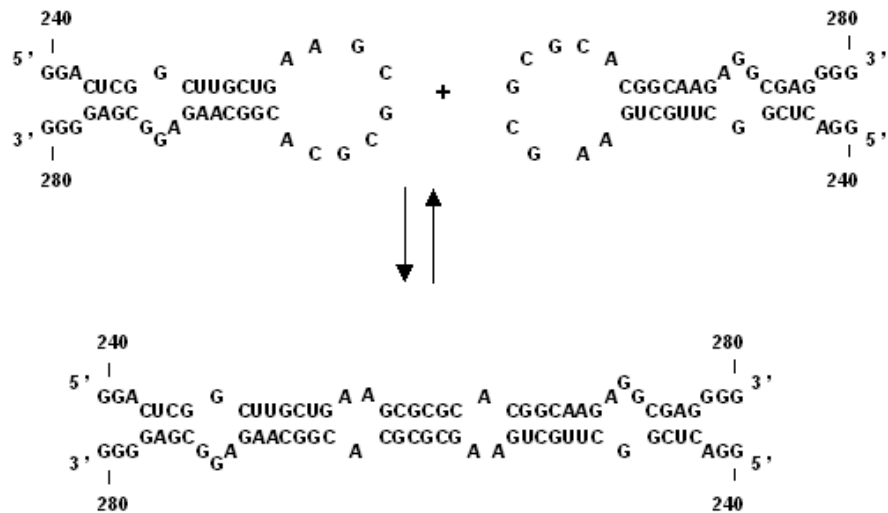
**Figure 7.** Structure secondaire du domaine d'encapsidation de la région leader de VIH-1Lai. **A/** Modèle proposé par Clever et al. (1995). L'extrémité de la tige boucle SL représente la séquence d'initiation de la dimérisation (DIS) ; **B/** Modèle proposé par Berkhout and van Wamel (2000), la tige boucle SL1 est une tige-boucle simple sans bulge à la base. La numérotation des résidus diffère en fonction que l'on considère ou non les régions LTR de la zone leader.

La séquence DIS de l'isolat Lai est légèrement différente: GCGCGC mais elle garde son auto-complémentarité (Laughrea and Jette 1994; Muriaux et al. 1995). Celle-ci est très conservée dans le génome rétroviral et n'est rencontrée qu'une seule fois. Par la suite, l'utilisation de sondes chimiques et enzymatiques couplée aux résultats de repliement par minimisation, ont montré que cette séquence se trouvait au sommet d'une des 4 tiges boucles (SL1) du domaine  $\psi$  d'encapsidation (Clever et al. 1995) (Figure 7).

### I-5-3 Initiation de la dimérisation par la formation d'un complexe boucle-boucle

Au vu de la complémentarité de la séquence apicale de la tige boucle SL1, un premier modèle a été proposé par l'équipe de Jacques Paoletti (Muriaux et al. 1996b). En effet, des oligonucléotides sens ou antisens de la séquence apicale de la tige boucle SL1 sont capables d'inhiber la dimérisation (Muriaux et al. 1995). Ces résultats ont pu être vérifiés par d'autres

équipes (Lodmell et al. 2000; Skripkin et al. 1996). Ce modèle implique une reconnaissance boucle-boucle des ARN viraux avant leur réarrangement en dimère double brin (Figure 8).



**Figure 8.** Modèle proposé pour la dimérisation de VIH-1Lai, d'après Muriaux et al. (1995).

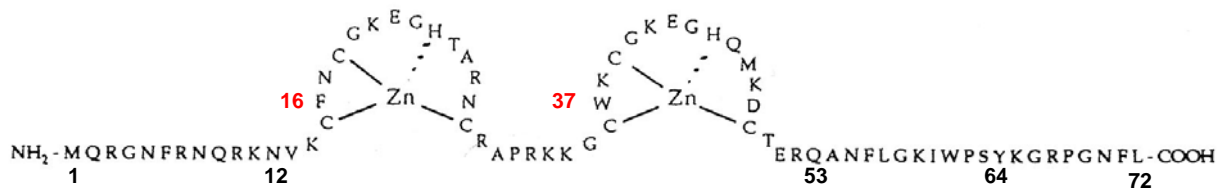
Cette interaction boucle-boucle dite en 'kissing complex' serait à l'origine de l'initiation de la dimérisation. Il a été également démontré que des mutations introduites au niveau des séquences auto complémentaires diminuent fortement la capacité à dimériser (Clever et al. 1996; Laughrea et al. 1999; Paillart et al. 1994) et que des mutations compensatoires permettent de rétablir la dimérisation *in vitro* (Paillart et al. 1994). Des études de sélection *in vitro* (Lodmell et al. 2001) montrent également qu'une boucle composée de 9 nucléotides, dont 6 auto-complémentaires, est optimale pour l'interaction boucle-boucle. Il est à noter que les trois purines qui sont adjacentes à la séquence auto-complémentaire sont très fortement conservées.

## I-5-4 Maturation du dimère boucle-boucle en dimère étendu par la protéine NCp7

Un grand nombre d'études ont pu mettre en évidence le rôle primordial de la protéine de nucléocapside NCp7 dans le phénomène de dimérisation de l'ARN génomique.

### I-5-4-a. Propriétés générales de la protéine de la nucléocapside NC

La protéine de nucléocapside NCp7 est issue du clivage par la protéase du précurseur polyprotéique Pr55gag. La séquence de la forme active de la protéine n'a toujours pas été clairement identifiée entre la NCp7(1-55) et la NCp7(1-72) (Morellet et al. 1992) (Figure 9). Elle comprend deux motifs hautement conservés [C-X2-C-X4-H-X4-C], où X est un acide aminé variable (Berg 1986; Covey 1986; Summers et al. 1990). Cette séquence s'organise autour d'un atome de zinc pour engendrer un complexe de coordination connu sous la dénomination de "doigt de zinc" (Cornille et al. 1990; Green and Berg 1990).



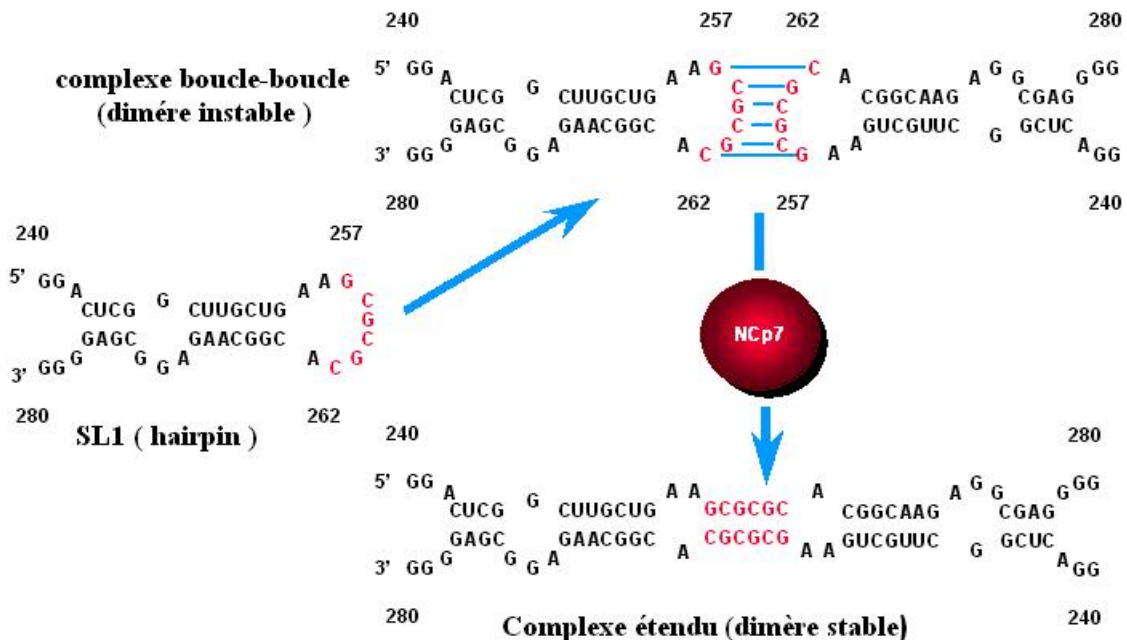
**Figure 9.** Structure secondaire en « doigt de zinc » de la NCp7(1-72).

La protéine possède également plusieurs régions basiques qui seraient susceptibles d'interagir avec l'ARN du rétrovirus.

### I-5-4-b. Rôle de la NCp7 dans la maturation du dimère

Des mutations au niveau de la protéase virale provoquent irrémédiablement la formation de dimères anormaux avec des propriétés physiques différentes des dimères observés dans le cas de virions sauvages (Fu et al. 1994). Par conséquent, la maturation du virion, indispensable à l'acquisition des propriétés de virulence, est provoquée par des protéines de clivage de précurseurs polyprotéiques et plus précisément le domaine NC de Pr55gag (Berkowitz et al.

1993; Darlix et al. 1990; Dorfman et al. 1993). Les équipes de Laughrea et Paoletti ont également pu mettre en évidence deux types de populations dimériques avec des stabilités thermiques différentes. Les interprétations faites pour expliquer la différence de stabilité entre les deux dimères suggèrent que le dimère le moins stable adopte une structure en ‘kissing complex’. Le dimère le plus stable, quant à lui, adopterait une structure en dimère étendu (Laughrea and Jette 1996a, b; Muriaux et al. 1996b). L’observation d’une telle transition a été démontrée en présence de la protéine NCp7 à 37° (Baba et al. 2001; Muriaux et al. 1996a; Takahashi et al. 2000a; Takahashi et al. 2001; Takahashi et al. 2000b) (Figure 10).

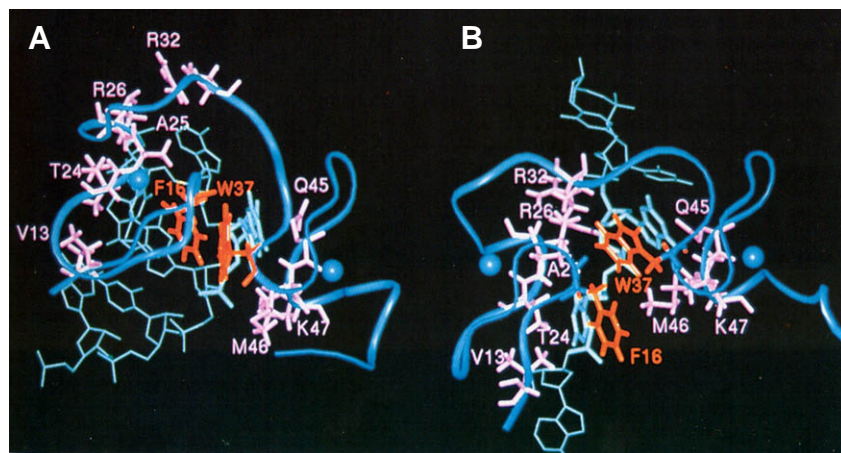


**Figure 10.** Séquence, structure secondaire et mécanisme de maturation du DIS (Muriaux et al. 1996a)

Les régions de la protéine NC intervenant dans l’interaction avec l’ARN génomique sont encore mal connues. L’intégrité des doigts de zinc semble être nécessaire pour avoir une encapsidation sélective de L’ARN (Aldovini and Young 1990; Dupraz et al. 1990; Gorelick et al. 1988; Meric and Spahr 1986; Morellet et al. 1994). Si les deux doigts de zinc sont enlevés et remplacés par des glycines, la protéine est toujours capable de se fixer à l’ARN mais non spécifiquement (De Rocquigny et al. 1992). Enfin, il apparaît que la délétion des acides aminés basiques localisés de part et d’autre des doigts de zinc conduit à la perte de la faculté de dimériser (Dannull et al. 1994; de Rocquigny et al. 1991; Lapadat-Tapolsky et al. 1993;



Surovoy et al. 1993). L'ensemble de ces résultats pourrait s'expliquer si les doigts de zinc avaient la capacité de maintenir une structure dans laquelle les résidus basiques interagiraient avec l'ARN (De Rocquigny et al. 1993). Néanmoins, l'existence d'une interaction unique entre les résidus de la tige boucle SL1 et les parties basiques de la protéine pour permettre l'isomérisation semble peu probable. En effet, les études du complexe nucléoprotéique par RMN entre la tige boucle SL3 et la NCp7 (De Guzman et al. 1998) montrent que les doigts de zinc sont en interaction avec les résidus de la tétra boucle (Figure 11).



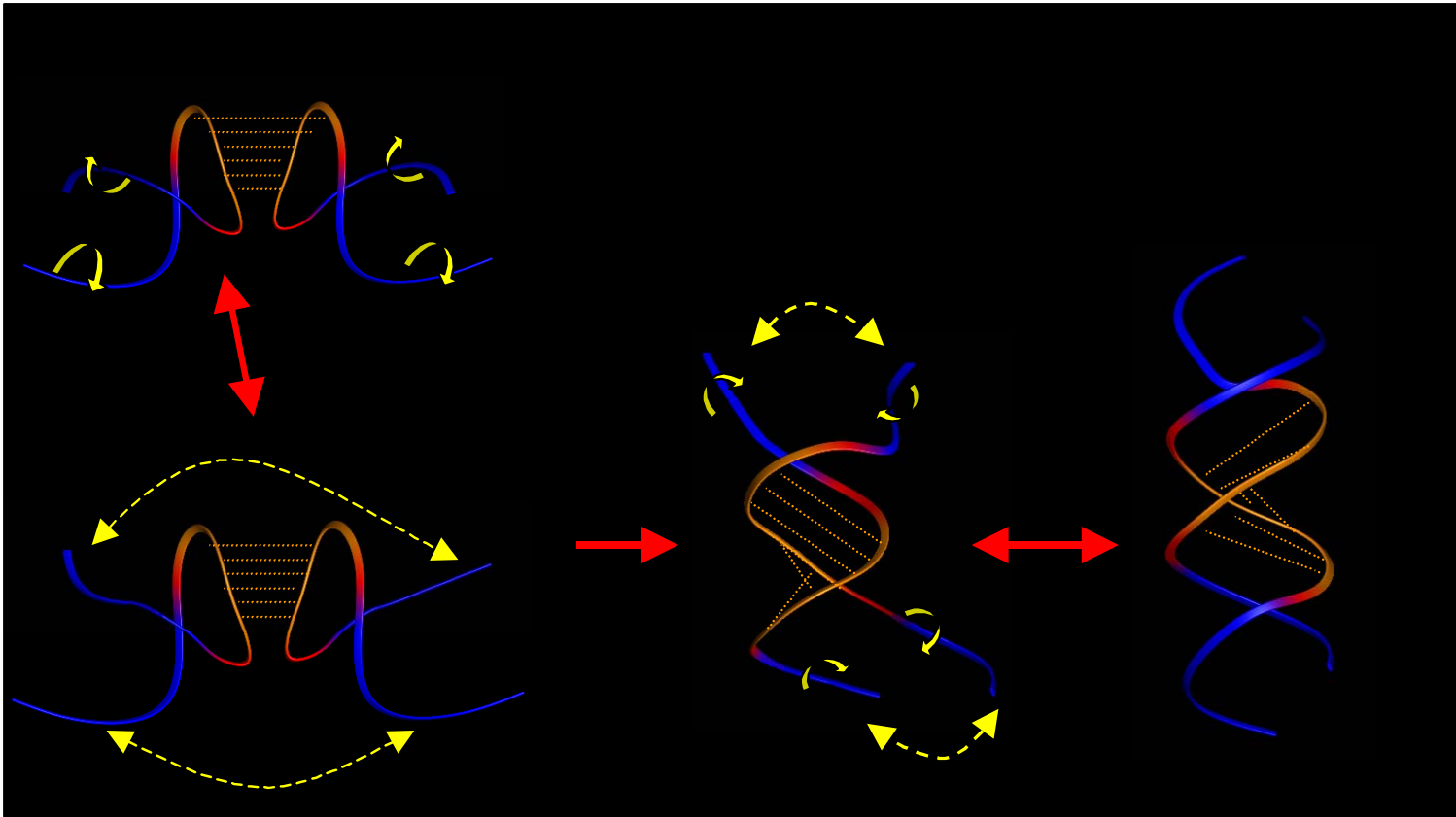
**Figure 11.** Comparaison de l'interaction NCp7(1-55)/SL3 et NCp7(12-53)/d(ACGCC). **A/** d'après De Guzman et al. (1998) ; **B/** d'après Morellet et al. (1998).

De plus, il a pu être mis en évidence le rôle du tryptophane 37 (W37) dans les interactions nucléoprotéiques par des études de fluorescence (Vuilleumier et al. 1999) et par des études RMN (Morellet et al. 1998) qui montrent une intercalation du tryptophane entre des résidus d'une séquence nucléique (ADN) (Figure 11). Notre équipe a également pu mettre en évidence la proximité du tryptophane W37 avec l'extrémité terminale de la tige SL123mer (Girard 2000). De plus, nous avons pu mettre en évidence les propriétés de polymorphisme du SL123mer. Ces observations s'opposent à celles publiées par Takahashi et al. (2000a) et Takahashi et al. (2000b) où les auteurs affirment que la dimérisation de la tige boucle nécessite la présence du bulge à la base de SL1.

#### I-5-4-c. Mécanisme de passage du dimère instable au dimère stable

Il existe deux mécanismes de transition pour expliquer le passage du dimère boucle-boucle en dimère étendu. Le premier mécanisme, proposé par Ennifar et al. (2001), est un mécanisme auto catalytique. Il résulte du passage d'une forme à une autre par double transestérification. Ce mécanisme repose sur le fait que le DIS est un site de recombinaison génétique (Balakrishnan et al. 2001; Lund et al. 1999; Mikkelsen et al. 1998, 2000; Mikkelsen and Pedersen 2000). Ce mécanisme rend effectivement compte des possibles recombinaisons qui peuvent avoir lieu. Il repose également sur la présence de résidus extra-hélicaux qui peuvent être à l'origine de coupures sur les brins d'ARN, notamment si la réaction est catalysée par des ions métalliques (Husken et al. 1996). La flexibilité introduite dans la chaîne d'ARN par le résidu non apparié permet d'obtenir, de manière transitoire, une structure en ligne favorable à la coupure. Ce type de structure avec des adénines extrahélicales et possédant la capacité d'un tel clivage a pu être mis en évidence par cristallographie (Tereshko et al. 2001).

Cependant, ce mécanisme semble peu vraisemblable et va à l'encontre de nombreux travaux. Le deuxième mécanisme (Theilleux-Delalande et al. 2000), proposé par notre équipe résulte de la fusion intramoléculaire des bases de la tige, en partant de la paire de bases la plus proche de la région boucle-boucle vers l'extrémité de la tige (Figure 12). Les paires de bases impliquées dans l'appariement boucle-boucle restent appariées tout au long du mécanisme. Ensuite, on assiste à un réappariement intermoléculaire des bases de la tige pour aboutir à la constitution d'un dimère étendu. Cette étude a pu être réalisée en suivant par RMN les résonances des protons imino qui ne peuvent être observées que s'ils sont appariés. Ce mécanisme a ensuite été validé par d'autres équipes et notamment par analyse par fluorescence (Rist and Marino 2002) et par clivage ribosomique (Windbichler et al. 2003).



**Figure 12.** Mécanisme de passage du dimère instable en dimère stable (Theilleux-Delalande et al. 2000)

## I-6 La dimérisation et ses implications dans le cycle de VIH-1

De nombreuses études démontrent que le caractère diploïde du génome est directement impliqué dans plusieurs étapes du cycle réplcatif et par conséquent dans l'infectiosité du virion. La région du DIS, responsable de l'initiation de la dimérisation, est située dans la région 5' terminale du génome qui contient un grand nombre d'éléments fonctionnels, tels que l'initiation de la transcription inverse, l'épissage ou la translation de l'ADN dans le noyau. En effet, la délétion du DIS ou des mutations au niveau de la séquence auto complémentaire diminue le taux de dimères et par conséquent l'infectiosité du virion (Bernacchi and Mely 2001; Clever and Parslow 1997; Haddrick et al. 1996; Laughrea et al. 1999; Liang et al. 1998; McBride et al. 1997; Paillart et al. 1996b; Paillart et al. 1996c; Shen et al. 2000). Cette baisse de la virulence s'accompagne de défauts d'encapsidation, ce qui prouve que les phénomènes de dimérisation et d'encapsidation sont étroitement liés.

## I-6-1 Dimérisation et Encapsidation

Les dernières étapes du cycle rétroviral impliquent une encapsidation sélective de l'ARN génomique dans les particules virales. L'ARN génomique, pourtant présent en très grande quantité par rapport aux ARNm, possèdent des signaux permettant d'être encapsidé de manière sélective. Ces signaux ont pu être localisés au niveau d'un domaine appelé  $\psi$  (Aldovini and Young 1990; Clavel and Orenstein 1990; Clever et al. 2002; Hayashi et al. 1992; Lever et al. 1989). Ce domaine d'encapsidation comprend les tiges boucles SL1 à SL4 (Figure 7A) et par conséquent comprend le DIS d'où une possible corrélation entre les phénomènes d'encapsidation et de dimérisation. En effet, une délétion individuelle de ces tiges perturbe le processus de dimérisation (Clever and Parslow 1997; Harrison et al. 1998; McBride and Panganiban 1996; McBride et al. 1997; Paillart et al. 1996b). Par contre, seule la structure de la tige boucle SL2 semble être importante d'un point de vue fonctionnel. La tige boucle SL3 représente quant à elle le site majeur de fixation de la NCp7 dans le domaine  $\psi$  (Amarasinghe et al. 2000a; Amarasinghe et al. 2000b).

La NC est très importante pour le processus d'encapsidation au même titre que pour la dimérisation puisque l'altération des doigts de zinc de la NC provoque une forte diminution de l'encapsidation de l'ARN génomique (Aldovini and Young 1990; Dorfman et al. 1993; Liang et al. 1998; Schwartz et al. 1997). La taille de la région basique séparant les deux doigts de zinc est également importante pour la fixation de la protéine sur l'ARN génomique (Poon et al. 1998; Poon et al. 1996; Schmalzbauer et al. 1996; Zhang and Barklis 1997).

## I-6-2 Dimérisation et Recombinaison génétique

La dimérisation facilite la variabilité génétique grâce à des phénomènes de recombinaison (Balakrishnan et al. 2001; Hu and Temin 1990). La délétion du DIS perturbe la synthèse de l'ADN proviral dans le virion, en particulier lors du second saut de brin (Paillart et al. 1996a) et affecte l'efficacité de la rétro-transcription entre la synthèse du brin (-) et de l'ADN proviral (Shen et al. 2000). Nous savons que la transcription inverse de l'ARN génomique en ADN proviral est un processus complexe du cycle avec l'existence de sauts de brins de 5' en 3'. Cette phase est réalisée sur l'un ou l'autre brin d'ARN associé en dimère. Ces sauts de brin intermoléculaire apparaissent comme une des origines des recombinaisons génétiques existant

chez les rétrovirus (Anderson et al. 1998; Jones et al. 1994). De plus, les sites impliqués dans le processus de dimérisation correspondent à des domaines préférentiels de recombinaison (Lund et al. 1999; Mikkelsen et al. 1996). Il est donc fortement probable que la dimérisation influe sur les phénomènes de recombinaison génétique mais en constitue seulement une des causes. La transcriptase inverse demeure la cause principale de la variabilité génétique des rétrovirus.

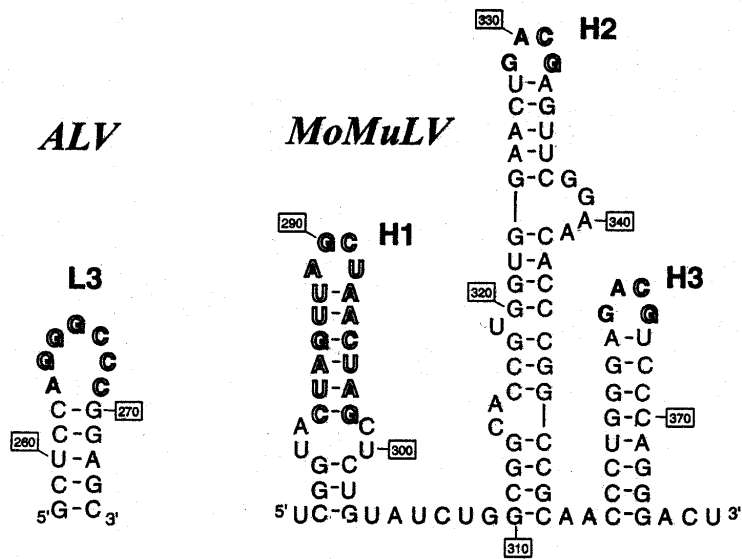
### I-6-3 Dimérisation et Traduction

La dimérisation de l'ARN génomique réduit l'efficacité de la traduction *in vitro* du précurseur gag pr55 (Baudin et al. 1993; Bieth et al. 1990; Mougel et al. 1993). La zone de dimérisation (DLS), qui recouvre une partie du site d'initiation de gag, perturberait la production des ARNm servant de messenger à la synthèse de la protéine. La traduction serait inhibée par la structure secondaire de la séquence en 5' terminal de la zone d'encapsidation  $\psi$  (Miele et al. 1996) qui masquerait le site d'initiation de la traduction : le codon AUG.

### I-6-4 Dimérisation chez d'autres rétrovirus

De même que pour VIH-1, les séquences impliquées dans la dimérisation des rétrovirus aviaires et murins ont également pu être identifiées et localisées dans la même région 5' terminale du génome. La dimérisation est également activée par la présence de NCp (Bieth et al. 1990; Bonnet-Mathoniere et al. 1996; Prats et al. 1990). Dans tous les cas étudiés, une tige boucle avec une séquence auto-complémentaire serait impliquée dans l'initiation de la dimérisation de façon analogue à ce qui est observé chez VIH-1.

Dans le cas du virus de la leucose aviaire (ALV) et du virus du sarcome de Rous (RSV), une tige boucle de 7 bases (Figure 13) avec une séquence auto complémentaire de 6 résidus a pu être identifiée comme responsable de l'initiation de la dimérisation (Banks and Linial 2000; Doria-Rose and Vogt 1998; Fosse et al. 1996; Polge et al. 2000). Cette séquence a pu être localisée comme pour le cas de VIH-1 en amont du site d'épissage. Deux formes de dimères ont pu être générées correspondant probablement à une forme en 'kissing complex' et une forme en dimère étendu (Polge et al. 2000). Il est également probable que la protéine de nucléocapside NCp12 intervienne dans l'isomérisation du génome diploïde de manière analogue à ce qui est observé chez VIH-1 (Meric et al. 1988; Meric and Spahr 1986).



**Figure 13.** Séquences responsables de l'initiation de la dimérisation chez ALV ( à gauche) et MoMuLV ( à droite)

Dans le cas de la leucémie Murine de Moloney ( MoMuV), le domaine  $\psi$  d'encapsulation est composé de 3 tiges boucles (De Tapia et al. 1998; Girard et al. 1995; Girard et al. 1996; Paoletti et al. 1993; Prats et al. 1990; Roy et al. 1990). Le site d'initiation de la dimérisation n'est pas aussi bien définie que dans le cas de VIH-1. En effet, le site semble être réparti sur les trois tiges boucles H1, H2 et H3 (Figure 13). Toutefois, la délétion de H1 inhibe le processus de dimérisation (Girard et al. 1996; Prats et al. 1990). Comme dans le cas de VIH et de ALV, la NCp10 intervient dans le processus de dimérisation et de maturation du virion.

# DETERMINATION STRUCTURALE DES ACIDES NUCLEIQUES PAR RMN

Il existe deux techniques principales pour la détermination structurale des macromolécules biologiques : la radiocristallographie de rayons X (RX) et la Résonance Magnétique Nucléaire (RMN).

La radiocristallographie de rayons X est la principale technique d'obtention de structures tridimensionnelles. Elle consiste à étudier une molécule sous sa forme cristalline et d'obtenir à partir des diagrammes de diffractions la position des différents atomes lourds qui la constituent. Dans le cas d'études de macromolécules biologiques, elle a permis de résoudre la structure de l'enzyme à tête de marteau (Pley et al. 1994; Scott et al. 1995) ou encore celle du ribosome (Cate et al. 1996). Néanmoins, cette technique présente un certain nombre de limites. En effet, un grand nombre de séquences d'ARN qui se présente sous forme de tige boucle unimoléculaire a tendance à ne pas cristalliser rendant leur étude par diffraction des rayons impossible (Cheong et al. 1990; Varani et al. 1991). De plus, certaines régions possèdent une forte flexibilité conformationnelle rendant ainsi leur étude par cristallographie très difficile.

La RMN apparaît comme une technique particulièrement adaptée à l'étude de telles régions. La RMN s'applique aux molécules en solution, dans des conditions plus proches des milieux biologiques par rapport à la radiocristallographie de rayons X. Elle permet également d'étudier la dynamique interne des molécules et constitue un excellent complément aux études cristallographiques.

## II-1 Stratégies d'attributions des acides nucléiques

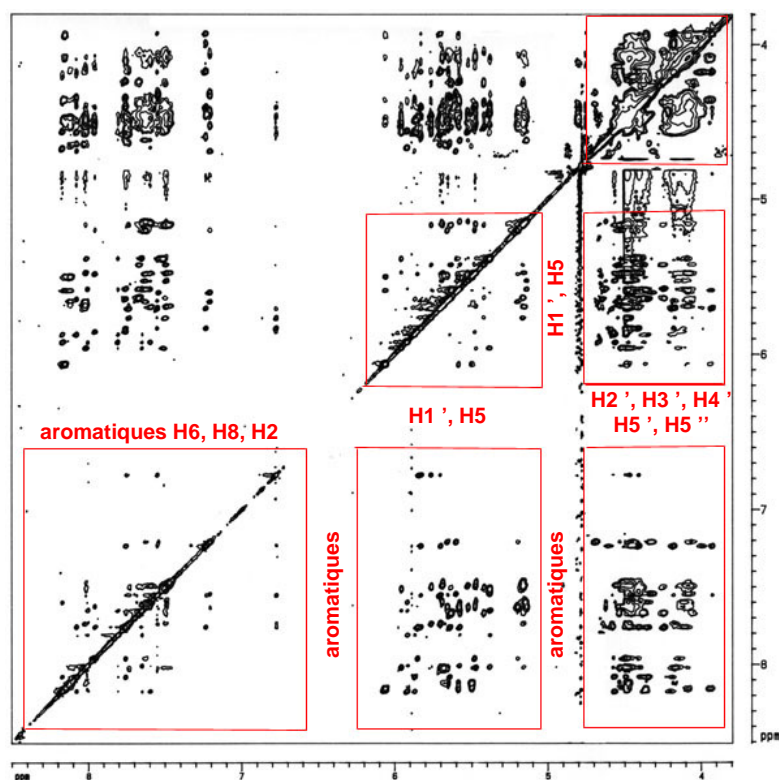
De nombreuses revues décrivent la démarche générale à employer pour l'établissement d'une structure tridimensionnelle d'acides nucléiques (Borer et al. 1997; Cromsig et al. 2001; Puglisi and Puglisi 1998; Wijmenga and van Buuren 1998). Dans la suite de ce chapitre, nous aborderons les difficultés rencontrées lors d'étude d'ARN ainsi que le principe de l'attribution séquentielle et les différents types d'informations qui peuvent être obtenues par RMN du proton avec des séquences non marquées par des isotopes stables ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^2\text{D}$ ).

### Complexité de l'étude des ARN par RMN

La résonance magnétique nucléaire est une technique relativement peu sensible nécessitant, pour l'étude structurale, une grande quantité de produit. Si la production d'ADN par synthèse chimique sur phase solide à partir de phosphoramidites est devenue courante, il en est tout autrement pour les ARN. En effet, la synthèse par voie chimique est excessivement onéreuse surtout pour des molécules marquées  $^{13}\text{C}$  et  $^{15}\text{N}$ . Dès lors, la production d'ARN par voie biochimique a été adoptée pour répondre aux exigences de la RMN. Cette méthode utilise la T7 polymérase pour transcrire une séquence d'ARN à partir d'une séquence d'ADN. Cette synthèse par voie biochimique est relativement sensible à la conformation de la séquence mais présente l'énorme avantage, pour l'étude par RMN, de permettre un marquage isotopique ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^2\text{D}$ ).

Une autre difficulté de l'étude par RMN des ARN par rapport aux ADN, réside dans le recouvrement important des raies de résonance des protons de la macromolécule. En effet, chaque proton du cycle ribofurannique de l'ARN est lié à un oxygène et, de ce fait, les déplacements chimiques des protons en 2',3',4',5' et 5'' sont groupés sur 1 ppm (Figure 14) alors qu'ils s'étendent sur 3 à 4 ppm dans le cas des ADN.





**Figure 14.** Spectre des protons non-échangeables du dimère stable de SL1-23mer VIH-1<sub>Lai</sub>. NOESY  $\tau_m=300\text{ms}$ , D2O, 25°C (Girard et al. 1999).

## II-2 Informations structurales révélées par RMN

### II-2-1 Mesure des distances

L'effet Overhauser nucléaire (nOe) représente la principale source d'information pour la détermination structurale en RMN. Ce phénomène repose sur le transfert d'aimantation à travers l'espace entre 2 protons. La vitesse de transfert peut être directement liée à la distance entre ces protons.

L'expression du nOe est donnée par la formule suivante (Abragam 1961) :

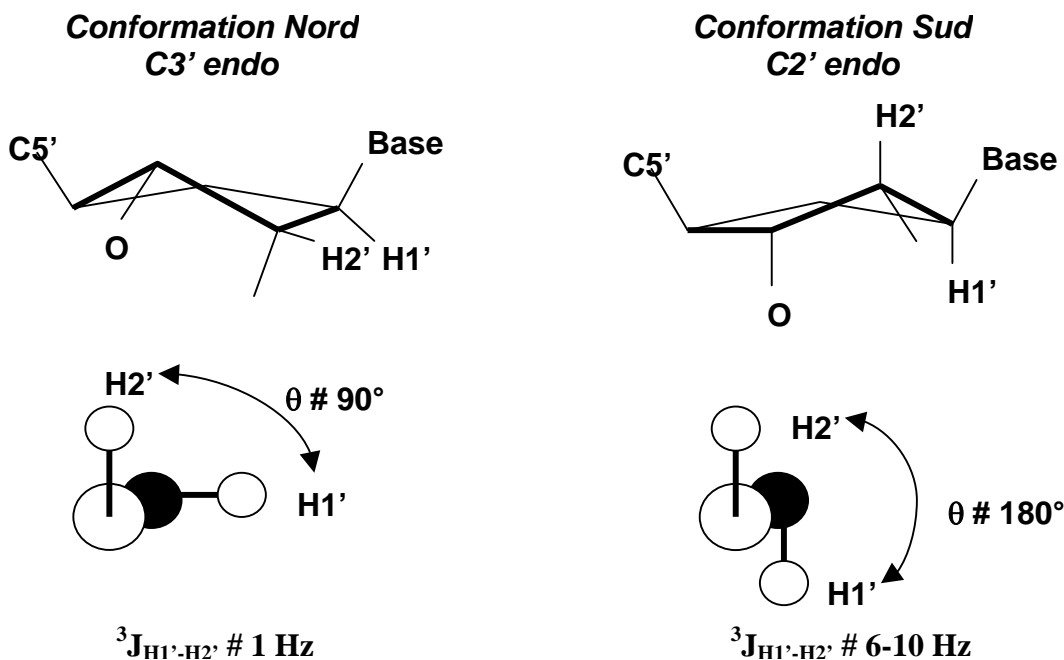
$$\eta_{ij} = K/r_{ij}^6$$

K est une constante qui dépend du temps de corrélation  $\tau_c$  de la molécule et  $r_{ij}$  représente la distance entre les deux spins. Seules les informations de courtes distances, c'est à dire inférieures à 5 Å, sont recueillies. Les effets nOe permettent d'avoir des données locales sur

la molécule. C'est la multiplication de ces données qui permettra d'obtenir la conformation globale de la molécule. Il est donc très important de recueillir un maximum de données pour affiner la structure de la molécule. Il est également possible de recueillir des informations sur la conformation globale de la molécule à condition de se placer dans des milieux orientés (Prestegard et al. 2000).

## II-2-2 Mesure des angles dièdres

La RMN permet également d'avoir accès aux couplages scalaires *via* les liaisons chimiques. En effet, les constantes de couplage dépendent de la valeur de l'angle de torsion qui décrit les rotations de la liaison joignant les atomes auxquels sont couplés les deux spins (Karplus 1959). Dans le cas des acides nucléiques, les angles dièdres vont nous permettre d'obtenir des informations sur la conformation du sucre de chaque nucléotide. En effet, la mesure de la constante de couplage  $^3J$  entre les protons H1' et H2' est directement reliée à la conformation du sucre (Figure 15).



**Figure 15.** Variation de la constante de couplage  $^3J$  H1'-H2' avec la conformation du sucre (Puglisi and Puglisi 1998).

Il est également possible d'obtenir grâce aux constantes de couplage scalaire  $^3J(^1\text{H}-^{31}\text{P})$  des informations sur la valeur des angles de torsion du squelette phosphodiester.

Ces informations sont également très utiles pour la détermination structurale puisque le squelette phosphodiester représente l'architecture même de l'acide nucléique. Malheureusement, étant donné le faible étalement des raies de résonance du phosphore, il n'est pas toujours possible de pouvoir identifier chaque raie de résonance.

### II-2-3 Le déplacement chimique du proton

Le déplacement chimique des raies de résonance des protons d'un acide nucléique s'étend sur une plage allant de 0 à 15 ppm. L'identification des différentes zones de résonance permet d'identifier les différents types de protons présents dans l'acide nucléique (Figure 14).

Comme nous l'avons vu précédemment, les ARN présentent un grand recouvrement de ces protons du sucre ce qui rend difficile l'obtention d'informations structurales pouvant être retirées à partir des spectres RMN.

### II-2-4 Le déplacement chimique du phosphore

La gamme de déplacement chimique du phosphore est relativement limitée (environ 1 ppm) ce qui rend son attribution assez complexe. Une déviation par rapport à la forme standard se traduit par des résonances 'anormales' vers les hauts ou les bas champs selon la déformation considérée. Ces déplacements peuvent être reliés aux angles du squelette phosphodiester (Delepierre et al. 1988; Giessner-Prettre and Pullman 1987; Gorenstein et al. 1984).

### II-2-5 Les liaisons hydrogène

Les protons échangeables, portés par les groupements amino et imino de l'acide nucléique, s'échangent beaucoup plus rapidement que ceux présents dans une protéine. Cependant, lorsque ces protons sont engagés dans des appariements de type liaison hydrogène, leur vitesse d'échange est ralentie et peut être observée par RMN. Les liaisons hydrogène sont visibles à condition d'utiliser des séquences qui permettent de réduire le temps d'échange des protons échangeables avec l'eau. Les séquences 'WATERGATE' (Piotto et al. 1992) ( $\tau_{\text{échange}}$

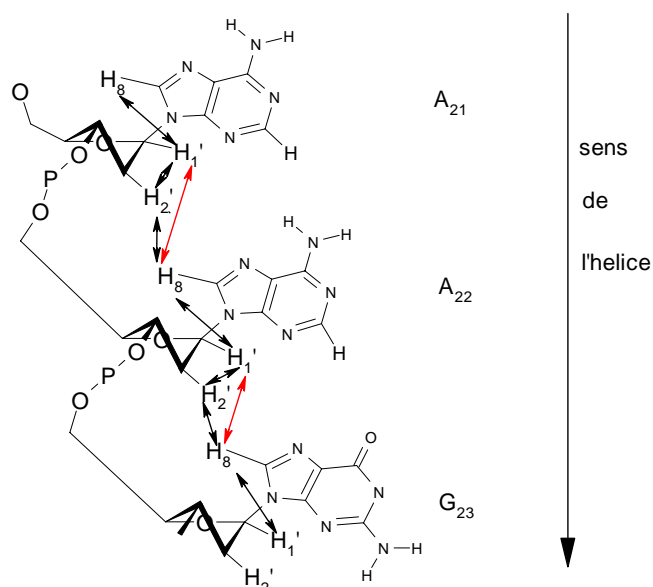
< 1ms) ou 'jump and return' (Plateau and Gueron 1982) ( $\tau_{\text{échange}} < 0,1\text{ms}$ ) répondent à cette exigence.

Les informations de type liaison hydrogène sont très importantes pour connaître la structure et le repliement secondaire de l'acide nucléique.

## II-2-6 Méthodologie d'attribution séquentielle

La première étape repose sur l'observation des résonances des protons imino et amino dans l'eau légère. Cette étape permet de mettre en évidence les différents appariements entre les bases de l'acide nucléique et de connaître par conséquent sa structure et son repliement secondaire. Sur un spectre NOESY dans  $\text{H}_2\text{O}$ , nous pouvons également réaliser l'attribution partielle de certains protons H2 et H5 en repérant leur connectivité avec les protons imino-amino.

La seconde étape consiste à étudier l'oligonucléotide dans  $\text{D}_2\text{O}$  pour identifier les résonances des protons non échangeables (H2, H6, H8, H1', H2', H3', H4', H5, H5' et H5''). Dans un premier temps, nous réalisons des expériences de transfert de cohérence de type TOCSY pour permettre de repérer les connectivités H5-H6 des cytosines ou des uraciles. Les expériences COSY permet de mettre en évidence les H1' et H2' des sucres et ainsi de pouvoir déterminer leur conformation via les constantes de couplage. Dans un deuxième temps, nous réalisons des expériences de type NOESY qui conduisent à l'attribution séquentielle des protons aromatiques, H1' et H2' de chaque résidu via les connectivités NOE (Figure 16).



**Figure 16.** Exemple d'attribution séquentielles des protons aromatiques, H1' et H2' de chaque résidu via les connectivités NOE.

## II-3 Les Nouvelles Approches

L'acquisition, au cours de ces quinze dernières années, de nouveaux spectromètres à haut champs (900 MHz), équipés de nouvelles sondes (cryosondes) permet de gagner en sensibilité et par conséquent de travailler avec des produits beaucoup moins concentrés. La résonance magnétique nucléaire occupe ainsi une place de choix parmi les techniques d'investigation structurale.

Néanmoins, ses applications pour la détermination structurale d'acides nucléiques ou de protéines de grandes tailles sont extrêmement limitées par deux facteurs :

- La complexité des spectres RMN à cause du fort recouvrement des raies de résonance
- La relaxation transversale (T2) rapide des spins
- L'échange chimique qui est favorisé par les hauts champs magnétiques.

### II-3-1 Marquage Isotopique

La principale technique pour permettre la simplification des spectres RMN, due au recouvrement des raies de résonance, est de marquer les macromolécules par des isotopes stables ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^2\text{D}$ ). L'enrichissement  $^{13}\text{C}$ ,  $^{15}\text{N}$  permet d'éliminer les ambiguïtés d'attribution, d'avoir des informations sur l'ensemble des angles dièdres de la molécule. Néanmoins, au

delà d'une certaine taille l'élargissement des raies de résonance dû à la relaxation transversale est telle que la détermination structurale est impossible. L'introduction du marquage sélectif au  $^2\text{D}$  des acides nucléiques (Cromsig et al. 2002) présente deux avantages celui de supprimer un certain nombre de raies de résonance (Glemarec et al. 1996; Yamakage et al. 1993) mais également de réduire les phénomènes de relaxation dipolaire et par conséquent la largeur de raies (Nikonowicz et al. 1997).

### II-3-2 TROSY

Les expériences TROSY (Transverse Relaxation Optimized Spectroscopy) ont pour but de réduire la relaxation transversale des spins d'une molécule (Pervushin et al. 1997). Cette technique est basée sur le fait que les corrélations croisées entre les interactions dipôle-dipôle et l'anisotropie de déplacement chimique engendrent différentes intensités et largeurs pour chaque raie de résonance du multiplet dans le cas de deux spins  $\frac{1}{2}$  comme par exemple  $^{15}\text{N}$ - $^1\text{H}$  ou  $^{13}\text{C}$ - $^1\text{H}$  (Pervushin et al. 1998; Pervushin et al. 1997; Salzman et al. 1998). L'expérience TROSY permet de sélectionner la raie la plus fine du multiplet ce qui permet de gagner en sensibilité et en résolution lors des expériences RMN.

### II-3-3 Les couplages dipolaires résiduels

Les distances mesurables à l'aide du nOe sont inférieures à 5 Å ce qui ne permet d'obtenir avec précision l'orientation respective de fragments éloignés dans l'espace. La mesure du couplage dipolaire résiduel permet de surmonter cette difficulté (Prestegard et al. 2000). Ce couplage est mesurable et dépend de l'orientation des vecteurs étudiés. En RMN du solide, ils se manifestent tous et leur valeur (plusieurs milliers de Hz) associée à leur nombre (des milliers) élargissent considérablement le spectre et détruisent son caractère haute résolution. En solution, la dispersion de l'orientation des molécules est moyennée à zéro. Par contre, si la molécule est orientée partiellement (environ une molécule sur 1000) à l'aide de bicelles (Tjandra et al. 1997) ou de phages (Clare et al. 1998), le couplage dipolaire "dit" résiduel apparaît et leurs valeurs (quelques Hz) ne compliquent pas le spectre. Cette valeur est une fonction de la distance ( $1/r^3$ ) et de l'orientation du vecteur considéré. Ces informations sont dites 'longues' distances (jusqu'à 10 Å). Ces distances ou ces angles fournissent des informations précieuses sur l'orientation relative du vecteur  $^{13}\text{C}$ -H ou  $^{15}\text{N}$ -H.

# MODELISATION MOLECULAIRE D'ACIDES NUCLEIQUES A PARTIR DE DONNEES RMN

Le but de la modélisation est d'obtenir un ensemble de structures, ou de jeux de coordonnées, répondant aux données RMN et correspondant à un minimum énergétique. Lors de ce travail de thèse, nous avons principalement utilisé les logiciels XPLOR (Brünger 1992) et CNS (Brünger et al. 1998). Ces logiciels utilisent les paramètres d'énergie potentielle de CHARMM (Brooks et al. 1983) améliorés pour les acides nucléiques.

Dans un premier temps, nous décrirons succinctement la fonction d'énergie potentielle utilisée par ces deux logiciels puis les différentes méthodologies de modélisation mises en jeu pour la détermination structure RMN.

## III-1 La fonction d'énergie potentielle des logiciels XPLOR et CNS

La résolution d'une structure RMN peut être formulée soit comme un problème de géométrie de distance soit comme un problème d'optimisation de l'énergie. Cette dernière approche utilise une fonction d'énergie qui peut se décomposer en deux principaux éléments, eux même décomposables. La fonction d'énergie totale peut s'écrire (1) :

$$E_{\text{totale}} = E_{\text{chimique}} + E_{\text{effective}} \quad (1)$$

Le terme d'énergie chimique,  $E_{\text{chimique}}$ , écrit les interactions covalentes dans la molécule ainsi que les interactions entre atomes non liés. Le terme d'énergie effective,  $E_{\text{effective}}$ , est obtenu à partir des données RMN, on parlera alors de contraintes énergétiques. Nous allons en premier lieu décrire les paramètres énergétiques.

### III-1-1 Les différents termes de l'énergie chimique

Le terme d'énergie chimique peut se décomposer en plusieurs autres termes (2) :

$$E_{\text{chimique}} = E_{\text{termes liés}} + E_{\text{termes non liés}} \quad (2)$$

Le terme  $E_{\text{termes liés}}$  représente la variation d'énergie due aux déformations de la molécule par rapport à la géométrie standard. Il comprend les énergies correspondant aux déformations de liaisons, des angles de valence, en dehors du plan et les termes de torsions.

