



**HAL**  
open science

# Etude par RMN et modélisation moléculaire des structures de la séquence ARN initiant la dimérisation chez VIH-1Lai et de son analogue ADN

Florent Barbault

► **To cite this version:**

Florent Barbault. Etude par RMN et modélisation moléculaire des structures de la séquence ARN initiant la dimérisation chez VIH-1Lai et de son analogue ADN. Autre. Université d'Orléans, 2001. Français. NNT: . tel-00009954

**HAL Id: tel-00009954**

**<https://theses.hal.science/tel-00009954>**

Submitted on 18 Aug 2005

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

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



# Thèse

Présentée à l'Université d'Orléans  
pour obtenir le grade de

*Docteur de l'Université d'Orléans*

Discipline

*Sciences de la vie*

Par

**Florent BARBAULT**

**Etude par RMN et modélisation moléculaire  
des structures de la séquence ARN initiant la  
dimérisation chez VIH-1<sub>Lai</sub> et de son analogue  
ADN.**

Soutenue le 09 novembre 2001, devant le jury :

|     |                   |                    |
|-----|-------------------|--------------------|
| M.  | Gérald Guillaumet | examineur          |
| M.  | Serge Femandjian  | rapporteur         |
| Mme | Brigitte Hartmann | rapporteur         |
| Mme | Françoise Vovelle | examineur          |
| M.  | Jacques Paoletti  | examineur          |
| M.  | Gérard Lancelot   | directeur de thèse |

A Anne, pour son indéfectible soutien...

*Ce travail de thèse s'est effectué dans l'équipe de "Structures et dynamiques des acides nucléiques et de leurs complexes par RMN" au Centre de Biophysique Moléculaire, du CNRS d'Orléans. Je tiens ici à remercier son directeur monsieur le Pr Paul Vigny, de m'avoir accueilli.*

*Je remercie le Dr Gérard Lancelot de m'avoir accepté au sein de son équipe et de la pleine latitude qu'il m'a donné pour m'exprimer dans ce sujet de recherche tout au long de ces trois années de thèse.*

*Je remercie chaleureusement le Dr Serge Femandjian de l'Institut Gustave Roussy et le Dr Brigitte Hartmann de l'Institut de Biologie Physico-Chimique d'avoir accepté la tâche ingrate de rapporteur du manuscrit de cette thèse.*

*Je remercie le Pr Gérald Guillaumet pour m'avoir accueilli au sein du DEA de "Chimie et physico-chimie des composés d'intérêt biologique" ainsi que d'avoir accepté de présider ce jury de thèse.*

*Je remercie le Dr Jacques Paoletti pour la collaboration fructueuse que nous avons menée et d'avoir accepté d'être parmi les membres de ce jury. De la même façon, j'adresse un vif remerciement au Pr Françoise Vovelle de m'avoir initié à la modélisation moléculaire et d'avoir bien voulu examiner ce manuscrit.*

*Je tiens également à remercier les équipes du Pr Tham Huynh Dinh, pour la synthèse des oligonucléotides rencontrés dans ce manuscrit, et du Dr Daniel Genest, pour la simulation de dynamique moléculaire du monomère d'ARN.*

*Je garde tout particulièrement une pensée reconnaissante pour Françoise Paquet qui, bien que son nom n'apparaisse pas dans la liste de mes publications, a contribué à ce travail par ses qualités scientifiques et humaines. Nos nombreuses discussions vont désormais me manquer.*

*Je tiens aussi à saluer chaleureusement Fabien Kieken et Frédéric Girard pour leur aide et leur soutien qu'ils m'ont apportés tout au long des trois années de thèse et ceci toujours dans une*

*ambiance de travail souriante et enthousiaste.*

*Une pensée amicale pour toutes les personnes que j'ai rencontrées lors des sacro saintes pause-café. Plus particulièrement les collègues RMNistes de l'équipe de Mme Françoise Vovelle : Céline, Laurence, Nicolas, Henri, Anita, Pedro et Denise. Mais aussi Christelle et Marc qui m'ont fait le plaisir d'être présent à ma soutenance malgré l'éloignement. Le dernier coup de pouce lors du tartinage matinal des toasts a été particulièrement apprécié.*

*Enfin ces lignes sont l'occasion de remercier Anne, mes parents, mon frère et sa femme qui m'ont encouragé tout au long de ce travail.*

*D'une façon générale, cette thèse est dédiée à toutes les personnes du CBM avec qui mes rapports furent aussi divers qu'enrichissants.*

## Préambule

Le virus de l'immunodéficience humaine (VIH) est le rétrovirus responsable du tristement célèbre syndrome d'immunodéficience acquise (SIDA). Son développement à l'échelle de la planète est tel qu'au terme d'épidémie, on a désormais substitué celui de pandémie. Aujourd'hui, 1,2% de la population mondiale des 15-49 ans sont porteurs de ce virus. Ce pourcentage s'élève à 6,2% pour le seul continent africain où il est la première cause de mortalité (source de l'institut national d'études démographiques au 03 août 2001). Depuis la découverte de ce virus, les scientifiques n'ont eu de cesse de rechercher des moyens de le combattre. Ces études ont notamment démontré l'efficacité des analogues nucléosidiques et des anti-protéases qui sont actuellement utilisés en polythérapie. L'efficacité de ces traitements est néanmoins limitée au long terme et ils demeurent très toxiques pour le malade. La recherche de nouvelles molécules et de cibles thérapeutiques alternatives est plus que jamais d'actualité.

Comme tous les rétrovirus, le génome de VIH est constitué de deux copies identiques d'ARN. Celles-ci sont liées de manière non-covalente au niveau de la région 5'-terminale. Cette dimérisation interfère avec bon nombre d'étapes clés du cycle rétroviral. L'inhibition du processus de dimérisation représente donc une nouvelle cible potentielle à de nouveaux traitements thérapeutiques.

La compréhension du mécanisme de dimérisation de l'ARN génomique nécessite de mettre en évidence les structures des espèces moléculaires mises en jeu. De plus, il nous semble intéressant d'élucider le rôle dans ces différentes structures que prend le groupement 2'OH, qui différencie l'ARN de l'ADN. Après avoir positionné ce travail dans son contexte biologique, nous rappellerons les différents types de structures des acides nucléiques, puis nous aborderons les grands principes de l'étude structurale des acides nucléiques par RMN. Enfin, nous décrirons les structures de la séquence ARN SL1 impliquées dans la dimérisation du rétrovirus HIV-1<sub>Lair</sub>, puis les structures de l'analogue désoxyribose de cette même séquence.

# Table des matières

---

## **INTRODUCTION** 09

### **L'ARN rétroviral du VIH-1 et la dimérisation**

- |  |    |
|--|----|
| 1. Généralités sur le rétrovirus VIH-1                           | 11 |
| 1.1 La particule virale  | 11 |
| 1.2 Le cycle rétroviral  | 12 |
| 1.3 Le génome rétroviral   | 13 |
| 2. La dimérisation et ses implications                           | 14 |
| 2.1 Importance biologique de la dimérisation                     | 14 |
| 2.2 Identification des séquences responsables de la dimérisation | 15 |
| 2.3 Modèles de dimérisation                                      | 16 |

### **Les différentes structures d'acides nucléiques**

- |                                     |    |
|-------------------------------------|----|
| 1. Structure primaire               | 20 |
| 2. Structures secondaires           | 20 |
| 2.1 Double hélice                   | 20 |
| 2.2 Mésappariement                  | 24 |
| 2.3 Epingle à cheveux               | 24 |
| 2.4 Boucle interne et bulge         | 25 |
| 2.5 Jonction                        | 26 |
| 3. Structures tertiaires            | 27 |
| 3.1 Triples hélices                 | 27 |
| 3.2 Pseudo-noeuds                   | 28 |
| 3.3 Interactions tige-boucle hélice | 29 |
| 3.4 Interactions boucle-boucle      | 29 |

### **Détermination structurale d'acides nucléiques par RMN**

- |  |    |
|--|----|
| 1. Stratégie d'attribution des acides nucléiques | 31 |
| 1.1 Le déplacement chimique du proton            | 32 |
| 1.2 Le déplacement chimique du $^{31}\text{P}$   | 33 |

|   |    |
|---|----|
| 1.3 Les liaisons hydrogène  | 33 |
| 1.4 Méthodologie d'attribution séquentielle                         | 34 |
| 2. Informations structurales et dynamiques révélées par RMN         | 35 |
| 2.1 La mesure de distance   | 35 |
| 2.2 L'élucidation du plissement du sucre                            | 37 |
| 2.3 La détermination de la conformation du squelette phosphodiester | 38 |
| 2.4 L'interprétation de la dynamique du système                     | 39 |
| 3. Les nouvelles approches  | 40 |
| 3.1 Enrichissement aux isotopes stables                             | 41 |
| 3.2 L'expérience TROSY  | 42 |
| 3.3 L'étude RMN dans les milieux orientés                           | 42 |

## **Modélisation moléculaire d'acides nucléiques à partir de données RMN**

|   |    |
|---|----|
| 1. Le champs de force des logiciels XPLOR et CNS                  | 45 |
| 1.1 Les termes empiriques   | 46 |
| 1.2 Les termes effectifs  | 49 |
| 2. Principaux algorithmes de modélisation à partir de données RMN | 51 |
| 2.1 Méthode de "géométrie des distances"                          | 51 |
| 2.2 Mécanique moléculaire   | 52 |
| 2.3 Dynamique moléculaire   | 53 |
| 2.4 Méthode dite du recuit simulé                                 | 56 |
| 2.4.1 Recuit simulé dans l'espace cartésien                       | 57 |
| 2.4.2 Recuit simulé dans l'espace des angles dièdres              | 58 |

---

## **PREMIERE PARTIE** 61

### **Etudes structurales des différentes conformations de SL1 VIH-1<sub>Lai</sub>**

|  |    |
|--|----|
| I. <b>Problématique et objectifs</b>           | 63 |
| II. <b>Le monomère SL1 VIH-1<sub>Lai</sub></b> | 67 |
| 1. Etudes physicochimiques                     | 67 |
| 2. Etudes RMN                                  | 68 |
| 2.1 Attribution dans l'eau légère              | 69 |
| 2.2 Attribution dans l'eau lourde              | 70 |



|   |            |
|---|------------|
| 3. Modélisation moléculaire                           | 72         |
| 3.1 Etablissement des contraintes                     | 72         |
| 3.2 Résolution de la structure tridimensionnelle      | 74         |
| 4. Analyses structurales de SL1*                      | 75         |
| 5. Simulation de dynamique moléculaire                | 80         |
| 6. Conclusions  | 81         |
| <b>III Les dimères de SL1-VIH<sub>Lai</sub></b>       | <b>83</b>  |
| 1. Etudes physicochimiques                            | 83         |
| 1.1 Mise en évidence des deux types de dimères        | 83         |
| 1.2 Préparation des dimères                           | 85         |
| 2. Etude RMN du dimère stable SL1                     | 88         |
| 2.1 Préparation de l'échantillon                      | 88         |
| 2.2 Attribution dans l'eau légère                     | 88         |
| 2.3 Attribution dans l'eau lourde                     | 90         |
| 3. Modélisation moléculaire                           | 92         |
| 3.1 Etablissement des contraintes                     | 92         |
| 3.2 Résolution de la structure tridimensionnelle      | 94         |
| 4. Analyses structurale                               | 98         |
| 4.1 Organisation des tiges                            | 98         |
| 4.2 Caractérisation du motif en fermeture éclair      | 100        |
| 5. Comparaison avec des études structurales analogues | 101        |
| 6. La structure du dimère instable                    | 104        |
| 7. Conclusions  | 108        |
| <b>IV Conclusions de la première partie</b>           | <b>109</b> |
| <hr/>   |            |
| <b><u>DEUXIEME PARTIE</u></b>                         | <b>111</b> |

**Etudes structurales des différentes conformations  
de l'analogue désoxyribose SL1 de VIH-1<sub>Lai</sub>**

|      |  |     |
|------|--|-----|
| I.   | <b>Problématique et objectifs</b>                      | 113 |
| II.  | <b>Le monomère dSL1</b>                                | 115 |
| 1.   | Préparation de l'échantillon                           | 115 |
| 2.   | Etudes RMN   | 116 |
| 3.   | Conclusions  | 117 |
| III. | <b>Le dimère instable de dSL1</b>                      | 119 |
| 1.   | Etudes physicochimiques                                | 119 |
| 2.   | Etude RMN du dimère instable dSL1                      | 121 |
|      | 2.1 Préparation de l'échantillon                       | 121 |
|      | 2.2 Attribution dans l'eau légère                      | 122 |
|      | 2.3 Attribution dans l'eau lourde                      | 126 |
|      | 2.4 Attribution du squelette phosphodiester            | 129 |
| 3.   | Modélisation moléculaire                               | 132 |
|      | 3.1 Etablissement des contraintes                      | 132 |
|      | 3.2 Résolution de la structure tridimensionnelle       | 135 |
| 4.   | Analyses structurales                                  | 139 |
|      | 4.1 Organisation globale des tiges                     | 139 |
|      | 4.2 Caractérisation des motifs stabilisants            | 143 |
| 5.   | Comparaison avec la structure kissing-complex de l'ARN | 145 |
| 6.   | Conclusions  | 147 |
| IV.  | <b>Le dimère stable de dSL1</b>                        | 149 |
| 1.   | Etudes physicochimiques                                | 149 |
| 2.   | Etude RMN du dimère stable dSL1                        | 150 |
|      | 2.1 Préparation de l'échantillon                       | 150 |
|      | 2.2 Attribution dans l'eau légère                      | 151 |
|      | 2.3 Attribution dans l'eau lourde                      | 152 |
|      | 2.4 Caractérisation du squelette phosphodiester        | 158 |
| 3.   | Modélisation moléculaire                               | 159 |
|      | 3.1 Etablissement des contraintes                      | 159 |
|      | 3.2 Résolution de la structure tridimensionnelle       | 162 |
|      | 3.3 Analyse du déplacement chimique en retour          | 164 |

|   |     |
|---|-----|
| 4. Analyses structurales                                  | 168 |
| 4.1 Organisation des tiges                                | 168 |
| 4.2 Etude des motifs stabilisant la structure             | 171 |
| 5. Comparaison avec différentes structures analogues      | 173 |
| 5.1 Comparaison avec le dimère instable                   | 173 |
| 5.2 Comparaison avec le kissing-complex de ColE1          | 175 |
| 6. Mécanisme de transition supposé entre les deux dimères | 177 |
| 7. Conclusions  | 179 |
| V. <b>Conclusions de la seconde partie</b>                | 183 |
| <hr/>   |     |
| <b><u>CONCLUSIONS GÉNÉRALES ET PERSPECTIVES</u></b>       | 185 |
| <hr/>   |     |
| <b><u>MATERIELS ET METHODES</u></b>                       | 191 |
| <hr/>   |     |
| <b><u>ANNEXES, ABREVIATIONS</u></b>                       | 195 |
| <hr/>   |     |
| <b><u>BIBLIOGRAPHIE</u></b>                               | 201 |
| <hr/>   |     |

# *Introduction*

# L'ARN rétroviral du VIH-1 et la dimérisation :

## 1. Généralités sur le rétrovirus VIH-1

Les rétrovirus appartiennent à la famille des retroviridae dont les caractéristiques principales sont les suivantes : ils possèdent une transcriptase inverse, ils transportent leur matériel génétique sous forme d'ARN et leurs génomes, rétrotranscrits en ADN viral, présentent des extrémités répétées (LTR).

Actuellement le virus du SIDA est le virus le plus étudié. La famille à laquelle nous nous intéressons est représentée principalement par deux souches qui sont VIH-1 et VIH-2. VIH-1 étant la souche prépondérante et la plus virulente, elle est en conséquence la mieux caractérisée. Par la suite, nous nous attacherons plus particulièrement à développer les propriétés et spécificités liées au VIH-1.

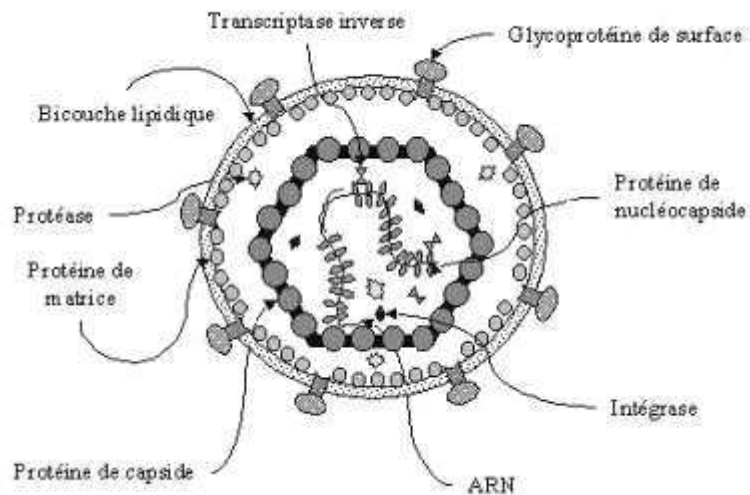
### 1.1 La particule virale

Le rétrovirus VIH se présente sous la forme d'une petite particule de 110 nm de diamètre, composée d'un noyau central entouré par une enveloppe bilipidique d'origine cellulaire. Un schéma de la particule rétrovirale est présentée figure 1.1.

La bicouche lipidique externe contient les glycoprotéines virales de surface (gp120, SU) et transmembranaires (gp41, TM). L'ensemble constitue le système protéique Env. La protéine gp120 contient les déterminants viraux qui se lient aux récepteurs de la cellule hôte. La protéine Gp41 est supposée initier la fusion membranaire entre le virus et la membrane cellulaire.

Le noyau central se décompose en deux couches : la capsid et la nucléocapsid. A l'intérieur de la capsid se trouvent les trois enzymes nécessaires à la réplication virale : la transcriptase inverse, l'intégrase et la protéase.

La protéine de nucléocapsid est une petite protéine très basique située dans le cœur du virion où elle est étroitement associée à l'ARN viral pour former un complexe ribonucléoprotéinique (Darlix et al., 1990). En outre, la protéine de nucléocapsid



intervient dans d'autres étapes essentielles au cycle rétroviral.

Figure 1.1 : Organisation schématique de la particule rétrovirale du VIH-1.

Pour une description plus précise de la particule rétrovirale du VIH et les fonctions des protéines associées, se reporter aux revues de Pavlakis (1996) et de Frankel et Young (1998).

## 1.2. Le cycle rétroviral

Le cycle de réplication des rétrovirus est le même quelle que soit la famille de rétrovirus. Une représentation schématique de ce cycle de réplication est réalisée sur la figure 1.2. Lors de ce cycle, les protéines de surface du rétrovirus vont reconnaître les récepteurs membranaires, permettant au virion de se fixer à la surface de la cellule. Après pénétration, l'ARN viral génomique est utilisé comme matrice par la transcriptase inverse qui synthétise un ADN proviral double brin. Lors de cette transcription, il existe des sauts de brins qui provoquent une grande variabilité génétique du virus rendant les traitements thérapeutiques très complexes. Cet ADN migre alors vers le noyau de la cellule où il s'intègre dans le génome cellulaire grâce à une autre enzyme : l'intégrase. Cet ADN proviral est ensuite transcrit en ARN. Une partie de cet ARN est traduite en protéines qui serviront à la constitution de nouvelles particules virales. L'autre partie s'associe en dimères qui seront encapsidés et serviront de support à l'information génétique du virion fils. Après maturation, le virion acquerra sa virulence et sera apte à contaminer d'autres

cellules.

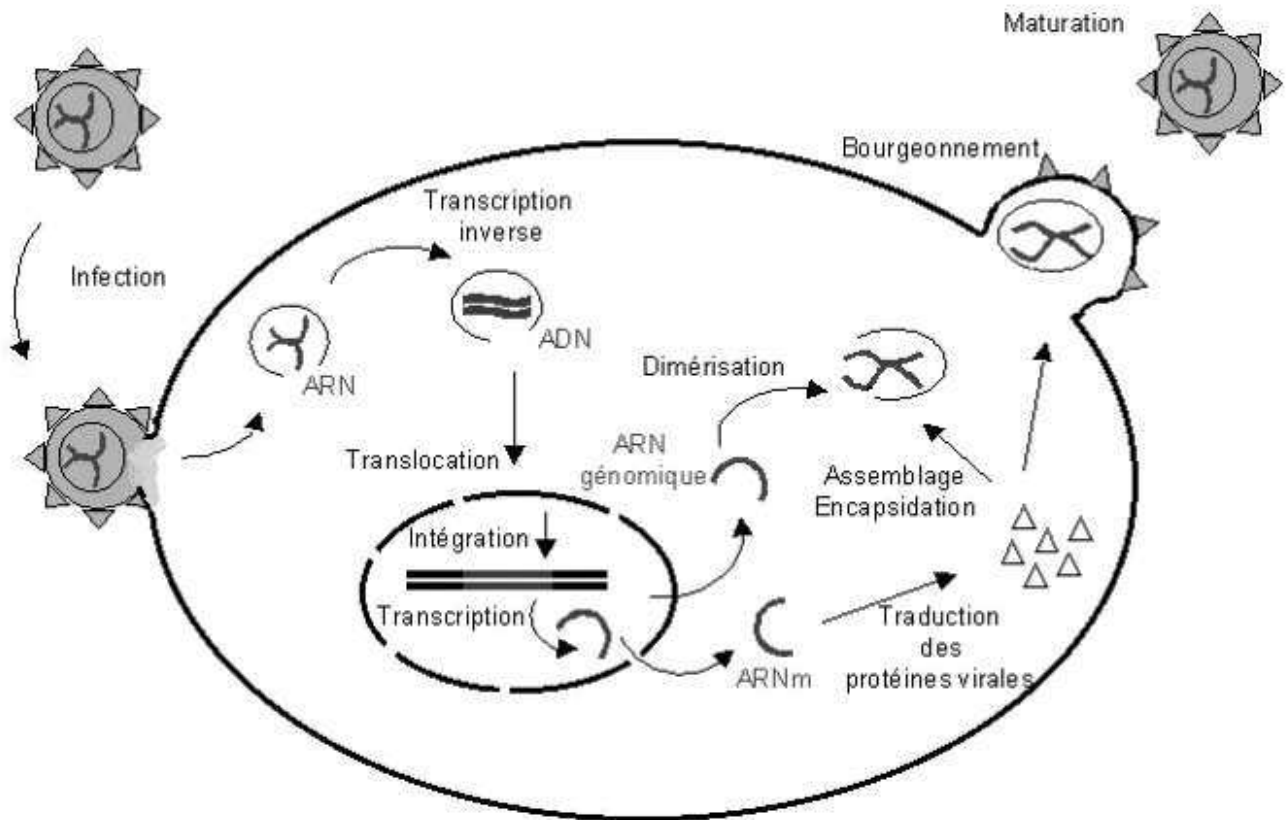


Figure 1.2 : Le cycle de répliation des rétrovirus.

### 1.3. Le génome rétroviral

Le génome des rétrovirus est composé de deux brins identiques d'ARN de même polarité. Sa taille pour le virus du VIH est d'environ 10 000 bases. La position de l'assemblage de ces deux brins a été largement étudiée par microscopie électronique et biochimie (Bender and Davidson, 1976; Murti et al., 1981; Prats et al., 1990). D'autre part, la microscopie électronique permet de visualiser nettement la capsid virale et l'état de maturité du génome. On montre que la structure dimérique des ARN rétroviraux est caractérisée par des interactions ARN/ARN. Les deux brins d'ARN sont liés près de leur extrémités 5' par une centaine de bases au niveau de séquences situées dans le domaine d'encapsidation appelé le domaine  $\Psi$  qui forment ce que l'on nomme en Anglais le : Dimer Linkage Structure (DLS).

Comme tous les rétrovirus, l'organisation génétique de l'ARN du VIH comprend une

large région codante composée des gènes Gag, Pol et Env qui permettent l'expression des trois principales polyprotéines. La polyprotéine Gag comprend les trois quarts des protéines de structures, le quart restant étant codé par le gène Env. La protéine Pol contient les enzymes nécessaires aux principales étapes du cycle rétroviral : intégrase, transcriptase inverse et protéase.

## 2) *La dimérisation et ses implications*

### 2.1. Importance biologique de la dimérisation

Le caractère diploïde et les propriétés de dimérisation associées sont directement impliqués dans le développement du cycle répliatif viral. En effet, la région 5' terminale intervient dans de nombreuses étapes essentielles à la réplication du rétrovirus :

- La transcription inverse, c'est-à-dire la synthèse d'un ADN à partir d'un ARN, est un processus complexe qui implique des transferts de brins 5' vers 3'. Cette phase est réalisée sur l'un ou l'autre brin d'ARN associé en dimère. Les sauts de brins intermoléculaires apparaissent de fait comme une des origines les plus probables de la variabilité des rétrovirus et notamment celui du VIH (Anderson et al., 1998; Jones et al., 1994). La structure dimérique de l'ARN génomique serait donc une des causes importantes de la variabilité rencontrée au sein de cette famille de virus.
- A la fin du cycle rétroviral une étape importante consiste en l'encapsidation des deux brins d'ARN. Le domaine  $\Psi$  recouvre la région impliquée dans la dimérisation, et correspond également à la zone identifiée comme étant responsable de l'encapsidation sélective du VIH (Aldovini and Young, 1990; Clavel and Orenstein, 1990; Lever et al., 1989). Le recouvrement de ces séquences suggère une interférence entre les propriétés d'encapsidation et de dimérisation (pour une revue voir (Rein, 1994)).
- L'ARN génomique peut à la fois former dans le dimère d'ARN, qui sera ensuite



encapsidé, ou être utilisé pour la traduction de Gag, Pol ou Env. Il a été suggéré que la formation du dimère d'ARN, en induisant une modification de conformation de l'ARN, pourrait inhiber la traduction (Baudin et al., 1993; Bieth et al., 1990; Mougel et al., 1993). Néanmoins ces résultats n'ont jamais été confirmés *in vivo*.

## 2.1 Identification des séquences responsables de la dimérisation :

Vu l'importance biologique de la dimérisation, de nombreuses équipes se sont intéressées à la structure secondaire responsable de la dimérisation. Le modèle de la structure secondaire de la région  $\Psi$  fut présenté par Clever en 1995 (Clever et al., 1995) voir figure 2.1.

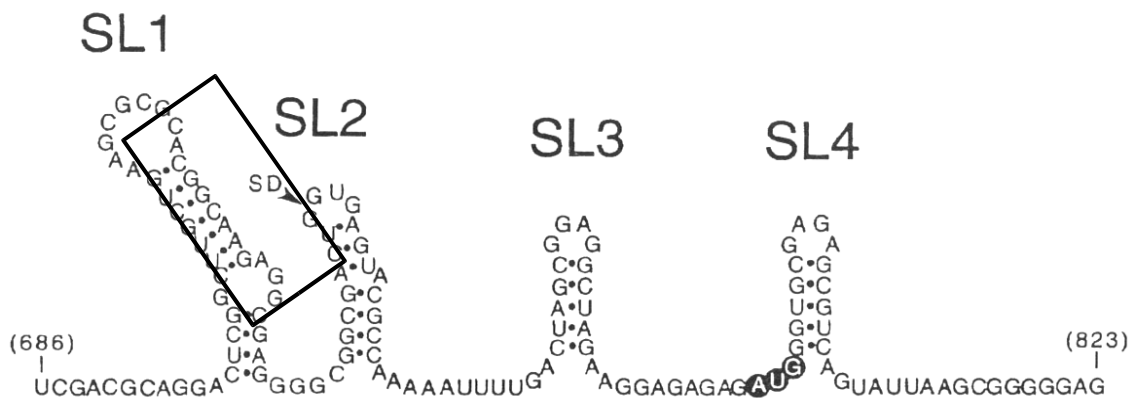


Figure 2.1 : structure secondaire du domaine d'encapsidation  $\Psi$  de la région leader du VIH-1 lai. La tige-boucle SL1 est encadrée (Clever et al., 1995).

Le DIS (ou SL1) se présente comme une tige-boucle de 23 résidus. La structure secondaire représente une tige de 7 paires de bases et une large boucle impliquant 9 bases. La séquence (figure 2.2) étant maintenant établie, il faut s'attacher à définir un modèle de dimérisation.

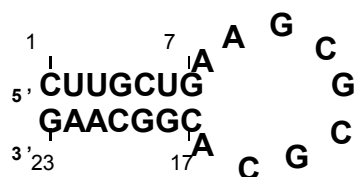


Figure 2.2 : Séquence impliquée dans la dimérisation du rétrovirus VIH-1 Lai.

### 2.3. Modèles de dimérisation

L'établissement d'un modèle de structure secondaire est l'étape logique qui suit la mise en évidence de la séquence initiant le processus de dimérisation du rétrovirus du VIH-1 Lai (voir figure 2.3). Le premier modèle proposé (Awang and Sen, 1993; Marquet et al., 1991; Sundquist and Heaphy, 1993; Weiss et al., 1993) faisait intervenir un jeu de quartet de purines et par conséquent le développement d'une quadruple hélice. Mais ce modèle a rapidement été mis en défaut puisque les séquences mutées privées de tract de polypurines conservaient les propriétés de dimérisation (Berkhout et al., 1993).

L'autre modèle propose que la dimérisation soit initiée par l'interaction des séquences auto-complémentaires de la boucle avant de former un dimère plus étendu par appariement intermoléculaire des nucléotides de la tige (Laughrea and Jette, 1994; Muriaux et al., 1995). L'étude *in vitro* du processus de dimérisation de l'ARN du VIH-1 Lai en fonction de la température a mis en évidence la formation de deux types de dimères (figure 2.3).

Un dimère instable est formé par appariement de la séquence autocomplémentaire GCGCGC de la boucle SL1 à une température inférieure ou égale à 37°C suivant la force ionique du tampon utilisé (Muriaux et al., 1996). On appelle communément ce dimère instable le "kissing-complex". Le deuxième dimère, qui est plus stable, met en jeu la totalité de la tige boucle SL1 dans un appariement intermoléculaire, et est formé à une température minimale de 55°C (Muriaux et al., 1996). De nombreuses études (Bonnet-Mathoniere et al., 1996; Darlix et al., 1990; De Rocquigny et al., 1992; Muriaux et al., 1996; Sakaguchi et al., 1993) menées sur les longs ou courts fragments d'ARN issus du domaine d'encapsidation, ont permis de démontrer l'implication de la protéine de nucléocapside dans ce mécanisme de maturation de l'ARN dimérique rétroviral.

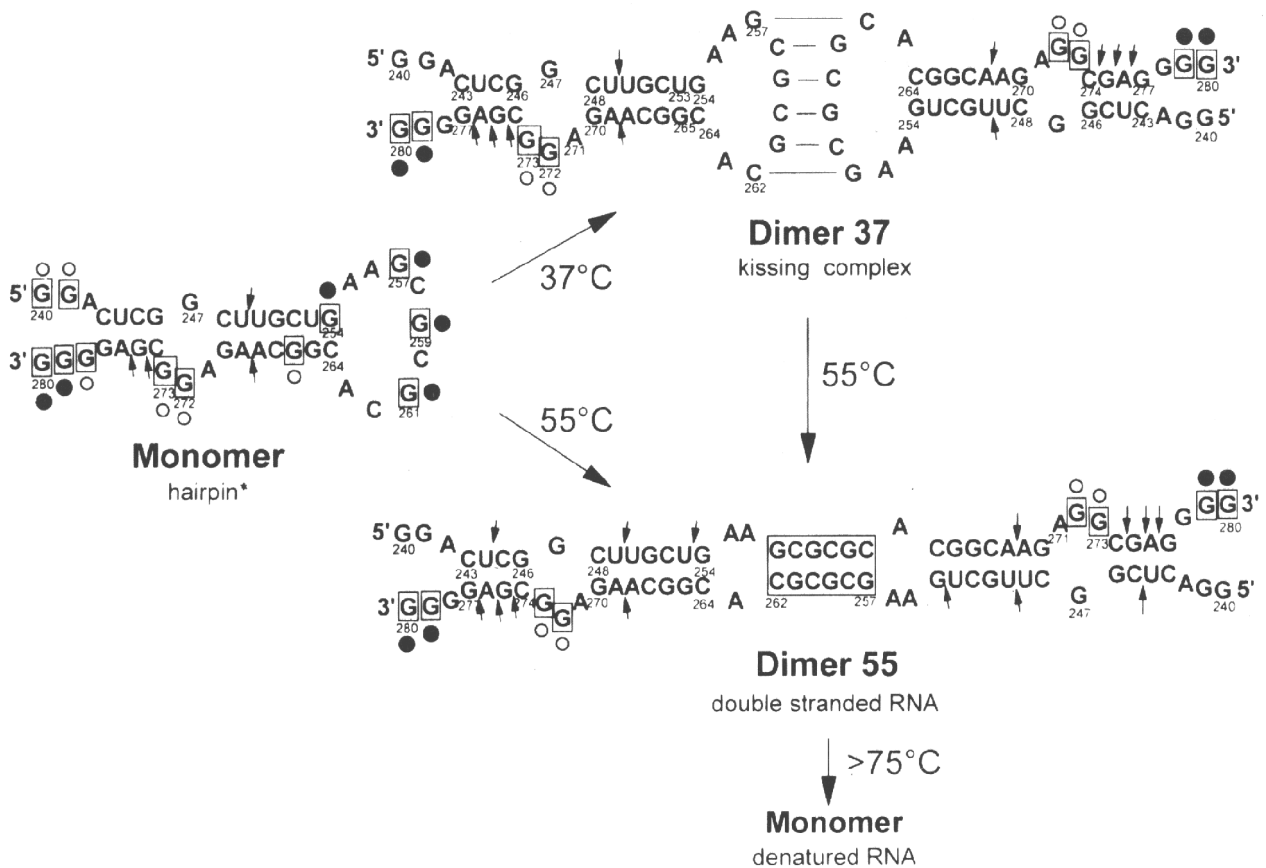


Figure 2.3 : Modèle de dimérisation impliquant l'existence d'un dimère instable de type "kissing-complex" ainsi qu'un dimère stable de type "étendu" (Muriaux et al., 1996)

Les études menées par Theilleux-Delalande (Theilleux-Delalande, 1998) sur de courts transcrits de SL1 VIH-1Lai ont permis de révéler des propriétés de dimérisation analogues : formation d'un dimère instable reposant sur une interaction de type boucle-boucle ("kissing-complex") et susceptible d'être mûré en un dimère stable ("dimère étendu") sous l'influence de la température. Ce travail a permis de développer un modèle simple pour l'étude du mécanisme de dimérisation, reposant sur une séquence courte de 23 nucléotides représentée sur la figure 2.2.

# Les différentes structures d'acides nucléiques

L'étude structurale des acides nucléiques a commencé il y a maintenant 48 ans avec le premier modèle de Watson Crick (Watson and Crick, 1953) qui continue à jouer un rôle primordial dans la compréhension de la régularisation génétique et de l'expression. L'organisation de l'ADN en une double hélice "miracle géométrique" ne doit pas masquer la complexité des structures rencontrées dans certains acides nucléiques. Un acide nucléique (ADN ou ARN) est un assemblage de nucléotides, décomposables à leur tour en éléments plus petits (voir en annexe la description des bases et des appariements possibles) : une base, un ose et un ester d'acide phosphorique.

La complexité et le degré de polymérisation, en bref la séquence, est à l'origine du patrimoine génétique. De plus certains nucléotides ont indépendamment des rôles biologiques clés : l'ATP est un réservoir d'énergie chimique et l'AMP cyclique est un régulateur. Néanmoins un ADN ne se résume pas à un "livre de la vie", son accessibilité, son mode de lecture et sa régulation en fait une molécule complexe où, comme dans le cas des protéines, la compréhension de la relation structure fonction est essentielle.

Les ARN quant à eux interviennent non seulement comme porteurs de l'information génétique mais participent aussi à toutes les étapes de l'expression des gènes. Ils sont capables de réaliser des interactions spécifiques avec des substrats (ARN, ADN ou protéines) et sont impliqués dans des étapes de régulations où ils peuvent être porteurs de propriétés catalytiques (ribosyme). Là encore, l'organisation structurale est directement liée à leur fonctionnalité.

Dans la suite de ce chapitre, nous allons développer les différentes organisations des acides nucléiques et les relier à leur intérêt biologique. On peut définir trois grands niveaux de structuration pour un acide nucléique : structure primaire, secondaire et tertiaire.

## **1. La structure primaire des acides nucléiques**

La structure primaire d'un acide nucléique est équivalente à la séquence. Elle repose sur l'alternance des différentes bases : A, T, C, G dans le cas d'un ADN et A, U, C, G dans le cas d'un ARN. Il faut noter que ces bases peuvent subir des mutations induites par la machinerie cellulaire (principalement des méthylations) ou bien par des rayonnements ionisants (oxydation des bases). Le repliement de la structure primaire conduit à la formation de la structure secondaire.

## **2. Les structures secondaires**

Nous allons présenter les principales différentes structures secondaires des ADN et des ARN afin de comprendre les différences entre les deux types d'acides nucléiques.

Pour comprendre et analyser les structures d'acides nucléiques il faut connaître les éléments structuraux qui les composent. Ces éléments structuraux (cf figure 2.1) s'appellent structures secondaires et ils sont très variés ; pour une revue se référer à (Gesteland et al., 1998; Neidle, 1999 ; Saenger, 1984). Nous allons décrire les principales structures secondaires rencontrées dans les acides nucléiques.

### **2.1 Double hélice**

Certainement aucune autre structure n'aura été autant étudiée que celle de la double hélice de l'ADN. L'ADN comporte trois grands types de familles de double hélices : A, B et Z. La figure 2.2 représente ces trois types d'hélice.

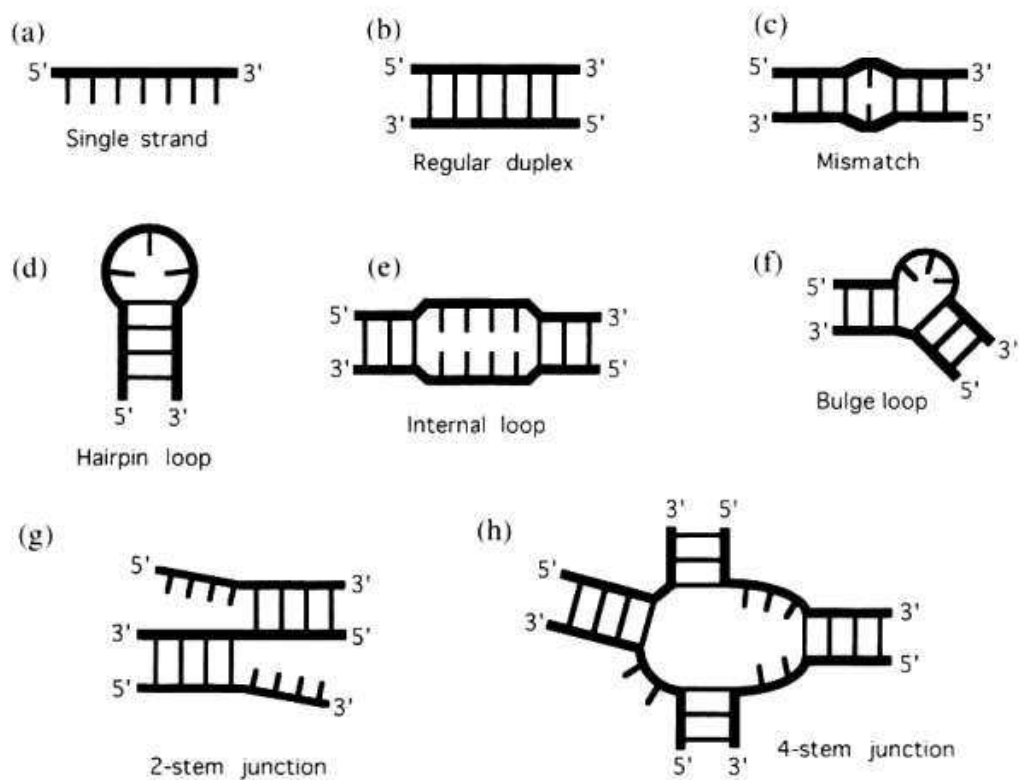


Figure 2.1 : Les principaux éléments de structures secondaires rencontrés dans les ARN et les ADN (Neidle, 1999).

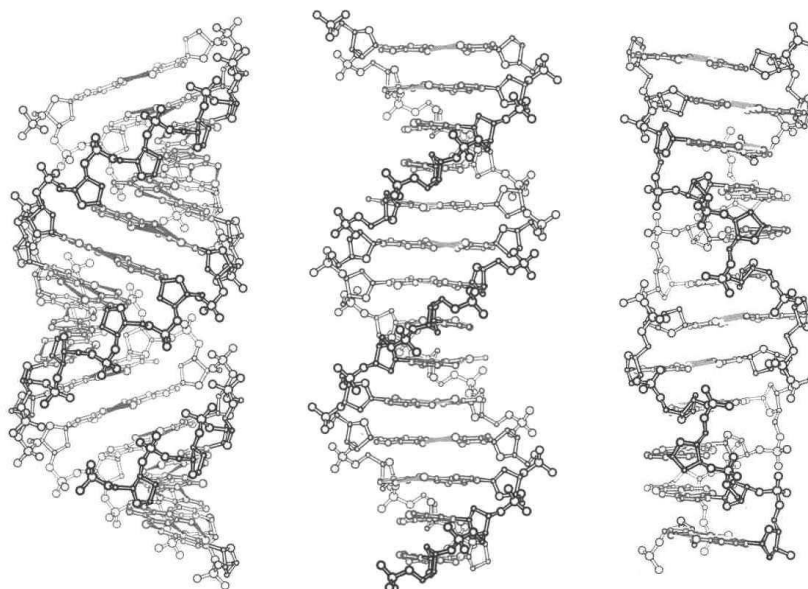


Figure 2.2 : représentation des trois types d'hélices d'ADN. de gauche à droite, l'hélice de type A, B et Z.

Historiquement, la première structure hélicoïdale d'ADN mise en évidence correspond à la forme A (Watson and Crick, 1953). Les hélices A reposent sur l'appariement de type Watson-Crick des bases, les sucres étant sous conformation  $C_3'$ -endo. Vu de profil la conformation A montre un "puits" autour duquel s'enroule l'ADN.

L'hélice A correspond à la forme déshydratée de l'ADN (Zimmerman and Pfeiffer, 1979). De ce fait, on ne la rencontre en solution que dans des conformations locales, voir par exemple l'article de Ban (Ban et al., 1994), ou lors d'hybrides ARN-ADN, l'article de Gyi (Gyi et al., 1998) en est un bon exemple.

La structure hélicoïdale de type B est la conformation "naturelle" de l'ADN en solution. Elle est couramment rencontrée en solution et montre une grande stabilité. La forme B est beaucoup plus compacte que l'hélice A et apparaît donc moins large. Le plissement du sucre est sous forme  $C_2'$ -endo pour l'ensemble des nucléotides et l'hélice comporte dix bases par tour. On observe de façon plus marquée que dans la forme A, deux sillons. Le grand sillon donne accès aux bases et le petit sillon permet l'accès aux sucres des nucléotides.

L'hélice de type Z possède la principale originalité d'être une hélice gauche alors que toutes les autres conformations d'hélice sont droites ; pour une revue se référer à Rich (Rich et al., 1984). La conformation Z se rencontre uniquement pour les séquences d'alternances purines-pyrimidines (principalement d[CG]<sub>n</sub>) où le résidu purique est en *syn* avec un plissement de sucre en  $C_3'$  endo, tandis que le résidu pyrimidique est en *anti* avec un plissement de sucre en  $C_2'$  endo. Cette alternance d'angles glycosidiques forme un "zigzag" du squelette phosphodiester qui est à l'origine du nom d'hélice Z. Ce type d'hélice se forme en solution avec d'importantes forces ioniques. Néanmoins, il a été montré que l'ADN Z a une réelle existence biologique (Rahmouni, 1992; Rich, 1995; Rich et al., 1984).

Enfin, de façon plus anecdotique, on peut observer un nouveau type d'hélice, nommé hélice P, lorsque l'on tire sur les deux extrémités de l'ADN (Allemand et al., 1998). Dans cette structure, qui n'a pas d'intérêt biologique démontré, les liaisons Watson-Crick disparaissent et les bases sont totalement exposées au solvant.

L'ARN présente beaucoup moins de possibilités d'hélices que l'ADN. La forme A est la double hélice la plus classiquement rencontrée dans les ARN. La structure est définie de la même façon que pour l'ADN. Elle correspond à des zones de fortes stabilités grâce notamment aux réseaux de liaisons hydrogènes impliquant le groupement 2'OH et l'eau (Cheatham et al., 1997; Egli et al., 1996). La forme d'hélice B, rencontrée couramment dans les ADN, n'a jamais été mis en évidence pour les ARN.

L'hélice gauche, ou forme Z, se rencontre dans les ARN uniquement pour les séquences de type poly-GC et dans des conditions de fortes forces ioniques ( $\text{NaClO}_4$  6M par exemple) (Hall et al., 1984). La conformation d'hélice Z rencontrée dans le cas des ARN est avoisinante de celle des ADN.

## 2.2 Mésappariement

Un mésappariement est rencontré lorsque deux bases, l'une face à l'autre, ne forment pas de liaisons Watson-Crick. Ces motifs perturbent l'organisation hélicoïdale et déstabilisent la structure. Une exception est faite pour le mésappariement GU de l'ARN qui forme une paire de base aussi stable que la paire AU (Sugimoto et al., 1986). Il faut cependant noter que lors de mésappariement la proximité des bases non appariées permettent généralement l'établissement de liaisons hydrogène et d'empilements de bases qui stabilisent ce motif.

## 2.3 Epingle à cheveux

On rencontre cette structure secondaire lorsqu'un simple brin s'auto-associe sur ces deux extrémités. La région appariée par des liaisons Watson-Crick s'appelle la tige et la jonction non-appariée la boucle, d'où le nom de tige-boucle. Si la taille des tiges en terme de nucléotides est variable, les boucles contiennent généralement de 2 à 10 nucléotides.

Ce motif de structure secondaire est très fréquent pour les ARN. La classe la plus importante et la plus étudiée est celle des tétra-boucles. Elles peuvent être groupées en trois familles : GNRA, UNCG et CUNG (N représente un nucléoside quelconque et R un résidu purique) (Woese et al., 1990). Les structures des divers membres ont été largement décrites par RMN (pour une revue se référer à (Varani, 1995)). Ces différents motifs



permettent un inversement de la direction du squelette phosphodiester et mettent en jeu des liaisons hydrogène impliquant les OH<sup>2'</sup> qui participent à la stabilisation de ces boucles (Jucker et al., 1996; SantaLucia et al., 1992). Les larges tiges-boucles existent mais elles comportent généralement des appariements de type Watson-Crick et des mésappariements qui font supposer que l'existence de boucles de plus d'une dizaine de résidus est peu probable.

D'une façon générale les "hairpins" peuvent constituer des motifs de reconnaissance à l'origine d'interaction ARN-ARN ou encore ARN-protéine. De plus, la dimérisation de tiges-boucles forment des complexes boucle-boucle appelés "kissing-complex".

Les épingles à cheveux d'ADN suscitent un intérêt grandissant du fait de leurs implications possibles dans plusieurs processus biologiques importants. Par exemple, les épingles à cheveux à petite boucle (inférieure à cinq résidus) peuvent être des sites d'interactions préférentielles de nombreuses protéines : topoisomérase II (Amir-Aslani et al., 1996; El Amri et al., 1999), l'intégrase du virus HIV-1 (Katz et al., 1998) ou encore la famille des protéines SMC (Structural Maintenance of Chromosomes) (Connolly et al., 1998). Du point de vue structural, chaque cas est différent mais nous pouvons noter le fait que l'hélice de la tige se trouve dans tous les cas en conformation B et non pas en conformation A.

## **2.4 Boucles internes et bulges**

Une boucle interne est formée à partir du moment où des nucléotides de chacun des brins d'une double hélice ne peuvent s'apparier. Si le nombre de nucléotides est équivalent de chaque côté des deux brins, la boucle interne est dite symétrique. Dans le cas où les nucléotides non-appariés se trouvent sur le même brin, on parle de "bulge" (voir (Hermann and Patel, 2000) pour une revue sur les bulges). La taille des boucles internes est variable et leur organisation est difficile à prévoir.

Les boucles internes interviennent de façon très nette dans la relation structure-fonction des ARN. Elles apportent un site de flexibilité locale autorisant une reconnaissance moléculaire. Elles peuvent apporter un site de liaisons avec des cations métalliques

(comme par exemple dans le cas de la séquence Tar de HIV-1 avec le cation  $\text{Ca}^{2+}$  (Ippolito and Steitz, 1998)) ou bien avec des protéines (par exemple la région 3' UTR en interaction avec la protéine humaine U1A (Gubser and Varani, 1996)) ou encore avec des antibiotiques ou des peptides ; on parle alors d'ARN aptamère (comme dans le cas de l'ATP aptamer (Jiang et al., 1996)).

Dans le cas des ADN les bulges et boucles internes sont plus rares et n'offrent pas de profondes différences avec l'ARN (Dornberger et al., 1999; Kalnik et al., 1989; Rosen et al., 1992).

## 2.5 Jonctions

Les jonctions s'observent lors de connexion entre plus de deux hélices. Différents types de jonctions sont possibles suivant le nombre d'hélices et la taille de la jonction. Les principaux motifs de jonctions sont représentés sur la figure 2.3.