Les termes non liés décrivent les contributions énergétiques dues aux interactions entre atomes qui ne sont pas liés de manière covalente. Les interactions non liées représentent les interactions de Van Der Waals, les interactions électrostatiques et les interactions de liaisons hydrogènes.

### III-1-2 Les termes effectifs

Grâce aux différentes expériences, nous avons en notre possession un ensemble de données expérimentales. Ces données vont apparaître dans la fonction d'énergie potentielle du logiciel sous la forme d'énergies de contraintes : les termes effectifs.

Le terme d'énergie effective peut se décomposer en plusieurs paramètres correspondant à des résultats expérimentaux particuliers.

$$E_{\text{effectif}} = W_{\text{nOe}} E_{\text{nOe}} + W_{\text{CDIH}} E_{\text{CDIH}} + W_{\text{plan}} E_{\text{plan}}$$

Le premier terme de cet équation correspond à l'énergie due aux contraintes nOe. Ces contraintes sont décrites, le plus couramment, par une fonction parabolique à fond plat avec une exponentielle douce.

Le deuxième terme de l'équation est l'énergie due aux contraintes d'angles dièdres déterminées par RMN,  $E_{\text{CDIH}}$ . Ces contraintes, dans le cadre d'une étude d'acides



nucléiques, permet d'imposer la conformation des sucres et du squelette du phosphodiester de la molécule. L'expression de son potentiel est proche de celui de l'effet nOe.

Contrairement au calcul de l'énergie nOe le calcul d'énergie dièdre, appelé  $E_{CDIH}$ , n'a qu'une seule option de calcul qui correspond à un potentiel de forme carrée.

Le troisième terme de l'équation représente l'énergie de plan. Cette énergie  $E_{plan}$  sert à favoriser les appariements des bases dans les acides nucléiques. Elle correspond à une énergie d'angle impropre et pénalise les atomes qui ne respectent les plans définis dans le fichier de contraintes.

## III-2 Algorithmes de modélisation

La RMN permet d'obtenir un certain nombre de caractéristiques structurales d'une molécule mais le nombre de ces informations est insuffisant pour trouver directement sa conformation. Les contraintes d'angles dièdres sont locales, tandis que les contraintes nOe sont limitées dans l'espace (inférieures à 0.5 nm). Le nombre de degré de liberté dans une macromolécule est très important générant ainsi un grand nombre de minima locaux qu'il faut explorer. Plusieurs protocoles ont ainsi pu être développés pour permettre cette exploration de l'espace conformationnelle.

### III-2-1 Mécanique Moléculaire

Les techniques de mécanique moléculaire sont généralement utilisées de manière itérative pour raffiner la conformation d'une molécule de manière à minimiser son énergie. Elle peut s'effectuer à partir d'un jeu de coordonnées internes comme pour le programme JUMNA (Lavery et al. 1995) ou cartésiennes pour le logiciel XPLOR. Cette minimisation est particulièrement utile pour effacer les contacts forts entre atomes. Il existe plusieurs méthodes de mécanique moléculaire, chacune avec sa force et ses faiblesses. Le principe général est pourtant le même. Une nouvelle conformation est générée à partir d'une conformation courante en utilisant la loi suivante :

$$X(\text{nouveau}) = X(\text{ancien}) + \lambda S$$

Où  $S$  est la direction de recherche et  $\lambda$  le pas de l'itération. Le pas contrôle l'éloignement dans la direction de recherche (c'est à dire le changement de la structure en une étape).

Si l'énergie de la nouvelle conformation est plus grande que celle de l'ancienne, on retient l'ancienne conformation et on utilise un pas plus petit. Par contre, si c'est le contraire qui se produit, on calcule alors une nouvelle direction de recherche en utilisant la nouvelle conformation et on applique un nouveau déplacement. Ce processus est répété jusqu'à ce que l'énergie ne varie pas de manière significative entre les itérations.

On utilise deux méthodes principales de mécanique moléculaire : la méthode dite de plus forte pente (Morse and Feshbach 1953) et celle du gradient conjugué (Powell 1977). La principale différence entre les techniques de mécanique réside dans le choix de la direction de recherche.

#### III-2-1-a. La méthode de plus forte pente

La force qui agit sur chaque atome est calculée à partir de la fonction d'énergie  $E$  en utilisant la relation :  $F = - dE / Dr$

La direction de recherche est donnée :  $S = g$  où  $g$  est le vecteur unitaire pointant dans la direction de la force.

$$g_i = \partial V / \partial x_i$$

$g_i$  étant le gradient de l'énergie potentielle  $V$  par rapport à la coordonnée cartésienne  $x_i$ .

La méthode de plus forte pente est une méthode rapide et robuste, très efficace dans les premières étapes où les énergies sont fortes. Cependant, les propriétés de convergence sont faibles rendant cette méthode très peu efficace dans les dernières étapes de la minimisation.

#### III-2-1-b. La méthode du gradient conjugué

Dans cette méthode, la direction de recherche est une combinaison de celle donnée par la force et de la direction de recherche précédente :

$$S = g + \gamma S (\text{ancien})$$

Cette technique est un peu plus lente que la précédente mais présente l'énorme avantage d'être beaucoup plus efficace dans les dernières étapes de minimisation.

Toutes les méthodes de mécanique moléculaire tendent à trouver le minimum local le plus proche du point de départ. Si un puit plus profond correspondant à une meilleure énergie minimum existe mais est séparé du point de départ par un maximum d'énergie, il ne sera pas trouvé. Les techniques de mécanique moléculaire sont incapables de faire de la recherche conformationnelle c'est à dire de s'affranchir de barrière énergétique pour trouver le minimum global de la molécule. Pour cela, nous devons utiliser un autre algorithme, la dynamique moléculaire.

### III-2-2 Dynamique Moléculaire

L'image vraie d'un système moléculaire est loin de l'image statique idéale donnée par la mécanique moléculaire. Les atomes sont en mouvement perpétuel. Le but de la dynamique moléculaire est de simuler le comportement dynamique d'un système. Cette technique est basée sur la fonction d'énergie potentielle  $V$ , à partir de celle-ci, on peut trouver les composantes des forces  $F_i$  de la force  $F$  agissant sur un atome :

$$F_i = - \partial V(r_1, r_2, \dots, r_n) / \partial r_i$$

La force résulte de l'accélération  $a$  selon l'équation du mouvement de Newton :

$$F_i = m_i a_i = m_i (d^2 r_i(t) / dt^2)$$

$r_i$  représente le vecteur position et  $m_i$  la masse.

L'intégration de ces équations permet, si l'on connaît l'accélération, de déterminer la vitesse d'un atome à l'étape suivante. Cette intégration se fait numériquement sur des petits intervalles de temps  $\Delta t$  de l'ordre de la femto seconde, pendant lequel on peut considérer que la force est constante.

Le moteur des programmes de dynamique moléculaire est l'algorithme d'intégration du temps, qui nécessite d'intégrer l'équation du mouvement de tous les atomes qui interagissent entre eux pour suivre la trajectoire.

Il existe plusieurs types d'algorithmes d'intégration ; mais dans le cadre de nos travaux, nous avons seulement utilisé l'algorithme de Verlet (Verlet 1967).

Cet algorithme est basé sur un développement de Taylor à l'ordre 3 des positions  $r(t)$ .

Si on appelle  $v$  les vitesses,  $a$  les accélérations et  $b$  les dérivés d'ordre 3 de  $r$  par rapport à  $t$ , on obtient :

$$r(t+\Delta t) = r(t) + v(t) \Delta t + (1/2)a(t) \Delta t^2 + (1/6) b(t) \Delta t^3 + O(\Delta t^4)$$

$$r(t-\Delta t) = r(t) - v(t) \Delta t + (1/2)a(t) \Delta t^2 - (1/6) b(t) \Delta t^3 + O(\Delta t^4)$$

La somme de ces deux expressions donne :

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + a(t) \Delta t^2 + O(\Delta t^4)$$

$$a(t) = -(1/m) \nabla V(r(t))$$

On peut donc calculer les vitesses à partir des positions en utilisant la relation suivante :

$$V(t) = (r(t+\Delta t) - r(t-\Delta t)) / 2\Delta t$$

L'algorithme de Verlet présente l'avantage de ne pas nécessiter les vitesses lors de l'intégration des équations de mouvement ; seule l'information de position est prise en compte. De plus, nous avons qu'une seule évaluation de la force par cycle d'intégration d'où un gain de temps dans la simulation.

Pour le contrôle de température, nécessaire dans le recuit simulé, on ajoute un autre terme :

$$F_i = m_i (d^2 r_i(t) / dt^2) = - \partial V(r_1, r_2, \dots, r_n) / \partial r_i + \beta_i ((T_0/T) - 1) v_i$$

$T_0$  est la température du bain dans lequel le système est couplé,  $\beta_i$  la constante de force,  $T$  est la température du système et  $v_i$  les vitesses de chaque atome.

### III-2-2-a. Méthode du recuit simulé

Le recuit simulé est une méthode de dynamique moléculaire permettant l'exploration de l'espace conformationnel (Nilges et al. 1988). Le système est chauffé artificiellement à haute

température grâce à l'introduction d'un bain thermique. Le terme d'énergie cinétique élevé permet alors de franchir des barrières énergétiques importantes contrairement à la mécanique moléculaire. Lorsque l'équilibre thermique est atteint, on refroidit progressivement le système. On peut ainsi se dégager des minima locaux proches de la structure initiale.

### III-2-2-b. Recuit dans l'espace cartésien

La méthode du recuit simulé utilise une force unique de répulsion  $F_{\text{repel}}$  qui remplace les forces non liantes (électrostatiques, Van der Waals et liaisons hydrogène) :

$$F_{\text{repel}} = K_{\text{VDW}} (R_{\text{min}}^2 - r^2)^2 \quad \text{si } r > r_{\text{min}}$$

$$F_{\text{repel}} = 0 \quad \text{si } r < r_{\text{min}}$$

$K_{\text{VDW}}$  est une constante de force variable au cours du cycle,  $r_{\text{min}}$  simule le rayon de Van Der Waals des atomes. Pendant la première phase, la constante de contraintes  $nOe$  est augmenté progressivement, tandis que la constante de répulsion  $K_{\text{VDW}}$  est maintenue très bas pour permettre les réarrangements structuraux. La constante de répulsion n'est augmentée à son tour jusqu'à sa valeur finale que lorsque la constante de contraintes  $nOe$  a atteint son maximum.

Pour permettre l'exploration de l'espace conformationnel, plusieurs techniques peuvent être mises en jeu :

- l'utilisation d'un grand jeu de coordonnées aléatoires de départ (Nilges et al. 1988)
- l'utilisation de vitesses initiales aléatoires (Mujeeb et al. 1993; SantaLucia and Turner 1993; Zhang and Patel 1991)

### III-2-2-c. Recuit dans l'espace des angles dièdres

Il est également possible de faire une simulation de dynamique moléculaire dans l'espace des angles dièdres. Pour cela, on calcule l'énergie cinétique  $E_{\text{cin}}$  et l'énergie potentielle  $E_{\text{pot}}$

pour former le lagrangien du système,  $L = E_{\text{cin}} - E_{\text{pot}}$  et on applique les équations de Lagrange :

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) = \frac{\partial L}{\partial q_i}$$

On obtient ainsi un jeu de N équations du mouvement (N correspond aux nombres d'angles dièdres). Ces équations sont intégrées de façon numérique comme dans le cas d'une dynamique cartésienne. Dans le cas du formalisme de Lagrange, on obtient un système à N équations reliant les accélérations entre elles qu'il faut résoudre à chaque pas, ce qui rend ce type de calcul très lourd. Pour éviter ce problème, on calcule de proche en proche chaque accélération par rapport à l'accélération de l'angle dièdre voisin. Cette méthode de calcul est intégrée dans les logiciels CNS et XPLOR (version 3.851).

Comme dans toute dynamique, le pas d'intégration doit être inférieur au temps correspondant au mouvement le plus rapide. Dans le cas d'une dynamique cartésienne, le mouvement le plus rapide correspond généralement à la vibration d'un atome d'hydrogène le long de la liaison covalente qui le lie à un atome lourd. Alors que dans le cas d'une dynamique dans l'espace des angles dièdres, tous ces mouvements rapides (élongation des liaisons, distorsion des angles dièdres) sont supprimés au profit de mouvements de rotation d'atomes ou de groupes d'atomes autour d'une liaison covalente, ce qui permet d'augmenter les temps de dynamique.

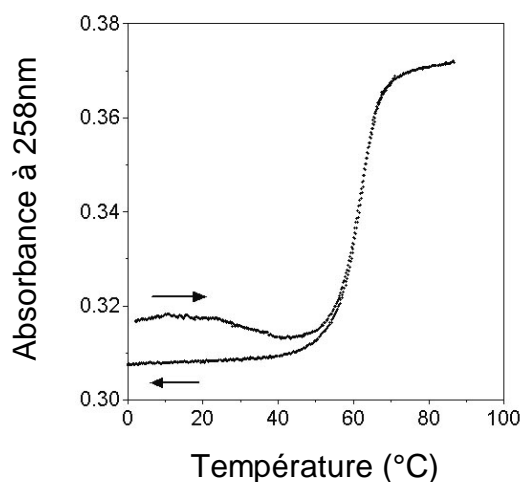
Le recuit simulé dans l'espace des angles dièdres apparaît comme la méthode la plus efficace pour obtenir une structure tridimensionnelle en RMN. L'équipe de Brünger (Stein et al. 1997) a comparé les trois méthodes d'obtention de structures (recuit simulé à partir de géométrie de distance, le recuit simulé dans l'espace cartésien et le recuit simulé dans l'espace des angles dièdres) avec cinq molécules d'intérêt biologique. Que ce soit pour un peptide, une protéine ou un acide nucléique, le recuit simulé dans l'espace des angles dièdres surpasse les deux autres formes de recuit .

# RESULTATS

# PROBLEMATIQUE

Dans l'introduction, nous avons remarqué que la tige boucle SL1 du domaine d'encapsulation était impliquée dans le phénomène d'initiation de la dimérisation de l'ARN génomique (Clever et al. 1996; Laughrea and Jette 1994; Muriaux et al. 1996b; Muriaux et al. 1995). Ces travaux, réalisés sur des fragments d'une centaine de nucléotides, ont montré que la séquence apicale auto-complémentaire de la tige boucle SL1 était nécessaire à la propriété d'initiation de la dimérisation. De plus, ces études ont permis de mettre en évidence deux types de dimère : un dimère instable et un dimère stable.

Plus récemment, les travaux de (Theilleux-Delalande 1998) rapportent que des séquences tronquées de la tige boucle SL1 de 23 nucléotides possédaient les mêmes propriétés que celles décrites sur de plus longs fragments à savoir la même température de fusion ou la présence de 2 types de dimères (Figure 17). Cette séquence de 23 nucléotides est particulièrement appropriée pour une étude structurale en Résonance Magnétique Nucléaire du fait de sa taille réduite.



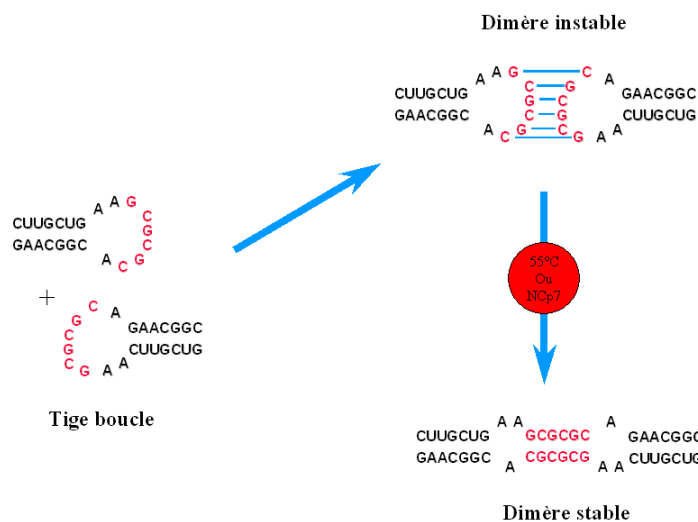
**Figure 17.** Courbe de dénaturation / renaturation du complexe boucle-boucle de SL1-23mer VIH-1<sub>Lai</sub>, d'après (Theilleux-Delalande et al. 2000). La force ionique et le pH sont fixés avec NaCl (100mM) et TrisHCl (50mM, pH 7,5). La variation de température repose sur un cycle de 15°C/heure, avec 5min d'attente après la phase de dénaturation.



A partir de ces travaux, la séquence SL1 de l'ARN génomique peut se présenter sous 3 conformations différentes:

- Une forme monomérique qui est à l'origine du phénomène de dimérisation.
- Une première forme dimérique dite en 'kissing' complexe qui résulte de l'appariement de deux monomères. Cette structure correspondrait à la forme immature dans le virion.
- Une deuxième forme dimérique dite en dimère étendu qui résulte du réarrangement de la forme en 'kissing' complexe et correspondrait à la forme mature du virion. Ce passage d'une forme à l'autre s'effectue grâce à la NCp7 (Feng et al. 1996b; Fu et al. 1994).

Le mécanisme de passage classiquement décrit pour le passage entre les différentes formes est représenté ci-dessous (Figure 18) (Girard et al. 1999; Muriaux et al. 1996a; Rist and Marino 2002):



**Figure 18.** Représentation des différentes structures possibles de la séquence 23mer de la tige boucle SL1.

La dimérisation des ARN génomiques est une étape clé du cycle rétroviral et n'est pas à l'heure actuelle une cible thérapeutique dans le traitement de la maladie. Afin d'élaborer de nouveaux traitements anti-viraux, il est important de mieux comprendre les différents

phénomènes et interactions qui régissent ce phénomène de dimérisation notamment d'un point de vue structural. Au cours de ce travail de thèse, nous avons donc étudié la forme monomérique et la forme dimérique instable, la forme dimérique stable ayant déjà été étudiée dans notre laboratoire (Girard et al. 1999).

# ETUDE DE LA FORME MONOMERIQUE DE LA TIGE BOUCLE SL1

La tige boucle SL1 est la première entité rencontrée lors du phénomène de dimérisation. En effet, c'est l'appariement de deux de ces tiges boucles qui va permettre la formation de la première forme dimérique : le 'kissing' complexe. Son étude structurale est donc essentielle pour permettre la compréhension de l'interaction boucle-boucle. La séquence sauvage (Figure 19B) présente la particularité de dimériser spontanément en solution, nous avons donc choisi de modifier la séquence auto-complémentaire en remplaçant la guanine 12 par une adénine afin de réaliser l'étude de la forme monomérique. Cette mutation présente le double avantage d'empêcher la dimérisation et de conserver la nature purique de la base. Cette nouvelle séquence est appelée SL1\* (Figure 19).



**Figure 19.** A/ Séquence SL123mer\* de VIH-1Lai.(la mutation apparaît en vert) B/ Séquence sauvage de VIH-1Lai.

## I-1 Préparation de l'échantillon

La tige boucle SL1\* a été synthétisée par voie chimique, par l'équipe de T. Huynh Dinh à l'institut Pasteur de Paris, et ensuite purifiée par chromatographie en phase liquide. Nous avons pu ainsi obtenir un échantillon de 500µl avec une concentration de 2mM.

Nous avons passé l'échantillon d'ARN sur une résine échangeuse de cations (Chelex 100) afin d'éliminer les éventuelles traces d'impuretés paramagnétiques susceptibles de détériorer la qualité des spectres.

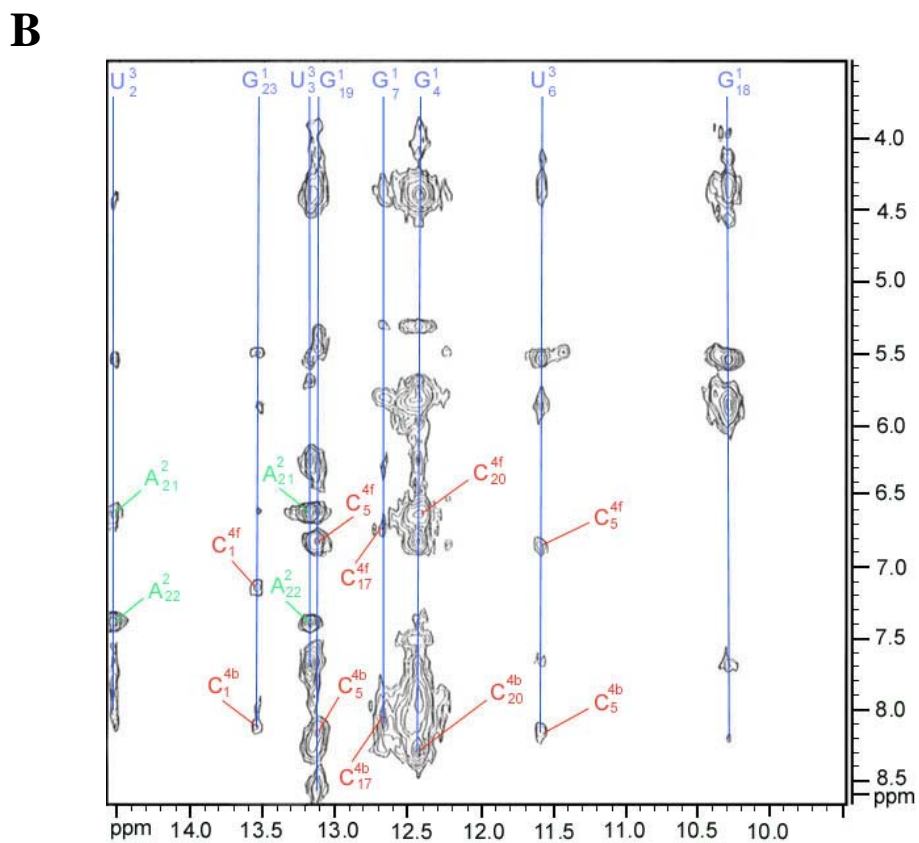
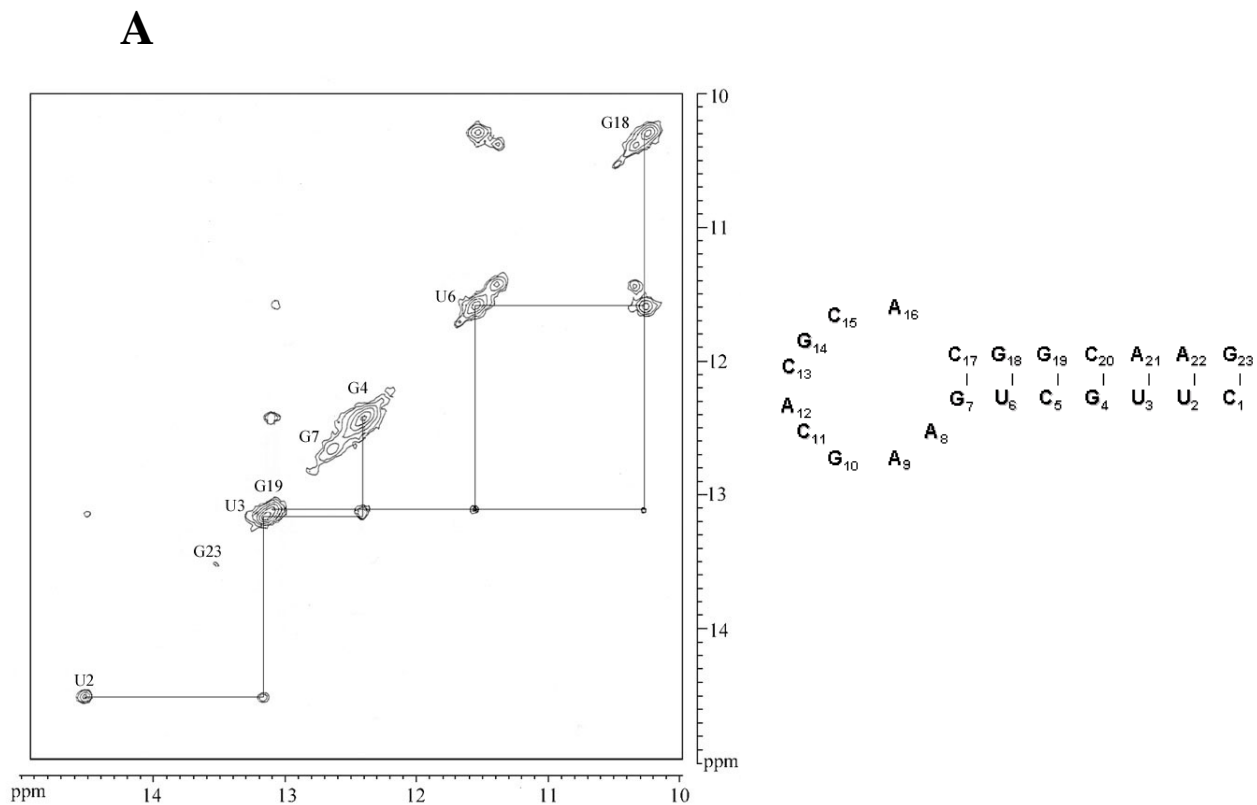
## I-2 Etude RMN

### I-2-1 Etude dans l'eau

L'échantillon est placé dans un mélange H<sub>2</sub>O/D<sub>2</sub>O (90/10) comprenant 10 mM de tampon phosphate pH=5.7 et 100mM NaCl.

Les spectres dans H<sub>2</sub>O ont été obtenus en utilisant soit une séquence « WATERGATE » (Piotto et al. 1992) soit une séquence 'Jump and Return' (Plateau and Gueron 1982) pour la suppression du solvant. Les spectres NOESY ont été enregistrés à différentes températures, entre 5 et 15°C, afin de lever des ambiguïtés dans les zones de recouvrement et à différents temps de mélange (80, 150 et 250ms).

Les paires A-U de la séquence SL1\* ont constitué le point de départ de l'attribution séquentielle. En effet, les connectivités entre NH<sub>3</sub>(U) et H<sub>2</sub>(A) de la même paire sont particulièrement intenses (Figure 20). Ceci est dû à la proximité de ces deux protons (des deux bases en interaction Watson-Crick), mais provient également du fait que les H<sub>2</sub>(A) ont classiquement des vitesses de relaxation très lentes ce qui se traduit par des raies de résonance fines. Les pics à 7.56 et 6.80 ppm ont ainsi été attribués respectivement à H<sub>2</sub>(A<sub>22</sub>) et H<sub>2</sub>(A<sub>21</sub>). Ces résonances - comme justifié plus haut - sont donc fortement corrélées avec NH<sub>3</sub>(U<sub>2</sub>) et NH<sub>3</sub>(U<sub>3</sub>). A partir de là, on peut attribuer séquentiellement (Figure 20) les résonances imino présentes dans la tige puisque ces protons sont proches dans l'espace.



**Figure 20.** Spectre NOESY dans H<sub>2</sub>O/D<sub>2</sub>O (90/10) à 9°C pH=5.7 **A/** Connectivités entre protons imino, temps de mélange de 150ms avec un schéma de la tige boucle SL1\* **B/** Connectivités entre protons imino et protons amino et aromatiques, temps de mélange 120ms.

On peut remarquer que les protons imino impliqués dans la paire U-G, dite ‘wobble pair’, présentent des résonances caractéristiques à 11,59 et 10,30 ppm. Nous avons pu mettre en évidence l’ensemble des connectivités imino et amino qui témoignent de l’appariement des bases de la tige.

## I-2-2 Etude dans l’eau lourde

Les expériences NOESY ont été réalisées à pH=7 à plusieurs températures variant entre 12 et 32°C et à plusieurs temps de mélange de 0 à 450ms. Un délai de 2s a été utilisé pour le retour à l’équilibre. La suppression de l’eau résiduelle a été effectuée en réalisant une pré-saturation basse fréquence.

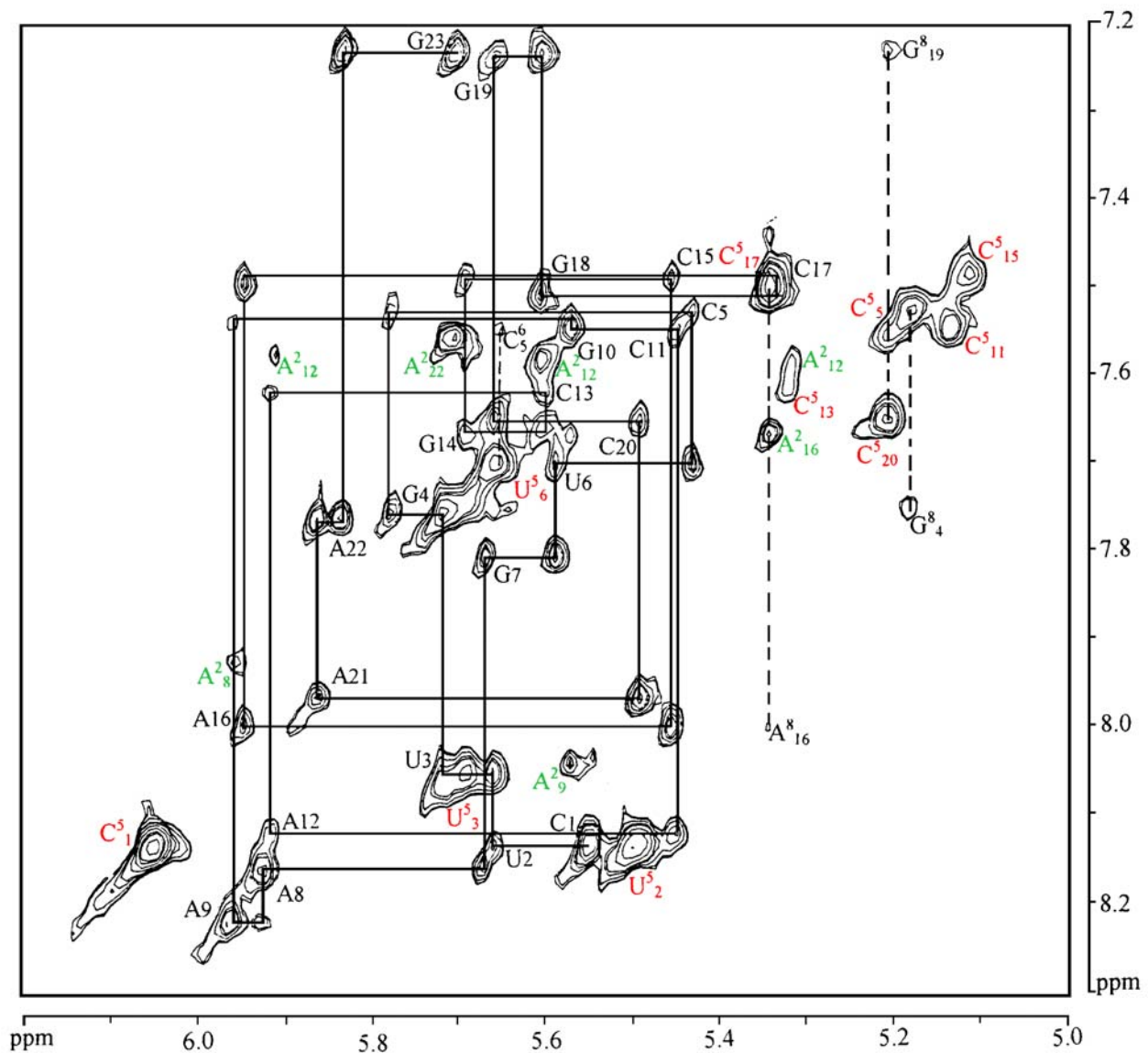
### I-2-2-a. Mise en évidence des H2 des adénines

Les protons H2 des adénines présentent la particularité de posséder un temps de relaxation longitudinale beaucoup plus long que les autres protons de l’oligonucléotide. Ce temps résulte de l’isolement de ce type de protons vis à vis des autres protons. Pour localiser les protons H2, nous avons réalisé des expériences de relaxation T1 utilisant un temps de relaxation de 3.8s et un temps de retour à l’équilibre de 20s. Nous avons ainsi pu localiser les H2 de l’oligonucléotide (Tableau 23).

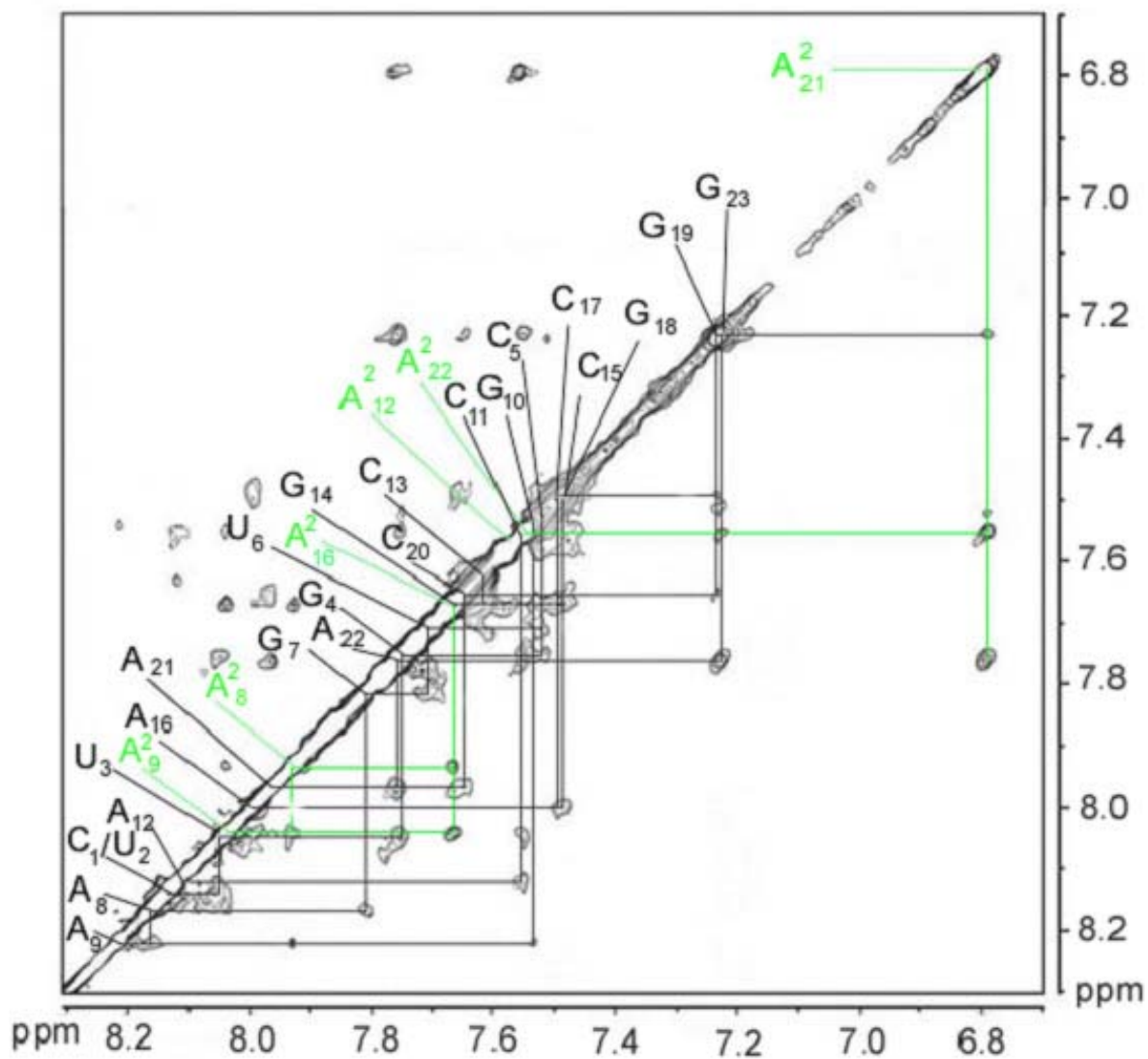
### I-2-2-b. Attribution séquentielle ribose-base

Les spectres NOESY nous ont permis de faire l’attribution séquentielle des protons aromatiques, H1’ et H2’ de l’oligonucléotide suivant la méthodologie décrite précédemment (Puglisi and Puglisi 1998; Wijmenga and van Buuren 1998).

L’attribution séquentielle des protons non-échangeables (Figure 21 et 22) a été réalisée en utilisant comme point de départ les H2 des adénines A21 et A22. Les incertitudes dans l’attribution dues aux recouvrements des signaux dans la région 7,6- 7,8 et 5,6-5,7 ppm ont pu être levées en effectuant des spectres NOESY à différentes températures. Les déplacements chimiques des différents protons de l’oligonucléotide sont reportés dans le tableau 23.



**Figure 21.** Spectre NOESY de la tige boucle SL1\* dans D2O à 27°C avec un temps de mélange de 450ms. Région des connectivités aromatiques /H1'. La position des protons H2 est indiquée en vert et celle des H5 en rouge.



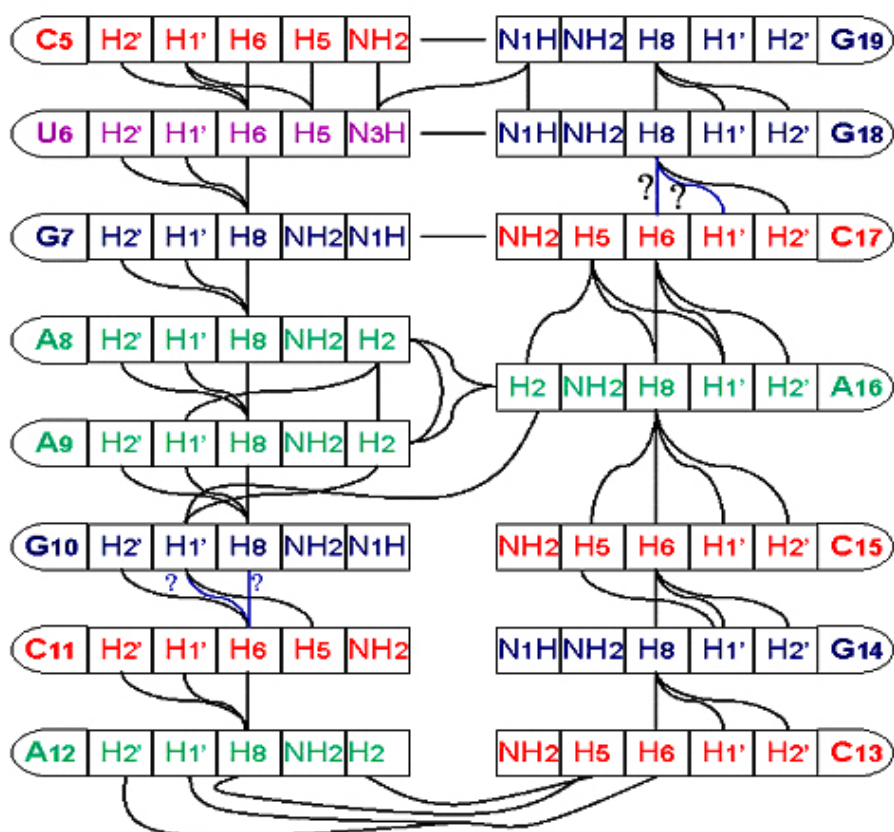
**Figure 22.** Spectre NOESY de la tige boucle SL1\* dans D2O à 27°C avec un temps de mélange de 250ms. Région des connectivités des protons aromatiques/aromatiques. La position des protons H2 est indiquée en vert .



	NH	NH2	H6/H8	H5	H2	H1'	H2'
C1		7,18/8,14	8,14	6,04		5,55	4,47
U2	14,53		8,14	5,50		5,65	4,53
U3	13,17		8,06	5,69		5,71	4,58
G4	12,44	7,84/7,94	7,76			5,77	4,47
C5		6,86/8,19	7,53	5,18		5,43	4,55
U6	11,59		7,70	5,65		5,58	4,22
G7	12,67	*	7,81			5,66	4,30
A8		*	8,17		7,93	5,92	4,64
A9		*	8,23		8,05	5,96	4,58
G10		*	7,54			5,57	4,64
C11		*	7,55	5,14		5,45	4,46
A12		*	8,13		7,59	5,91	4,41
C13		*	7,62	5,32		5,60	4,27
G14		*	7,67			5,69	4,58
C15		*	7,49	5,11		5,45	4,52
A16		*	8,00		7,67	5,94	4,53
C17		6,74/8,08	7,49	5,34		5,31	4,44
G18	10,30	*	7,51			5,60	4,68
G19	13,11	*	7,24			5,66	4,45
C20		6,85/8,32	7,66	5,21		5,49	4,51
A21		7,72/6,27	7,97		6,80	5,86	4,58
A22		*	7,76		7,56	5,83	4,41
G23	13,54	*	7,23			5,70	3,93

**Tableau 23.** Déplacements chimiques de la tige boucle SL1\*. Les protons échangeables sont attribués à pH 5,7 dans H2O:D2O (90:10) à 9°C et les non-échangeables à 27°C dans D2O, pH 7.0. (\* non attribué).

Les résultats obtenus en effectuant des spectres NOESY dans D2O et H2O indiquent que la séquence SL1\* est vraisemblablement sous la forme d'une tige boucle dans laquelle les 7 premières paires de bases sont appariées. La boucle SL1\* apparaît relativement structurée. En effet, nous avons pu effectuer l'ensemble de l'attribution séquentielle aromatique/H1' interbrin et avons trouvé de nombreuses connectivités entre les résidus A8 ,A9,G10,A16 (Figure 24). Cependant, les connectivités impliquant les H5 des cytosines de la boucle avec les protons aromatiques de leurs voisins en 5' sont très faibles voir inexistantes. Ceci peut être dû à un arrangement particulier des bases dans la boucle ou à une dynamique importante. D'autre part, les spectres COSY ont révélé qu'il n'existait pas de connectivités entre les H1' et H2' pour les différents résidus de l'oligonucléotide à l'exception des résidus G23, A8 et A9 indiquant une conformation nord de la majorité des sucres.

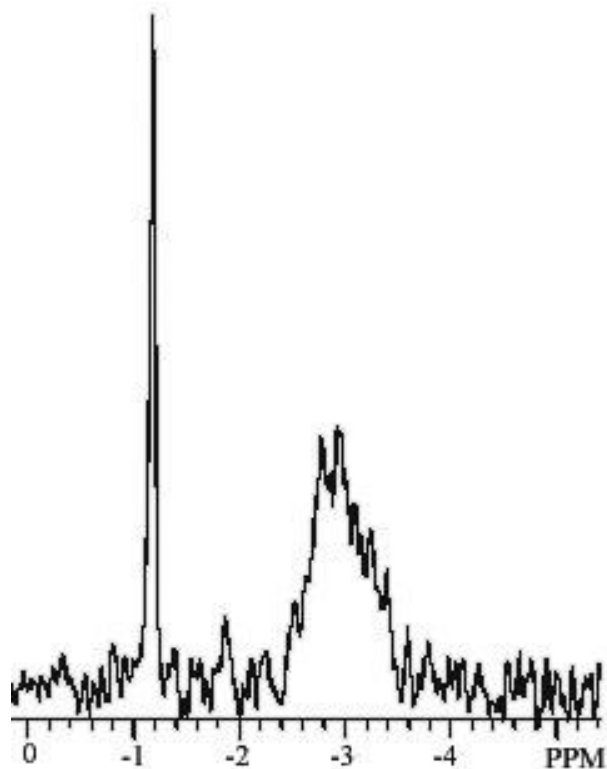


**Figure 24.** Ensemble des contraintes de distances sur la tige boucle ARN (? indique un recouvrement entre plusieurs connectivités).

### I-2-3 Spectre Phosphore

Les spectres phosphore peuvent apporter des informations sur la présence ou non de déformation du squelette phosphodiester. Dans le cadre de notre étude du monomère SL1\*, la quasi totalité des raies de résonance du  $^{31}\text{P}$  est concentrée sur une faible gamme de déplacement chimique de  $-1.6$  à  $-3,4$  ppm qui sont des valeurs standard dans le cas d'hélice régulière (Legault and Pardi 1994). Seule une raie de résonance est déplacée vers les hauts champs ( $-1,9\text{ppm}$ ). Ceci montre que la conformation du squelette phosphodiester ne diffère pas pour tous les oligonucléotides à l'exception d'un.

Malheureusement, le recouvrement important des corrélations  $^1\text{H}_{3',5',5''}\text{-}^{31}\text{P}$  dans les expériences hétéro-TOCSY ne nous ont permis d'attribuer avec exactitude cette raie de résonance (Figure 25).



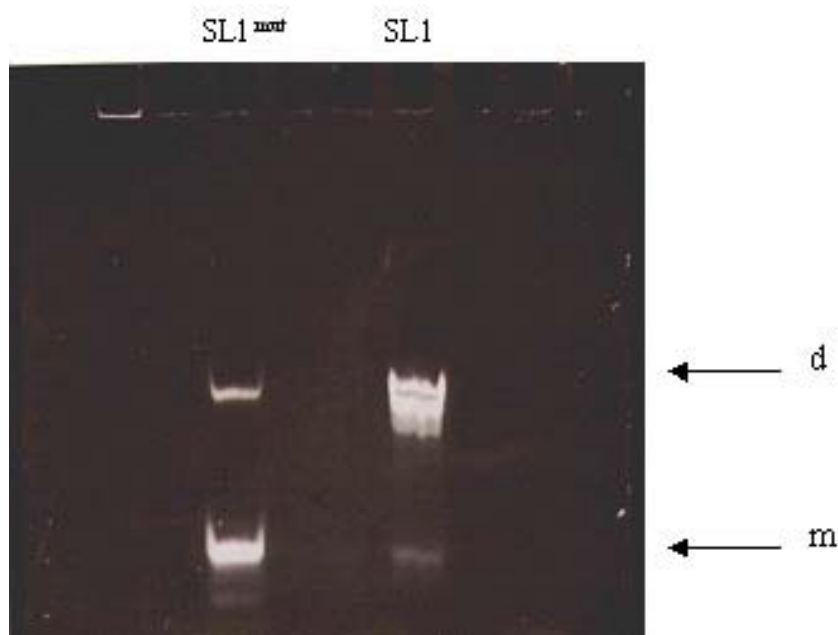
**Figure 25.** Spectre  $^{31}\text{P}$  de la tige boucle SL1\* à 27°C dans D2O. La raie à -1,1ppm correspond à la résonance du tampon phosphate.

### I-3 Structure du monomère SL1\*

#### I-3-1 SL1\* est monomérique

Il existe plusieurs arguments qui nous permettent d'affirmer que la tige boucle SL1\* se trouve sous forme monomérique.

Premièrement, nous avons réalisé des gels natifs (Figure 26)(voir matériels et méthodes) en comparant la tige boucle SL1 sauvage avec la séquence mutée SL1\* (Figure 19). La quantité de forme dimérique est de 90% pour la séquence sauvage tandis qu'on en retrouve moins de 15 % pour la séquence mutée. La présence de cette faible quantité de dimère ne gêne pas l'étude RMN du monomère SL1\*.



**Figure 26.** Gel d'électrophorèse à 4°C. SL1\* : Séquence mutée A12, SL1 : Séquence sauvage. (d : dimère et m : monomère). Les échantillons sont chauffés à 90°C pendant 3 min et laissés incuber pendant 15min dans la glace.

Deuxièmement, lors de l'étude RMN dans l'eau, à haute concentration, nous n'avons pas pu mettre en évidence d'appariement de type Watson-Crick pour les paires de bases centrales (Figure 20).