Les jonctions sont des éléments structuraux très importants qui déterminent la forme globale des acides nucléiques. Leur structure est difficilement prévisible. On rencontre les jonctions principalement dans les ARN, notamment dans l'ARNt (se référer à la revue de Söll (Söll, 1993) pour des renseignements complémentaires). Elles contribuent à l'organisation des motifs de structures secondaires les uns par rapport aux autres et donc à la structuration tertiaire de l'ARN.

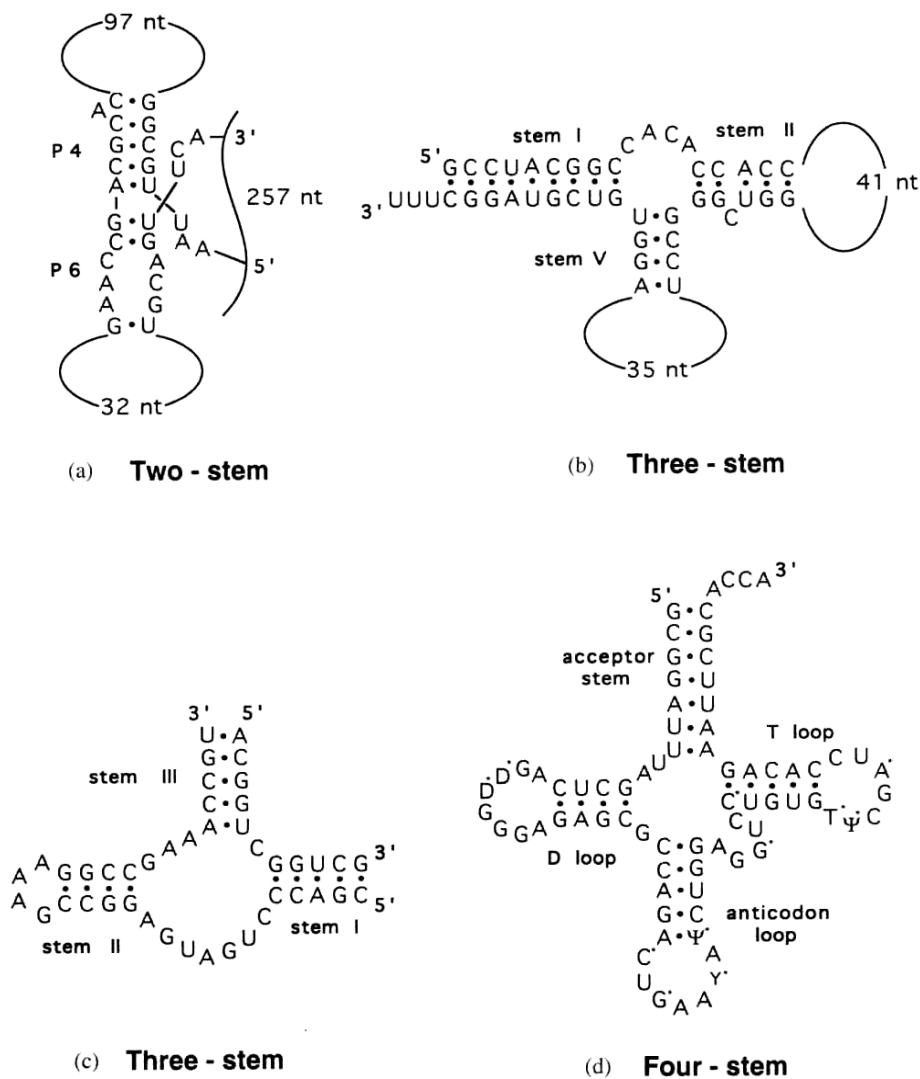


Figure 2.3 : différents types de jonctions d'ARN d'après Neidle (Neidle, 1999).

Dans le cas de l'ADN, les jonctions sont impliquées dans les phénomènes temporaires de recombinaison génétique. En effet le modèle de jonction à quatre branches, appelé "four way DNA junction", permettrait l'incorporation d'un gène dans l'ADN (Sobell, 1972). De plus les jonctions d'ADN sont aussi rencontrées lors d'événements génétiques tels la réplication de l'ADN (Kemper and Janz, 1976).

### 3. Les structures tertiaires

La combinaison des structures secondaires va constituer la structure tridimensionnelle de l'acide nucléique. C'est ce niveau d'organisation que l'on définit comme structure tertiaire. Il existe quatre motifs de base de ces structures tertiaires (figure 3.1) : les triples hélices, les

pseudo-nœuds, les interactions tige-boucle hélice et les interactions boucle-boucle. Nous allons décrire brièvement ces éléments structuraux.

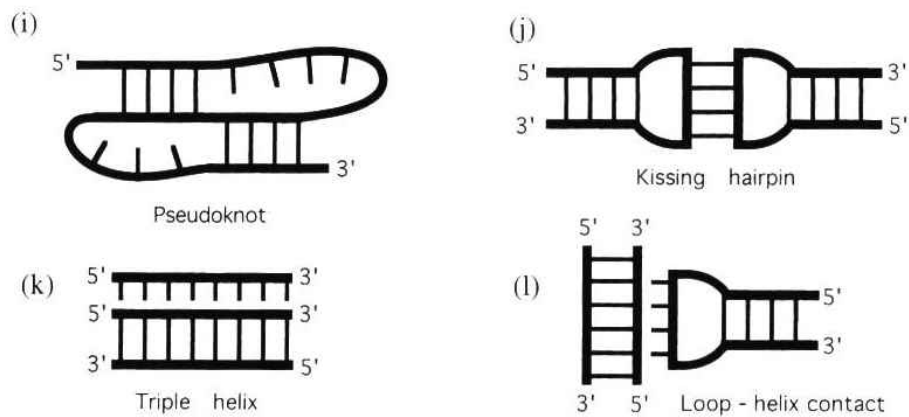


Figure 3.1 : Les principaux différents éléments de structures tertiaires. i) pseudonoeud ; j) interaction boucle-boucle ; k) triple hélices et l) interaction tige-boucle hélice (d’après (Neidle, 1999)).

### 3.1 Triples hélices

Les triples hélices s’observent lorsque un simple brin vient s’enrouler dans le grand sillon de l’hélice (hélice A pour l’ARN et B pour l’ADN). L’intérêt biologique de ces structures est considérable dans le cadre de la stratégie anti-gène ; se référer à la revue du Pr Hélène (Hélène, 1991) pour une revue sur ce sujet. La structure tertiaire est stabilisée par un réseau de liaisons hydrogène entre les bases de l’hélice et celles de l’adduit. Ces différents types de liaisons hydrogène induisent différentes familles de triple hélice, suivant le sens d’interaction du simple brin, sa séquence ou le pH de la solution ; voir l’ouvrage de Neidle (Neidle, 1999) pour une description approfondie des différentes possibilités de triples hélices.

### 3.2 Pseudo-nœuds

Un pseudo-nœud se développe lorsqu’un simple brin se replie sur une tige-boucle contiguë et issue du même brin (voir figure 3.1). Ces structures ont été identifiées chez de nombreux types d’ARN et démontrent un rôle important dans l’initiation de la réplication,

le contrôle des phénomènes de translation ou encore dans l'organisation du noyau catalytique de certains ARN (Dam et al., 1992; Pleij, 1995). La première résolution de la structure d'un pseudo-nœud a été réalisée par l'équipe de Hilbers (Kolk et al., 1998).

Une structure de pseudo-nœud a également été élucidée par RMN dans le cas de l'ADN (Ulyanov et al., 1998). L'importance biologique de ce type de structure n'est pas totalement démontrée, mais ce travail montre que l'ADN est capable d'adopter, lui aussi, des structures éloignées de la conformation hélicoïdale.

### **3.3 Interactions tige-boucle hélice**

A ce jour il existe un seul exemple de ce type d'interaction entre une tige-boucle et une hélice. Il s'agit de la structure cristalline de l'ARN ribosomique à tête de marteau où une des trois tétra-boucles GAAA est en interaction dans le petit sillon de l'hélice II (Pley et al., 1994). On peut constater que les structures de ces éléments, seuls ou en interaction, sont quasiment identiques. L'interaction est stabilisée par une liaison hydrogène par base triple et trois autres liaisons hydrogène impliquant le groupement 2'OH.

### **3.4 Interactions boucle-boucle**

L'appariement boucle/boucle de deux épingles à cheveux constitue un motif en "kissing-complex" (figure 3.1). Pour illustrer ce type de structure on peut rappeler les travaux de Crothers sur le système RNA Ii/RNA Iii du plasmide ColE1 chez *Escherichia coli* (Lee and Crothers, 1998; Marino et al., 1995) (figure 3.2). Ce type d'interaction est impliqué ici dans la régulation de la réplication du plasmide. Les structures en "kissing-complex" sont stabilisées par des liaisons de type Watson-Crick entre les deux tiges boucles mais aussi par des motifs stabilisants non classiques (motifs en fermeture éclair, liaisons hydrogène, stacking) difficiles à cataloguer.

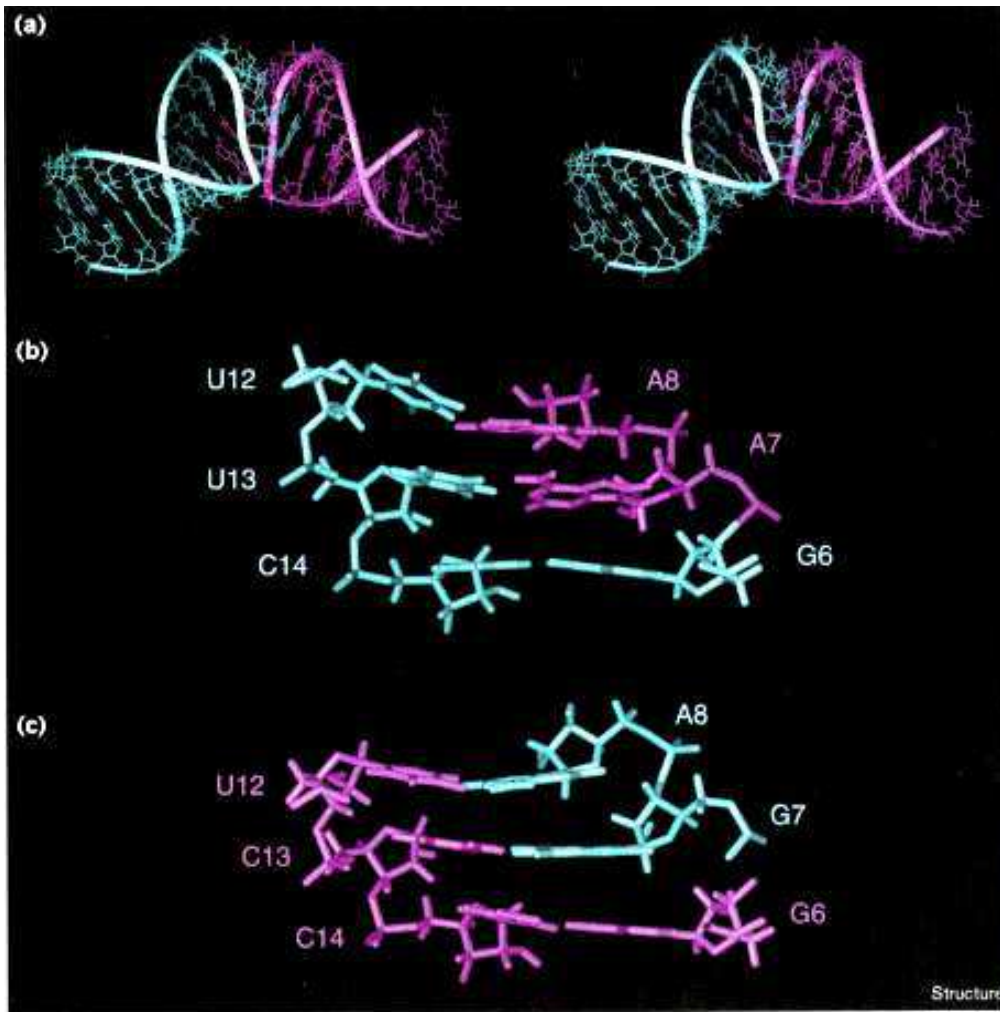


Figure 3.2 : a) Vue stéréoscopique du complexe boucle-boucle d'ARN de ColE1, d'après Lee et Crothers (1998). b) et c) représentation des liaisons Watson-Crick de la région d'interaction entre les 2 tiges-boucles.

# Détermination structurale d'acides nucléiques par RMN

Les deux techniques expérimentales les mieux adaptées pour la détermination structurale des macromolécules biologiques sont la radiocristallographie des rayons X (RX) et la résonance magnétique nucléaire (RMN). Ces deux techniques sont complémentaires. La radiocristallographie des rayons X étudie la forme cristalline et donne des informations sur la position des atomes lourds dans l'espace. La RMN s'applique aux molécules en solution donc a priori dans des conditions plus proches des milieux biologiques, et donne des informations sur les distances inter-atomiques, les angles dièdres et la dynamique interne de la molécule. En combinant une de ces deux méthodes aux techniques de modélisation moléculaire, nous sommes en mesure de déterminer la conformation qu'adopte la molécule étudiée.

La limitation en taille est un des obstacles les plus importants de l'étude par RMN de structures de macromolécules biologiques. Elle est directement liée au problème du recouvrement des raies de résonance des protons. De plus, des problèmes de relaxation liés au temps de corrélation des molécules de grandes tailles interviennent et l'utilisation de champs magnétiques puissants (750, 800 voir 900 MHz), difficilement accessibles, devient nécessaire. La RMN demeure néanmoins une des principales méthodes permettant l'accès à la structure et aux paramètres dynamiques des acides nucléiques.

## **1. Stratégies d'attribution des acides nucléiques**

De nombreuses revues décrivent la démarche générale à employer pour l'établissement de structures tridimensionnelles d'ADN ou d'ARN (Borer et al., 1997; Puglisi and Puglisi, 1998; Wijmenga and van Buuren, 1998). Nous aborderons ici uniquement le protocole d'attribution des séquences non-marquées aux isotopes stables.

## 1.1 Le déplacement chimique du proton

Les déplacements chimiques des protons d'un acide nucléique s'échelonnent sur une plage de déplacement chimique allant de 0 à 15ppm. L'identification des différentes zones de résonance, correspondant aux différents types de protons de l'acide nucléique étudié, constitue la première étape essentielle à l'interprétation de la structure tridimensionnelle. Les différentes zones sont représentées sur la figure 1.1.

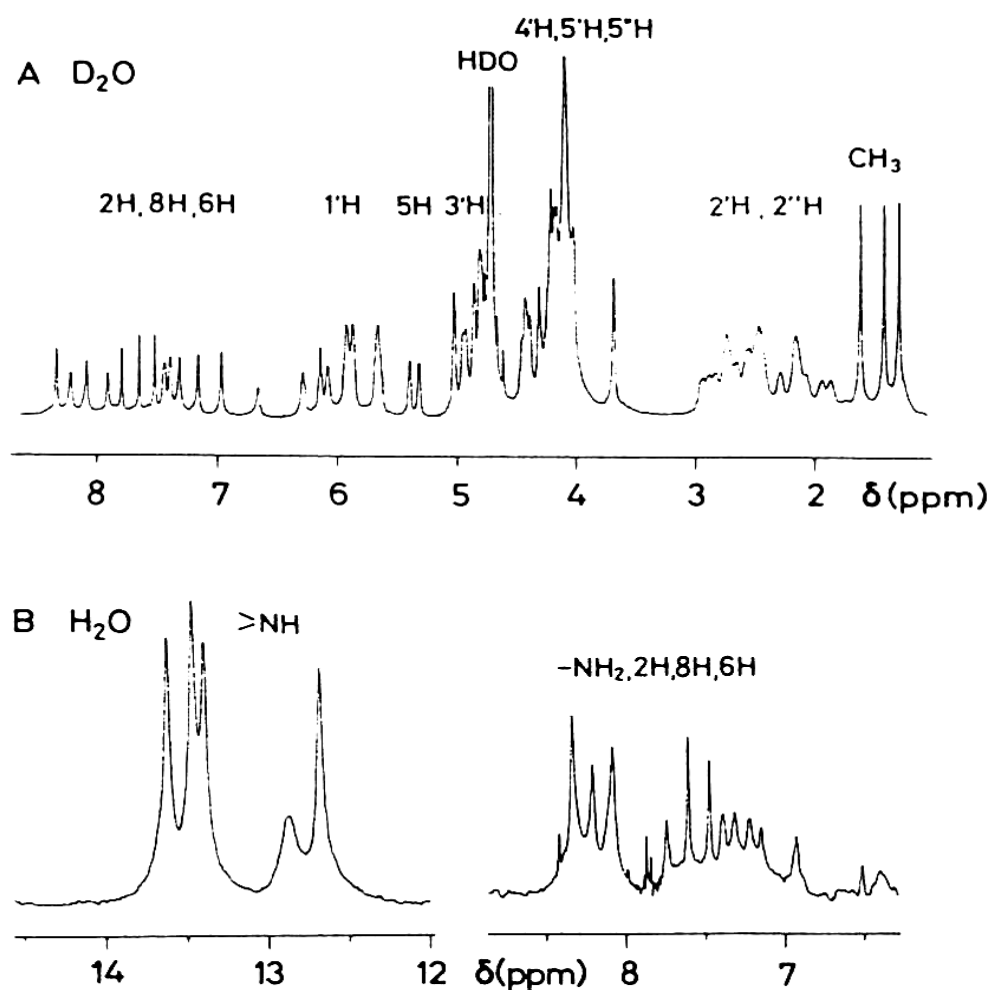


Figure 1.1 : déplacements chimiques standards des protons des acides nucléiques. Spectre proton de d(CGATTAATGC)<sub>2</sub> à 500 MHz. (A) Spectre dans l'eau lourde à 28°C (B) Portions de spectre effectué dans l'eau légère à 15°C.

Dans le cas d'un ADN, mis à part les protons du sucre 4', 5' et 5'' dont le recouvrement est important, chaque type de proton est facilement identifiable par sa plage de résonance. En revanche dans le cas d'un ARN, chaque atome de carbone du cycle ribofurannique est lié



à un atome d'oxygène et, de ce fait, les déplacements chimiques des protons 2', 3', 4', 5' et 5'' s'étendent sur une plage d'un ppm environ (pour 3 à 4 ppm dans le cas d'un ADN).

Cette différence fondamentale a beaucoup d'incidence dans l'établissement de structure d'ARN par RMN. En effet, la non-attribution de la plupart des protons du sucre a pour conséquence d'obtenir moins de contraintes de distances et entraîne un manque de précision par rapport à l'ADN.

## 1.2 Le déplacement chimique du phosphore

Pour les oligonucléotides, la gamme des déplacements chimiques du phosphore, qui a un spin  $\frac{1}{2}$  à l'abondance naturelle, est assez réduite (environ un ppm) ce qui rend l'attribution souvent difficile. Une déviation par rapport à la forme standard se traduit par des résonances "anormales" à haut champ ou bas champ selon la déformation considérée. Ces déplacements ont été reliés à la détorsion de l'ADN (Wilson and Jones, 1982) et aux angles de torsion du squelette (Delepierre et al., 1988; Giessner-Prettre and Pullman, 1987; Gorenstein, 1984). Lorsque l'on repère un étalement des résonances du phosphore, il est alors intéressant d'effectuer des corrélations hétéronucléaires de types HSQC, HMQC ou hétérotocsy pour essayer d'attribuer ces résonances.

## 1.3 Les liaisons hydrogène

A la différence de ceux des protéines, les protons labiles des acides nucléiques s'échangent en général trop rapidement avec l'eau pour être observés, même dans une solution H<sub>2</sub>O très acide. Cependant, dès qu'ils sont engagés dans une liaison hydrogène, leur vitesse est ralentie et on peut les observer. Notons qu'il ne faut pas éliminer le signal de l'eau par présaturation, mais par des séquences d'impulsion comme le "jump and return" (Plateau and Guéron, 1982) ou le "watergate" (Piotto et al., 1992). En effet la présaturation égalise les populations des niveaux  $\alpha$  et  $\beta$  du signal de l'eau, dès lors on ne peut observer l'échange que fait celle-ci avec les protons échangeables.

Les liaisons hydrogène apportent des renseignements structuraux très riches,

notamment les protons iminos (thymines, ou uraciles dans le cas des ARN, et guanines) qui résonnent à très bas champ s'ils sont appariés dans des liaisons de type Watson-Crick

#### 1.4 Méthodologie d'attribution séquentielle

Il est toujours délicat de présenter une procédure d'identification des raies de résonance d'un oligonucléotide, en ce sens qu'il n'existe pas réellement de "bonne et définitive" méthode pour l'attribution totale d'un spectre. Nous pouvons néanmoins présenter une méthodologie classique d'attribution où l'utilisateur devra faire de constants va et vient entre chaque étape pour déterminer l'ensemble des résonances.

La première étape repose sur l'observation dans l'eau légère des résonances des protons échangeables (imino et amino). Cette étape permet de mettre en évidence l'appariement des bases et donc la structure secondaire. La mise en jeu de l'effet NOE permet une attribution partielle de certains protons. En effet à partir des résonances imino d'un uracile ou d'une thymine on peut identifier le proton H<sub>2</sub> de l'adénine qui lui est appariée ainsi que les protons aminos de cette adénine. De la même façon, la proximité du proton imino de la guanine avec le proton H<sub>5</sub> et les protons aminos de la cytosine qui lui est appariée permet l'identification de ces derniers.

La seconde étape se déroule généralement dans D<sub>2</sub>O. Les expériences utilisant le transfert de cohérence à travers les liaisons (expériences COSY ou TOCSY par exemple) permettent l'observation des corrélations H<sub>5</sub> et H<sub>6</sub> rencontrées uniquement pour les cytosines ou les uraciles. Dans les ADN, ces mêmes expériences permettent l'identification des CH<sub>3</sub> des thymines par observation du couplage scalaire H<sub>6</sub> et CH<sub>3</sub>. La conformation du sucre est aussi déterminée par ces expériences et notamment l'expérience TOCSY (le chapitre suivant détaillera la portée structurale de ces informations).

Dans la troisième étape les expériences NOESY effectuées à différents temps de mélanges conduisent à l'attribution séquentielle des protons non-échangeables via les connectivités NOE rencontrées entre le proton H<sub>1'</sub> et l'aromatique H<sub>6</sub> ou H<sub>8</sub> de son propre résidu ainsi qu'avec l'aromatique H<sub>6</sub> ou H<sub>8</sub> du résidu voisin en 3'. De la même façon, à haut temps de

mélange, l'attribution séquentielle dans le cas des ADN peut se faire également avec les protons  $H_2'$ ,  $H_2''$  et  $H_3'$ . Pour les ARN, le recouvrement de la région des protons  $H_2'$ ,  $H_3'$ ,  $H_4'$ ,  $H_5'$  et  $H_5''$  empêche généralement l'identification complète de tous les protons des sucres et diminue de façon conséquente la précision des structures.

Enfin, les expériences de corrélations hétéronucléaires phosphore/proton permettent la confirmation, voir la détermination des protons  $H_3'$ ,  $H_4'$ ,  $H_5'$  et  $H_5''$ . Néanmoins, la faible dispersion en déplacement chimique du  $^{31}\text{P}$  conduit généralement à des recouvrements pouvant rendre cette démarche inefficace.

## 2. Informations structurales et dynamiques révélées par RMN

### 2.1 La mesure de distance

Considérons deux spins homonucléaires proches dans l'espace, chacun d'eux exerce sur l'autre un champ magnétique local. Ce phénomène est plus spécifiquement appelé relaxation dipolaire croisée ou effet Overhauser. Le système de spins est décrit par quatre niveaux d'énergie (figure 2.1)

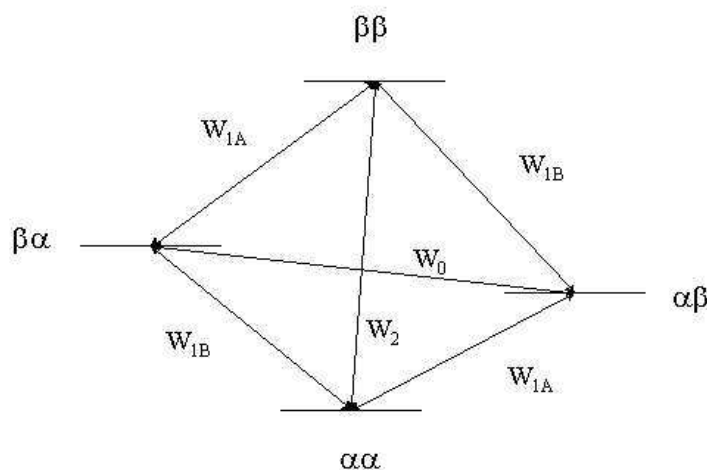


Figure 2.1 : Schéma des transitions possibles pour deux spins proches dans l'espace.

Les équations de Solomon (voir l'ouvrage d'Abragam (Abragam, 1961)) décrivent la contribution de la relaxation croisée sur la variation de la magnétisation longitudinale  $M_i$

du spin i "relaxé" par le spin j :

$$\frac{dM_i}{dt} = -\rho_i(M_i - M_{i0}) - \sigma_{ij}(M_j - M_{j0})$$

- $M_{i0}$  et  $M_{j0}$  représentent les magnétisations des spins i et j à l'équilibre.
- $\sigma_{ij} = W_2(ij) - W_0(ij)$
- $\rho_{ij} = W_0(ij) - 2W_1(i) + W_2(ij)$

$W_0$ ,  $W_1$  et  $W_2$  sont les probabilités de transition à 0,1 et 2 quanta dues à l'interaction dipole-dipole des deux spins étudiés. Elles sont toutes proportionnelles à  $1/r_{ij}^6$  et donc effectives en pratique pour des distances inférieures à 5 Å (en raison de cette dépendance en  $1/r_{ij}^6$ ). L'effet Overhauser nucléaire (NOE) représente la perturbation de la magnétisation du spin I due à la proximité du spin S.

La mesure du NOE s'effectue soit en saturant sélectivement un spin et en mesurant les changements d'intensités de la magnétisation des spins qui lui sont couplés dipolairement, ou par des expériences à deux dimensions (NOESY). Les différentes probabilités de transition s'expriment en fonction de la densité spectrale et donc en faisant intervenir le temps de corrélation de la molécule étudiée. L'expression du NOE est donnée par la formule suivante :

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

K dépend du temps de corrélation  $\tau_c$  du mouvement du vecteur joignant les deux spins. Cette relation n'est plus valable pour les temps de mélange longs, où le proton j va transférer l'aimantation issue du proton i sur un troisième proton k qui lui est proche (à j), phénomène bien connu sous le nom de diffusion de spin. Il apparaît un NOE entre les proton i et k sans que ceux-ci soient nécessairement proches, tandis que l'intensité du NOE (i/j) est diminuée et ne satisfait plus à la relation.

Seules les informations de courtes distances, inférieures à 5 voire 6 Å, sont recueillies. De ce fait, seules les distances locales sont accessibles avec une bonne précision. La conformation globale sera déterminée par un nombre conséquent de distances locales,

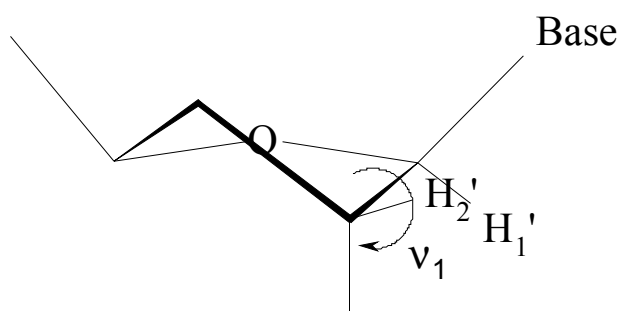
d'où l'intérêt d'avoir un maximum d'informations sur les interactions dipolaires intra et inter nucléotidiques (Wijmenga and van Buuren, 1998).

## 2.2 L'élucidation du plissement du sucre

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 qui nous intéressent sont ceux qui vont permettre d'obtenir la conformation de l'unité ribose ou désoxyribose. En fait, la différence de conformation sud et nord peut être déterminée par la seule valeur de l'angle dièdre  $\nu_1$  donnée par la constante de couplage  ${}^3J_{H1'-H2'}$  (figure 2.2).

Cette détermination du plissement du sucre par l'analyse qualitative de la seule constante de couplage  ${}^3J_{H1'-H2'}$  est principalement utilisée dans le cas des ARN où les protons  $H_3'$  et  $H_4'$  sont rarement élucidés (du fait du recouvrement). D'autres constantes de couplage permettent la détermination de la conformation du sucre ribose ou désoxyribose. Les valeurs types de ces constantes de couplage, qui sont généralement déterminées uniquement dans le cas des ADN, sont résumées dans le tableau 2.1.

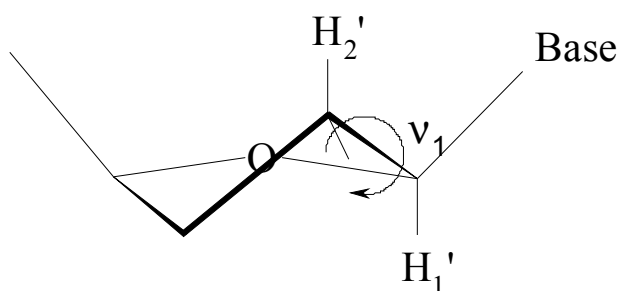
### Conformation $C_3'$ endo



$$\nu_1 \cong 90^\circ$$

$${}^3J_{H_1'H_2'} \cong 1 \text{ Hz}$$

### Conformation $C_2'$ endo



$$\nu_1 \cong 180^\circ$$

$${}^3J_{H_1'H_2'} \cong 10 \text{ Hz}$$

Figure 2.2 : Variation de la constante de couplage  ${}^3J_{H1'-H2'}$  avec la conformation du sucre.

| Constantes de Couplage (Hz) | C <sub>2</sub> '-endo (Sud) | C <sub>3</sub> '-endo (Nord) |
|-----------------------------|-----------------------------|------------------------------|
| H1'-H2'                     | 10,0                        | 1,8                          |
| H1'-H2''                    | 5,7                         | 8,0                          |
| H3'-H2'                     | 5,8                         | 7,3                          |
| H3'-H2''                    | 1,2                         | 9,7                          |
| H3'-H4'                     | 1,4                         | 8,1                          |

Tableau 2.1 : Tableau des valeurs types des constantes de couplage dans les conformations sud et nord des acides nucléiques.

### 2.3 La détermination de la conformation du squelette phosphodiester

Le squelette phosphodiester est l'architecture même de l'acide nucléique. Malheureusement, peu d'informations peuvent être facilement tirées de l'étude simple de la RMN du proton, suite au manque de protons dans cette région. Comme nous l'avons vu précédemment, l'étalement faible des déplacements chimiques du phosphore ne nous permet pas l'interprétation fine de la conformation du squelette et des 6 angles de torsion qui le composent. Les angles de torsions sont représentés sur la figure 2.3.

La mesure des constantes de couplage scalaire hétéronucléaire  $^3J(^1\text{H}-^{31}\text{P})$  permet d'établir la valeur des angles de torsions  $\beta, \gamma$  (Wijmenga et al., 1993) et  $\epsilon$  (Kim et al., 1992). Les angles de torsions  $\alpha$  et  $\zeta$  ne sont pas définis du fait de l'absence directe de constantes de couplage ou de NOE à ce niveau. L'angle de torsion  $\delta$  et l'angle de torsion  $\nu_3$  du sucre étant intimement corrélés, la valeur de l'angle  $\delta$  est dépendante de l'angle de pseudorotation du cycle ribose. La constante de couplage  $^3J(\text{H}_3'-\text{H}_4')$  caractérise donc la conformation de cet angle dièdre.

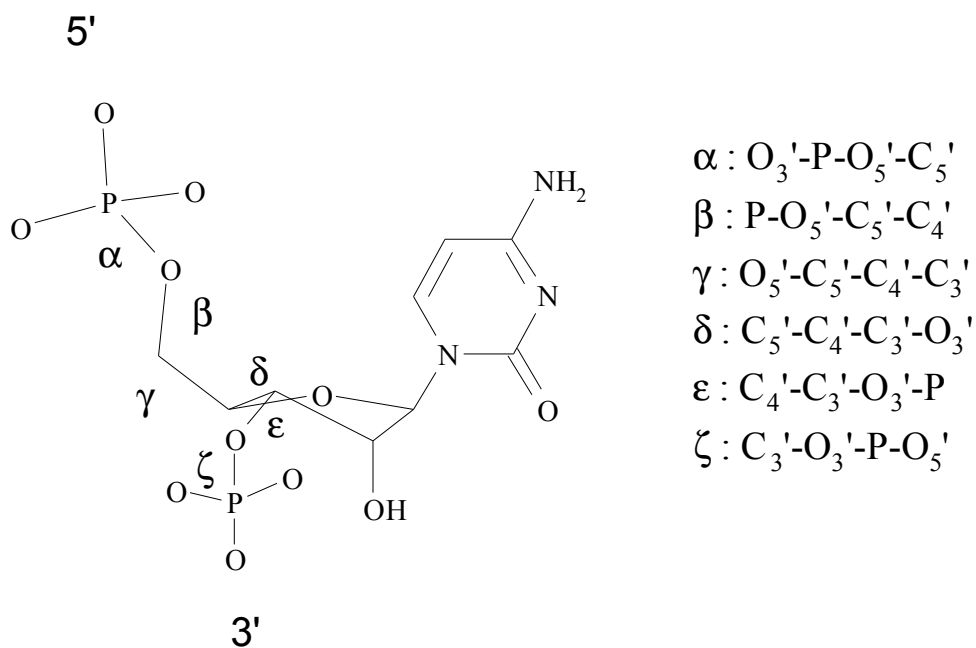


Figure 2.3 : Angles de torsion définissant la conformation des acides nucléiques.

## 2.4 L'interprétation de la dynamique du système

Bien plus que la simple détermination d'angles et de distances inter-atomiques, renseignant sur la structure moyenne en solution de la macromolécule étudiée, la RMN permet d'obtenir des précisions sur la dynamique de la macromolécule et par conséquent sur le comportement de la macromolécule.

La mesure des vitesses de relaxation associées aux différents phénomènes de la relaxation va permettre de décrire la dynamique globale et interne des molécules. La distribution des fréquences, correspondant aux différences de niveaux d'énergie, est représentée par des fonctions de densités spectrales qui elles-mêmes dépendent des temps de corrélation associés aux différents mouvements. Les zones d'interaction entre acides nucléiques ou entre un acide nucléique et une protéine ont généralement une dynamique particulière que les études de relaxation permettent de mettre en évidence.

De plus, la détermination du temps de corrélation global de la molécule, c'est à dire le temps que met la molécule étudiée à faire une rotation sur elle-même, indique de façon qualitative la "taille" de cette molécule. Bien évidemment, sa taille influe sur le temps de

rotation de celle-ci et permet, très efficacement, de différencier un monomère d'un dimère ou d'autre degré de polymérisation.

### ***3 Les nouvelles approches***

Les expériences de RMN à deux dimensions permettent de surmonter en grande partie le problème de la superposition des signaux, et de pouvoir réaliser l'étape essentielle d'attribution séquentielle à l'aide des seuls spectres du proton. Il est généralement admis que l'attribution séquentielle en  $^1\text{H}$ , des acides nucléiques contenant plus de 30 nucléotides (10 000 Daltons), est difficile.

Au cours des dix dernières années, différentes avancées technologiques ont conduit à accroître les possibilités d'études de macromolécules biologiques par RMN. L'évolution de l'électronique permet maintenant d'accéder à des spectromètres travaillant jusqu'à 900 MHz. De plus, le design des sondes et la baisse continue du rapport signal sur bruit, permet d'étudier des échantillons de moins en moins concentrés et de générer des séquences d'impulsions de plus en plus complexes et spécifiques ; c'est le cas notamment des nouvelles cryo-sondes qui permettent de diminuer très sensiblement le "bruit blanc" inhérent à l'électronique et autorise l'étude des échantillons à 0,3mM à 600MHz.

Au delà des performances matérielles, de nouvelles approches et méthodologies sont actuellement disponibles. Le but de ce chapitre n'est pas de faire une liste exhaustive de ces différents points mais de montrer les nouveaux outils qui doivent améliorer grandement la détermination de structures d'acides nucléiques.

#### **3.1 L'enrichissement aux isotopes stables**

Le développement, lors de cette dernière décennie, des techniques de synthèse biomoléculaire, autorise désormais l'obtention de séquences d'ARN uniformément marqué aux isotopes stables ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) et permet de surmonter les principales difficultés dues au recouvrement. Néanmoins, lors d'étude d'ARN de plus de 40 nucléotides, le



recouvrement devient important et l'élargissement des raies de résonance empêche la détermination structurale. L'introduction sélective de deutérium, en éliminant certaines raies de résonance, permet de clarifier le spectre (Glemarec et al., 1996; Yamakage et al., 1993) mais également de réduire notablement la relaxation dipolaire et donc la largeur des raies (Nikonowicz et al., 1997; Nikonowicz et al., 1997). Le développement de ces techniques permet aussi d'augmenter la sensibilité des transferts de polarisation  $^{13}\text{C} \rightarrow ^{15}\text{N}$  ou bien  $^{13}\text{C} \rightarrow ^{31}\text{P}$  dans des molécules uniformément marquées (Cromsig et al., 2000).

Dans le cas d'études d'ADN, le marquage isotopique est moins développé et ce pour plusieurs raisons :

- L'ADN offre une plus grande plage de déplacement chimique par rapport à un ARN. En conséquence, il existe moins d'ambiguïtés dans l'attribution des spectres protons-protons.
- La synthèse biochimique par PCR (polymerase chain reaction) ou par voie enzymatique (Smith et al., 1997) de grandes quantités d'ADN est plus difficile à mettre en œuvre.

Notons cependant que certaines études d'ADN enrichis isotopiquement commencent à émerger, pour exemple voir (Ohki and Kainosho, 1999).

### 3.2 L'expérience TROSY

Les expériences TROSY (transverse relaxation optimized spectroscopy) ont été développées par l'équipe du professeur Wüthrich (Pervushin et al., 1997). Cette méthode, tout d'abord élaborée pour les protéines, a été ensuite adaptée pour les acides nucléiques. Si l'on prend par exemple le cas d'un couple  $^1\text{H}-^{15}\text{N}$ , un spectre COSY classique découplé conduit à l'observation de raies de résonance larges dues à la grande constante de couplage  $^{15}\text{N}-^1\text{H}$  qui est d'environ 95 Hz. Si on ne découple pas le système, on observe alors naturellement un multiplet. Pour certaines raies de résonance, les contributions de l'anisotropie de déplacement chimique (CSA) et de l'interaction dipôle-dipôle s'opposent, entraînant dans ce multiplet un affinement d'une raie. L'expérience TROSY permet de sélectionner la raie la plus fine parmi les différentes composantes du multiplet et d'augmenter la sensibilité et la résolution des spectres  $^1\text{H}-^{15}\text{N}$  ou  $^1\text{H}-^{13}\text{C}$ .

### 3.3 L'étude RMN dans les milieux orientés

Une des approches les plus prometteuses est l'étude du couplage résiduel dipolaire dans les milieux orientés. Contrairement aux autres nouvelles méthodes, celle-ci n'a pas pour but de faciliter l'attribution séquentielle, mais offre la possibilité d'une nouvelle source de renseignements structuraux. Cette nouvelle technique a fait l'objet d'une récente revue (Prestegard et al., 2000).

La macromolécule biologique est introduite dans un milieu cristallin liquide (constitué par exemple de bicelles (Tjandra and Bax, 1997) ou de bactériophages (Clare et al., 1998)) qui ont la particularité de s'aligner en présence d'un champ magnétique. Les molécules s'orientent partiellement dans le milieu pour des raisons stériques et cette orientation, qui est anisotrope, permet l'expression de constantes de couplages dipolaires résiduels.

La mesure de ces faibles constantes de couplage permet de déterminer des angles entre des vecteurs  $^1\text{H}-^{15}\text{N}$  ou  $^1\text{H}-^{13}\text{C}$  qui présentent une interaction dipolaire observable. L'équation suivante représente l'expression analytique de ce couplage dipolaire résiduel entre les noyaux P et Q.

$$D_{PQ}(\theta, \Phi) = -S \frac{\mu_0}{4\pi} \gamma_P \gamma_Q \left[ \frac{A_A(3\cos(\theta)^2 - 1) + \frac{3}{2} A_R \sin(\theta)^2 \cos(2\Phi)}{4\pi^2 r_{PQ}^3} \right]$$

S représente le paramètre d'ordre généralisé pour le mouvement interne du vecteur PQ dont la distance est donnée par le terme  $r_{PQ}$ ;  $\mu_0$  est la perméabilité magnétique dans le vide. Les termes  $\gamma_P$  et  $\gamma_Q$  sont les rapports gyromagnétiques des deux spins, et  $\theta$  et  $\Phi$  sont les coordonnées sphériques du vecteur PQ selon l'axe A.

Ce couplage étant proportionnel à  $1/r^3$ , il permet de considérer des angles entre des vecteurs éloignés séquentiellement. Ce type d'information longue distance est le seul type d'information considéré jusqu'alors inaccessible par RMN. Ces informations angulaires servent soit de mesure complétant les contraintes d'effet NOE et d'angle dièdres (elles sont d'ailleurs maintenant incorporables facilement dans les scripts de modélisation

moléculaire du logiciel CNS) ou sont utilisables seules pour la détermination de structures (Delaglio et al., 2000).

# Modélisation moléculaire d'acides nucléiques à partir de données RMN

Le but de la modélisation moléculaire est d'obtenir une "image", ou plus précisément un jeu de coordonnées de la molécule étudiée, répondant aux données RMN. Nous avons, lors de ce travail de thèse, utilisé les logiciels XPLOR (Brünger, 1992) et CNS (Brünger et al., 1998) qui utilisent les paramètres d'énergie potentielle de CHARMM (Brooks et al., 1983) améliorés pour les acides nucléiques (Nilsson and Karplus, 1986).

Dans un premier temps, nous allons décrire brièvement la fonction d'énergie potentielle utilisée par ces deux logiciels puis nous exposerons les différentes méthodologies de modélisation pour élucider une structure RMN.

## **1. La fonction d'énergie potentielle des logiciels XPLOR et CNS**

Le calcul de l'énergie de la molécule étudiée peut se décomposer en deux principaux éléments, eux mêmes décomposables. L'énergie totale s'écrit donc (1) :

$$E_{TOT} = E_{CHEM} + E_{EFFECTIF} \quad (1)$$

Le terme d'énergie chimique  $E_{CHEM}$  décrit les interactions covalentes dans la molécule ainsi que les interactions entre atomes non liés. Le terme d'énergie effective  $E_{EFFECTIF}$  est donné par les mesures RMN ; on parle alors de contraintes énergétiques. Nous allons en premier lieu décrire les paramètres chimiques.

### **1.1 Les termes chimiques**

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

$$E_{\text{CHEM}} = \sum_{P=1}^N \left\{ w_{\text{BOND}}^P E_{\text{BOND}} + w_{\text{ANGL}}^P E_{\text{ANGL}} + w_{\text{DIHE}}^P E_{\text{DIHE}} + w_{\text{IMPR}}^P E_{\text{IMPR}} + w_{\text{VDW}}^P E_{\text{VDW}} + w_{\text{ELEC}}^P E_{\text{ELEC}} + w_{\text{HBOND}}^P E_{\text{HBOND}} \right\} \quad (2)$$

E correspond à chaque terme énergétique et à sa propre expression, expression que l'on peut modifier dans le programme XPLOR ou CNS. Le poids (w) est donné pour chaque terme énergétique.

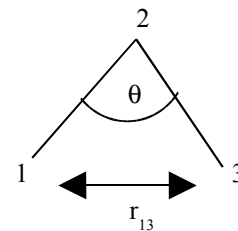
Le terme d'énergie de liaison,  $E_{\text{BOND}}$  correspond à l'énergie covalente entre deux atomes liés par une liaison chimique. Il est défini par l'équation (3)

$$E_{\text{BOND}} = \sum_{\text{BOND}} k_{\text{BOND}} (r - r_0)^2 \quad (3) \quad \text{A} \text{---}^r \text{---} \text{B}$$

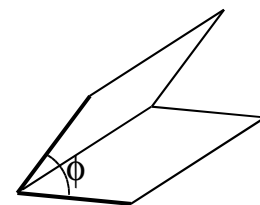
Le terme  $k_{\text{BOND}}$  est défini empiriquement, principalement par des mesures précises de spectroscopie vibrationnelle de type IR ou RAMAN. Les valeurs de  $r_0$  sont obtenues à partir de données cristallographiques sur de petites molécules.

Les termes d'énergie angulaire et d'angles dièdres ( $E_{\text{ANGL}}$  et  $E_{\text{DIHE}}$ ) sont définis d'une façon similaire dans les équations (4,5). L'angle dièdre est l'angle formé par deux plans. Cette énergie représente donc une énergie de torsion.

$$E_{\text{ANGL}} = \sum_{\text{ANGLES}} \left\{ k_{\theta} (\theta - \theta_0)^2 \right\} \quad (4)$$



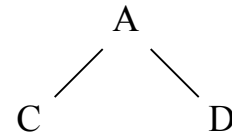
$$E_{\text{DIHE}} = \sum_{\text{DIEDRES}} \sum_{i=1,m} k_{\varphi_i} (1 + \cos(n\varphi_i + \delta_i)) \quad (5)$$



n représente la périodicité et  $\delta$  la phase à l'origine.

Le paramètre d'angle impropre se rapproche de la fonction énergétique d'angle dièdre (équation 6). L'angle dièdre impropre est défini comme étant l'angle entre les plans contenant les atomes A,B,C et B,C,D.

$$E_{\text{IMPR}} = \sum_{\text{IMPROPRES}} \sum_{i=1,m} k_{\varphi_i} (\varphi_i - \varphi_0)^2 \quad (6)$$



Le terme d'énergie de van der Waals ( $E_{\text{VDW}}$ ) correspondant aux interactions entre atomes non liés (équation 7), est défini par la fonction de Lennard-Jones (figure 1.1) qui simule de façon la plus exacte le phénomène d'attraction et de répulsion traduisant les interactions à distance des atomes non-liés. Ce paramètre énergétique rend compte de l'encombrement stérique des atomes.

$$E_{\text{VDW}} = \sum_{ij} -4E_{\text{MIN}} \left[ \left( \frac{R_{\text{MIN}}}{r\sqrt{2}} \right)^{12} - \left( \frac{R_{\text{MIN}}}{r\sqrt{2}} \right)^6 \right] \quad (7)$$

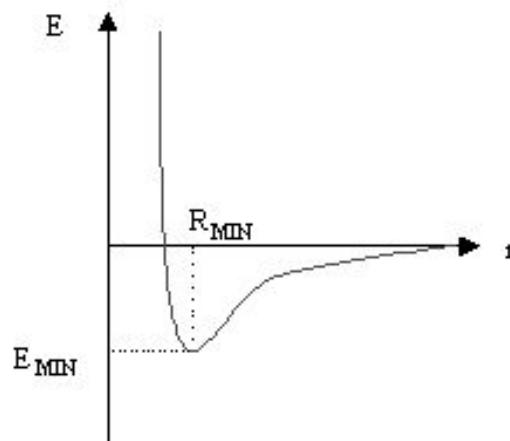


Figure 1.1 : représentation de la fonction de Lennard-Jones

Notons, de plus, que dans le cas des méthodes de recuit simulé le terme d'énergie de van der Waals est simplifié pour être purement répulsif.

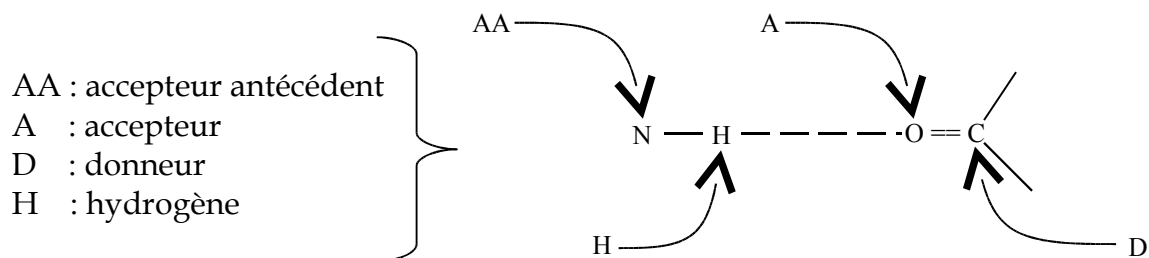
La plupart des atomes liés de façon covalente portent une charge partielle  $q_i$ , même si la

molécule est neutre, parce que les liaisons covalentes entre des atomes de types différents conduisent à une distribution électronique asymétrique. L'équation (8) rend compte de l'énergie électrostatique de la molécule.

$$E_{\text{ELEC}} = \sum_{ij} \left( \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} \right) \quad (8)$$

Le dernier terme d'énergie correspond à celui de la liaison hydrogène (9), il représente une interaction entre un donneur de proton et un atome accepteur. Dans une telle interaction, les 2 atomes s'approchent d'une distance inférieure à la somme des rayons de van der Waals. La plus grande partie de l'énergie d'une liaison hydrogène est de nature électrostatique. Cependant les fonctions de van der Waals doivent être modifiées afin de permettre l'approche des deux atomes. Dans XPLOR le terme  $E_{\text{HBON}}$  est égal à :

$$E_{\text{HBON}} = \sum_{i < j} \left\{ \left( \frac{A_{AD}}{r_{AD}^i} - \frac{B_{AD}}{r_{AD}^j} \right) \cos^m(\theta_{AHD}) \cos^n(\theta_{AAAH}) \right\} \quad (9)$$



L'utilisation d'un terme d'énergie de liaison hydrogène permet le contrôle et l'orientation des atomes impliqués dans la liaison. Mais il représente simplement un cas particulier de l'énergie électrostatique et de l'énergie de van der Waals, si bien que ce terme a été abandonné dans le logiciel CNS (qui correspond à la version récente et améliorée de XPLOR).