Troisièmement, nous avons calculé le temps de corrélation de l'oligonucléotide en utilisant le programme RELAZ (Lancelot et al. 1989). Ce programme permet de relier les intensités nOe des H5-H6 des cytosines C5, C20, C17 observées à différents temps de mélange, au temps de corrélation global de la molécule. Nous avons obtenu un temps de corrélation moyen de 3 ns pour les différentes connectivités, ce qui est tout à fait en accord avec une espèce moléculaire de 7000 Da. A titre d'information, pour le dimère étendu, nous avons obtenu un temps de corrélation de 6 ns dans les mêmes conditions (Girard et al. 1999).

## I-3-2 Modélisation moléculaire

### I-3-2-a. Etablissement des contraintes

#### *Contraintes sur les angles*

Les spectres COSY effectués sur la tige boucle nous ont permis de déterminer les conformations des différents sucres de l'oligonucléotide. Une constante de couplage quasi nulle (1 à 0 Hz) indique que le sucre se trouve en conformation nord ou C3' endo. Dans le cadre de notre étude, tous les nucléotides à l'exception des résidus A8,A9 et G23 se trouvent sous cette conformation nord, les autres se trouvant sous une conformation du sucre de type C2' endo. Un jeu de 46 contraintes d'angles dièdres ( $\nu_1$  et  $\nu_2$ ) a été créé pour rendre compte de ces plissements du sucre.

Nous avons pu voir précédemment que la quasi-totalité des raies de résonance du phosphore était regroupée sur un faible écart de déplacement indiquant une de très faible variation de la conformation des phosphates. Nous avons choisi de contraindre les angles du squelette phosphodiester ( $\alpha, \beta, \gamma, \epsilon, \xi$ ) avec des valeurs classiques d'hélice A (Saenger 1984).

#### *Contraintes de distance*

Les contraintes de distance sont issues des volumes des connectivités nOe intégrées à plusieurs temps de mélange (0,50,100,150,300,450ms). Pour cela, nous avons utilisé le programme RELAZ (Lancelot et al. 1989) qui permet de calculer la matrice de relaxation pour traduire les nOe en distances tout en tenant compte de la diffusion de spin.

Nous avons ajouté un jeu de 36 contraintes de distance pour rendre compte de l'appariement de type Watson-Crick des 7 premières paires de bases de la tige correspondant à l'observation des protons échangeables (imino-imino et imino-amino). De plus pour favoriser les appariements des bases, nous avons ajouté 6 contraintes de plan pour les 6 premières paires de bases lors des premiers cycles de minimisation. Ces contraintes sont appliquées avec une force faible de  $25 \text{ kcal.mol}^{-1} \cdot \text{Å}^{-2}$  (pour comparaison, chaque contrainte nOe a une force de  $50 \text{ kcal.mol}^{-1} \cdot \text{Å}^{-2}$ ).

Du fait de la forte dynamique de la boucle, plusieurs structures différentes peuvent respecter toutes les données RMN. Cependant, une analyse de toutes les distances inférieures à  $5 \text{ Å}$ , à l'intérieur de chacune des structures, indiquent que certaines proximités ne sont pas réelles puisque les corrélations ne sont pas observables sur les spectres RMN. Cette constatation nous a amené à définir de nouvelles contraintes, dites d'exclusion, supérieures à une distance de

5,5Å. Un total de 63 contraintes d'exclusion a été nécessaire pour obtenir une bonne résolution de la structure tridimensionnelle. L'ensemble des contraintes est présenté dans le tableau 27.

Contraintes RMN pour la structure de la tige boucle SL1*	
<b>A. Contraintes de distance</b>	
Intra-résiduelles	50
Inter-résiduelles	163 (dont 63 d'exclusion)
Liaison Hydrogène	36
Total des contraintes de distance	249
NOE / résidu	10,8
Constante de force	$50 \text{ kcal.mol}^{-1} \cdot \text{Å}^{-2}$
<b>B. Contraintes d'angle dièdre</b>	
Plissement du sucre	46
Squelette phosphodiester	109
Total des contraintes d'angle	155
Angles / résidu	6,7
Constante de force	$50 \text{ kcal.mol}^{-1} \cdot \text{Å}^{-2}$
<b>C. Total</b>	
Nombre de contraintes	404
Contraintes / résidu	17,6

**Tableau 27.** Résumé de l'ensemble des contraintes RMN utilisées pour la modélisation de la tige boucle SL1\*

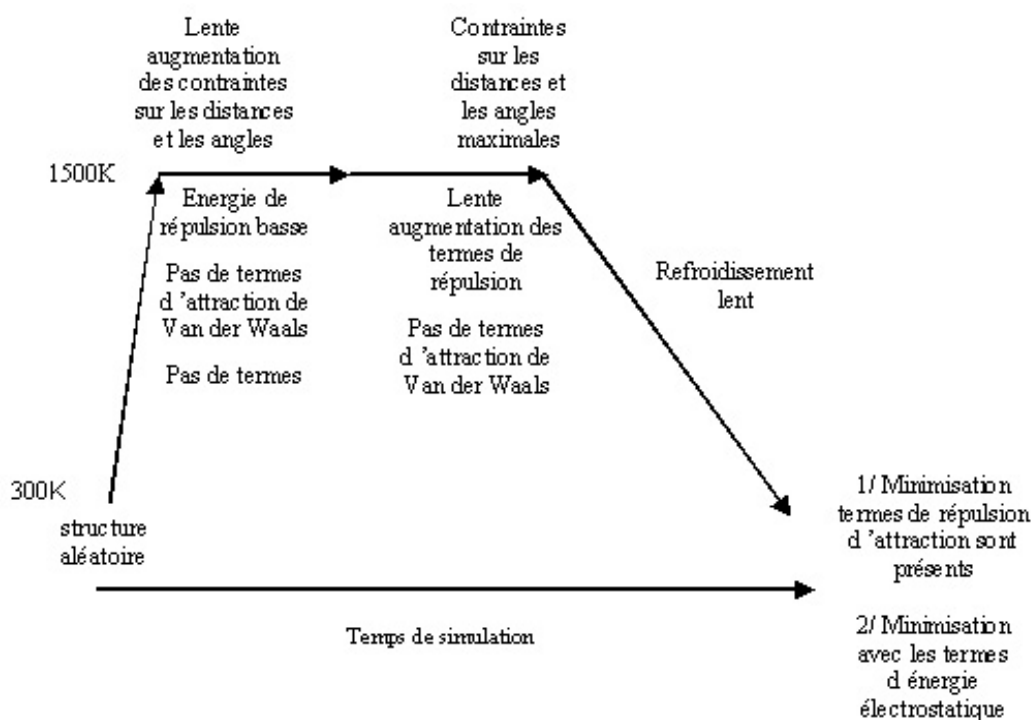
### I-3-2-b. Etablissement des incertitudes sur les contraintes

La résonance magnétique nucléaire ne permet pas de connaître avec exactitude la valeur des différents angles et des différentes distances. Au cours de la modélisation moléculaire, toutes les contraintes sont entrées dans le logiciel sous la forme d'un intervalle plus ou moins grand. De plus, il peut exister des différences locales du temps de corrélation ou de l'échange chimique entre différentes formes, même si celui-ci n'est pas visible sur nos spectres. Ces effets peuvent altérer les distances calculées à partir des effets nOe pour les protons possédant une grande flexibilité. Nous avons donc choisi différents intervalles de contraintes suivant qu'il s'agisse de contraintes d'angle dièdre ou de contraintes de distance. Nous avons aussi

découpé notre molécule en trois zones : la tige, les résidus d'entrée de boucle (A8,A9,A16) et les résidus de la séquence auto complémentaire. Pour les angles dièdres nous avons choisi quatre intervalles par rapport à la valeur moyenne de l'hélice A : +/- 5°,15°,30°,45°. Pour les distances, nous avons deux types d'intervalles par rapport à la valeur (d) calculée par le programme RELAZ : une erreur de 15% ou une erreur comprise entre  $0,5d < r < 1,3d$ . Ce dernier type d'écart très large se rapproche plus d'un classement des distances entre trois catégories fort, moyen, faible que l'on appelle des contraintes UPL pour « upper lower limit ».

### I-3-2-c. Résolution de la structure tridimensionnelle

La détermination de la structure finale a été effectuée en utilisant la méthode de recuit simulé dans l'espace cartésien à l'aide du logiciel XPLOR ( version 3,851) (Brünger 1992). Nous avons généré 200 structures de départ aléatoires, respectant la covalence des liaisons. Le protocole de recuit simulé (Figure 28) commence par 1000 pas de minimisation, selon l'algorithme de Powell (Mécanique moléculaire avec la méthode du gradient conjugué), de façon à éliminer les contacts stériques. Une période de 10ps de dynamique à 1500K permet l'exploration de l'espace conformationnel sous contraintes RMN. Le champs de force est simplifié en tenant compte uniquement des longueurs de liaison, des angles de liaison, des angles impropres, du potentiel répulsif de Van Der Waals, des contraintes de distance, d'angle et de plan. Puis, pendant 8 ps, les masses et les tailles des atomes sont lentement amenés à leurs valeurs classiques. La phase de refroidissement (1500K à 300K) se fait suivant une dynamique de 20 ps. Enfin, une dernière étape de minimisation de 1000 pas par la méthode du gradient conjugué finit le protocole. Les contraintes RMN sont introduites progressivement à chaque cycle de calcul. Dans un premier temps, nous ajoutons les contraintes de type Watson-Crick et les inter-résidus pour finir par les contraintes d'exclusion.



**Figure 28.** Schéma du déroulement des étapes de recuit simulé et de minimisation des structures

Finally, the resulting structures are refined by 10000 steps of energy minimization according to the Powell method using a complete potential energy function including the attractive parameters of Van Der Waals, the electrostatic energy and the hydrogen bonds. The structures thus obtained are analyzed and selected according to their number of violations of NMR constraints, their global energy and by the calculation of chemical shifts in return with the help of the software NUCHEMICS (Wijmenga et al. 1997). This calculation in return makes it necessary to use the semi-empirical parameters developed by (Giessner-Prettre and Pullman 1987). The general equation for the calculation of the chemical shift for a nucleus 1 in an oligonucleotide containing N atoms is defined by the equation :

$$\delta_1 = \delta_{1,int ra} + \delta_{1,1CA} + \sum_{j=1}^N [\delta_{1,1jrc} + \delta_{1,1jma} + \delta_{1,1jE} + \delta_{1,1jCT}]$$

The first term of the equation represents the chemical shift of the hydrogen 1 in the fragment A in the absence of other nucleotides, this term is purely empirical. The second



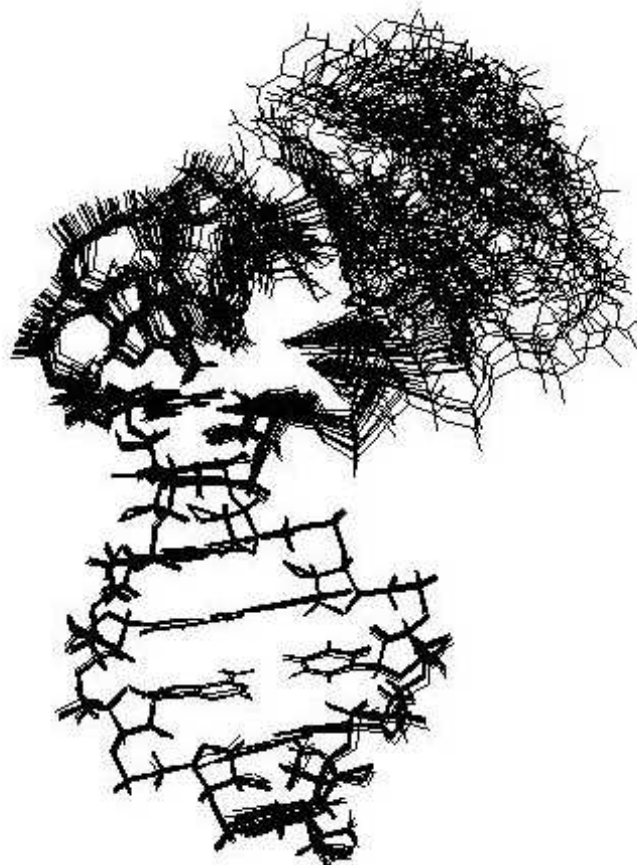
correspond à la contribution due aux changements conformationnels du résidu A en l'absence des autres résidus.

Les autres paramètres représentent les contributions apportées par l'environnement extérieur du noyau 1, par ordre d'apparition dans l'équation nous avons en premier l'effet dû au courant de cycle. Cet effet est essentiel dans les acides nucléiques. Le second paramètre correspond à l'anisotropie magnétique locale due par exemple à une distribution asymétrique des électrons tandis que le troisième rend compte de la polarisation issue du déplacement des électrons le long des liaisons chimiques au voisinage du noyau considéré. Le quatrième représente le transfert de charge lors d'une liaison chimique.

### I-3-3 Analyse structurale de la tige boucle SL1\*

A partir des 200 structures de départ, nous avons effectué une première pré-sélection sur ces structures en effectuant une modélisation utilisant les contraintes de distance avec des incertitudes de +/-15% sur les valeurs calculées par RELAZ et les angles dièdres avec des incertitudes de +/- 45° sur la valeur classique d'une hélice. Nous avons pré-sélectionné 19 structures sur lesquelles nous avons effectué d'autres étapes de modélisation en utilisant différentes gammes d'incertitude sur les angles dièdres et sur les distances. Le but est de déterminer les plages d'incertitude les mieux adaptées pour la modélisation de la tige boucle d'ARN. Les meilleurs résultats ont été obtenus avec des contraintes sur la tige de 15% sur les  $nOe$  et 15° sur les angles, pour les 3 adénines (A8, A9, A16) d'entrée de boucle des contraintes de type UPL sur les distances et 15° sur les angles et pour la séquence apicale des incertitudes de 45° sur les angles et des distances sous forme UPL.

Nous avons obtenu 19 structures (Figure 29) qui ne présentent pas de violation de contraintes supérieure à 0,3Å et présentant des déplacements chimiques calculés en retour compatibles avec ceux observés expérimentalement. Tous les protons des 19 structures sont en accord avec ceux calculés avec une différence de 0,5 ppm. Néanmoins, aucune des structures ne présentent simultanément des déplacements chimiques calculés pour les H5 des cytosines de la boucle en accord avec ceux expérimentaux. Ceci est probablement dû à la grande flexibilité de la boucle. Néanmoins, les déplacements chimiques observés pour ces H5 de cytosines (5,1-5,3 ppm) sont compatibles avec un empilement des cytosines avec les résidus voisins. Une cytosine libre présente un déplacement chimique de 6,1ppm. De plus, nos 19 structures présentent un déplacement chimique des H2 pour les adénines d'entrée de boucle (A8,A9,A16) compatible avec ceux calculés (Tableau 30).



**Figure 29.** Superposition des 19 structures de la tige boucle SL1\*. Les structures sont superposées sur les atomes lourds de la tige.

<u>Précision des structures</u>	
Nombre de structures	19
Nombres de violation nOe >0,3Å	0
Nombres de violation dièdre >5°	0
RMSD global	1,39 ± 0,70 Å
RMSD tige (bases 1-7,17-23)	0,15 ± 0,08 Å
RMSD boucle (bases 8-16)	1,02 ± 0,39 Å
<u>Energies</u>	
Energie totale	-338 ± 20 kcal.mol <sup>-1</sup>
Energie de liaison	+13 ± 1 kcal.mol <sup>-1</sup>
Energie angulaire	+188 ± 5 kcal.mol <sup>-1</sup>
Energie torsion	+395 ± 7 kcal.mol <sup>-1</sup>
Energie dièdre impropre	+25 ± 4 kcal.mol <sup>-1</sup>
Energie Van Der Waals	-369 ± 4 kcal.mol <sup>-1</sup>
Energie électrostatique	-469 ± 11 kcal.mol <sup>-1</sup>
Energie liaison Hydrogène	-132 ± 4 kcal.mol <sup>-1</sup>
Energie nOe	+12 ± 4 kcal.mol <sup>-1</sup>
Energie dièdre	+0,5 ± 0,1 kcal.mol <sup>-1</sup>

**Tableau 30.** Résumé des énergies et des RMSD calculés sur les atomes lourds par rapport à la moyenne non minimisée, pour la tige boucle SL1\*.

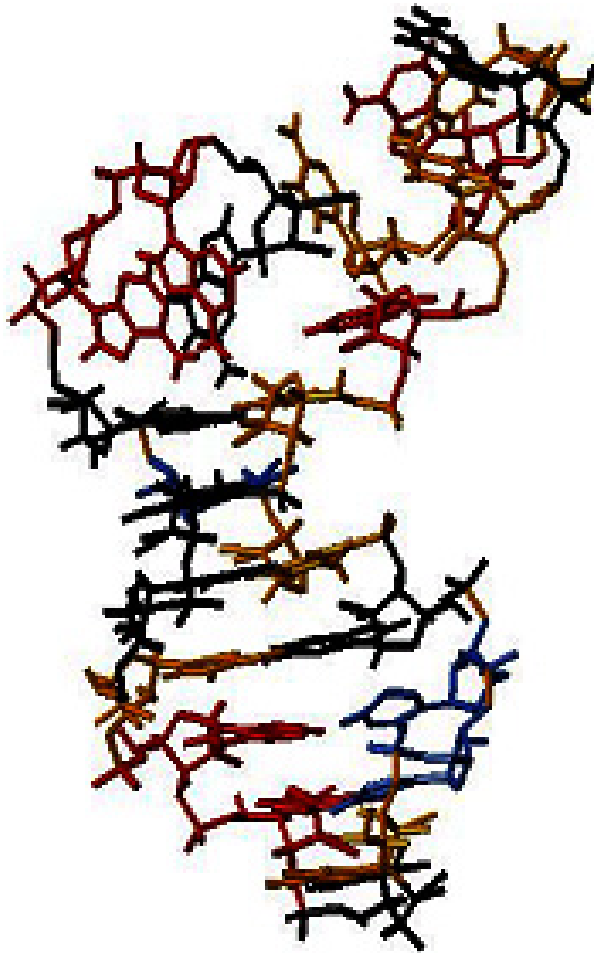
La structure tridimensionnelle obtenue (Figure 31) est une tige boucle présentant plusieurs zones de structuration.

Premièrement, une tige très bien structurée est présente sous la forme d'une double hélice. Les paramètres de cette double hélice ont été calculés à l'aide du logiciel CURVES 5.1 (Lavery and Sklenar 1988, 1989) et correspondent à ceux présents dans une double hélice de type A, classiquement rencontrée dans les ARN. Tous les appariements de type Watson-Crick sont présents, y compris pour la paire G7-C17 dont les contraintes RMN sont imposées avec une constante de force plus faible que les autres paires de bases.

Deuxièmement, les adénines A8,A9,A16 et dans une certaine mesure G10 apparaissent structurées. Cette région correspond à la jonction entre la boucle et l'hélice de la tige. L'adénine A8 se place perpendiculairement à l'axe de l'hélice tandis que l'adénine A9 s'empile entre l'adénine A8 et la guanine G10. Cet empilement est stable et induit une torsion du squelette. La boucle apparaît donc inclinée par rapport à l'axe de l'hélice.

Troisièmement, l'empilement des adénines de la jonction implique également que les bases A8,G10,A16 se trouvent à l'intérieur de la boucle et, de ce fait, empêchent les résidus de la séquence apicale de s'auto appairer de façon intra-moléculaire. Ces résidus se retrouvent expulsés vers l'extérieur de la boucle et exposés au solvant. Bien que les résidus de la séquence apicale forment une zone très flexible dans nos structures, les cytosines centrales (C11, C13,C15) restent empilées avec la purine adjacente. Cet empilement est compatible avec la valeur des déplacements chimiques des H5 aux environs de 5,2ppm.

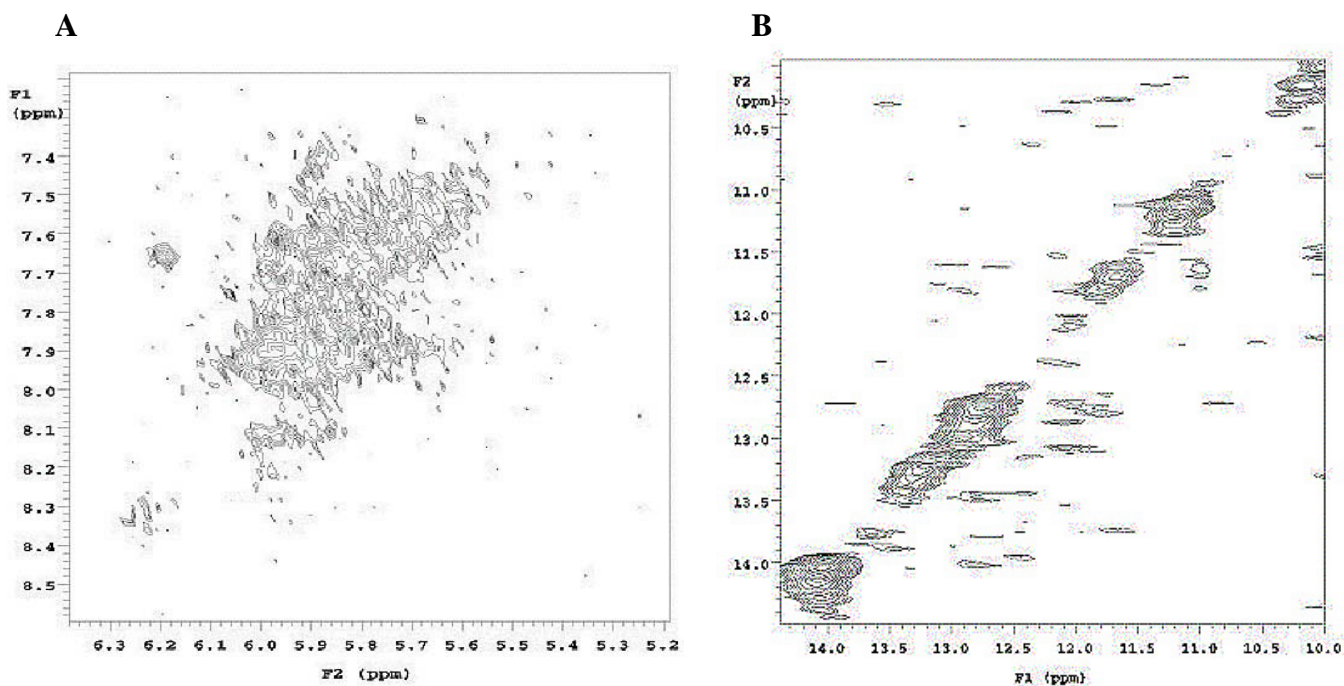
Nous avons pu obtenir un ensemble de structures de la tige boucle SL1 qui montre une grande flexibilité du sommet de la boucle sans pour autant obtenir une très bonne résolution de celle-ci. En effet, la RMN ne donne qu'une image moyenne de la structure en solution. Afin de mieux comprendre les phénomènes dynamiques intervenant dans la boucle, nous avons complété cette étude par une simulation de dynamique moléculaire avec solvant et ions explicites.



**Figure 31.** Structure tridimensionnelle de la tige boucle SL1\*. (en bleu, les uraciles ; en rouge, les adénines ; en noir, les guanines et en or les cytosines).

## I-4 Simulation de dynamique moléculaire

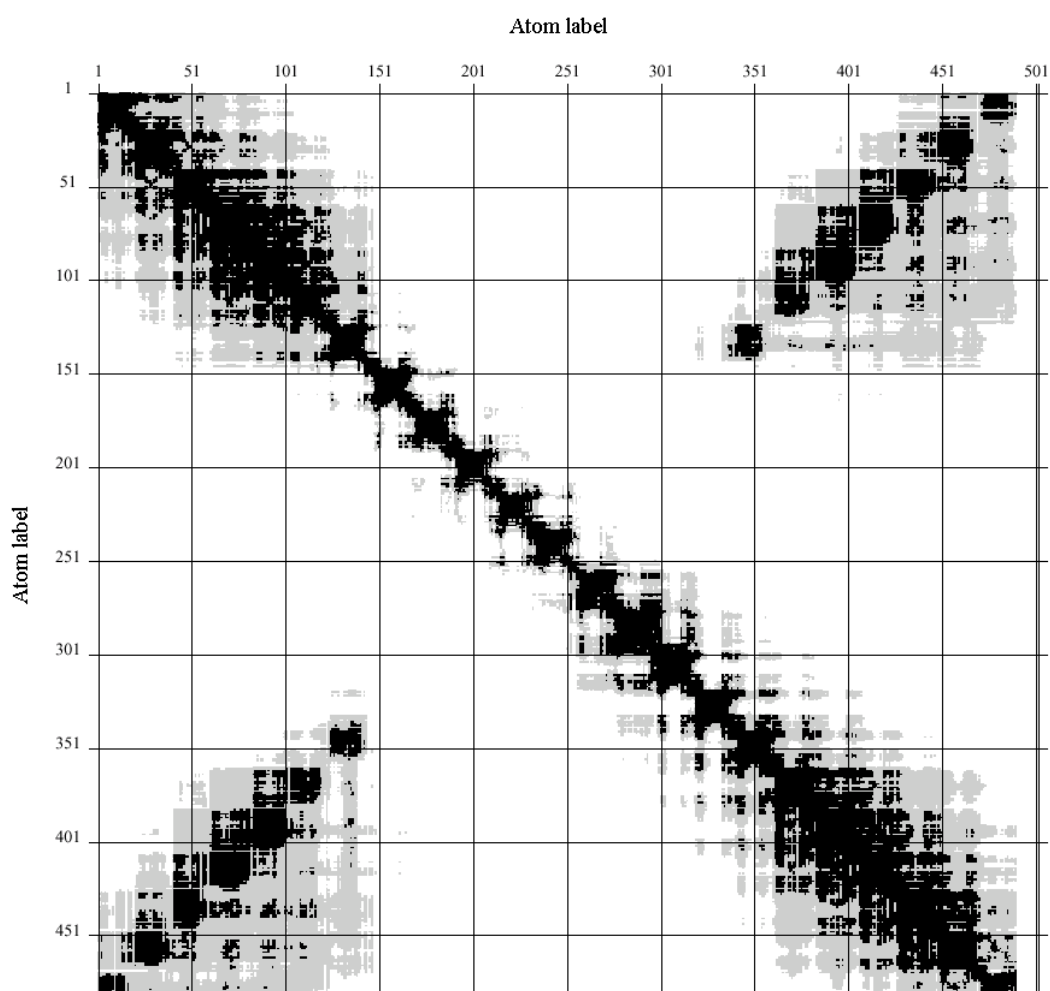
La boucle de SL1\* présente une forte dynamique d'après son étude par RMN mais reste relativement stable et organisée contrairement à son homologue ADN (Figure 32). En effet, l'étude de l'ADN SL1\* montre un comportement structural différent de celui de l'ARN et apparaît comme déstructuré.



**Figure 32.** Spectre de la tige boucle SL1\* ADN à 600MHz: **A/** Connectivités entre protons aromatiques H1' avec un temps de mélange de 300ms à 27°C dans D2O, **B/** Régions des protons imino avec un temps de mélange de 120ms à 15°C dans H2O.

Pour comprendre cette différence de stabilité entre les deux molécules, nous avons voulu connaître les facteurs stabilisants pouvant intervenir dans la structure de l'ARN notamment connaître le réseau de liaisons hydrogène entre l'oligonucléotide et le solvant. Nous avons donc collaboré avec l'équipe du Dr Genest qui a effectué une simulation dynamique à partir de la structure moyenne de la tige boucle SL1\* obtenue par RMN. Cette simulation a été réalisée à l'aide du logiciel AMBER 5.0 (Case et al. 1997). Durant cette simulation, la molécule d'ARN est intégrée dans une 'boite' comportant 4313 molécules d'eau et 22 ions  $\text{Na}^+$  de manière à mimer le milieu physiologique. Après une période d'équilibration thermique à 300K, une dynamique de 10ns a pu être effectuée avec un stockage des coordonnées toutes les 0.5ps pour l'oligonucléotide et de 3ps pour le solvant.

Le calcul des RMSD des atomes des structures obtenues par rapport à la structure initiale permet d'établir des cartes de flexibilité. Cette carte permet de représenter les mouvements sur les différentes régions de l'oligonucléotide (Figure 33). Conformément aux résultats exposés par l'étude RMN, la tige semble relativement stable et comporte une amplitude de mouvement classique pour une double hélice d'ARN. La boucle est, quant à elle, animée de nombreux mouvements. L'orientation générale de la tige vis à vis de la boucle, au cours de la simulation, varie entre 95 et 145°. Il est également à noter que la paire de bases G7-C17 est toujours appariée bien qu'elle présente de petites fluctuations. Ceci peut s'expliquer par le fait qu'il s'agit de la jonction de la boucle avec la tige.



**Figure 33.** Carte de flexibilité de la tige boucle SL1\*. Le premier brin de la tige correspond aux atomes 1-143, la boucle 144-337 et le second brin de la tige de 338 à 490. Le noir, le gris et le blanc représentent respectivement les zones rigides, semi-rigides et flexibles.

Lors de cette étude, l'équipe du Dr Genest a pu également définir un réseau de liaisons hydrogène entre l'oligonucléotide et le solvant intervenant directement dans la stabilisation de la structure de la tige boucle SL1\* (Tableau 34). Une liaison hydrogène est définie par une distance entre les atomes donneurs et accepteurs de moins de 3,5Å et un angle entre le donneur, l'hydrogène et l'accepteur supérieur à 45°.

Liaison Hydrogène	C1	U2	U3	G4	C5	U6	G7	A8	A9	G10	C11	A12
% présence intramoléculaire	64,3	67,8	0,0	0,0	0,0	42,3	82,9	79,0	26,3	69,3	22,4	69,3
Nombre moyen avec le soluté donneur	1,09	1,15	1,12	1,22	1,18	1,15	1,15	0,82	1,17	1,32	1,28	0,86
Nombre moyen avec le soluté accepteur	1,17	0,44	0,59	0,74	0,77	0,40	0,30	0,27	0,66	0,40	0,20	0,52

Liaison Hydrogène	C13	G14	C15	A16	C17	G18	G19	C20	A21	A22	G23
% présence intramoléculaire	47,6	59,0	100,0	22,6	79,2	0,0	20,5	0,0	0,0	61,3	0,0
Nombre moyen avec le soluté donneur	1,16	0,93	1,13	1,18	1,15	1,24	1,22	1,17	1,21	1,26	1,15
Nombre moyen avec le soluté accepteur	0,50	0,19	0,23	0,53	0,19	0,77	0,73	0,63	0,66	0,47	1,59

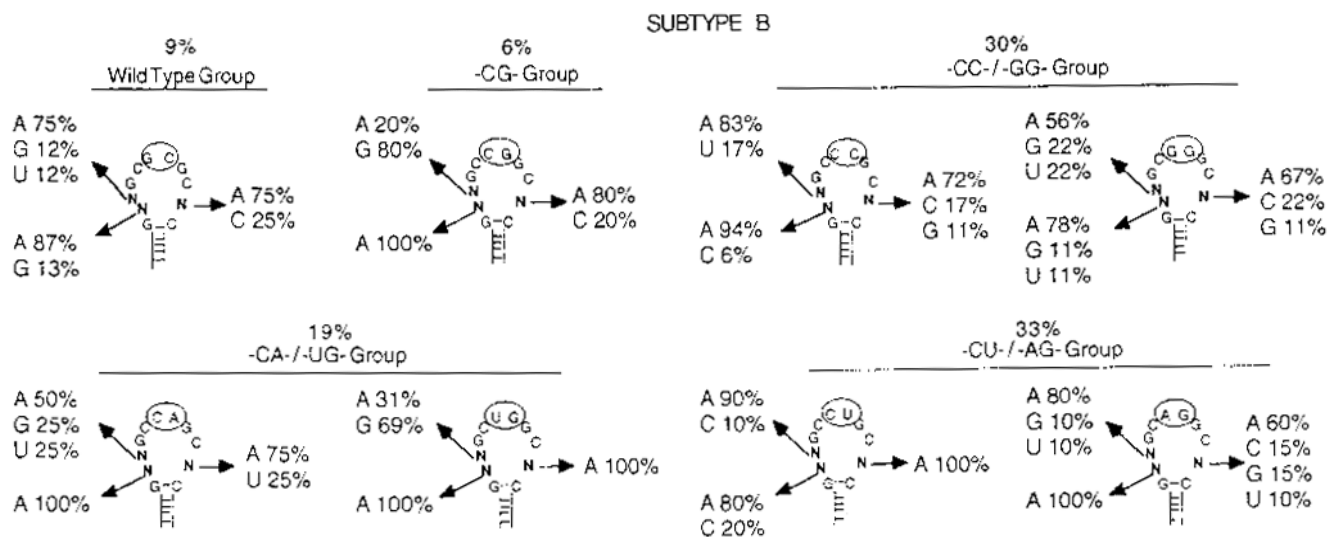
**Tableau 34.** Liaisons hydrogène impliquant le groupement 2'OH pour l'ARN lors des 10ns de dynamique moléculaire avec solvant et ions explicites.

Les O2' forment des liaisons hydrogène intramoléculaires essentiellement avec les O4' et O5' du résidu n+1. Les O2' interviennent également dans la solvataion de la tige boucle en formant des liaisons hydrogène avec les molécules d'eau environnantes soit comme donneur ou comme accepteur. Le rôle stabilisant des 2'OH pourrait alors expliquer la différence de stabilité entre le monomère ADN et ARN et aussi pourquoi les tiges boucles sont plus fréquentes chez les ARN.

Une deuxième simulation de 6ns a été réalisée avec la même structure de départ mais en remplaçant l'adénine A12 par une guanine, pour retrouver la séquence sauvage de la tige boucle. Cette simulation présente les mêmes caractéristiques que la séquence mutée, preuve que le comportement de la structure n'est pas affecté par cette mutation.

## I-5 Discussion

L'étude aléatoire des séquences capables de provoquer un phénomène de dimérisation pour la tige boucle SL1 de l'isolat Lai (Lodmell et al. 2000, 2001; Lodmell et al. 1998), indique que les résidus situés au sommet de la boucle sont essentiels. Mais plus étonnant, il existe une très forte proportion de résidus puriques et plus particulièrement d'adénines au niveau des 3 résidus qui entourent la séquence auto complémentaire (Figure 35). Par conséquent, la seule présence d'une séquence auto complémentaire au sommet de la boucle ne suffit pas à la dimérisation des ARN. La présence de résidus puriques doit entraîner un réarrangement spatial de la séquence auto complémentaire nécessaire au processus de dimérisation.



**Figure 35.** Résumé des résultats de la sélection in vitro des séquences capables de dimériser pour le sous-type B de VIH-1 (isolat Lai) (Lodmell et al., 2001).

Si on observe la structure de la tige boucle SL1\*, on s'aperçoit que ces adénines sont impliquées dans un phénomène d'empilement. Cet empilement a deux conséquences majeures pour la structure. D'une part il provoque une torsion du squelette si bien que la boucle apparaît inclinée par rapport à l'axe de la double hélice. D'autre part, les bases 8, 9, 10, 16 se trouvent à l'intérieur de la boucle et, de ce fait, empêchent les résidus de la séquence apicale de s'auto appairer de façon intra-moléculaire. Ces résidus se retrouvent expulsés vers l'extérieur de la boucle et exposés au solvant. A partir de ces constatations, il est possible d'imaginer un processus de dimérisation simple qui résulte de la simple approche des deux



monomères. La position de la boucle par rapport à la tige permet le rapprochement d'une deuxième entité et les résidus auto complémentaires sont tournés vers l'extérieur permettant la formation de liaisons hydrogène nécessaires à la formation d'un dimère. Le phénomène de dimérisation doit également être facilité par une certaine dynamique de la boucle pour permettre l'appariement entre les deux monomères.

## I-6 Conclusion

Les tiges boucles sont des motifs structuraux particulièrement abondants dans les ARN. Les données thermodynamiques ont pu montrer que les boucles comprenant jusqu'à 10 résidus conduisent à des structures stables. Les analyses phylogéniques indiquent que les boucles comprenant 4 résidus (tétra-boucles) sont les boucles les plus fréquentes dans les grands ARN tels que L'ARN<sub>r</sub> ou L'ARN des génomes viraux (Woese et al. 1990). Les tétra-boucles les plus étudiées proviennent de deux familles : les GNRA et les UNCG, où N est un nucléotide quelconque et R représente une purine. Ce genre de boucle possède une grande stabilité par la formation d'appariement non canonique de type AG ou UG et également par des interactions de courant de cycle avec les 2 résidus restants (Abdelkafi et al. 1998; Jucker et al. 1996; Jucker and Pardi 1995; Leulliot et al. 1999; Menger et al. 2000). Ces tiges boucles constituent un motif de reconnaissance pour les protéines ou d'autres ARN comme pour la dimérisation des ARN rétroviraux (Chang and Tinoco 1997; Chang and Varani 1997; Ennifar et al. 2001; Mujeeb et al. 1998). Pour mieux comprendre le processus d'interaction entre les deux brins d'ARN lors du processus de dimérisation et de maturation par la NCp7, nous avons étudié la première entité présente dans le cytoplasme : la forme monomérique de la tige boucle. L'interaction de deux de ces tiges boucles aboutit à la formation d'un dimère. Pour cela, nous avons muté une guanine par une adénine étant donné que la tige boucle sauvage dimérise spontanément à la concentration utilisée en RMN. Cette mutation présente néanmoins l'avantage de garder le caractère purique de la base. L'étude RMN de la tige boucle SL1\* avec ces 9 résidus dans la boucle fait partie des rares résolutions structurales de tiges boucles possédant une boucle de grande taille (Colvin et al. 1993; Huang et al. 1996; Leeper et al. 2002; Sich et al. 1997; Wang et al. 1996).

Nous avons ainsi pu mettre en évidence une tige boucle avec une tige composée de 7 paires de bases organisée en double hélice A. Ce type de double hélice est classiquement observé chez les ARN. Il existe également une région structurée en entrée de boucle correspondant aux résidus A8, A9, G10 et A16, empilés les uns par rapport aux autres. Enfin, une séquence

apicale beaucoup moins structurée par rapport au reste de la molécule car soumise à une forte dynamique dont les résidus sont également largement exposés au solvant.

Nous avons également pu effectuer une simulation de dynamique moléculaire avec solvant et ions permettant d'obtenir des informations sur les mouvements et le réseau de liaisons hydrogène, notamment avec les 2'OH, impliqués dans la stabilisation de l'édifice moléculaire. Ce travail a fait l'objet d'une publication (Kieken et al. 2002).

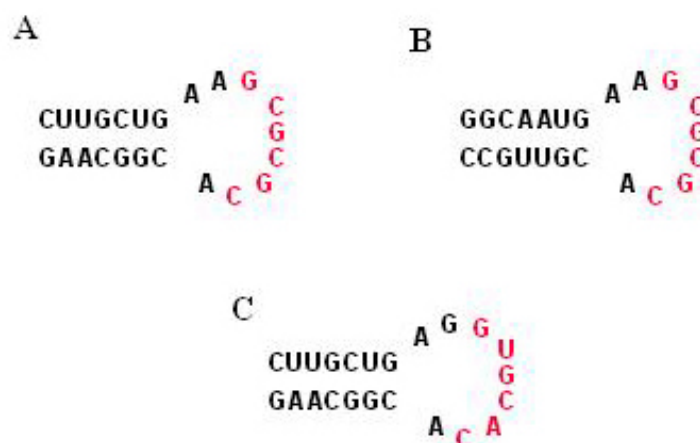
# ETUDE DU DIMERE INSTABLE DU DOMAINE SL1 DE VIH

## II-1 Bilan des études structurales menées sur le domaine SL1 de VIH

### II-1-1 Etude des dimères stables

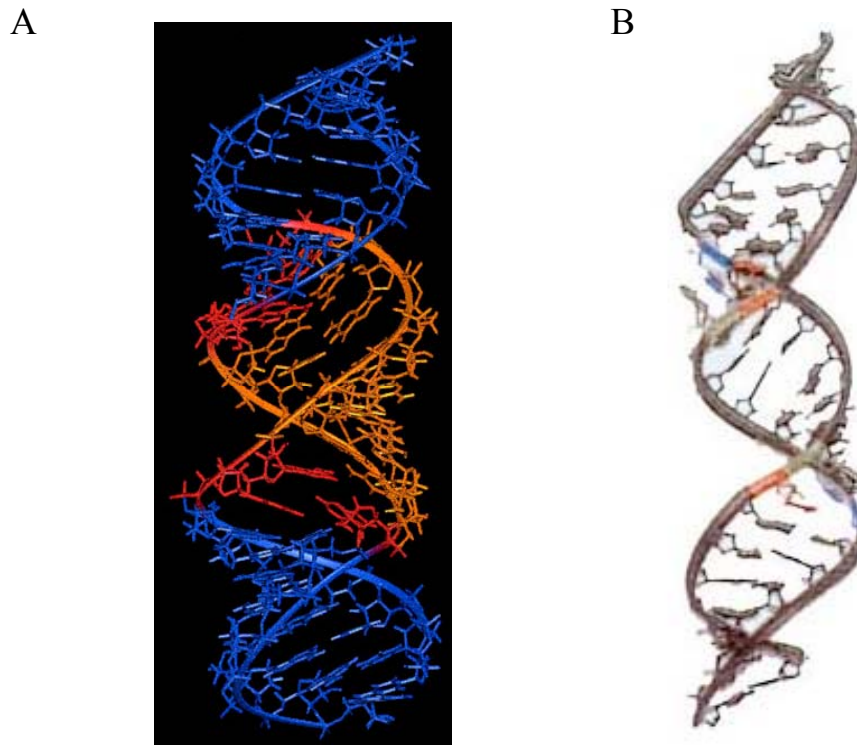
Trois études structurales ont été menées sur les formes stables de la tige boucle SL1 de VIH-1; deux par RMN (Girard et al. 1999; Mujeeb et al. 1999) et une par cristallographie (Ennifar et al. 1999).

Les études RMN ont été menées soit sur la séquence sauvage (Figure 36A) soit sur une séquence fortement mutée (Figure 36B). La forme stable, dont la structure a été élucidée dans notre équipe par F. Girard correspond exactement à l'extrémité de la tige boucle SL1 du domaine d'encapsidation (Berkhout and van Wamel 2000; Clever et al. 1995). La séquence étudiée par l'équipe de T. James possède une tige largement mutée (4 paires de bases sur les 7 pouvant exister), la boucle et l'entrée de celle-ci sont identiques dans les deux études.



**Figure 36.** Séquence de la tige SL1 de VIH utilisée pour les résolutions structurales. A/ Séquence de l'isolat Lai sauvage. B/ Séquence mutée de l'isolat Lai. C/ Séquence de l'isolat Mal sauvage. (en rose, les séquences auto-complémentaires).

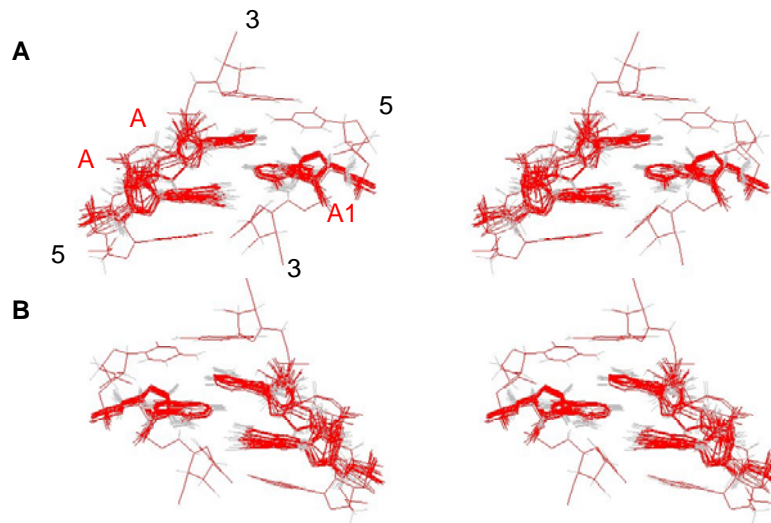
Les structures obtenues pour les deux séquences sont quasiment identiques (Figure 37). La forme dimérique stable adopte une conformation étendue. La structure comprend trois zones hélicoïdales. Deux zones symétriques comprenant les résidus 1 à 7 et 17 à 23 où tous les résidus sont appariés et forment une double hélice A et une troisième zone correspondant à la séquence auto-complémentaire (résidus 10-15), elle aussi sous la forme d'une double hélice A.



**Figure 37.** Structure du dimère étendu de VIH-1 **A/** Structure obtenue sur la séquence sauvage (Girard et al. 1999) **B/** Structure obtenue sur la séquence mutée (Mujeeb et al. 1999).

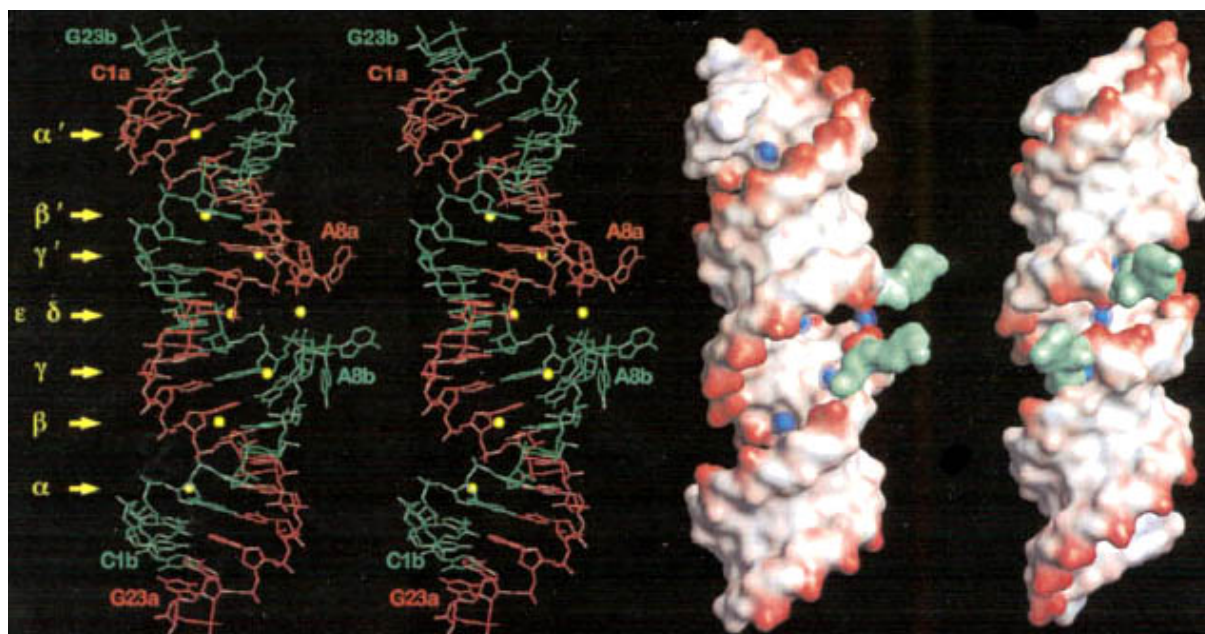
La particularité de la structure réside en un motif en fermeture éclair des 3 adénines (A8, A9, A16). Le résidu A16, du brin opposé, vient se glisser entre les résidus A8 et A9 provoquant un empilement partiel des adénines (Figure 38).

La présence de ce motif permet de ne pas générer de motif déstabilisant dans la structure. Au contraire, la position de ces adénines de jonction induit une stabilisation de la structure par cet empilement partiel et permet une organisation globale du dimère en double hélice A. Cette structure possède un maximum d'appariement entre les deux entités monomériques (20 au total) ce qui confère à l'ensemble une grande stabilité.



**Figure 38.** Le bulge d'adénines organisé en fermeture éclair. **A/** Coté sillon peu profond (petit sillon) ; **B/** Coté sillon profond (grand sillon). La figure présente une superposition des adénines du bulge des 15 meilleures structures du dimère sauvage de VIH-1 Lai (Girard et al. 1999).

L'étude cristallographique du dimère linéaire a été effectuée par (Ennifar et al. 1999) sur la séquence sauvage de l'isolat Mal (Figure 39). Les conditions de cristallisation favorisent la forme la plus stable, le dimère étendu. Cette structure, obtenue en présence de 100 mM de magnésium, présente la particularité de posséder l'adénine 8 complètement tournée vers l'extérieur par rapport à la double hélice A. De plus, il existe un appariement non classique G9-A16 permettant de stabiliser la région en amont de la séquence auto-complémentaire. Les auteurs supposent qu'il existe un équilibre entre une forme où l'adénine est tournée vers l'extérieur et une autre forme où celle-ci est intercalée dans le bulge. Les conditions expérimentales utilisées pour la cristallisation semblent déplacer cet équilibre vers la structure où A8 est expulsée vers l'extérieur ce qui permet des appariements entre les complexes dimériques.



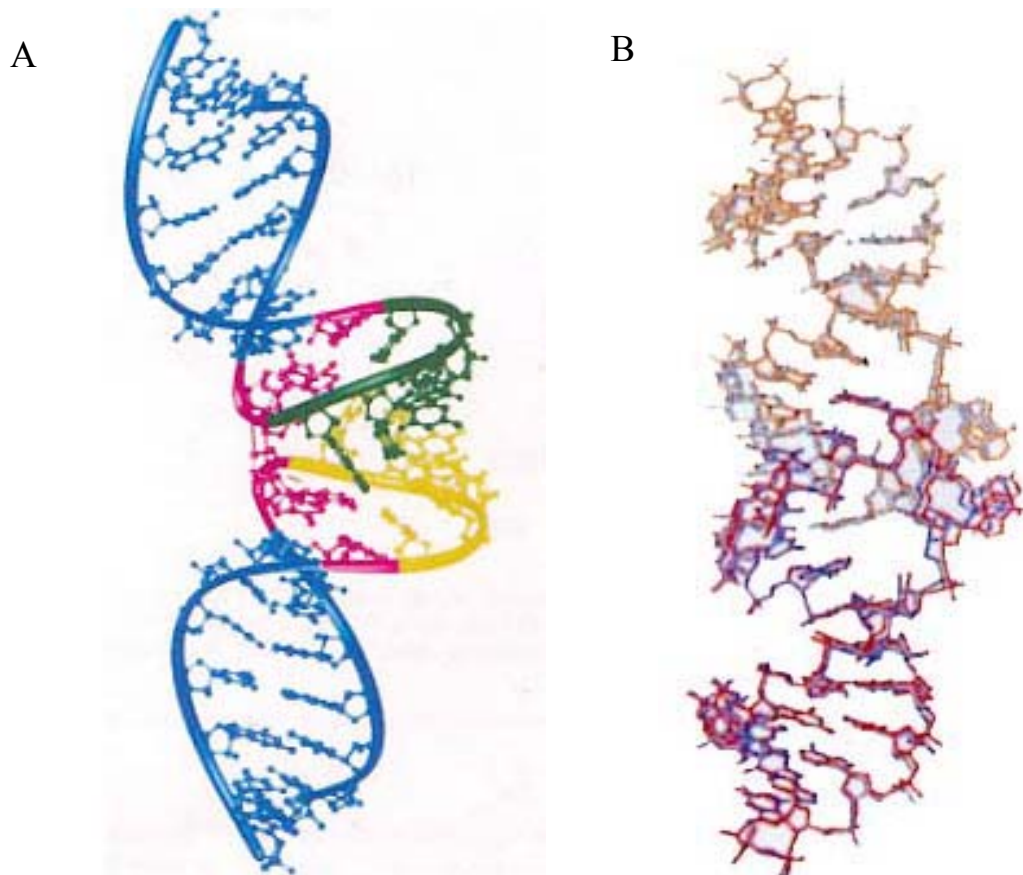
**Figure 39.** Structure de SL1-23mer de VIH-1<sub>Mal</sub> (Ennifar et al. 1999).

## II-1-2 Etude sur les dimères instables

Les études sur la forme instable du dimère de la tige boucle SL1 sur les isolats Mal et Lai sont au nombre de quatre, deux réalisées par RMN (Dardel et al. 1998; Mujeeb et al. 1998) et deux autres réalisées par cristallisation sur l'isolat Lai et Mal (Ennifar et al. 2001). L'étude structurale réalisée par RMN a été effectuée par l'équipe de T.James (Mujeeb et al. 1998). La séquence est une séquence fortement mutée de la tige boucle (Figure 40).

L'étude RMN met en évidence un homodimère symétrique où les tiges sont organisées en double hélice de type A canonique. Il est à noter que la dernière paire de bases G7-C17 de la tige, qui précède les adénines de jonction, n'est pas appariée et présente une déviation importante par rapport à la forme canonique.

Le point essentiel sur lequel repose l'organisation de la structure concerne l'architecture de la zone charnière entre les tiges, composée des résidus G7,A8,A9,A16,C17. Le résidu A9 du premier brin s'intercale entre les résidus A8' et C17' du second brin (l'apostrophe désigne le second brin). L'empilement de ces trois bases permet de stabiliser le domaine de jonction entre la tige et la séquence auto-complémentaire.



**Figure 40.** Structures du ‘kissing complex’ du domaine SL1 de VIH publiées à ce jour. **A/** Structure RMN de SL1\* Lai publiée par (Mujeeb et al. 1998) **B/** Superposition des structures cristallographiques de SL1 Lai ( en rouge et orangé) et SL1 Mal ( bleu foncé et bleu clair) (Ennifar et al. 2001).

Les études cristallographiques effectuées sur l’isolat Lai et l’isolat Mal par Ennifar et al. (2001), révèlent des structures assez comparables (Figure 40). Ces deux structures ne diffèrent que par la conformation de la purine en position 9, dans l’isolat Lai l’adénine est en position *anti* alors que pour l’isolat Mal la guanine est en position soit *anti*, soit *syn*.

Cette structure possède une tige avec 7 paires de bases appariées. Ces appariements sont de type canonique à l’exception de l’avant-dernière base U6-G18 qui présente une torsion hélicoïdale. L’interaction entre les deux entités monomériques se réalise par l’appariement des 6 résidus de la séquence auto-complémentaire. La particularité structurale des formes instables des deux isolats est un motif ‘purine zipper’. Les 2 résidus puriques de l’entrée de boucle sont exclus de la double hélice et sont empilés l’un sur l’autre. Il existe un espace libre entre ces deux purines qui permet l’intercalation de deux autres purines provenant d’un autre

dimère. La forme instable n'existerait donc pas sous forme de dimère mais de dimère de dimères. De plus, il est vraisemblable que ce motif 'purine zipper' dirige la cristallisation de la forme cristalline.

## II-2 Pourquoi s'intéresser de nouveau à la forme instable de VIH-1 Lai ?

Nous avons décidé d'étudier à nouveau la structure correspondant à la conformation la moins stable de la tige boucle SL1 du domaine d'encapsidation de VIH même si celle-ci a déjà été obtenue par RMN (Mujeeb et al. 1998). En effet, il existe certaines contradictions entre les résultats menés par l'équipe de T. James et les résultats obtenus dans notre équipe lors de la résolution du mécanisme de passage entre le dimère stable et le dimère instable (Theilleux-Delalande et al. 2000). Les raisons qui nous ont poussées à entreprendre l'étude de la forme instable sont les suivantes :

- La séquence étudiée pour la résolution structurale en RMN par l'équipe de T. James est une séquence fortement mutée au niveau de la tige (Figure 36). La séquence de la tige a été réarrangée pour se terminer par trois paires GC consécutives. Les auteurs affirment que cette mutation permettrait d'optimiser les rendements de production de l'ARN lors de la transcription enzymatique. Nous pensons que cette mutation engendre une plus grande stabilité de la tige et probablement celle du complexe boucle-boucle, facilitant ainsi son étude par RMN. Néanmoins, il existe un subtil équilibre entre les contraintes relatives à l'appariement des bases dans la tige et l'association des boucles et toute modification qui vise à stabiliser l'édifice moléculaire risque d'entraîner des modifications sur la structure globale de la molécule. Il est donc nécessaire de pouvoir étudier la séquence de la tige boucle sous sa forme native et sans réarrangement des appariements.
- L'obtention de la forme a été obtenue par refroidissement rapide de l'échantillon suite à sa dénaturation par la chaleur. Ce mode de refroidissement s'il n'est pas couplé à une dilution de l'échantillon au préalable ne permet pas d'isoler seulement la forme instable (Theilleux-Delalande et al. 2000) car il existe une compétition entre l'association de deux brins dénaturés et l'association entre 2 tiges boucles.



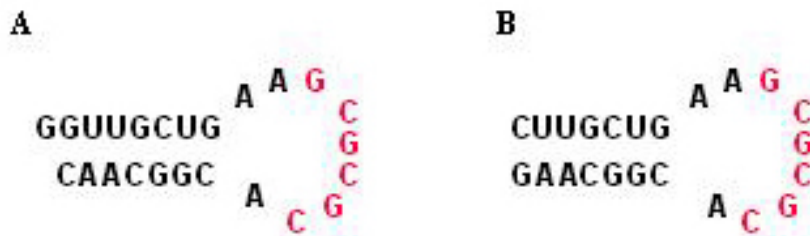
- Comme nous l'avons vu dans le paragraphe précédent, la dernière paire de bases, qui précède les régions des adénines de jonction, n'est pas appariée. Il est même spécifié dans l'article que tous les essais lors de la modélisation moléculaire pour appairier ces deux dernières bases se sont traduits par une destruction du complexe boucle-boucle. Or, lors de l'étude du mécanisme de passage entre le dimère instable et le dimère stable sur la séquence sauvage (Theilleux-Delalande et al. 2000), il a été observé que la paire de bases G7-C17 était appariée. Dès lors, il est fortement possible que la zone de jonction entre la tige et la séquence auto-complémentaire soit complètement différente entre la structure présentée par (Mujeeb et al. 1998) et celle de l'isolat sauvage.
  
- La structure obtenue par (Mujeeb et al. 1998) présente également quelques défauts sérieux. En effet, il existe notamment, sur la structure moyenne présentée, une distance entre deux oxygènes chargés négativement appartenant au squelette phosphodiester de chaque brin (atome O1P du résidu A9) de 1,7 Å. Cet éloignement (inférieur à la somme des rayons de Van Der Waals) est très faible et certainement dommageable pour l'énergie électrostatique du système. De plus, il existe un certain nombre de violations de Van Der Waals supérieures à 0,5 Å ce qui se traduit par une énergie totale pour la structure moyenne très faiblement négative (-5 +/- 75 kcal.mol<sup>-1</sup>).

## II-3 Préparation de l'échantillon

### II-3-1 Synthèse de l'oligonucléotide

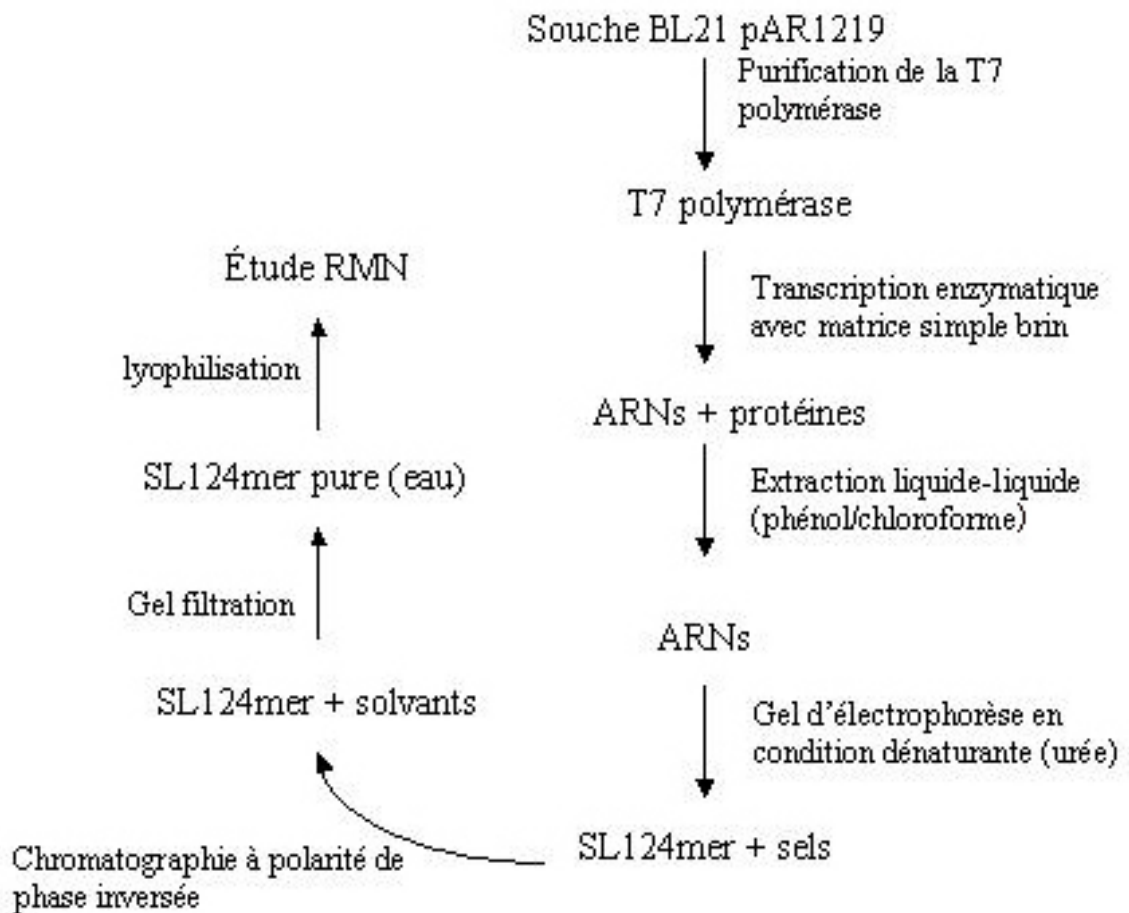
Comme nous l'avons vu précédemment, la modification de l'agencement des appariements dans la tige peut avoir une répercussion sur la stabilité et par conséquent sur la structure de la molécule. Nous avons donc choisi de préserver l'intégrité de la séquence SL123mer. Dès le début de cette étude, nous avons choisi d'établir un marquage <sup>13</sup>C et <sup>15</sup>N de la molécule. Dès lors, la synthèse de l'oligonucléotide ne pouvait se réaliser que par voie biochimique car la synthèse par voie chimique d'un oligonucléotide marqué est toujours excessivement onéreuse. Nous avons choisi une séquence de 24 nucléotides (Figure 41) où l'ordre des appariements est respecté. Nous avons juste rajouté un résidu guanine en 5' et décidé d'inverser les résidus 1 et

23 par rapport à la séquence sauvage. La transcription enzymatique nécessite la présence de deux résidus guanine en 5' pour s'effectuer correctement.



**Figure 41.** A/ Séquence SL124mer de VIH-1Lai. B/ Séquence SL123mer sauvage de VIH-1Lai.

Pour les besoins de la transcription, nous avons préparé l'enzyme T7 polymérase (Davanloo et al. 1984), point crucial pour la synthèse enzymatique (voir Matériels et méthodes). La souche bactérienne nous a été fournie par l'équipe de M. Kochoyan. Nous avons transcrit l'ARN à partir d'une matrice simple brin (Milligan et al. 1987), et l'avons ensuite purifié par extraction liquide-liquide, gel de polyacrylamide, colonne de polarité de phase inversée et gel filtration (Figure 42) (voir Matériels et méthodes). Cette méthodologie de synthèse nous a permis de synthétiser deux échantillons, un premier non marqué avec une concentration de 2mM dans un volume de 300µl, et un deuxième échantillon marqué <sup>13</sup>C et <sup>15</sup>N avec une concentration de 0,6mM également dans 300µl.



**Figure 42.** Protocole d'obtention de la tige boucle SL124mer par voie biochimique.

### II-3-2 Préparation de l'échantillon

La première étape de l'étude structurale consiste à trouver les conditions d'hybridation et de salinité adéquates afin d'éviter d'éventuels mélanges conformationnels.

La suite de l'étude montre que l'on peut obtenir deux conformations isolées à l'aide de deux protocoles de préparation différents. Ces conformations sont stables dans des conditions expérimentales différentes. La conformation, qui transite le plus facilement par un refroidissement rapide, est le complexe instable (formé par cinétique rapide). L'autre forme, obtenue par refroidissement lent, peut être également considérée comme une forme cinétique (formation par cinétique lente). Nous la nommerons le complexe stable. Nous n'avons pas employé les termes de dimère cinétique et de dimère thermodynamique car ils peuvent prêter

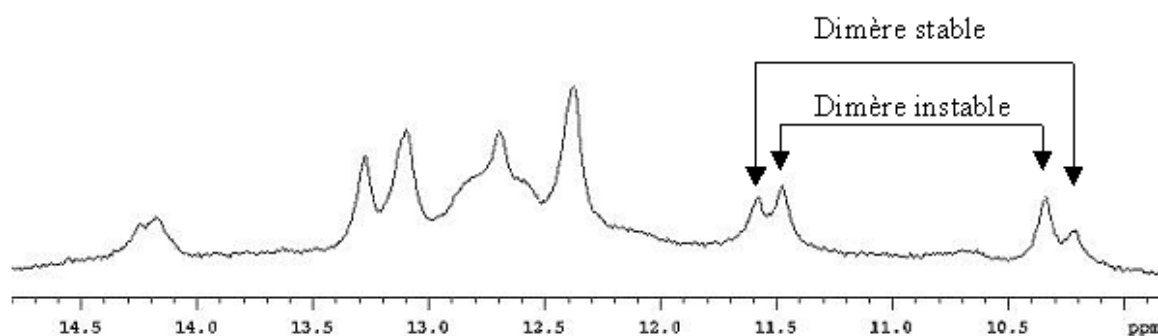
à confusion. Dans la suite du manuscrit, nous utiliserons comme vocabulaire pour ces deux conformations, le terme de dimère stable et de dimère instable.

### II-3-2-a. Sélection de la forme stable et instable

Pour l'hybridation du brin d'ARN SL1 24mer nous avons suivi les protocoles décrits dans la littérature pour l'étude de la forme la moins stable (Theilleux-Delalande et al. 2000). Le protocole comprend une phase de dénaturation à 90°C pendant 3 minutes en absence de sels. La sélection de l'une ou l'autre forme repose en grande partie sur les conditions de refroidissement. Plus le refroidissement est rapide de 90 à - 20°C, plus on favorisera la formation de la forme instable. En effet, par ce biais on isole la conformation qui se forme le plus rapidement. A l'inverse, si le refroidissement est lent, la molécule a le temps de se réarranger et d'adopter la conformation la plus stable (Figure 44).

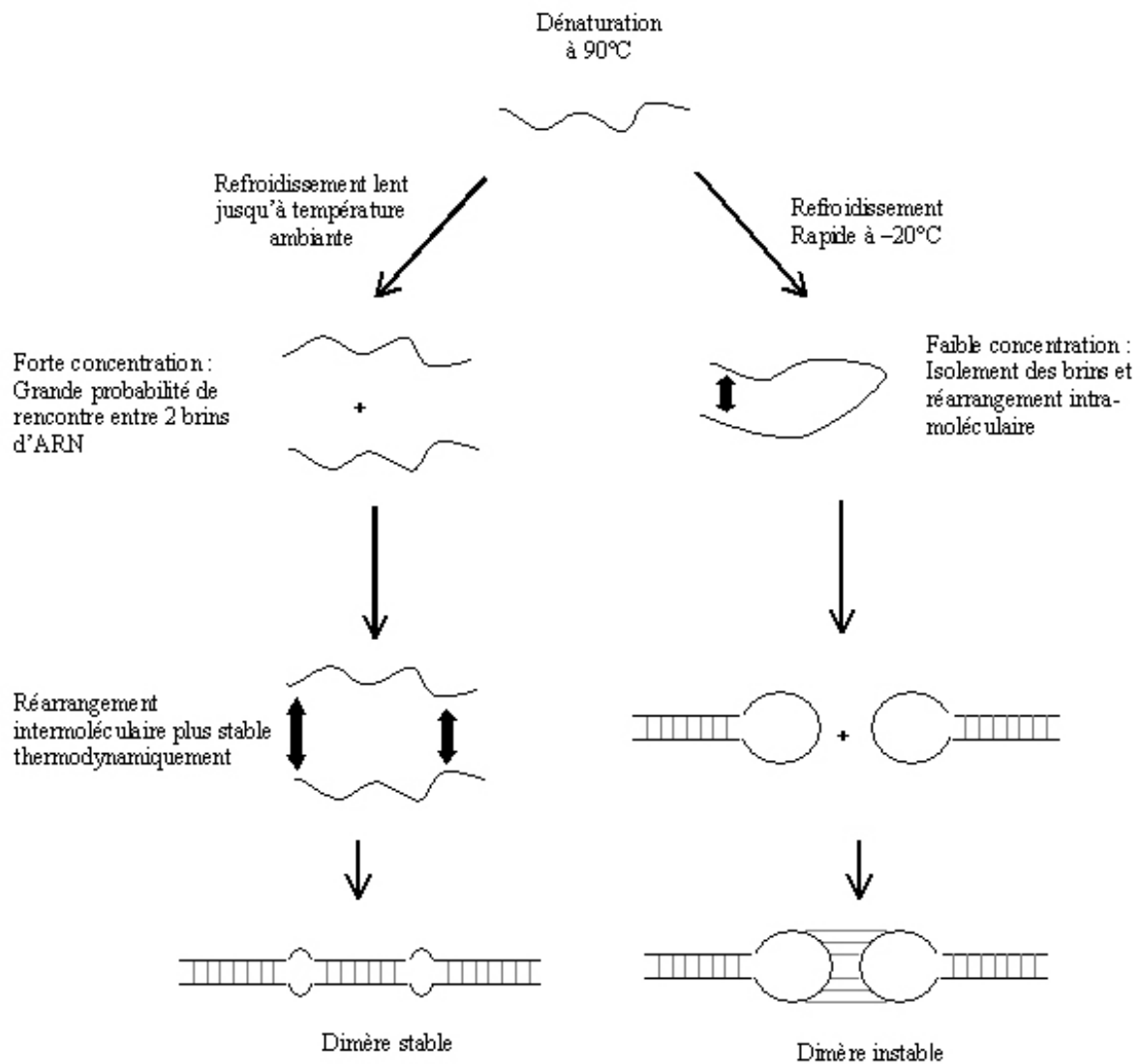
Si on applique le processus ci-dessus sans autre précaution pour isoler la forme instable sur l'échantillon à 2mM, on obtient un mélange entre plusieurs formes (Figure 43). Un autre facteur intervient de façon prépondérante : la concentration en soluté. En effet, si la concentration en soluté est trop importante, on favorise les interactions intermoléculaires qui sont prépondérantes dans le dimère stable. En diminuant cette concentration, on diminue la probabilité de rencontre entre deux brins d'ARN et on favorise alors leur repliement en épingle à cheveux. Pour isoler la forme instable, nous avons travaillé avec une concentration en soluté < 10µM lors de l'étape de refroidissement.

De plus, lors de l'hybridation, il convient d'avoir un échantillon parfaitement dessalé. Les sels comme NaCl ou MgCl<sub>2</sub> masquent les charges portées par les groupements phosphate. Ils stabilisent les structures et par conséquent favorisent la formation de la forme stable, ce qui va se traduire par la présence d'un mélange de formes lors du protocole d'hybridation.



**Figure 43.** Spectre 1D dans H<sub>2</sub>O du brin SL1 24mer sous forme de mélange de forme instable et forme stable (Résonance à 10,2ppm de G18 sous forme d'un doublet).

Les conditions d'hybridation sont reprises ci-dessous (Figure 44)



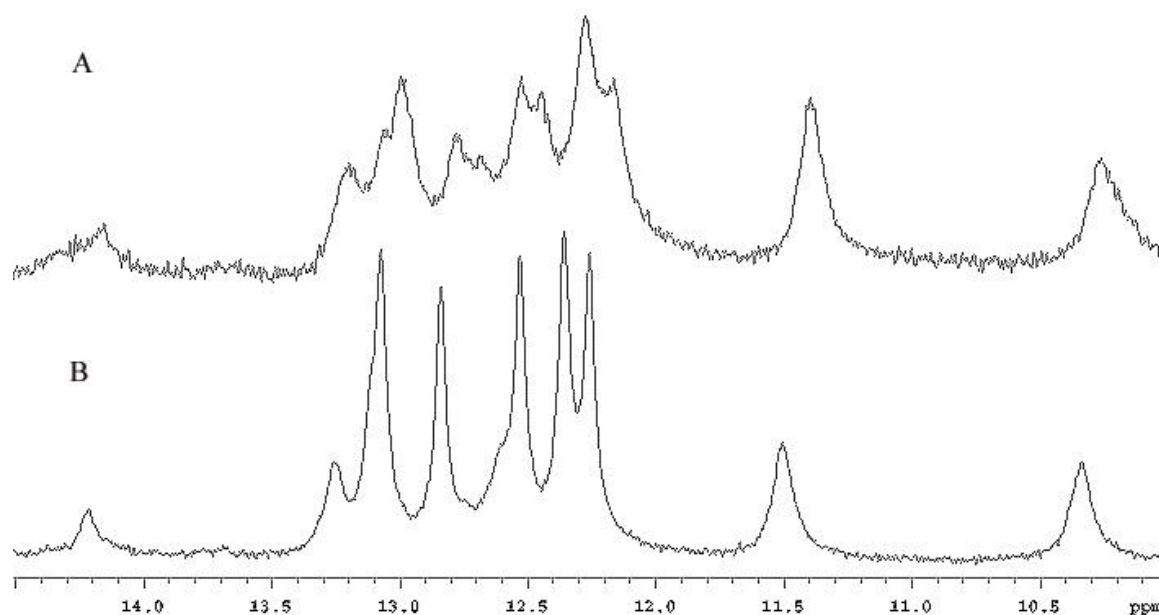
**Figure 44.** Représentation schématique de la formation *in situ* des deux dimères.

Après lyophilisation, l'ARN est repris dans l'eau ou dans l'eau lourde pour obtenir un volume total de 300µl, volume minimal pour l'utilisation de tube shigemi 5mm. Le pH est ajusté à 5,7 avec NaOH ou HCl dilué. Lors de l'étude de la forme instable, toutes les étapes doivent s'effectuer à basse température (entre 10 et 20°C) pour garder l'échantillon sous sa forme instable.

### II-3-2-b. Ajout de sels

Les études menées sur la forme instable par cristallographie indiquent le rôle déterminant du magnésium. Il semble que sans présence de magnésium la cristallisation de la forme instable ne puisse se faire. Les concentrations en magnésium utilisées pour l'étude cristallographique sont de 100mM. Lors des travaux menés par F. Girard (Theilleux-Delalande et al. 2000) sur le mécanisme de transition, la forme instable a pu être isolée en l'absence d'ion magnésium. L'ajout du magnésium (2mM) entraîne un élargissement des raies de résonance et une précipitation partielle de l'échantillon.

Dans le cadre de notre étude du brin SL1 24mer, nous avons pu également isoler la forme instable en l'absence de sel. La forme "instable" s'est avérée stable jusqu'à des températures de 20°C permettant son étude par RMN. Nous avons essayé de stabiliser la forme instable par l'ajout de NaCl. Cet ajout se traduit par un élargissement considérable de la largeur de raie (Figure 45) et facilite la transition vers la forme stable à une concentration de 1mM.

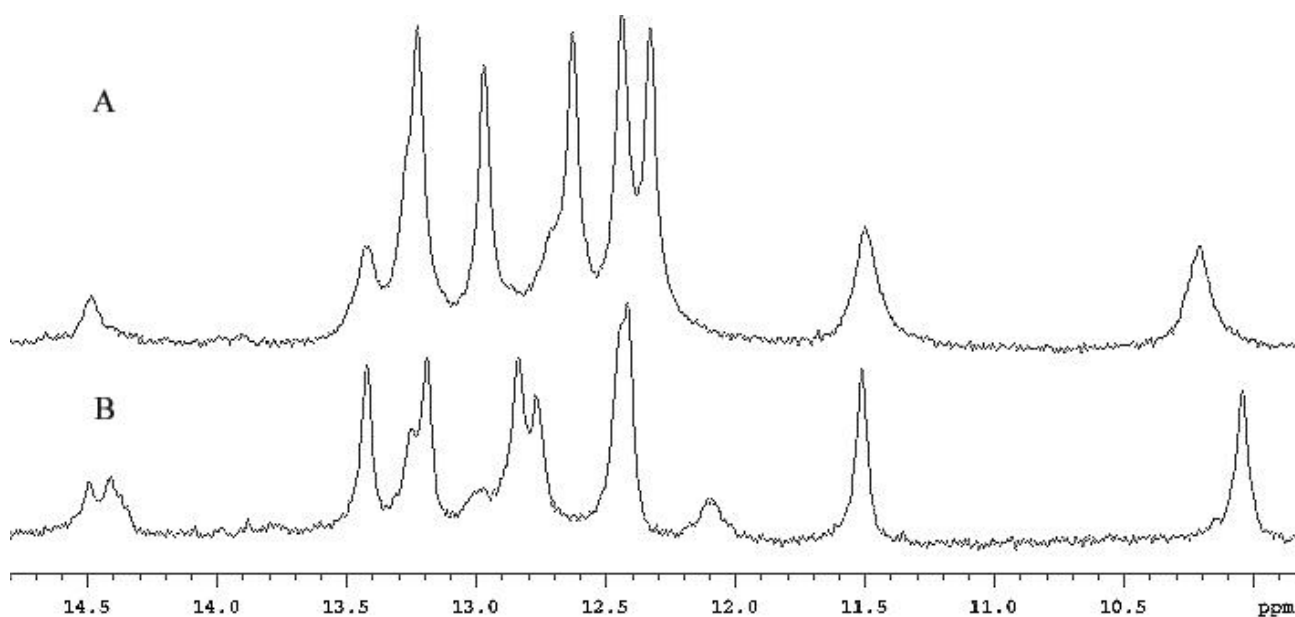


**Figure 45.** Spectre 1D dans H<sub>2</sub>O de la région imino de SL1 24mer de VIH-1 Lai à 10°C pH 5,7. A/ Spectre de la forme instable avec 20mM NaCl. B/ Spectre de la forme instable sans sels.

Nous avons également essayé de stabiliser la forme instable par l'ajout de  $MgCl_2$ . Mais après quelques jours la forme instable parvient à transiter vers la forme stable ce qui nécessite une nouvelle hybridation de l'échantillon. Cette hybridation nécessite alors une élimination complète du magnésium sous peine de dégradation de l'échantillon. Malheureusement, cette élimination est difficile à obtenir même en effectuant des dialyses répétées avec une grande concentration en sel.

### II-3-2-c. Caractérisation spectrale des deux formes de dimères

Le protocole décrit ci-dessus permet d'isoler chaque dimère indépendamment l'un de l'autre. Leurs spectres RMN, particulièrement dans la région des protons imino (Figure 46), sont différents et permettent leur distinction très rapidement. Ces spectres présentent des raies fines et bien résolues.



**Figure 46.** Spectre 1D de la région imino de SL1 24mer de VIH-1 Lai à 10°C pH 5,7. **A/** Spectre de la forme instable. **B/** Spectre de la forme stable.

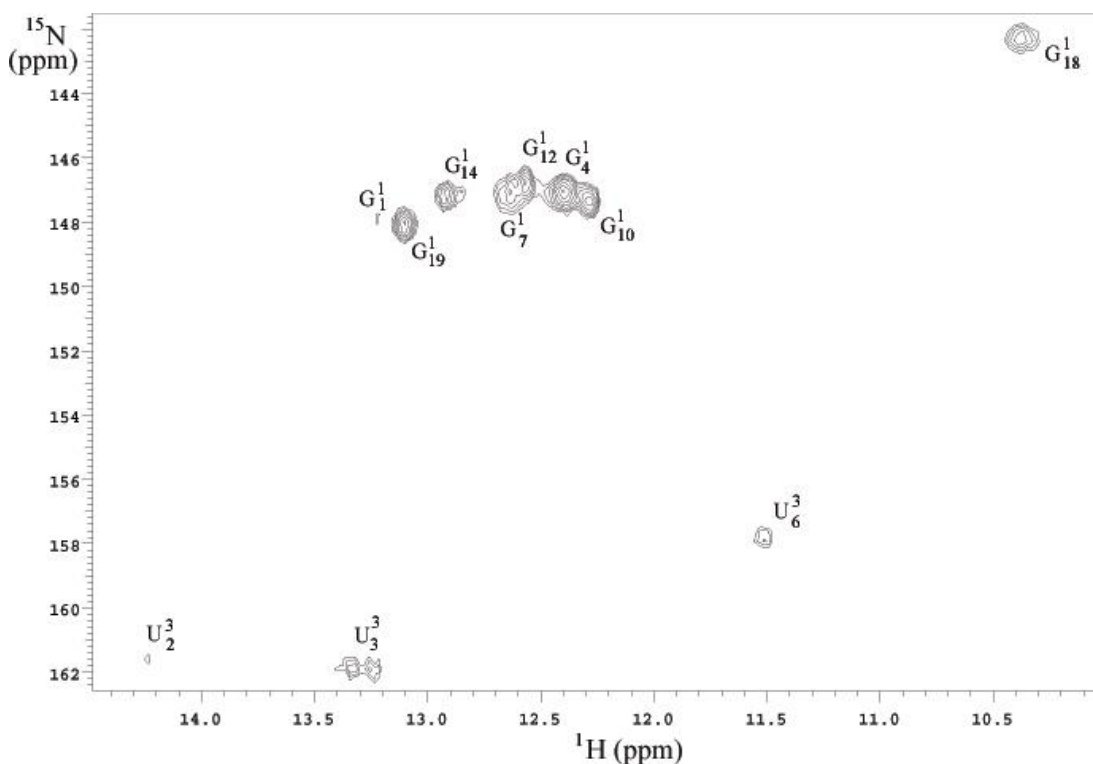
## II-4 Etude RMN

Selon les spectres envisagés, l'échantillon non marqué ou marqué  $^{15}\text{N}$  et  $^{13}\text{C}$  est dissout soit dans un mélange  $\text{H}_2\text{O}:\text{D}_2\text{O}$  (90 :10), soit dans  $\text{D}_2\text{O}$ . L'échantillon est ensuite mis dans un tube shigemi<sup>®</sup> de 5mm dans un volume total de 300 $\mu\text{l}$ . Le pH de la solution pour l'étude structurale dans l'eau légère est de 5,7 et de 6,2 dans l'eau lourde.

### II-4-1 Etude dans l'eau légère

#### II-4-1-a. Les spectres HSQC $^{15}\text{N}$

Les spectres HSQC  $^{15}\text{N}$  permettent d'observer les résonances des  $^{15}\text{N}$  impliquées dans les groupements amino et imino(Figure 47). Les résonances des  $^{15}\text{N}$  des groupements imino et amino des cytosines ont pu être attribuées avec certitude. Le manque de résolution dû à la relaxation importante de la forme instable rend l'attribution des groupements amino des adénines et des guanines très incertaine et pratiquement impossible.



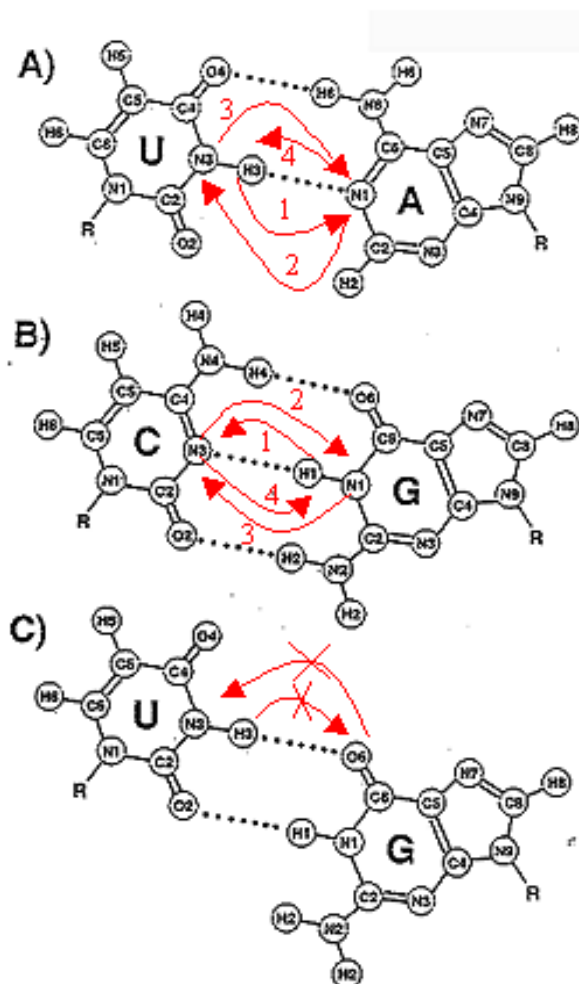
**Figure 47.** Spectre HSQC  $^{15}\text{N}$  de la forme instable du brin SL1 24mer. Région des connectivités  $^{15}\text{N}1\text{-H}$  (guanines) et  $^{15}\text{N}3\text{-H}$  (uridine).



## II-4-1-b. Spectres HNN-COSY

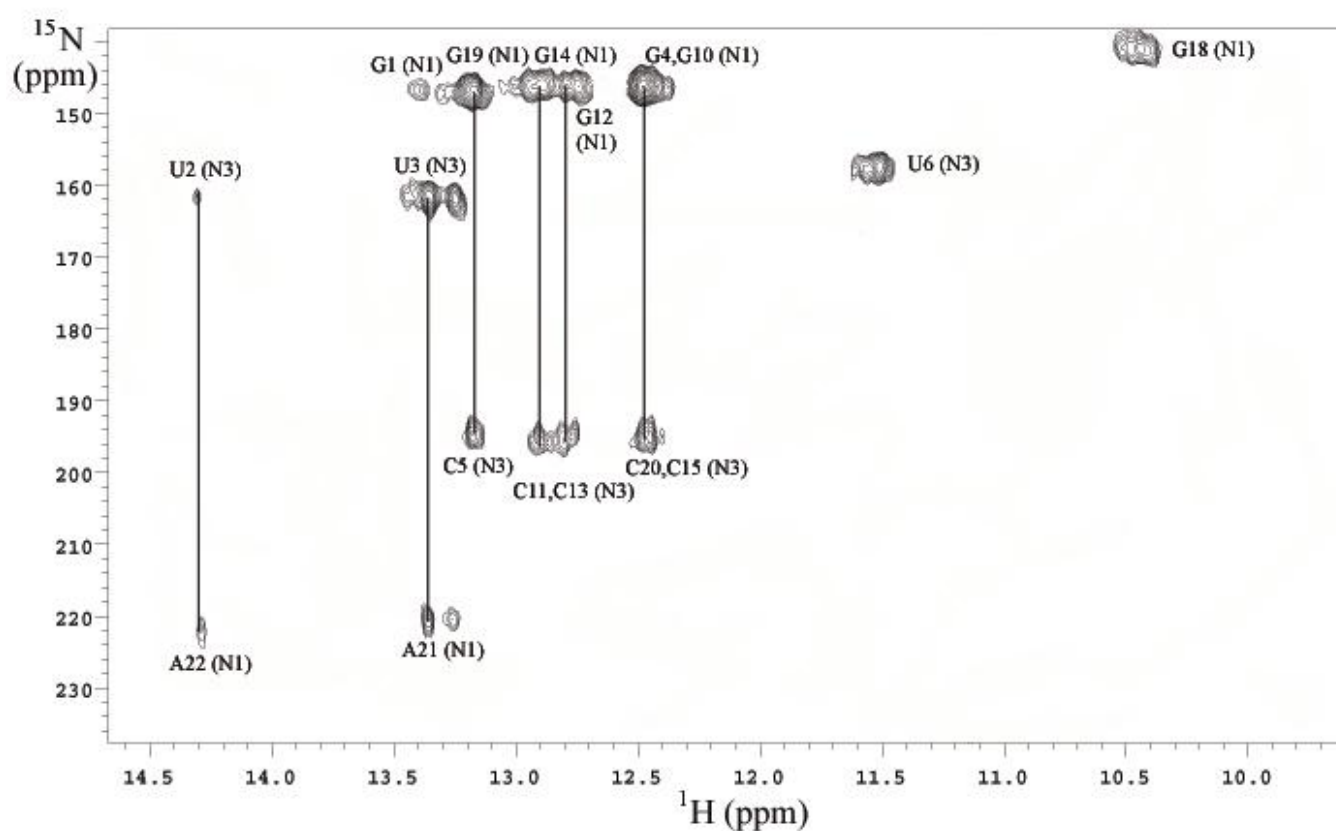
Nous avons également effectué des spectres HNN-COSY (Dingley and Grzesiek 1998; Majumbar et al. 2001) qui permettent d'observer directement les protons engagés dans une liaison hydrogène de type Watson-Crick via la constante de couplage entre le proton imino des guanines et des uraciles avec l'azote accepteur de proton, N3 des cytosines et N1 des adénines.

Cette séquence transfère la polarisation du proton NH sur le  $^{15}\text{N}$  de l'autre base appariée (étape 1 de la figure 48) ( $^1J_{\text{H-}^{15}\text{N}} \# 14\text{Hz}$ ). Puis, on retransfère cette polarisation sur l'azote de l'autre base appariée via le couplage scalaire entre les deux azotes ( $^2J^{15}\text{N}^{15}\text{N} \# 7\text{Hz}$ ) (étape 2 de la figure 48), on étiquette cet azote et on retransfère la polarisation sur le proton pour que celui-ci soit détectée en effectuant le chemin inverse (étape 3 et 4 de la figure 48).



**Figure 48.** Transfert de polarisation lors d'une expérience HNN-COSY. La numérotation indique les différentes étapes du transfert.

On peut mettre en évidence les deux paires AU de la séquence où le proton imino de l'uracile est couplé avec deux azotes différents, preuve qu'il est engagé dans une liaison hydrogène. On peut également remarquer que pour la "woobble" paire U-G, le transfert de polarisation entre le proton imino de l'uracile et l'azote accepteur n'est pas possible car la liaison hydrogène se forme avec un oxygène de la guanine (Figure 48 et 49). Nous avons préféré présenter ici un spectre réalisé sur la forme stable de la tige boucle SL1 24mer à 20°C. Le spectre HNN-COSY a été également effectué sur la forme instable donnant les mêmes résultats que la forme stable mais sa qualité est moindre par rapport à celui présenté.

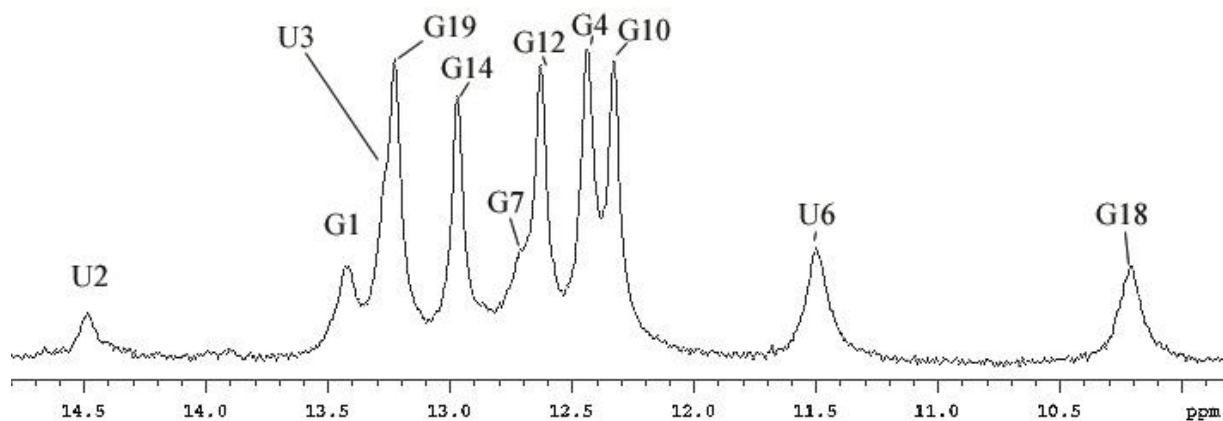


**Figure 49.** Spectre HNN-COSY de la tige boucle SL1 24mer sous sa forme stable à 20°C dans H<sub>2</sub>O/D<sub>2</sub>O (90/10).

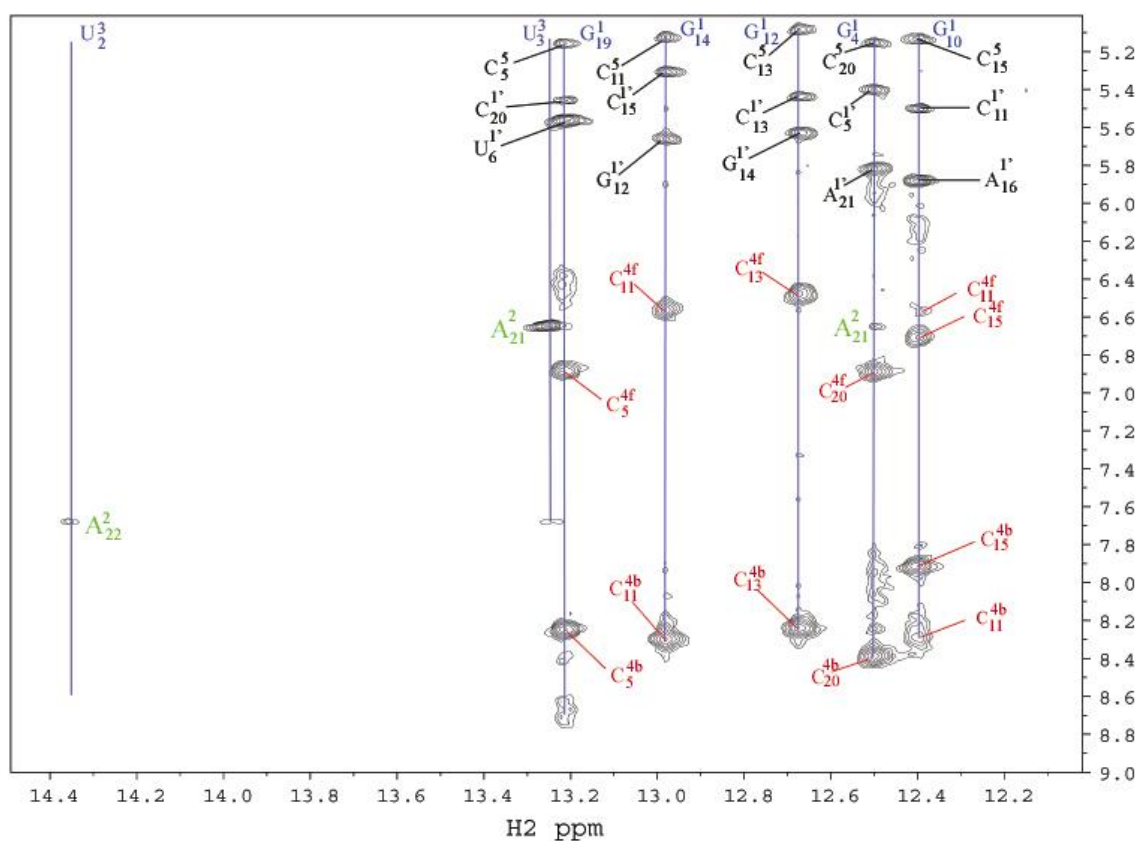
#### II-4-1-c. Attribution des protons échangeables

Les spectres dans l'eau (Figure 50) ont été obtenus en utilisant une séquence WATERGATE (Piotto et al. 1992) pour la suppression du solvant. L'attribution des protons imino et amino (Figure 51) a pu être réalisée en utilisant des spectres NOESY à 80 ms et des spectres HSQC <sup>15</sup>N (Figure 47). A partir des connectivités H2 des adénines A21 et A22, nous avons pu

attribuer les résonances des protons échangeables, imino et amino en utilisant la même méthodologie que celle utilisée dans le cadre de l'étude de la forme monomérique (cf I-2-1).