## 1.2 Les termes effectifs

Lorsque l'on détermine une structure à partir de données expérimentales, il est important de faire apparaître ces informations expérimentales sous forme d'énergies de contrainte (voir équation 2) : ce sont les termes effectifs.

Le terme d'énergie effective peut se décomposer en plusieurs paramètres correspondant à des résultats expérimentaux particuliers (10).

$$E_{\text{EFFECTIF}} = w_{\text{NOE}} E_{\text{NOE}} + w_{\text{CDIH}} E_{\text{CDIH}} + w_{\text{PLAN}} E_{\text{PLAN}} \quad (10)$$

Le premier terme de cet équation est l'énergie NOE. Il représente sans aucun doute l'énergie de contrainte la plus importante pour la détermination structurale. Les distances interprotons déduites des NOE sont exprimées sous la forme d'un intervalle de valeurs ( $d_{\text{inf}}, d_{\text{sup}}$ ). Il existe différentes options permettant de modifier la forme du potentiel de contrainte NOE. Pour la détermination des structures, un potentiel harmonique à fond plat (option square-well) est le plus couramment utilisé (11).

$$E_{\text{NOE}}^{\text{CLASS}} = \min(\text{ceil}, k)\Delta \quad (11)$$

avec

$$\Delta \begin{cases} (R - d_{\text{sup}})^2 & \text{si } R > d_{\text{sup}} \\ 0 & \text{si } d_{\text{inf}} < R < d_{\text{sup}} \\ (d_{\text{inf}} - R)^2 & \text{si } R < d_{\text{inf}} \end{cases}$$

Nous pouvons tracer la courbe de potentiel correspondant à l'option SQUARE-well (figure 1.2).



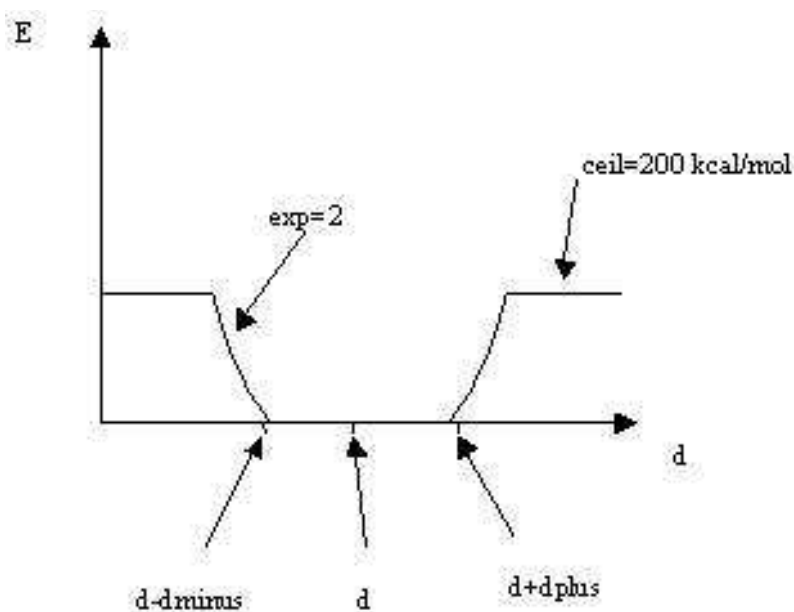


Figure 1.2 : représentation schématique de la courbe de la fonction d'énergie potentielle NOE.

Dans les étapes de recuit simulé, les structures initiales sont assez éloignées du minimum recherché, l'expression analytique de la pseudo-énergie de contraintes conduit à des valeurs très élevées qui risquent de conduire à des instabilités lors du calcul de la trajectoire. Pour s'en prémunir, on adoucit la forme du potentiel en prenant une asymptote linéaire à l'infini (équation 12).

$$E_{\text{NOE}} = k \left[ a + \frac{b}{(d_{\text{sup}} - R)^2} + k_{\text{ASYM}}(d_{\text{sup}} - R) \right] \quad \text{si } R > d_{\text{ASYM}} \quad (12)$$

Les paramètres  $a$  et  $b$  sont calculés de telle sorte qu'il y ait continuité de la fonction et les mêmes dérivés en  $d_{\text{max}}$  et  $d_{\text{min}}$ . La pente  $k_{\text{ASYM}}$  peut être ajustée afin que le poids de la contrainte soit plus ou moins fort pour de grandes valeurs de déviation.

Le deuxième terme de l'équation (10) est l'énergie due aux contraintes d'angles dièdres déterminées par la RMN. Dans la détermination de structures d'acides nucléiques il permet d'imposer les différentes conformations du sucre et la conformation du backbone. L'expression de son potentiel de contrainte est très proche de celui de l'effet NOE (13).

$$E_{CDIH} = S \sum C.well(modulo_{2\pi}(\phi - \phi_0), \Delta\phi)^{ed} \quad (13)$$

$$well(\phi, \phi_0) \begin{cases} \phi - (\phi_0 + \Delta) & \text{si } \phi > (\phi_0 + \Delta) \\ 0 & \text{si } (\phi_0 - \Delta) < \phi < (\phi_0 + \Delta) \\ (\phi_0 - \Delta) - \phi & \text{si } \phi < (\phi_0 - \Delta) \end{cases}$$

Le troisième terme de l'équation 10 est l'énergie de plan. En effet, dans les logiciels XPLOR et CNS, nous avons la possibilité d'imposer un terme d'énergie de plan (14). Il correspond à une énergie d'angle impropre et pénalise les atomes qui sortent des plans définis dans un fichier de contraintes. Elle est très utile dans la modélisation des acides nucléiques pour favoriser les appariements de bases lors des premiers cycles de modélisation.

$$E_{PLAN} = \sum_{g \in \text{group}} w_{PLAN} \sum_{i \in g} g_i^2 \quad (14)$$

$g_i$  correspond à la distance orthogonale de l'atome  $i$  par rapport au plan défini par tous les atomes appartenant au même groupe.

## 2. Principaux algorithmes de modélisation à partir de données RMN

Dans la dernière décennie, d'importants efforts ont été développés afin d'obtenir des structures de molécules à partir de données RMN. Il existe d'ailleurs différentes revues et articles relatant cet intérêt (Brünger and Nilges, 1993; Guntert, 1998; James, 1994; Nilges, 1996). Les principales difficultés viennent de la taille importante des macromolécules, ce qui implique un nombre important de minima locaux, et de la nature des contraintes RMN. Les contraintes d'angles dièdres sont locales et peu précises, et celles impliquant les effets NOE sont à faibles portées (inférieures à 5 voire 6Å). Il existe aujourd'hui plusieurs protocoles permettant d'obtenir une structure de macromolécule sous contraintes RMN, chaque méthode ayant ses avantages et ses inconvénients.

## 2.1 Méthode de géométrie des distances

Cette méthode n'a pas pour but de déterminer la structure finale de la macromolécule, même si cela a déjà été réalisé par le passé dans le cas d'un ADN (Pardi et al., 1988). Toute minimisation d'énergie nécessite un jeu de coordonnées atomiques initiales et le but des méthodes de géométrie des distances est de générer ce jeu de coordonnées en accord avec les données RMN. De cette façon, la seconde étape de dynamique moléculaire ou de mécanique est beaucoup plus rapide car les structures initiales répondent déjà aux contraintes RMN. Il existe deux types de méthodes de géométrie des distances : la méthode de la matrice métrique (utilisant les coordonnées cartésiennes) (Havel and Wuthrich, 1985; Sippl and Scheraga, 1986) et la méthode de la fonction cible (utilisant les coordonnées internes) (Braun and Go, 1985; Guntert et al., 1991a; Guntert et al., 1991b; Gunter et al, 1997). La méthode de géométrie des distances devient de moins en moins employée, l'utilisateur préférant souvent utiliser des structures aléatoires comme points de départ.

## 2.2 Mécanique moléculaire

La mécanique moléculaire consiste à trouver pour une molécule la structure qui correspond au minimum du potentiel décrit dans la première partie. Les coordonnées des atomes peuvent être internes (programme JUMNA pour les acides nucléiques (Lavery et al., 1995)) ou cartésiennes (comme dans le cas de CNS ou XPLOR). Il n'existe pas de méthodes analytiques ou numériques permettant de déterminer la position du minimum absolu. Une méthode consiste à effectuer un grand nombre de minimisations à partir de points de départ différents, puis à choisir le minimum ayant la plus faible valeur. Différentes méthodes sont disponibles pour trouver le minimum de la fonction d'énergie potentielle : la méthode de la plus grande pente et celle du gradient conjugué.

La méthode de la plus grande pente, ou "Steepest descent" (SD) (Wiberg, 1977), consiste à rechercher la direction de la plus grande pente au cours de laquelle la fonction d'énergie décroît le plus rapidement. La direction de minimisation suivie sera celle indiquée par l'opposé du gradient. Cette méthode permet en général de déterminer rapidement une

direction de minimisation, mais converge mal au voisinage des minima.

La méthode du gradient conjugué, ou “Conjugated Gradient” (CG) (Powell, 1977), permet de prendre en compte les valeurs du gradient mais aussi la direction de déplacement de l’étape précédente. Cette méthode donne en général une meilleure convergence que la méthode de la plus grande pente.

Ces deux méthodes sont complémentaires mais elles ne peuvent explorer en totalité l’espace conformationnel disponible d’une macromolécule. Il est donc *a priori* impossible d’être certain d’avoir atteint le minimum absolu de l’hypersurface avec ces méthodes de minimisation. Pour explorer davantage l’espace conformationnel nous avons recours à d’autres méthodes comme celle de la dynamique moléculaire ou du recuit simulé.

### 2.3 Dynamique moléculaire

La dynamique moléculaire appliquée aux macromolécules biologiques permet d’échantillonner des régions de l’espace conformationnel qui ne sont pas atteintes par des méthodes de mécanique moléculaire. Les applications qui en résultent renseignent sur la structure et la dynamique interne de ces molécules. Les toutes premières simulations de macromolécules biologiques ont été réalisées à la fin des années 70 (McCammon et al., 1977). Le formalisme de la dynamique moléculaire utilise la seconde loi de Newton pour décrire le mouvement d’une molécule en fonction du temps (15):

$$\mathbf{F}_i = m_i \cdot \mathbf{a}_i \quad (15)$$

Où  $\mathbf{F}_i$  est le vecteur force agissant sur l’atome  $i$  de masse  $m_i$ ,  $\mathbf{a}_i$  est le vecteur accélération de l’atome  $i$ .

Comme le montre l’équation de Newton, la vitesse et le sens du mouvement atomique dépendent des forces qui s’exercent entre les atomes. En pratique, les atomes possèdent une vitesse initiale propre qui se conforme à l’énergie cinétique totale du système. Cette énergie dépend elle-même de la température de simulation désirée. Elle est habituellement obtenue en élevant lentement la température d’un système moléculaire placé initialement au zéro absolu et en amenant progressivement les atomes qui le constituent à l’énergie

d'équilibre. La dynamique moléculaire permet de calculer la force s'exerçant sur chaque atome et de déterminer différentes informations liées à la connaissance de cette donnée (vitesse et position des atomes pendant un intervalle de temps spécifique, de l'ordre de 1 à 4000 ps, suivant les systèmes moléculaires).

La force  $\vec{F}_i$  qui s'exerce sur un atome  $i$  se trouvant en position  $\vec{r}_i(t)$  est déterminée par la dérivation de la fonction énergie potentielle au voisinage de la position  $\vec{r}_i(t)$  (16) :

$$\vec{F}_i = -\overrightarrow{\text{grad}}_{\vec{r}_i} E(\vec{r}_1, \dots, \vec{r}_n) \quad (16)$$

L'énergie potentielle totale du système,  $E(r)$ , se calcule en utilisant les techniques de mécanique moléculaire. La dynamique moléculaire permet d'étudier la trajectoire des atomes à partir de connaissance des forces s'exerçant sur chaque atome et des masses atomiques. Pour l'ensemble des atomes définissant une structure moléculaire, l'équation de Newton suivante est vérifiée (17):

$$\vec{F}_i = m_i \frac{d^2 \vec{r}_i(t)}{dt^2} \quad (17)$$

Cependant, par suite d'un défaut de solution analytique, les trajectoires ne sont pas directement déterminées par résolution de l'équation de Newton. Une approche différente est utilisée. Elle consiste à calculer :

- Les accélérations atomiques à partir des forces s'exerçant sur chaque atome  $\vec{F}_i = m_i \cdot \vec{a}_i$
- Les vitesses de chaque atome à partir des accélérations atomiques :  $\vec{a}_i = \frac{d\vec{v}_i}{dt}$
- Les positions des atomes à partir des vitesses atomiques :  $\vec{v}_i = \frac{d\vec{r}_i}{dt}$

L'intégration de ces équations se fait en subdivisant la trajectoire en une série d'états discrets, séparés par des intervalles de temps très courts, dont la longueur définit le pas d'intégration  $\Delta t$ . Il est alors possible de déterminer l'accélération à partir de la seconde loi de Newton et la vitesse propre à chaque atome (18) :

$$\vec{v}_i(t + \frac{\Delta t}{2}) = \vec{v}_i(t - \frac{\Delta t}{2}) + \vec{a}_i(t) \Delta t \quad (18)$$

La détermination de la vitesse permet d'obtenir la position de l'atome à l'aide de l'équation précédemment énoncée à l'instant. Le déplacement d'un atome pendant l'intervalle de temps  $\Delta t$  est donné par l'équation (19) :

$$\vec{r}_i(t+\Delta t) - \vec{r}_i(t) = \frac{d\vec{r}_i(t^\circ)}{dt} \Delta t = \vec{v}_i(t^\circ) \Delta t \quad (19)$$

La valeur de l'instant  $t^\circ$ , fournie par la méthode de Verlet (Verlet, 1967), est le milieu de  $\Delta t$  et non  $(t + \Delta t/2)$  car la vitesse n'est pas constante sur cet intervalle de temps. La répétition de cette procédure à des intervalles de temps discrets, en fonction de la vitesse, aboutit à l'identification de la trajectoire (voir figure 2.1).

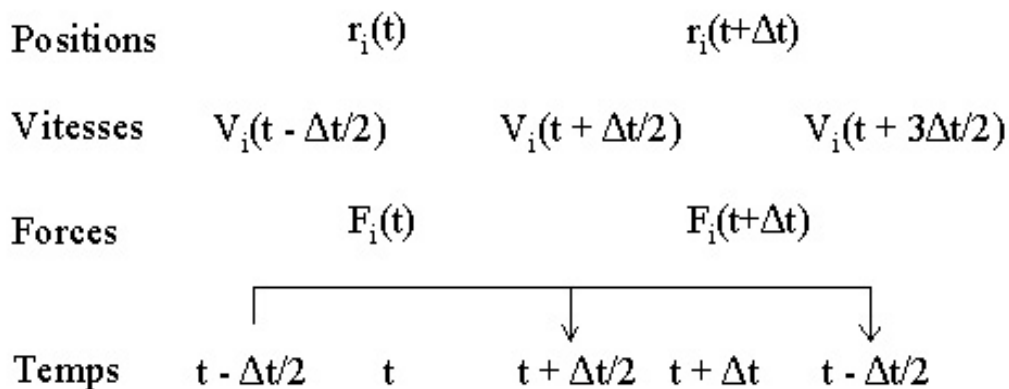


Figure 2.1 : résumé des différents éléments impliqués dans la détermination des trajectoires de dynamique moléculaire.

Notons que la cohérence des résultats obtenus par la méthode de dynamique moléculaire sera directement liée à deux éléments : la fiabilité des potentiels et la valeur donnée au pas d'intégration  $\Delta t$  : l'intervalle de temps choisi devra être plus petit que la période du mouvement de la plus haute fréquence (vibration d'un atome d'hydrogène le long de la liaison covalente, par exemple). Son ordre de grandeur est de l'ordre de 1 fs, ce qui permet de simuler correctement les mouvements des hydrogènes.

## 2.4 Méthode dite du recuit simulé

La méthode du recuit simulé (Kirkpatrick et al., 1983) tire son nom de l'analogie avec le procédé métallurgique du recuit. Si, par exemple, on fait fondre un métal et qu'on le laisse ensuite refroidir doucement, les atomes du liquide qui sont initialement agités de mouvements désordonnés vont ralentir graduellement et s'ordonner en un cristal régulier. Ce cristal est dans l'état d'énergie minimale du système, son minimum global. Si, en revanche, on refroidit rapidement le liquide, par trempage, le système va se retrouver prisonnier d'un minimum local qui n'est pas l'état le plus stable du système. On peut obtenir ainsi un état vitreux ou quelquefois un autre état cristallin d'énergie plus élevée. Une manière d'éviter ce problème consiste à réchauffer le système (le recuire) pour lui permettre de s'extraire du minimum local puis de le laisser refroidir suffisamment doucement pour que les atomes aient le temps de s'organiser en cristal. Il est important de noter que le système peut, si on lui en laisse l'opportunité, trouver de lui même le minimum global de l'énergie. Il suffit pour cela de s'assurer que le refroidissement soit suffisamment lent.

Les méthodes de minimisation correspondent, d'une certaine façon, à un trempage du système. Par analogie avec le recuit, il faut donc laisser au système la possibilité de quitter le minimum local où il se trouve en acceptant éventuellement des configurations où la valeur de la fonction est plus élevée. On peut subdiviser les algorithmes de recuit simulé en deux sous-groupes : ceux travaillant dans l'espace des coordonnées cartésiennes et ceux travaillant dans l'espace des angles dièdres.

### 2.4.1 Recuit simulé dans l'espace cartésien

Initialement, la méthode de recuit simulé a été introduite dans le domaine de la RMN comme raffinement de structures générées par des programmes de géométrie de distance (Nilges et al., 1988b). La différence avec la dynamique à haute température sous contraintes provient de la forme du potentiel : les forces dérivant des potentiels et les forces non liantes (électrostatiques, van der Waals et liaisons hydrogène) sont remplacées par une force unique de répulsion  $F_{\text{repel}}$ , dont l'expression est représentée dans l'équation

(20).

$$\left\{ \begin{array}{ll} F_{\text{REPEL}} = K_{\text{VDW}}(r_{\text{min}}^2 - r^2)^2 & \text{si } r \geq r_{\text{min}} \\ F_{\text{REPEL}} = 0 & \text{si } r < r_{\text{min}} \end{array} \right. \quad (20)$$

$K_{\text{VDW}}$  est une constante de force variable au cours du cycle,  $r_{\text{min}}$  simule le rayon de van der Waals des atomes. Souvent cette valeur est égale à 80% du rayon de van der Waals pour favoriser la formation des liaisons. Pendant la première phase, la constante des contraintes NOE est augmentée 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 pénalité des violations de NOE a atteint son maximum.

La méthode de recuit simulé a été affinée dans un second temps pour permettre l'affranchissement complet par rapport aux structures de départ : une matrice aléatoire de coordonnées peut servir de structure de base (Nilges et al., 1988a). On peut considérer l'état de refroidissement comme un minimiseur efficace car elle correspond à la conversion de l'énergie potentielle en excès en énergie cinétique, qui est éliminée à son tour par couplage à un "bain" thermique externe.

Les vitesses initiales des atomes de la molécule sont générées aléatoirement, ce qui permet une exploration de l'espace conformationnel : on peut utiliser un seul fichier de coordonnées atomiques de départ et on fait varier, pour chaque structure, la première valeur du générateur de nombre aléatoire ("seed") permettant le calcul des vitesses initiales différentes (Mujeeb et al., 1993; SantaLucia and Turner, 1993; Wimberly et al., 1993; Zhang and Patel, 1991). On générera autant de fichiers de structures que l'on a de valeurs différentes de vitesses de départ.

#### 2.4.2 Recuit simulé dans l'espace des angles dièdres

Formellement, il est 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  $E_{\text{pot}}$  en fonction des variables et des vitesses généralisées, on forme le



Lagrangien du système,  $L=E_{\text{cin}}-E_{\text{pot}}$ , et on applique les équations de Lagrange (21) :

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right)=\frac{\partial L}{\partial q_i} \quad (21)$$

On obtient ainsi un jeu de N équations du mouvement (N correspond aux nombres d'angles dièdres), ces équations doivent être intégrées numériquement comme dans le cas d'une dynamique cartésienne. Dans le cas du formalisme de Lagrange on obtient pour les accélérations généralisées un système de N équations reliant toutes les accélérations entre elles qu'il faut à chaque pas résoudre, ce qui rend les calculs très lourds. Une méthode de calcul très intéressante (Stein et al., 1997) évite ce problème et permet de calculer de proche en proche chaque accélération par rapport à l'accélération de l'angle dièdre voisin. Cette méthode 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 de la molécule (voir la section précédente). 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. Dans une représentation utilisant les angles dièdres, tous ces mouvements rapides (élongation des liaisons, distorsions des angles dièdres) sont supprimés et dès lors il ne reste que des mouvements faisant intervenir des rotations d'atomes ou de groupe d'atomes autour des liaisons covalentes, ce qui permet d'augmenter les temps de dynamique de façon notable.

Le recuit simulé dans l'espace des angles dièdres apparaît comme la méthode la plus efficace pour obtenir une structure RMN. L'équipe de Brünger (Stein et al., 1997) a comparé les trois différentes méthodes d'obtention de structures (recuit simulé à partir de géométrie de distance, recuit simulé simple et recuit simulé dans l'espace des angles dièdres) avec cinq macromolécules biologiques. Dans le cas des protéines et des peptides, le résultat est indéniable mais il est encore plus impressionnant dans le cas d'un ADN de 12 paires de bases : après un cycle de calcul, le recuit simulé dans l'espace des angles dièdres donne 52% de convergence tandis que les deux autres méthodes ne donnent aucun résultat.

# Première partie

## *Etudes structurales des différentes conformations de SL1 VIH-1<sub>Lai</sub>*

### III. Problématique et objectifs

Dans l'introduction, nous avons remarqué que la tige-boucle SL1 est impliquée dans l'initiation à la dimérisation de l'ARN génomique de VIH (Clever, et al., 1996; Laughrea and Jette, 1994; Muriaux, et al., 1996b; Muriaux, et al., 1995). La large boucle de SL1 (comportant neuf nucléotides) comprend une séquence auto-complémentaire GCGCGC chez la souche VIH-1<sub>Lai</sub>. C'est cette portion de séquence qui supporte la propriété d'initiation de la dimérisation. Les travaux des équipes de M. Laughrea (Laughrea and Jette, 1996a; Laughrea and Jette, 1996b) et de M. Paoletti (Muriaux, et al., 1996b) ont montré que de longs transcrits d'ARN, contenant SL1, formaient deux types de dimères : un dimère stable et un instable.

L'utilisation de sondes chimiques et enzymatiques a permis de mettre en évidence les types d'interactions engagées dans l'une et l'autre forme et de définir leur structure secondaire. Le complexe instable est supposé faire intervenir une interaction boucle-boucle mettant en jeu des liaisons hydrogène. Chez le dimère stable, l'ensemble des nucléotides seraient en interaction inter-brin tout au long de la séquence.

Les travaux de Theilleux-Delalande (Theilleux-Delalande, 1998) rapportent des propriétés analogues pour une séquence plus courte de SL1 (cf figure I.1). Cette nouvelle séquence apparaît comme un modèle plus simple pour l'étude structurale et mécanistique du phénomène de dimérisation observé chez VIH-1. La taille réduite du fragment, 23 nucléotides, permet une étude par RMN. On retrouve pour cette séquence tronquée les mêmes caractéristiques de dimérisation que pour la séquence complète. Nous pouvons donc, par extension des précédents travaux, définir les structures secondaires impliquées dans le phénomène de dimérisation (cf figure I.1). Ces structures secondaires sont au nombre de trois :

3. La première est le monomère (cf figure I.1). Cette structure s'apparente à une épingle à cheveux et il nous faut définir son implication au phénomène de dimérisation en tant qu'espèce de pré-dimérisation.

4. La seconde structure (en haut à droite de la figure I.1) représente le dimère instable et il semble essentiel de connaître sa structure tridimensionnelle pour comprendre son comportement métastable.
5. Enfin, la dernière espèce représente le dimère stable. Son étude est fondamentale pour comprendre les différences de stabilité et la transition de conformation.

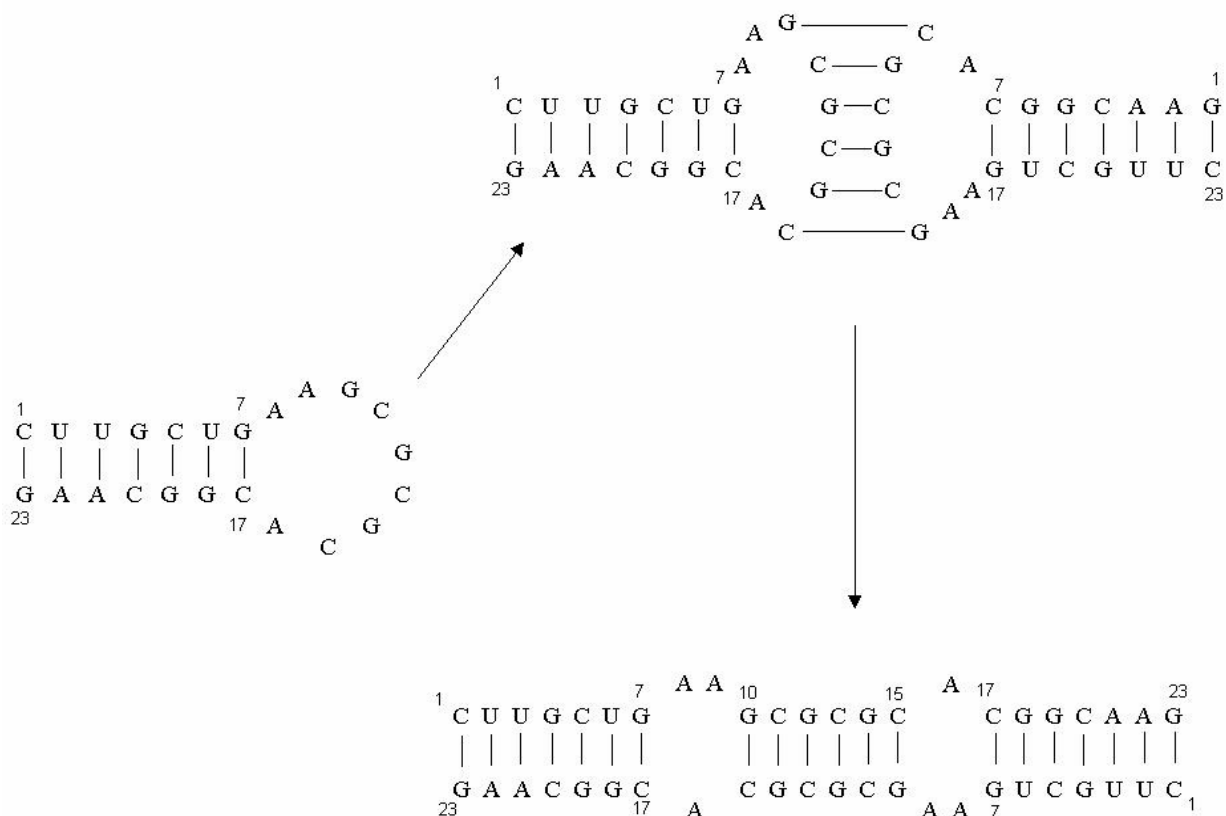


Figure I.1 : Représentation schématique des différentes structures secondaires rencontrées pour la séquence tronquée de SL1 VIH-1<sub>Lai</sub>.

Tout d'abord, il nous faut savoir si les structures secondaires représentées sur la figure I.1 existent réellement. Pour cela, il est nécessaire de séparer efficacement chaque espèce pour son étude structurale. Rappelons que le processus de dimérisation est lié à de nombreuses étapes du cycle rétroviral et que l'évolution d'un dimère instable vers un dimère stable semble correspondre à la maturation de l'ARN de HIV-1<sub>Lai</sub> (Muriaux, et al., 1996a). Comprendre toutes les étapes de ce phénomène biologique est essentiel pour la conception future de molécules inhibitrices de la dimérisation. Et c'est dans ce but que l'étude

structurale a été menée.

# Le monomère SL1 de VIH-1<sub>Lai</sub>

La tige boucle SL1 est la première espèce impliquée dans la dimérisation. Son étude structurale est essentielle car elle peut permettre la compréhension de l'interaction boucle-boucle rencontrée dans le kissing-complex. Le principal problème rencontré dans l'étude du monomère est qu'il dimérise spontanément en solution. Pour empêcher cette dimérisation, il a été choisi de modifier la séquence auto-complémentaire (figure II.1). Par conséquent, le résidu G<sub>12</sub> est remplacé par le résidu A<sub>12</sub>. Cette mutation permet de garder le caractère purique de la base G<sub>12</sub>. Afin d'éviter les confusions, on appellera SL1\* cette séquence mutée.

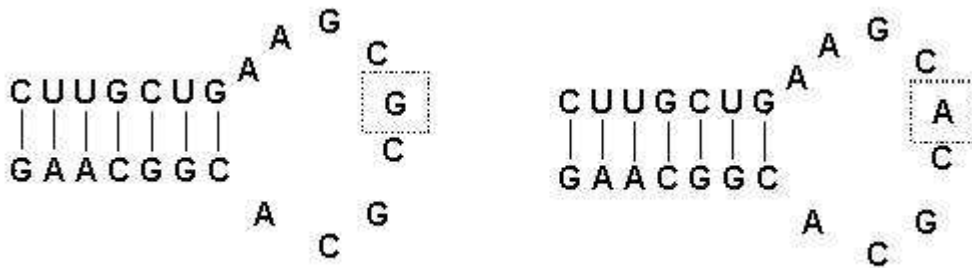


Figure II.1 : Séquence ARN des 23 nucléotides impliqués dans l'initiation à la dimérisation. A gauche : séquence sauvage, à droite le résidu G<sub>12</sub> (pointillé) est remplacé par une adénine.

## *1. Etudes physico-chimiques*

La molécule d'ARN a pu être synthétisée par voie chimique par l'équipe de T.Huynh Dinh à l'institut Pasteur de Paris. Elle a été purifiée par chromatographie en phase liquide et la concentration de l'échantillon a été déterminée par la mesure d'absorbance à 260nm. Nous avons pu obtenir un échantillon de 500  $\mu$ L de concentration finale de 2mM. L'effet de la mutation peut être observé par électrophorèse sur gel de polyacrylamide à 4°C (figure II.2). Les deux échantillons, SL1-23mer sauvage (à droite) et SL1-23mer muté A<sub>12</sub> (à

gauche), sont chauffés à 90°C pendant 3 minutes puis refroidis rapidement à 4°C pendant 10 minutes. Ils sont ensuite mis en présence d'un tampon de dimérisation (50mM Tris ; pH 7.5 ; 100mM NaCl ; [ARN]=10µM). Le gel d'électrophorèse (figure II.2) montre que la séquence existe principalement sous forme monomérique, bien qu'il existe une petite partie (estimée à 10%) organisée en dimère. Son étude structurale par RMN semble possible.

Les impuretés paramagnétiques sont éliminées par traitement sur colonne échangeuse de cations (chelex 100). Cette opération est indispensable car ces impuretés, en se fixant sur les phosphates, provoquent un élargissement des raies et l'extinction des effets NOE.

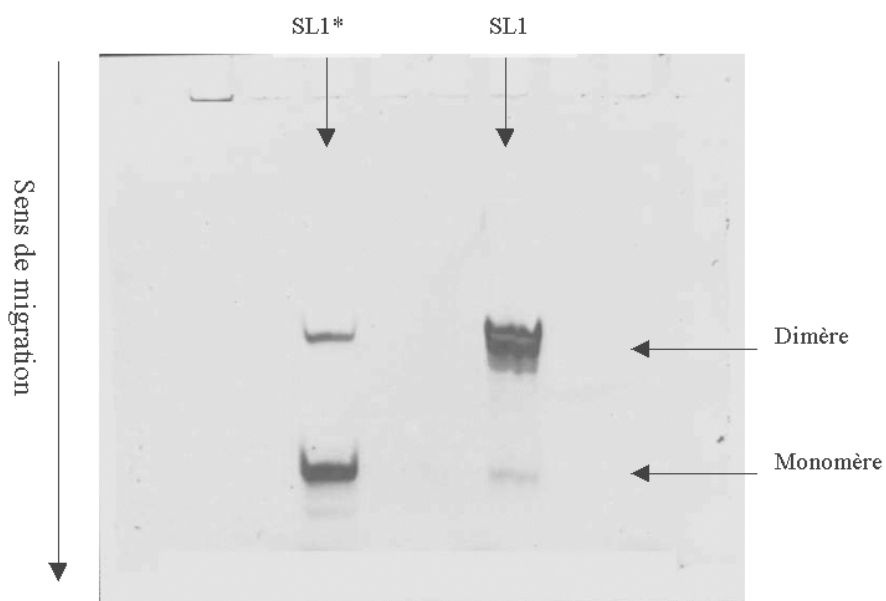


Figure II.2 : Gel d'électrophorèse à 4°C. SL1\* : séquence mutée A12, SL1 séquence sauvage (vue en négatif pour des raisons de clarté)

## 2. Etudes RMN

Pour l'étude RMN, l'échantillon a d'abord été préparé dans H<sub>2</sub>O avant d'être lyophilisé afin de le dissoudre dans D<sub>2</sub>O. Les spectres ont été enregistrés à différentes températures afin de lever des ambiguïtés dans les zones de recouvrement.

## 2.1 Attribution dans l'eau légère

La figure II.3 montre la région des protons imino. Seuls les protons situés dans la tige peuvent être observés. En effet, dans cette zone organisée en hélice, l'accessibilité de l'eau est moindre et implique un échange lent entre les protons imino et les protons du solvant. On a donc bien un monomère. Le schéma 1.4 résume les connectivités imino-imino observées. Le point de départ utilisé pour l'attribution fait intervenir une particularité de la séquence : la proximité des adénines 21 et 22 implique un fort effet NOE des protons H2. Lorsque ces protons sont attribués, on repère les imino des uraciles 2 et 3 du fait de leurs proximités avec les H2 des adénines.

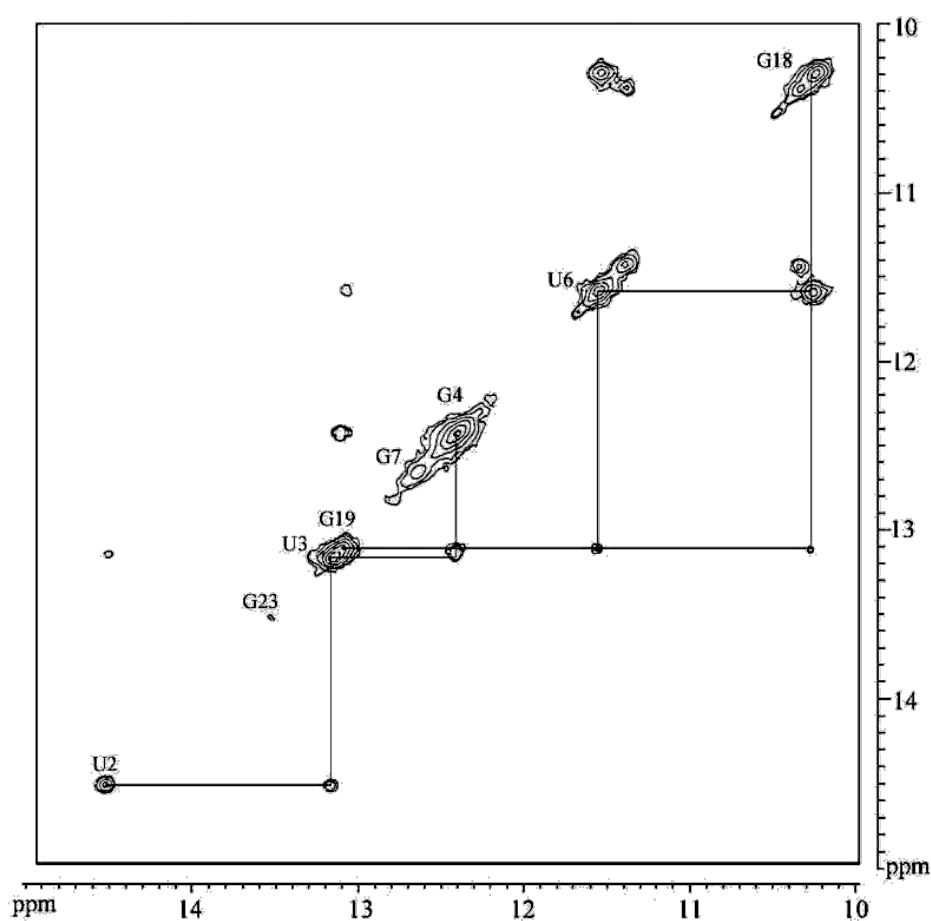


Figure II.3 : Spectre NOESY de la tige boucle SL1\* dans H<sub>2</sub>O/D<sub>2</sub>O (90|10) (pH=7, T<sub>m</sub>=150ms T=282K), agrandissement de la région des protons imino.



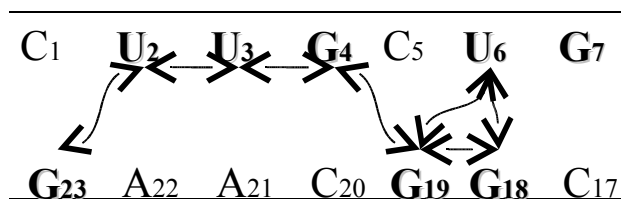


Figure II.4 : Schéma représentant les couplages dipolaires observés entre les protons imino dans la tige. Les résidus contenant un proton imino sont indiqués en gras.

On ne repère pas de couplage dipolaire du proton imino de la base G<sub>7</sub> avec les autres protons imino. Au premier abord, on peut donc supposer que l'appariement G<sub>7</sub>C<sub>17</sub> n'existe pas. Mais il existe un couplage dipolaire entre le proton imino de G<sub>7</sub> et les protons amino et H5 de C<sub>17</sub>. Ce couplage montre que cette paire de bases est bien présente dans la structure bien que l'on n'observe pas d'effet NOE dans la région imino pour G<sub>7</sub>, cela montre que cette paire de bases est plus exposée au solvant que les autres. Nous n'observons aucune corrélation mettant en évidence des appariements intra-boucle bien qu'il soit fréquent de rencontrer ce phénomène pour des boucles de cette taille.

### 6.1 Attribution dans l'eau lourde

L'étude des spectres NOESY dans D<sub>2</sub>O permet de faire l'attribution séquentielle complète du système de spin. La figure II.5 représente la région aromatique-H1'. Le point de départ consiste, comme précédemment, à repérer les H2 d'adénines. Le H2 du A<sub>22</sub> montre une forte intensité NOE avec le H1' en 3' (G<sub>23</sub>). L'étude du spectre TOCSY permet de localiser les protons H5 des uraciles ou des cytosines. Les ambiguïtés provenant de la région s'étalant de 7,6ppm à 7,8ppm et 5,6ppm et 5,7ppm ont pu être levées par différents spectres NOESY à diverses températures.

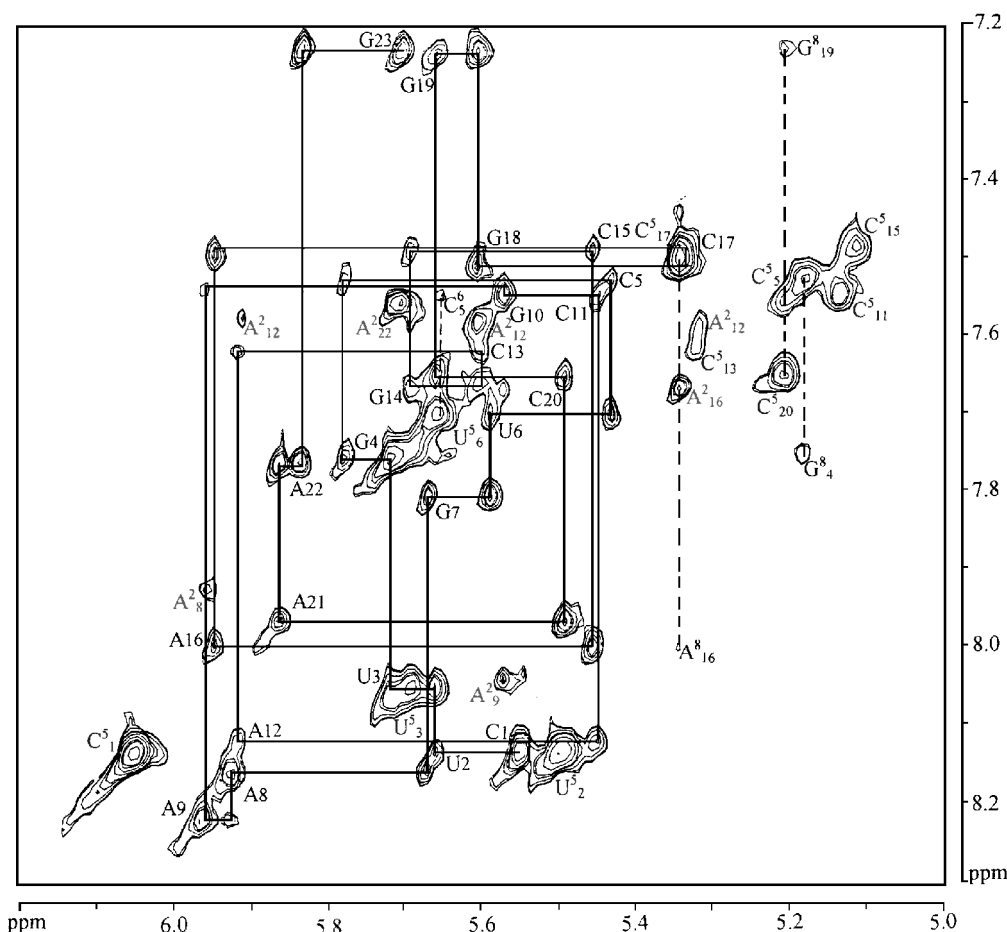


Figure II.5 : Agrandissement du spectre NOESY de la région aromatique/H1' dans D<sub>2</sub>O à 27°C et 450ms.

La mesure des intensités NOE, correspondant aux connectivités H5-H6 des résidus C<sub>5</sub> et C<sub>20</sub> à différents temps de mélange, permet d'établir une courbe de build-up dont la comparaison avec celle calculée par le programme RELAZ (Lancelot, et al., 1989), nous permet de déterminer le temps de corrélation de la molécule. Pour cette molécule le temps de corrélation est de 3,1ns à 27°C ; ce qui est cohérent avec un monomère de cette taille d'ARN. Le programme RELAZ prend en compte les intensités NOE à différents temps de mélange et nécessite un fichier de coordonnées s'approchant de la structure finale. Il calcule alors la matrice de relaxation complète pour les spins considérés, et donne leurs nouvelles intensités NOE corrigées du phénomène de diffusion de spins.

L'observation continue de corrélations NOE aromatique/H1' et aromatique/aromatique montre que la boucle n'est pas totalement déstructurée. Cependant, les connectivités H5

(C<sub>11</sub>, C<sub>13</sub>, C<sub>15</sub>) ou H8 (G<sub>10</sub>, A<sub>12</sub>, G<sub>14</sub>) et leurs protons H1' (figure II.5) sont faibles, voire inexistantes. Nous pouvons en déduire que cette région, correspondant à la boucle, présente une forte dynamique. Un certain nombre de connectivités intéressantes implique les protons H2 des adénines situées à la jonction de la tige et de la boucle (A<sub>8</sub>, A<sub>9</sub> et A<sub>16</sub>). La figure II.6 résume ces renseignements.

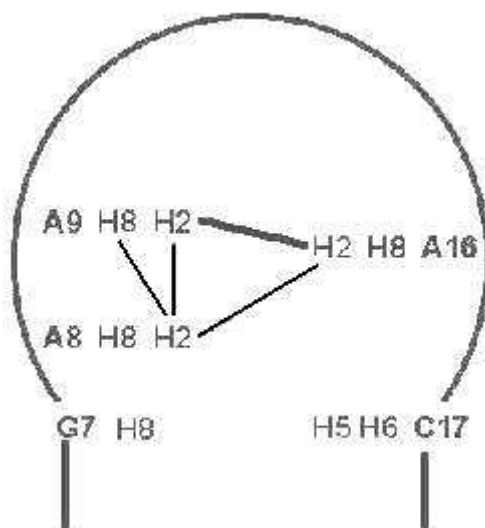


Figure II.6 : Schéma représentant les différentes connectivités impliquant les protons H2 des adénines de jonction. En trait plein : connectivité d'intensité moyenne. En trait fin : intensité faible.

### 3. Modélisation moléculaire

#### 3.1 Etablissement des contraintes

Le spectre TOCSY nous a permis de déterminer les conformations des différentes unités riboses de SL1\*. La constante de couplage scalaire entre les protons H1' et H2' est quasi nulle (0 à 1Hz) lorsque le sucre se trouve en conformation nord. Ainsi, l'absence de corrélation (ou une corrélation très faible) pour ce couple de protons révèle une conformation nord (ou C3'-endo). Pour notre échantillon, tous les résidus sont sous cette conformation, excepté A<sub>8</sub>, A<sub>9</sub> et G<sub>23</sub> dont le plissement du sucre est 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 de sucres.

Les contraintes imposées sur le squelette phosphodiester résultent de l'étude du spectre

phosphore à une dimension. Sur ce spectre, nous remarquons un étalement en déplacement chimique très faible pour l'ensemble des phosphates (de -1,4 à -2,0ppm). Ce faible étalement indique qu'il n'y a pas de phosphore adoptant une conformation très différentes des autres phosphates. Nous avons choisi de contraindre les angles du backbone ( $\alpha, \beta, \gamma, \epsilon$  et  $\zeta$ ) avec des valeurs classiques des hélices A. Cependant, pour rendre compte de la dynamique de la boucle, les contraintes du squelette phosphodiester impliquant la boucle et les bases de jonctions ont été imposées avec une incertitude de 45° (contre 15° dans la tige).

Les contraintes de distance issues des effets NOE ont été intégrées à plusieurs temps de mélange (0, 50, 100, 150, 300 et 450ms). L'utilisation du programme RELAZ (Lancelot, et al., 1989), qui calcule la matrice de relaxation, traduit les effets NOE en distances tout en tenant compte de la diffusion de spin. Les distances sont ensuite introduites avec une incertitude de 15%.

Les NOE observés pour les protons échangeables (imino-imino ou amino-imino) permettent de définir les appariements de types Watson-Crick. Il nous a semblé plus intéressant de ne pas quantifier ces derniers NOE et d'imposer l'appariement. Un total de 36 contraintes de distance rend compte des appariements de la tige. De plus, pour favoriser les appariements des bases de la tige, il nous a paru judicieux de rajouter 6 contraintes de plan pour les 6 premières paires de bases lors des premiers cycles de minimisation. Ces contraintes ont une constante de force faible de 25 kcal.mol<sup>-1</sup>.Å<sup>-2</sup> (pour comparaison, chaque contrainte NOE a une constante de force de 50 kcal.mol<sup>-1</sup>.Å<sup>-2</sup>).

Du fait de la forte dynamique de la boucle, de nombreuses structures respectant toutes les données RMN ont été obtenues. Néanmoins, une analyse du réseau de distances interprotons inférieures à 5Å, pour chacune des structures, indique différentes corrélations qui 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 II.1.

| <i>Contraintes RMN pour la structure SL1*</i> |  |                       |
|---|--|-----------------------|
| <u>Contraintes de distance</u>                |  |                       |
| Intra-résidu                                  | 50   |                       |
| Inter-résidu                                  | 163  | (dont 63 d'exclusion) |
| Watson-Crick                                  | 36   |                       |
| Total distances                               | 249  |                       |
| NOE/résidus                                   |  | 10,8                  |
| Constante de force                            | 50 kcal.mol <sup>-1</sup> .Å <sup>-2</sup>   |                       |
| <u>Contraintes d'angle dièdre</u>             |  |                       |
| Plissement du sucre                           | 46   |                       |
| Backbone                                      | 109  |                       |
| Total angles                                  | 155  |                       |
| Angles/résidu                                 | 6,7  |                       |
| Constante de force                            | 50 kcal.mol <sup>-1</sup> .rad <sup>-2</sup> |                       |
| <u>Total</u>                                  |  |                       |
| Nombre de contraintes                         | 404  |                       |
| Contraintes/résidu                            | 17,6   |                       |

Tableau II.1 : Résumé de l'ensemble des contraintes RMN utilisées dans la modélisation de la molécule SL1\*.

### 3.2 Résolution de la structure tridimensionnelle

La détermination de la structure finale a été effectuée en utilisant la méthode du recuit simulé dans l'espace cartésien à l'aide du logiciel XPLOR (version 3.851) (Brünger, 1992). Nous avons généré 200 structures aléatoires de départ, respectant la covalence des liaisons. Le protocole de recuit simulé 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 champ de force utilisé est simplifié en tenant compte uniquement des longueurs de liaisons, des angles de liaisons, des angles impropres, du potentiel répulsif de van der Waals, des contraintes de distance, des contraintes d'angle dièdre et des contraintes de plan. Puis, pendant 8ps, les masses et les tailles des atomes sont lentement amenées à leurs valeurs classiques. La phase de refroidissement (1500K à 300K) se fait suivant une dynamique de 20ps. Enfin, une dernière minimisation de 1000 pas par la méthode du gradient conjugué finit ce protocole. Les contraintes RMN sont introduites progressivement à chaque cycle de calcul, dans un premier temps les plus sûrs (Watson-Crick, inter-résidus), pour finir par les

contraintes d'exclusions. A ce niveau 37 structures ont été sélectionnées selon leurs faibles violations de contraintes RMN et leurs énergies globales.

Finalement, les structures résultantes sont affinées par 10000 pas de minimisation d'énergie selon la méthode de Powell en utilisant une fonction d'énergie potentielle complète incluant les paramètres attractifs de van der Waals, l'énergie électrostatique et les liaisons hydrogène.

Les 37 structures sont ensuite examinées graphiquement. Leur superposition montre une famille de tiges boucles avec une boucle de forte dynamique. La structure moyenne et minimisée de ces structures sert de point de départ à l'étude dynamique en présence de solvant et d'ions explicites.

Pour obtenir la structure moyenne, les structures sont raffinées par comparaison des déplacements chimiques calculés en retour à l'aide du logiciel NUCHEMICS (Wijmenga, et al., 1997). Le protocole sera explicité plus profondément dans le chapitre III-3.3 de la seconde partie. Un ensemble de 19 structures finales, respectant l'ensemble des données RMN et montrant des violations de déplacements chimiques inférieures à 0,3ppm, définit la structure de SL1\*. La superposition des 19 structures est représentée sur la figure II.7. Les valeurs énergétiques (tableau II.2) montrent que la structure tridimensionnelle est correctement définie.

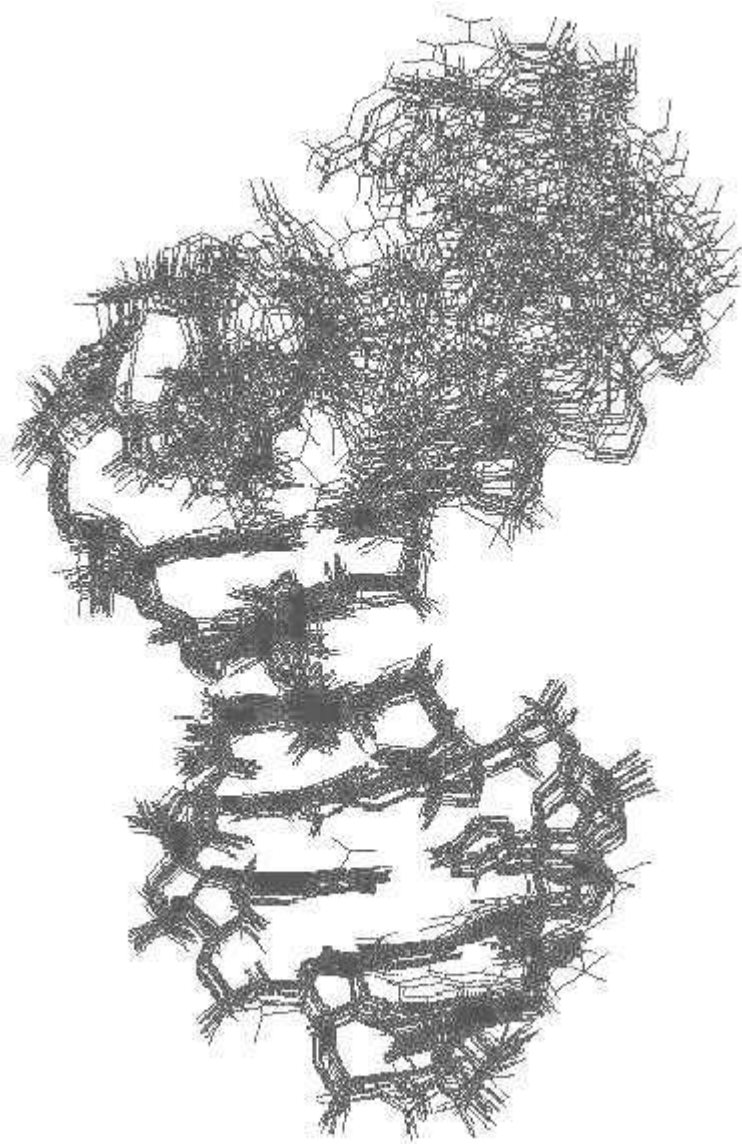


Figure II.7 : Superposition des 19 structures de SL1\*. Les structures sont superposées sur les atomes lourds de la tige (C1-G7 et C17-G23).

#### **4. Analyses structurales de SL1\***

Les paramètres d'hélice de la tige de SL1\* sont calculés à l'aide du logiciel CURVES 5.1 (Lavery and Sklenar, 1988; Lavery and Sklenar, 1989; Lavery and Sklenar, 1990; Ravishanker, et al., 1989) et comparés aux valeurs classiquement rencontrées dans les ARN (Saenger, 1984). Les paramètres hélicoïdaux sont définis en annexe. Comme nous le supposions, nous sommes en présence d'une hélice de type A classiquement rencontrée dans les ARN. Tous les appariements de type Watson-Crick sont présents dans la structure hélicoïdale, y compris pour la paire G<sub>7</sub>-C<sub>17</sub> dont les contraintes RMN sont

imposées plus “souplement” que les autres paires de bases.

La superposition correcte des 19 structures finales (figure II.7) nous permet de calculer une structure moyenne et de définir les RMSD (“Root Mean Square Deviation”) pour chaque région (tableau II.2). La mesure du RMSD par résidu permet de déterminer les zones comportant une plus forte dynamique. Le graphique (figure II.8) montre le RMSD pour chaque résidu.

| <u>Précision des structures</u>   |      |              |                        |
|---|------|--------------|------------------------|
| Nombre de structures  |      | 19           |                        |
| Nombre de violation NOE>0.3Å  |      | 0            |                        |
| Nombre de violation dièdre>5°   |      | 0            |                        |
| RMSD <sup>1</sup> global  |      | 1.57 ±0.75 Å |                        |
| RMSD tige (C <sub>1</sub> -G <sub>7</sub> ,C <sub>17</sub> -G <sub>23</sub> ) |      | 0.54 ±0.17 Å |                        |
| RMSD jonction (A <sub>8</sub> ,A <sub>9</sub> et A <sub>16</sub> )            |      | 1.01 ±0.30 Å |                        |
| <u>Energie</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> |

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

Tableau II.2 : Résumé des énergies et des RMSD pour SL1\*.

Nous sommes étonnés de constater que les résidus A<sub>8</sub>,A<sub>9</sub> et A<sub>16</sub> sont structurés. Cette région correspond à la jonction entre la boucle et l’hélice de la tige. L’adénine 8 se place perpendiculairement à l’axe de l’hélice tandis que l’adénine 9 s’empile entre l’adénine 8 et la guanine 10. Cet empilement (figure II.9) est stable et induit une torsion du squelette. Cette torsion et cet empilement impliquent que les bases 8,9 et 16 sont situées à l’intérieur de la boucle et, de ce fait, empêchent les résidus de la séquence palindromique GCGCGC de s’auto-apparier. Ces résidus se retrouvent expulsés vers l’extérieur de la boucle et exposés au solvant, permettant ainsi la dimérisation du monomère.



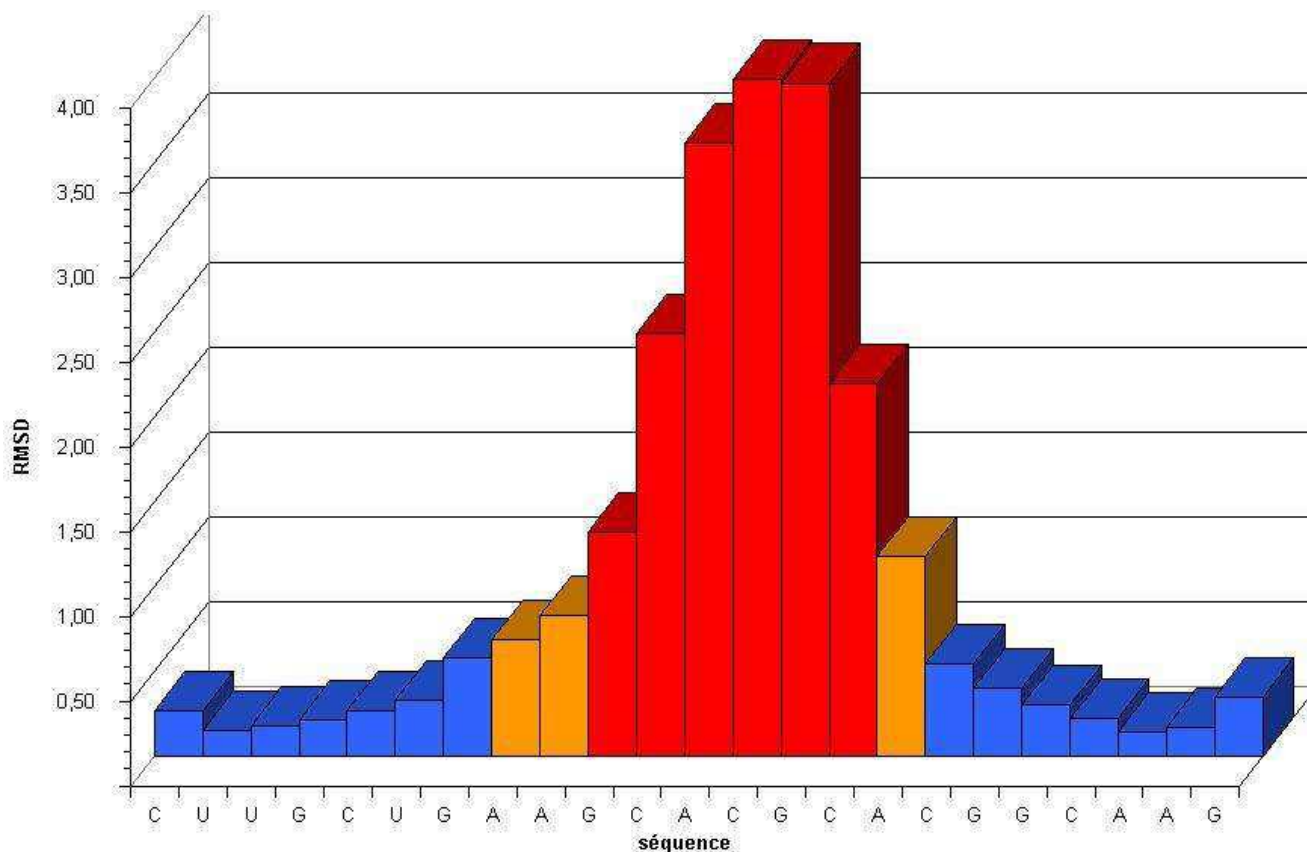


Figure II.8 : RMSD par résidu, calculé sur les 19 structures à l'aide du logiciel MOLMOL (version 2K.1) la superposition est faite sur les résidus de la tige. Les résidus de l'hélice sont représentés en bleu, ceux des adénines de jonction en orange foncé et enfin les résidus de la boucle en rouge.

Les cytosines centrales ( $C_{11}$ ,  $C_{13}$  et  $C_{15}$ ) de nos structures se trouvent empilées avec la purine adjacente. Ce "stacking" est confirmé par le déplacement chimique de ces cytosines. A titre d'exemple, les protons H5 de ces cytosines se trouvent dans une fenêtre de déplacements chimiques de 5,1 à 5,3ppm. Le déplacement chimique du proton H5 d'une cytosine seule en solution est de 6,13ppm. Cette différence (de 0,7 à 0,9ppm) est classiquement due à l'empilement des bases. Lors du calcul en retour du déplacement chimique, les protons aromatiques H6 des trois cytosines étaient totalement corrélés. Cependant, aucune de nos structures parmi les 37 ne montre de corrélation correcte avec les déplacements chimiques des protons H5 des cytosines centrales. Cette information suggère qu'en solution et à température ambiante, l'oligonucléotide SL1\* comporte une boucle (de  $G_{10}$  à  $C_{15}$ ) avec une grande variabilité de conformations et donc une forte dynamique.

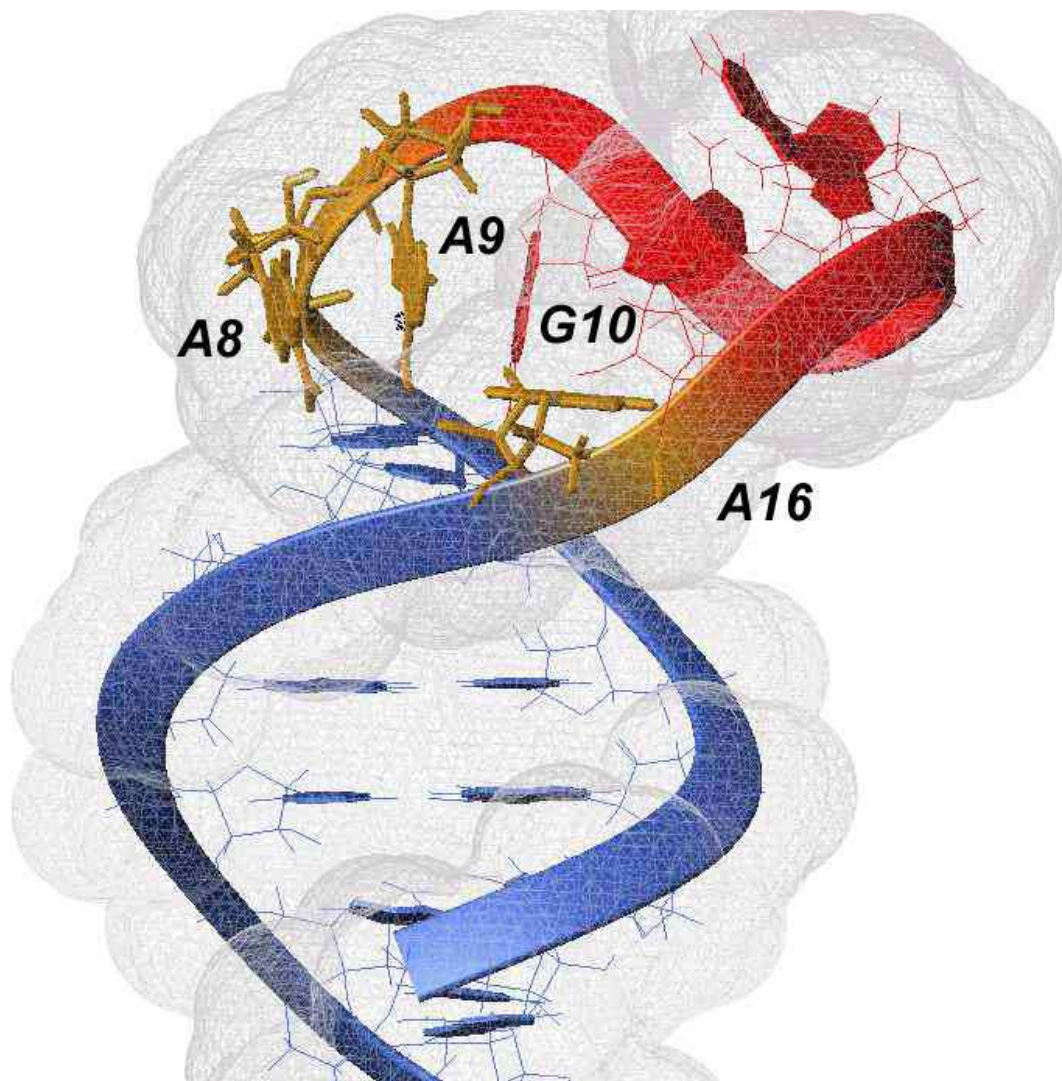


Figure II.9 : Représentation du motif stabilisant des adénines de jonction. L'adénine 9 est empilée entre l'adénine 8 et la guanine 10, induisant une forte torsion du backbone. Les différentes régions sont indiquées par un code couleur : la tige est en bleu, les bases de jonction en orange et la boucle en rouge. La surface d'accessibilité au solvant permet de rendre compte du volume de chaque motif.

Malheureusement, une structure RMN ne représente qu'une image moyenne en solution. Pour apprécier le comportement dynamique de la boucle, il faut donc compléter cette étude par une simulation de dynamique moléculaire avec solvant et ions explicites.

### 5. Simulation de dynamique moléculaire

L'équipe du Dr Genest, avec laquelle nous collaborons étroitement, a effectué la simulation dynamique de notre séquence SL1\* dont nous avons fourni la structure

moyenne. Cette simulation s'est 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 reproduire le comportement du solvant. Après une période d'équilibration thermique à 300K, une dynamique de 10ns est démarrée avec un stockage des coordonnées toutes les 0,5ps pour l'oligonucléotide et 3ps pour le solvant.

L'analyse des trajectoires dynamiques est réalisée sur l'ensemble des 10ns, excepté les 200 premières picosecondes qui sont écartées. L'analyse consiste en un calcul du RMSD des atomes par rapport à la structure de départ et à l'analyse des liaisons hydrogène. L'analyse au cours du temps de simulation des RMSD établit des cartes de rigidité et définit les régions comportant le plus de mouvements. Bien que 10ns de dynamique représente un temps de calcul important, nous ne pouvons définir une structure à l'équilibre. Cependant, nous pouvons affirmer, comme nous l'avions supposé, que la tige semble stable et comporte une amplitude de mouvements classiques pour une hélice d'ARN. La boucle est, quant à elle, animée de nombreux mouvements. Néanmoins, elle est toujours orientée selon la même face de la tige. Au cours des 10ns de la simulation, la paire de bases  $G_7-C_{17}$  est toujours présente bien que comportant de légères fluctuations puisqu'elle est la base de jonction de la boucle.

Lors de cette étude, une liaison hydrogène est définie par une distance entre les atomes donneurs et accepteurs de moins de  $3,5\text{\AA}$  et un angle (donneur, hydrogène, accepteur) supérieur à  $45^\circ$ . Les  $\text{O}2'$  forment des liaisons hydrogène intramoléculaires essentiellement avec les  $\text{O}4'$  et  $\text{O}5'$  du résidu  $n+1$ . Les  $\text{O}2'$  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  $2'\text{OH}$  pourrait alors expliquer la raison pour laquelle les tiges boucles sont plus fréquentes chez les ARN que chez les ADN. Nous pouvons ajouter qu'une deuxième simulation de 6ns a été réalisée avec la même structure de départ mais avec la base  $A_{12}$  mutée en  $G_{12}$ . Cette simulation de la séquence sauvage comporte les mêmes caractéristiques que la séquence mutée, preuve que le comportement n'est pas affecté par cette petite mutation.

## 6. Conclusions

L'étude du monomère de SL1 est le premier pas permettant la compréhension complète du mécanisme de dimérisation du rétrovirus VIH-1<sub>Lai</sub>. L'étude d'une séquence mutée permet d'éviter le phénomène de dimérisation spontanée tout en gardant une bonne homogénéité de séquence. L'étude RMN montre une structure comportant trois principales régions :

- Une tige composée de 7 paires de bases organisée en hélice A ; ce qui est classiquement observé dans les ARN.
- Une région de jonction impliquant les trois adénines 8,9 et 16. Ces résidus se trouvent empilés ; ce qui est favorable thermodynamiquement et provoque une exclusion des autres résidus de la boucle vers le solvant.
- Une dernière région, constituée des résidus GCACGC centraux (résidus de 10 à 15). Cette boucle montre une forte dynamique et est largement exposée au solvant. Aucun appariement de cette région est observé sur nos structures RMN.

L'accomplissement d'une simulation de dynamique moléculaire de 10ns, avec solvant et ions, renseigne sur les mouvements et la stabilité de ces motifs. Dans cette étude, les données RMN et la simulation convergent et montrent que le comportement dynamique du monomère correspond à ce qui a été cité précédemment. De plus, cette simulation a permis de mettre en évidence le rôle stabilisant du groupement 2'OH dans l'organisation structurale. Dans la structure du monomère SL1 de VIH-1<sub>Lai</sub>, les bases de la boucle G<sub>10</sub>CGCGC<sub>15</sub> (pour l'oligonucléotide sauvage) s'orientent vers l'extérieur et se trouvent en position favorable pour générer des liaisons Watson-Crick avec un autre monomère. Ceci explique sans doute pourquoi cet oligonucléotide dimérise spontanément, en solution, sous forme de kissing-complex et possède donc une prédisposition structurale à la dimérisation.

## IV. Les dimères de SL1-VIH<sub>Lai</sub>

Pour cette étude structurale, il est essentiel de connaître le comportement de notre séquence. La séquence SL1-23mer sauvage (non mutée en position 12, cf figure III.1 ) doit être isolable à la fois sous forme de dimère instable et de dimère stable, et cela dans la gamme de concentration requise pour la RMN. De ce fait, nous allons tout d'abord nous concentrer sur les études physicochimiques de la séquence SL1 puis nous aborderons l'étude structurale.

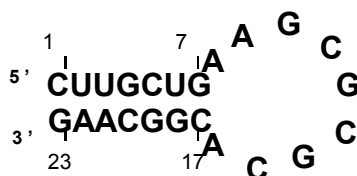


Figure III.1 : Séquence SL1-23mer étudiée, présentée ici sous la forme du monomère.

La taille de l'ARN est une des limites liées à l'utilisation de la RMN. Il est donc capital de s'assurer du comportement structural et/ou dynamique d'un système plus court (23 nucléotides dans notre cas, soit une masse molaire de 14879g.mol<sup>-1</sup> sous forme dimérique).

### *1. Etudes physicochimiques de la dimérisation*

#### **1.1 Mise en évidence des deux types de dimères**

La différenciation des deux types de dimères a été réalisée par électrophorèse sur gel de polyacrylamide (Theilleux-Delalande, 1998). Les conditions de dimérisation sont celles définies par Muriaux et al (Muriaux, et al., 1996b), et les résultats sont présentés sur la figure III.2.

La séquence d'ARN SL1-23mer est portée à 90°C pendant trois minutes. A cette température tous les appariements par liaisons hydrogène disparaissent et l'échantillon est dit "dénaturé". Puis le milieu est refroidi le plus rapidement possible à 4°C. L'échantillon est alors incubé dans un tampon de dimérisation (50mM Tris ; pH 7,5 ; 100mM NaCl ; [ARN]=3,4μM) pendant 15 minutes pour deux températures différentes : 4°C pour le puits numéro 2 et 40°C pour le puits numéro 3. Un autre puits (numéro1) comportant l'échantillon d'ARN sans tampon de dimérisation sert de référence.

La migration sur gel est réalisée à 4°C ou 20°C (respectivement portion A et B de la figure III.2).

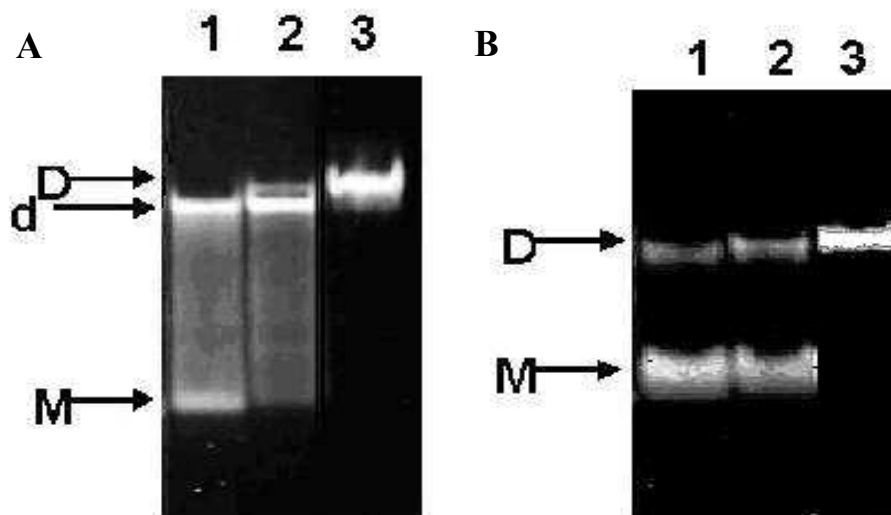


Figure III.2 : Stabilité des dimères de SL1-23mer formés à 4°C ou 40°C. D : dimère stable ; d : dimère instable ; M : monomère. Puits 1 : ARN en absence de tampon d'incubation. Puits 2 : (ARN + tampon d'incubation) à 4°C. Et puits 3 : (ARN + tampon d'incubation) à 40°C. Le gel électrophorétique est réalisé à puissance constante ; à 4°C pour le gel A et à 20°C pour le gel B.

Sur chaque gel, la ligne numéro 1 montre que l'ARN n'est pas totalement structuré en absence de tampon de dimérisation. Ce dernier sert de référence car il nous permet de localiser le monomère de SL1.

Sur le gel électrophorétique réalisé à 4°C (A), les échantillons incubés à 4°C et 40°C apparaissent majoritairement sous forme de dimère. Notons cependant que la ligne 2 (incubée à 4°C) donne une bande légèrement plus basse (d) que dans le cas de l'incubation

à 40°C (ligne 3). De plus cette ligne (2) montre une traînée, ce qui nous indique un relatif manque de stabilité de cet échantillon.

Lors de la migration à 20°C (B), le dimère du puits numéro 2 se dissocie et migre majoritairement sous forme d'un monomère. Quant au dimère formé à 40°C (ligne 3), il demeure stable durant cette migration.

Ces expériences démontrent sans ambiguïté, qu'en fonction de la température d'incubation utilisée, nous génèrons deux types de dimère : un stable et un moins stable. Ces propriétés définies pour notre séquence de 23 nucléotides sont analogues à celles démontrées auparavant pour de longs fragments d'ARN de SL1 (Laughrea and Jette, 1994; Muriaux, et al., 1996a; Muriaux, et al., 1996b). Le modèle de 23 nucléotides contient donc l'ensemble des informations structurales nécessaires à l'étude du phénomène de dimérisation. Notons cependant, qu'en raison du nombre de paires de bases plus petits dans notre cas, les températures d'incubation nécessaires à la dénaturation des dimères sont diminuées.

## 1.2 Préparation des dimères

La validation du système étant réalisée, il faut maintenant pouvoir isoler chaque forme dimérique dans des quantités observables par RMN. L'étude d'un mélange est très délicat. La première étape de l'étude structurale consiste donc à trouver les protocoles de préparation de l'échantillon suivant le type de dimère envisagé.

La dimérisation du brin d'ARN SL1-23mer est réalisée en utilisant les principes de préparation décrits précédemment par Muriaux et al (Muriaux, et al., 1996b). L'ARN est tout d'abord "dénaturé" à 90°C pendant 3 minutes en absence de sels. A cette température aucun appariement n'est envisageable.

Si l'on refroidit lentement l'échantillon, de 90°C à 4°C pendant 3 à 4 heures, le système se trouve à tout moment dans un état de réversibilité thermodynamique. Les interactions inter ou intra moléculaire peuvent alors s'établir et s'échanger. Cette condition adiabatique permet l'obtention du dimère thermodynamiquement le plus stable, par extension : le dimère stable.

Au contraire, si l'on refroidit rapidement l'échantillon, de 90°C à 4°C en quelques minutes, le système n'est pas dans un état de réversibilité. Les conformations les plus rapides à se former se rencontrent et aucun équilibre conformationnel n'est possible. Cette condition favorise la formation du dimère le plus rapide à se former, c'est à dire le dimère cinétique qui n'est pas forcément le plus stable.

Pour réaliser cette expérience, plusieurs conditions doivent être contrôlées. La salinité est un élément important. En effet, les charges positives font écrans aux groupements phosphates ou forment des ponts salins entre les charges éloignées. Ces ions permettent alors la stabilisation des structures, ce qui induit une augmentation de leur température de dénaturation et perturbe le protocole de dimérisation. Rajoutons, de plus, que l'utilisation de sels de magnésium, qui sont classiquement réputés pour stabiliser les ARN, détruisent ces derniers à haute température.

Un autre facteur intervient de façon prépondérante : la concentration en ARN. En effet, l'utilisation d'un échantillon concentré en ARN (>100µM) va favoriser les interactions intermoléculaires qui sont prépondérantes dans le dimère stable. A contrario, travailler à basse concentration va favoriser les interactions intramoléculaires. Le dimère instable est constitué de deux sous-unités tige-boucle. L'isolement de ces brins d'ARN les uns des autres va favoriser la formation de ces tiges-boucles. La figure III.5 résume les différents types de formation des dimères et les différents mécanismes mis en jeu.



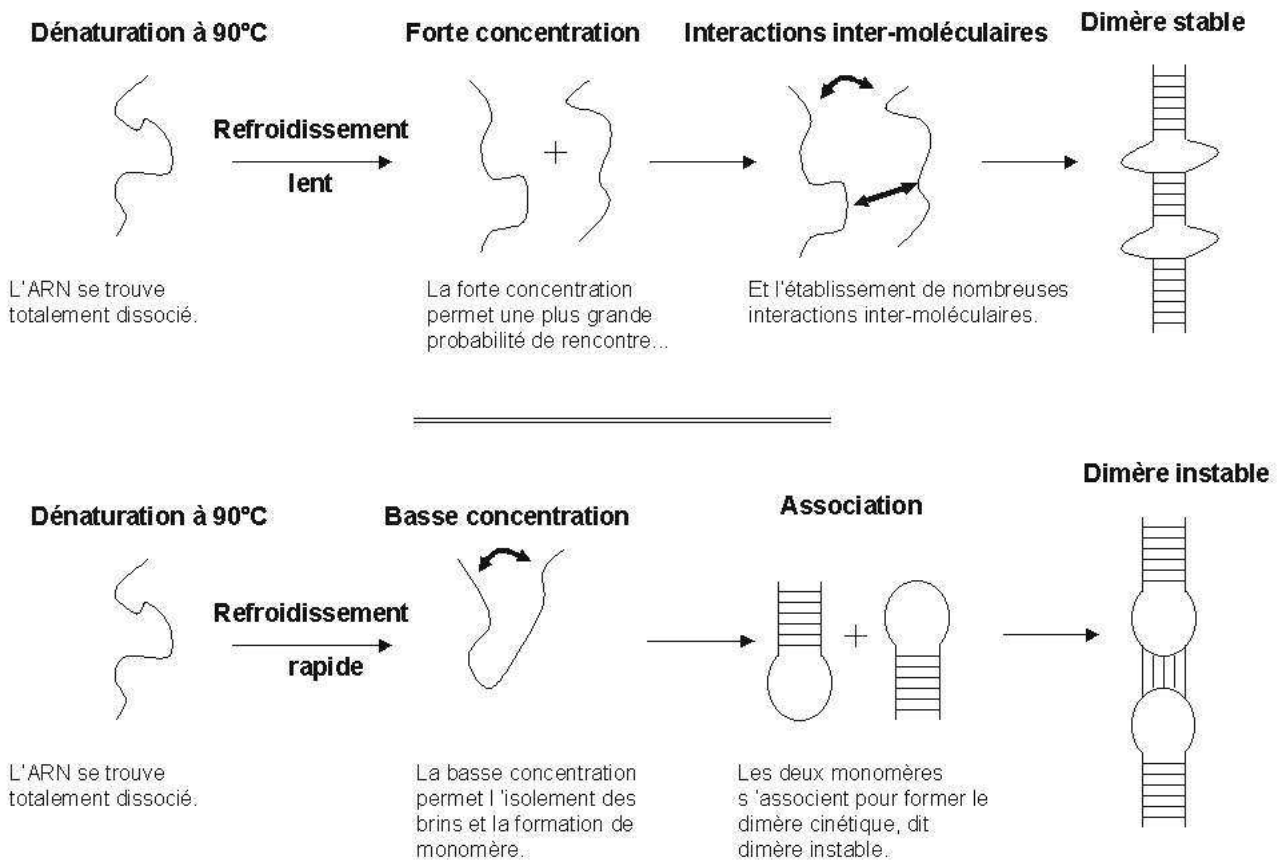


Figure III.5 : Représentation schématique des deux méthodes de préparation des dimères. En haut, obtention du dimère thermodynamique ou dimère stable. En bas, obtention du dimère cinétique appelé dimère instable.

Une analogie intéressante peut être faite entre le principe de préparation des dimères et la méthode du recuit simulé (voir l'introduction sur la modélisation moléculaire). En effet, l'obtention du dimère instable correspondrait au "trempage" du système, et l'accession au dimère stable à un deuxième recuit à partir de la conformation du premier dimère.

Après lyophilisation, l'ARN est repris dans l'eau à la concentration désirée pour l'étude RMN (2 mM dans notre cas). Il est toutefois nécessaire d'ajouter des sels (NaCl et Tampon phosphate) pour stabiliser la forme dimérique désirée. Cela induit l'affinement des raies de résonance sur un spectre RMN et autorise une étude sur une plage plus large de température. De plus, nous cherchons à recréer des conditions les plus proches possible du pH et de la salinité physiologique.

## 2. Etude RMN du dimère stable SL1

### 2.1 Préparation de l'échantillon

L'échantillon d'ARN, obtenu par synthèse chimique, a tout d'abord été élué sur une résine échangeuse d'ions (chelex 100 Bio Rad) afin d'éliminer d'éventuelles impuretés paramagnétiques susceptibles de détériorer la qualité des spectres RMN. Il a ensuite été dialysé contre un tampon phosphate (10mM, pH=6) en présence d'EDTA (2mM) puis contre l'EDTA seul et enfin contre de l'eau pure. L'échantillon est concentré par dialyse, repris dans 500 $\mu$ L d'un mélange 90% H<sub>2</sub>O et 10%D<sub>2</sub>O et dégazé à l'argon. L'ARN est par la suite hybridé sous forme de dimère stable : il est chauffé à 90°C pendant 3 minutes puis refroidi lentement jusqu'à température ambiante. Nous fixons la force ionique ([NaCl] =100mM) et le pH avec un tampon phosphate (10mM, pH=5.7).

### 2.2 Attribution dans l'eau légère

Les spectres dans l'eau ont été obtenus en utilisant une séquence d'impulsion de type "Jump & return" afin de supprimer le signal de l'eau (Plateau and Guéron, 1982). Ils permettent l'étude des protons échangeables imino et amino. Les corrections de ligne de base sont réalisées selon la méthode proposée par Marion (Marion, et al., 1989). La figure III.6 représente un zoom de la région des connectivités imino-imino.

Si nous comparons ce spectre NOESY avec celui obtenu dans les mêmes conditions pour le monomère de SL1\*, nous observons de grandes similarités. Ces similarités correspondent à l'organisation de la tige (C<sub>1</sub>-G<sub>7</sub>, C<sub>17</sub>-G<sub>23</sub>). Nous remarquons cependant, des pics supplémentaires attribués au protons iminos des bases G<sub>10</sub>, G<sub>12</sub> et G<sub>14</sub>. Ces corrélations supplémentaires prouvent que nous sommes en présence d'un dimère.

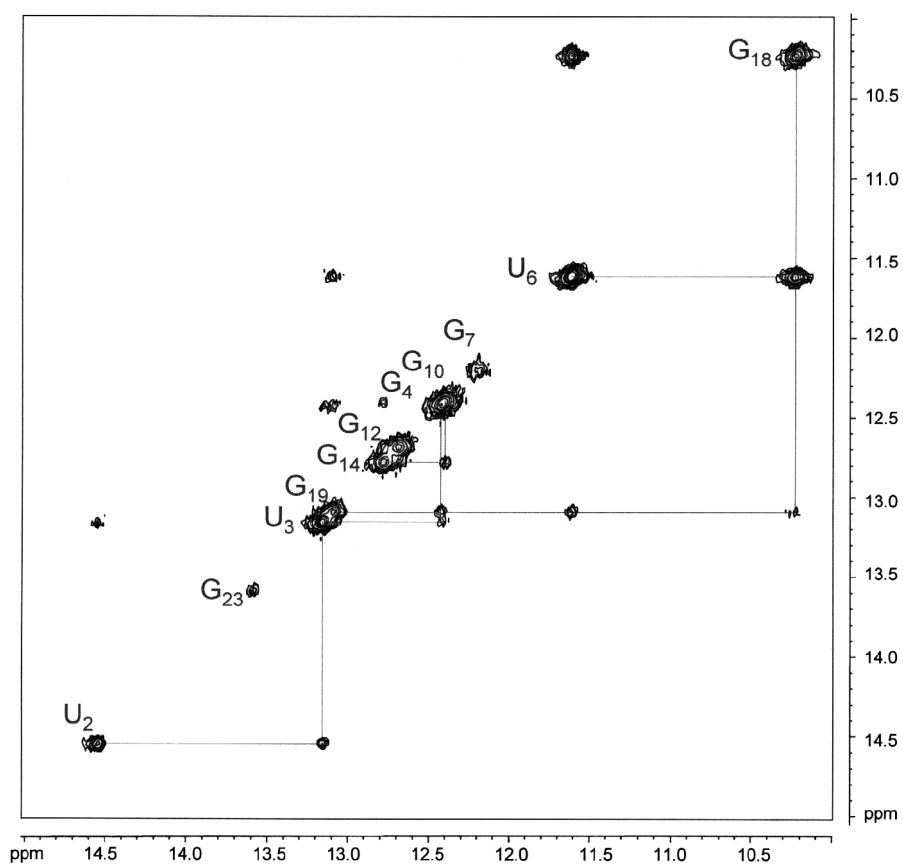


Figure III.6 : Spectre NOESY de la région imino-imino du dimère stable de VIH-1<sub>Lai</sub> (150ms, 5°C).

Les connectivités entre les protons iminos des uraciles et les protons H2 des adénines appartenant aux paires AU de la séquence constituent, là encore, le point de départ de l'attribution séquentielle. Des mesures du temps de relaxation longitudinale sont utilisées pour repérer les protons H2 des adénines. La figure III.7 résume les connectivités NOE entre protons iminos présentes pour notre échantillon. Notons que, comme dans le cas du monomère, nous n'observons pas de corrélations dipolaires entre protons iminos de la base G<sub>7</sub>.

Chaque proton imino, précédemment repéré, établit une corrélation NOE avec les protons aminos des résidus qui leur font face. De cette manière, nous définissons les appariements de type Watson-Crick pour l'ensemble des tiges. Tous les résidus de la tige sont engagés dans un appariement, y compris le résidu G<sub>7</sub>. Nous remarquons cependant,

un élargissement notable du pic imino de G<sub>7</sub> qui semble indiquer un appariement plus lâche pour la paire G<sub>7</sub>C<sub>17</sub>. La tige centrale (G<sub>10</sub>-C<sub>15</sub>,G<sub>10'</sub>-C<sub>15'</sub>) est totalement appariée.

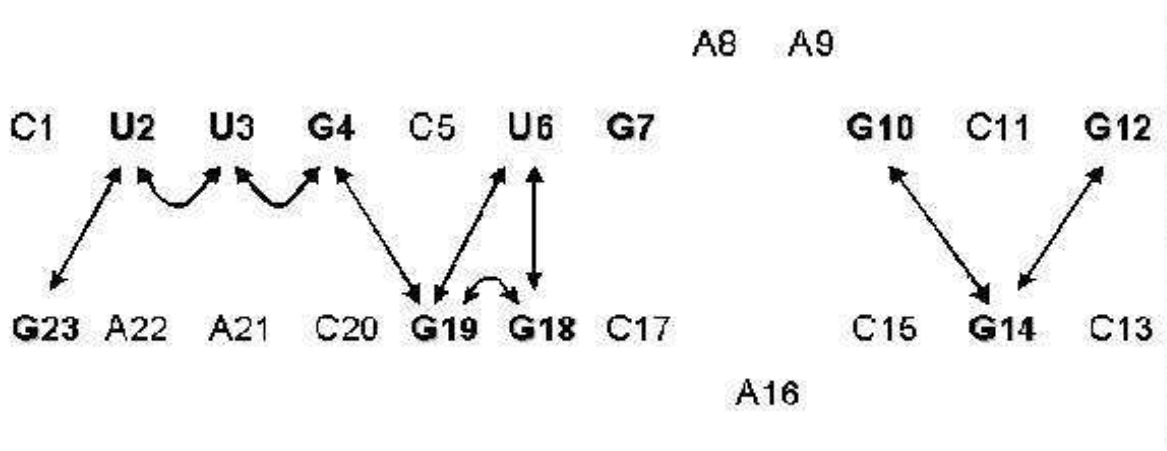


Figure III.7: Représentation schématique des connectivités entre protons iminos rencontrés lors de l'étude du dimère stable SL1 23-mer. La figure ne représente que les 12 premières bases, les autres étant symétriques. En gras sont indiquées les bases possédant un proton imino.

Nous n'observons aucun proton amino des adénines de jonction (8,9 et 16) sur les spectres NOESY (à 5°C ou 15°C). Ces protons sont donc échangés rapidement avec le solvant. Par conséquent, il semblerait qu'ils ne soient pas engagés dans des liaisons hydrogène.

### 2.3 Attribution dans l'eau lourde

L'échantillon d'ARN, utilisé pour l'étude précédente, est lyophilisé deux fois contre D<sub>2</sub>O avant d'être repris dans 500µL de D<sub>2</sub>O et scellé sous atmosphère d'argon. Les expériences NOESY sont réalisées à 25°C. La figure III.8 représente le spectre de la région aromatique/H1' qui a été utilisée pour l'attribution séquentielle.

La grande majorité des connectivités est bien résolue et ne présente que rarement des recouvrements. Les ambiguïtés restantes ont été levées par l'enregistrement de spectres à 20 et 30°C. Aucun dédoublement des résonances attribuées n'est observé, ce qui indique l'absence d'une dynamique d'échange lent et nous conforte dans l'idée que nous sommes

en présence du dimère thermodynamique.

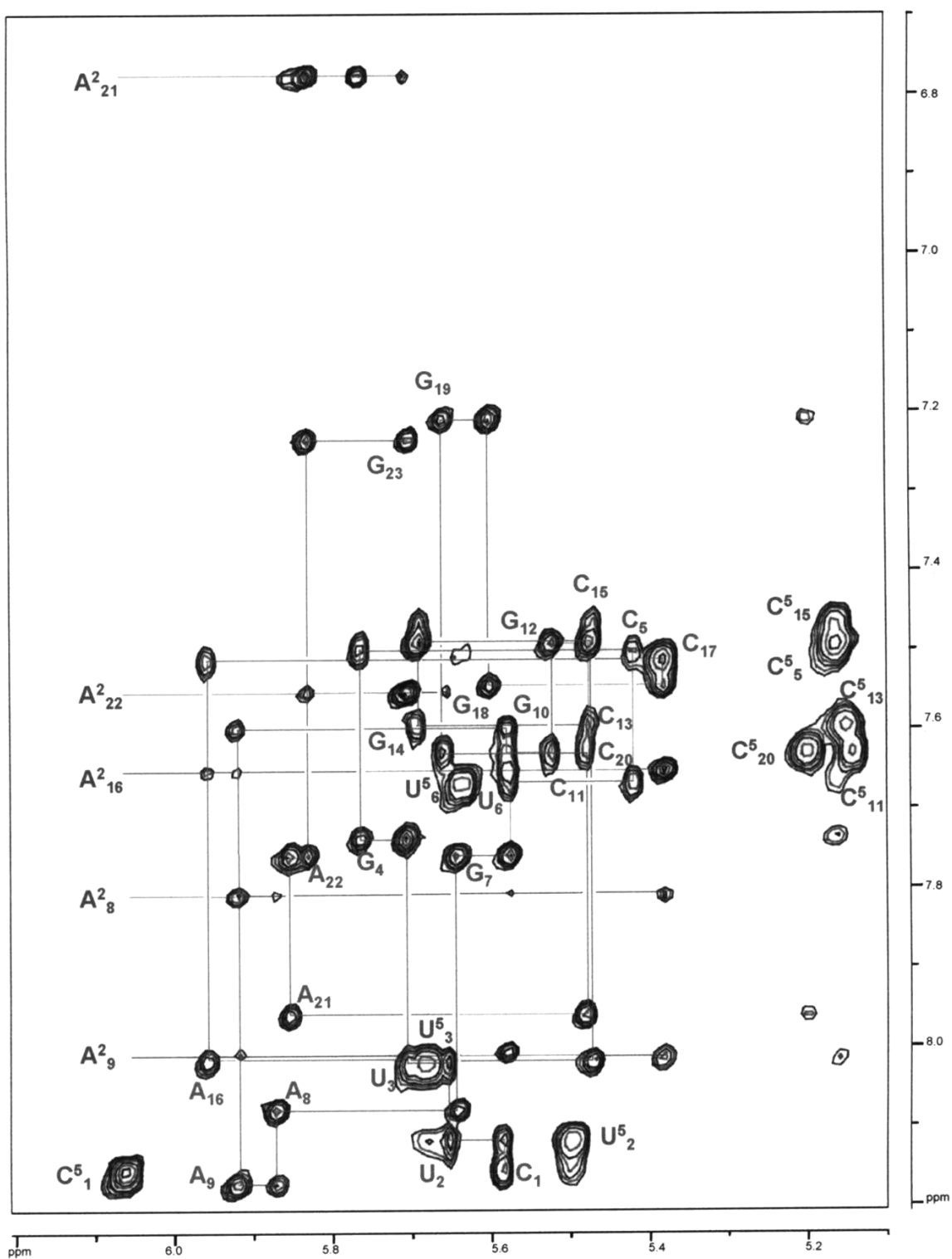


Figure III.8 : Spectre NOESY du dimère stable de SL1-23 mer de VIH-1<sub>Lai</sub>. Agrandissement de la région anomérique/aromatique. Conditions : 300ms, D<sub>2</sub>O, 25°C et pH=5,7.

Des informations intéressantes sont apportées par le comportement des protons H2 des adénines 8,9 et 16. Les résonances des protons H2 ont été attribuées grâce à leur proximité relative des protons anomériques de leur propre résidu et de leur voisin en 3'. On observe une forte connectivité H2 (A<sub>9</sub>) et H2 (A<sub>16</sub>) ainsi qu'une plus faible pour H2 (A<sub>8</sub>) et H2 (A<sub>16</sub>). De plus, nous constatons un faible NOE entre H2 (A<sub>8</sub>) et H5 (C<sub>17</sub>). Tous ces renseignements nous amènent à suggérer que la proximité des résidus A<sub>8</sub>, A<sub>9</sub> et A<sub>16</sub> est faible. Au vue des spectres RMN, la résolution de la structure tridimensionnelle semble possible.

### 3 Modélisation moléculaire

#### 3.1 Etablissement des contraintes

Dans le cas du dimère stable d'ARN SL1, les résonances <sup>31</sup>P sont concentrées dans une fenêtre de déplacement chimique relativement étroite, de l'ordre de 1,5ppm. Le faible étalement en déplacement chimique ne permet pas d'identifier leurs résonances (par l'établissement d'expérience de type HSQC par exemple). Néanmoins, l'observation qualitative du spectre unidimensionnelle révèle une absence de déviation importante de la conformation du squelette phosphodiester tout au long de la séquence. En ce sens, il a été choisi de contraindre les angles  $\alpha, \beta, \gamma, \epsilon$  et  $\zeta$  avec les valeurs classiques d'une hélice de type A (Saenger, 1984) avec une incertitude de 15°. Cependant, les régions du squelette englobant les adénines de jonction A<sub>8</sub>, A<sub>9</sub> et A<sub>16</sub> sont laissées sans contraintes, autorisant une plus grande liberté de conformation de ces résidus. Remarquons que l'angle  $\delta$  n'est classiquement pas imposé lors de la modélisation, puisqu'il rend compte du plissement du sucre et serait redondant avec ces contraintes.

L'étude du spectre TOCSY nous renseigne sur la conformation des unités riboses des résidus. Comme dans le cas du monomère, l'absence ou la faiblesse des constantes de couplage entre le proton H1' et le proton H2' sur le spectre TOCSY indique que les sucres sont en conformation nord. De cette façon, le plissement du sucre a pu être déduit. Tous les résidus se trouvent sous forme C3' endo, excepté les résidus G<sub>7</sub>, A<sub>8</sub>, A<sub>9</sub> et G<sub>23</sub> qui apparaissent sous forme sud ( $^3J_{H1'-H2'} \sim 7\text{Hz}$ ). Nous avons contraint tous les angles  $\nu_1$  et  $\nu_2$

avec des valeurs correspondantes à une conformation nord ( $\nu_1 = -26,0 \pm 10^\circ$  et  $\nu_2 = +24,7 \pm 10^\circ$ ), sauf pour les résidus 7,8,9 et 23 ( $\nu_1 = 37 \pm 10^\circ$  et  $\nu_2 = -34,8 \pm 10^\circ$ ).