**Figure 50.** Spectre 1D de la forme instable, région des protons imino à 10°C pH 5,7.



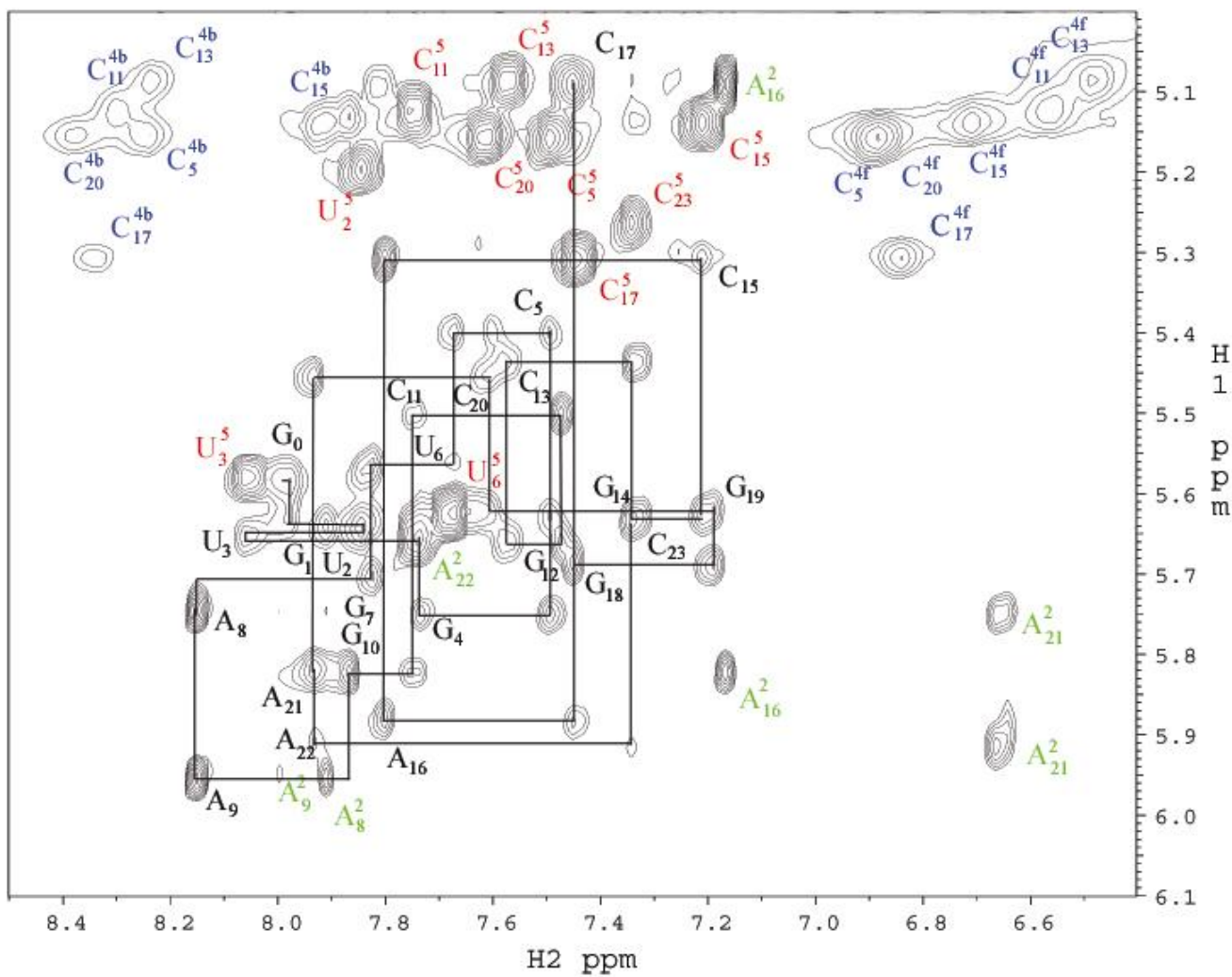
**Figure 51.** Spectre NOESY dans H<sub>2</sub>O/D<sub>2</sub>O (90/10) à 10°C pH=5.7 Connectivités entre protons imino et protons amino et aromatiques, temps de mélange 80 ms.

#### II-4-1-d. Attribution des protons non échangeables

Un spectre NOESY ( $\tau_m = 80\text{ms}$ ) de la forme instable permet de faire aisément l'attribution séquentielle des protons H1' - aromatique sans passer dans une solution dans D<sub>2</sub>O (Figure 52). Nous avons utilisé comme point de départ les H2 des A21 et A22. Les incertitudes dans l'attribution des signaux ont pu être levées, nous le verrons par l'utilisation de spectres hétéro-nucléaires sur l'échantillon marqué <sup>13</sup>C et <sup>15</sup>N. Ces incertitudes concernent principalement l'attribution des résonances des H2 des adénines situées à la jonction de la tige avec la boucle. Ces protons résonnent à la même fréquence que les protons H6 et H8 des autres résidus.

#### II-4-2 Etude dans l'eau lourde

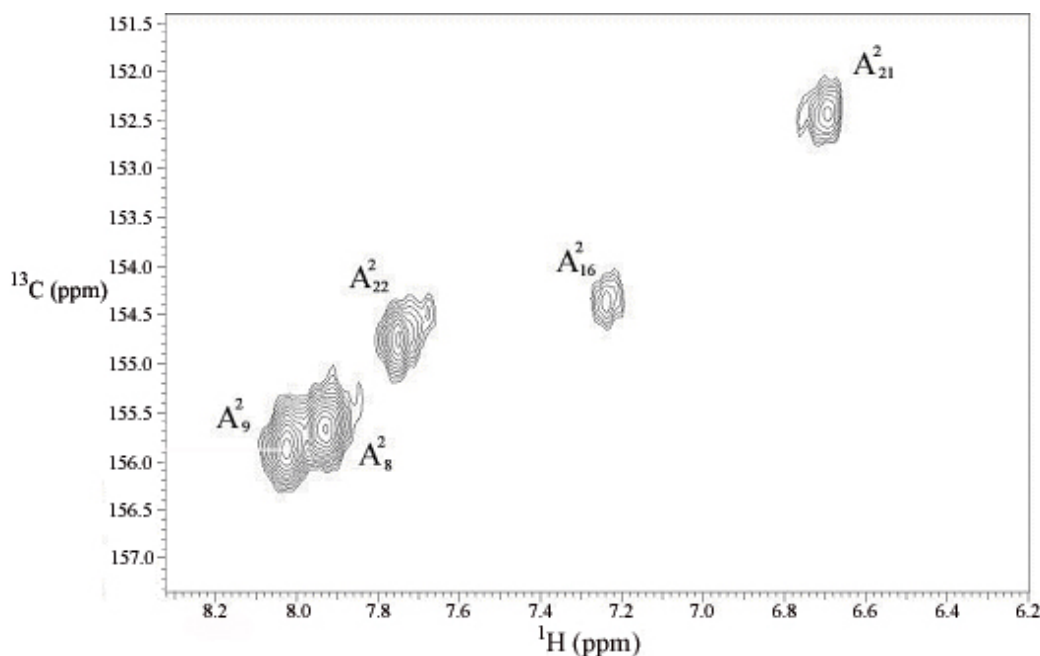
Nous avons effectué l'étude de la forme instable de la tige boucle SL1 24mer dans D<sub>2</sub>O. La qualité des spectres lors de ce passage s'est nettement dégradée par rapport à l'étude dans H<sub>2</sub>O. En effet, nous étudions une forme instable qui est stabilisée par un réseau de liaisons hydrogène. Le passage entre H<sub>2</sub>O et D<sub>2</sub>O induit un remplacement des hydrogènes impliqués dans les liaisons hydrogène par des deutériums. Les liaisons deutérium sont énergétiquement plus faibles que les liaisons hydrogène ce qui induit une stabilisation plus faible de la forme instable dans D<sub>2</sub>O. Ce déficit énergétique provoque un échange plus important entre les différentes conformations et se traduit par un élargissement des pics à mi-hauteur et une perte de résolution sur les spectres dans D<sub>2</sub>O (Figure 53 et 54). Nous avons néanmoins pu réaliser certaines expériences qui nous ont permis de vérifier les informations recueillies dans H<sub>2</sub>O.



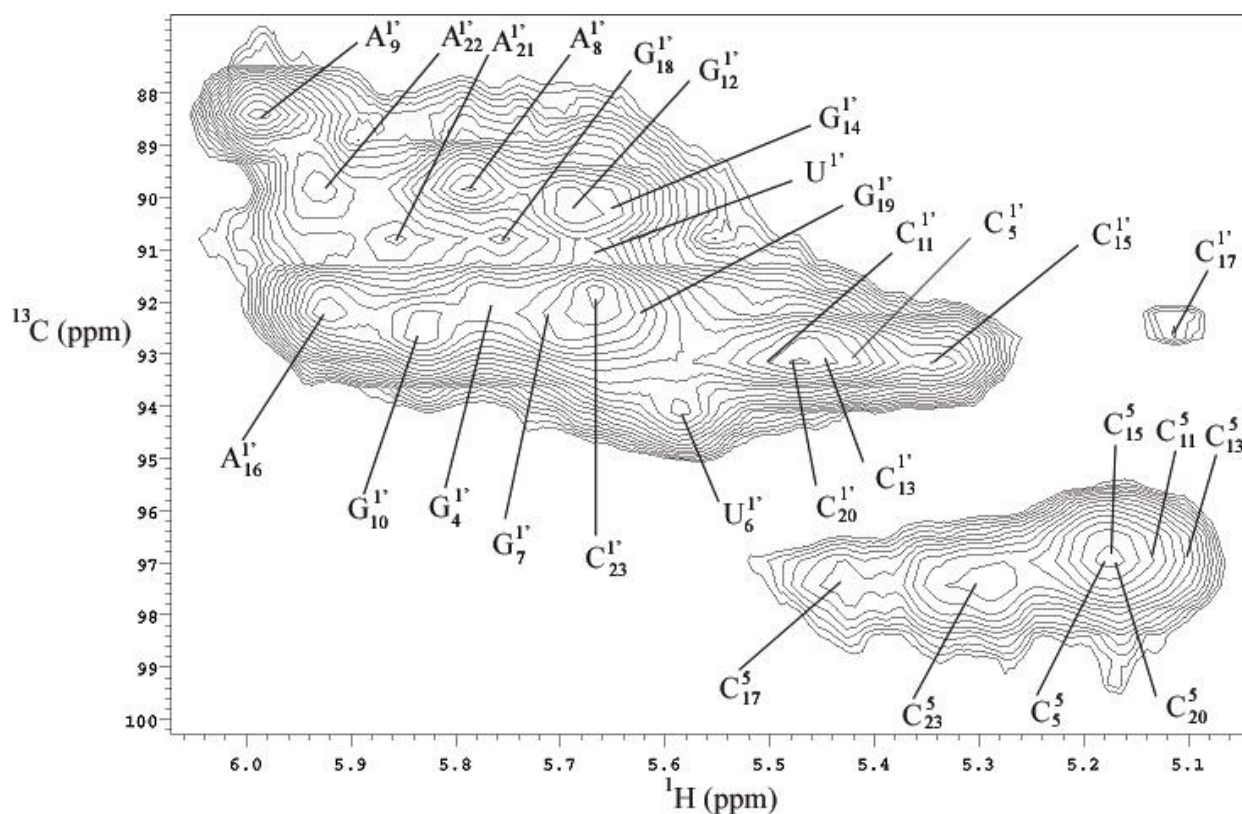
**Figure 52.** Spectre NOESY de la forme instable dans H<sub>2</sub>O à 10°C avec un temps de mélange de 80ms. Région des connectivités aromatiques/H1<sup>?</sup>. La position des protons H2 est indiquée en vert, celle des H5 en rouge et celle des protons iminos en bleu.

#### II-4-2-a. Spectre HSQC $^{13}\text{C}$

La principale difficulté qui se présente lors de la résolution de l'attribution séquentielle des protons H1'-aromatiques (Figure 52) est d'identifier la résonance des protons H2 et H8 des adénines de l'oligonucléotide. Pour connaître leur position, il est possible d'effectuer des expériences de relaxation T1 (cf I-2-2-a). Les H2 des adénines possèdent des temps de relaxation en général plus longs que les autres protons de l'oligonucléotide. Il existe dans le cas de la forme instable du dimère SL1 24mer trois adénines en entrée de boucle. Il est possible que ces trois adénines soient engagées dans une structure particulière puisque le temps de relaxation de leur H2 est plus court qu'en règle générale. Pour s'affranchir de cette difficulté, nous avons effectué des spectres HSQC  $^{13}\text{C}$  (Figure 53) de la tige boucle SL1 24mer. Les carbones C2 des adénines possèdent une gamme de résonance particulière (de 152 à 156 ppm) par rapport aux autres carbones aromatiques C6 et C8 (de 136 à 144 ppm) (Wijmenga et al., 1998). Nous avons ainsi pu mettre en évidence sans le moindre doute les cinq H2 des différentes adénines de l'oligonucléotide. Le spectre HSQC de la région des C1'-H1' et C5-H5, nous a permis de mettre en évidence une résonance d'un proton H1' déplacé vers les hauts champs (par rapport aux autres résonances). Cette résonance a été attribuée au C1'-H1' du résidu C17 (Figure 54).



**Figure 53.** Spectre HSQC  $^{13}\text{C}$  de la forme instable du brin SL1 24mer, région des connectivités  $^{13}\text{C}2$ -H2 dans  $\text{D}_2\text{O}$  à  $20^\circ\text{C}$ .



**Figure 54.** Spectre HSQC  $^{13}\text{C}$  de la forme instable du brin SL1 24mer, région des connectivités  $^{13}\text{C}5\text{-H}5$  et  $^{13}\text{C}1'\text{-H}1'$ .

#### II-4-2-b. Attribution des protons H2' et H3'

La conformation majoritaire des riboses de l'oligonucléotide sous sa forme instable est de type nord. Cette conformation ne permet pas l'utilisation d'expériences COSY pour attribuer les protons H2' à partir du H1' car  $J_{(\text{H}1'\text{-H}2')} \neq 0$ . L'attribution des protons H2' a donc été effectuée grâce à des spectres HCCH qui permettent le passage à travers les liaisons du proton H1' du sucre vers le proton H2' via les carbones C1' et C2'. Cette expérience permet d'obtenir directement la position des différents H2'. Ensuite par comparaison avec un spectre NOESY à 50 ms dans D2O, nous avons pu trouver par différence les résonances des différents protons H3', étant donné que la distance entre le proton aromatique du résidu n et le proton H3' du résidu n-1 est faible ( $\sim 3,4\text{\AA}$ ).

résidu	NH	N1/N3 (imino)	NH2 (amino)	N2/6/4 (amino)	H6/H8	C6/C8
G0					8,02	
G1	13,27	147,88			8,00	
U2	14,22	161,60			7,87	143,02
U3	13,24	162			8,06	142,45
G4	12,36	147,12			7,75	136,50
C5			6,88/8,25	82,1	7,52	140,24
U6	11,51	157,95			7,71	141,44
G7	12,6	147,17			7,85	138,29
A8					8,17	140,64
A9					8,18	142,00
G10	12,27	147,53			7,9	138,23
C11			6,56/8,30	80,7	7,77	141,41
G12	12,53	146,81			7,5	136,02
C13			6,49/8,24	82,6	7,6	140,44
G14	12,84	147,32			7,35	135,64
C15			6,71/7,92	80,9	7,24	140,24
A16					7,82	138,88
C17			6,86/8,36	81,6	7,47	140,53
G18	10,36	142,54			7,47	137,21
G19	13,07	148,24			7,24	136,32
C20			6,89/8,38	81,1	7,63	141,00
A21					7,95	139,90
A22					7,91	139,71
C23			7,00/8,26	81,2	7,37	141,24

résidu	H5/H2	C5/C2	H1'	C1'	H2'	H3'
G0			5,62			
G1			5,68			
U2	5,26	103,06	5,69	91,20	4,46	4,5
U3	5,62	103,88	5,69	91,20	4,52	4,57
G4			5,78	92,20	4,46	4,53
C5	5,19	97	5,44	93,20	4,53	4,35
U6	5,63	104,87	5,60	94,00	4,12	4,46
G7			5,74	92,30	4,14	4,35
A8	7,95	155,65	5,79	89,86	4,48	4,3
A9	8,02	155,84	6,00	88,53	4,88	4,54
G10			5,87	92,60	4,84	4,01
C11	5,16	97	5,53	93,20	4,43	4,54
G12			5,69	90,40	4,31	4,4
C13	5,13	97	5,53	93,20	4,39	4,46
G14			5,66	90,26	4,40	4,36
C15	5,18	97	5,35	93,20	4,43	4,31
A16	7,19	154,33	5,90	92,26	4,56	4,39
C17	5,34	97,54	5,15	92,48	4,45	4,31
G18			5,73	90,86	4,44	4,34
G19			5,66	92,26	4,44	4,52
C20	5,24	97	5,49	93,20	4,55	4,52
A21	6,7	152,43	5,86	91,00	4,51	4,55
A22	7,78	154,76	5,93	89,87	4,28	4,43
C23	5,3	97,54	5,68	92,00	4,11	3,93

**Tableau 55.** Déplacements chimiques de la tige boucle SL1 24mer sous sa forme instable. Les protons échangeables sont attribués à pH 5,7 dans H<sub>2</sub>O:D<sub>2</sub>O (90:10) à 10°C et les non-échangeables à 20°C dans D<sub>2</sub>O, pH 6,2.

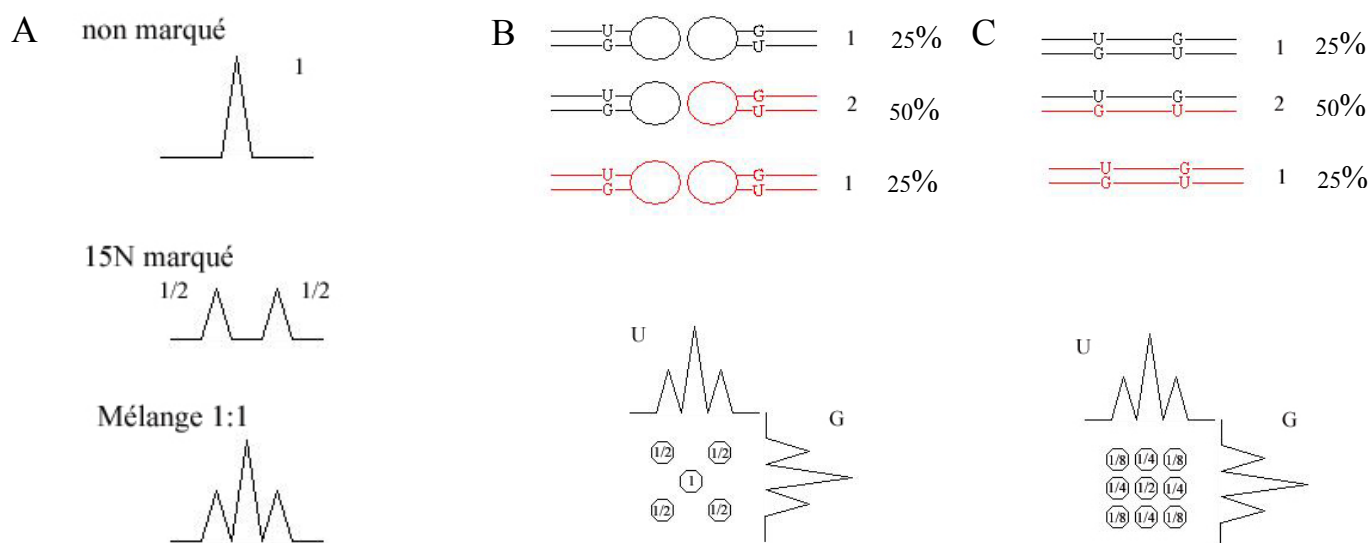


## II-4-2-c. Le dimère instable est un dimère boucle-boucle

### II-4-2-c-a Spectre NOESY non irradié d'un mélange marqué- non marqué (1/1)

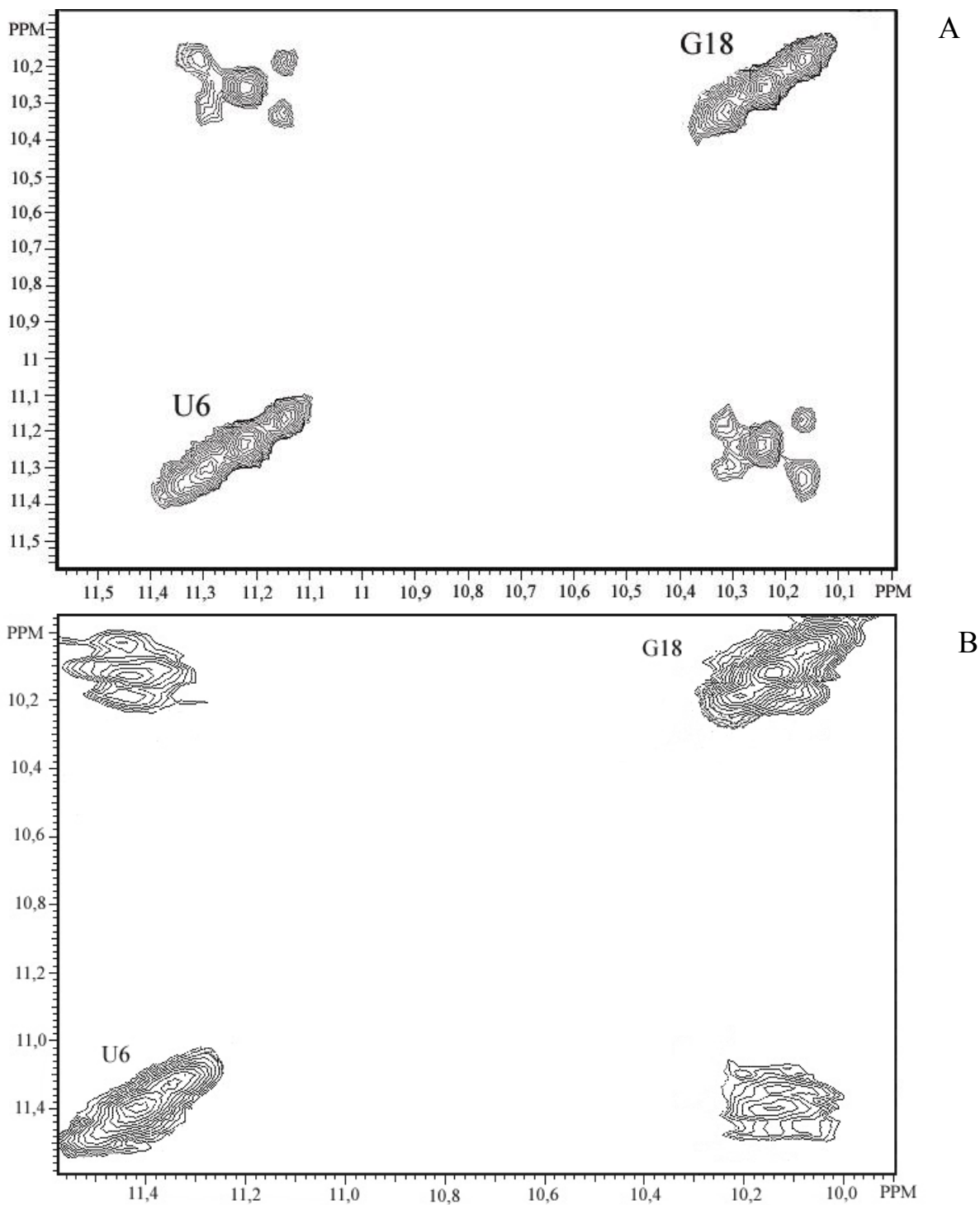
Nous avons pu mettre en évidence la nature dimérique des différentes formes. Nous avons ensuite cherché à savoir si le dimère instable était bien le résultat d'une interaction entre deux tiges boucles. Pour cela, nous avons suivi deux méthodologies pour différencier une tige formée de manière intra- ou inter-moléculaire, une première approche basée sur les travaux de (Takahashi et al. 2000a) et une deuxième basée sur des expériences demi filtre.

Afin de mettre en évidence un dimère étendu ou un complexe boucle-boucle, nous avons utilisé la structure fine de la connectivité  $N_3H(U6) - N_1H(G18)$  dans un mélange marqué ( $^{15}N, ^{13}C$ ) et non marqué dans des proportions 1 :1. Le signal du proton  $^{15}NH$  apparaît sous la forme d'un doublet alors que celui d'un  $^{14}NH$  sort sous la forme d'un singlet (Figure 56A). Les connectivités  $^{14}N_3H(U6) - ^{15}N_1H(G18)$  n'existent que pour les dimère étendu (Figure 56B et C) ce qui permet la discrimination entre les deux types de dimères.



**Figure 56.** Schéma de la structure fine des connectivités NOESY NH-NH de la paire UG, au sein d'un mélange équimolaire de tiges boucles marquées  $^{15}N$  et non marquées (Takahashi et al. 2000a). En noir, L'ARN non marqué et en rouge l'ARN marqué. A/ schéma de la structure fine pour une connectivité NH sur un spectre 1D. B/ Cas d'un complexe boucle-boucle. C/ Cas d'un dimère étendu.

L'étude de la structure fine des connectivités sur les deux formes de dimères indique clairement que le dimère préparé dans des conditions de cinétique rapide résulte de l'interaction de 2 tiges boucles (Figure 57).



**Figure 57.** Spectre NOESY d'un mélange d'un brin SL1 24mer marqué  $^{15}\text{N}$  et  $^{13}\text{C}$  et d'un brin SL1 24mer non marqué. A/ forme instable. B/ forme stable.

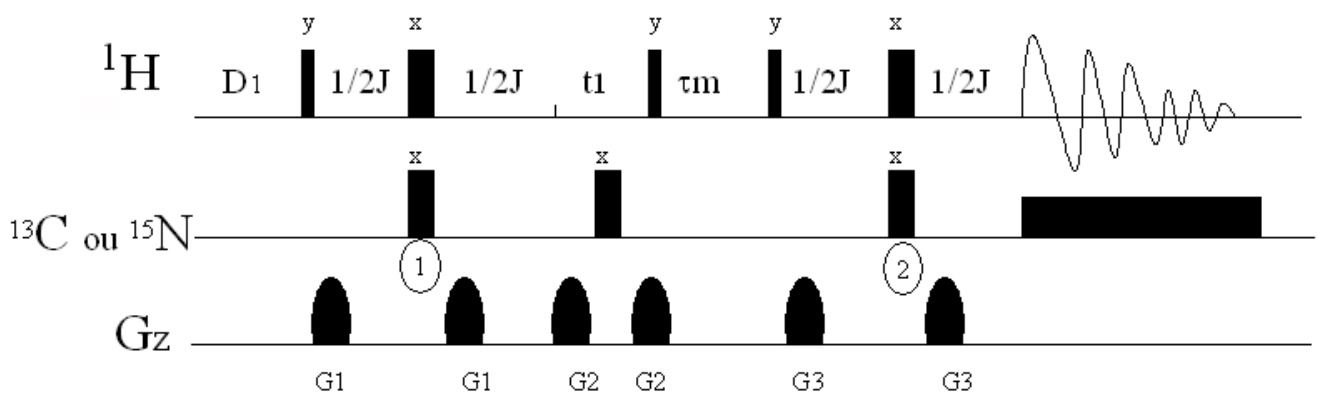
#### *II-4-2-c-b Expérience de demi-filtre*

Les expériences demi-filtre permettent la séparation des résonances des protons liés à un hétéroatome X de tous les autres protons. Par exemple, un spectre NOESY utilisant des demi-

filtres  $^{13}\text{C}/^{15}\text{N}$  effectué sur un complexe nucléo-protéique, où seule la protéine est marquée  $^{13}\text{C}/^{15}\text{N}$ , permet d'obtenir dans un sous spectre l'ensemble des connectivités NOESY et dans un autre sous-spectre les connectivités de la protéine seule. Nous nous sommes intéressés de la même façon que précédemment à la paire de bases non canonique U6-G18 présente dans la tige dans un mélange marqué ( $^{15}\text{N}, ^{13}\text{C}$ ) et non marqué dans des proportions 1 :1.

Nous avons mis au point une séquence (Figure 58) (Otting and Wuthrich 1989) qui permet de sélectionner les cohérences appartenant à  $\text{H}^{13}\text{C}$  ou  $\text{H}^{12}\text{C}$  pour chaque type de transfert entre molécule marquée et molécule non marquée à savoir : molécule marquée vers molécule marquée, molécule marquée vers molécule non marquée, molécule non marquée vers molécule marquée et molécule non marquée vers molécule non marquée.

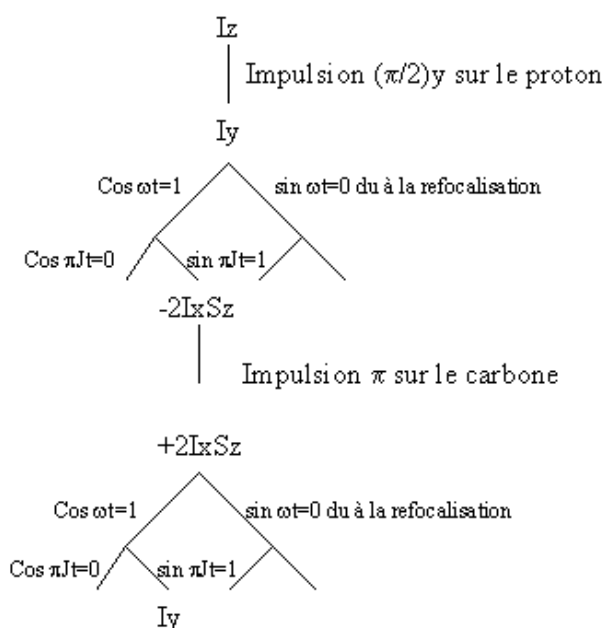
Pour la mise au point des expériences demi-filtres, nous avons utilisé un oligonucléotide ADN  $\text{d}(\text{CGCAAATTTGCG})_2$  dont les 3 thymines sont marqués  $^{13}\text{C}$  sur le carbone C1' et C5' (Figure 62).



**Figure 58.** NOESY équipé de deux demi-filtres  $^{13}\text{C}$  ou  $^{15}\text{N}$ .

Les deux impulsions sur le canal de l'hétéro-atome permettent d'obtenir des composantes  $I_y$  pour les différents transferts soit positifs soit négatifs (Figure 59 et tableau 60).

### Pour une liaison $^{13}\text{C-H}$



Si on retire l'impulsion  $\pi$  sur le carbone, l'opérateur  $I_y$  devient  $-I_y$ .  
 Pour une liaison  $^{12}\text{C-H}$ , l'impulsion  $\pi$  du carbone n'a pas d'influence sur l'évolution de l'opérateur densité  $I_y$ .

**Figure 59.** Evolution de l'opérateur densité lors du premier filtre de la séquence demi-filtre.

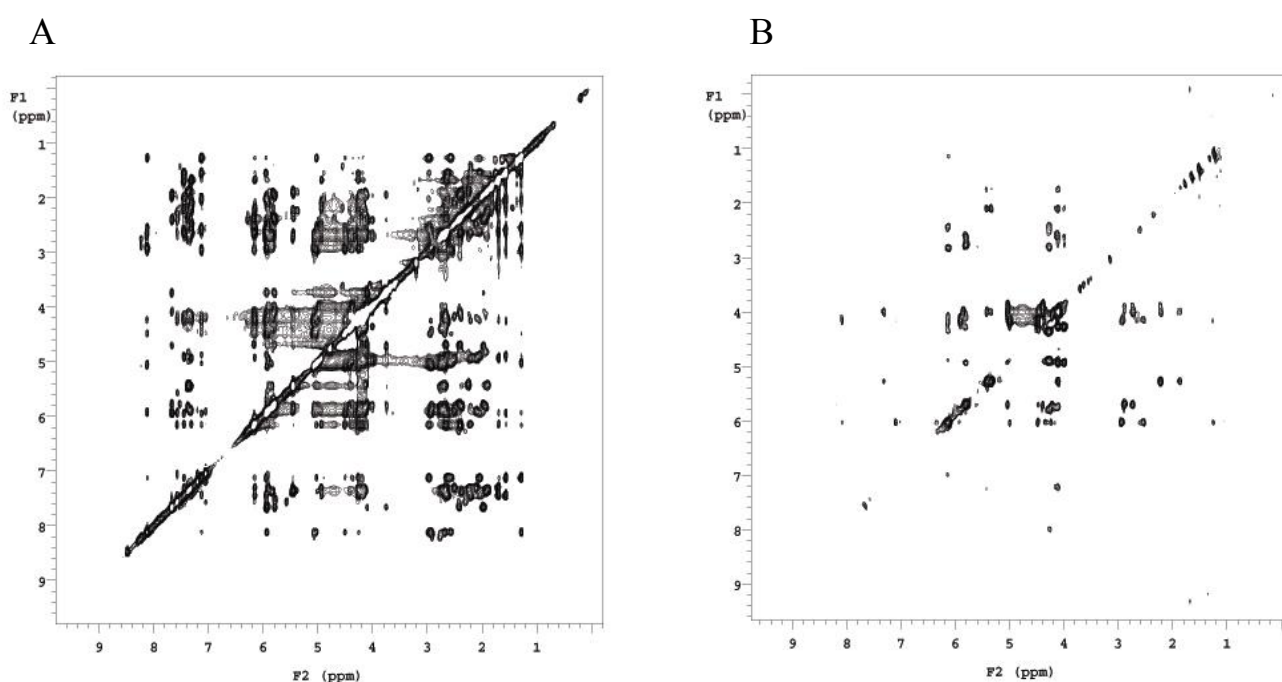
présence des impulsions $180^\circ$ 1 et 2	$\text{H}^{13}\text{C-H}^{13}\text{C}$	$\text{H}^{12}\text{C-H}^{12}\text{C}$	$\text{H}^{13}\text{C-H}^{12}\text{C}$	$\text{H}^{12}\text{C-H}^{13}\text{C}$	
oui /oui	$+I_y$	$+I_y$	$-I_y$	$+I_y$	spectre 1
non/non	$+I_y$	$+I_y$	$+I_y$	$-I_y$	spectre 2
oui/non	$-I_y$	$+I_y$	$-I_y$	$+I_y$	spectre 3
non/oui	$-I_y$	$+I_y$	$+I_y$	$-I_y$	spectre 4

**Tableau 60.** Evolution de la composante  $I_y$  avec la présence ou non des impulsions  $\pi$  sur le carbone lors de la séquence demi-filtre.

En effectuant une combinaison linéaire de 4 expériences, on peut obtenir différents sous-spectres dans lesquels se manifeste un seul type de transfert (tableau 61 et figure 62).

combinaison linéaire des expériences	H <sup>13</sup> C-H <sup>13</sup> C	H <sup>12</sup> C-H <sup>12</sup> C	H <sup>13</sup> C-H <sup>12</sup> C	H <sup>12</sup> C-H <sup>13</sup> C
1+2+3+4	0	4	0	0
1+2-3-4	4	0	0	0
-1+2-3+4	0	0	4	0
-1+2+3-4	0	0	0	4

**Tableau 61.** Résultat de quelques combinaisons linéaires possibles des différents sous spectres.

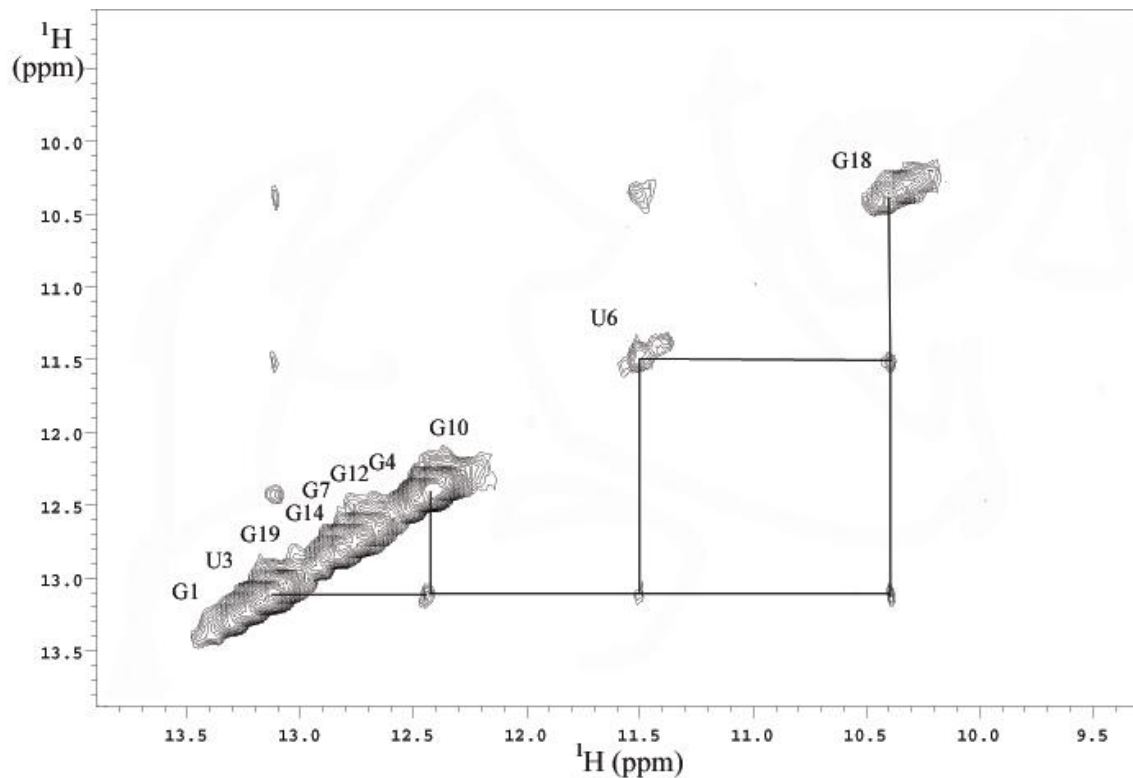


**Figure 62.** Spectre d'un NOESY équipé de deux demi-filtres RMN d'un oligonucléotide ADN d(CGCAAATTTGCG)<sub>2</sub>. A/ sous-spectre <sup>12</sup>C → <sup>12</sup>C. B/ Sous-spectre <sup>12</sup>C → <sup>13</sup>C.

Seuls les carbones C1' et C5' des 3 thymines de l'oligonucléotide ADN sont marqués <sup>13</sup>C. On observe nettement les passages <sup>13</sup>C-H1' vers <sup>13</sup>C-H6/8 et <sup>13</sup>C-H1' vers <sup>13</sup>C-H2' alors que les passages <sup>12</sup>C-H1' vers <sup>12</sup>C-H6/8 et <sup>12</sup>C-H1' vers <sup>12</sup>C-H2' sont absents (Figure 62B).

Après cette mise au point, nous avons réalisé cette expérience de NOESY équipée de deux demi-filtres <sup>15</sup>N sur la forme instable de la tige boucle SL1 24mer. Si le dimère instable est le résultat d'une interaction boucle-boucle, l'appariement entre les deux boucles se fera de façon intermoléculaire alors que l'appariement des bases au niveau de la tige se fera de façon intramoléculaire. Nous pouvons mettre en évidence sur un sous-spectre <sup>14</sup>N vers <sup>14</sup>N les connectivités NH-NH dues aux protons imino de la tige et nous n'observons pas les

connectivités des protons imino de la boucle (Figure 63). La forme instable résulte bien d'une interaction boucle-boucle. De plus, si l'on effectue un sous-spectre  $^{14}\text{N}$  vers  $^{15}\text{N}$ , on observe une disparition de la paire U-G de la tige, preuve supplémentaire que la forme instable est un complexe boucle-boucle. En effet, la paire U-G ne peut se faire que de façon intramoléculaire et par conséquent si nous sélectionnons le passage  $^{14}\text{N}$  vers  $^{15}\text{N}$  cette interaction doit disparaître.



**Figure 63.** Sous- Spectre  $^{14}\text{N} \rightarrow ^{14}\text{N}$  NOESY équipé de deux demi-filtres du dimère instable SL1 24 mer.

Le dimère instable est donc un complexe boucle-boucle appelé 'kissing complex' qui résulte d'un appariement inter-moléculaire de deux monomères. Ces résultats vont à l'encontre des résultats publiés par l'équipe de G. Kawai (Takahashi et al. 2000a; Takahashi et al. 2000b) qui n'avait réussi à isoler qu'une seule forme de dimère (le dimère stable) sur l'extrémité de la tige boucle SL1(23mer). L'équipe de G. Kawai avait néanmoins réussi à isoler deux types de dimères sur la totalité de la tige boucle SL1 comprenant un bulge en amont. Nos résultats indiquent que l'extrémité de la tige boucle SL1 est suffisante pour adopter différentes formes impliquées dans le phénomène de dimérisation et de maturation (Feng et al. 1996b; Fu et al. 1994). La présence du bulge en amont de la boucle n'est pas capitale pour la dimérisation,

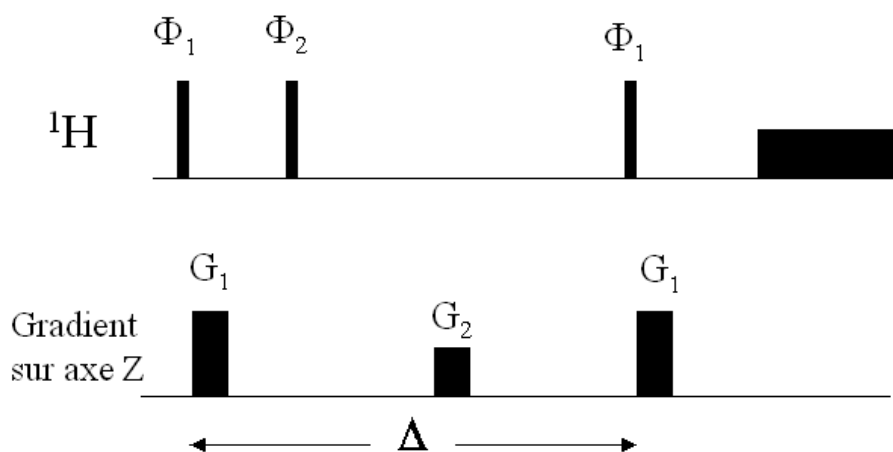
cette hypothèse a pu être vérifiée par d'autres travaux dans notre équipe (Theilleux-Delalande et al. 2000).

#### II-4-2-d. Sommes nous en présence d'un dimère ou d'un tétramère?

Nous avons pu isoler deux types de structures suivant les conditions de préparation de l'échantillon. Nous avons également pu mettre en évidence l'existence d'appariements au niveau des paires de bases auto complémentaires de la boucle. Nous pouvons donc en conclure que nous sommes en présence d'une structure qui se trouve au minimum à l'état dimérique. Néanmoins, d'après la structure cristallographique publiée par Ennifar et al. (2001), le complexe instable dit en kissing complex présente un motif 'purine-zipper'. Ce motif impliquerait alors une structure sous la forme d'un tétramère. Pour vérifier cette hypothèse en solution, nous avons réalisé des expériences de diffusion sur les deux types de structures.

En effet, tout paramètre changeant l'hydrodynamique de la molécule induit un changement de son coefficient de diffusion. Par conséquent, le coefficient de diffusion dépend directement de la taille de la molécule étudiée (Lapham et al. 1997).

Pour calculer ce coefficient de diffusion, nous avons utilisé des expériences RMN d'écho de gradient à champ pulsé (PGF). La séquence de cette expérience est représentée par la figure 64.



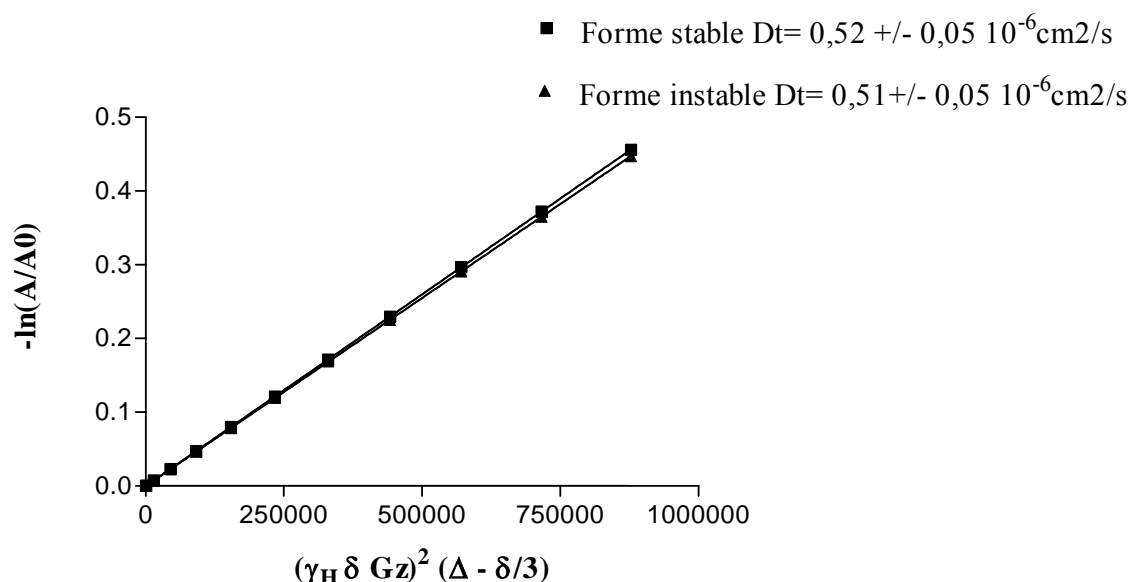
**Figure 64.** Séquence d'écho de gradient à champ pulsé (PGF).  $\Delta$  la durée entre les deux gradients. Le cyclage de phase est le suivant :  $\Phi_1$  ( $x,-x,-y,y$ ),  $\Phi_2$  ( $-x,x,y,-y$ ), la réception s'effectue avec la même phase que  $\Phi_2$ . (Lapham et al. 1997).