Les effets NOE correspondant aux protons échangeables ont été rangés en trois classes : fort ( $1,8 \text{ \AA} < \text{distance} < 3,5 \text{ \AA}$ ), moyen ( $1,8 \text{ \AA} < \text{distance} < 5 \text{ \AA}$ ) et faible ( $1,8 \text{ \AA} < \text{distance} < 6 \text{ \AA}$ ). Ces approximations sont réalisées dans le souci de rendre compte de la vitesse d'échange avec le solvant. Les effets NOE correspondant aux protons non-échangeables (attribués dans l'eau lourde) ont été mesurés pour différents temps de mélange (0, 50, 90, 120 et 300ms). Le calcul de la matrice de relaxation, qui est réalisé à l'aide du programme RELAZ (Lancelot, et al., 1989), traduit les effets NOE en distances. Les distances sont ensuite introduites sous forme d'un intervalle de 15% autour de la valeur calculée. Le résumé des connectivités NOE inter- et intra-résidus observées lors de l'étude du dimère stable est présenté sur la figure III.9. Le tableau III.1 résume les différentes contraintes imposées dans la modélisation.

| <u>Contraintes de distances</u>                                       |       |
|---|-------|
| intra-résidu  | 156   |
| inter-résidu  | 326   |
| protons échangeables et WC  | 108   |
| Total contraintes de distance   | 590   |
| Contraintes NOE/résidu  | 12,83 |
| <u>Contraintes d'angle</u>  |       |
| Plissement du sucre ( $\nu_0$ et $\nu_2$ )                            | 92    |
| Backbone ( $\alpha$ , $\beta$ , $\gamma$ , $\varepsilon$ et $\zeta$ ) | 174   |
| Total contraintes d'angle   | 266   |
| Total contraintes   | 856   |
| Contraintes/résidu  | 18,61 |

Tableau III.1 : Résumé des différentes contraintes incorporées dans la modélisation moléculaire.

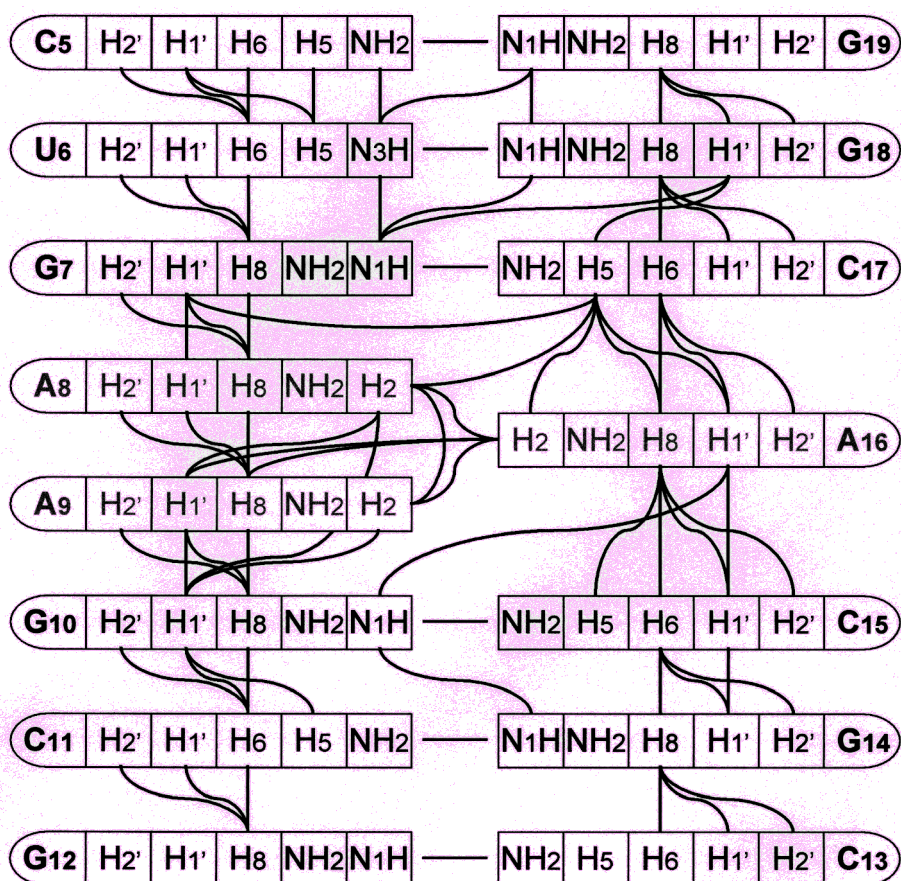


Figure III.9 : Résumé des données NOE déterminées pour la zone  $C_5$ - $G_{12}$ / $C_{13}'$ - $G_{19}'$ .

Similairement à la modélisation précédente, des contraintes de plans ont été ajoutées pendant les premiers cycles de calcul. Elles favorisent l'appariement des bases et concernent les plans :  $C_1G_{23}$ ,  $U_2A_{22}$ ,  $U_3A_{21}$ ,  $G_4C_{20}$ ,  $C_5G_{19}$ ,  $C_{11}G_{14}$ ,  $G_{12}C_{13}$  et leurs symétriques. La constante de force pour chaque plan est de  $10 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$ .

### 3.2 Résolution de la structure tridimensionnelle

A ce point de l'étude, il n'est pas possible de trancher entre les deux types de dimère par simple considération qualitative des données fournies par la RMN. Bien que le réseau de contraintes de distance (figure III.9) semble indiquer un dimère stable, il ne permet pas de conclure définitivement sur un modèle de structure secondaire. L'utilisation d'un échantillon non-marqué ne permet malheureusement pas de déterminer les distances



inter-brin d'un dimère. La seule information de la structure secondaire dont nous sommes sûrs est qu'il s'agit bien d'un dimère. En effet, le temps de corrélation calculé pour notre échantillon est de 6,0 ns, ce qui est compatible avec un dimère (rappelons que celui du monomère est de 3,1ns).

L'étape de modélisation moléculaire est cruciale. Afin de mener une exploration rigoureuse et de ne prendre aucun a priori, nous avons décidé de considérer les deux modèles possibles. Par conséquent, le premier modèle fixe un appariement inter-brin tout au long de la séquence ( $C_1-G_{23}', \dots, G_{23}'-C_1'$  ; l'apostrophe désigne le brin complémentaire) et correspond au dimère linéaire. Le second modèle fixe un appariement intra-brin pour chaque tige ( $C_1-G_{23}, \dots, G_7-C_{17}$  et  $C_1'-G_{23}', \dots, G_7'-G_{23}'$ ) et inter-brin pour la zone complémentaire ( $G_{10}-C_{15}', \dots, C_{15}'-G_{10}'$ ). Il définit le dimère de type "kissing-complex". La sélection du modèle se fait selon le nombre et l'amplitude des violations de distances et d'angles obtenus pour chaque famille de structures.

Un échantillonnage de 100 structures aléatoires a été créé en utilisant le logiciel DYANA 1.5 (Guntert, et al., 1997). L'intérêt d'utiliser ce logiciel est qu'il génère très rapidement des structures aléatoires qui respectent la covalence des liaisons. Ces structures sont créées avec les angles de torsion du squelette aléatoire. Elles sont utilisées comme coordonnées de départ pour les deux modèles considérés. Pour chaque modèle, les structures sont soumises à un protocole de dynamique moléculaire utilisant la méthode du recuit simulé dans l'espace cartésien à l'aide du logiciel XPLOR (version 3.851) (Brünger, 1992). Le champ de force est simplifié de la même façon que pour la modélisation du monomère. Les constantes de force ont été fixées à  $50 \text{ kcal.mol}^{-1}.\text{Å}^{-2}$  pour les contraintes de distances et  $25 \text{ kcal.mol}^{-1}.\text{rad}^{-2}$  pour les contraintes d'angles dièdres.

Le protocole de calcul débute par une minimisation de 1000 pas suivant l'algorithme du gradient conjugué (Powell, 1977). Les interactions de contact et de nature non-covalente sont de ce fait éliminées. Les vitesses initiales sont ensuite attribuées suivant une distribution statistique de Maxwell tandis que les masses sont uniformément attribuées à 10ua. Une dynamique de 20ps à 2000 K permet l'exploration conformationnelle sous contraintes RMN. S'en suit une période de refroidissement (de 2000 K à 300 K) de 10ps de

dynamique. Enfin, comme au début, une minimisation par mécanique moléculaire de 1000 pas termine ce protocole. Les structures sont alors examinées en fonction de leurs énergies, leurs violations (distances et angles dièdres) puis graphiquement.

Dans le cas envisagé d'un dimère de type "kissing-complex", aucune structure n'a convergé. Dans le meilleur des cas, nous observons une structure avec 14 violations de distances, toutes observées dans la région des résidus G<sub>7</sub>, A<sub>8</sub>, A<sub>9</sub>, G<sub>10</sub>, A<sub>16</sub>, C<sub>17</sub>. Pour cette dernière structure, il faut noter que les tiges sont bien définies et les angles dièdres ne montrent aucune violation. Nous pouvons conclure que les contraintes de distances impliquant la région des résidus de jonction (7, 8, 9, 10, 15, 16 et 17) sont capitales et que les principales différences structurales entre les deux dimères s'observeront dans cette région.

En revanche, dans le cas d'un modèle de dimère étendu, 20% des structures de départ convergent vers ce modèle. Une dernière phase de 1000 pas de minimisation par mécanique moléculaire est alors réalisée. Le champ de force utilisé ajoute les termes du potentiel attractif de van der Waals (fonction de Lennard-Jones), le potentiel d'énergie électrostatique et le potentiel de liaison hydrogène. Un ensemble de 19 structures finales, ne comportant aucune violation de distance supérieure à 0,3Å et aucune violation d'angle dièdre supérieure à 5°, est généré. La figure III.10 représente la superposition de ces 15 structures. Le tableau III.2 résume les énergies et RMSD de cette ensemble structural. Les RMSD sont calculés par comparaison avec une structure moyenne non-minimisée.

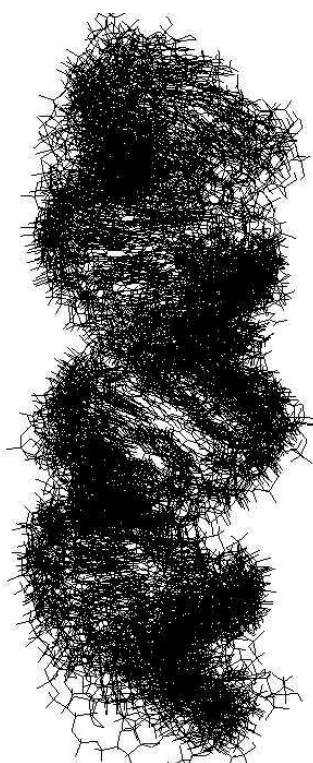


Figure III.10 : Superposition des 15 structures convergentes sur la structure moyenne.

| <u>Précision des structures</u>   |      |              |                        |
|---|------|--------------|------------------------|
| Nombre de structures  |      | 15           |                        |
| Nombre de violations NOE>0.5Å   |      | 0            |                        |
| Nombre de violations NOE>0.3Å   |      | 8 ±4         |                        |
| Nombre de violations dièdre>5°  |      | 0            |                        |
| RMSD <sup>1</sup> global  |      | 2.44 ±0.82 Å |                        |
| RMSD tige1 (C1-G7,C17'-G23')  |      | 0.86 ±0.36 Å |                        |
| RMSD tige2 (C1'-G7',C17-G23)  |      | 0.88 ±0.35 Å |                        |
| RMSD tige3 (G10-C15 ,G10'-C15')   |      | 0.97 ±0.58 Å |                        |
| RMSD jonction1 (A8,A9 et A16')  |      | 0.78 ±0.31 Å |                        |
| RMSD jonction2 (A8',A9' et A16)   |      | 0.89 ±0.50 Å |                        |
| <u>Energie</u>  |      |              |                        |
| Energie totale  | -359 | ±45          | kcal.mol <sup>-1</sup> |
| Energie de liaison  | +49  | ±2           | kcal.mol <sup>-1</sup> |
| Energie angulaire   | +433 | ±9           | kcal.mol <sup>-1</sup> |
| Energie torsion   | +805 | ±8           | kcal.mol <sup>-1</sup> |
| Energie dièdre impropre   | +84  | ±6           | kcal.mol <sup>-1</sup> |
| Energie van der Waals   | -712 | ±13          | kcal.mol <sup>-1</sup> |
| Energie électrostatique   | -895 | ±19          | kcal.mol <sup>-1</sup> |
| Energie liaison hydrogène   | -282 | ±11          | kcal.mol <sup>-1</sup> |
| Energie NOE   | +146 | ±6           | kcal.mol <sup>-1</sup> |
| Energie dièdre  | +13  | ±2           | kcal.mol <sup>-1</sup> |
| <sup>1</sup> : RMSD calculé sur les atomes lourds par rapport à la moyenne non minimisée. |      |              |                        |

Tableau III.2 : Résumé des énergies et des RMSD pour SL1\*.

## 4 Analyses structurales

La première observation que nous pouvons remarquer sur cette structure du dimère étendu, est qu'elle s'apparente à une structure classique d'hélice. En effet, si l'on compare notre structure à son homologue auto-complémentaire organisé en hélice A (cf figure III.11) une grande similarité s'offre à nos yeux. Il est essentiel de noter que la stabilité de ce dimère provient sans aucun doute de cette homogénéité de structure ou plutôt de cette absence "d'accident" de structure.

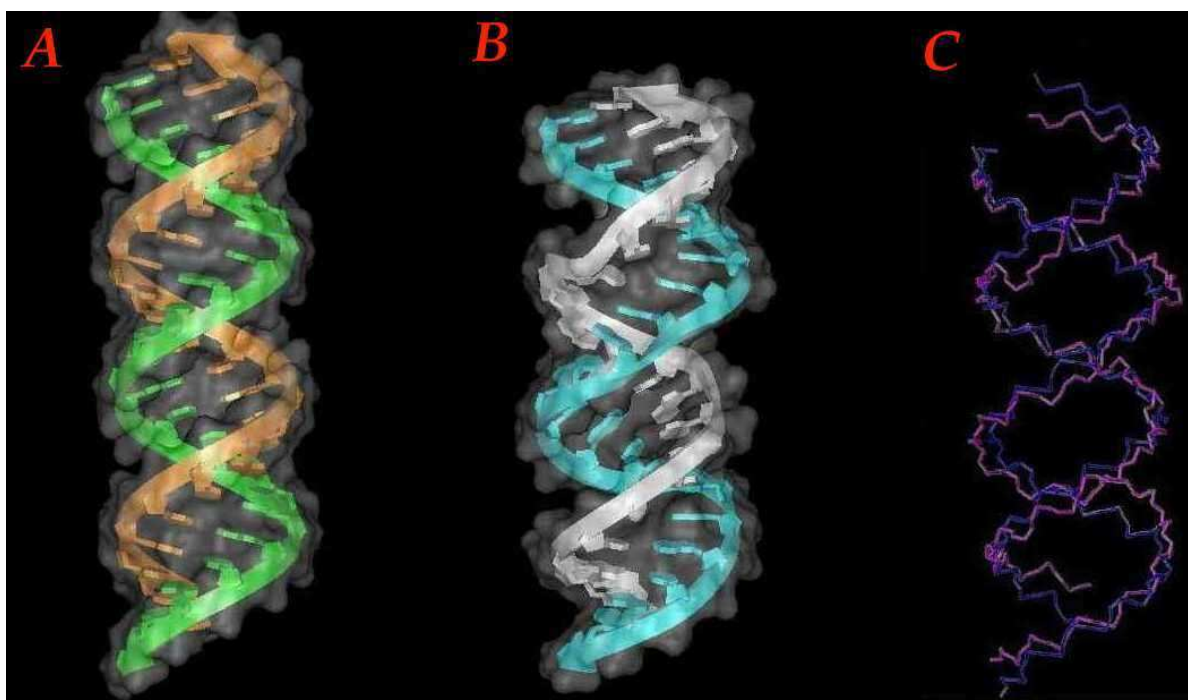


Figure III.11 : Organisation globale du dimère stable de SL1-23 mer de VIH-1<sub>Lai</sub>. A) Représentation en hélice A de l'ARN complémentaire de la séquence SL1 (brin SL1 + brin complémentaire : CUUGCCGUGCGCGCUUCAGCAAG) cette séquence sert de référence. B) Structure du dimère stable, orientée identiquement à la l'hélice de référence. C) Superposition des deux squelettes phosphodiester des structures A (bleue) et B (violet), les structures sont superposées en sélectionnant uniquement les atomes de phosphore.

### 4.1 Organisation des tiges

La structure du dimère étendu de SL1-23mer comporte trois zones hélicoïdales, appelées tiges. Deux de ces trois zones sont symétriques et donc identiques. Elles comprennent les

résidus :  $C_1-G_7, C_{17}'-G_{23}'$  et  $C_{17}-G_{23}, C_1'-G_7'$  (l'apostrophe désigne le brin opposé). L'autre tige implique la zone auto-complémentaire des résidus  $G_{10}-C_{15}, G_{10}'-C_{15}'$ . L'ensemble des paramètres d'hélices a été déterminé en utilisant le logiciel CURVES (version 5.1) (Lavery and Sklenar, 1988; Lavery and Sklenar, 1989; Lavery and Sklenar, 1990; Ravishanker, et al., 1989).

Le calcul du RMSD de la première tige (et de son symétrique) montre une valeur faible (0,86 et 0,88 Å, cf tableau III.2) reflétant une bonne convergence de cette région. Le "rise" axial moyen (se référer à l'annexe A1 pour les différentes définitions des paramètres hélicoïdaux) est de 2,83Å avec une rotation moyenne de 30° par résidu. Le petit sillon, qui est peu profond dans le cas des ARN, présente une largeur moyenne de 9,75Å tandis que le sillon profond n'est que de 2,99Å. Ces paramètres sont en accord avec ceux définissant les hélices d'ARN classiques de types A (Saenger, 1984). On notera toutefois un manque de convergence observé pour la paire de bases  $C_1-G_{23}$  qui est justifié par la diminution des forces d'empilement des bases du fait de la position terminale de cette paire de bases. Cette tige (et par là-même son symétrique) caractérise une zone de forte stabilité pour la structure sous forme de dimère étendu.

Cependant, la présence d'une paire UG n'est pas sans influence sur la géométrie hélicoïdale. En effet, la paire de bases  $U_6-G_{18}$  située presque en bout de tige provoque certaines disparités dans l'enchaînement des paires de bases. L'uracile pointant vers le petit sillon génère un creusement du grand sillon. Comme il a déjà été observé (Mizuno and Sundaralingam, 1978), la paire  $5'U-G3'$  présente un recouvrement du au stacking très faible avec la paire de bases qui la précède. De ce fait, la paire  $U_6-G_{18}$  déstabilise légèrement la série d'empilement des paires de bases en affaiblissant la région terminant cette tige, ainsi la paire de bases suivante est moins stable et présente une valeur de "buckle" ( $\alpha=25^\circ$ ) importante (cf figure III.12). Cette caractéristique est bien reflétée par les données RMN : les spectres NOESY dans l'eau légère ont montré que le proton imino de la guanine 7 présente un échange important avec le solvant.

La tige centrale présente un RMSD faible indiquant une convergence correcte des structures générées (0,97Å pour les résidus  $G_{10}-C_{15}$  et  $G_{10}'-C_{15}'$  sur les 19 structures).

L'organisation des paires de bases est particulièrement régulière mis à part les résidus des extrémités (G<sub>10</sub> et C<sub>15</sub> de chaque brin) qui présentent des valeurs de roll ( $\rho$ ) supérieures à la moyenne (cf figure III.12). Ceci est la conséquence des contraintes générées par le mésappariement voisin. Cette hélice présente un faible angle de courbure d'une dizaine de degrés. La conformation globale de la séquence est de type A.

#### 4.2 Caractérisation du motif en fermeture éclair

Le mésappariement représente la principale originalité de cette structure. Ce motif, pour des raisons de symétrie, est présent à deux reprises bien qu'aucun critère de symétrie n'ait été employé lors de l'étape de modélisation moléculaire. La figure III.12 représente l'arrangement structural de ces résidus.

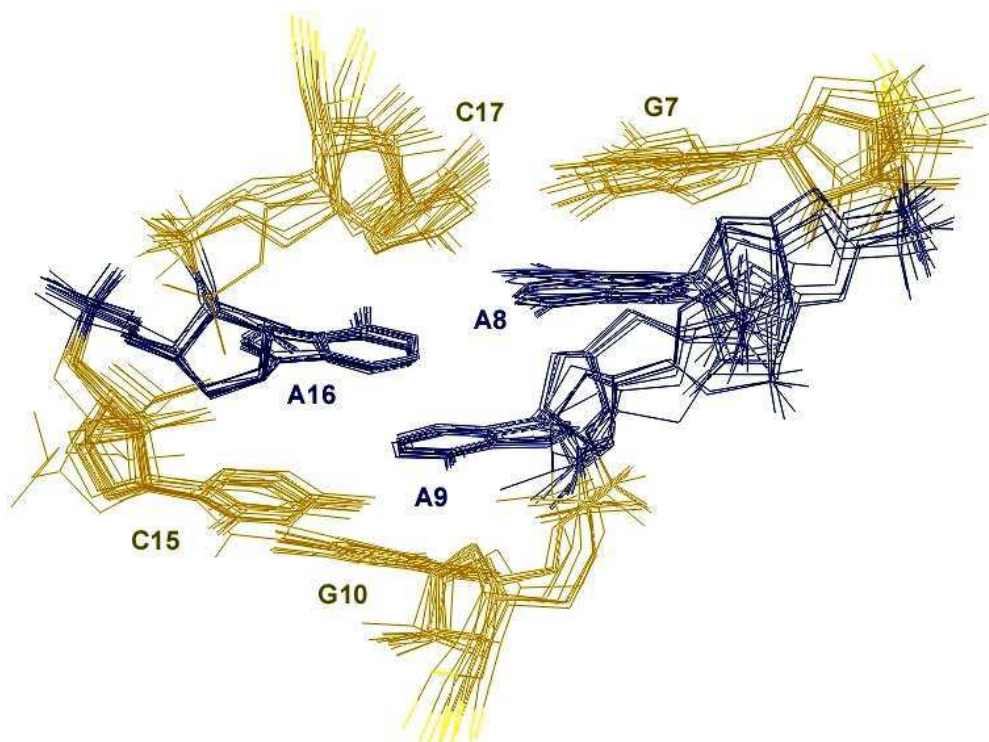


Figure III.12 : Agrandissement de la région impliquant les adénines de jonction (8,9 et 16). La superposition des 15 meilleures structures montre un motif en fermeture éclair qui stabilise le dimère. Nous pouvons aussi noter le buckle présent pour la paire de bases G<sub>7</sub>-C<sub>17</sub> situé au dessus de la base A<sub>8</sub>.

Ce mésappariement fait subir des distorsions importantes au squelette phosphodiester. Les sucres des résidus A<sub>8</sub> et A<sub>9</sub> adoptent une conformation C<sub>2</sub>'-endo. Le grand sillon se rétrécit tandis que le petit s'élargit. La valeur du "rise" entre les bases 8 et 9 est

significativement augmentée d'environ 2Å. Le résidu A<sub>16</sub>, du brin opposé, vient alors se glisser entre les résidus A<sub>8</sub> et A<sub>9</sub> provoquant ainsi un stacking partiel (cf figure III.12). Comme nous l'avons constaté précédemment, les paires de bases voisines subissent l'effet perturbant de ce motif. En effet, la tension du squelette phosphodiester est générée par l'absence d'un résidu en 3' de la base C<sub>17</sub>, ce qui conduit à une rotation du plan de cette base.

L'effet déstabilisant est cependant compensé par l'augmentation du stacking engendrée par le motif en fermeture éclair. Ce type d'interaction d'empilement inter-brin a déjà été mis en évidence dans d'autres structures notamment pour la jonction de la boucle T et B de l'ARNt (Kim, et al., 1974; Robertus, et al., 1974) et l'ARN aptamère de la théophylline (Zimmermann, et al., 1997).

En conclusion, la présence du "bulge" d'adénines dans la structure du dimère étendu SL1-23mer de VIH-1<sub>Lai</sub> ne génère pas un motif structural déstabilisant. Au contraire, la position de ces adénines de jonction en un empilement inter-brin sous forme de fermeture éclair induit une stabilisation de la structure par un stacking partiel et permet une organisation globale du dimère en hélice A. Ceci explique pourquoi ce dimère est dit dimère stable. En effet, nous pouvons difficilement envisager une autre structure pour ce dimère présentant une telle stabilité structurale.

## 5 Comparaison avec des études structurales analogues

Plusieurs mois après la publication de nos résultats, l'équipe du professeur T.L. James a publié un travail similaire sur une souche mutée de SL1-23mer (Mujeeb, et al., 1999). Cette séquence est modifiée au niveau de la tige (cf figure III.13).

Notons tout d'abord que le nombre de contraintes de distances issu de l'attribution RMN est loin d'être le même : 394 contraintes dans ce cas (soit 8,6 contraintes par résidu) contre 590 pour notre souche (soit 12,8 contraintes par résidu). Ceci semble préjudiciable car une structure RMN est principalement dépendante de la qualité et surtout du nombre des contraintes de distances (Liu, et al., 1992). Leur nombre impressionnant de contraintes

d'angles dièdres provient principalement des angles impliqués dans le plissement du sucre. En effet, dans leur étude, cinq angles définissent la conformation du cycle ribose (soit un total de 230 contraintes) tandis que nous imposons simplement deux angles dièdres par cycle ribose (soit un total de 96 contraintes d'angles dièdres) ce qui, géométriquement, est suffisant pour décrire l'ensemble des plissements sans pour autant augmenter artificiellement l'énergie totale de la molécule.

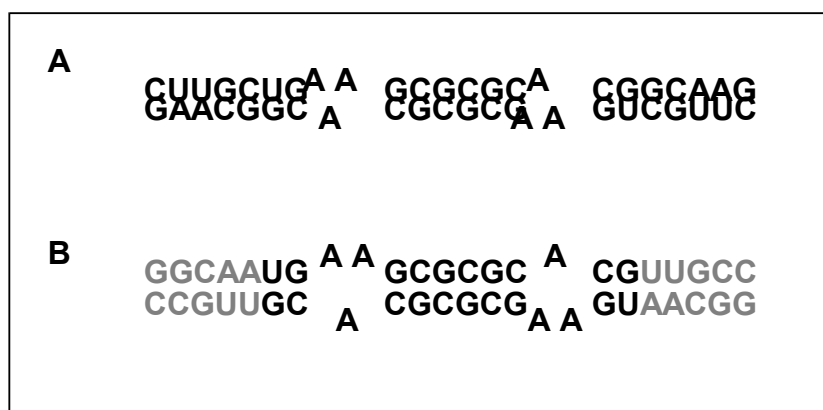


Figure III.13 : Séquences SL1-23mer de VIH-1<sub>Lai</sub>. A/ souche sauvage B/ souche mutée, les résidus en gris représentent les zones modifiées.

La structure tridimensionnelle s'organise, là aussi, sous la forme d'un dimère linéaire avec des propriétés quasi-analogues à celles décrites auparavant. Les caractéristiques du bulge d'adénines et de la séquence auto-complémentaire semblent comparables à notre description.

Une autre structure de dimère stable a également été récemment élucidée (Ennifar, et al., 1999). Il s'agit de la structure cristallographique du dimère linéaire SL1-23 mer de la souche VIH-1<sub>Mal</sub> dont la séquence est représentée sur la figure III.15. Dans le cas de cette souche, il semblerait que le dimère de type kissing-complex soit présent en solution uniquement (Paillart, et al., 1996; Dardel, et al., 1998). Un modèle tridimensionnelle, issu d'études de sondes chimiques, a été proposé (Paillart, et al., 1997) mais il n'a jamais été confirmé. En effet une étude RMN a été réalisée mais sa structure tridimensionnelle n'a jamais été déterminée (Dardel, et al., 1998). L'étude montrait néanmoins des caractéristiques concluant sur l'existence d'un dimère de type kissing-complex.





## 6 *La structure du dimère instable*

A ce jour, la seule étude structurale publiée du kissing-complex de SL1 reste le travail de l'équipe de James (Mujeeb, et al., 1998). Cette étude repose sur le fragment muté, au niveau de la tige, de SL1-23mer dont la séquence a déjà été présentée sur la figure III.13. La mutation est nécessaire, d'après les auteurs, à l'optimisation du rendement de synthèse par voie enzymatique. Cette mutation engendrerait également une meilleure stabilité de la tige, et donc celle du complexe boucle-boucle.

L'étude RMN met en évidence un homodimère symétrique où les tiges sont organisées en hélice de type A canonique. La paire  $U_6-G_{18}$  présente un appariement identique à celui que nous avons décrit auparavant. Par contre, la paire de bases  $G_7-C_{17}$ , qui précède la région des adénines de jonction, est totalement rompue. La séquence auto-complémentaire est entièrement appariée, mais présente une déviation importante par rapport à la forme canonique. Les auteurs décrivent un sous enroulement global de l'hélice de  $18^\circ$ , ainsi que de hautes valeurs de roll ( $\sim 26^\circ$  en moyenne) pour les résidus de la séquence palindromique. La figure III.18 représente cette structure.

Les deux tiges comprennent les résidus de 1 à 6 et de 18 à 23, contrairement à la structure du duplexe étendu où les résidus  $G_7$  et  $C_{17}$  appartiennent à ces tiges car ils sont appariés. La valeur du RMSD est faible ( $\text{RMSD}=0,87\pm 0,36 \text{ \AA}$ ) et indique une bonne résolution de ces régions. Cependant, les auteurs rapportent une fluctuation de  $30^\circ$  entre les axes respectifs de chacune des tiges (cf figure III.18-b). Cette flexibilité est décrite comme étant issue des variations des angles de torsion impliquant les résidus de la jonction  $G_7, A_8, A_9, A_{16}, C_{17}$ . La tige centrale, où siège l'interaction boucle-boucle entre les deux tiges-boucles, forme une hélice de type A bien qu'aucune contrainte en ce sens ne soit imposé.

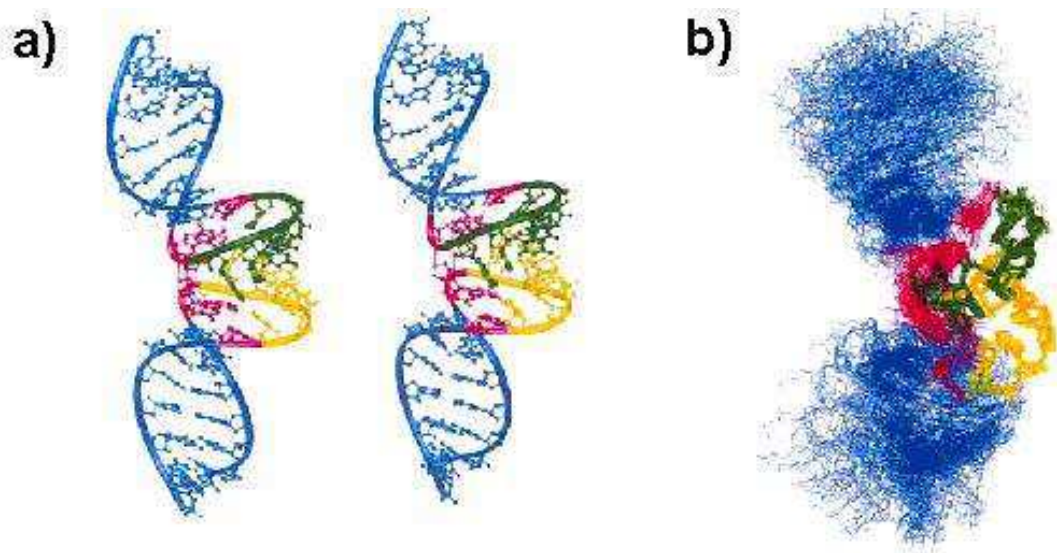


Figure III.18 : Structure du dimère en kissing-complex de SL1 mutée, d'après Mujeeb et al (Mujeeb, et al., 1998). a) Vue stéréoscopique globale de la molécule. b) Superposition des 34 structures ayant convergé.

Le point essentiel sur lequel repose l'organisation de la structure concerne l'architecture de la zone charnière entre les tiges. Elle concerne les résidus  $G_7$ ,  $A_8$ ,  $A_9$ ,  $A_{16}$ ,  $C_{17}$ . L'intercalation du résidu  $A_9$  du premier brin, entre les résidus  $A_8'$  et  $C_{17}'$  du second brin (l'apostrophe désigne l'autre brin), crée un motif de stacking à trois bases stabilisant la région et compensant la rupture de la liaison  $G_7-C_{17}$ . Ce motif est représenté sur la figure III.19. Néanmoins, aucun spectre RMN du proton n'a jamais été présenté. De plus, l'article ne décrit pas la nature des contraintes permettant de définir le motif d'intercalation. Même si depuis quelques années, les structures de type "kissing-complex" se multiplient, le nombre de structures décrites et publiées demeure faible.

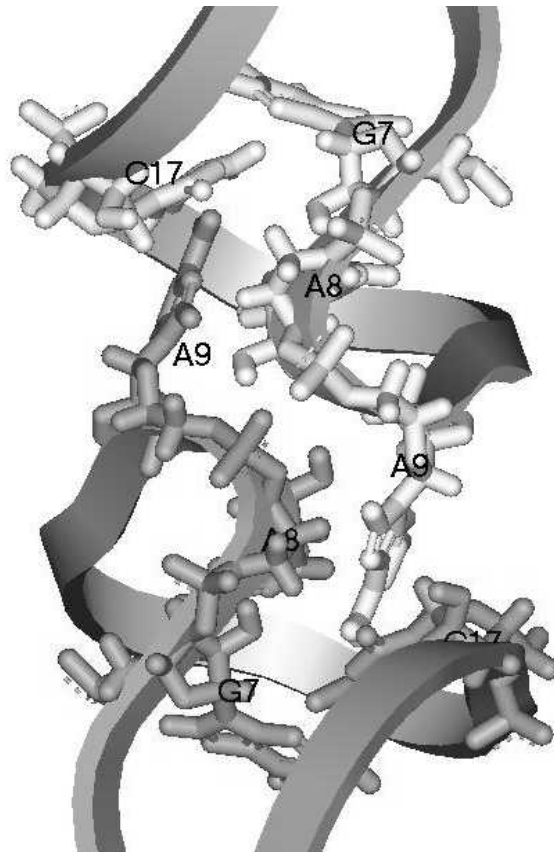


Figure III.19 : Représentation du motif de stacking inter-brin de la structure kissing-complex de SL1. Le résidu A<sub>9</sub> est intercalé entre les résidus C<sub>17</sub> et A<sub>8</sub> du brin opposé.

Plusieurs études mettent en doute la véracité de cette structure. Les études sur le mécanisme de transition (Theilleux-Delalande, et al., 2000) faites par notre équipe montrent un appariement complet de la paire de bases G<sub>7</sub>-C<sub>17</sub>, et cela quelque soit le dimère envisagé. Cette information est capitale, car la rupture de cette liaison est le point central du motif inter-brin rencontré dans la structure de Mujeeb. La présence de cette paire de bases a été confirmée par d'autres études, notamment celle du professeur Kawai (Takahashi, et al., 2000) et lors d'études RMN NOESY sur un échantillon marqué (Girard, 2000) où la paire de bases apparaît sans ambiguïté. De plus, nous pouvons noter, sur la structure proposée (cf figure III.19), la présence de deux oxygènes chargés négativement appartenant aux squelettes phosphodiester de chaque brin (atome O1P du résidu A<sub>9</sub>) et éloignés de simplement 1,7Å. Cette éloignement (inférieur à la somme des rayons de van der Waals) nous semble très faible et est certainement dommageable pour l'énergie électrostatique du système. Enfin, plus surprenant encore, l'analyse des violations du

rayon de van der Waals montre 72 violations supérieures à 0,5Å, ce qui est très étonnant pour une structure minimisée à l'aide du logiciel AMBER avec un champ de force étendu. Ceci explique peut-être la valeur, très faiblement négative, de l'énergie totale calculée (-5,041±75 kcal.mol<sup>-1</sup>).

Très récemment, l'équipe Ehresmann (Strasbourg) a déterminé la structure cristallographique des dimères instables des SL1 isolats Lai et Mal. Bien que ces résultats ne soient pas à l'heure actuelle encore publiés, leur présentation orale (Orléans, 17 mai 2001) décrivait un kissing-complex dans les deux cas. On peut constater que leur structure n'est pas du tout semblable à celle décrite par Mujeeb et al (Mujeeb, et al., 1998) pour l'isolat Lai : en particulier les adénines A<sub>8</sub>, A<sub>9</sub> et A<sub>16</sub> sont dans des positions très différentes, la paire G7C17 est formée et les deux boucles ne sont pas perpendiculaires aux tiges mais dans leur prolongement. Dans le cas de la souche Mal, la structure cristallographique est complètement différente de celle proposée par l'équipe de Westhof (Paillart, et al., 1997).

Seulement trois autres structures de kissing-complex viennent compléter ces études. Ces structures ont été élucidées par RMN : dans le système de régulation du plasmide Cole1 (Lee and Crothers, 1998; Marino, et al., 1995) ; dans le système TAR-TAR\* du VIH (Chang and Tinoco, 1994; Chang and Tinoco, 1997) ; et dans le cas du dimère de la séquence H3 du virus de la maladie murine de Moloney (MoMuLV) (Kim and Tinoco, 2000). Ces différentes structures publiées ne présentent pas réellement de caractéristiques communes. Notons que dans tous les cas nous observons des distorsions de l'hélice, induites par l'interaction boucle-boucle. Celles-ci dépendent de la séquence considérée.

## ***7 Conclusions***

L'étude présentée dans ce chapitre a permis de mettre en évidence l'existence de deux formes dimériques distinctes de la séquence SL1-23mer de VIH-1<sub>Lai</sub>. Ces deux formes sont isolables, suivant le mode de préparation envisagé, et permettent l'élucidation structurale des deux types de dimères.

Le dimère stable adopte une forme linéaire étendue. Sa principale caractéristique est de

développer deux motifs en fermeture éclair engageant les trois adénines de jonction (8,9 et 16). Ces motifs, dont l'empilement est favorable énergétiquement, assurent une organisation globale en hélice de type A du dimère. De cette façon, le dimère adopte un profil hélicoïdal quasi-parfait (cf figure III.11) en minimisant la distorsion du squelette phosphodiester. La comparaison avec des études structurales analogues confirme largement notre étude.

L'étude du dimère instable a été réalisée par l'équipe de James (Mujeeb, et al., 1998). La structure présentée montre un motif symétrique métastable impliquant une rupture de la paire de bases G<sub>7</sub>-C<sub>17</sub>. Cette structure a été plusieurs fois critiquée et son étude est actuellement poursuivie au laboratoire au moyen d'un échantillon marqué isotopiquement à l'azote 15 et au carbone 13.

## IV. Conclusions

Le fragment SL1-23mer de VIH-1<sub>Lai</sub>, comprenant 23 nucléotides est un bon modèle du comportement macroscopique de l'ARN rétroviral dans l'initiation à la dimérisation du rétrovirus VIH-1<sub>Lai</sub>. L'étude présentée dans ce chapitre a permis de caractériser structurellement, par résonance magnétique nucléaire, les trois espèces présentes dans ce processus biologique.

La première structure à réaliser correspondait au monomère SL1. Cette espèce se rencontre avant le processus de dimérisation. Il a été nécessaire d'utiliser un échantillon muté, pour éviter le phénomène parasite de dimérisation spontanée. Néanmoins, cette mutation n'implique pas un comportement différent du fragment sauvage. La structure s'organise comme une épingle à cheveux. La tige, grande de 7 paires de bases, permet de stabiliser la boucle. L'entrée de boucle montre un relatif degré de structuration, dont rend compte le calcul du RMSD par résidu. Cette entrée de boucle, impliquant les trois adénines de jonction (8, 9 et 16) forme un motif stabilisant et expulse la séquence auto-complémentaire G<sub>10</sub>C<sub>11</sub>G<sub>12</sub>C<sub>13</sub>G<sub>14</sub>C<sub>15</sub> hors de la boucle. Cette forte dynamique a ensuite été confirmée par une simulation de dynamique moléculaire en présence d'un solvant explicite. L'exposition de la séquence autocomplémentaire au solvant a déjà été supposée lors d'étude d'inhibiteurs d'oligonucléotides anti-dimérisation (Lodmell, et al., 1998). Ainsi, le monomère SL1 est sans doute la cible la plus intéressante pour une stratégie anti-sens anti-dimérisation.

Le mode de préparation des dimères permet de sélectionner sans ambiguïté l'espèce envisagée pour l'étude. La seconde structure que nous avons réalisée, représente le dimère stable de la séquence SL1. Cette structure se rapproche d'une structure hélicoïdale classique, ce qui explique la grande stabilité de ce dimère. Cette organisation structurale est principalement due aux deux motifs en fermeture éclair impliquant, là encore, les trois adénines de jonction. Ce motif permet un stacking partiel des bases et minimise la

distorsion du squelette phosphodiester. Les études postérieures faites sur des échantillons de même taille, mutés ou d'une autre souche, confirment notre étude et apportent des données intéressantes sur la mobilité du motif.

Enfin, la dernière structure correspond au dimère instable d'une séquence mutée de SL1. Cette structure en "kissing-complex" représente une interaction boucle-boucle, avec un appariement incomplet de la tige. Elle montre un motif inter-brin impliquant les adénines 8 et 9 ainsi que la cytosine 17. Cet empilement métastable des trois bases explique la très relative stabilité du dimère. Cependant, cette structure a été largement critiquée et est en cours de réalisation dans notre équipe.

Un des points intéressants que nous avons voulu étudier est le rôle qu'apporte le groupement 2'OH dans ces différentes structures d'ARN. Ce groupement différencie, à lui seul, un ADN d'un ARN (l'importance du groupement méthyl sur les thymines n'explique pas les nombreuses différences structurales rencontrées pour ces deux espèces). La seconde partie de cette thèse va donc s'attacher à déterminer structurellement le comportement de la séquence analogue à SL1 en désoxyribose.



## Deuxième partie

*Etudes structurales des  
différentes conformations  
de l'analogue désoxyribose  
SL1 de VIH-1<sub>Lai</sub>*

# I. Problématique et objectifs

Bien que les ARN et les ADN ont des rôles très différents du point de vue biologique, leurs formules moléculaires sont très proches. La première différence que nous pouvons noter est la présence de la base thymine présente dans les ADN et de la base uracile dans les ARN. Cette différence, au vu de leurs formules moléculaires, semble peu suffisante pour expliquer la diversité structurale de ces deux macromolécules. Le groupement  $\text{CH}_3$  présent sur le squelette de la thymine n'a qu'un effet stérique limité et une influence électrostatique quasi nulle. De plus, il est localisé dans le grand sillon de l'ADN, et de ce fait ce groupement est peu engagé dans la structuration des acides nucléiques.

La principale différence entre l'ADN et l'ARN se résume donc quasi uniquement à la présence ou non du groupement hydroxyle OH en 2' du sucre ribose. Ce groupement joue un double rôle en étant donneur ou accepteur de liaisons hydrogène. Structuralement, il influe notamment lors de l'hydratation de l'ARN (Cheatham, et al., 1997; Egli, et al., 1996). Son rôle est essentiel pour guider le plissement du sucre (Plavec, et al., 1994; Thibaudeau, et al., 1994).

Les interactions entre les molécules d'ARN sont basées sur la reconnaissance structurale de ces molécules. Beaucoup de ces interactions font intervenir le groupement 2'OH dans un réseau de liaisons hydrogène directes ou avec l'eau de solvation. Le rôle biologique et structural de ce groupement est généralement élucidé en remplaçant celui-ci par l'analogue désoxyribose. Citons pour exemple l'élucidation du rôle du groupement 2'OH dans les propriétés catalytiques du ribozyme (Narlikar, et al., 1997; Perreault, et al., 1991), ou encore son importance primordiale dans la reconnaissance moléculaire du site d'initiation à l'encapsidation de l'ARN du virus de l'hépatite B (Schaaf, et al., 1999).

Nous avons vu, dans la première partie de ce travail de thèse, que la séquence SL1-23mer adoptait, en solution, différentes conformations. Pour comprendre l'influence possible du groupement 2'OH dans ces différentes structures, il nous a semblé intéressant d'étudier les

conformations que peut adopter l'analogue désoxyribose de SL1-23mer (cf figure I.1). Pour plus de commodité, cette dernière séquence sera notée dSL1.

C<sub>1</sub>T<sub>2</sub>T<sub>3</sub>G<sub>4</sub>C<sub>5</sub>T<sub>6</sub>G<sub>7</sub>A<sub>8</sub>A<sub>9</sub>G<sub>10</sub>C<sub>11</sub>G<sub>12</sub>C<sub>13</sub>G<sub>14</sub>C<sub>15</sub>A<sub>16</sub>C<sub>17</sub>G<sub>18</sub>G<sub>19</sub>C<sub>20</sub>A<sub>21</sub>A<sub>22</sub>G<sub>23</sub>

Figure I.1 : Séquence de l'analogue désoxyribose SL1-23mer de VIH-1<sub>Lai</sub>.

L'intérêt biologique paraît lointain lorsque l'on étudie l'ADN d'un rétrovirus, mais cette séquence est néanmoins rencontrée dans l'ADN proviral. Une recherche dans les banques de données génomiques à l'aide du logiciel BLAST (Altschul, et al., 1997) nous indique que cette séquence est largement conservée dans le génome proviral de HIV-1<sub>Lai</sub> puisque représentée à 100% dans plus de 305 bases de données génomiques de HIV (résultat au 6 août 2001). D'une façon plus générale encore, si les structures en épingle à cheveux d'ARN commencent à être bien représentées (voir l'introduction structurale) peu de choses sont connues dans le cas des ADN et aucune structure de type "kissing-complex" n'est décrite. Il nous semble intéressant de savoir si les conformations rencontrées dans le cas de SL1 sont observables pour un désoxyribose, et quelles sont les différences observées.

Comme nous pouvons le constater les interrogations sont nombreuses et la seconde partie de ce travail de thèse va s'articuler autour de ce fragment d'ADN pour en comprendre toutes ses caractéristiques structurales possibles.

## II. Le monomère dSL1

Dans le but d'étudier l'ensemble des conformations possibles adoptées par la séquence dSL1, il nous a paru intéressant de commencer, comme dans le cas de l'ARN, par caractériser la structure monomérique. Pour cela, nous avons étudié une séquence mutée en position 12 par une adénine. De cette façon, le fragment d'ADN est quasi-identique, au niveau de sa séquence, à celui utilisé pour la détermination structurale du monomère SL1 (cf figure II.1). Par analogie, nous appellerons dSL1\* cette séquence de dSL1 mutée en position 12.

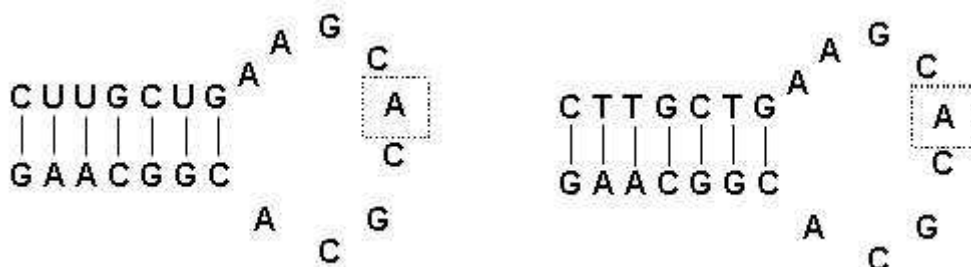


Figure II.1 : Comparaison des séquences SL1\* et dSL1\*. La structure secondaire supposée est celle d'une tige-boucle analogue à celle rencontrée précédemment.

### *1. Préparation de l'échantillon*

L'équipe de T.Huynh Dinh de l'institut Pasteur de Paris a synthétisé la molécule d'ADN par voie chimique. Elle a été purifiée par chromatographie en phase liquide et la concentration finale de l'échantillon a été déterminée par la mesure de l'absorbance à 260nm. Nous avons pu obtenir un échantillon de 17,7 mg de concentration finale de \*mM. Il est alors dialysé plusieurs fois contre de l'eau osmosée afin d'enlever les sels résiduels. Enfin, les impuretés paramagnétiques sont éliminées par traitement sur colonne échangeuse de cations (chelex 100 Bio Rad). La concentration finale est de 2.3 mM, dans une solution aqueuse de 500µL.

## 2. Etude RMN

Lors de l'étude RMN, l'échantillon a d'abord été observé dans l'eau légère. Le spectre NOESY de la région des protons iminos/iminos est représenté sur la figure II.2.

Ce spectre est à comparer avec celui obtenu pour l'ARN SL1\* (figure II.3 de la première partie). Nous remarquons que les protons iminos sont moins bien résolus que dans le cas de l'ARN. Ceci indique que les liaisons de type Watson-Crick sont absentes ou peu stables (dans le cas de dSL1\*). Afin de favoriser ces liaisons, nous avons essayé de modifier le pH sur une gamme de 7,2 à 4,8. De plus, toujours dans cette même optique, nous avons utilisé différents contre-ions pour diverses concentrations (NaCl, KCl, MgCl<sub>2</sub>, NaClO<sub>4</sub>). Il est admis que l'ajout de cations fait écran avec les charges des groupements phosphates et favorise l'appariement des molécules d'ADN. Malgré tous nos efforts, aucune de ces méthodes n'a permis d'obtenir un spectre mieux résolu que celui présenté sur la figure II.2 (gauche).