Cette expérience présente deux impulsions de type gradient. La premier gradient déphase l'aimantation transversale suivant l'axe z et le troisième gradient permet de refocaliser cette aimantation. Si une molécule diffuse le long de l'axe z durant le temps séparant les deux gradients, son aimantation ne sera pas complètement refocalisée. Ainsi, si une molécule diffuse rapidement, l'atténuation de ses résonances sera importante, au contraire si une molécule diffuse très peu, cette atténuation sera faible. La relation suivante relie le coefficient de diffusion avec les paramètres mesurables par RMN :

$$A/A_0 = \exp[-Dt (\gamma_H \delta Gz)^2 (\Delta - \delta/3)]$$

où A est l'intensité du pic mesuré,  $A_0$  l'intensité maximale du pic, Dt le coefficient de diffusion de translation en  $\text{cm}^2/\text{s}$ ,  $\gamma_H$  le rapport gyromagnétique du proton,  $\delta$  la durée du gradient,  $\Delta$  la durée entre les deux gradients et  $Gz$  la force du gradient en gauss/cm. Il suffit de tracer  $-\ln(A/A_0)$  en fonction  $(\gamma_H \delta Gz)^2 (\Delta - \delta/3)$  pour déterminer le coefficient de diffusion de la molécule.

Nous avons réalisé cette expérience sur la forme instable, le kissing complex, et sur le dimère stable. Nous avons fait varier la puissance du gradient de 0 à 66 gauss/cm (Figure 64). Les données ont été fittées par la méthode des moindres carrés non linéaire. Nous avons ainsi pu déterminer le coefficient de diffusion. En effet,  $A_0$  est mal connu à cause de la présence de courants induits pour les autres points. Les courbes  $\ln(A/A_0) = f(Gz)$  pour les deux types de dimères sont représentées sur la figure 65.

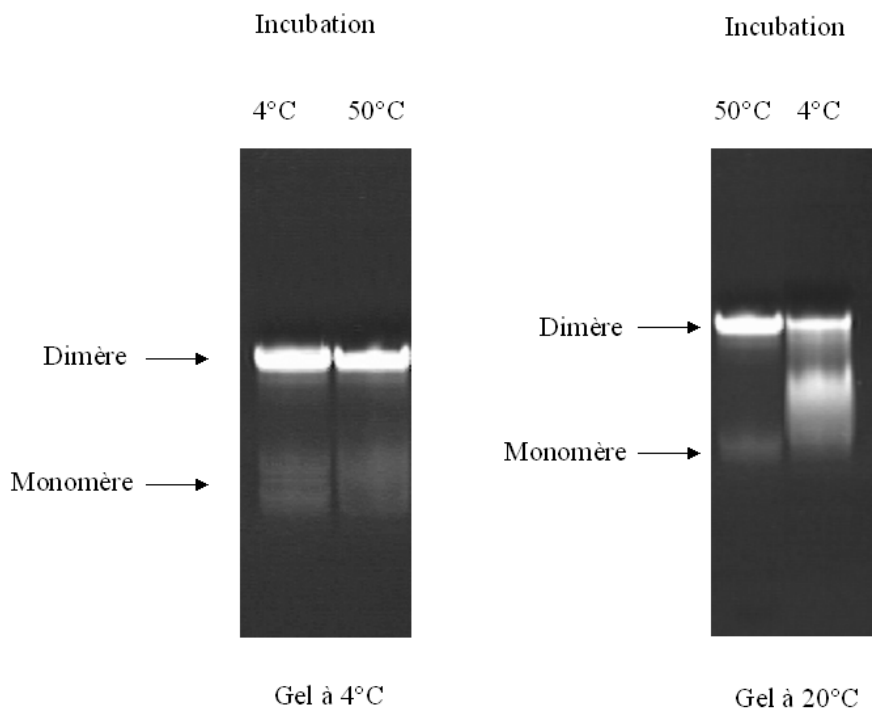


**Figure 65.** Mesure des coefficients de diffusion à 10°C de la forme stable et instable de la tige boucle SL1 24mer .



Nous obtenons des valeurs de coefficients de diffusion assez proches. Les deux formes de la tige boucle SL1 24mer diffusent approximativement à la même vitesse. On en déduit donc qu'elles ont la même masse moléculaire. Sachant que la forme stable est un dimère (Ennifar et al. 1999; Girard et al. 1999; Mujeeb et al. 1999), on peut en déduire que la forme instable adopte elle-aussi une structure dimérique en solution. On exclut ainsi la présence d'un tétramère dans nos conditions expérimentales.

Nous avons également réalisé des gels de polyacrylamide en condition native à 4°C (Figure 66). Nous avons isolé les deux types de dimère. Le forme instable est obtenue par passage rapide de 90°C à 4°C suivi d'une période d'incubation de 15 min à 4°C tandis que la forme stable est obtenue par passage rapide de 90°C à 4°C suivi d'une période d'incubation de 15 min à 50°C. D'après les résultats du gel d'électrophorèse (Figure 66), la forme instable et la forme stable migrent quasiment à la même vitesse. Nous pouvons donc en conclure que les différentes formes migrent sur gel également sous forme dimérique et non tétramérique.



**Figure 66.** Gel natif de la tige boucle SL1 24mer sous sa forme instable et stable.

Les expériences de diffusion ou de gel en condition native indiquent que la forme instable, comme la forme stable, sont sous une forme dimérique.

La présence d'un motif 'purine zipper', qui stabilise une forme tétramérique de la tige boucle SL1 en condition cristallographique semble donc peu vraisemblable en solution. Il est fortement possible que la présence de ce motif qui existe dans le cristal soit favorisée par l'empilement cristallin nécessaire pour une bonne cristallisation.

## II-6 Modélisation moléculaire

### II-6-1 Etablissement des contraintes

#### *Contraintes sur les angles*

Les spectres COSY et TOCSY effectués sur la forme instable nous ont permis de déterminer les conformations des différents sucres de l'oligonucléotide. Une constante de couplage quasi nulle (1 à 0 Hz) indique que le sucre se trouve en conformation nord ou C3' endo. Dans le cadre de notre étude, tous les nucléotides à l'exception des résidus 8, 9, 16 se trouvent sous cette conformation nord, les autres se trouvent sous une conformation du sucre de type sud. Un jeu de 96 contraintes d'angles dièdres ( $\nu_1$  et  $\nu_2$ ) a été créé pour rendre compte des plissements du sucre. Ces valeurs d'angles, tirées de l'ouvrage de Saenger (Saenger 1984) sont imposées avec une incertitude de 15°.

Nous avons pu voir précédemment que la quasi-totalité des raies de résonance du phosphore était regroupée sur un faible écart de déplacement chimique indiquant une faible variation de la conformation des phosphates. Nous avons choisi de contraindre les angles du squelette phosphodiester ( $\alpha, \beta, \gamma, \epsilon, \xi$ ) avec des valeurs classiques d'hélice A. Nous avons imposé ces contraintes avec une incertitude de 30° sur tous les résidus de l'oligonucléotide à l'exception de G7, A8, A9, G10, C15, A16, C17, résidus qui sont impliqués dans la jonction de la tige avec la séquence auto-complémentaire. La valeur des incertitudes sur ces résidus est de +/- 90° pour ne pas contraindre le système au niveau de la jonction.

Nous avons introduit des contraintes sur l'angle  $\chi$ , qui correspond à l'angle formé entre le sucre et la base de chaque nucléotide. Sa valeur peut être déduite de l'observation de l'intensité des connectivités H1'- aromatique intra-résiduel. Tous les résidus sont sous forme *anti* à l'exception des résidus A8 et A9 qui possèdent une forte connectivité à 50ms entre le

proton anomérique et le proton aromatique et se trouvent dans une conformation proche de *syn*.

### *Contraintes de distances*

Les contraintes de distance entre les différents protons non échangeables ont pu être estimées à partir des spectres NOESY à différents temps de mélange (50,100,150ms) et ont été rangées en plusieurs classes : fort ( $2\text{\AA} < \text{distance} < 3,5\text{\AA}$ ), moyen ( $3,5\text{\AA} < \text{distance} < 5\text{\AA}$ ), et faible ( $4,5\text{\AA} < \text{distance} < 6\text{\AA}$ ). Les contraintes de distance impliquant des protons échangeables sont estimées entre 3 et 6Å afin de tenir compte des éventuelles erreurs dues à la diffusion de spin. Nous avons ajouté un jeu de 108 contraintes de distance pour rendre compte de l'appariement de type Watson-Crick des 7 premières paires de bases de la tige et de l'appariement des deux séquences auto complémentaires des deux brins d'ARN. Des contraintes de plan ont également été utilisées lors des premiers cycles de calcul pour favoriser l'appariement des bases. La constante de force pour chaque plan est de  $15 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$  (pour comparaison, chaque contrainte NOE a une force de  $50 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ ).

L'ensemble des contraintes expérimentales pour la modélisation est résumé dans le tableau 67.

Il existe une ambiguïté dans l'attribution des différents protons de l'oligonucléotide. Nous avons réussi à distinguer les différents protons H2 de l'oligonucléotide grâce à des spectres HSQC  $^{13}\text{C}$  mais l'analyse de leur réseau de connectivités NOE dans l'oligonucléotide ne nous a pas permis de distinguer les résonances des protons H2 des résidus A8 et A9. En effet, ces protons possèdent dans cette structure un environnement chimique très voisin ce qui se traduit par un ensemble de connectivités très similaire. Nous avons donc choisi de résoudre la structure de l'oligonucléotide pour les deux types d'attribution possible et de tester chacune d'elle par modélisation moléculaire.

## Contraintes RMN pour la structure du dimère instable

<i>A. Contraintes de distance</i>		
Intra-résiduelles		252
Inter-résiduelles		116
Liaison Hydrogène		108
Total des contraintes de distance		476
nOe / résidu		9,9
Constante de force		50 kcal.mol <sup>-1</sup> . Å <sup>-2</sup>
<i>B. Contraintes d'angle dièdre</i>		
Plissement du sucre		144
Squelette phosphodiester		220
Total des contraintes d'angle		364
Angles / résidu		7,6
Constante de force		25 kcal.mol <sup>-1</sup> . Å <sup>-2</sup>
<i>C. Total</i>		
Nombre de contraintes		840
Contraintes / résidu		17,5

**Tableau 67.** Résumé de l'ensemble des contraintes RMN utilisées pour la modélisation du dimère instable de SL1.

### II-6-2 Résolution de la structure tridimensionnelle

Nous avons généré à l'aide du logiciel CNS version 1.0 (Brünger et al. 1998) 100 structures aléatoires respectant la covalence des liaisons. Ces structures servent de point de départ pour la modélisation moléculaire de la forme instable de SL1 24mer.

Les structures aléatoires sont soumises à un recuit simulé dans l'espace des angles dièdres à l'aide du même logiciel (Stein et al. 1997). La fonction d'énergie potentielle incluse dans ce protocole est simplifiée pour prendre en compte les termes de : liaisons, angles, angles dièdres, angles impropres, potentiel de Van der Waals sous sa forme répulsive, et les termes expérimentaux de distances NOE, d'angles dièdres et de plans. Les premiers contacts de Van der Waals, appelés 'close contact', sont éliminés par une minimisation de 200 pas de mécanique moléculaire en utilisant la méthode du gradient conjugué (Powell 1977). Les vitesses initiales sont ensuite attribuées selon une distribution statistique de Maxwell à 20000K. Une première période de dynamique moléculaire est effectuée dans l'espace des angles dièdres à température constante de 20000K pendant 60ps. Elle sert à explorer l'espace

conformationnel. La constante de force de Van Der Waals est maintenue à  $0,1 \text{ kcal.mol}^{-1} \cdot \text{\AA}^{-2}$ . Les constantes de forces expérimentales des NOE et des angles dièdres sont respectivement à  $50 \text{ kcal.mol}^{-1} \cdot \text{\AA}^{-2}$  et  $5 \text{ kcal.mol}^{-1} \cdot \text{rad}^{-2}$ . Le refroidissement du système, de 20000 à 0K, se réalise à l'aide d'une dynamique de 60 ps dans l'espace des angles de torsion. La constante de force de la fonction répulsive de Van der Waals évolue alors progressivement de 0,1 à 1,0  $\text{kcal.mol}^{-1} \cdot \text{\AA}^{-2}$  tandis que celle des NOE et des angles dièdres sont à  $50 \text{ kcal.mol}^{-1} \cdot \text{\AA}^{-2}$ . Pour une convergence plus rapide des structures, nous effectuons une deuxième dynamique moléculaire dans l'espace cartésien de 10 ps à 1500K. La phase de refroidissement (1500K à 300K) se fait suivant une dynamique de 10ps où la constante de force de Van der Waals passe de 1 à 4  $\text{kcal.mol}^{-1} \cdot \text{\AA}^{-2}$ . Enfin, le protocole est suivi d'une étape de mécanique moléculaire de 400 pas avec l'algorithme de Powell.

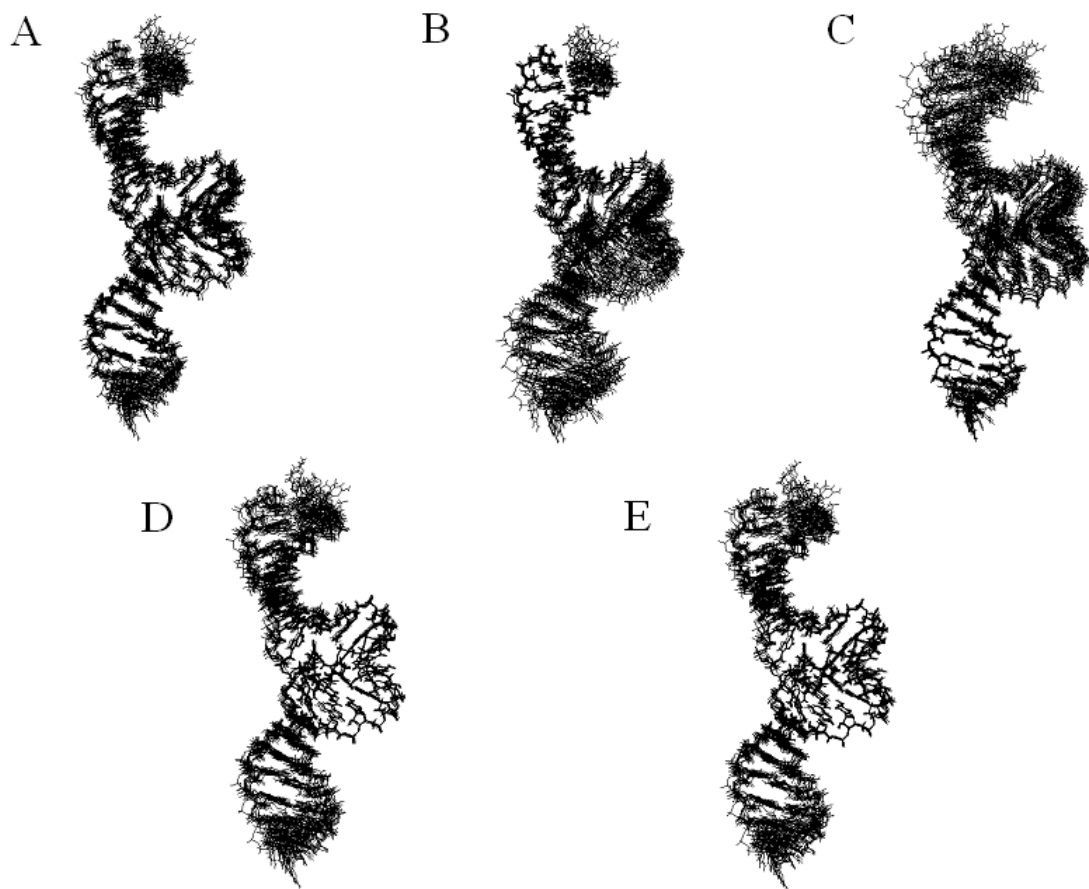
A ce niveau, les structures sont sélectionnées selon le nombre de leurs violations NOE et d'angles dièdres ainsi que leur énergie totale. Pour finir, les structures résultantes sont examinées graphiquement.

Dans ce protocole, les contraintes de distances ont été introduites de la façon suivante. Au départ, nous avons pris en compte seulement les appariements entre bases et les angles dièdres, ensuite, les contraintes de distance pour les résidus intervenant dans la formation de la tige. Nous obtenons ainsi un jeu de structures où la structure des tiges et l'appariement des deux séquences auto complémentaires sont respectés. Sur ces structures, nous avons effectué de nouvelles étapes de dynamique moléculaire en ajoutant les contraintes entre les différentes bases qui sont impliquées dans la jonction entre la tige et la séquence auto complémentaire (A8, A9, A16). Ces contraintes de distance peuvent être soit inter soit intra-moléculaires et ont été injectées une à une sous leurs deux formes dans la modélisation. La contrainte (inter ou intra des H2/H8 des résidus A8 et A9) permettant d'obtenir le minimum de violation structurale est alors retenue. Une fois que toutes les contraintes de la jonction ont été injectées et sélectionnées de façon inter ou intra moléculaire, nous avons introduit le reste des contraintes RMN.

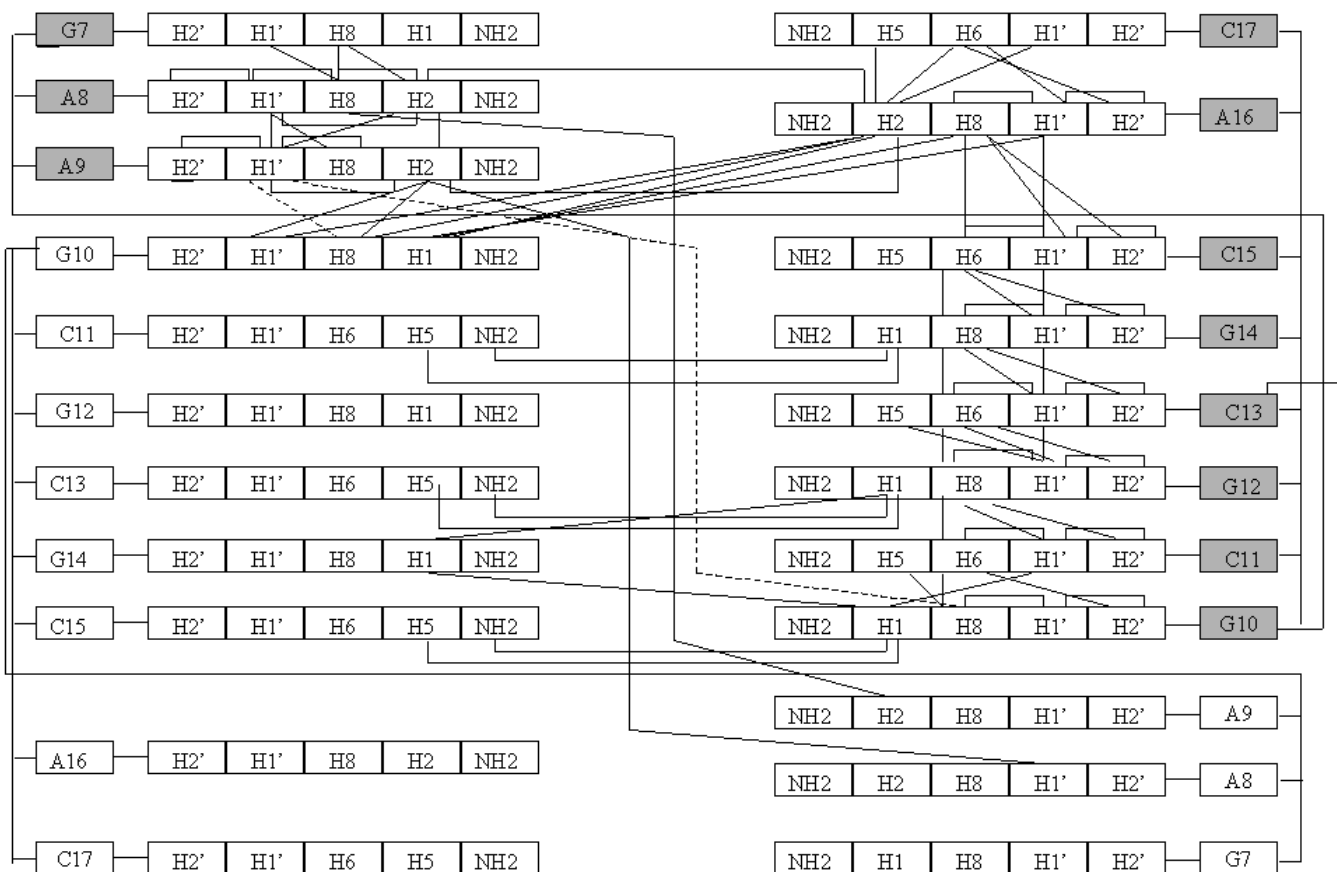
Nous obtenons ainsi deux ensembles de structures qu'on appellera ensemble 1 et ensemble 2 (Figures 68 et 71) répondant à toutes les contraintes expérimentales. Ces structures ne présentent aucune violation de distance supérieure à  $0,5 \text{\AA}$  et aucune violation d'angle dièdre supérieure à  $5^\circ$ . Les tableaux 70 et 72 résument l'ensemble des termes énergétiques et les RMSD obtenus pour ces deux ensembles.

### *Organisation structurale de l'ensemble 1*

Nous avons pu obtenir un premier ensemble de 10 structures (Figure 68) respectant ces différentes contraintes. Les différentes contraintes de distance introduites dans la modélisation moléculaire de l'ensemble 1 sont représentées dans la figure 69. Le tableau 70 résume les énergies et les RMSD de cette ensemble structural. Les RMSD sont calculés par comparaison avec une structure moyenne non minimisée.



**Figure 68.** Superposition des 10 structures du dimère instable de la séquence SL1 de l'ensemble 1. A) superposition globale ; B),C),D) superposition sur les tiges 1, la tige 2, la partie centrale; E) superposition sur les adénines de jonction.



**Figure 69.** Schéma des différentes contraintes de distance introduites pas à pas dans la modélisation moléculaire de SL1 du dimère instable de l'ensemble 1. Les résidus de la tige A sont représentés en grisé, les résidus de la tige B apparaissent en clair. Seules les contraintes impliquant les résidus de la tige A ont été représentées.

### Précision des structures

Nombre de structures	10
Nombre de violations NOE > 0,5Å	0
Nombre de violations NOE > 0,3Å	4
Nombre de violation dièdre > 5°	0
RMSD <sup>1</sup> global	1,01 ± 0,33 Å
RMSD tige 1 (G1-G7,C17-C23)	0,58 ± 0,11 Å
RMSD tige 2 (G1'-G7',C17'-C23')	0,59 ± 0,11 Å
RMSD tige 3 (G10-C15,G10'-C15')	0,39 ± 0,06 Å
RMSD adénines de jonction	0,44 ± 0,10 Å

### Energie

Energie totale	-1163 ± 30 kcal.mol <sup>-1</sup>
Energie géométrique <sup>2</sup>	242 ± 9 kcal.mol <sup>-1</sup>
Energie NOE	47 ± 4 kcal.mol <sup>-1</sup>
Energie Van der Waals	-685 ± 10 kcal.mol <sup>-1</sup>
Energie électrostatique	-767 ± 25 kcal.mol <sup>-1</sup>

<sup>1</sup> RMSD calculé sur les atomes lourds par rapport à la moyenne non minimisée.

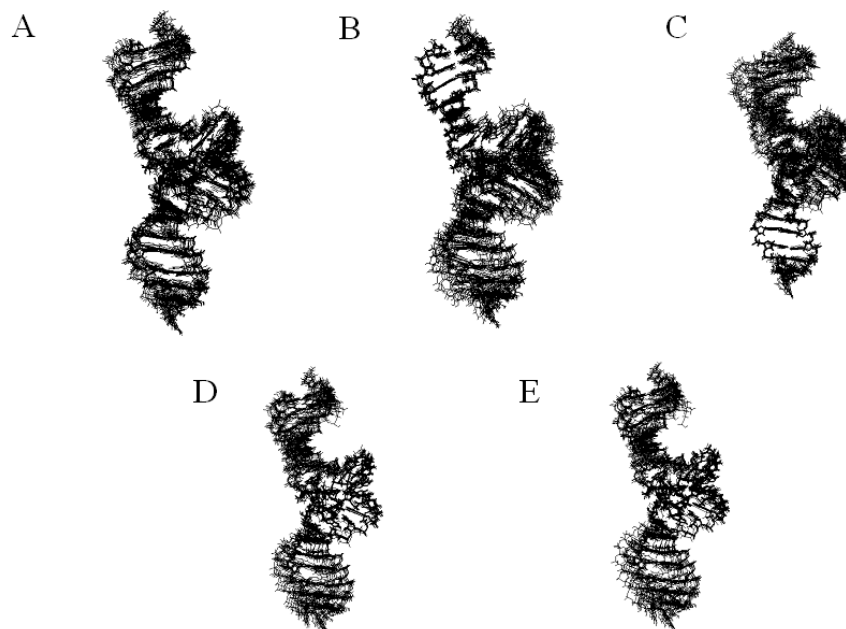
<sup>2</sup> Terme géométrique incluant les paramètres de liaisons, angles, angles de torsion, angles impropres et angles dièdres.

**Tableau 70.** Résumé des énergies et des RMSD pour la modélisation du dimère instable de la séquence SL1 de l'ensemble 1.

### *Organisation structurale de l'ensemble 2*

Nous avons ensuite inversé l'attribution des protons H2 des adénines A8 et A9. Nous avons pu sélectionner, grâce à la modélisation, un deuxième jeu de contraintes inter et intramoléculaire pour les résidus impliqués dans la jonction (Figure 71). Nous avons pu obtenir un deuxième ensemble de 12 structures correspondant à nos contraintes expérimentales. L'ensemble des différentes contraintes impliquant les résidus de la boucle est représenté dans la figure 73. Le tableau 72 résume les énergies et les RMSD de cette ensemble structural. Les RMSD sont calculés par comparaison avec une structure moyenne non minimisée.





**Figure 71.** Superposition des 12 structures du dimère instable de la séquence SL1 de l'ensemble 2. A) superposition globale ; B),C),D) superposition sur les tiges 1, la tige 2, la partie centrale; E) superposition sur les adénines de jonction.

### Précision des structures

Nombre de structures	12
Nombre de violations NOE > 0,5Å	0
Nombre de violations NOE > 0,3Å	0
Nombre de violation dièdre > 5°	0
RMSD <sup>1</sup> global	1,35 ± 0,52 Å
RMSD tige 1 (G1-G7,C17-C23)	0,62 ± 0,11 Å
RMSD tige 2 (G1'-G7',C17'-C23')	0,61 ± 0,11 Å
RMSD tige 3 (G10-C15,G10'-C15')	0,29 ± 0,16 Å
RMSD adénines de jonction	0,41 ± 0,21 Å

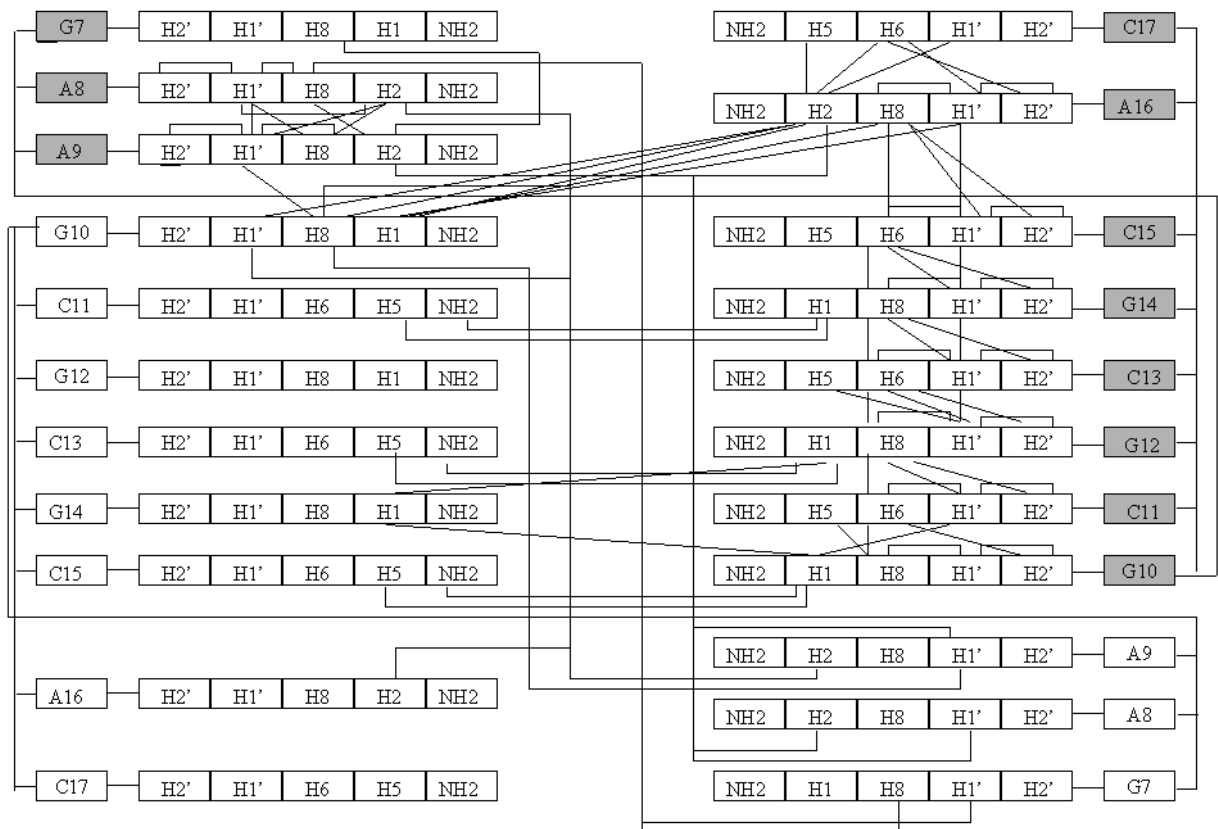
### Energie

Energie totale	-1243 ± 50 kcal.mol <sup>-1</sup>
Energie géométrique <sup>2</sup>	219 ± 10 kcal.mol <sup>-1</sup>
Energie NOE	12 ± 2 kcal.mol <sup>-1</sup>
Energie Van der Waals	-724 ± 11 kcal.mol <sup>-1</sup>
Energie électrostatique	-750 ± 38 kcal.mol <sup>-1</sup>

<sup>1</sup> RMSD calculé sur les atomes lourds par rapport à la moyenne non minimisée.

<sup>2</sup> Terme géométrique incluant les paramètres de liaisons, angles, angles de torsion, angles impropres et angles dièdres.

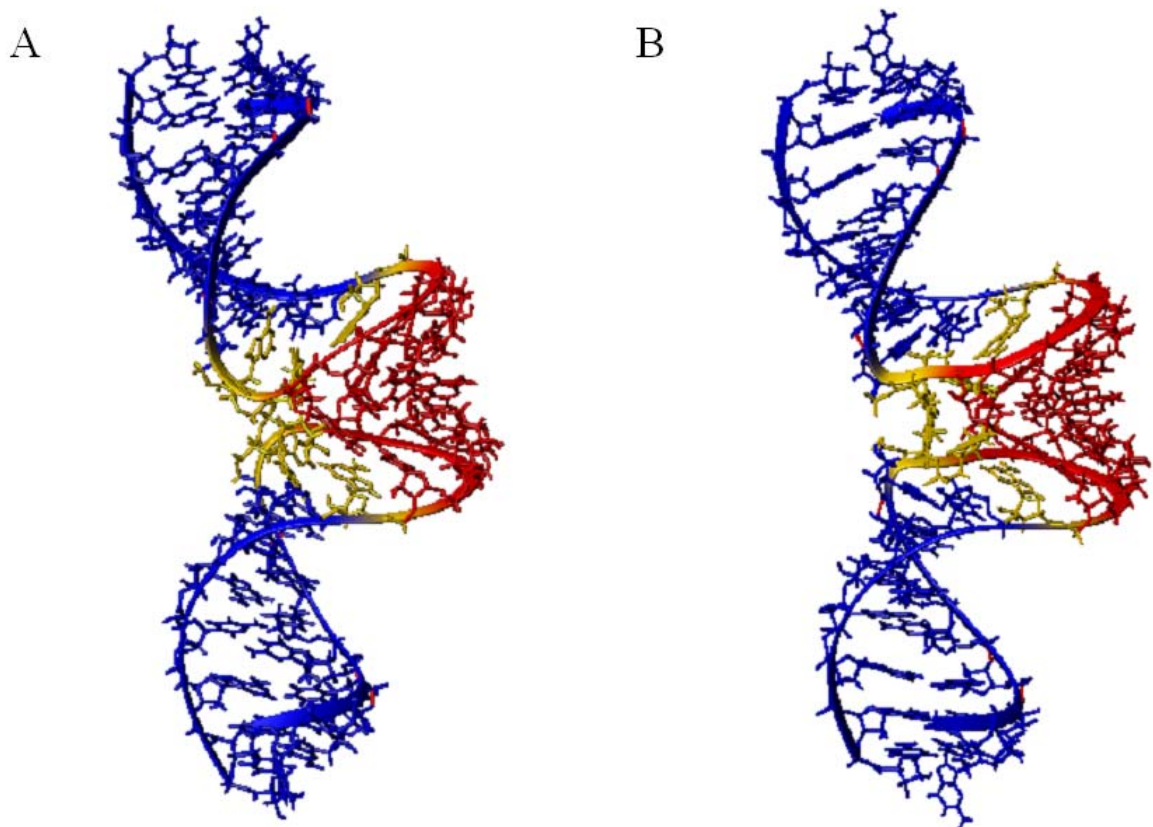
**Tableau 72.** Résumé des énergies et des RMSD pour la modélisation du dimère instable de la séquence SL1 de l'ensemble 2.



**Figure 73.** Schéma des différentes contraintes de distance introduites dans la modélisation moléculaire de SL1 du dimère instable de l'ensemble 2. Les résidus de la tige A sont représentés en grisé, les résidus de la tige B apparaissent en clair. Seules les contraintes impliquant les résidus de la tige A ont été représentées.

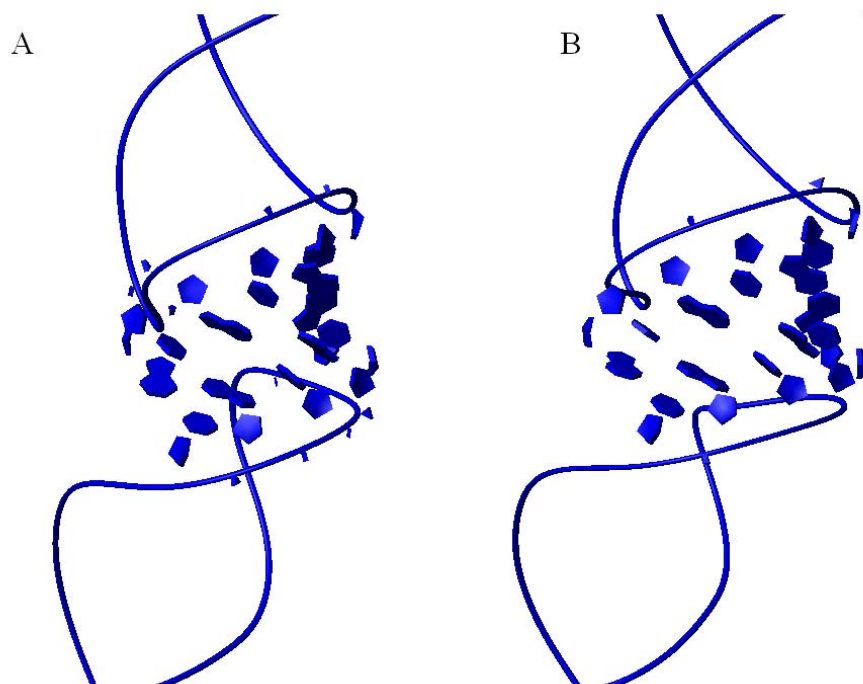
### *Organisation générale des deux ensembles de structure*

La structure moyenne du dimère instable de la tige boucle SL1 de chaque ensemble de structure (Figure 74) se présente sous la forme d'un homo-dimère symétrique. Les tiges de chaque monomère adoptent une structure en double hélice de type A. Chaque hélice est composée d'un ensemble de 7 appariements de type Watson-Crick entre les résidus G1 à G7 et de C17 à C23. L'analyse structurale a été effectuée à l'aide de la version 5.1 du logiciel CURVES (Lavery and Sklenar 1988, 1989) et du logiciel MOLMOL (Koradi et al. 1996).



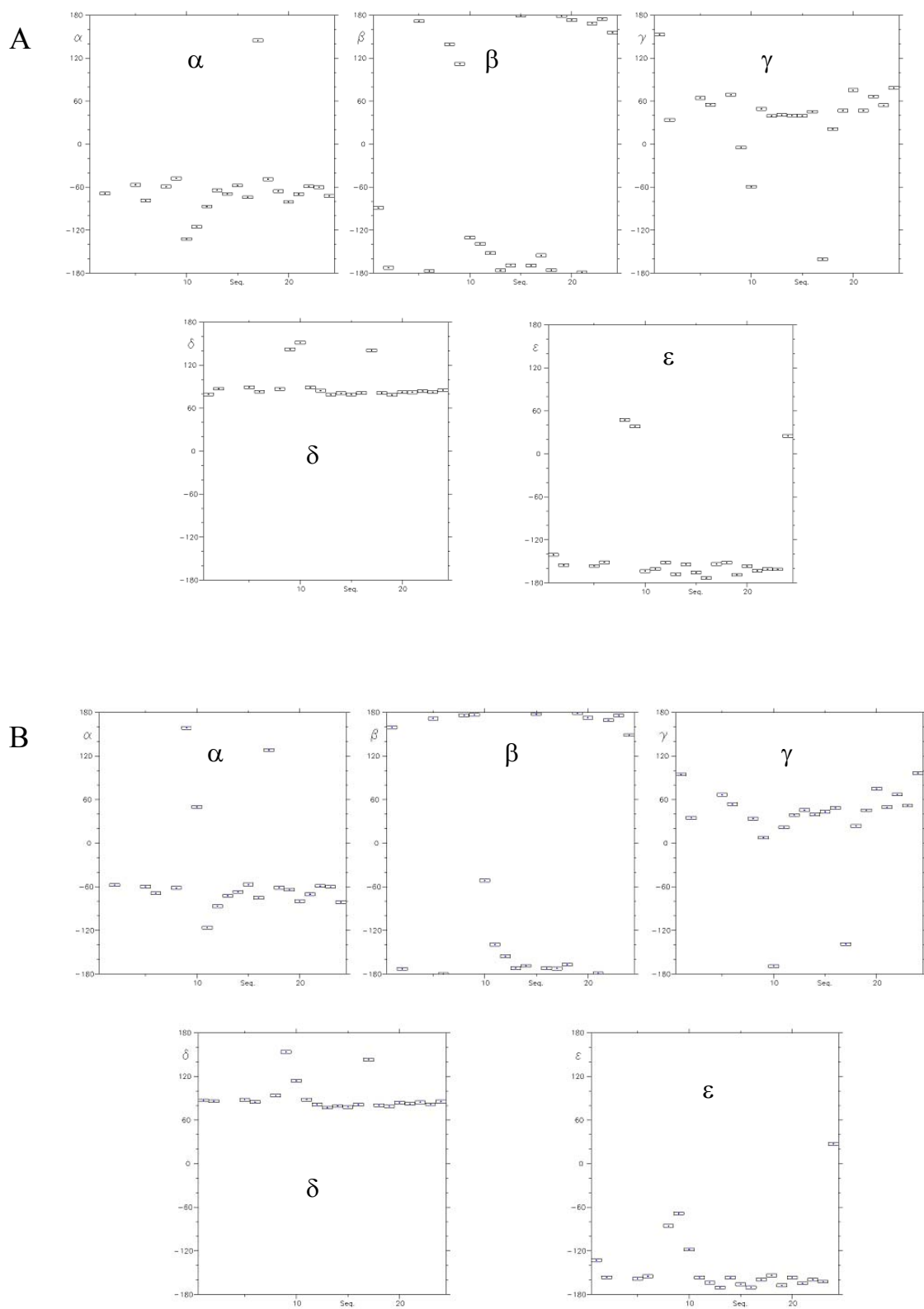
**Figure 74.** Structure moyenne du dimère instable de la séquence SL1 pour l'ensemble 1 (A) et pour l'ensemble 2 (B).

L'appariement central entre les deux tiges boucles monomériques est composé de 6 paires de bases appariées G-C pour les deux ensembles de structures. L'agencement de ces appariements n'est pas hélicoïdal (Figure 75) et se présente sous la forme de plans empilés avec une inclinaison d'environ  $50^\circ$  vers la gauche par rapport à l'axe de l'hélice. Les plans des bases appariées sont enroulés autour de l'interaction boucle-boucle en suivant un demi-cercle. Il existe un effet d'empilement entre les différentes paires de bases appariées de la séquence auto-complémentaire permettant une plus grande stabilisation de l'interaction boucle-boucle.



**Figure 75.** Organisation de l'appariement entre les deux entités monomériques du dimère instable de SL1 pour l'ensemble 1 (A) et l'ensemble 2 (B).

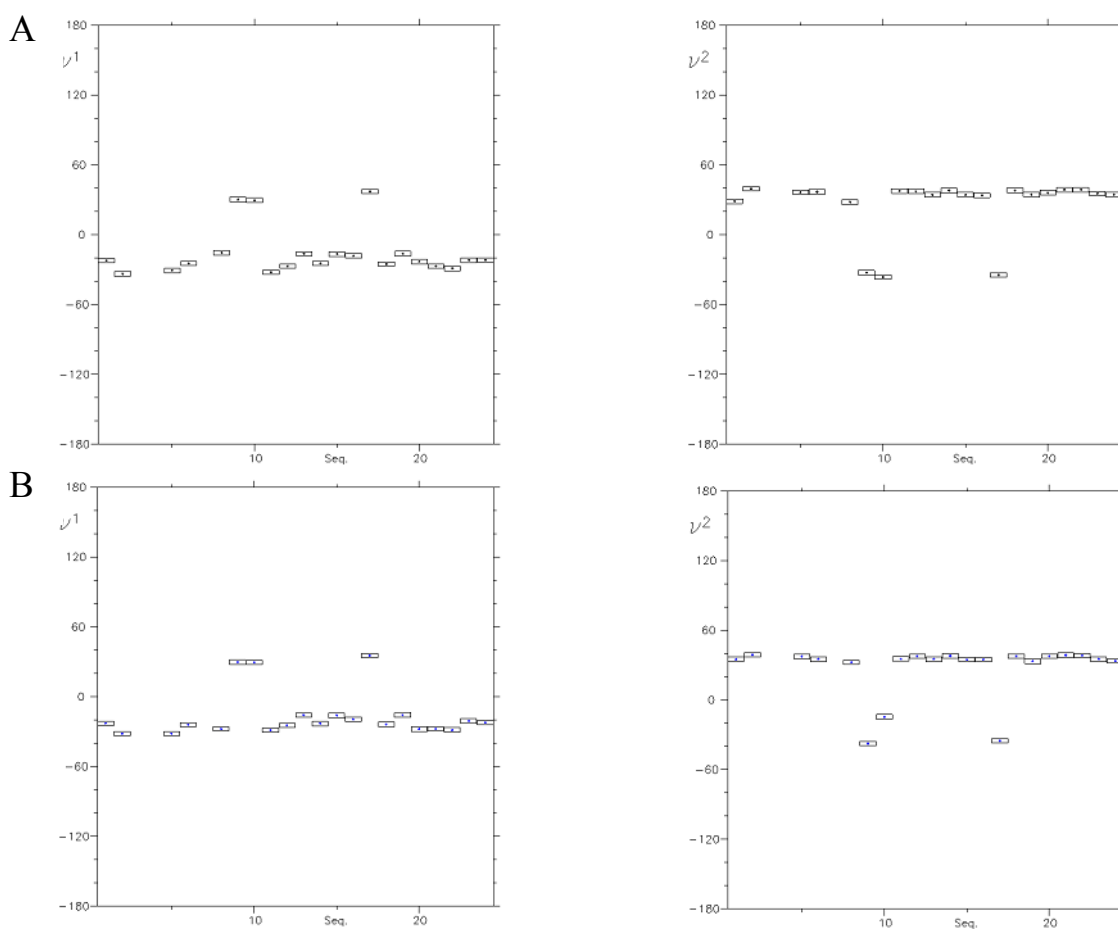
Le squelette phosphodiester montre de fortes distorsions pour les deux ensembles structuraux (Figure 76), au niveau des résidus situés dans la jonction entre la tige et la séquence auto-complémentaire (A8,A9,G10,A16). Ces variations indiquent que le squelette phosphodiester de la forme instable de la tige boucle SL1 24mer est soumis dans la région de la jonction à un ensemble de contraintes géométriques pour permettre l'appariement de la séquence auto-complémentaire et la reformation de la double hélice. L'ensemble 2 présente au niveau de son squelette phosphodiester une cassure, au niveau de la jonction en A8. Cette cassure peut être visualisée en observant la valeur de l'angle  $\alpha$  qui est de  $160^\circ$  pour le deuxième ensemble de structures tandis que pour le premier ensemble cette valeur de  $-40^\circ$  (valeur classique pour une hélice A  $-70^\circ$ ). Cette valeur est la conséquence directe de la position de l'adénine A8 très particulière dans le deuxième ensemble de structures.



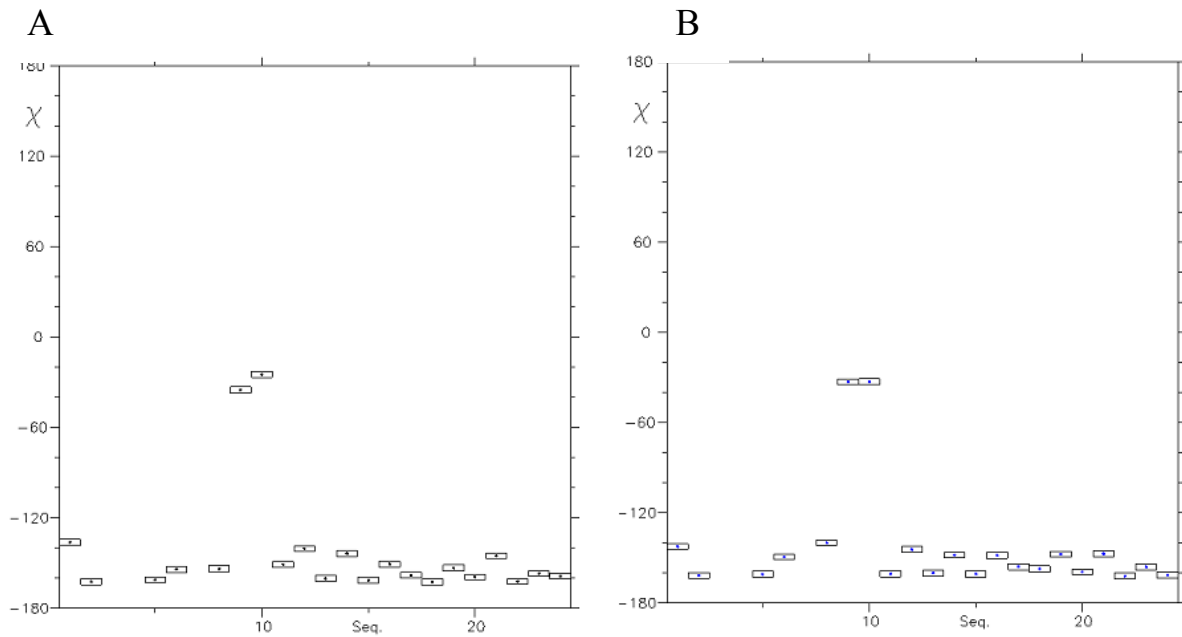
**Figure 76.** Variation des angles du squelette phosphodiester en fonction de la séquence (abscisse) pour l'ensemble 1 (A) et l'ensemble 2 (B).

### Organisation structurale de la jonction des deux ensembles de structure

La jonction se compose de trois adénines (A8,A9 et A19). Bien que l'agencement spatial des adénines A8 et A9 soit différent entre les deux ensembles de structures, les structures présentent néanmoins des caractéristiques communes pour ces trois résidus. Leurs sucres se présentent sous une forme C2'-endo tandis que pour le reste de la molécule, les sucres se trouvent tous sous une conformation de type C3'-endo (nord) (Figure 77). Les positions des bases par rapport aux sucres des adénines A8 et A9 indiquent que celles-ci se trouvent en configuration *syn* contrairement aux autres résidus qui sont en configuration *anti* (Figure 78).

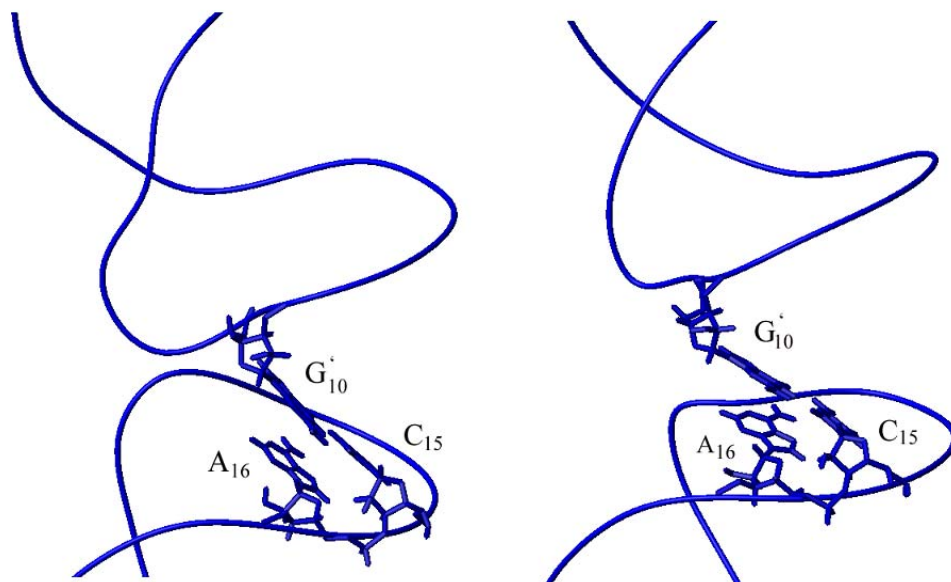


**Figure 77.** Variation des angles  $\nu^1$  et  $\nu^2$  du sucre en fonction de la séquence (abscisse) pour l'ensemble 1 et l'ensemble 2.



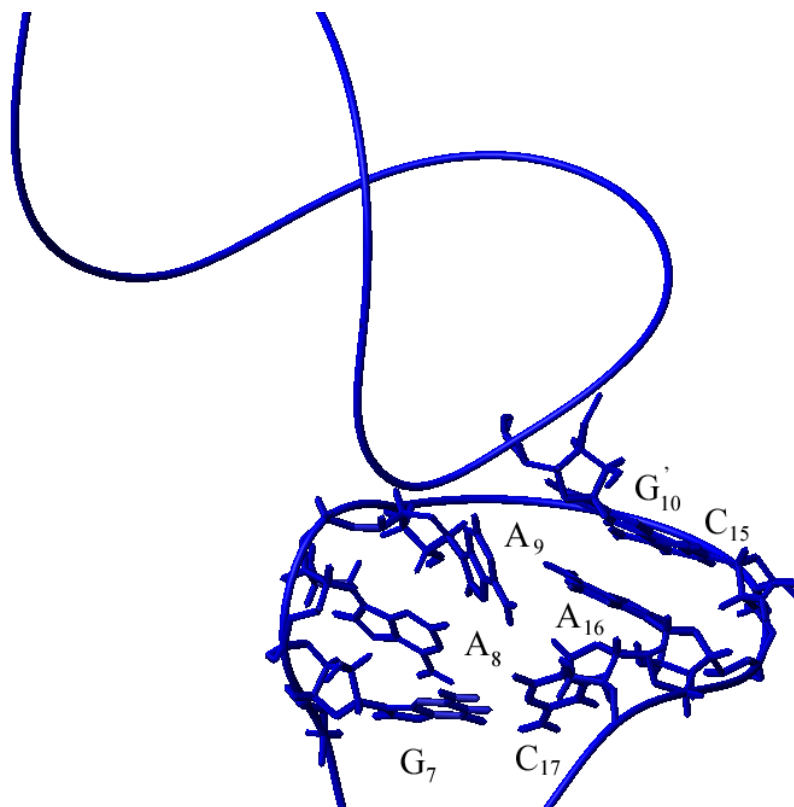
**Figure 78.** Variation de l'angle  $\chi$  entre la base et le sucre de chaque résidu de la séquence.

Une caractéristique structurale importante et commune aux deux ensembles de structures est la stabilisation de l'édifice moléculaire par l'empilement de l'adénine A16 avec la guanine G10 de l'autre brin (Figure 79). La guanine G10 est donc à la fois impliquée dans des liaisons hydrogène avec la cytosine C15 de l'autre brin et dans un empilement avec l'adénine A16 également de l'autre brin.



**Figure 79.** Représentation du motif d'empilement impliquant la paire de bases inter-brin G10'-C15 et l'adénine A16 pour l'ensemble 1 (A) et l'ensemble 2 (B).

Pour l'ensemble 1, les résidus A8 et A9 de chaque monomère se tournent vers l'intérieur de la boucle (Figure 80). L'adénine A8 est partiellement empilée avec la guanine G7 du même brin et forme une liaison hydrogène avec celle-ci (Tableau 81). L'adénine A9 peut elle aussi former une liaison hydrogène avec la cytosine C17 permettant d'accroître la stabilité de la structure.



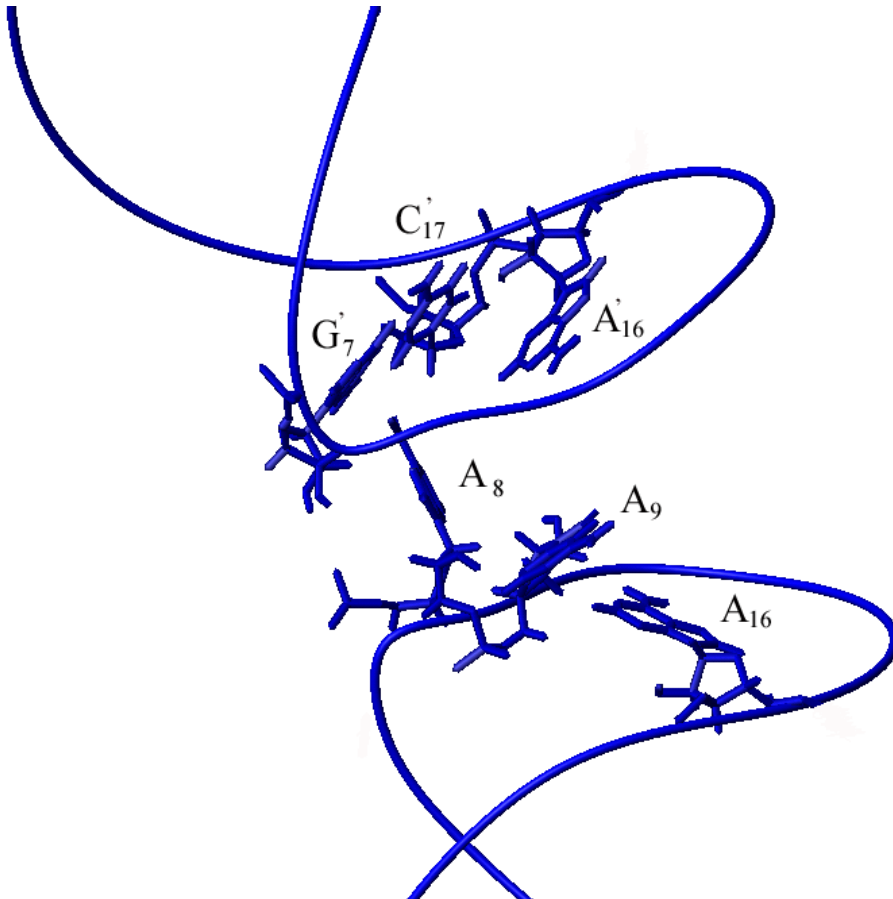
**Figure 80.** Représentation de l'organisation de l'entrée de boucle du dimère instable pour la structure moyenne de l'ensemble 1.

Donneur	Accepteur	Distance en Å	Angle en °
8 ADE NH2	7 GUA O6	3,00	28,8
8 ADE HO2	9 ADE O2P	2,87	19,6
9 ADE NH2	17 CYT O2	2,47	28,5
15 CYT HO2	16 ADE O5'	2,10	26,2

**Tableau 81.** Liste des liaisons hydrogène impliquant les adénines de l'entrée de boucle dans la structure moyenne du dimère instable de SL1 24mer de l'ensemble 1. Les liaisons hydrogène ont été définies par une proximité inférieure à 3Å et un angle entre l'atome donneur et accepteur de liaison hydrogène inférieur à 40°.



Pour l'ensemble 2, la position des adénines A8 et A9 est complètement opposée (Figure 82). L'adénine A8 est tournée vers l'extérieur du monomère pour interagir avec la cytosine C17 de l'autre monomère via des liaisons hydrogène. L'adénine A9, quant à elle, reste tournée vers l'intérieur de la boucle et forme des liaisons hydrogène avec la guanine 10 et l'adénine 16 du même monomère. La liste des différentes liaisons hydrogènes impliquant les différentes adénines d'entrée de boucle est représenté dans le tableau 83.



**Figure 82.** Représentation de l'organisation de l'entrée de boucle du dimère instable pour la structure moyenne de l'ensemble 2.

Hydrogène	Accepteur	Distance en Å	Angle en °
8 ADE NH2	17' CYT O2	2,65	20,3
8 ADE HO2	9 ADE O4'	2,72	30,2
8 ADE HO2	9 ADE N3	2,14	28,6
9 ADE NH2	10' GUA N7	2,26	35,8
9 ADE HO2	9 ADE O3'	1,99	39,6
16 ADE NH2	9 ADE N6	2,59	35,1
16 ADE HO2	17 CYT O5'	1,75	30,6

**Tableau 83.** Liste des liaisons hydrogène impliquant les adénines de l'entrée de boucle dans la structure moyenne du dimère instable de SL1 24mer de l'ensemble 2. Les liaisons hydrogène ont été définies par une proximité inférieure à 3Å et un angle entre l'atome donneur et accepteur de liaison hydrogène inférieur à 40°.

#### *Choix de la structure la plus probable du dimère instable de la tige SL1 24mer*

Nous avons pu obtenir deux ensembles de structures qui convergent avec nos données RMN lorsqu'on utilise l'une des deux attributions possibles de la résonance des protons H2 des adénines A8 et A9 de l'oligonucléotide. Nous avons choisi les structures obtenues dans l'ensemble 2 pour être la structure finale du dimère instable pour deux raisons. L'ensemble 2 ne fait aucune violation sur les angles dièdres mais surtout sur les contraintes NOE, même supérieure à 0,3Å alors qu'elles sont toujours au minimum au nombre de 4 pour l'ensemble 1. La deuxième raison porte sur l'arrangement des adénines dans l'entrée de la boucle, celles-ci sont impliquées dans un plus grand nombre de liaisons hydrogène dans l'ensemble 2 que dans l'ensemble 1. L'énergie totale de l'ensemble 2 est effectivement plus basse que celle obtenue pour l'ensemble 1. Nous avons effectué des expériences de demi-filtre  $^{13}\text{C}$  qui nous ont permis de mettre en évidence la proximité de l'adénine A16 d'un brin et de la guanine G10 de l'autre brin. Malheureusement, du fait de la faible concentration en oligonucléotide marqué, nous n'avons pas pu mettre en évidence des connectivités entre les adénines A8 et A9 avec le reste de l'oligonucléotide. Sur la base de l'étude structurale, nous en déduisons que la deuxième structure est la plus probable pour la conformation du dimère instable de la tige boucle SL1 24mer.

## II-7 Discussion

La structure RMN (Mujeeb et al. 1998) et la structure cristallographique (Ennifar et al. 2001) du complexe boucle-boucle, qui ont été préalablement publiées, sont différentes de la structure que nous avons obtenue.

Notre structure du complexe boucle-boucle se présente sous la forme d'un homo-dimère composé de deux tiges en double hélice de type A. Ces tiges en double hélice possèdent 7 appariements entre les différentes bases (G1-C23, U2-A22, U3-A21, G4-C20, C5-G19, U6-G18, G7-C17) comme dans le cas de la structure cristallographique (Tableau 84). La formation de la paire de bases G7-C17 a pu être vérifiée sur des spectres HSQC-<sup>15</sup>N et cette observation va en l'encontre de la structure publiée par RMN par l'équipe de Mujeeb. En effet, dans cette structure cet appariement n'existe pas et les auteurs indiquent que tous les efforts afin d'apparier les bases G7-C17 conduisent à des violations structurales importantes. De plus, il est indiqué que seule la guanine G7 présente une conformation sud comme dans l'étude cristallographique. Ces observations structurales sur la conformation du sucre de la base G7 n'ont pas été confirmées lors de notre étude qui indiquent plutôt que celle-ci se trouve en conformation nord.

	Kieken et al.	Mujeeb et al.	Ennifar et al.
tige	7 appariements W. C. sucres en nord	6 appariements W.C. G7 en sud	7 appariements W.C. G7 en sud
entrée de boucle	A8, A9 ,A16 en sud A8 et A9 en <i>syn</i>	A8, A9, A16 en nord A8 et A9 en <i>anti</i>	A9 en sud et A8 , A16 en nord A8 et A9 en <i>anti</i>
séquence autocomplémentaire	6 appariements W.C.	6 appariements W.C.	6 appariements W.C.

**Tableau 84.** Tableau comparatif des caractéristiques structurales des trois structures de 'kissing complex' de VIH-1 ( W. C. indiquent des appariements de type Watson-Crick).

L'appariement central entre les deux séquences auto-complémentaires, comme dans le cas des deux structures publiées, concerne les 6 paires de bases G-C de la boucle. Ces différents appariements ne constituent pas une structure hélicoïdale et se présentent sous la forme de plans empilés avec une inclinaison de 52° vers la gauche par rapport à l'axe de l'hélice.

La partie structurale qui diffère le plus entre les différentes structures publiées, est la zone d'entrée de boucle composée de 3 adénines. Ces trois adénines possèdent un sucre en conformation sud alors que seule l'adénine A9 se trouve dans cette conformation dans la structure cristallographique. Dans l'étude RMN de Mujeeb, tous les sucres de l'oligonucléotide à l'exception de G7 sont en conformation nord. Nous trouvons également que les bases A8 et A9 ont une configuration *syn* alors que dans les structures publiées tous les résidus sont en position *anti*.

Nous trouvons un empilement de l'adénine A16 avec la guanine G10 de l'autre brin. Cet empilement est également présent dans la structure cristallographique mais absent de la structure RMN publiée par Mujeeb. Après vérification du fichier de contraintes utilisé pour la résolution structurale de la structure RMN, aucune connectivité NOE n'existe entre le H2 de l'adénine A16 et les protons H1' et H8 de la guanine 10 (qui sont nettement visibles sur nos spectres figure 51) ; ce qui indique que celle-ci adopte une conformation différente.

Dans notre structure, la position des adénines A8 et A9 est complètement opposée. L'adénine A8 est tournée vers l'extérieur du monomère pour interagir avec la cytosine C17 de l'autre monomère via des liaisons hydrogène. L'adénine A9, quant à elle, reste tournée vers l'intérieur de la boucle pour former des liaisons hydrogène avec la guanine 10 et l'adénine 16 du même monomère. L'équipe de James a pu mettre en évidence un empilement de trois bases qui permet de stabiliser le domaine de jonction entre la tige et la séquence auto-complémentaire. Le résidu A9 du premier brin s'intercale entre les résidus A8' et C17' du second brin (l'apostrophe désigne le second brin). Ce type d'empilement implique un certain nombre de proximités entre protons qui doivent se traduire sur un spectre RMN par des effets NOE si cette distance est inférieure à 5Å. Or nous n'observons aucun NOE entre le proton H2 de l'adénine A9 et les protons H1' et H6 de la cytosine C17 alors que la distance entre ces différents protons est inférieure à 5Å dans la structure publiée par Mujeeb. Ennifar a montré un motif 'purine zipper' dans la structure du dimère instable de la tige boucle SL1. En cristallographie, les 2 adénines de l'entrée de boucle A8 et A9 sont exclues de la double hélice et sont empilées l'une sur l'autre. Si ce motif est présent en solution dans notre étude, il doit se traduire par une rupture de l'attribution séquentielle des protons H1'-aromatiques entre le résidu G7 et A8. Ce passage existe pourtant sur nos spectres NOESY à 100ms dans D<sub>2</sub>O. On peut donc conclure que ce motif n'est pas présent en solution.

Il apparaît que la structure du dimère instable de la tige SL1 24mer non mutée ne présente pas les mêmes caractéristiques structurales que les dimères instables déjà publiées. La structure RMN du 'kissing complex' en solution a été effectuée sur une séquence fortement mutée dans

la tige (*GGCAAUGAAGCGCGCACGUUGCC*). Ces mutations engendrent une plus grande stabilité de la tige. Nous savons que la formation d'un complexe boucle-boucle est le résultat d'un équilibre entre les contraintes relatives à l'appariement des bases dans la tige et l'association des boucles. Toute modification qui vise à stabiliser l'édifice moléculaire risque donc d'entraîner des modifications sur la structure globale de la molécule. Cet équilibre subtil expliquerait pourquoi deux structures étudiées en solution par la même technique et sur des oligonucléotides dont la séquence est très proche possèdent des caractéristiques structurales très éloignées l'une de l'autre.

Dans le cas de la structure cristallographique du complexe boucle-boucle, il existe un motif 'purine zipper'. Ce motif présente un espace libre entre deux adénines qui permet l'intercalation de deux autres adénines provenant d'un autre dimère. La forme instable n'existerait donc pas sous forme de dimère mais de dimère de dimère dont nous avons prouvé l'inexistence en solution via les expériences de diffusion. La RMN permet d'accéder à la structure moyenne en solution, alors que l'étude cristallographique nous renseigne sur une structure unique sélectionnée lors de l'étape de cristallisation. Les conditions de cristallisation (concentration d'ARN élevée et haute force ionique) influent sur la discrimination entre les différents type de dimère. Il est vraisemblable que ce motif 'purine zipper' permet d'avoir une colinéarité du système et dirige la cristallisation de la forme cristalline.

Plusieurs travaux (Lodmell et al. 2001; Paillart et al. 1997) obtenus en solution, à basse concentration, semblent en accord avec notre structure. Il a été démontré que la position N1 de l'adénine A16 est réactive aux sondes chimiques alors que la position N7 du même résidu n'est pas réactive. La base de l'adénine A16 forme une liaison hydrogène entre son NH2 et le N6 de l'adénine A9 mais n'est pas appariée, laissant ainsi sa position N1 libre pour réagir avec les sondes chimiques. La non réactivité de la position N7 peut s'expliquer par le fait que cette adénine est empilée avec la guanine G10 de l'autre brin, empêchant ainsi l'accessibilité de cette position. On peut également méthyler la position N1 des adénines A8 et A9 indiquant que ces deux résidus ne sont pas impliqués dans des appariements de type Watson-Crick comme l'indique notre structure.

Comme nous l'avons vu dans la discussion de la partie monomérique de ce travail de thèse, il existe une très forte proportion de résidus puriques et plus particulièrement d'adénines dans les séquences d'aptamers capables de dimériser. Par conséquent, ces adénines doivent intervenir dans le cas de l'isolat Laï, soit dans la reconnaissance, soit dans la stabilisation du 'kissing complex'. Effectivement, ces adénines sont impliquées dans différentes interactions

visant à stabiliser l'édifice moléculaire. L'adénine A16 est stackée avec la guanine G10 de l'autre brin et forme des liaisons hydrogène avec l'adénine A9. L'adénine A8 forme également des liaisons hydrogène avec la cytosine C17 de l'autre brin.

## II-8 Conclusion

Même si depuis 10 ans les études de 'kissing complex' se multiplient, le nombre de structures décrites et publiées demeure faible. La nature de ces 'kissing complex' peut être un hétérodimère (ARN/ARN ou ARN/ADN) (Chang and Tinoco 1994, 1997; Collin et al. 2000; Lee and Crothers 1998) ou comme dans le cadre de notre étude un homodimère (Ennifar et al. 2001; Kim and Tinoco 2000; Mujeeb et al. 1998). La forme instable de l'extrémité de la tige boucle SL1 24mer de VIH-1<sub>lat</sub> a déjà fait l'objet de 2 publications, une par RMN sur une séquence mutée et une autre par cristallographie. Les structures sont différentes l'une de l'autre. Forts de nos connaissances sur le monomère et le dimère étendu, représentant la forme stable du dimère, nous avons poursuivi notre étude en étudiant la structure de la forme instable de la séquence sauvage.

Nous avons mis au point un protocole de purification d'ARN en grosse quantité pour pouvoir effectuer l'étude par RMN. Nous avons ensuite isolé sans ambiguïté le complexe boucle-boucle par un protocole d'hybridation basé sur la température et la concentration en soluté. La préparation de cet ARN a été réalisée d'une part sur une séquence non marquée, mais également sur une séquence doublement marquée uniformément <sup>13</sup>C et <sup>15</sup>N. Grâce à l'hybridation contrôlée, nous obtenons des complexes non marqués, totalement marqués ou hybrides. Une étude RMN homo-nucléaire et hétéro-nucléaire nous a permis d'obtenir une nouvelle structure de la forme instable différente des deux structures déjà publiées. Notre structure est en accord avec les données biologiques qui nous sont connues. Elle permet également de montrer l'importance de l'agencement des paires de bases de la tige dans la structure finale d'un complexe boucle-boucle. En effet, notre structure tridimensionnelle diverge totalement de celle obtenue par Mujeeb et al. alors que les séquences diffèrent de seulement quelques paires de bases dans la tige. Nous montrons que les trois adénines, situées à l'entrée de la boucle, participent à la stabilisation de l'édifice moléculaire. Ceci pourrait expliquer pourquoi elles apparaissent très souvent dans les séquences d'aptamers capables de dimériser.

# CONCLUSION et PERSPECTIVES

Le phénomène de dimérisation interfère avec différentes étapes clés du cycle rétroviral. Il est entre autre corrélé aux étapes d'encapsidation de l'ARN génomique et de maturation des particules virales responsables de l'infectiosité du virus. La séquence SL1 de l'ARN génomique est hautement conservée et peut adopter une structure en tige boucle avec une séquence autoccomplémentaire GCGCGC dans la boucle. Il a été montré (Muriaux et al. 1996a; Muriaux et al. 1996b) *in vitro* que la formation du dimère est initiée *via* une interaction boucle-boucle entre deux molécules d'ARN pour aboutir à un complexe boucle-boucle instable appelé 'kissing complex', qui sous l'effet de la température ou de la protéine de nucléocapside NCp7 se réarrange pour donner un autre complexe plus stable appelé duplexe étendu. Ce réarrangement est responsable de la maturation du virion et par conséquent de l'infectiosité de celui-ci. Il était donc essentiel de connaître la structure des différentes conformations de la séquence SL1 afin de mieux comprendre les différents mécanismes responsables de l'infectiosité du virion.

Pour connaître les caractéristiques qui favorisent l'interaction boucle-boucle en solution, notre premier travail a été la détermination des caractéristiques structurales et dynamiques qui interviennent lors de cette interaction. Notre première étude a consisté à étudier l'extrémité de la tige boucle SL1 (23 nucléotides) de l'ARN génomique de VIH-1 mutée en position 12 en remplaçant une guanine par une adénine. Cette mutation a pour avantage d'empêcher la dimérisation tout en conservant la nature purique de la base. Nous avons ainsi pu mettre en évidence une tige boucle dont la tige est composée de 7 paires de bases organisés en double hélice A. Ce type de double hélice est classiquement observé chez les ARN. Il existe également une région structurée en entrée de boucle correspondant aux résidus A8, A9, G10 et A16, empilés les uns par rapport aux autres. Enfin, une séquence apicale beaucoup moins structurée par rapport au reste de la molécule car soumise à une forte dynamique dont les résidus sont également largement exposés au solvant. Nous avons également pu effectuer une simulation de dynamique moléculaire avec solvant et ions permettant d'obtenir des informations sur les mouvements et le réseau de liaisons hydrogène, notamment avec les 2'OH, impliqués dans la stabilisation de l'édifice moléculaire.

La deuxième partie de notre travail a porté sur l'étude de la forme la moins stable du dimère de la tige boucle SL1 non mutée. Pour cette étude, nous avons mis au point la méthodologie pour préparer de l'ARN en grosse quantité (quelques milligrammes). Nous avons ainsi pu synthétiser par voie biochimique deux échantillons de l'extrémité de la tige boucle SL1



sauvage (24 nucléotides) : un échantillon non marqué et un échantillon marqué uniformément en  $^{13}\text{C}$  et  $^{15}\text{N}$ . Cette synthèse consiste à transcrire une matrice ADN en ARN par l'intermédiaire de la T7 polymérase préalablement préparée et purifiée. Nous avons mis au point un protocole d'hybridation de l'échantillon pour l'obtention de la forme la moins stable du dimère afin de permettre son étude par RMN à basse température ( $< 20^\circ\text{C}$ ). Nous avons pu mettre en évidence l'association boucle-boucle de chaque monomère. Nous pouvons également montrer par des expériences de diffusion que la forme instable est seulement présente sous une structure dimérique en solution. La structure cristallographique (Ennifar et al. 2001) de la forme instable de la même séquence possède un motif 'purine zipper'. Ce motif implique alors une structure sous la forme d'un tétramère. Nos résultats indiquent que la forme instable en solution diffère de celle obtenue dans le cristal. A partir des données NOESY et COSY, nous avons pu obtenir un ensemble d'informations structurales (distances et angles) sur la forme instable de la séquence SL1. Nous avons pu déterminer un ensemble de structures qui est cohérent avec nos contraintes expérimentales. Cet ensemble ne correspond pas aux différentes structures de dimère instable de SL1 déjà publiées (Ennifar et al. 2001; Mujeeb et al. 1998) dans la littérature. Nous avons confirmé que la structure de la forme instable est un dimère avec une interaction boucle-boucle entre les deux tiges boucles. La tige de chaque monomère se présente sous la forme d'une double hélice de type A avec 7 paires de bases appariées via des liaisons hydrogène de type Watson-Crick. La zone de jonction entre la tige et la séquence auto complémentaire de chaque monomère, comprenant 3 adénines, présente des empilements et des liaisons hydrogène entre ces bases et les bases environnantes. L'adénine A8 est tournée vers l'extérieur du monomère pour interagir avec la cytosine C17 de l'autre monomère via des liaisons hydrogène. L'adénine A9, quant à elle, reste tournée vers l'intérieur de la boucle et forme des liaisons hydrogène avec la guanine 10 et l'adénine 16 du même monomère. L'adénine A16 est également impliquée dans des phénomènes d'empilement avec la guanine G10 de l'autre monomère. Ces adénines interviennent directement dans la stabilisation de l'édifice moléculaire, ce qui pourrait expliquer pourquoi il existe une très forte proportion de résidus puriques et plus particulièrement d'adénines dans les séquences d'aptamers capables de dimériser (Lodmell et al. 2001).

Au sein de notre équipe, nous avons pu isoler et caractériser les différentes conformations que peut adopter l'extrémité de la séquence SL1 à savoir une forme monomérique, une forme dimérique instable résultant d'un complexe boucle-boucle et une deuxième forme dimérique étendue (Girard et al. 1999). Nous avons uniquement des connaissances sur des structures de

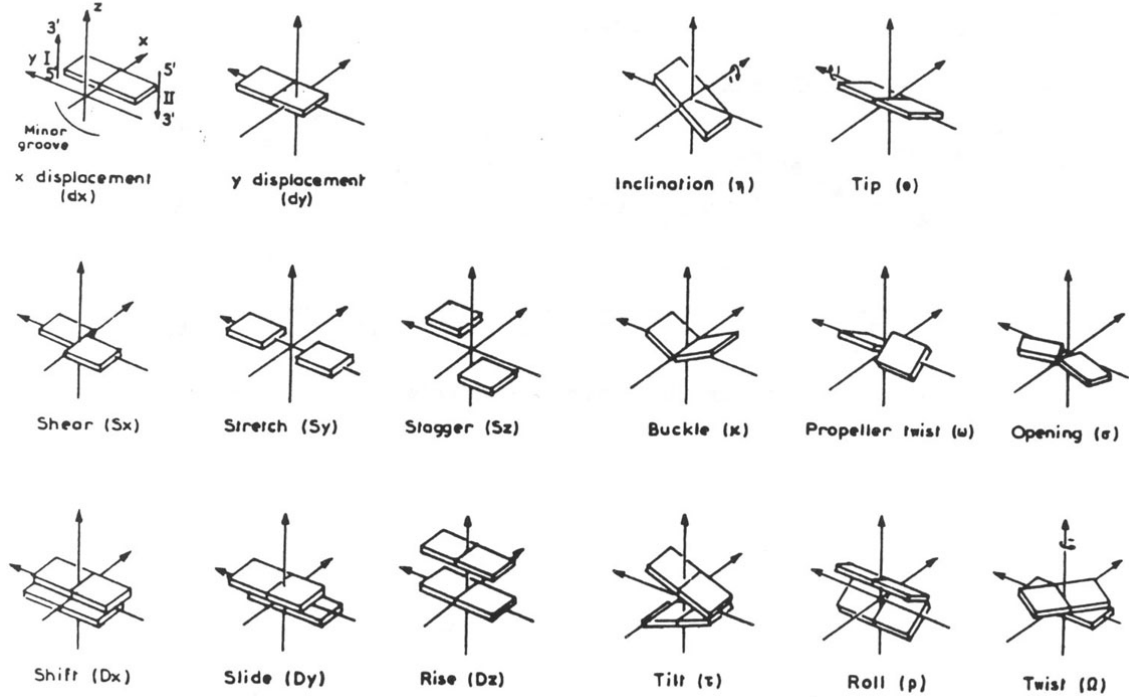
l'extrémité de la tige boucle SL1 en solution. Il serait intéressant de connaître l'interaction des différentes conformations de la tige boucle SL1 avec la protéine de NCp7. Bien que notre équipe ait déjà montré par RMN que l'interaction de la NCp7 avec le 'kissing complex' de SL1 Laï provoque un changement conformationnel vers le dimère étendu, l'étude oligo-protéique complète ne peut se faire que si l'un des deux partenaires est marqué  $^{13}\text{C}$  et  $^{15}\text{N}$ . Notre équipe dispose également des connaissances pour marquer les ARN. Il semble donc logique que la prochaine étape de ce travail consiste à étudier la séquence SL1 24mer marquée en interaction avec la protéine NCp7. Il est également envisageable d'étudier la totalité de la séquence SL1, 35 nucléotides, sous ses différentes conformations pour connaître l'importance du bulge en amont dans la stabilisation de l'édifice moléculaire. En effet, tout comme la séquence apicale auto-complémentaire, les résidus du bulge sont des résidus très bien conservés parmi les différents isolats de VIH (Kuiken et al. 2000). Si l'on effectue des mutations au niveau des résidus de ce bulge, cela se traduit par des défauts d'encapsidation (Clever and Parslow 1997). Plusieurs équipes indiquent que ce bulge est nécessaire pour le passage d'une forme dimérique à une autre (Lawrence et al. 2003; Takahashi et al. 2001). Bien que nous ayons démontré le contraire dans ce travail, on peut envisager que la tige complète facilite l'existence du 'kissing complex'. En effet, l'association boucle-boucle génère des tensions dans les conformations des résidus à la charnière tige boucle, tensions qui sont en partie compensée par un changement conformationnel de la tige qui se transmet de G7-C17 vers C1-G23. La présence d'un bulge pourrait stopper cette propagation des changements de configuration. On pourrait enfin essayer de connaître des interactions entre la tige boucle SL1 avec les autres tiges du domaine Psi d'encapsidation. En effet, les mécanismes de dimérisation et d'encapsidation de l'ARN génomique sont des mécanismes qui sont étroitement liés. Pour cela, on peut envisager de synthétiser de longs ARN comprenant les différentes tiges ou seuls certains résidus seraient marqués afin de connaître localement certaines interactions.

Ces différentes investigations structurales constituent une base d'étude dans les stratégies de drug design impliquant le développement d'inhibiteurs susceptibles d'interférer avec les sites d'interactions ARN/ARN.

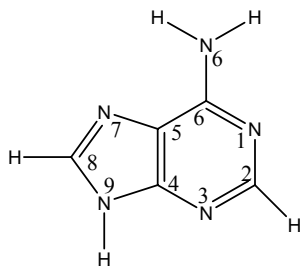
# ANNEXE

# Annexe

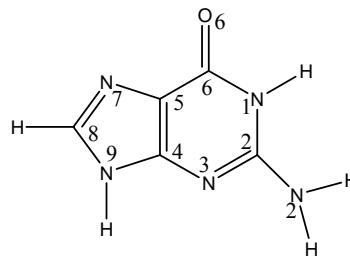
Définition des paramètres d'hélice.



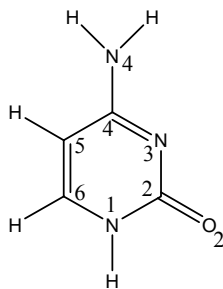
Structure et numérotation des bases nucléiques dans un ARN



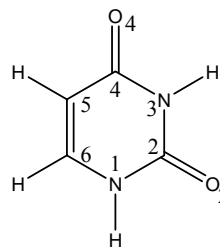
Adénine



Guanine



Cytosine



Uracile

# Matériels et méthodes

## Les études réalisées par RMN

### *Etude structurale du monomère de la tige boucle SL1\**

Les études ont été réalisées à 500MHz sur un spectromètre AMX BRUKER<sup>®</sup>. Les spectres dans H<sub>2</sub>O et D<sub>2</sub>O ont été obtenus à partir d'un échantillon de 500µL d'ARN à 2mM, en présence de NaCl (100mM) et de tampon phosphate (10mM, pH 5,7). Les spectres ont été enregistrés avec un nombre de points complexes de 800 et 1024 pour les dimensions t<sub>1</sub> et t<sub>2</sub> respectives. Toutes les expériences dans D<sub>2</sub>O ont été effectuées avec une largeur spectrale de 4502 Hz et 4002Hz pour les dimensions F<sub>1</sub> et F<sub>2</sub> respectives. La résonance de l'eau résiduelle a été irradiée par une pré-saturation faible de 1s. Pour les expériences dans H<sub>2</sub>O, la largeur spectrale est de 11112Hz et 11003Hz dans les dimensions F<sub>1</sub> et F<sub>2</sub> et la dernière impulsion  $\pi/2$  est remplacée par une excitation sélective de type 'Jump and Return' (Plateau and Gueron 1982) pour éliminer le signal de l'eau. Les séquences 2D de type DQF-COSY (Aue et al. 1976; Marion and Wuthrich 1983), TOCSY (Braunschweiler and Ernst 1983; Davis and Bax 1985) ou NOESY (Macura et al. 1981; Macura et al. 1982) ont été enregistrées en mode STATES (Marion et al. 1989). Les spectres NOESY ont été enregistrés à différents temps de mélange 0, 50, 100, 200 et 450ms à 27°C dans D<sub>2</sub>O et 80 et 150ms dans H<sub>2</sub>O à 9°C avec 2 accumulations et un temps de retour à l'équilibre de 2s.

### *Etude structurale du complexe boucle-boucle de SL1-24mer VIH-1<sub>Lai</sub>*

Les études dans l'eau ont été réalisées sur un spectromètre VARIAN<sup>®</sup> INOVA à 600MHz sur un échantillon d'ARN à 2mM pour l'oligonucléotide non marqué et 0,6 mM pour l'échantillon marqué dans un tube Shigemi<sup>®</sup> de 300µl. Le pH de la solution est ajusté à 5,7 par ajout de NaOH ou HCl. L'échantillon ne contient aucun sel et aucun tampon.

### *Spectre sur l'échantillon non marqué*

Les spectres ont été enregistrés avec un nombre de points complexes de 400 et 2368 pour les dimensions t<sub>1</sub> et t<sub>2</sub> respectives. Toutes les expériences dans D<sub>2</sub>O ont été effectuées avec une largeur spectrale de 6000 Hz et 6000Hz pour les dimensions F<sub>1</sub> et F<sub>2</sub> respectives. La résonance de l'eau résiduelle a été irradiée par une pré-saturation faible de 1s. Pour les expériences dans H<sub>2</sub>O, la largeur spectrale est de 12999Hz et 13008Hz dans les dimensions

F1 et F2. Le signal de l'eau est éliminé par une séquence 'WATERGATE'(Piotto et al. 1992). Les séquences 2D de type DQF-COSY, TOCSY ou NOESY ont été enregistrés en mode STATES. Les spectres NOESY ont été enregistrés à quatre temps de mélange 50, 100, 150 et 300ms à 20°C dans D<sub>2</sub>O et 80 et 150 ms dans H<sub>2</sub>O à 10°C avec 64 accumulations et un temps de retour à l'équilibre de 1s dans H<sub>2</sub>O et de 3s dans D<sub>2</sub>O.

### *Spectre sur l'échantillon marqué <sup>13</sup>C et <sup>15</sup>N*

Les spectres ont été enregistrés avec un nombre de points complexes de 256 et 1216 pour les dimensions t1 (hétéroatome) et t2 (proton) pour un spectre <sup>15</sup>N-<sup>1</sup>H et de 256 et de 1472 pour un spectre <sup>13</sup>C-<sup>1</sup>H. La largeur spectrale classiquement utilisée pour des spectres dans la dimension de l'hétéro-atome est de 6080Hz pour le <sup>15</sup>N et de 9100 Hz pour le <sup>13</sup>C. Les expériences HSQC <sup>13</sup>C et <sup>15</sup>N comportent classiquement 32 ou 128 accumulations pour chacun des noyaux respectifs et un temps de retour à l'équilibre de 1s. La détection en quadrature sur la dimension indirecte utilise la méthode States-TPPI et les impulsions composites pour découpler durant l'acquisition est de type WALTZ-16 (Shaka et al. 1983) avec une puissance de 44dB(pour le <sup>13</sup>C) et GARP (Shaka et al. 1985) avec une puissance de 45dB( pour le<sup>15</sup>N) (la puissance maximale délivrée par le spectromètre est de 63dB). La valeur du  $\pi/2$  est de 36,8 $\mu$ s pour le <sup>15</sup>N et de 15,2 $\mu$ s pour le <sup>13</sup>C avec une puissance de 59dB.

## Préparation de la T7 polymérase

L'ARN polymérase T7 est isolée à partir de la souche d'*Escherichia coli* BL 21 contenant le plasmide pAR1219, fournie par l'équipe de M. Kochoyan. Suite à de nombreux problèmes lors de la transcription, nous avons suivi un protocole de purification de l'enzyme modifié par rapport à celui décrit classiquement par (Davanloo et al. 1984).

### *Culture des cellules*

Deux milieux de culture de BL21/pAR1219 de 500 ml sont été incubés à 37°C après ensemencement au 500<sup>ème</sup> par un inoculum, qui a préalablement poussé dans des erlens de 2L contenant un milieu enrichi, dans le cas présent LB (Luria-Bertani) (toutes les solutions sont décrites à la fin du paragraphe), en présence d'ampicilline (40ug/mL). Lorsque les cultures se trouvent bien en phase de croissance exponentielle et présentent une absorbance à 600nm comprise entre 0,5 et 0,8, on ajoute une concentration de  $\beta$ -D-thio-galactopyranoside (IPTG) afin d'obtenir une concentration finale de 1mM. L'IPTG permet de 'dé-réprimer' le promoteur Lac UV5, induisant l'accumulation de la T7 ARN polymérase dans les cellules. Nous avons ensuite testé les meilleures conditions de température et de durée d'induction. Les meilleurs résultats ont été obtenus en effectuant une induction toute la nuit à 25°C. Ces conditions permettent d'obtenir les quantités d'enzyme optimums et de laisser à la protéine le temps de se replier correctement. Après vérification de l'induction sur un gel de protéine (voir matériels et méthodes), les cellules sont ensuite récupérées par centrifugation (4000tr/min, 10min, 4°C). Toujours à 4°C, les cellules sont lavées deux fois avec 100 mL de tampon A. Les cellules sont reprises dans 10mL de tampon B et placées à -20°C pour y être congelées.

### *Lyse des cellules*

Afin d'éviter l'utilisation de lysosyme qui est reconnu pour être un inhibiteur de la T7 ARN polymérase (Ikeda and Bailey 1992) nous avons choisi d'effectuer la lyse des cellules par sonification. Pour cela, nous avons ajouté au culot bactérien congelé 10 mL de Tampon B contenant 2 inhibiteurs de protéase, 20 $\mu$ L de leupeptine à 5mg/mL et 50 $\mu$ L de PMSF (phenylmethylsulfonylfluoride) à 20 $\mu$ g/mL. Toutes ces étapes sont effectuées à 4°C.



Le culot bactérien congelé avec la solution de tampon B est sonifié à 4°C à l'aide d'un sonificateur Branson<sup>®</sup> avec des séries d'impulsions de 3 secondes d'une puissance de 50 durant 1 minute. Entre chaque série d'impulsion, on mélange bien le glaçon bactérien avec la solution pour permettre à l'ensemble de bien rester à 4°C. Nous recommençons la sonification jusqu'à disparition du glaçon. On ajoute ensuite 5 mL de (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 2M et on complète jusqu'à 50 mL par du tampon B.

#### *Précipitation à la Polymine P*

On additionne ensuite 5 mL de Polymine P 10% (ajusté à pH 8,0 par addition de HCl concentré) lentement en agitant dans la glace. Après 20 min, le précipité formé contenant les acides nucléiques est éliminé par centrifugation (14000rpm, 15min, 4°C).

#### *Précipitation des protéines par (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> saturé*

Le surnageant est mélangé lentement dans la glace avec 45mL de sulfate d'ammonium saturé. Après 15 min d'agitation dans la glace, le précipité est récupéré par centrifugation (6000 rpm, 10min, 4°C) et récupéré dans 15 mL de tampon C et mis à dialyser toute la nuit contre un litre de tampon C additionné de 100mM NaCl. Le dialysât est ensuite centrifugé (4000 rpm, 10min, 4°C). On récupère le surnageant et on vérifie sa conductivité. Celle-ci doit être identique à celle du tampon C contenant 50 mM NaCl. Si ce n'est pas le cas on ajuste la conductivité avec du tampon C.

#### *Purification de la T7 ARN polymérase*

La dernière étape de la préparation de l'enzyme consiste à effectuer une colonne échangeuse de cations, Trisacryl-SP(SIGMA<sup>®</sup>). Pour cela, nous réalisons une colonne de 25 mL. Nous enlevons les éventuelles impuretés avec du tampon C avec un fort pourcentage en NaCl (200mM). Nous équilibrons ensuite la colonne avec 4 fois son volume en tampon C avec 25mM NaCl. Nous chargeons l'échantillon et nous lavons de nouveau la colonne avec 4 fois son volume de tampon C contenant 50mM NaCl pour décrocher un maximum de protéines ne correspondant pas à notre enzyme. L'enzyme est ensuite éluée avec du tampon C contenant 200mM NaCl. Les fractions présentant les absorbances maximales sont analysées sur gel d'électrophorèse (voir matériels et méthodes). Ces fractions sont regroupées et concentrées

par dialyse contre 1L de solution contenant 50% de tampon C à 100mM NaCl et 50% de glycérol pendant 48H. On récupère l'enzyme qui est aliquotée et stockée à -20°C.

Solutions :

LB (Sigma®):

20 g dans un litre d'eau

Tampon A :

Tris 1M pH 8.1	10mL	final 20mM	
NaCl 4M	2.5mL	final 20mM	
EDTA 500mM	2mL	final 2mM	
QSP H2O UP 500mL			à autoclaver

Tampon B :

Tris 1M pH 8.1	15mL	final 50mM	
NaCl 4M	1.5mL	final 20mM	
EDTA 500mM	1.2mL	final 2mM	
QSP H2O UP 300mL			à autoclaver

Avant utilisation

DTT 1M	300µL	final 1mM	
--------	-------	-----------	--

Tampon C :

Na2HPO4	10.72g		
EDTA 500mM	4mL		
Glycérol	100mL		
QSP H2O UP 2L			à autoclaver

Ajuster le pH à 7.7 avec HCl fumant

Avant utilisation

DTT 1M		final 1mM	
PMSF 20mg/mL		final 20µg/mL	

## Synthèse d'ARN par voie biochimique

La synthèse des ARN est réalisée par transcription à partir d'une matrice ADN simple brin par la T7 ARN polymérase, selon la méthodologie décrite par (Milligan et al. 1987) et (Wyatt et al. 1991). Cette transcription nécessite également un ADN dit promoteur qui se fixe en 3' de l'ADN matrice pour former un double brin nécessaire à l'activité de la T7 ARN polymérase.

### *Transcription de l'ARN*

Nous avons utilisé différentes conditions pour réaliser nos transcriptions. Nous avons pu mettre en évidence qu'au delà de 3h de transcription, les rendements n'étaient pas améliorés. Il est également préférable d'hybrider ensemble la matrice ADN, le promoteur ADN en chauffant à 90°C pendant 3 min et ensuite à 4°C pendant 3 min, plutôt que de les ajouter séparément à la transcription. Un excès d'enzyme lors de la transcription ne permet pas obligatoirement d'avoir de meilleurs rendements. De plus, la quantité de magnésium est essentielle pour les quantités d'ARN transcrites. Il convient d'effectuer avant toute production importante des transcriptions tests pour connaître la concentration optimum en magnésium (0,5 à 2,5 équivalent par rapport à la concentration globale de rXTP).

Les conditions générales pour la transcription d'ARN sont décrites dans le tableau 85 pour un échantillon non marqué et dans le tableau 86 pour l'échantillon marqué doublement  $^{13}\text{C}/^{15}\text{N}$ , pour un volume final de transcription de 100 $\mu\text{l}$ . Pour de plus gros volumes, les quantités utilisées sont simplement extrapolées.