Malheureusement, l'étude dans D<sub>2</sub>O a confirmée nos craintes sur la structuration de cet échantillon. La figure II.2 (droite) représente l'agrandissement de la région des protons aromatiques/anomériques à 27°C.

Là encore, ce spectre est à comparer avec celui obtenu pour l'échantillon SL1\* (cf figure II.5 de la première partie). La comparaison est éloquente, le comportement structural est complètement différent. Nous pouvons affirmer que l'analogue désoxyribose de la séquence SL1, muté en position 12, est totalement déstructuré. Il s'apparente à un simple brin désordonné.

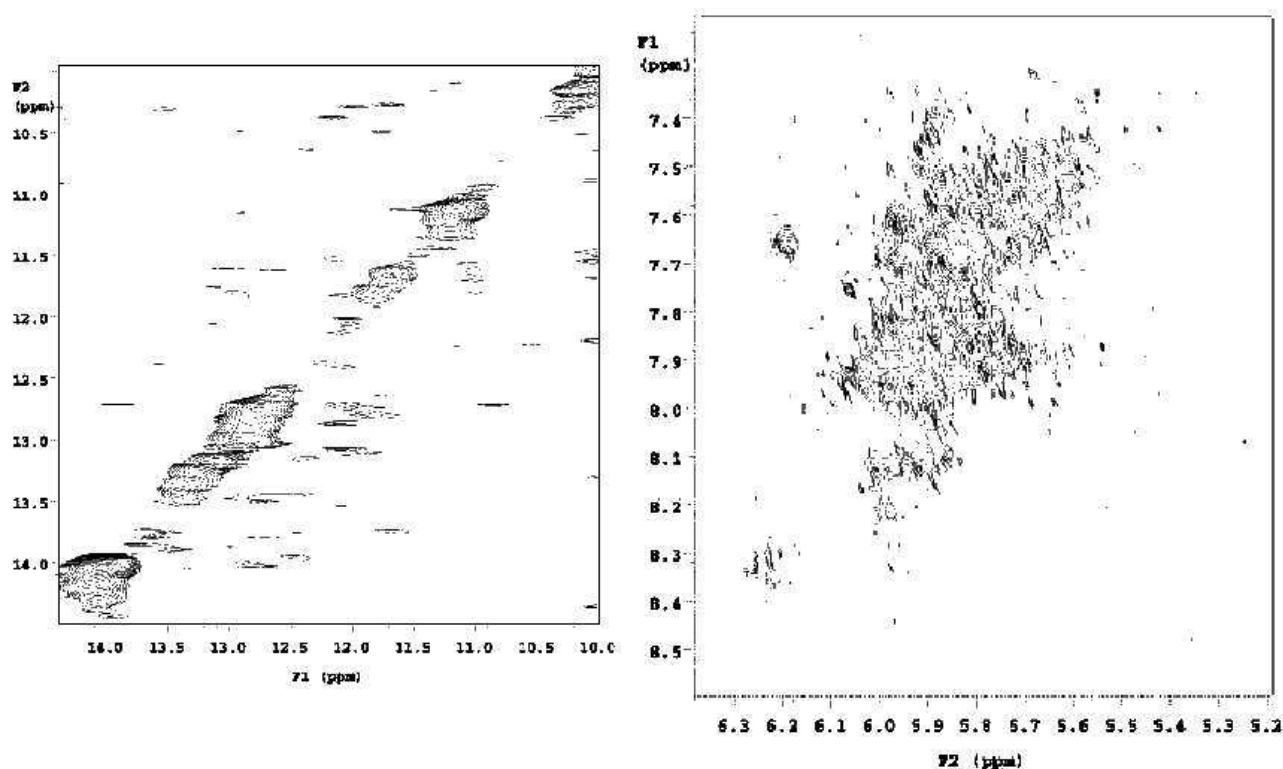


Figure II.2 : Spectre RMN de dSL1\*. A gauche, agrandissement de la région imino/imino du spectre NOESY dans l'eau légère (90%) de 120ms à 15°C. A droite, Agrandissement de la région des protons aromatiques/H1' du spectre NOESY dans D<sub>2</sub>O (T=27°C ; T<sub>m</sub>=300ms).

### 3. Conclusions

Rappelons que dans le cas du monomère SL1 (se référer au chapitre II de la première partie) la structure s'organisait autour d'une tige-boucle : la tige, composée de 7 paires de bases assure la stabilité de la large boucle. Dans l'échantillon dSL1\*, la mutation de la guanine 12 en adénine, empêche là aussi la dimérisation spontanée. Mais la tige-boucle seule n'est pas stable et est complètement déstructurée. La principale conclusion que nous pouvons en déduire est, qu'indéniablement, le groupement 2'OH joue un rôle fondamental dans la stabilisation du monomère ARN présent en aval de la dimérisation.

Une étude cristallographique d'un fragment d'ARN (r(CCCCGGG)<sub>2</sub>) à haute résolution (1,46Å) a démontré le rôle du groupement 2'OH dans l'hydratation des ARN (Egli, et al., 1996). Dans cette étude, le groupement 2'OH est impliqué dans une liaison hydrogène, localisée dans le petit sillon, avec l'eau . Ces liaisons hydrogène ne sont pas rencontrées

dans l'ADN et expliquent la différence d'hydratation de ces deux macromolécules (Cheatham, et al., 1997). Dans le cas de l'échantillon dSL1\*, l'absence de cette hydratation énergétiquement favorable est préjudiciable pour la stabilité de la tige. Or cette tige est essentielle dans la structure en épingle à cheveux.

De plus, comme nous l'avons étudié dans la première partie, l'équipe de Genest a réalisé la simulation de dynamique moléculaire avec solvant et ions explicites de la structure du monomère SL1. Cette dynamique a montré que le groupement 2'OH est engagé dans de nombreuses liaisons hydrogène intramoléculaires avec, essentiellement, les atomes O4' et O5' du résidu n+1. Ce rôle stabilisant est rencontré tout au long de la séquence, y compris dans la région de la boucle. Là encore, l'absence du groupement 2'OH dans l'échantillon dSL1\* interdit ces liaisons hydrogènes qui sont stabilisantes pour la structure en tige-boucle du monomère.

### III. Le dimère instable de dSL1

Après l'étude du monomère dSL1\*, nous nous sommes intéressés à la possibilité ou non d'obtenir un, ou deux, dimères de dSL1 et d'en décrire le ou les structures par RMN.

#### *1. Etudes physico-chimiques*

La séquence dSL1 de 23 nucléotides (cf figure I.2) a été synthétisée par voie chimique par l'équipe de Tam Huynh-Dinh de l'institut Pasteur. L'échantillon, de masse finale de 15,5mg, est dialysé plusieurs fois contre de l'eau afin d'enlever les sels résiduels. Les impuretés paramagnétiques sont éliminées par traitement sur colonne échangeuse de cations (chelex 100 Bio Rad). La concentration finale de l'acide nucléique est de 2,7mM pour un volume de 600 $\mu$ L.

Le processus de dénaturation/renaturation de dSL1 est suivi par spectroscopie UV. La courbe obtenue apporte des renseignements intéressants sur le comportement de l'ADN.

L'espèce de départ est préparée suivant un protocole analogue à celui du dimère instable de l'ARN (cf figure III.5 de la première partie) : 1,2mL d'une solution aqueuse dSL1 à 10 $\mu$ M est portée à 89°C pendant deux minutes. La solution est alors plongée rapidement dans la glace. On laisse la température s'équilibrer pendant 5 minutes. Puis, on ajoute un tampon de dimérisation, préalablement mis dans la glace, composé de TRIS et de NaCl (les concentrations finales sont : [TRIS]= 50mM ; [NaCl]=100mM et [dSL1]=10 $\mu$ M). On place alors la solution dans le spectromètre à 6°C, on laisse le système s'équilibrer pendant 10 minutes, puis on enregistre l'absorbance pendant la montée en température (0,4°C par minute). A 80°C, la solution est équilibrée pendant 5 minutes, puis la température de l'échantillon redescend à la même vitesse qu'en montée. La courbe est représentée sur la figure III.1.



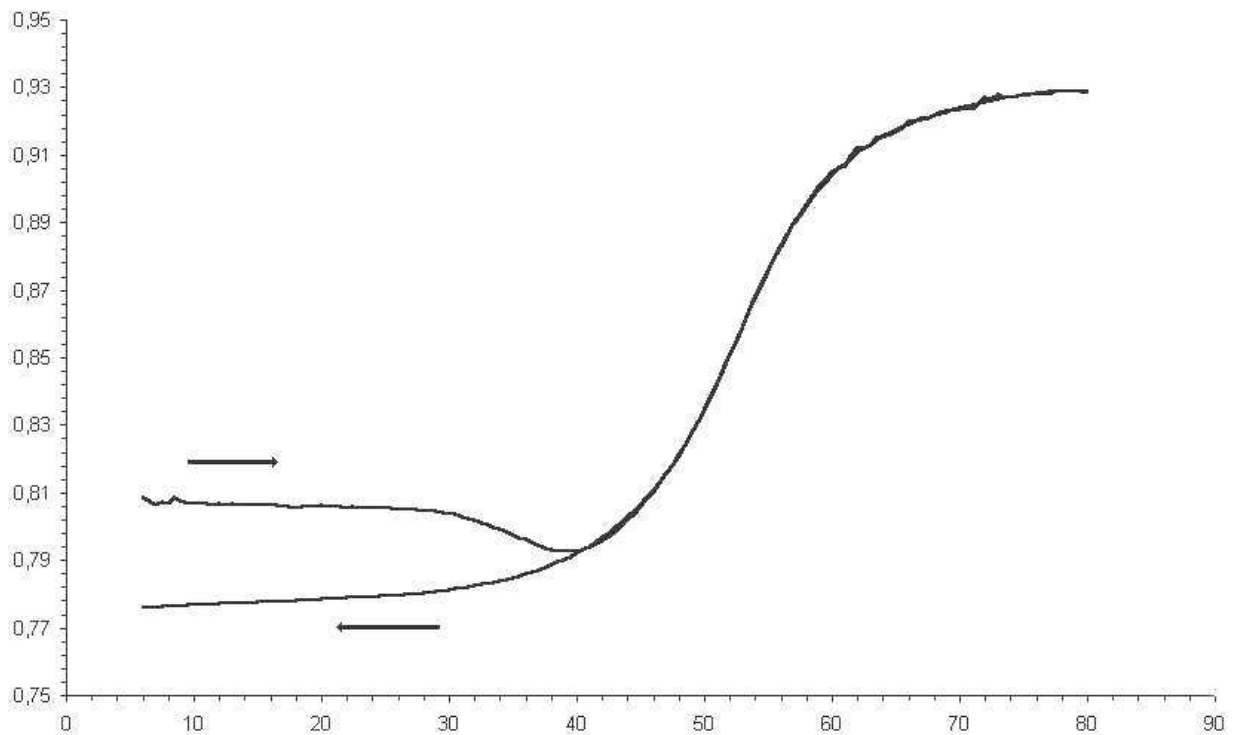


Figure III.1 : Courbe de dénaturation / renaturation de dSL1. La courbe de départ représente la montée en température, tandis que l'autre représente la descente. (Abscisse : Température en °C ; Ordonnée : Absorbance à 256nm)

On observe, sur cette courbe, deux transitions distinctes. La première, intervenant entre 30°C et 40°C ( $T_m=35^\circ\text{C}$ ), est légèrement hypochromique. La seconde transition est hyperchromique et apparaît entre 40°C et 80°C ( $T_m=55^\circ\text{C}$ ). Elle correspond à la fusion de la structure. En redescendant lentement à 6°C, la présence d'une absorbance différente, montre l'existence de deux conformations différentes de dSL1. On peut supposer que l'espèce de départ est l'espace instable, avec une température de fusion ( $T_m$ ) de 35°C, et que l'espèce finale est celle de la conformation stable.

Ce comportement est très proche de celui observé dans les mêmes conditions expérimentales pour SL1-23mer de VIH-1<sub>Lai</sub> (cf figure III.2). Cependant, rien ne nous permet d'affirmer complètement l'existence de deux types de dimères pour dSL1. La conclusion que nous apporte cette étude spectroscopique est qu'il existe deux conformations différentes et isolables de dSL1 selon son mode de préparation.

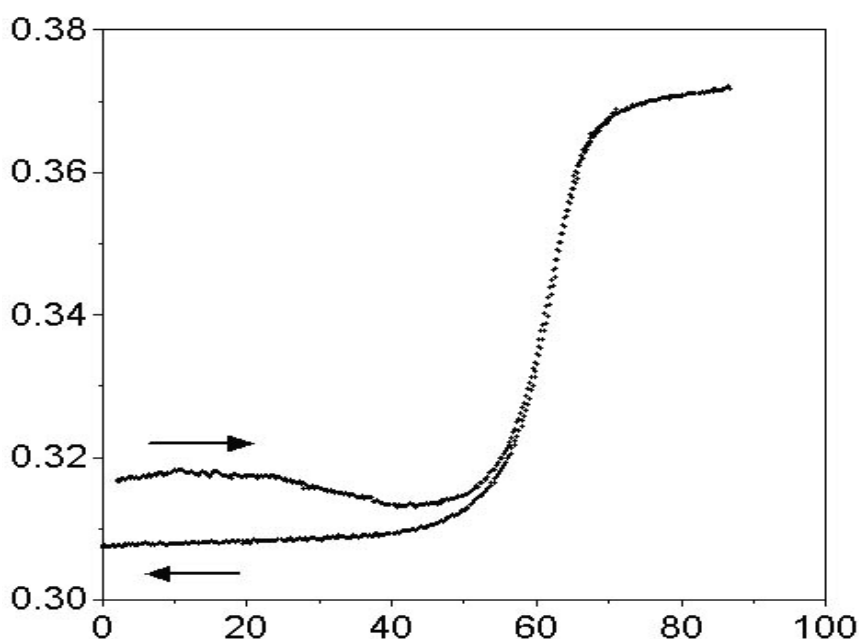


Figure III.2: Courbe de dénaturation / renaturation. D'après Theilleux-Delalande (Theilleux-Delalande, et al., 2000). Le dimère instable est préparé de la même façon que celle décrite auparavant. La conformation de départ est celle du "kissing-complex" et celle d'arrivée correspond au dimère stable de SL1-23mer de HIV-1<sub>Lai</sub>.

## 2. Etudes RMN du dimère instable dSL1

### 2.1 Préparation de l'échantillon

Nous avons vu, dans le chapitre précédent, que nous pouvions obtenir deux conformations différentes pour dSL1. Nous allons donc suivre scrupuleusement ce protocole de préparation pour l'obtention de la conformation instable. Un des points essentiel de cette préparation est d'être en solution diluée. Nous diluons donc l'échantillon dans 16 ballons comportant 10mL d'une solution aqueuse de dSL1 à environ 10 $\mu$ M. On chauffe chaque ballon à 90°C pendant 3 minutes puis on le refroidit brusquement à -20°C (glace+sel+acétone). On reconcentre alors l'échantillon en effectuant plusieurs lyophilisations. Il est important de toujours manipuler l'échantillon avec un matériel à 4° C.

L'étude RMN se fait dans une solution de 600 $\mu$ L, dégazée à l'argon, d'eau légère à 90% ou d'eau lourde à 100%.

## 2.1 Attribution dans l'eau légère

Les spectres dans l'eau ont été obtenus en utilisant la séquence d'impulsions "watergate" (Piotto, et al., 1992) avec une fenêtre spectrale de 13000Hz. La figure III.3 montre l'agrandissement de la région des protons iminos. Le point de départ de l'attribution de ces protons consiste à repérer les signaux des thymines qui sont proches de leur groupement  $\text{CH}_3$ . L'attribution complète de la tige (résidus de 1 à 7 et 17 à 23) est facilement résolue, elle indique un total appariement de celle-ci. Les protons iminos supplémentaires appartiennent aux guanines de la séquence palindromique. Par conséquent, l'échantillon étudié n'est sans doute pas le monomère tige-boucle mais un dimère ; cette hypothèse est confirmée par le calcul du temps de corrélation de la molécule dont la valeur ( $\tau_c=7,5\text{ns}$ ) est cohérente avec celle d'un dimère à  $27^\circ\text{C}$ .

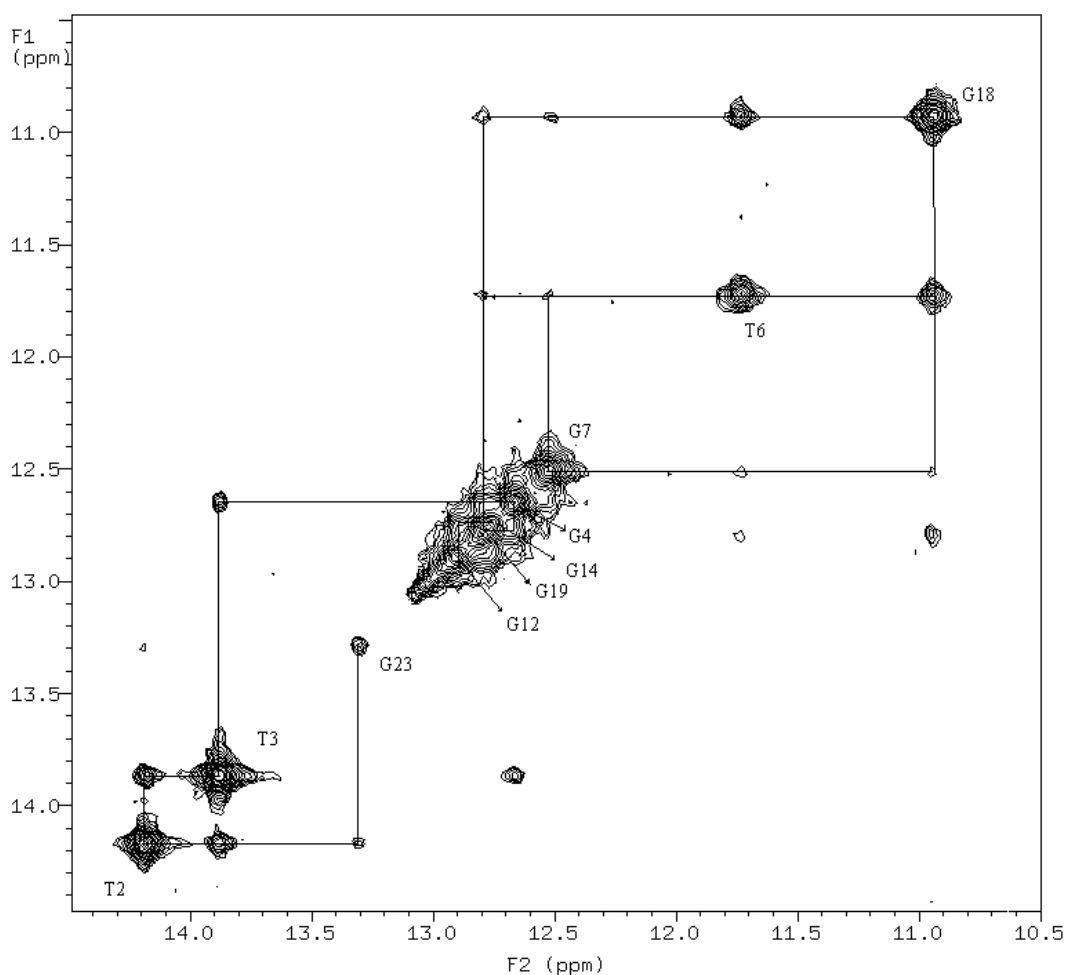


Figure III.3 : Spectre NOESY de dSL1 dans  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (90 | 10) (pH=6,8  $T_m=150\text{ms}$   $T=5^\circ\text{C}$ ). Agrandissement de la région des connectivités imino/imino.

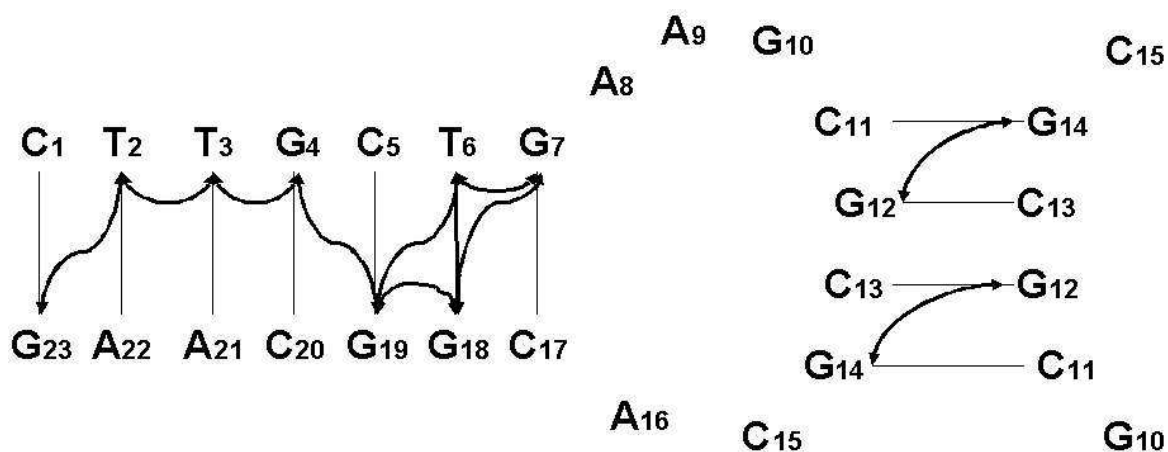


Figure III.4 : Schéma représentant les connectivités NOESY des protons imino/imino.

Les protons iminos impliqués dans la paire  $T_6G_{18}$  présentent une connectivité NOE très intense due à l'architecture de cette paire de bases. En effet, les protons iminos de ce type de paire de bases sont très proches dans l'espace (environ  $2,5\text{\AA}$ ). Un seul proton imino n'est pas observable, il s'agit du proton  $H_1$  de la guanine 10. Le schéma des corrélations NOE est représenté sur la figure III.4.

Il semblerait donc que la guanine 10 ait un comportement particulier : soit elle n'est pas engagée dans une liaison de type Watson-Crick soit son appariement est faible (plus faible, par exemple, que la paire de bases  $G_{23}C_1$  situé à l'extrémité de la tige). Les agrandissements des connectivités imino/amino (cf figure III.5) et amino/amino permettent de définir sans ambiguïté les appariements Watson-Crick respectifs.

Les connectivités iminos des guanines corrélées aux protons aminos des cytosines sont observables à  $5^\circ\text{C}$  et à  $15^\circ\text{C}$  pour la paire de bases  $G_7C_{17}$ . En revanche, ces corrélations ne sont observables qu'à  $5^\circ\text{C}$  pour la paire de bases  $C_1G_{23}$ . L'appariement de la paire de bases située à l'extrémité de la tige est plus faible, car soumis à la dynamique des deux brins. Les protons aminos des cytosines sont repérés à l'aide de leurs proximités avec leurs protons  $H_5$ , eux même déterminés dans  $D_2O$ . Aucun proton amino n'est identifié comme appartenant à la cytosine 15. Si l'on rapproche cette dernière information au fait que la résonance de la guanine 10 n'est pas observable, nous pouvons conclure que les bases  $G_{10}$  et  $C_{15}$  ne sont pas appariées.

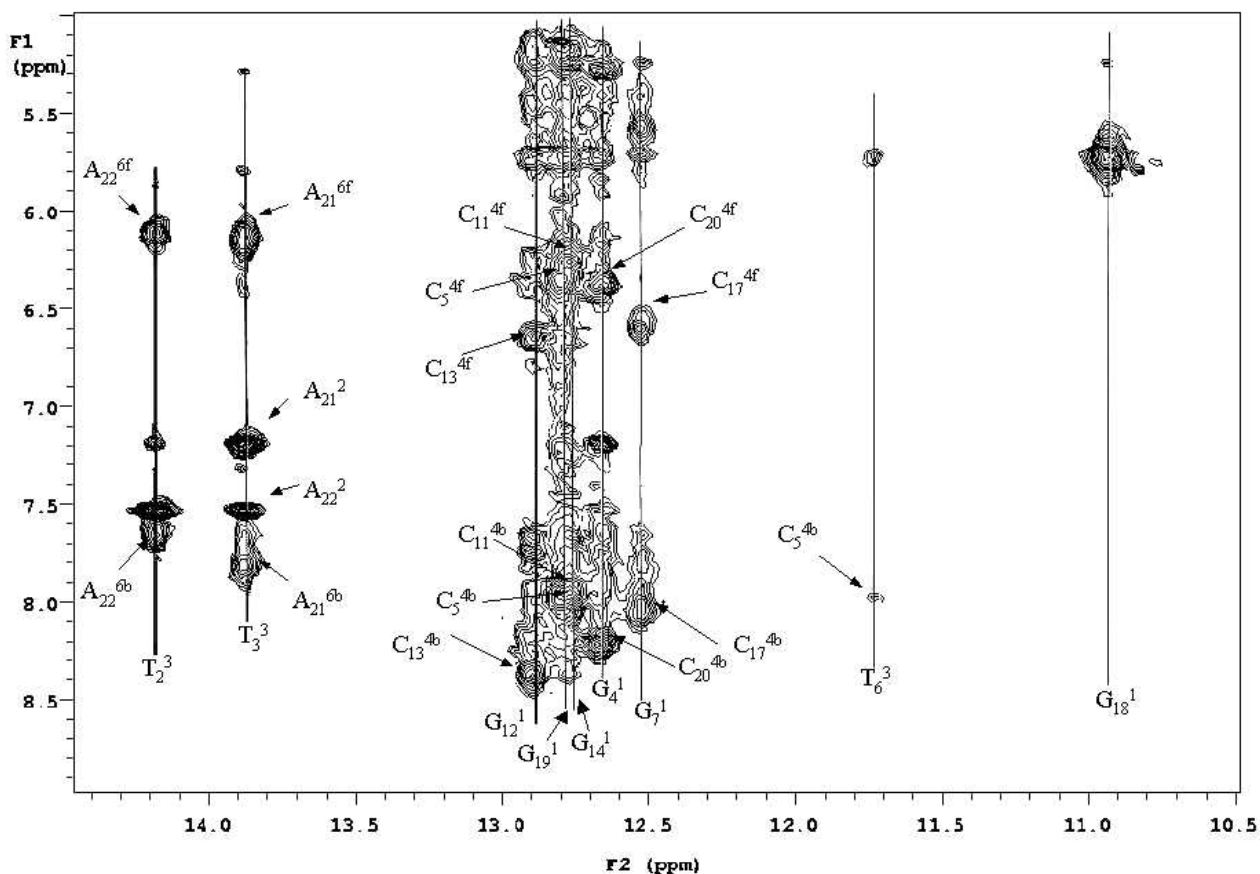


Figure III.5: Agrandissement du spectre NOESY de la région des connectivités imino/amino (pH=6,8 ; Tm=150ms ; T=15°C ; H<sub>2</sub>O/D<sub>2</sub>O 90 | 10).

L'observation sur le spectre NOESY de la région des protons amino/amino (cf figure III.6) montre une corrélation entre le proton H8 de l'adénine 8 et un groupement amino de A<sub>8</sub>, A<sub>9</sub> ou A<sub>16</sub>. Cette corrélation, d'intensité moyenne, nous suggère l'existence d'une paire de bases adénine-adénine. En effet, la distance entre le proton H8 et le proton NH2 d'une même adénine est de l'ordre de 4,9Å et n'est pas détectable pour les adénines A<sub>22</sub> et A<sub>21</sub> dont les protons aminos sont repérés par les paires de bases AT qu'elles forment. En revanche, lors d'une paire de bases symétrique adénine-adénine de type II (se référer à l'annexe pour les types d'appariements canoniques et non canoniques possibles) ces protons ne sont séparés que de 3,4Å, ce qui correspond aux distances observables pour cette intensité (cf figure III.7).

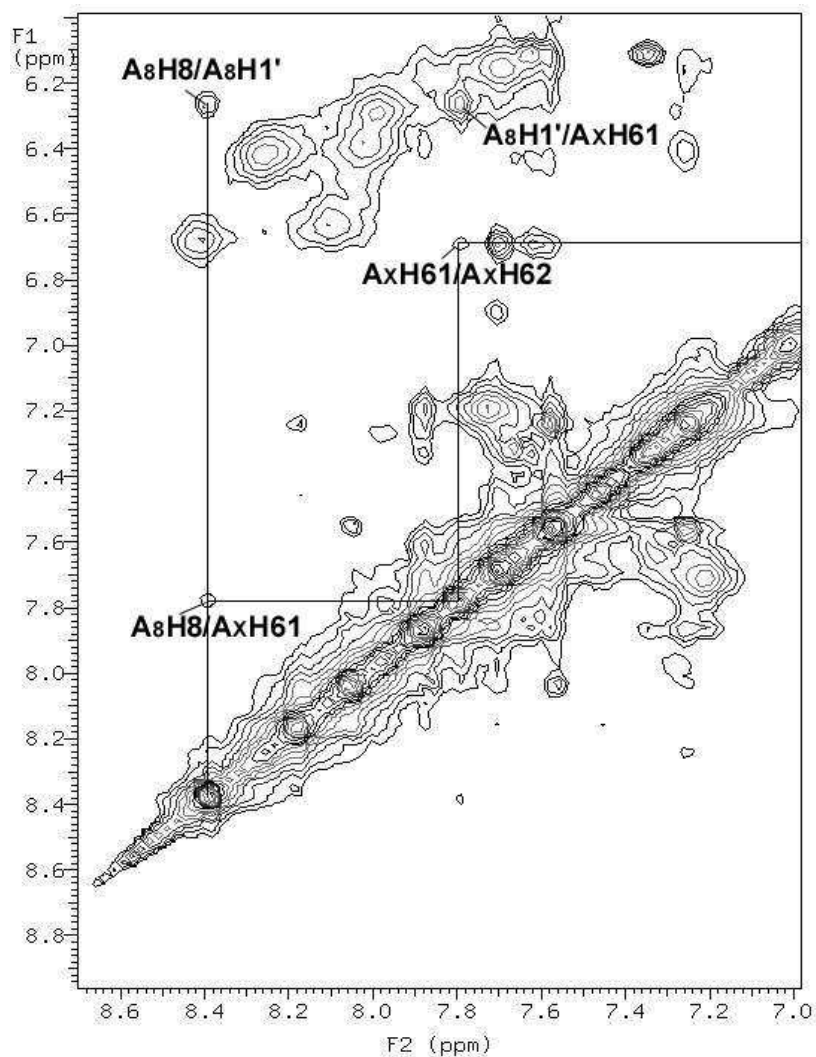


Figure III.6 : Spectre NOESY dans H<sub>2</sub>O à 15°C de la région amino/amino avec un temps de mélange de 100ms. Les connectivités impliquant la proximité du proton H8 de l'adénine 8 et les protons NH<sub>2</sub> d'une autre adénine (x) sont indiquées sur la figure.

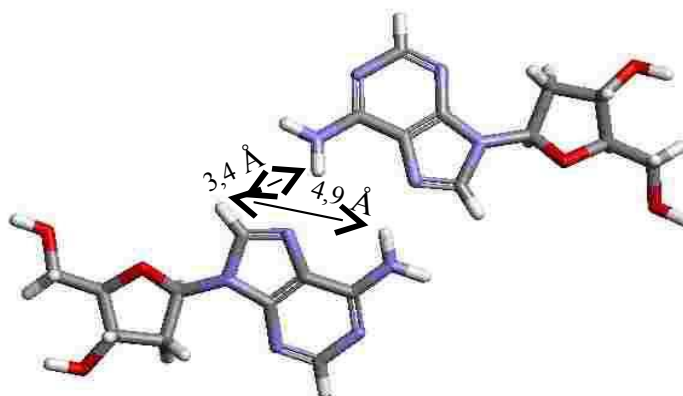


Figure III.7 : Schéma d'un appariement adénine/adénine. Les distances montrent les proximités éventuelles inter ou intra résidus entre un proton H8 et un proton amino.

Malheureusement, rien ne nous permet de savoir si l'amino appartient à l'adénine 8, 9 ou 16. Cette contrainte ne sera donc pas imposée directement dans la modélisation moléculaire.

### 2.3 Attribution dans l'eau lourde

L'échantillon dSL1 utilisé pour l'étude dans l'eau est ensuite lyophilisé deux fois contre D<sub>2</sub>O avant d'être repris dans 600µL de D<sub>2</sub>O et scellé sous atmosphère d'argon. Là encore, il est nécessaire de prendre des précautions en travaillant avec du matériel "froid" (mis au réfrigérateur) pour empêcher la transition entre les deux dimères.

Les expériences NOESY ont été réalisées à 27°C pour différents temps de mélange : 0, 50, 90, 150 et 300ms. Elles ont été effectuées avec une accumulation de 32 scans pour chaque incréments t1. Un délai de 2 secondes a été utilisé en début de séquence pour optimiser le rapport signal/bruit de l'ensemble des protons. Une présaturation à basse puissance permet d'éliminer le signal d'eau résiduelle sur les spectres. La figure III.8 représente l'agrandissement de la région des connectivités aromatique/H1'. Presque toutes les connectivités sont bien résolues et ne présentent que rarement des recouvrements. La relativement faible largeur de raie à mi-hauteur et la dispersion en déplacement chimique observée tout au long de la séquence traduisent une structuration importante du dimère.

Le point de départ de l'attribution, consiste à repérer la position des CH<sub>3</sub> des trois thymines et des sept protons H5 des cytosines. La position de ces résonances est révélée par leurs connectivités scalaires avec leurs protons H6 sur les spectres COSY-DQF et TOCSY à la même température. Sur le spectre NOESY au temps de mélange de 300ms, toutes les corrélations intra et inter-résidus des protons H6 ou H8 avec les protons H1' sont observables. Il en est de même pour les protons H6 ou H8 avec les protons H2' et H2''. Ceci permet une attribution séquentielle avec peu d'ambiguïtés à cette température.

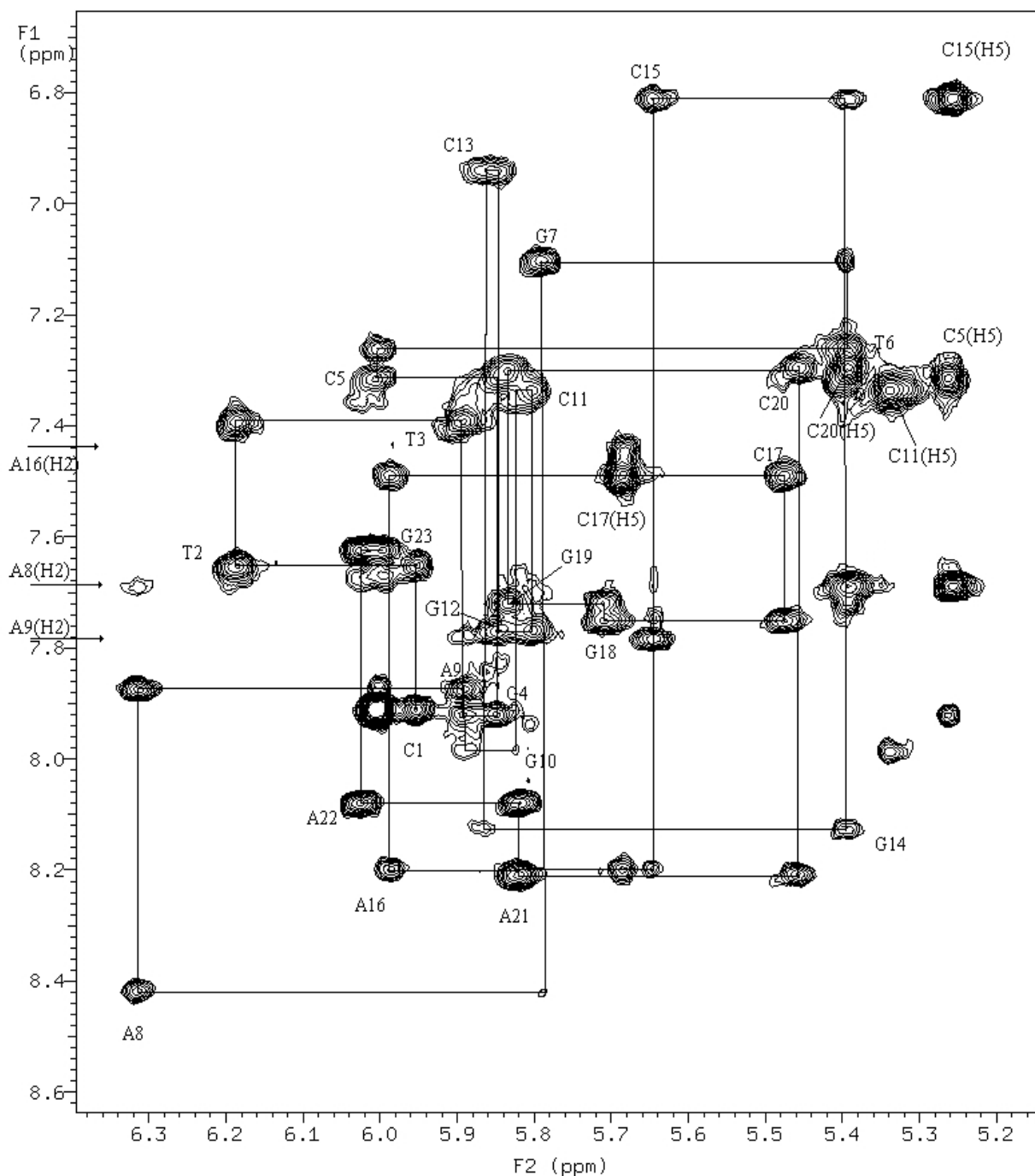


Figure III.8 : Spectre NOESY à 150ms du dimère instable dSL1. Région des connectivités aromatique/H1'. Les flèches indiquent la position des protons H2 des adénines 16,9 et 8.

L'attribution des cinq protons H2 des adénines a été réalisée en utilisant leurs proximités relatives (de 4 à 5Å) avec leurs protons anomériques. Les corrélations NOE des protons H2 de A<sub>8</sub> montre une proximité spatiale importante avec les protons : H1' de G<sub>14</sub>; H5 et H6 de C<sub>15</sub>. De plus, le proton H2 de A<sub>9</sub> se trouve proche des protons H1' et H6 de C<sub>15</sub>. Ces connectivités indiquent une organisation tridimensionnelle spéciale à cet endroit de la



séquence. Enfin, le fort déplacement chimique du proton H8 de l'adénine 8 (~ 8,42ppm) indique que son cycle aromatique n'est sans doute pas empilé comme dans le cas d'une hélice de type B.

La position de la base C<sub>15</sub> semble, elle aussi, particulière. En effet, en plus des corrélations déjà citées avec les bases 8 et 9, nous n'observons pas de proximité du proton H5 de C<sub>15</sub> avec le proton H8 de la guanine 14. Cette dernière donnée confirme le fait que cette base n'est pas appariée classiquement. Le tableau III.1 résume l'ensemble des déplacements chimiques des protons de dSL1.

|            | H <sub>8</sub> /H <sub>6</sub> | H <sub>5</sub> /H <sub>2</sub> /CH <sub>3</sub> | H <sub>1</sub> ' | H <sub>2</sub> ' | H <sub>2</sub> '' | H <sub>3</sub> ' | H <sub>4</sub> ' | H <sub>5</sub> ' | H <sub>5</sub> '' | Imino | NH <sub>2</sub> lié | NH <sub>2</sub> libre |
|------------|--------------------------------|---|------------------|------------------|-------------------|------------------|------------------|------------------|-------------------|-------|---------------------|-----------------------|
| <b>C1</b>  | 7,91                           | 5,99  | 5,94             | 2,63             | 2,26              | 4,68             | 4,14             | 3,86             | 3,81              |       | 7,23                | 5,94                  |
| <b>T2</b>  | 7,64                           | 1,52  | 6,17             | 2,61             | 2,25              | 4,90             | 4,26             | 4,12             | 4,08              | 14,20 |                     |                       |
| <b>T3</b>  | 7,39                           | 1,52  | 5,88             | 2,51             | 2,14              | 4,90             | 4,37             | 4,17             | 4,14              | 13,90 |                     |                       |
| <b>G4</b>  | 7,92                           |   | 5,84             | 2,53             | 2,48              | 4,91             | 4,38             | 4,22             | 4,10              | 12,67 |                     |                       |
| <b>C5</b>  | 7,32                           | 5,24  | 6,00             | 2,45             | 1,91              | 4,98             | 4,37             |                  |                   |       | 8,04                | 6,42                  |
| <b>T6</b>  | 7,26                           | 1,54  | 5,38             | 1,91             | 1,64              | 4,22             |                  | 3,98             | 3,88              | 11,75 |                     |                       |
| <b>G7</b>  | 7,10                           |   | 5,78             | 1,29             | 1,20              | 4,64             | 4,19             | 3,88             |                   | 12,55 |                     |                       |
| <b>A8</b>  | 8,42                           | 7,69  | 6,31             | 3,22             | 2,97              | 5,06             | 4,36             |                  |                   |       | 7,80                | 6,69                  |
| <b>A9</b>  | 7,87                           | 7,78  | 5,88             | 2,53             | 1,51              | 4,86             | 4,30             | 4,08             | 4,04              |       |                     |                       |
| <b>G10</b> | 7,98                           |   | 5,81             | 2,62             | 2,53              | 4,89             | 4,28             | 4,06             | 4,02              |       |                     |                       |
| <b>C11</b> | 7,33                           | 5,32  | 5,79             | 2,09             | 2,29              | 5,31             | 4,28             |                  |                   |       | 8,02                | 6,33                  |
| <b>G12</b> | 7,77                           |   | 5,84             | 2,62             | 2,43              | 4,84             | 4,46             |                  |                   | 12,85 |                     |                       |
| <b>C13</b> | 6,94                           | 5,04  | 5,85             | 2,03             | 1,46              | 4,90             | 4,14             | 4,09             | 4,04              |       | 8,44                | 6,73                  |
| <b>G14</b> | 8,12                           |   | 5,38             | 2,82             | 2,76              | 4,46             |                  |                  |                   | 12,78 |                     |                       |
| <b>C15</b> | 6,81                           | 5,24  | 5,64             | 1,79             | 0,43              | 4,56             | 4,19             | 3,98             | 3,94              |       |                     |                       |
| <b>A16</b> | 8,19                           | 7,46  | 5,98             | 2,76             | 2,65              | 4,93             | 4,57             | 4,30             |                   |       |                     |                       |
| <b>C17</b> | 7,49                           | 5,67  | 5,46             | 2,34             | 2,12              | 4,94             | 4,30             | 4,12             | 4,03              |       | 8,11                | 6,67                  |
| <b>G18</b> | 7,75                           |   | 5,69             | 2,74             | 2,55              | 4,94             | 4,32             |                  |                   | 10,96 |                     |                       |
| <b>G19</b> | 7,72                           |   | 5,81             | 2,69             | 2,48              | 4,94             | 4,33             |                  |                   | 12,80 |                     |                       |
| <b>C20</b> | 7,30                           | 5,38  | 5,44             | 2,23             | 1,86              | 4,93             | 4,44             |                  |                   |       | 8,29                | 6,45                  |
| <b>A21</b> | 8,21                           | 7,37  | 5,81             | 2,83             | 2,70              | 5,01             | 4,34             | 4,08             | 3,97              |       | 7,89                | 6,19                  |
| <b>A22</b> | 8,08                           | 7,66  | 6,02             | 2,82             | 2,57              | 5,01             | 4,40             | 4,19             | 4,10              |       | 7,71                | 6,18                  |
| <b>G23</b> | 7,62                           |   | 5,98             | 2,35             | 2,25              | 5,01             | 4,79             |                  |                   | 13,32 |                     |                       |

Tableau III.1 : Déplacements chimiques observés dans le dimère instable dSL1.

Le plissement des unités désoxyribose a été évalué à l'aide du spectre TOCSY. La présence continue des corrélations H<sub>1</sub>'/H<sub>2</sub>' et H<sub>1</sub>'/H<sub>2</sub>'' indique une conformation sud de tous les résidus. De plus, la mesure (lorsqu'elle est possible) des constantes de couplage sur le spectre COSY-DQF entre ces derniers protons donne des valeurs typiques de

conformation C2'-endo.

Enfin, la faible proximité générale intra-résidus des protons H8/6 - H1' en regard de la forte intensité des protons H8/6 - H2' indique une position de la base par rapport au sucre en anti (cf figure III.9).

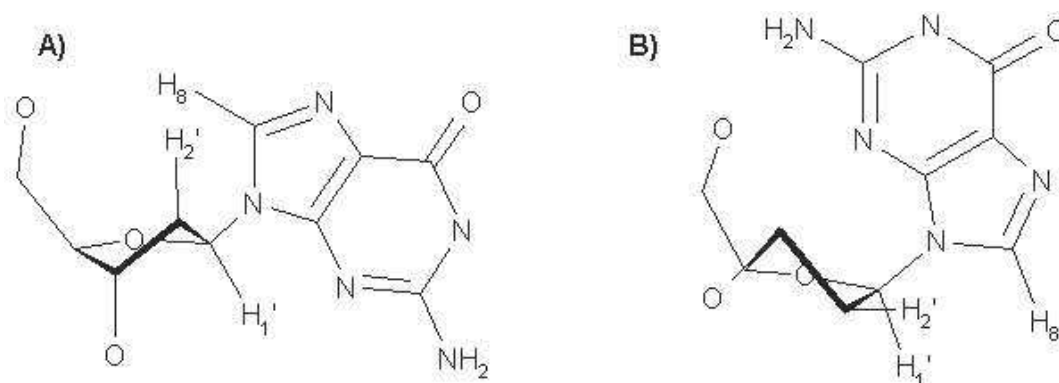


Figure III.9 : Conformation anti (A) et syn (B) de la guanosine. La proximité du proton H1' avec le proton H8 est de l'ordre de 3,6Å en conformation anti (A) pour environ 2,2Å en conformation syn (B).

#### 2.4 Attribution du squelette phosphodiester

Le spectre phosphore de dSL1 (cf figure III.10) a été effectué à la même température (27°C). Il montre un massif comportant la majorité des résonances du  $^{31}\text{P}$  concentrées dans une étroite fenêtre de déplacement chimique (environ 1ppm) et cinq raies qui s'écartent plus ou moins de ce massif. Nous pouvons en déduire que la majorité des phosphates ont la même conformation. Cependant, il faut élucider à quel phosphore appartiennent les raies de résonance qui s'éloignent du massif.

La quasi totalité des résonances des protons H3' et H4' étant acquise, il nous paraissait intéressant d'effectuer un spectre HSQC  $^{31}\text{P}$ - $^1\text{H}$  pour localiser les résonances des phosphates et élucider la conformation du squelette phosphodiester. La figure III.11 représente donc le spectre HSQC à 27°C de dSL1. La localisation des différents protons H3' et H4' a permis d'attribuer ces résonances.

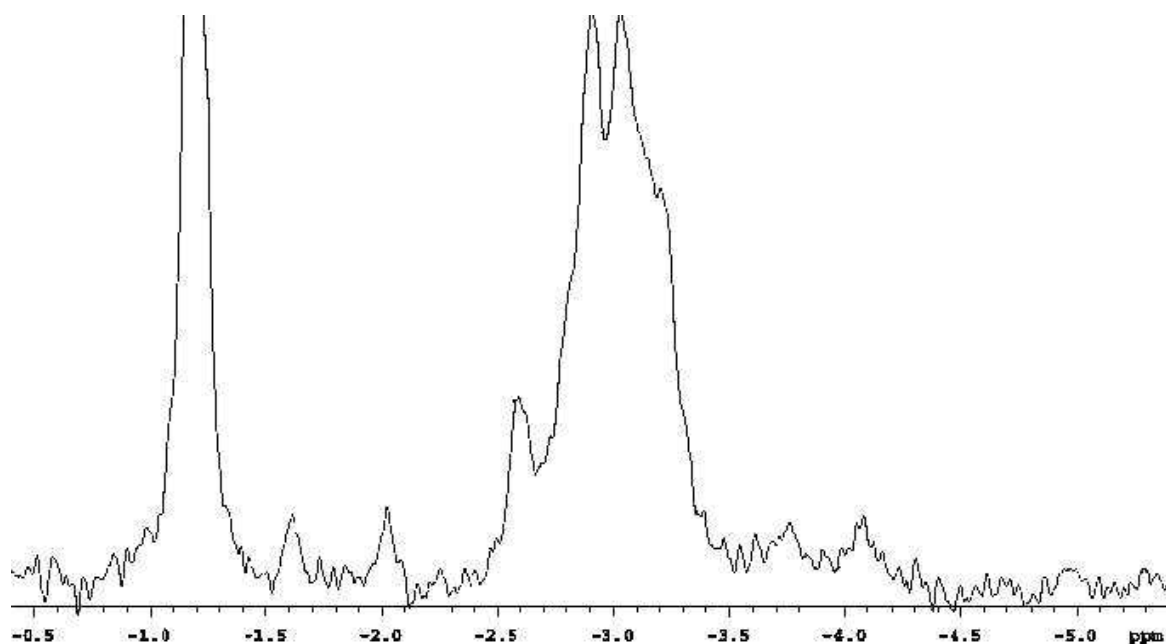


Figure III.10 : Spectre  $^{31}\text{P}$  du dimère instable dSL1 à 27°C. La raie la plus à gauche correspond au tampon phosphate.