Le réaction enzymatique est réalisée à 37°C pendant 3h. On additionne ensuite 1 volume d'EDTA (500mM, pH 8,0) pour 10 volumes de transcrit. On contrôle ensuite le rendement de la réaction à l'aide d'un gel d'électrophorèse analytique (polyacrylamide 20%/Urée 7M) suivi d'une étape de coloration au stain-all (voir matériels et méthodes).

solution	volume	concentration finale
Tampon 10X: 400mM Tris pH 8,1 10mM spermidine 0,1% (v/v) triton X-100 50mM DTT	10 $\mu$ L	1X 40mM Tris pH 8,1 1mM spermidine 0,01% (v/v) triton X-100 5mM DTT
H2O QSP	15,2 $\mu$ L	
PEG 8000 (500mg/ml)	20 $\mu$ L	100 mg/mL
rATP (100mM)	4 $\mu$ L	4 mM
rCTP (100mM)	4 $\mu$ L	4 mM
rGTP (100mM)	4 $\mu$ L	4 mM
rUTP (100mM)	4 $\mu$ L	4 mM
Mgcl2 (100 mM)	28,8 $\mu$ L	28,8 mM
T7 polymérase (~)	4 $\mu$ L	
ADN matrice (10uM)+ ADN promoteur (10uM)	3 $\mu$ L 3 $\mu$ L	300 nM 300 nM

chauffé à 90°C 3min, 4°C 3min

**Tableau 85.** Conditions générales de transcription d'ARN non marqué pour un volume réactionnel de 100  $\mu$ L.

solution	volume	concentration finale
Tampon 10X: 400mM Tris pH 8,1 10mM spermidine 0,1% (v/v) triton X-100 50mM DTT	10 $\mu$ L	1X 40mM Tris pH 8,1 1mM spermidine 0,01% (v/v) triton X-100 5mM DTT
H2O QSP	37,6 $\mu$ L	
PEG 8000 (500mg/ml)	20 $\mu$ L	100 mg/mL
rATP (100mM)	2 $\mu$ L	2 mM
rCTP (100mM)	2 $\mu$ L	2 mM
rGTP (100mM)	2 $\mu$ L	2 mM
rUTP (100mM)	2 $\mu$ L	2 mM
Mgcl2 (100 mM)	6,4 $\mu$ L	6,4 mM
T7 polymérase (~)	4 $\mu$ L	
ADN matrice (10uM)+ ADN promoteur (10uM)	3 $\mu$ L 3 $\mu$ L	300 nM 300 nM

chauffé à 90°C 3min, 4°C 3min

**Tableau 86.** Conditions générales de transcription d'ARN marqué  $^{13}\text{C}/^{15}\text{N}$  pour un volume réactionnel de 100  $\mu$ L. (les XTP doublement marqués  $^{13}\text{C}/^{15}\text{N}$  ont été synthétisés par Silantes<sup>®</sup>)

### *Purification des ARN*

Le milieu réactionnel est extrait au phénol:chloroforme:alcool isoamylique (25:24:1), volume à volume, 2 fois de suite; les phases aqueuses sont finalement extraites par un mélange chloroforme:alcool isoamylique (24:1) équilibré avec du tampon TE (un volume pour cinq volumes de solution). L'ARN est mis à précipiter durant la nuit à -20°C en présence d'acétate de sodium (300mM, pH 5,2) et d'éthanol (3 volumes pour 1 volume de solution). On

centrifuge (13000tr/min, 1H, 4°C), on sèche le culot au speed-vac pendant 15 min et on reprend le culot avec un mélange eau/urée8M (v/v).

On purifie la séquence d'ARN sur gel dénaturant de polyacrylamide 20% / urée (7M). On repère par UV shadowing la bande d'ARN majoritaire correspondant à la séquence désirée et on la découpe. L'ARN est extrait des bandes par diffusion dans l'eau sous agitation durant la nuit. On change de temps en temps l'eau pour la diffusion. On vérifie que les bandes ne contiennent plus d'ARN par UV shadowing. L'eau contenant l'ARN est ensuite concentrée par évaporation sous pression réduite. Pour enlever l'urée et les éventuelles traces de sels dans notre échantillon, nous avons effectué une colonne C18 en phase inverse. Ce type de purification (voir matériels et méthodes) a pu être effectué sur une colonne C18 de type HPLC (Beckman<sup>®</sup> ultrasphère 5µm) ou sur un système SPE (solid phase extraction) utilisant des cartouches de C18 Se-pak (Waters<sup>®</sup>). Il est nécessaire de rajouter au préalable dans l'échantillon 10mM tampon phosphate pH 5,8 afin de ne pas détériorer la colonne phase inverse au cours de la purification. Les fractions récupérées sont évaporées sous pression réduite pour enlever les traces de solvant organique et ensuite lyophilisées. Le produit lyophilisé est ensuite repris dans 3 mL et passé ensuite sur une colonne de gel filtration DG 10 (Biorad<sup>®</sup>) (voir matériels et méthodes). Cette colonne permet de compléter la colonne C18 en éliminant les impuretés possédant un poids moléculaire inférieur à 3000 Da. Les fractions contenant l'ARN sont récupérées et lyophilisées. Le produit ainsi obtenu peut être repris dans de l'eau pour son étude structurale par RMN.

Le rendement en ARN non marqué est de 20 DO pour 10 ml de transcription et en ARN marqué de 7 DO pour 10 ml de transcription.

## GELS

### *Gel dénaturant polyacrylamide 20%/ urée 8M*

L'échantillon lors du chargement doit contenir au moins 50 % urée ou 70% de formamide.

L'échantillon est dénaturé à 90°C pendant 3min et mis ensuite à 4°C pendant la glace.

	<i>Analytique</i>	<i>Préparatif</i>
Acryl/Urée 25%	5,6mL	56 mL
TBE 10X	0,7 mL	7 mL
Urée 8M	0,53 mL	5,3 mL
APS 10%	50 µL	500 µL
Temed	5 µL	50 µL
Migration	TBE 1X 300V, 10W	TBE 1X 450V, 30W
Prérin	15 min	60 min

Révélation: 1) 5 min dans le BET suivi de 2 lavages de 5 min dans l'eau. La révélation s'effectue sur UV.

2) 15 min dans le stain all suivi d'un lavage avec de l'eau. La décoloration se fait grâce à la lumière blanche, ensuite on sèche le gel sous pression réduite.

### *Gel natif polyacrylamide 15 %*

L'échantillon lors du chargement doit contenir au moins 50 % de glycérol.

	<i>Analytique</i>	<i>Préparatif</i>
Acrylamide/bisacrylamide(38 :2)	2,8 mL	28 mL
TBE 10X	0,75 mL	7,5 mL
H2O	4 mL	40 mL
APS 10%	75 µL	750 µL
Temed	7,5 µL	75 µL
Migration	TBE 1X 250V, 10W	TBE 1X 250V, 10W

Révélation: 5 min dans le BET suivi de 2 lavages de 5 min dans l'eau. La révélation s'effectue sur UV.

### *Gel de protéine (SDS PAGE)*

L'échantillon (1 ou 2 µL) doit être mélangé avec 10µL de bleu de charge (2X LSB). Avant injection, l'ensemble est chauffé à 90°C pendant 3 min et mis ensuite à 4°C dans la glace.

#### Gel de stacking 5%

Acrylamide/bisacrylamide(38 :2)	1,3 mL
Tampon Top	2,5 mL
H2O	6,2 mL
APS 10%	100 µL
Temed	20 µL

#### Gel de séparation 10%

Acrylamide/bisacrylamide(38 :2)	2,6 mL
Tampon Bottom	2,5 mL
H2O	4,9 mL
APS 10%	100 µL
Temed	20 µL

Migration	Running buffer 1X
	200V, 10W

Révélation : 15 min avec du bleu de coomasie. La décoloration s'effectue en plusieurs lavages avec de l'eau chaude.

#### *Tampon pour le gel*

##### **Bleu/Urée**

420mg d'urée ultrapure  
10µL de bleu de bromophénol (2.5%)  
10µL de xylène cyanol (2.5%)  
On complète à 1mL avec H2Oqsp.

##### **Bleu/Glycérol 50%**

500µL H2O  
500µL glycérol  
20µL bleu de bromophénol (2.5%)  
20µL xylène cyanol

**Bleu/Formamide**

98% formamide désionisé, 10mM EDTA, 0.025% xylène cyanol, 0.025% bleu de bromophénol:

9.8mL formamide ultrapure (SERVA) désionisé 5min dans bain marie à ultra sons

200µL EDTA 0.5

25µL xylène cyanol (10%)

100µL bleu de bromophénol (2.5%)

**Bleu de bromophénol 2.5%**

250mg Bleu de bromophénol

On complète à 10mL avec H<sub>2</sub>O qsp.

ajouter 100µL NaOH 1M pour neutraliser (passage rouge à bleu).

**BET (bromure d'éthidium)**

solution à 10mg/mL

DANGER: porter gants + masque!

**Polyacrylamide/8M urée 25%**

480.48g d'urée en poudre

625mL acrylamide/bisacrylamide rotophorèse 40

mise en solution à 50°C sous agitation, puis filtrer sur 0.44µm

**Stain-all:** 1-ethyl-2-[3-(1-ethylnaphto[1,2-d]thiazolin-2-ylidene)-2-methylpropenyl]naphto-[1,2-d]thiazolium bromide

0.25g de stain-all

500mL DMF 50%

dissoudre d'abord le stain all dans le DMF puis ajouter l'eau.

ATTENTION ne pas inhaler le DMF

**TBE 10X:** 890mM acide borique, 890mM Trizabase, 25mM EDTA

55g acide borique

108g Tris

9.3g EDTA

on complète à 1L avec H<sub>2</sub>O qsp.

**Urée 8M**

480.48g urée (MW 60.06g/mol)

on complète à 1L avec H<sub>2</sub>O qsp sous agitation à 50°C, ensuite on filtre sur 0.45µm.

**SDS 10%** : sodium dodécyl sulfate

10g SDS

on complète à 100 ml avec H<sub>2</sub>O qsp.

**Tampon 4X Bottom**

181,6g Trisma base

4g Sodium dodecyl sulphate (SDS)

On ajuste à pH 8,8 avec HCl concentré et on complète à 1L avec H<sub>2</sub>O qsp.



**Tampon Top**

30,28g Trisma base

15 ml HCl concentré

450 ml H<sub>2</sub>O qsp

On ajuste à pH 6,8 avec HCl concentré et on complète à 500 ml avec H<sub>2</sub>O qsp.

**Running Buffer**

60,56g Trisma base

288,5g Glycine

20g SDS

on complète à 2 L avec H<sub>2</sub>O qsp.

**2X LSB**

6,25 ml Tris.HCl (0,5M pH 6,8)

2g SDS

9 ml Glycerol

3 ml 2-Mercaptoéthanol

20mg Bleu de bromophenol

on complète à 50 ml avec H<sub>2</sub>O qsp.

# Chromatographie

## *Séparation sur C18 beckman<sup>®</sup> ultrasphère 5 µm (HPLC)*

Il s'agit d'une colonne semi-préparative, le débit utilisé classiquement utilisé est de 3ml/min. Toutes les solutions utilisées sur un système doivent être parfaitement dégazées et filtrées même l'échantillon.

On lave au préalable la colonne avec ACN (acétonitrile) 100% pour enlever les éventuelles impuretés.

On équilibre ensuite la colonne en passant une solution à 10 mM de tri-éthyl ammonium acétate (TEAA).

On charge ensuite l'échantillon dans la colonne. Pour cela, on a deux possibilités. La première consiste à charger par la boucle d'injection, la deuxième directement par la pompe quand on a un grand volume à charger.

On lave l'échantillon avec une solution à 10 mM de TEAA. Il ne faut surtout pas passer un solvant organique directement sur la colonne contenant l'échantillon au risque de voir précipiter les sels si ce dernier contenait des sels.

On décroche l'ARN par un gradient H<sub>2</sub>O/ACN de 100% H<sub>2</sub>O à 80/20 H<sub>2</sub>O/ACN en 20 min. La détection de l'ARN se fait à 280 nm pour éviter de saturer le détecteur UV.

Une fois l'échantillon récupéré, on lave la colonne avec 100% ACN et avec 100% MeOH.

## *Séparation sur C18 Sé-pak Waters<sup>®</sup> (SPE)*

A l'aide d'une seringue de 10 mL, on fait passer 10 mL ACN pour hydrater et laver la cartouche contenant la C18.

On équilibre la cartouche avec 10 mL H<sub>2</sub>O et 2 mL de TEAA (0,5M) pour masquer les fonctions silanols résiduelles.

On charge l'échantillon (volume total inférieur à 20 mL).

On enlève les sels en faisant passer 10 à 15 mL d'eau.

On élue finalement l'ARN par 3 mL MeOH/H<sub>2</sub>O (60/40)

Le méthanol est ensuite évaporé sous pression réduite.

# BIBLIOGRAPHIE

- Abdelkafi M, Leulliot N, Baumruk V, Bednarova L, Turpin PY, Namane A, Gouyette C, Huynh-Dinh T, Ghomi M (1998) Structural features of the UCCG and UGCG tetraloops in very short hairpins as evidenced by optical spectroscopy. *Biochemistry* 37:7878-84
- Abragam A (1961) Principles of nuclear magnetism. Carendon Press, Oxford
- Aldovini A, Young RA (1990) Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J Virol* 64:1920-6
- Amarasinghe GK, De Guzman RN, Turner RB, Chancellor KJ, Wu ZR, Summers MF (2000a) NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J Mol Biol* 301:491-511
- Amarasinghe GK, De Guzman RN, Turner RB, Summers MF (2000b) NMR structure of stem-loop SL2 of the HIV-1 psi RNA packaging signal reveals a novel A-U-A base-triple platform. *J Mol Biol* 299:145-56
- Anderson JA, Teufel RJ, 2nd, Yin PD, Hu WS (1998) Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination. *J Virol* 72:1186-94
- Aue WP, Bartholdi E, Ernst RR (1976) *J Chem Phys* 64:2229-2246
- Awang G, Sen D (1993) Mode of dimerization of HIV-1 genomic RNA. *Biochemistry* 32:11453-7
- Baba S, Takahashi K, Nomura Y, Noguchi S, Koyanagi Y, Yamamoto N, Takaku H, Kawai G (2001) Conformational change of dimerization initiation site of HIV-1 genomic RNA by NCp7 or heat treatment. *Nucleic Acids Res Suppl*:155-6
- Balakrishnan M, Fay PJ, Bambara RA (2001) The kissing hairpin sequence promotes recombination within the HIV-1 5' leader region. *J Biol Chem* 276:36482-92
- Banks JD, Linial ML (2000) Secondary structure analysis of a minimal avian leukosis-sarcoma virus packaging signal. *J Virol* 74:456-64
- Baudin F, Marquet R, Isel C, Darlix JL, Ehresmann B, Ehresmann C (1993) Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J Mol Biol* 229:382-97
- Bender W, Davidson N (1976) Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. *Cell* 7:595-607
- Berg JM (1986) Potential metal-binding domains in nucleic acid binding proteins. *Science* 232:485-7
- Berkhout B, van Wamel JL (2000) The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *Rna* 6:282-95
- Berkowitz RD, Luban J, Goff SP (1993) Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J Virol* 67:7190-200
- Bernacchi S, Mely Y (2001) Exciton interaction in molecular beacons: a sensitive sensor for short range modifications of the nucleic acid structure. *Nucleic Acids Res* 29:E62-2
- Bieth E, Gabus C, Darlix JL (1990) A study of the dimer formation of Rous sarcoma virus RNA and of its effect on viral protein synthesis in vitro. *Nucleic Acids Res* 18:119-27
- Bonnet-Mathoniere B, Girard PM, Muriaux D, Paoletti J (1996) Nucleocapsid protein 10 activates dimerization of the RNA of Moloney murine leukaemia virus in vitro. *Eur J Biochem* 238:129-35
- Borer PN, Pappalardo L, Kerwood DJ, Pelczer I (1997) NMR-based structure determination for unlabelled RNA and DNA. *Adv Biophys Chem* 6:173-216

- Braunschweiler L, Ernst RR (1983) Coherence transfer by isotopic mixing: application to proton correlation spectroscopy. *J Magn Reson* 53:521-528
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J Comp Chem* 4:187-217
- Brünger AT (1992) X-PLOR: a system for X-Ray crystallography and NMR. Yale University Press
- Brünger AT, Adams PD, Clore GM, De Lano WL, Gros P, Grosse-Kuntzele RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905-21
- Bukrinsky MI, Haggerty S, Dempsey MP, Sharova N, Adzhubel A, Spitz L, Lewis P, Goldfarb D, Emerman M, Stevenson M (1993a) A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature* 365:666-9
- Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, Stevenson M (1993b) Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proc Natl Acad Sci U S A* 90:6125-9
- Canaani E, Helm KV, Duesberg P (1973) Evidence for 30-40S RNA as precursor of the 60-70S RNA of Rous sarcoma virus. *Proc Natl Acad Sci U S A* 70:401-5
- Case DA, Pearlman DA, Caldwell JW, Cheatham TE, Ross WS, Simmerling CL, Darden TA, Merz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Ferguson DM, Radmer RJ, Seibel GL, Singh UC, Weiner PK, Kollman PA (1997) AMBER 5. University of California
- Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273:1678-85
- Chang KY, Tinoco I, Jr. (1994) Characterization of a "kissing" hairpin complex derived from the human immunodeficiency virus genome. *Proc Natl Acad Sci U S A* 91:8705-9
- Chang KY, Tinoco I, Jr. (1997) The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement. *J Mol Biol* 269:52-66
- Chang KY, Varani G (1997) Nucleic acids structure and recognition. *Nat Struct Biol* 4 Suppl:854-8
- Cheong C, Varani G, Tinoco I, Jr. (1990) Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature* 346:680-2
- Clavel F, Orenstein JM (1990) A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J Virol* 64:5230-4
- Clever J, Sasseti C, Parslow TG (1995) RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J Virol* 69:2101-9
- Clever JL, Mirandar D, Jr., Parslow TG (2002) RNA structure and packaging signals in the 5' leader region of the human immunodeficiency virus type 1 genome. *J Virol* 76:12381-7
- Clever JL, Parslow TG (1997) Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol* 71:3407-14
- Clever JL, Wong ML, Parslow TG (1996) Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J Virol* 70:5902-8
- Clore GM, Starich MR, Gronenborn AM (1998) Measurement of residual dipolar couplings of macromolecules aligned in a nematic phase of a colloidal suspension of rod-shaped viruses. *J Am Chem Soc* 120:10571-2

- Coffin JM (1997) Retroviruses. Cold Spring Harbor Laboratory Press, New York, pp 1-69; 121-334
- Collin D, van Heijenoort C, Boiziau C, Toulme JJ, Guittet E (2000) NMR characterization of a kissing complex formed between the TAR RNA element of HIV-1 and a DNA aptamer. *Nucleic Acids Res* 28:3386-91
- Colvin RA, White SW, Garcia-Blanco MA, Hoffman DW (1993) Structural features of an RNA containing the CUGGGA loop of the human immunodeficiency virus type 1 trans-activation response element. *Biochemistry* 32:1105-12
- Cornille F, Mely Y, Ficheux D, Savignol I, Gerard D, Darlix JL, Fournie-Zaluski MC, Roques BP (1990) Solid phase synthesis of the retroviral nucleocapsid protein NCp10 of Moloney murine leukaemia virus and related "zinc-fingers" in free SH forms. Influence of zinc chelation on structural and biochemical properties. *Int J Pept Protein Res* 36:551-8
- Covey SN (1986) Amino acid sequence homology in gag region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. *Nucleic Acids Res* 14:623-33
- Cromsigt J, Schleucher J, Gustafsson T, Kihlberg J, Wijmenga S (2002) Preparation of partially <sup>2</sup>H/<sup>13</sup>C-labelled RNA for NMR studies. Stereo-specific deuteration of the H5" in nucleotides. *Nucleic Acids Res* 30:1639-45
- Cromsigt J, van Buuren B, Schleucher J, Wijmenga S (2001) Resonance assignment and structure determination for RNA. *Methods Enzymol* 338:371-99
- Dannull J, Surovoy A, Jung G, Moelling K (1994) Specific binding of HIV-1 nucleocapsid protein to PSI RNA in vitro requires N-terminal zinc finger and flanking basic amino acid residues. *Embo J* 13:1525-33
- Dardel F, Marquet R, Ehresmann C, Ehresmann B, Blanquet S (1998) Solution studies of the dimerization initiation site of HIV-1 genomic RNA. *Nucleic Acids Res* 26:3567-71
- Darlix JL, Gabus C, Nugeyre MT, Clavel F, Barre-Sinoussi F (1990) Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J Mol Biol* 216:689-99
- Davanloo P, Rosenberg AH, Dunn JJ, Studier FW (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc Natl Acad Sci U S A* 81:2035-9
- Davis DG, Bax A (1985) Assignment of complex <sup>1</sup>H NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy. *J Am Chem Soc* 107:2820-2821
- De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* 279:384-8
- De Rocquigny H, Ficheux D, Gabus C, Allain B, Fournie-Zaluski MC, Darlix JL, Roques BP (1993) Two short basic sequences surrounding the zinc finger of nucleocapsid protein NCp10 of Moloney murine leukemia virus are critical for RNA annealing activity. *Nucleic Acids Res* 21:823-9
- de Rocquigny H, Ficheux D, Gabus C, Fournie-Zaluski MC, Darlix JL, Roques BP (1991) First large scale chemical synthesis of the 72 amino acid HIV-1 nucleocapsid protein NCp7 in an active form. *Biochem Biophys Res Commun* 180:1010-8
- De Rocquigny H, Gabus C, Vincent A, Fournie-Zaluski MC, Roques B, Darlix JL (1992) Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc Natl Acad Sci U S A* 89:6472-6
- De Tapia M, Metzler V, Mougel M, Ehresmann B, Ehresmann C (1998) Dimerization of MoMuLV genomic RNA: redefinition of the role of the palindromic stem-loop H1

- (278-303) and new roles for stem-loops H2 (310-352) and H3 (355-374). *Biochemistry* 37:6077-85
- Delepierre M, Roques BP, Langlois D'Estaintot B, Igolen J (1988) A general procedure for assigning the 31P spectra of drug-oligonucleotide complexes. *J Biomol Struct Dyn* 5:933-7
- Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S, Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR, Landau NR (1996) Identification of a major co-receptor for primary isolates of HIV-1. *Nature* 381:661-6
- Dingley AJ, Grzesiek S (1998) Direct observation of Hydrogen bonds in nucleic acid base pairs by internucleotide  $^2J_{NN}$  Couplings. *J Am Chem Soc* 120:8293-97
- Dorfman T, Luban J, Goff SP, Haseltine WA, Gottlinger HG (1993) Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* 67:6159-69
- Doria-Rose NA, Vogt VM (1998) In vivo selection of Rous sarcoma virus mutants with randomized sequences in the packaging signal. *J Virol* 72:8073-82
- Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA, Cayanan C, Maddon PJ, Koup RA, Moore JP, Paxton WA (1996) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* 381:667-73
- Dupraz P, Oertle S, Meric C, Damay P, Spahr PF (1990) Point mutations in the proximal Cys-His box of Rous sarcoma virus nucleocapsid protein. *J Virol* 64:4978-87
- Ennifar E, Walter P, Ehresmann B, Ehresmann C, Dumas P (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol* 8:1064-8
- Ennifar E, Yusupov M, Walter P, Marquet R, Ehresmann B, Ehresmann C, Dumas P (1999) The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure Fold Des* 7:1439-49
- Feng Y, Broder CC, Kennedy PE, Berger EA (1996a) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* 272:872-7
- Feng YX, Copeland TD, Henderson LE, Gorelick RJ, Bosche WJ, Levin JG, Rein A (1996b) HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro. *Proc Natl Acad Sci U S A* 93:7577-81
- Fosse P, Motte N, Roumier A, Gabus C, Muriaux D, Darlix JL, Paoletti J (1996) A short autocomplementary sequence plays an essential role in avian sarcoma-leukosis virus RNA dimerization. *Biochemistry* 35:16601-9
- Frankel AD, Young JA (1998) HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* 67:1-25
- Fu W, Gorelick RJ, Rein A (1994) Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J Virol* 68:5013-8
- Gelderblom HR, Hausmann EH, Ozel M, Pauli G, Koch MA (1987) Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. *Virology* 156:171-6
- Giessner-Prettre C, Pullman B (1987) Quantum mechanical calculations of NMR chemical shifts in nucleic acids. *Q Rev Biophys* 20:113-72
- Girard F (2000) Etude par RMN de la structure des dimères de la séquence SL1 initiant la dimérisation chez VIH-1. *Biophysique Moléculaire*, Orleans
- Girard F, Barbault F, Gouyette C, Huynh-Dinh T, Paoletti J, Lancelot G (1999) Dimer initiation sequence of HIV-1Lai genomic RNA: NMR solution structure of the extended duplex. *J Biomol Struct Dyn* 16:1145-57

- Girard PM, Bonnet-Mathoniere B, Muriaux D, Paoletti J (1995) A short autocomplementary sequence in the 5' leader region is responsible for dimerization of MoMuLV genomic RNA. *Biochemistry* 34:9785-94
- Girard PM, de Rocquigny H, Roques BP, Paoletti J (1996) A model of PSI dimerization: destabilization of the C278-G303 stem-loop by the nucleocapsid protein (NCp10) of MoMuLV. *Biochemistry* 35:8705-14
- Glemarec C, Kufel J, Foldesi A, Maltseva T, Sandstrom A, Kirsebom LA, Chattopadhyaya J (1996) The NMR structure of 31mer RNA domain of Escherichia coli RNase P RNA using its non-uniformly deuterium labelled counterpart [the 'NMR-window' concept]. *Nucleic Acids Res* 24:2022-35
- Gorelick RJ, Henderson LE, Hanser JP, Rein A (1988) Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a "zinc finger-like" protein sequence. *Proc Natl Acad Sci U S A* 85:8420-4
- Gorenstein DG, Lai K, Shah DO (1984) 31P and two-dimensional 31P/1H correlated NMR spectra of Duplex d(Ap[17O]Gp[18O]Cp[16O]T) and assignment of 31P signals in d(ApGpCpT)-actinomycin D complex. *Biochemistry* 23:6717-23
- Green LM, Berg JM (1990) Retroviral nucleocapsid protein-metal ion interactions: folding and sequence variants. *Proc Natl Acad Sci U S A* 87:6403-7
- Haddrick M, Lear AL, Cann AJ, Heaphy S (1996) Evidence that a kissing loop structure facilitates genomic RNA dimerisation in HIV-1. *J Mol Biol* 259:58-68
- Harrison GP, Miele G, Hunter E, Lever AM (1998) Functional analysis of the core human immunodeficiency virus type 1 packaging signal in a permissive cell line. *J Virol* 72:5886-96
- Hayashi T, Shioda T, Iwakura Y, Shibuta H (1992) RNA packaging signal of human immunodeficiency virus type 1. *Virology* 188:590-9
- Hu WS, Temin HM (1990) Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci U S A* 87:1556-60
- Huang S, Wang YX, Draper DE (1996) Structure of a hexanucleotide RNA hairpin loop conserved in ribosomal RNAs. *J Mol Biol* 258:308-21
- Husken D, Goodall G, Blommers MJ, Jahnke W, Hall J, Haner R, Moser HE (1996) Creating RNA bulges: cleavage of RNA in RNA/DNA duplexes by metal ion catalysis. *Biochemistry* 35:16591-600
- Ikeda RA, Bailey PA (1992) Inhibition of T7 RNA polymerase by T7 lysozyme in vitro. *J Biol Chem* 267:20153-8
- Jones JS, Allan RW, Temin HM (1994) One retroviral RNA is sufficient for synthesis of viral DNA. *J Virol* 68:207-16
- Jucker FM, Heus HA, Yip PF, Moors EH, Pardi A (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol* 264:968-80
- Jucker FM, Pardi A (1995) Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry* 34:14416-27
- Karplus M (1959) Contact electron-spin coupling of nuclear magnetic moments. *J Chem Phys* 30:11-5
- Kieken F, Arnoult E, Barbault F, Paquet F, Huynh-Dinh T, Paoletti J, Genest D, Lancelot G (2002) HIV-1(Lai) genomic RNA: combined use of NMR and molecular dynamics simulation for studying the structure and internal dynamics of a mutated SL1 hairpin. *Eur Biophys J* 31:521-31
- Kim CH, Tinoco I, Jr. (2000) A retroviral RNA kissing complex containing only two G.C base pairs. *Proc Natl Acad Sci U S A* 97:9396-401



- Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graphics* 14:51-55
- Kuiken C, Thakallapalli R, Esklid A, de Ronde A (2000) Genetic analysis reveals epidemiologic patterns in the spread of human immunodeficiency virus. *Am J Epidemiol* 152:814-22
- Kung HJ, Bailey JM, Davidson N, Nicolson MO, McAllister RM (1975) Structure, subunit composition, and molecular weight of RD-114 RNA. *J Virol* 16:397-411
- Lancelot G, Guesnet JL, Vovelle F (1989) Solution structure of the parallel-stranded duplex oligonucleotide alpha-d(TCTAAAC)-beta-d(AGATTTG) via complete relaxation matrix analysis of the NOE effects and molecular mechanics calculations. *Biochemistry* 28:7871-8
- Lapadat-Tapolsky M, De Rocquigny H, Van Gent D, Roques B, Plasterk R, Darlix JL (1993) Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Res* 21:831-9
- Lapham J, Rife JP, Moore JP, Crothers DM (1997) Measurement of diffusion constants for nucleic acids by NMR. *Journal of Biomolecular NMR* 10:255-262
- Laughrea M, Jette L (1994) A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry* 33:13464-74
- Laughrea M, Jette L (1996a) HIV-1 genome dimerization: formation kinetics and thermal stability of dimeric HIV-1Lai RNAs are not improved by the 1-232 and 296-790 regions flanking the kissing-loop domain. *Biochemistry* 35:9366-74
- Laughrea M, Jette L (1996b) Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248-271 are dispensable for dimer formation. *Biochemistry* 35:1589-98
- Laughrea M, Shen N, Jette L, Wainberg MA (1999) Variant effects of non-native kissing-loop hairpin palindromes on HIV replication and HIV RNA dimerization: role of stem-loop B in HIV replication and HIV RNA dimerization. *Biochemistry* 38:226-34
- Lavery R, Sklenar H (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn* 6:63-91
- Lavery R, Sklenar H (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J Biomol Struct Dyn* 6:655-67
- Lavery R, Zakrzewska K, Sklenar H (1995) JUMNA: junction minimisation of nucleic acids. *Comp Phys Comm* 91:255-62
- Lawrence DC, Stover CC, Noznitsky J, Wu Z, Summers MF (2003) Structure of the intact stem and bulge of HIV-1 Psi-RNA stem-loop SL1. *J Mol Biol* 326:529-42
- Lee AJ, Crothers DM (1998) The solution structure of an RNA loop-loop complex: the Cole1 inverted loop sequence. *Structure* 6:993-1005
- Leeper TC, Martin MB, Kim H, Cox S, Semenchenko V, Schmidt FJ, Van Doren SR (2002) Structure of the UGAGAU hexaloop that braces Bacillus RNase P for action. *Nat Struct Biol* 9:397-403
- Legault P, Pardi A (1994) <sup>31</sup>P chemical shift as a probe of structural motifs in RNA. *J Magn Reson B* 103:82-6
- Leulliot N, Baumruk V, Abdelkafi M, Turpin PY, Namane A, Gouyette C, Huynh-Dinh T, Ghomi M (1999) Unusual nucleotide conformations in GNRA and UNCG type tetraloop hairpins: evidence from Raman markers assignments. *Nucleic Acids Res* 27:1398-404

- Lever A, Gottlinger H, Haseltine W, Sodroski J (1989) Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J Virol* 63:4085-7
- Liang C, Rong L, Laughrea M, Kleiman L, Wainberg MA (1998) Compensatory point mutations in the human immunodeficiency virus type 1 Gag region that are distal from deletion mutations in the dimerization initiation site can restore viral replication. *J Virol* 72:6629-36
- Lodmell JS, Ehresmann C, Ehresmann B, Marquet R (2000) Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site. *Rna* 6:1267-76
- Lodmell JS, Ehresmann C, Ehresmann B, Marquet R (2001) Structure and dimerization of HIV-1 kissing loop aptamers. *J Mol Biol* 311:475-90
- Lodmell JS, Paillart JC, Mignot D, Ehresmann B, Ehresmann C, Marquet R (1998) Oligonucleotide-mediated inhibition of genomic RNA dimerization of HIV-1 strains MAL and LAI: a comparative analysis. *Antisense Nucleic Acid Drug Dev* 8:517-29
- Lund AH, Mikkelsen JG, Schmidt J, Duch M, Pedersen FS (1999) The kissing-loop motif is a preferred site of 5' leader recombination during replication of SL3-3 murine leukemia viruses in mice. *J Virol* 73:9614-8
- Macura S, Huang Y, Suter D, Ernst RR (1981) Two-dimensional chemical exchange and cross-relaxation spectroscopy of coupled nuclear spins. *J Magn Reson* 43:259-281
- Macura S, Wuthrich K, Ernst RR (1982) Separation and suppression of coherent transfer effects in two-dimensional NOE and chemical exchange spectroscopy. *J Magn Reson* 46:269-282
- Majumbar A, Gosser Y, Patel DJ (2001) 1H-1H correlations across N-H...N hydrogen bonds in nucleic acids. *Journal of Biomolecular NMR* 21:289-306
- Mangel WF, Delius H, Duesberg PH (1974) Structure and molecular weight of the 60-70S RNA and the 30-40S RNA of the Rous sarcoma virus. *Proc Natl Acad Sci U S A* 71:4541-5
- Marion D, Ikura M, Tschudin R, Bax A (1989) Rapid recording of 2D NMR spectra without phase cycling: application to the study of hydrogen exchange in proteins. *J Magn Reson* 85:393-399
- Marion D, Wuthrich K (1983) Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurement of 1H-1H spin spin coupling constants in proteins. *Biochem Biophys Res Commun* 113:967-974
- Marquet R, Baudin F, Gabus C, Darlix JL, Mougel M, Ehresmann C, Ehresmann B (1991) Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and possible mechanism. *Nucleic Acids Res* 19:2349-57
- Marquet R, Paillart JC, Skripkin E, Ehresmann C, Ehresmann B (1994) Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site. *Nucleic Acids Res* 22:145-51
- McBride MS, Panganiban AT (1996) The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J Virol* 70:2963-73
- McBride MS, Schwartz MD, Panganiban AT (1997) Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. *J Virol* 71:4544-54
- Menger M, Eckstein F, Porschke D (2000) Dynamics of the RNA hairpin GNRA tetraloop. *Biochemistry* 39:4500-7
- Meric C, Gouilloud E, Spahr PF (1988) Mutations in Rous sarcoma virus nucleocapsid protein p12 (NC): deletions of Cys-His boxes. *J Virol* 62:3328-33

- Meric C, Spahr PF (1986) Rous sarcoma virus nucleic acid-binding protein p12 is necessary for viral 70S RNA dimer formation and packaging. *J Virol* 60:450-9
- Miele G, Mouland A, Harrison GP, Cohen E, Lever AM (1996) The human immunodeficiency virus type 1 5' packaging signal structure affects translation but does not function as an internal ribosome entry site structure. *J Virol* 70:944-51
- Mikkelsen JG, Lund AH, Duch M, Pedersen FS (1998) Recombination in the 5' leader of murine leukemia virus is accurate and influenced by sequence identity with a strong bias toward the kissing-loop dimerization region. *J Virol* 72:6967-78
- Mikkelsen JG, Lund AH, Duch M, Pedersen FS (2000) Mutations of the kissing-loop dimerization sequence influence the site specificity of murine leukemia virus recombination in vivo. *J Virol* 74:600-10
- Mikkelsen JG, Lund AH, Kristensen KD, Duch M, Sorensen MS, Jorgensen P, Pedersen FS (1996) A preferred region for recombinational patch repair in the 5' untranslated region of primer binding site-impaired murine leukemia virus vectors. *J Virol* 70:1439-47
- Mikkelsen JG, Pedersen FS (2000) Genetic reassortment and patch repair by recombination in retroviruses. *J Biomed Sci* 7:77-99
- Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* 15:8783-98
- Morellet N, de Rocquigny H, Mely Y, Jullian N, Demene H, Ottmann M, Gerard D, Darlix JL, Fournie-Zaluski MC, Roques BP (1994) Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by 1H NMR. *J Mol Biol* 235:287-301
- Morellet N, Demene H, Teilleux V, Huynh-Dinh T, de Rocquigny H, Fournie-Zaluski MC, Roques BP (1998) Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC). *J Mol Biol* 283:419-34
- Morellet N, Jullian N, De Rocquigny H, Maigret B, Darlix JL, Roques BP (1992) Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR. *Embo J* 11:3059-65
- Morse PM, Feshbach H (1953) *Methods of Theoretical Physics, Part I*. McGraw-Hill, New York
- Mougel M, Tounekti N, Darlix JL, Paoletti J, Ehresmann B, Ehresmann C (1993) Conformational analysis of the 5' leader and the gag initiation site of Mo-MuLV RNA and allosteric transitions induced by dimerization. *Nucleic Acids Res* 21:4677-84
- Mujeeb A, Clever JL, Billeci TM, James TL, Parslow TG (1998) Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat Struct Biol* 5:432-6
- Mujeeb A, Kerwin SM, Kenyon GL, James TL (1993) Solution structure of a conserved DNA sequence from the HIV-1 genome: restrained molecular dynamics simulation with distance and torsion angle restraints derived from two-dimensional NMR spectra. *Biochemistry* 32:13419-31
- Mujeeb A, Parslow TG, Zarrinpar A, Das C, James TL (1999) NMR structure of the mature dimer initiation complex of HIV-1 genomic RNA. *FEBS Lett* 458:387-92
- Muriaux D, De Rocquigny H, Roques BP, Paoletti J (1996a) NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J Biol Chem* 271:33686-92
- Muriaux D, Fosse P, Paoletti J (1996b) A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. *Biochemistry* 35:5075-82

- Muriaux D, Girard PM, Bonnet-Mathoniere B, Paoletti J (1995) Dimerization of HIV-1 RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide. *J Biol Chem* 270:8209-16
- Murti KG, Bondurant M, Tereba A (1981) Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. *J Virol* 37:411-19
- Nikonowicz EP, Michnicka M, Kalurachchi K, DeJong E (1997) Preparation and characterization of a uniformly  $^2\text{H}/^{15}\text{N}$ -labeled RNA oligonucleotide for NMR studies. *Nucleic Acids Res* 25:1390-6
- Nilges M, Clore GM, Gronenborn AM (1988) Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett* 229:317-24
- Otting G, Wuthrich K (1989) Extended heteronuclear editing of 2D  $^1\text{H}$  NMR spectra of isotope-labeled proteins, using the  $X(\omega_1, \omega_2)$  double half filter. *J Magn Reson* 85:586-94
- Paillart JC, Berthoux L, Ottmann M, Darlix JL, Marquet R, Ehresmann B, Ehresmann C (1996a) A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J Virol* 70:8348-54
- Paillart JC, Marquet R, Skripkin E, Ehresmann B, Ehresmann C (1994) Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *J Biol Chem* 269:27486-93
- Paillart JC, Marquet R, Skripkin E, Ehresmann C, Ehresmann B (1996b) Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie* 78:639-53
- Paillart JC, Skripkin E, Ehresmann B, Ehresmann C, Marquet R (1996c) A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci U S A* 93:5572-7
- Paillart JC, Westhof E, Ehresmann C, Ehresmann B, Marquet R (1997) Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *J Mol Biol* 270:36-49
- Paoletti J, Mougél M, Tounekti N, Girard PM, Ehresmann C, Ehresmann B (1993) Spontaneous dimerization of retroviral MoMuLV RNA. *Biochimie* 75:681-6
- Pervushin K, Ono A, Fernandez C, Szyperski T, Kainosho M, Wuthrich K (1998) NMR scalar couplings across Watson-Crick base pair hydrogen bonds in DNA observed by transverse relaxation-optimized spectroscopy. *Proc Natl Acad Sci U S A* 95:14147-51
- Pervushin K, Riek R, Wider G, Wuthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* 94:12366-71
- Piotto M, Saudek V, Sklenar V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2:661-5
- Plateau P, Gueron M (1982) Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J Am Chem Soc* 104:7310-11
- Pley HW, Flaherty KM, McKay DB (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature* 372:68-74
- Polge E, Darlix JL, Paoletti J, Fosse P (2000) Characterization of loose and tight dimer forms of avian leukosis virus RNA. *J Mol Biol* 300:41-56
- Poon DT, Li G, Aldovini A (1998) Nucleocapsid and matrix protein contributions to selective human immunodeficiency virus type 1 genomic RNA packaging. *J Virol* 72:1983-93

- Poon DT, Wu J, Aldovini A (1996) Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *J Virol* 70:6607-16
- Powell MJD (1977) Restart procedures for the conjugate gradient method. *Mathematical Programming* 12:241-54
- Prats AC, Roy C, Wang PA, Erard M, Housset V, Gabus C, Paoletti C, Darlix JL (1990) cis elements and trans-acting factors involved in dimer formation of murine leukemia virus RNA. *J Virol* 64:774-83
- Prestegard JH, al-Hashimi HM, Tolman JR (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q Rev Biophys* 33:371-424
- Puglisi EV, Puglisi JD (1998) Nuclear magnetic resonance spectroscopy of RNA. Cold Spring Harbor Laboratory Press, New York
- Rist MJ, Marino JP (2002) Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry* 41:14762-70
- Roy C, Tounekti N, Mougél M, Darlix JL, Paoletti C, Ehresmann C, Ehresmann B, Paoletti J (1990) An analytical study of the dimerization of in vitro generated RNA of Moloney murine leukemia virus MoMuLV. *Nucleic Acids Res* 18:7287-92
- Saenger W (1984) Principles of Nucleic Acid Structure. Springer-Verlag, New York
- Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K (1998) TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci U S A* 95:13585-90
- SantaLucia J, Jr., Turner DH (1993) Structure of (rGGCGAGCC)<sub>2</sub> in solution from NMR and restrained molecular dynamics. *Biochemistry* 32:12612-23
- Schmalzbauer E, Strack B, Dannull J, Guehmann S, Moelling K (1996) Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding in vitro. *J Virol* 70:771-7
- Schwartz MD, Fiore D, Panganiban AT (1997) Distinct functions and requirements for the Cys-His boxes of the human immunodeficiency virus type 1 nucleocapsid protein during RNA encapsidation and replication. *J Virol* 71:9295-305
- Scott WG, Finch JT, Klug A (1995) The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* 81:991-1002
- Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wide band applications and low level operation. *J Magn Reson* 64:547-552
- Shaka AJ, Keeler J, Freeman R (1983) Evolution of a new broadband decoupling sequence: WALTZ-16. *J Magn Reson* 53:313-340
- Shen N, Jette L, Liang C, Wainberg MA, Laughrea M (2000) Impact of human immunodeficiency virus type 1 RNA dimerization on viral infectivity and of stem-loop B on RNA dimerization and reverse transcription and dissociation of dimerization from packaging. *J Virol* 74:5729-35
- Sich C, Ohlenschlager O, Ramachandran R, Gorlach M, Brown LR (1997) Structure of an RNA hairpin loop with a 5'-CGUUUCG-3' loop motif by heteronuclear NMR spectroscopy and distance geometry. *Biochemistry* 36:13989-4002
- Skripkin E, Paillart JC, Marquet R, Blumenfeld M, Ehresmann B, Ehresmann C (1996) Mechanisms of inhibition of in vitro dimerization of HIV type I RNA by sense and antisense oligonucleotides. *J Biol Chem* 271:28812-7
- Skripkin E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C (1994) Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc Natl Acad Sci U S A* 91:4945-9
- Stein EG, Rice LM, Brunger AT (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J Magn Reson* 124:154-64

- Summers MF, South TL, Kim B, Hare DR (1990) High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach. *Biochemistry* 29:329-40
- Sundquist WI, Heaphy S (1993) Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc Natl Acad Sci U S A* 90:3393-7
- Surovoy A, Dannull J, Moelling K, Jung G (1993) Conformational and nucleic acid binding studies on the synthetic nucleocapsid protein of HIV-1. *J Mol Biol* 229:94-104
- Takahashi K, Baba S, Hayashi Y, Koyanagi Y, Yamamoto N, Takaku H, Kawai G (2000a) NMR analysis of intra- and inter-molecular stems in the dimerization initiation site of the HIV-1 genome. *J Biochem (Tokyo)* 127:681-6
- Takahashi K, Baba S, Koyanagi Y, Yamamoto N, Takaku H, Kawai G (2001) Two basic regions of NCp7 are sufficient for conformational conversion of HIV-1 dimerization initiation site from kissing-loop dimer to extended-duplex dimer. *J Biol Chem* 276:31274-8
- Takahashi KI, Baba S, Chattopadhyay P, Koyanagi Y, Yamamoto N, Takaku H, Kawai G (2000b) Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *Rna* 6:96-102
- Tereshko V, Wallace ST, Usman N, Wincott FE, Egli M (2001) X-ray crystallographic observation of "in-line" and "adjacent" conformations in a bulged self-cleaving RNA/DNA hybrid. *Rna* 7:405-20
- Theilleux-Delalande V (1998) Les interactions ARN/protéine de nucléocapside chez VIH-1: leurs implications dans les relations structure-fonction in vitro. Université René Descartes, Paris
- Theilleux-Delalande V, Girard F, Huynh-Dinh T, Lancelot G, Paoletti J (2000) The HIV-1(Lai) RNA dimerization. Thermodynamic parameters associated with the transition from the kissing complex to the extended dimer. *Eur J Biochem* 267:2711-9
- Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A (1997) Use of dipolar <sup>1</sup>H-<sup>15</sup>N and <sup>1</sup>H-<sup>13</sup>C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat Struct Biol* 4:732-8
- Turner BG, Summers MF (1999) Structural biology of HIV. *J Mol Biol* 285:1-32
- Varani G, Cheong C, Tinoco I, Jr. (1991) Structure of an unusually stable RNA hairpin. *Biochemistry* 30:3280-9
- Verlet L (1967) Computer experiments on classical fluids. Thermodynamical properties of Lennard-Jones molecules. *Phys Rev* 159:98-105
- Vuilleumier C, Bombarda E, Morellet N, Gerard D, Roques BP, Mely Y (1999) Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry* 38:16816-25
- Wang YX, Huang S, Draper DE (1996) Structure of a U.U pair within a conserved ribosomal RNA hairpin. *Nucleic Acids Res* 24:2666-72
- Weiss S, Hausl G, Famulok M, Konig B (1993) The multimerization state of retroviral RNA is modulated by ammonium ions and affects HIV-1 full-length cDNA synthesis in vitro. *Nucleic Acids Res* 21:4879-85
- Wijmenga S, Kruithof M, Hilbers CW (1997) Analysis of <sup>1</sup>H chemical shifts in DNA: Assessment of the reliability of <sup>1</sup>H chemical shift calculations for use in structure refinement. *Journal of Biomolecular NMR* 10:337-50
- Wijmenga S, van Buuren B (1998) The use of NMR methods for conformational studies of nucleic acids. *Prog Nucl Magn Reson Spec* 34:287-387

- Windbichler N, Werner M, Schroeder R (2003) Kissing complex-mediated dimerisation of HIV-1 RNA: coupling extended duplex formation to ribozyme cleavage. *Nucleic Acids Res* 31:6419-27
- Woese CR, Winker S, Gutell RR (1990) Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proc Natl Acad Sci U S A* 87:8467-71
- Wyatt JR, Chastain M, Puglisi JD (1991) Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques* 11:764-9
- Yamakage SI, Maltseva TV, Nilson FP, Foldesi A, Chattopadhyaya J (1993) Deuteration of sugar protons simplify NMR assignments and structure determination of large oligonucleotide by the <sup>1</sup>H-NMR window approach. *Nucleic Acids Res* 21:5005-11
- Zhang XL, Patel DJ (1991) Solution structure of the luzopeptin-DNA complex. *Biochemistry* 30:4026-41
- Zhang Y, Barklis E (1997) Effects of nucleocapsid mutations on human immunodeficiency virus assembly and RNA encapsidation. *J Virol* 71:6765-76