Deux des raies situées hors du massif correspondent aux résidus de jonction ( $\text{G}_7\text{-P-A}_8$ ;  $\text{A}_8\text{-P-A}_9$ ). La distorsion du squelette phosphodiester de ces résidus semble d'ailleurs logique dans le cas d'un dimère boucle-boucle. Les deux autres raies situées hors du massif correspondent aux phosphores localisés entre les résidus 13 et 14 et entre les résidus 14 et 15. Cette dernière information confirme que les bases  $\text{G}_{10}$  et  $\text{C}_{15}$  ne sont pas appariées et montre que celles-ci participent à la jonction des tiges. Enfin, le phosphore situé entre le résidu 16 et le résidu 17 a un déplacement chimique compris dans le massif. Nous supposons donc que ce phosphore comporte les mêmes caractéristiques de conformation que ceux des tiges. La figure III.12 résume les informations structurales issues de ce spectre.

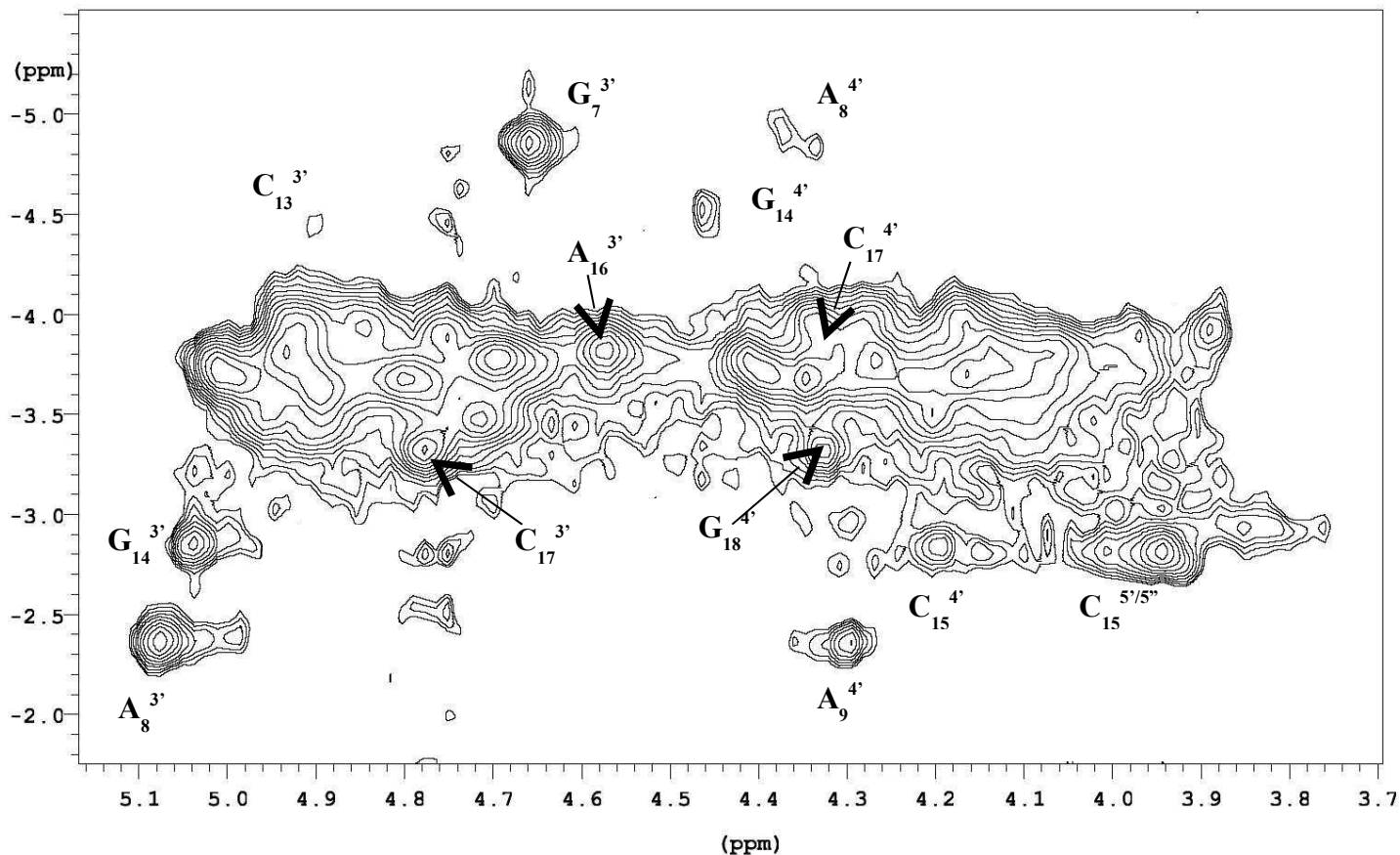


Figure III.11 : Spectre HSQC  $^{31}\text{P}$ - $^1\text{H}$  du dimère instable dSL1 à 27°C. Les résonances des phosphores (ordonnée) sont déterminées par leurs corrélations avec leur proton H3' et le proton H4' du résidu i+1.

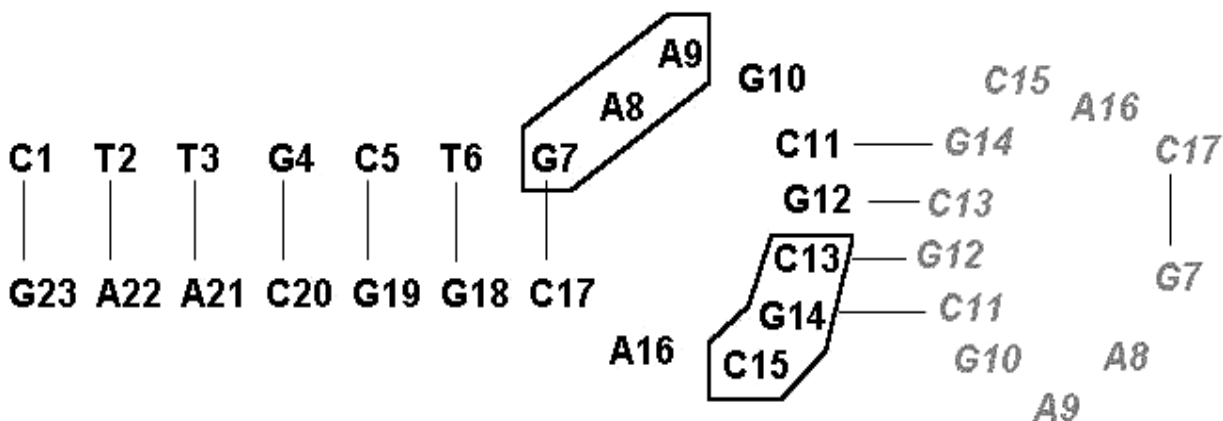


Figure III.12 : Résumé des informations structurales données par les spectres phosphores. L'ensemble du squelette phosphodiester adopte une conformation classique d'hélice B à l'exception des résidus encadrés.

### 3. Modélisation moléculaire

#### 3.1 Etablissement des contraintes

Les spectres TOCSY et COSY-DQF nous renseignent sur la conformation des unités désoxy-riboses des résidus. La mesure des constantes de couplages H1'-H2' et H1'-H2'' n'est pas toujours possible du fait du recouvrement. Néanmoins, l'observation de ces corrélations sur le spectre TOCSY pour tous les résidus montre qu'ils sont en conformation sud. Trois angles dièdres ( $\nu_1$ ,  $\nu_2$ , et  $\nu_3$ ) permettent de contraindre le plissement des cycles riboses; seuls deux angles sont nécessaires mais l'ajout d'un troisième angle permet une convergence plus rapide. Ces valeurs d'angles, tirées de l'ouvrage de Saenger (Saenger, 1984), sont imposées avec une incertitude de 10°.

Nous avons vu, sur les spectres phosphore 1D et HSQC  $^{31}\text{P}$ - $^1\text{H}$ , que la majorité du squelette phosphodiester apparaît sous la même conformation, excepté pour les résidus 8,9,10,15 et 16. De plus, nous savons que toutes les conformations des sucres sont en C2'-endo et que la position des bases par rapport aux sucres est en anti. Ceci indique une conformation d'hélice de type B pour l'ensemble des résidus, excepté les résidus 8,9,10,15 et 16 qui sont laissés libres de contraintes. Dès lors, cinq angles dièdres ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\epsilon$  et  $\zeta$ ) sont contraints à des valeurs classiques d'ADN B. Ces valeurs sont tirées de l'ouvrage de Saenger (Saenger, 1984) et sont imposées avec une incertitude de 15°.

Les corrélations NOE issues des études dans D<sub>2</sub>O ont été intégrées à plusieurs temps de mélange (0, 50, 90, 150 et 300ms). Là encore, l'utilisation du programme RELAZ (Lancelot, et al., 1989) permet de traduire les effets NOE en distance tout en tenant compte de la diffusion de spin. Les distances sont introduites dans un fichier de contraintes avec une incertitude de 15% autour de la valeur de distance calculée. Cependant, l'étude de dimère symétrique pose un problème sur l'établissement des contraintes inter-brins: s'agit-il de corrélation intra ou inter monomères? L'utilisation de la fonction "SUM" (Nilges, 1993), présente dans les logiciels CNS et XPLOR, permet d'imposer les contraintes dont on ne peut décider *a priori* leur caractère intra ou inter monomère. Nous l'avons donc employé

avec succès pour l'ensemble de nos données inter-brins ambigus.

Les NOE observés dans l'eau légère sont introduits de deux façons différentes. Ceux qui permettent de définir l'appariement de type Watson-Crick entre les bases sont introduits sous forme de contraintes de deux distances pour chaque liaison hydrogène (par exemple 6 contraintes sont définies pour une paire GC tandis que 4 contraintes de distances définissent une paire AT). Les autres NOE sont introduits sous forme d'un intervalle d'inclusion ( $1,8 < d < 5,5 \text{ \AA}$ ) pour s'affranchir des vitesses d'échanges avec le solvant. La figure III.13 schématise les contraintes de distances introduites pour la région de l'interaction boucle-boucle et l'entrée de tige, les contraintes inter-brin ont été déduites de l'étude. Au tout début de la modélisation moléculaire, aucune contrainte n'est établie pour rendre compte de l'appariement adénine/adénine suggérée par l'étude dans H<sub>2</sub>O. Enfin le tableau III.2 résume l'ensemble des contraintes et leurs nombres.

Pour finir, et seulement dans les premiers cycles de modélisation, 18 contraintes de plans sont créées. Elles servent à maintenir les bases appartenant à un même appariement de type Watson-Crick dans le même plan ( $C_1G_{23}$ ,  $T_2A_{22}$ ,  $T_3A_{21}$ ,  $G_4C_{20}$ ,  $C_5G_{19}$ ,  $T_6G_{18}$ ,  $G_7C_{17}$ ,  $C_{11}G_{14}$ ,  $G_{12}C_{13}$  et leurs symétriques). Ces contraintes sont réellement utiles pour obtenir une convergence plus rapide de la structure globale mais doivent être abandonnées dans les cycles de modélisation suivants, car elles ne sont pas réellement issues de nos données expérimentales. La constante de force de cette contrainte est de  $10 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ , à titre de comparaison la constante de force d'un effet NOE est de  $50 \text{ kcal.mol}^{-1}.\text{\AA}^{-2}$ .

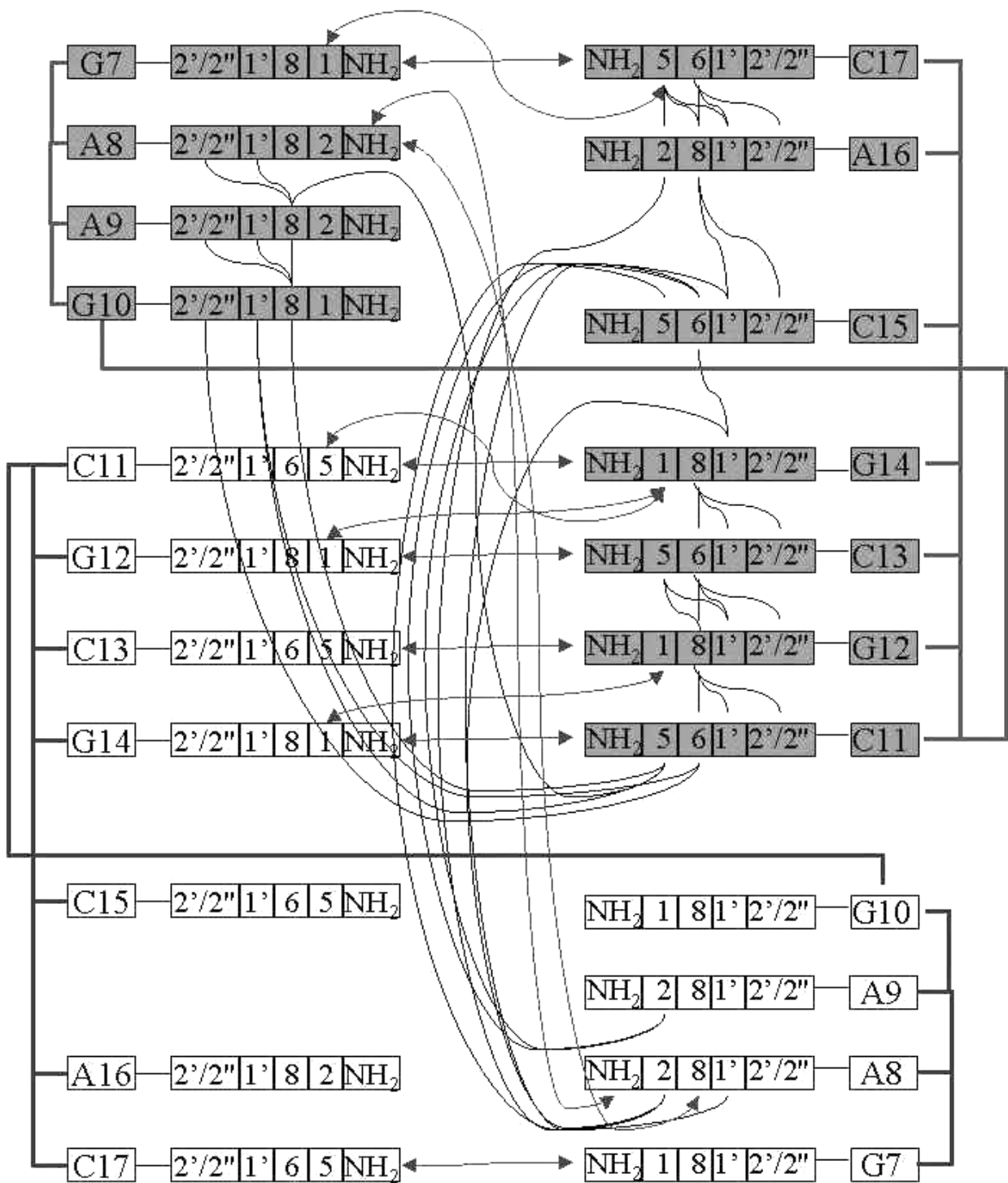


Figure III.13 : Schéma des différentes contraintes de distances introduites dans la modélisation moléculaire de dSL1 dimère instable. Les deux brins sont différenciés par une couleur plus ou moins foncée. Dans un souci de clarté, les contraintes ne sont portées que pour un seul des monomères.

| <u>Contraintes de distances</u>                                       |      |
|---|------|
| Issues de l'étude dans D <sub>2</sub> O                               |      |
| intra-résidu  | 172  |
| inter-résidu  | 322  |
| Issues de l'étude dans H <sub>2</sub> O                               |      |
| contraintes Watson-Crick  | 102  |
| protons échangeables  | 88   |
| Total contraintes de distances  | 684  |
| NOE/résidu  | 14,9 |
| <u>Contraintes d'angles</u>   |      |
| Plissement du sucre (v1, v2 et v3)                                    | 138  |
| backbone ( $\alpha$ , $\beta$ , $\gamma$ , $\varepsilon$ et $\zeta$ ) | 120  |
| Total contraintes d'angles  | 258  |
| Angles/résidu   | 5,6  |
| Total contraintes   | 942  |
| Contraintes/résidu  | 20,5 |

Tableau III.2 : résumé des différentes contraintes imposées dans la modélisation moléculaire de dSL1 en dimère instable.

### 3.2 Résolution de la structure tridimensionnelle

Comme dans le cas de l'ARN, il n'est pas possible à la simple vue des corrélations NOE, de discriminer une structure secondaire d'une autre. De ce fait, les deux types possibles de dimères (cf figure III.14) sont tentés sans aucun *a priori*.



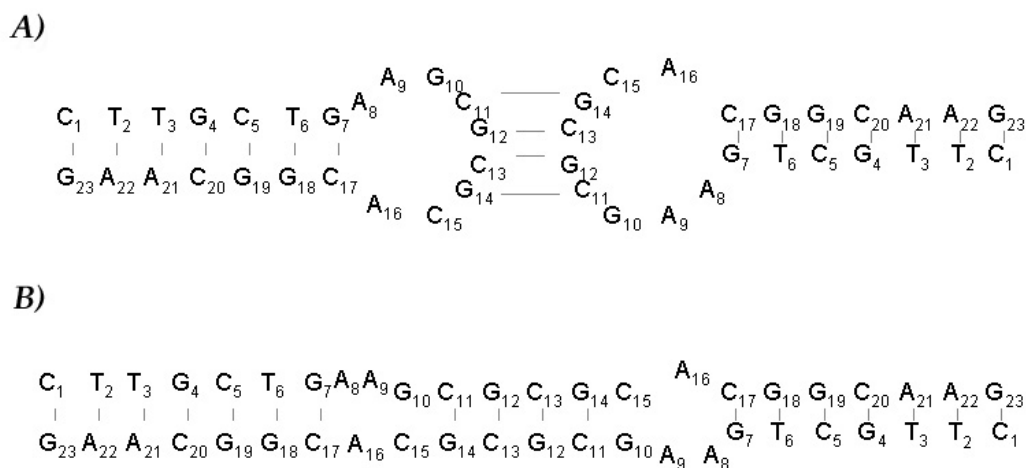


Figure III.14 : Les deux types de structures secondaires essayés lors de la modélisation moléculaire.

Une centaine de structures aléatoires respectant la covalence des liaisons est créée à l'aide du logiciel CNS version 1.0 (Brunger, et al., 1998). Ces structures servent de point de départ pour la modélisation moléculaire des deux modèles de dimère.

Dans les deux cas, les structures aléatoires sont soumises à un recuit simulé dans l'espace des angles dièdres à l'aide du même logiciel. 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 uniquement 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 par 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 conformationnelle. La constante de force de van der Waals est maintenue à seulement  $0,1\text{kcal.mol}^{-1}.\text{\AA}^{-2}$  pour favoriser les réarrangements et l'exploration de la macromolécule. Les constantes de forces expérimentales des NOE et des angles dièdres sont respectivement à  $50\text{kcal.mol}^{-1}.\text{\AA}^{-2}$  et  $5\text{kcal.mol}^{-1}.\text{rad}^{-2}$ . Le refroidissement du système, de 20000 à 0K, se réalise à l'aide d'une dynamique de 60ps dans l'espace des angles de torsions. La constante de force de la fonction répulsive de van

der Waals évolue alors progressivement de 0,1 à 1kcal.mol<sup>-1</sup>.Å<sup>-2</sup> tandis que celles des NOE et des angles dièdres sont à 50kcal.mol<sup>-1</sup>. Comme cela est souvent le cas lors de modélisation d'acides nucléiques ou de grandes protéines, il est nécessaire, pour obtenir une convergence plus rapide des calculs, de faire un second refroidissement dans l'espace cartésien. Celui-ci correspond à une dynamique de 10ps où la constante de force de van der Waals augmente de 1 à 4kcal.mol<sup>-1</sup>.Å<sup>-2</sup>. Enfin, le protocole est suivi d'une étape de mécanique moléculaire de 400pas avec l'algorithme de Powell. La fonction d'énergie potentielle employée englobe tous les paramètres précédemment cités, avec une fonction de Lennard-Jones pour rendre compte du caractère attractif du potentiel de van der Waals et la fonction d'énergie électrostatique. Dans le logiciel CNS la fonction d'énergie potentielle de liaison hydrogène, présent dans le logiciel XPLOR, est incluse dans le potentiel de van der Waals.

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 le cas de la modélisation sous forme de dimère stable, c'est à dire sous forme de duplexe étendu, aucune convergence n'est observée. Le nombre de violations NOE supérieures à 1,0Å est important dans la région des jonctions de tiges et les énergies totales sont largement supérieures à celles observées pour le dimère boucle-boucle. Cette hypothèse est donc abandonnée, le dimère instable est donc une structure de type kissing-complex.

Dans ce dernier cas, une information intéressante apparaît dans les premiers cycles de recuit simulé : nous observons une proximité entre les adénines 8 de chaque brin qui nous laisse supposer une possible paire AA. En effet, la superposition de toutes les structures observant un minimum de violations NOE inférieure à 0,5Å (de 2 à 10 violations) montre ce positionnement de l'adénine 8 (cf figure III.15).

Cet appariement avait été supposé par les expériences dans l'eau légère mais il n'avait pas été contraint. La position de ces adénines indique qu'une paire de bases inter-brin A<sub>8</sub>A<sub>8</sub> est possible. Des essais pour obtenir une paire de bases A<sub>8</sub>A<sub>9</sub> et A<sub>8</sub>A<sub>16</sub> intra et inter-brin sont tentés. Dans tous les cas, nous observons de fortes violations de distances NOE

ce qui n'est pas le cas de l'appariement  $A_8A_8$ . Dès lors, nous avons contraint cette paire de bases à l'aide de deux distances NOE supplémentaires.

Un ensemble de 17 structures convergentes est obtenu. Ces structures ont subi une sélection supplémentaire par le calcul en retour de leurs déplacements chimiques. Ce protocole sera expliqué dans le chapitre suivant. Ces structures sont superposées sur chaque région dans la figure III.16. De plus, le tableau III.3 résume les termes énergétiques et les RMSD obtenus pour ces 17 structures convergentes.

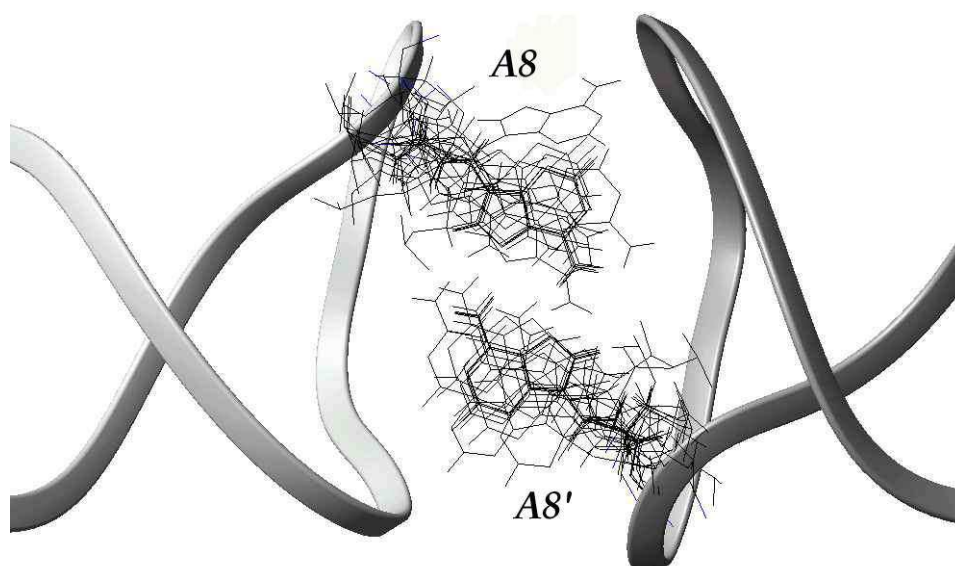


Figure III.15 : Positionnement des adénines 8 de chaque brin sur un ensemble de 18 premières structures ne présentant que peu de violations NOE. Le positionnement laisse supposer un appariement homopurique symétrique de type II de ces adénines.

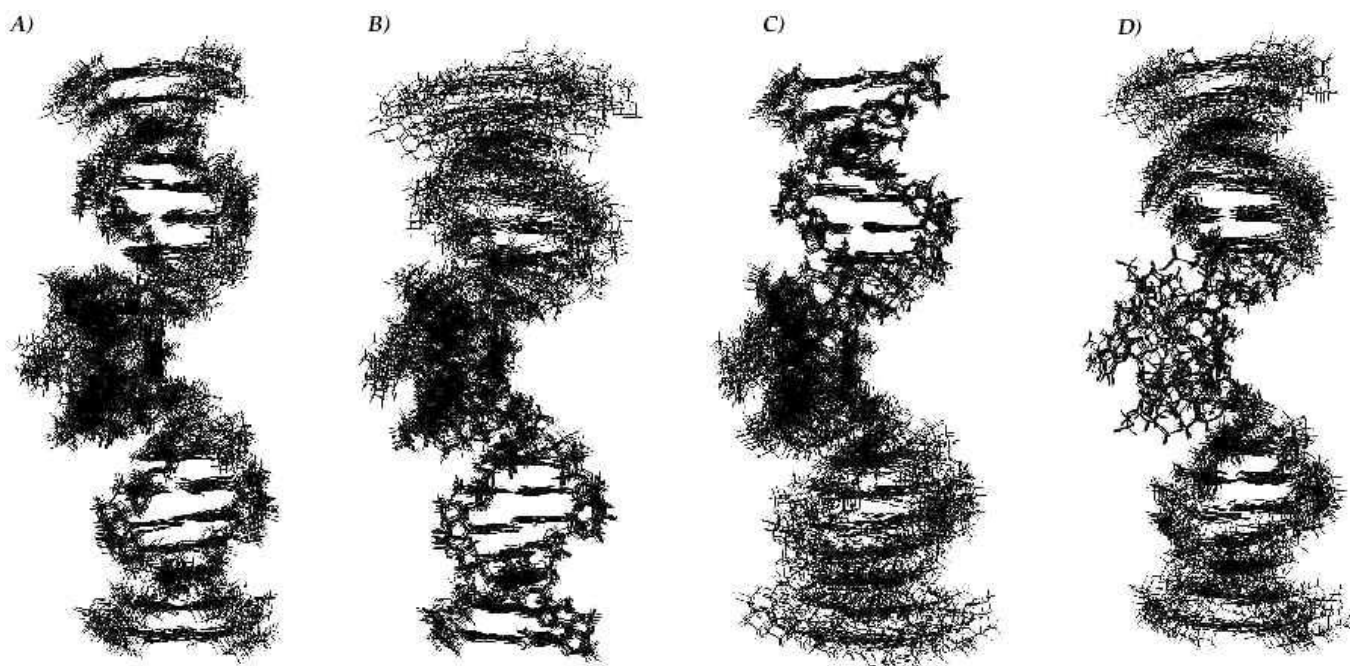


Figure III.16 : Superposition des 17 structures résultantes sur la moyenne. Les superpositions sont réalisées sur la structure moyenne non minimisée. A) superposition globale. B) superposition sur la tige 1 (de  $C_1$  à  $G_7$  et de  $C_{17}$  à  $G_{23}$ ). C) superposition sur la tige 2 (de  $C_1'$  à  $G_7'$  et de  $C_{17}'$  à  $G_{23}'$ ). D) superposition sur la tige centrale (de  $C_{11}$  à  $G_{14}$  et de  $C_{11}'$  à  $G_{14}'$ ).

| <u>Précision des structures</u>  |              |                        |
|--|--------------|------------------------|
| Nombre de structures   | 17           |                        |
| Nombre de violations NOE > 0.5 Å   | 0            |                        |
| Nombre de violations NOE > 0.3 Å   | 3 ± 2        |                        |
| Nombre de violations dièdre > 5°   | 0            |                        |
| RMSD <sup>1</sup> global   |              | 1,72 ± 0,58 Å          |
| RMSD tige1 (C1-G7, C17'-G23')  |              | 0,44 ± 0,17 Å          |
| RMSD tige2 (C1'-G7', C17-G23)  |              | 0,41 ± 0,20 Å          |
| RMSD tige3 (C11-G14, C11'-G14')  |              | 0,32 ± 0,11 Å          |
| RMSD jonction (A8-G10, C15-A16 et A8'-G10', C15'-A16')   |              | 1,11 ± 0,35 Å          |
| <u>Energie</u>   |              |                        |
| Energie totale   | -613,02 ± 70 | kcal.mol <sup>-1</sup> |
| Energie géométrique <sup>2</sup>   | +309,31 ± 47 | kcal.mol <sup>-1</sup> |
| Energie van der Waals  | -530,77 ± 18 | kcal.mol <sup>-1</sup> |
| Energie électrostatique  | -391,55 ± 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 torsions, angles impropres, distances NOE et angles dièdres. |              |                        |

Tableau III.3 : Résumé des énergies et des RMSD pour dSL1 sous forme de dimère instable.

## 4. Analyses structurales

### 4.1 Organisation globale des tiges

Dans la structure en kissing-complex de dSL1 nous avons trois zones distinctes en appariement de type Watson-Crick. Deux de ces zones sont identiques car symétriques l'une de l'autre. Il s'agit des tiges comportant les résidus  $C_1$  à  $G_7$  et  $C_{17}$  à  $G_{23}$  et donc leurs symétriques  $C'_1$  à  $G'_7$  et  $C'_{17}$  à  $G'_{23}$ . Les paramètres hélicoïdaux ont été déterminés à l'aide de la version 5.1 du logiciel CURVES (Lavery and Sklenar, 1988; Lavery and Sklenar, 1989; Lavery and Sklenar, 1990; Ravishanker, et al., 1989). L'organisation moyenne de ces deux tiges est représentée sur la figure III.17.

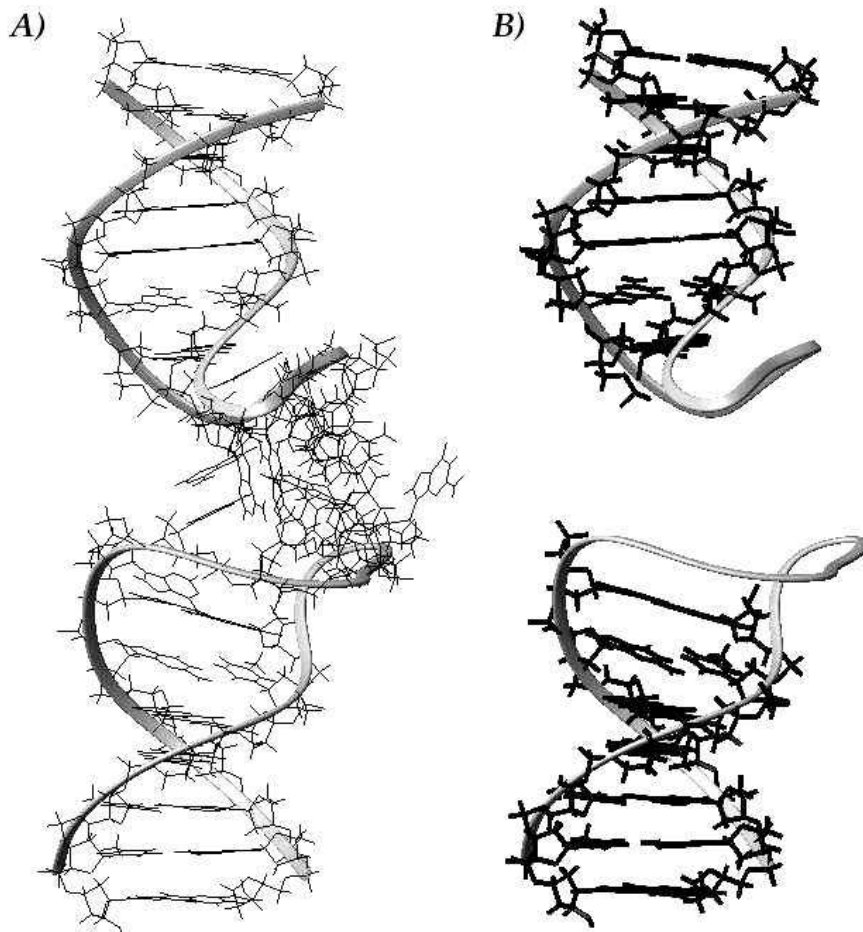


Figure III.17 : Organisation des deux tiges symétriques  $C_1$ - $G_7$  et  $C_{17}$ - $G_{23}$ . A) représentation globale de la molécule. B) Représentation des deux tiges symétriques.

Les deux tiges sont identiques, bien qu'aucune contrainte en ce sens n'ait été imposée. Elles présentent les caractéristiques classiques d'un ADN de type B. La structure de type kissing-complex ne nécessite donc pas la présence d'hélice de type A pour cadenasser l'interaction boucle-boucle. Nous pouvons affirmer sans aucun doute que ces deux tiges permettent la stabilisation de l'interaction boucle-boucle. Leur géométrie hélicoïdale classique, et donc stable, permet de "bloquer" la dissociation des deux brins et autorise l'appariement de la tige centrale.

La tige centrale comprend quatre paires de bases :  $C_{11}G_{14}'$  ;  $G_{12}C_{13}'$  ;  $C_{13}G_{12}'$  et  $G_{14}C_{11}'$ . Son organisation est loin d'être hélicoïdale, elle se présente plutôt comme un éventail (cf figure III.18).

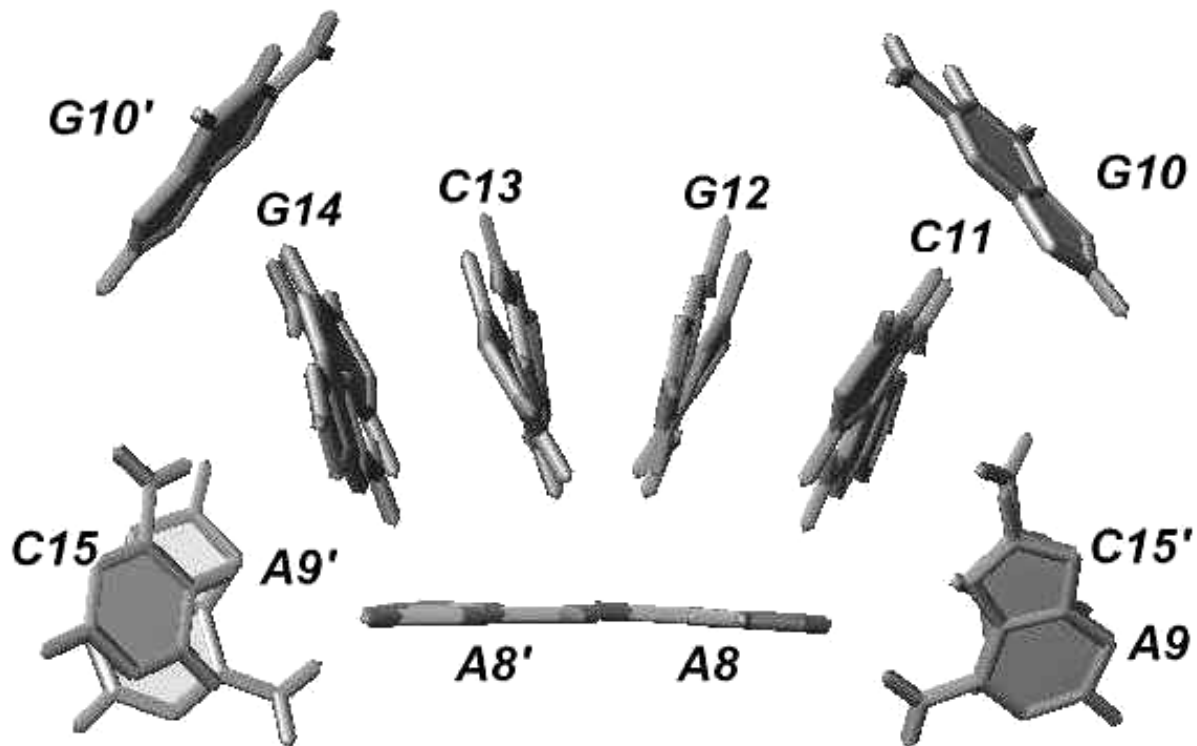


Figure III.18 : Vue orthogonale de l'interaction boucle-boucle. Pour des raisons de clarté, ne sont représentés que les plans des bases. Le motif représente un éventail ou les quatre paires de bases de l'interaction sont perpendiculaires au lecteur.

Les quatre plans de bases de ce motif en éventail sont régulièrement espacés d'un angle d'environ  $20^\circ$ . Le plan des adénines  $A_8$  de chaque brin est perpendiculaire à cet éventail.

Sur la structure, la guanine 10 est expulsée vers l'extérieur (cf figure III.18). Ceci confirme nos données RMN sur le non-appariement de cette base et le déplacement chimique vers les bas champs observé pour son proton H8. Tous les essais, lors de la modélisation moléculaire, consistant à contraindre la paire de bases G<sub>10</sub>C<sub>15</sub> se sont montrés infructueux. Par conséquent, quatre paires de bases suffisent au niveau de l'interaction boucle-boucle et cela parce qu'il existe deux motifs qui vont stabiliser cette structure métastable.

La jonction entre le motif en éventail et la tige hélicoïdale fait intervenir un réarrangement important du squelette phosphodiester. La principale distorsion intervient pour le résidu A<sub>8</sub>. Ce résidu est expulsé, comme nous allons le voir dans le chapitre suivant, et l'angle  $\gamma$  (O5'-C5'-C4'-C3') situé entre les résidus 8 et 9 passe d'une conformation gauche + à une conformation gauche -. La figure III.19 représente ces deux conformations possibles pour l'angle  $\gamma$ .

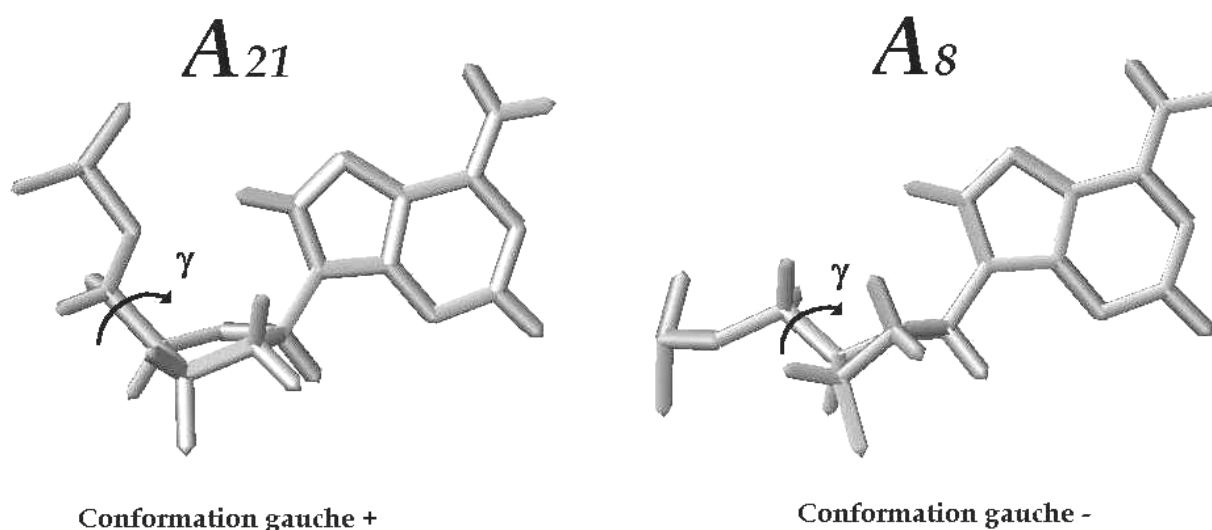


Figure III.19 : La distorsion du squelette phosphodiester est en partie due au passage de la conformation gauche + à gauche - de l'angle  $\gamma$ . La figure représente ces deux conformations rencontrées dans la structure du dimère instable de dSL1. Le résidu A<sub>21</sub> (à gauche) exhibe une conformation gauche +, classiquement rencontrée pour cet angle, par contre le résidu A<sub>8</sub> (à droite) montre une conformation gauche -.

L'autre distorsion du squelette phosphodiester s'observe "au retour" du motif en éventail. Là encore, les angles  $\gamma$  des résidus 15 et 16 rendent compte de cette distorsion. La

figure III.20 montre les valeurs que prend cet angle au cours de la séquence. Ces modifications du squelette phosphodiester sont nécessaires pour présenter, aux deux monomères, la séquence palindromique  $C_{11}G_{12}C_{13}G_{14}$  face à face et permettre ainsi l'interaction boucle-boucle.

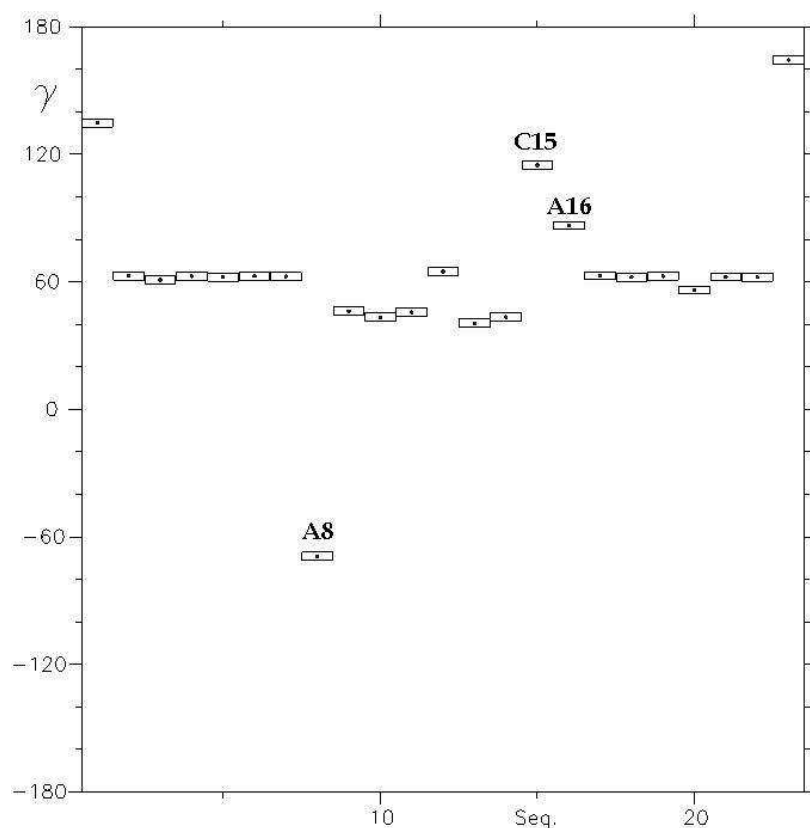


Figure III.20 : Les différentes valeurs que prend l'angle  $\gamma$  (en ordonnée) tout au long de la séquence. Les résidus 8, 15 et 16 sont, à eux seuls responsables de la distorsion du backbone de dSL1 en dimère instable.

#### 4.2 Caractérisation des motifs stabilisants

Deux motifs particuliers stabilisent la structure en kissing-complex de dSL1. Le premier est une paire de bases non canonique adénine/adénine entre les bases  $A_8$  de chaque brin. Du fait de la jonction entre la tige et l'interaction boucle-boucle, le résidu  $A_8$  est expulsé sous le motif en éventail. Le plan de sa base est, de ce fait, perpendiculaire à l'axe de la tige et présente ses atomes N7 et H61 pour former une liaison hydrogène symétrique (cf figure



III.21).

Lors du calcul des liaisons hydrogène nous avons défini celles-ci par une proximité inférieure à 2,4Å et un angle, entre l'atome donneur et accepteur de liaison hydrogène, inférieur à 35° (Billeter, et al., 1990). Une liaison hydrogène est particulièrement bien conservée, sur les 17 structures, et vient conforter la position spatiale de la base A<sub>8</sub>. En effet, l'atome N3 de cette dernière est engagé avec l'atome H61 de A<sub>9</sub> pour réaliser, là encore, cette liaison hydrogène stabilisante (cf figure III.21). Cette dernière n'est pas observable sur les spectres RMN ce qui indique qu'elle est échangée rapidement avec le solvant.

Le second motif stabilisant le dimère cinétique de dSL1 fait intervenir un empilement inter-brin : l'adénine 9 est en position de stacking avec la cytosine 15 du brin opposé. La figure III.22 montre cet empilement. Ce motif est confirmé par le déplacement vers les bas champs des protons aromatiques de ces deux résidus. Il conforte la position des résidus de jonction et stabilise l'interaction boucle-boucle.

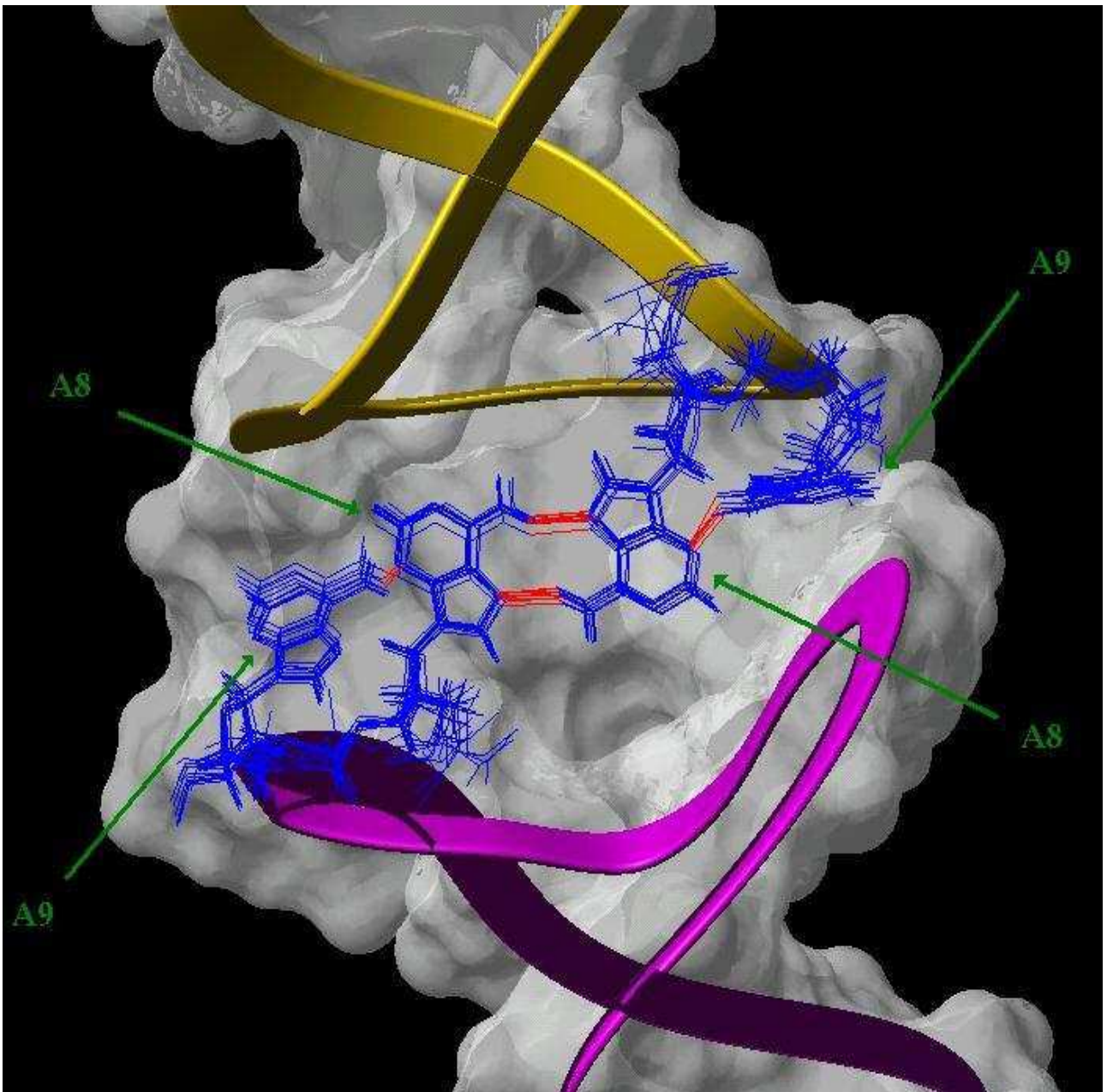


Figure III.21 : Figure représentant l'appariement  $A_8/A_8$  et les liaisons hydrogène avec le résidu  $A_9$ . La figure représente la superposition sur les 17 structures des résidus 8,9,8' et 9' incluse dans la surface et le backbone de la structure moyenne.

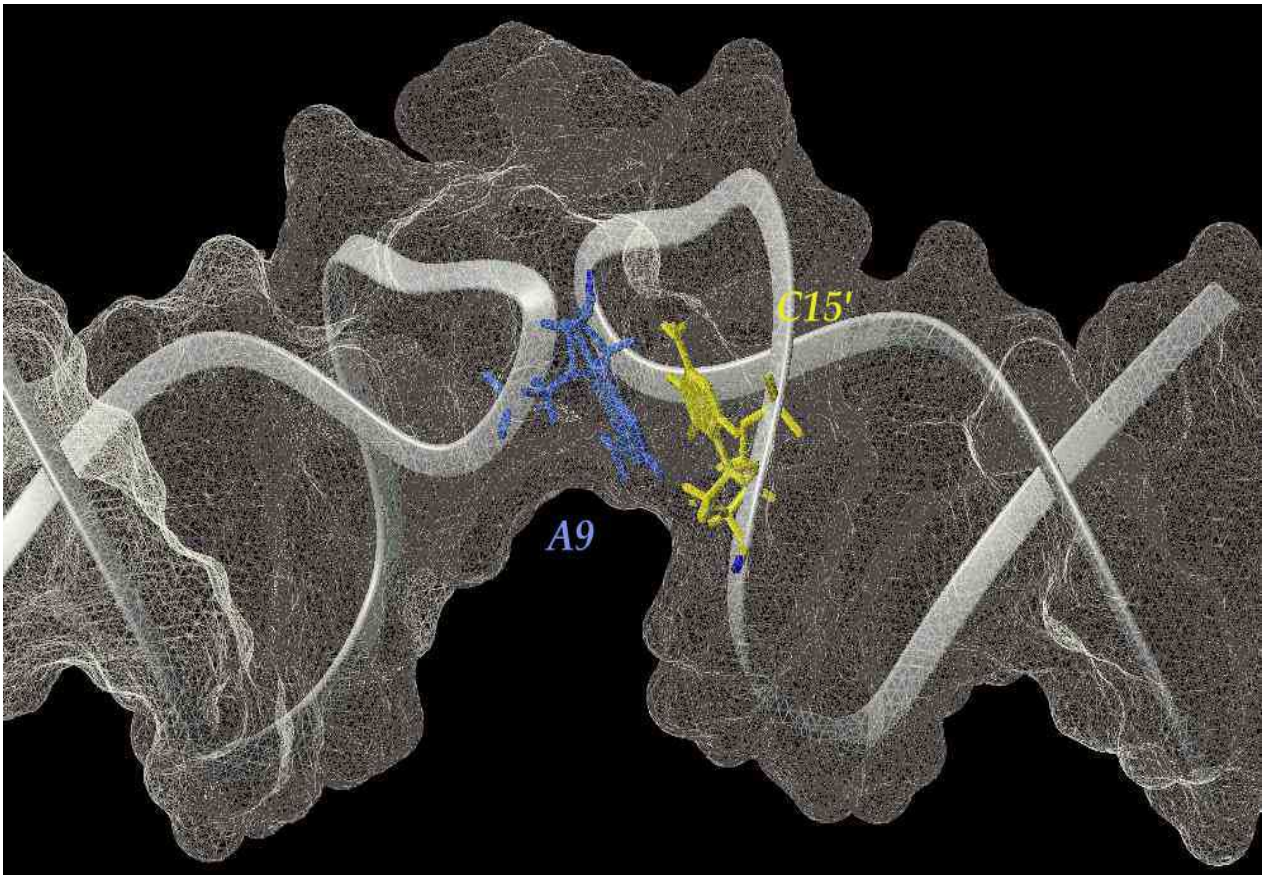


Figure III.22 : Représentation de l'empilement inter-brin entre le résidu A<sub>9</sub> et C<sub>15'</sub> du brin opposé. Ce stacking conforte la position des résidus de la jonction entre la tige et la boucle.

### 5. Comparaison avec la structure kissing-complex de l'ARN

On peut rapprocher la structure du dimère instable de dSL1 avec celle obtenue par l'équipe de James sur l'ARN (Mujeeb, et al., 1998). Cette dernière structure a été présentée dans le chapitre III-6.2 de la première partie.

Dans les deux cas, la structure globale représente un homodimère symétrique de type kissing-complex (cf figure III.23). Les deux structures présentent une courbure importante et quasi identique de l'interaction boucle-boucle avec l'axe de leurs tiges respectives (environ 100°). L'interface entre les monomères fait intervenir des appariements de type Watson-Crick, mais aussi des interactions inter-brin avec les bases de jonctions. Le squelette phosphodiester des résidus A<sub>9</sub> à C<sub>17</sub>, présente une forme en U permettant l'appariement entre les tiges-boucles.

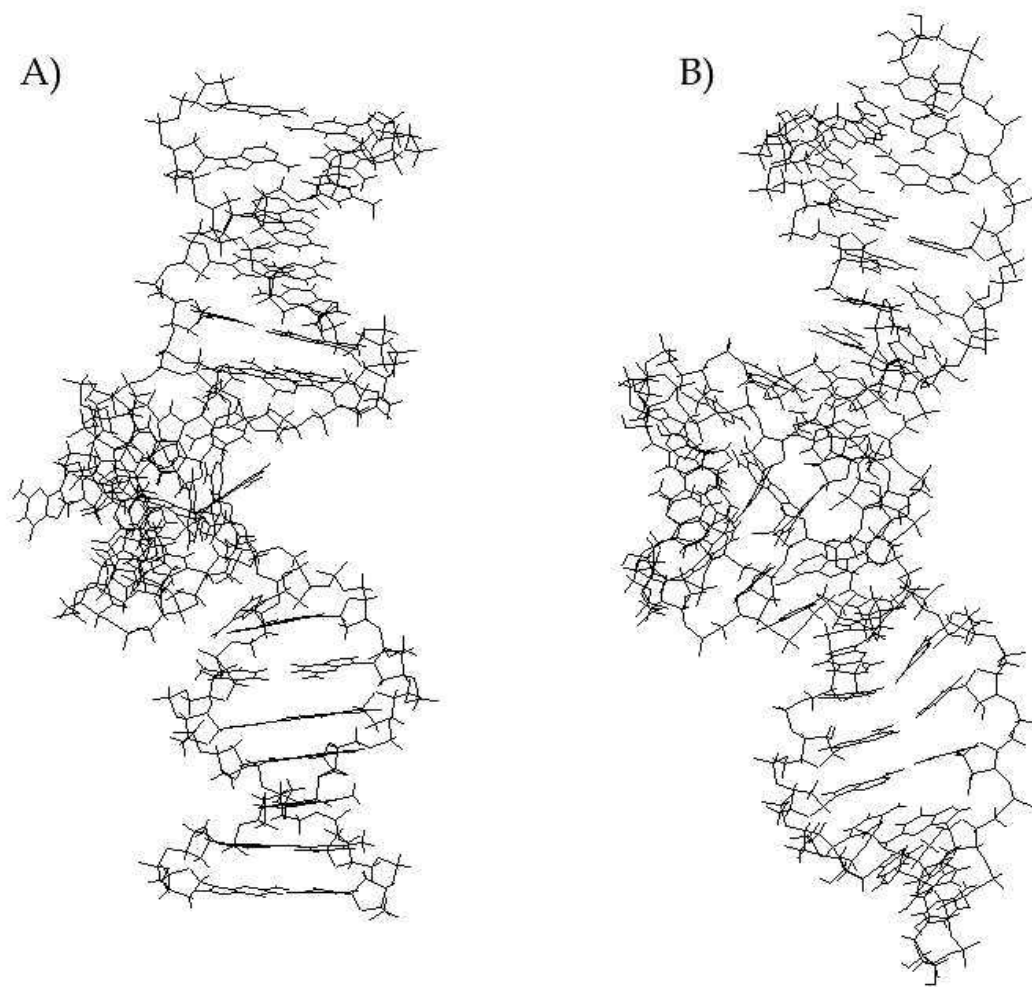


Figure III.23 : Comparaison de la structure globale du dimère instable dSL1 (A) et SL1 (B).

Dans les deux structures, la boucle est déviée vers le grand sillon de leurs tiges respectives. Enfin, les deux structures globales présentent d'abruptes discontinuités d'empilements et plusieurs bases non-appariées dans la région des résidus de jonction. Enfin, la stabilité des deux structures, calculée par spectroscopie UV, est quasi-identique : 32°C dans le cas de l'ARN et 35°C dans le cas de l'ADN.

Malgré tous ces points communs, il existe des différences entre les deux structures. La conformation de la tige obtenue pour l'ADN est de type B contrairement à l'ARN qui est de type A. Cette différence peut paraître sans intérêt mais prouve cependant que la structure en kissing-complex ne nécessite pas la formation d'hélice A.

La distorsion, induite par l'interaction boucle-boucle, du kissing-complex ARN entraîne la rupture de la paire de bases  $G_7C_{17}$ . Dans le cas de dSL1, cette paire de bases est conservée. Au niveau de l'interaction boucle-boucle, 6 appariements GC sont observables pour l'ARN pour seulement 4 dans le cas de l'ADN. De plus, dans la structure de dSL1 les adénines  $A_8$  des deux brins sont appariées, tandis que pour l'ARN, ces bases sont impliquées dans un stacking inter-brin à trois résidus : l'adénine  $A_9$  est expulsée du monomère puis empilée entre les résidus  $A_8$  et  $C_{17}$  de l'autre brin. Ce motif est stabilisé en solution par la présence d'une liaison hydrogène inter-brin entre l'atome O5' de  $A_9$  et le groupement 2'OH du résidu  $A_8$  du brin opposé. Bien évidemment, cette liaison hydrogène disparaît dans la structure ADN et l'adénine  $A_9$  est alors impliquée dans un empilement inter-brin avec la cytosine  $C_{15}$ , ce qui expulse vers le solvant le résidu  $G_{10}$ .

## 6. Conclusions

L'étude de l'analogue désoxyribose du fragment ARN SL1-23mer de VIH-1<sub>Lair</sub>, appelé dSL1, montre que deux conformations, de stabilité différente, existent.

L'étude RMN de l'espèce instable, montre un dimère comportant des caractéristiques structurales rarement rencontrées dans les études d'ADN. La présence de corrélations NOE entre résidus éloignés séquentiellement, l'observation de variations importantes du déplacement chimique du phosphore et la mise en évidence d'une paire de bases AA font partie de ces "bizarreries" rencontrées dans cette étude.

L'étape de modélisation moléculaire était cruciale car, pour être impartial, il fallait envisager la possibilité d'obtenir un dimère étendu malgré nos fortes présomptions. Lors de cette modélisation, la paire de bases AA a été finalement attribuée à la paire inter-brin  $A_8A_8$ .

La structure tridimensionnelle se présente sous forme globale d'un homodimère symétrique de type kissing-complex. Le squelette phosphodiester est distordu au niveau des résidus de jonction, avec notamment une transition gauche + vers gauche - de l'angle  $\gamma$  de  $A_8$ . L'interaction boucle-boucle ne présente que 4 appariements GC organisés en éventail. La structure présente deux motifs stabilisants : une paire  $A_8A_8$  inter-brin

renforcée d'une liaison hydrogène avec le résidu A<sub>9</sub> et un empilement avec le résidu C<sub>15</sub> et l'adénine A<sub>9</sub> du brin opposé. Malgré cela, la structure conserve un aspect métastable cohérent avec ce type de préparation (obtention du dimère cinétique). La comparaison avec la structure en kissing-complex de l'ARN SL1 montre quelques points communs et quelques différences mais ne permet pas de conclure sur le rôle du groupement 2'OH.

Nous pouvons néanmoins remarquer que l'organisation structurale des adénines 8,9 et 16 présentes dans les structures ARN de SL1 et maintenant dans la structure ADN du dimère instable, joue un rôle essentiel dans la stabilité de ces structures. Aucune de ces structures ne présente la même organisation de ses résidus. En fait, indépendamment de la nature ribose ou désoxyribose du fragment SL1, ces adénines comportent à elles seules le polymorphisme rencontré pour ces structures. La comparaison avec les structures élucidées pour la souche Mal, où l'adénine 9 devient une guanine, n'échappe pas à cette règle.

Cette structure représente, à notre connaissance, la première structure tridimensionnelle de type kissing-complex pour un ADN. Elle vient donc élargir l'ensemble des structures secondaires possibles pour les acides nucléiques. La compréhension complète du rôle du groupement 2'OH ne peut se faire qu'avec une simulation de dynamique newtonienne en présence d'un solvant explicite. Celle-ci est en cours de réalisation, dans l'équipe de Genest avec laquelle nous collaborons étroitement.

## IV. Le dimère stable de dSL1

Après la détermination structurale du dimère instable de dSL1, nous nous sommes tout naturellement intéressés à l'élucidation du comportement de son dimère stable.

### *1. Etudes physico-chimiques*

Nous avons mis en évidence, dans le chapitre III.1, l'existence de deux types distincts de dimères pour la séquence dSL1. Le premier de ces dimères, que nous avons étudié, correspond au dimère instable et représente le dimère cinétique. Sa préparation demande un soin tout particulier. Lorsque nous nous intéressons au dimère thermodynamique, nous nous concentrons sur le dimère le plus stable de cette séquence. L'obtention de ce dimère nécessite donc de favoriser les réarrangements conformationnels. Pour cela, l'échantillon est dénaturé complètement en chauffant celui-ci à 85°C, dans un bain thermostaté, pendant 3 minutes. L'acide nucléique n'est pas dilué comme dans le cas du dimère instable pour autoriser un maximum de réarrangements. Ce dernier point évite également bon nombre de manipulations de l'échantillon comme dans le cas de la préparation précédente. La descente en température s'effectue le plus lentement possible : l'échantillon est refroidi jusqu'à température ambiante, l'opération dure 6 heures. De cette façon, l'échantillon se trouve continuellement en condition adiabatique et favorise le dimère stable.

Une bonne méthode pour déterminer si la conformation souhaitée est celle du dimère stable, est de suivre par spectroscopie UV son processus de dénaturation/renaturation en fonction de la température. Une petite partie de l'échantillon que nous venons de préparer est diluée dans le même tampon de dimérisation ([TRIS]= 50mM, [NaCl]=100mM et [dSL1]=10 $\mu$ M) que celui utilisé pour mettre en évidence les deux types de dimères (cf chapitre précédent). La figure IV.1 représente cette courbe.

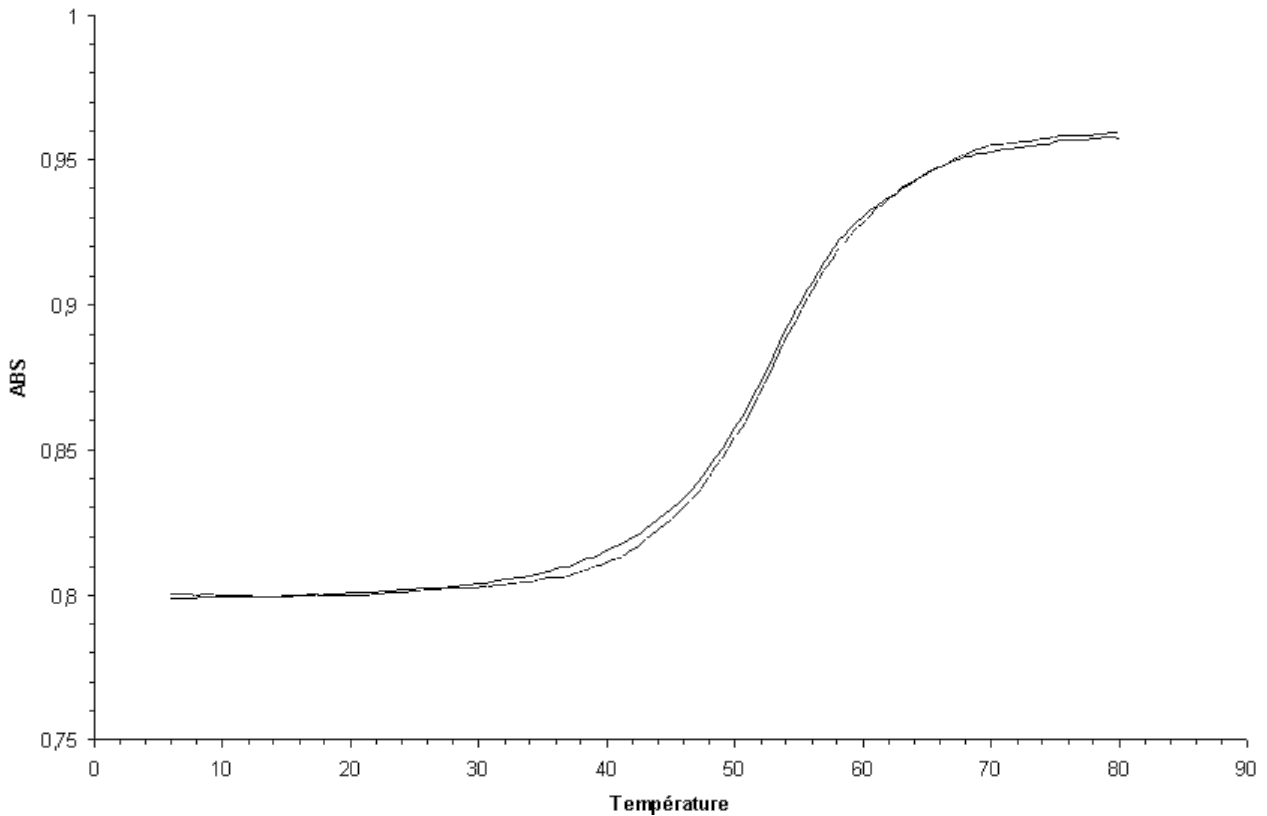


Figure IV.1 : Courbe de dénaturation/renaturation en fonction de la température de dSL1, préparé dans l'optique d'un dimère stable, suivi par spectroscopie UV à 256nm.

Sur cette courbe, que nous pouvons directement comparer à la figure III.1, une seule transition est visible. Elle correspond à la dénaturation du dimère stable et se rapproche aux courbes classiquement rencontrées pour un ADN ne présentant pas de polymorphisme.

## 2. Etudes RMN du dimère stable dSL1

### 2.1 Préparation de l'échantillon

L'obtention du dimère stable ne pose donc pas réellement de problème. L'échantillon est chauffé et refroidi comme précédemment. Il est alors lyophilisé, dilué dans une solution à 600 $\mu$ L d'eau légère à 90% ou d'eau lourde à 100%, puis dégazé à l'argon.



## 2.2 Attribution dans l'eau légère

La totalité des spectres obtenus dans l'eau légère a nécessité l'utilisation de la séquence d'impulsions "watergate" (Piotto, et al., 1992) avec une fenêtre spectrale de 13000Hz. La figure IV.2 montre l'agrandissement de la région des corrélations des protons iminos/iminos.

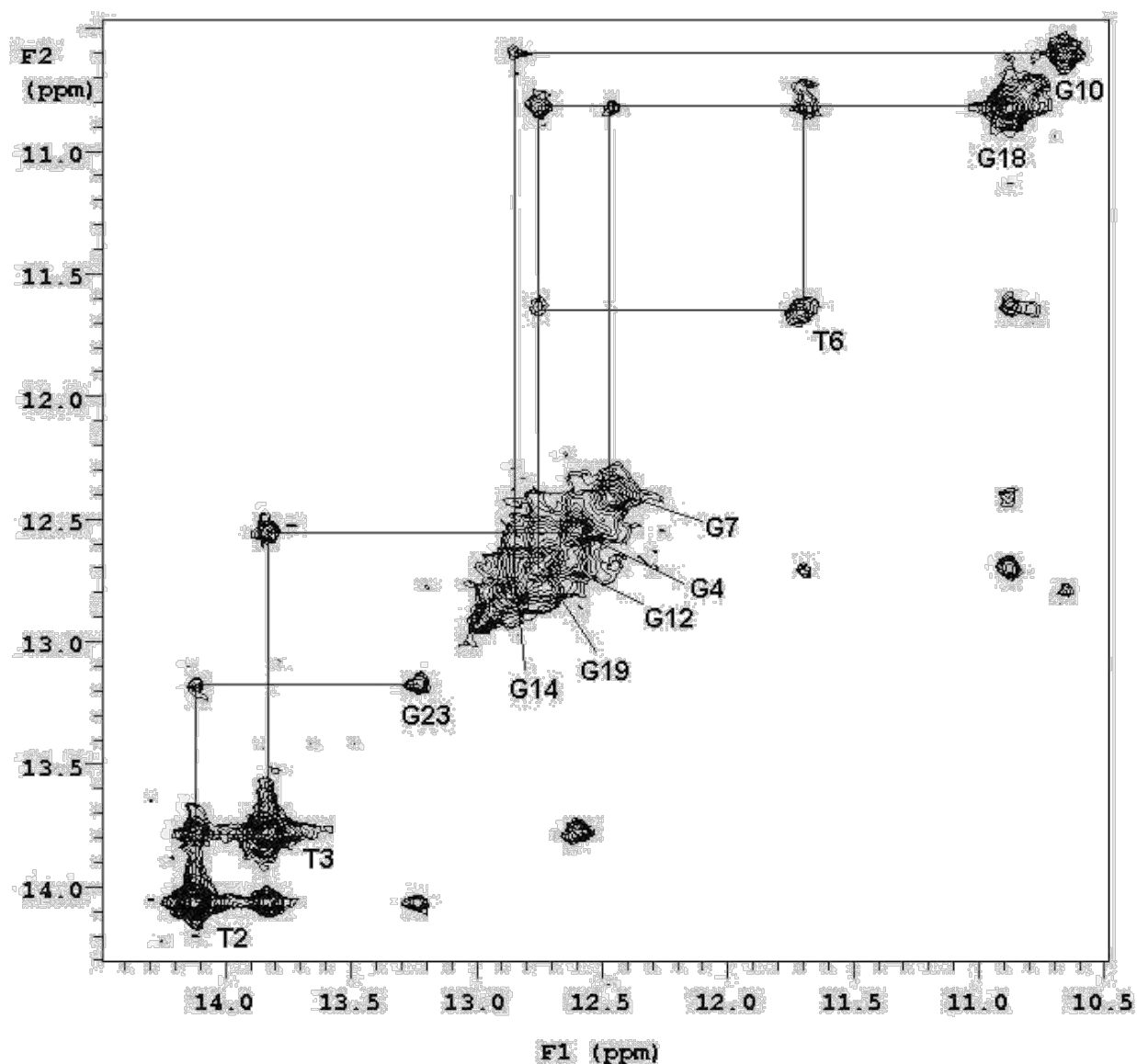


Figure IV.2 : Agrandissement du spectre NOESY ( $\text{H}_2\text{O}/\text{D}_2\text{O}$  90 :10 ;  $T_m=150\text{ms}$  ;  $\text{pH}=6,8$  ;  $T=278\text{K}$ ) de la région des connectivités des protons iminos/iminos.

Le point de départ et le cheminement de l'attribution sont identiques à ceux du dimère instable (figure IV.3). Cependant, lorsque l'on compare ce spectre à celui réalisé dans les

mêmes conditions pour le dimère cinétique (figure III.3), une résonance supplémentaire apparaît. Il s'agit du signal du proton imino de la guanine 10. Celui-ci a pu être attribué grâce à la connectivité que réalise son proton imino avec le proton imino de G<sub>14</sub> mais aussi à l'aide des corrélations entre le proton imino et les proton amino de C<sub>15</sub>.

Par conséquent, la première conclusion que nous pouvons déduire de cette étude est que nous sommes en présence d'une structure différente de celle élucidée précédemment. De plus, contrairement à ce que nous avons découvert auparavant, dans cette nouvelle structure de dimère tous les appariements de type Watson-Crick sont repérables.

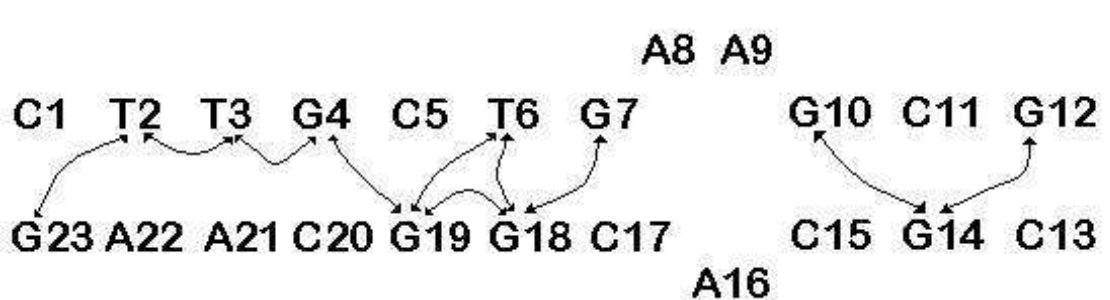


Figure IV.3 : Schéma des connectivités entre protons iminos rencontrés pour le dimère stable dSL1.

Le spectre des corrélations iminos/aminos et aminos/aminos ne révèle pas beaucoup de différences par rapport à l'étude précédente. Notons, néanmoins, que les résonances des protons aminos de la cytosine C<sub>15</sub> sont maintenant discernables. Ceci prouve que la paire G<sub>10</sub>C<sub>15</sub>, absente de la structure kissing-complex, est présente pour cette conformation. De plus, les signaux révélant la possibilité d'une paire de bases non-canonique adénine/adénine disparaissent. Une corrélation intéressante du point de vue structural apparaît dans l'étude de ce dimère stable. Il s'agit d'une corrélation NOE faible entre le proton imino de G<sub>10</sub> et le proton H2'' de A<sub>8</sub>. Cette connectivité n'était pas observable dans l'étude précédente.

### 2.3 Attribution dans l'eau lourde

Au vu de ces différences spectrales dans l'eau légère, il a été décidé de s'intéresser aux spectres effectués dans l'eau lourde. Pour cela, l'échantillon a été lyophilisé deux fois contre D<sub>2</sub>O avant d'être repris dans 600µL de D<sub>2</sub>O dans un tube scellé sous atmosphère

d'argon.

Dans un but de comparaison avec les données spectrales obtenues pour le dimère instable, les expériences NOESY ont été réalisées dans les mêmes conditions que précédemment. Elles ont été réalisées à 27°C pour différents temps de mélange : 0, 50, 90, 150 et 300ms, avec une accumulation de 32 scans pour chaque incrément t1. Un délai de 2 secondes a été utilisé en début de séquence pour optimiser le rapport signal/bruit de l'ensemble des protons. Une présaturation à basse puissance permet d'éliminer le signal d'eau résiduelle sur les spectres.

Le point de départ de l'attribution utilise les corrélations scalaires H6-H5 et H6-CH3 rencontrées sur les spectres TOCSY ou COSY pour les cytosines et les thymines. Il est nécessaire d'apporter un regard neuf sur cette attribution séquentielle afin de ne pas être influencé par l'attribution précédente. Sur le spectre NOESY au temps de mélange de 300ms et à la même température, toutes les corrélations intra et inter-résidus des protons H6 ou H8 avec les protons H1' sont observables. Il en est de même pour les protons H6 ou H8 avec les protons H2' et H2''. Ceci permet une attribution séquentielle avec peu d'ambiguïtés à cette température.

La figure IV.4 représente l'agrandissement de la région des connectivités des protons aromatiques/anomériques. Si l'on compare celui-ci au spectre de la même région pour le dimère instable (cf figure III.5) nous remarquons que de très légères différences.

Ceci peut paraître étrange, néanmoins dans le cas des deux dimères SL1 de l'ARN, les différences rencontrées sur les spectres RMN sont minimales (Girard, 2000; Mujeeb, et al., 1999). Dans le cas de l'analogue désoxyribose qu'est dSL1, on remarque néanmoins quelques différences marquantes : la résonance du proton H8 de A<sub>16</sub> est largement déplacée vers les hauts champs, l'intensité de la corrélation entre le proton H1' de G<sub>7</sub> et le proton H8 de A<sub>8</sub> indique une proximité spatiale plus importante de ces deux protons, le proton H2 de A<sub>8</sub> fait une corrélation moyenne avec le proton H1' de C<sub>15</sub>, et avec le proton H8 de G<sub>7</sub>.

De plus, on observe un cross-peak entre le proton H2 de A<sub>9</sub> et le proton H1' de A<sub>16</sub> et avec le proton H2 de A<sub>16</sub> (cf figure IV.5), ces corrélations ne sont pas observables pour le dimère cinétique.

Le tableau IV.1 résume l'ensemble des résonances établi pour le dimère stable. Ces résonances, doivent être directement comparées à celles observées pour le dimère instable.

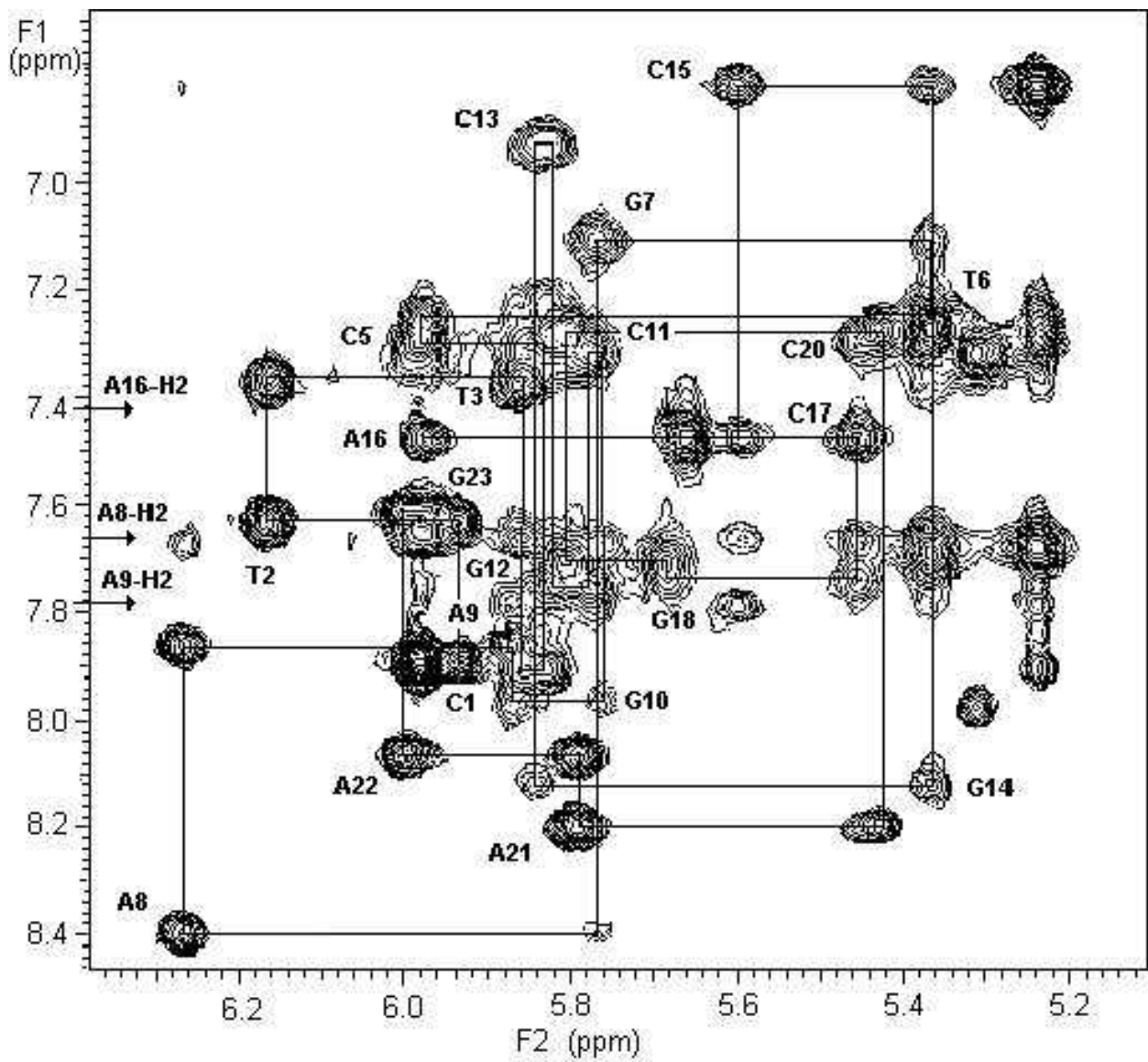


Figure IV.4 : Agrandissement du spectre NOESY à 300ms du dimère stable de dSL1.

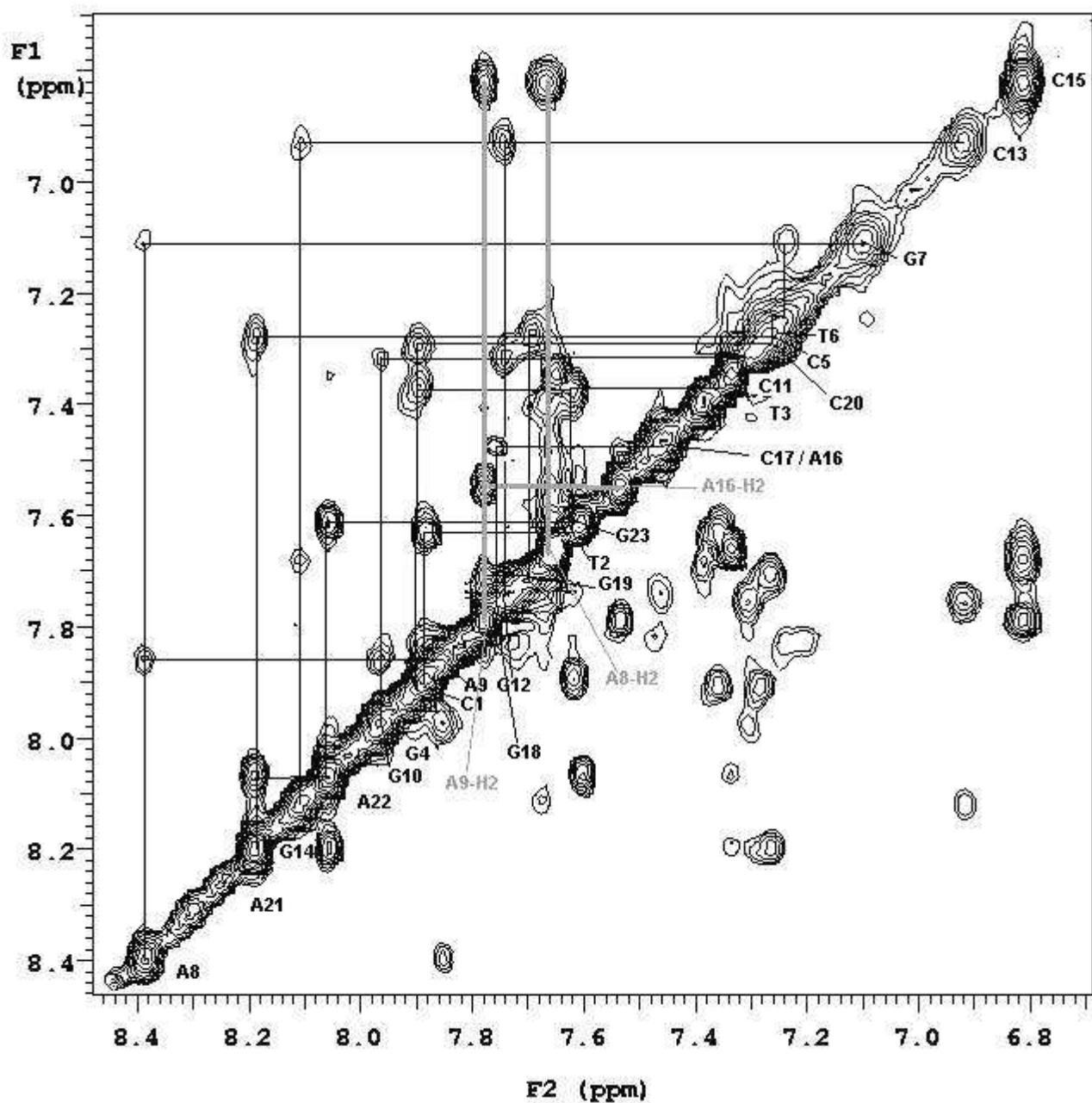


Figure IV.5 : Agrandissement du spectre NOESY du dimère stable ( $T_m=300\text{ms}$  ;  $T=27^\circ\text{C}$ ) de la région des corrélations aromatiques/aromatiques.

En effet, le déplacement chimique du proton est directement lié à l'environnement électronique, et par conséquent à la structure tridimensionnelle de la macromolécule. Les différences de déplacements chimiques sont indiquées par les histogrammes des figures IV.6, IV.7 et IV.8.

|            | H8/H6 | H5/H2/CH3 | H1'  | H2'  | H2'' | H3'  | H4'  | H5'  | H5'' | imino | NH2 lié | NH2 libre |
|------------|-------|-----------|------|------|------|------|------|------|------|-------|---------|-----------|
| <b>C1</b>  | 7,88  | 5,95      | 5,90 | 2,60 |      | 4,65 | 4,10 | 3,82 | 3,78 |       | 7,30    | 6,00      |
| <b>T2</b>  | 7,62  | 1,67      | 6,13 | 2,60 | 2,22 | 4,87 | 4,22 | 4,12 | 4,06 | 4,24  |         |           |
| <b>T3</b>  | 7,35  | 1,67      | 5,83 | 2,46 | 2,10 | 4,86 | 4,22 |      |      | 3,95  |         |           |
| <b>G4</b>  | 7,88  |           | 5,81 | 2,12 | 2,62 | 4,94 | 4,35 |      |      | 2,71  |         |           |
| <b>C5</b>  | 7,26  | 5,22      | 5,96 | 2,42 | 1,90 | 4,94 | 4,22 | 3,85 | 3,80 |       | 8,18    | 6,50      |
| <b>T6</b>  | 7,22  | 1,69      | 5,34 | 2,48 | 1,60 | 4,89 | 4,30 |      |      | 1,81  |         |           |
| <b>G7</b>  | 7,08  |           | 5,74 | 1,27 | 1,18 | 4,61 | 4,16 | 3,85 | 3,79 | 2,58  |         |           |
| <b>A8</b>  | 8,37  | 7,66      | 6,24 | 3,17 | 2,91 | 5,03 | 4,30 | 4,00 | 3,70 |       |         |           |
| <b>A9</b>  | 7,84  | 7,76      | 5,85 | 2,49 | 1,49 | 4,82 | 4,25 |      |      |       |         |           |
| <b>G10</b> | 7,96  |           | 5,76 |      |      | 4,84 | 4,24 | 4,13 | 4,00 | 0,77  |         |           |
| <b>C11</b> | 7,3   | 5,29      | 5,75 | 2,41 | 2,05 | 4,79 | 4,23 |      |      |       | 8,07    | 6,34      |
| <b>G12</b> | 7,73  |           | 5,79 | 2,58 | 2,38 | 4,87 | 4,24 |      |      | 2,83  |         |           |
| <b>C13</b> | 6,92  | 5,01      | 5,82 | 2,00 | 1,43 | 4,42 | 4,23 | 4,08 | 4,00 |       | 8,48    | 6,78      |
| <b>G14</b> | 8,10  |           | 5,34 | 2,75 | 2,68 | 5,00 | 4,42 |      |      | 2,96  |         |           |
| <b>C15</b> | 6,8   | 5,22      | 5,57 | 1,56 | 0,42 | 4,54 | 4,16 | 3,93 | 3,88 |       | 8,50    | 6,79      |
| <b>A16</b> | 7,45  | 7,52      | 5,94 | 2,70 | 2,63 | 4,89 | 4,33 |      |      |       |         |           |
| <b>C17</b> | 7,45  | 5,63      | 5,43 | 2,31 | 2,09 | 4,90 | 4,27 |      |      |       | 8,10    | 6,72      |
| <b>G18</b> | 7,72  |           | 5,67 | 2,72 | 2,52 | 4,91 | 4,30 |      |      | 0,99  |         |           |
| <b>G19</b> | 7,68  |           | 5,78 | 2,20 | 1,82 | 4,99 | 4,43 |      |      | 2,86  |         |           |
| <b>C20</b> | 7,26  | 5,34      | 5,41 | 2,28 | 1,82 | 4,80 | 4,10 |      |      |       | 8,29    | 6,45      |
| <b>A21</b> | 8,17  | 7,32      | 5,76 | 2,80 | 2,66 | 4,97 | 4,31 | 4,04 | 3,93 |       | 7,90    | 6,20      |
| <b>A22</b> | 8,04  | 7,64      | 5,97 | 2,66 | 2,54 | 4,97 | 4,37 | 4,22 | 4,15 |       | 7,71    | 6,17      |
| <b>G23</b> | 7,60  |           | 5,94 | 2,34 | 2,24 | 4,97 | 4,46 |      |      | 3,35  |         |           |

Tableau IV.1 : Tableau des déplacements chimiques du proton pour le dimère stable de dSL1.

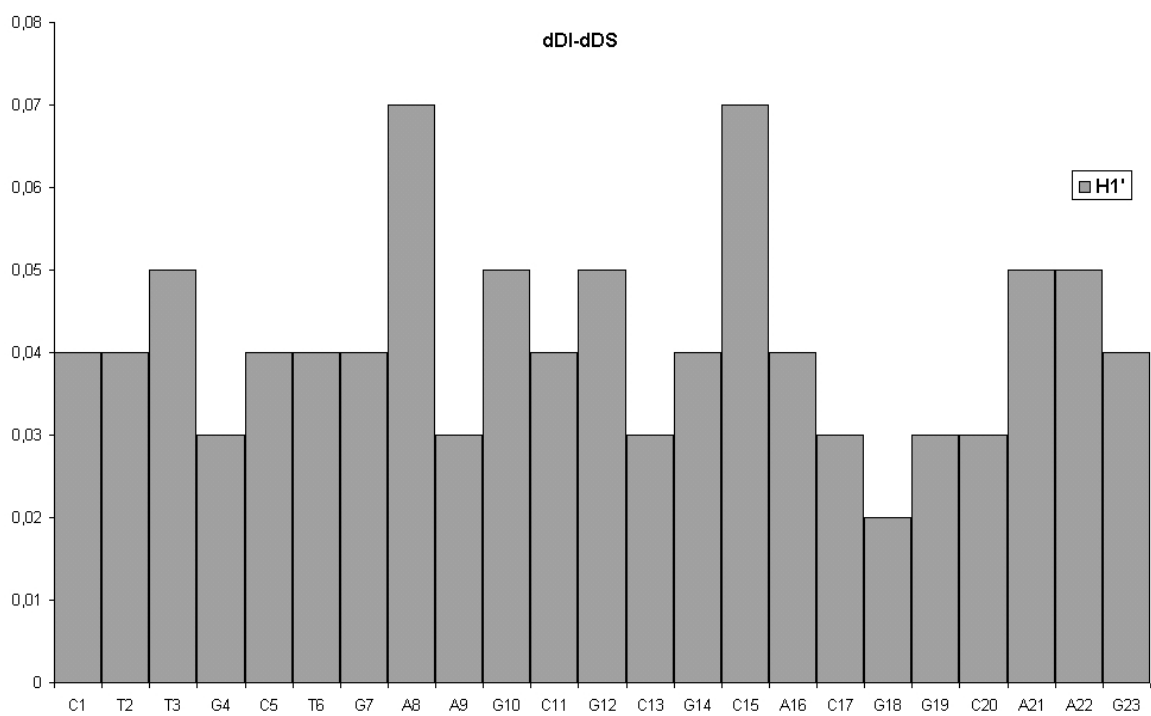


Figure IV.6 : Différences de déplacements chimiques pour les protons H1' entre les deux conformères de dSL1.

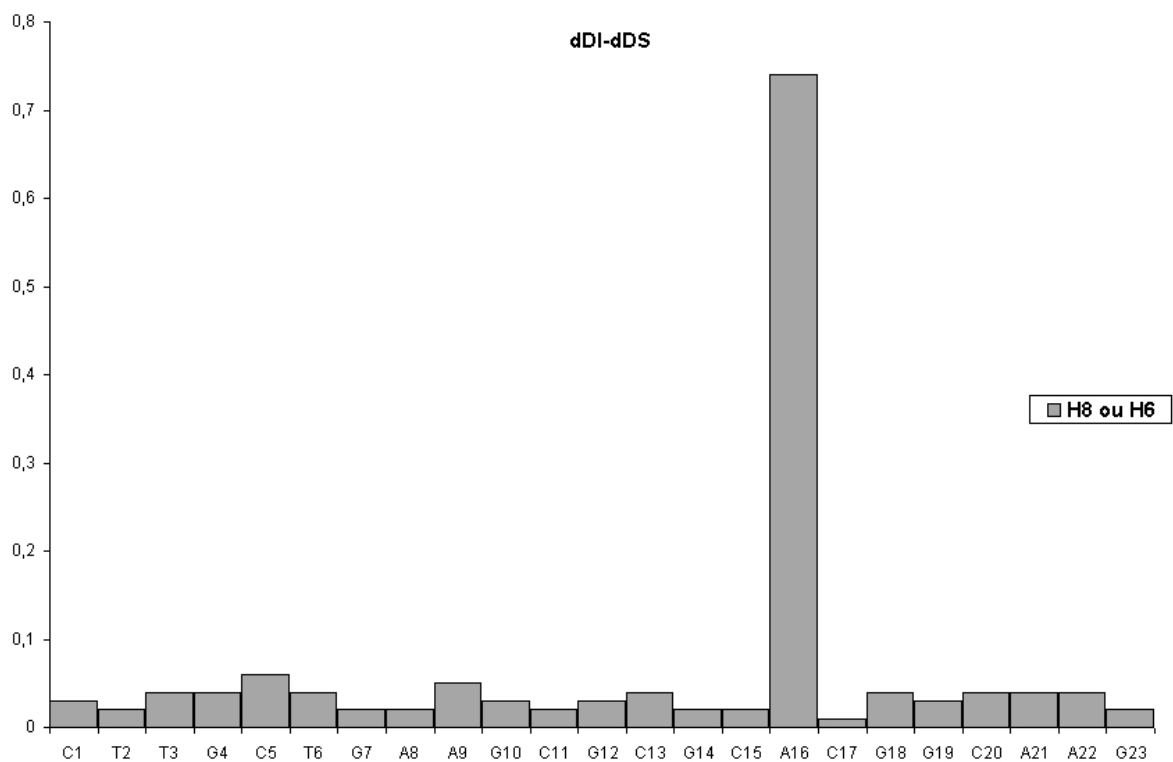


Figure IV.7 : Différences de déplacements chimiques pour les protons H8 ou H6 entre les deux conformères de dSL1.

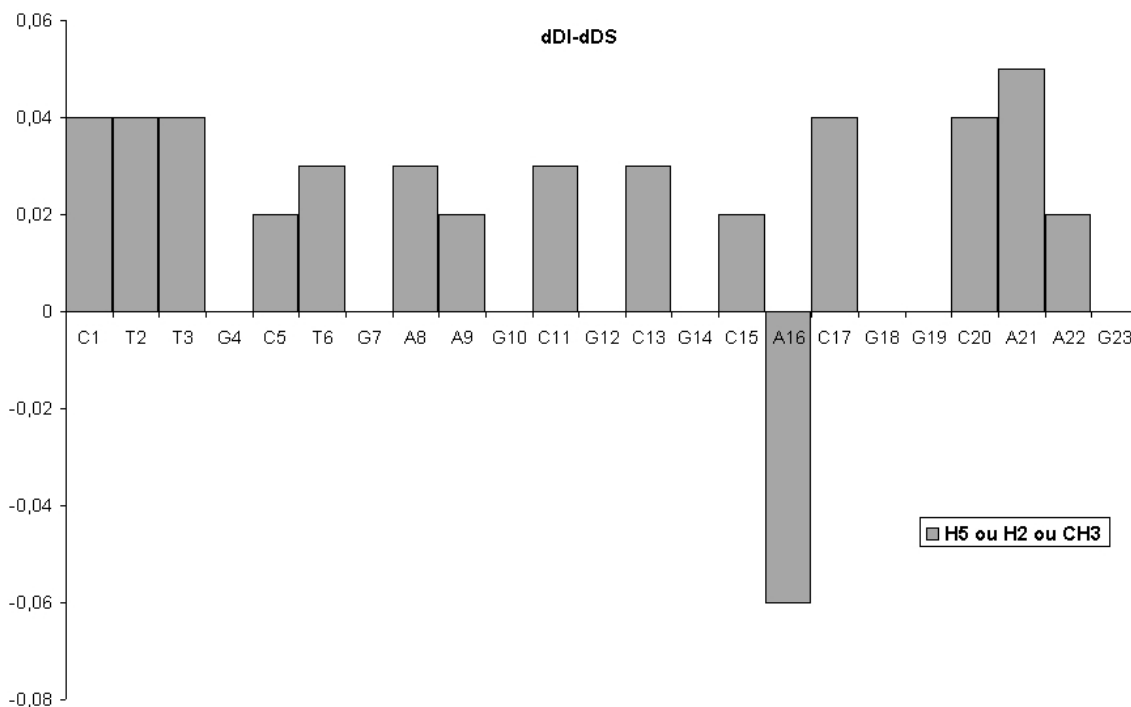


Figure IV.8 : Différences de déplacements chimiques pour les protons H5 (C), H2(A) ou CH3 (T) entre les deux conformères de dSL1.



Le calcul du temps de corrélation de la molécule, à l'aide du programme RELAZ (Lancelot, et al., 1989) nous donne une valeur de 7,5ns cohérente avec un dimère et identique à la précédente conformation. La présence d'une paire de bases supplémentaire à la structure précédente et de corrélations dipolaires non rencontrées précédemment sont des arguments structuraux forts qui indiquent une nouvelle conformation pour ce dimère.

Cependant, certaines connectivités NOE, présentes dans le dimère de type "kissing-complex", sont toujours existantes dans cette nouvelle structure. Notamment, la présence d'une corrélation forte entre le proton H2 de A<sub>8</sub> et le proton H1' de G<sub>14</sub>, qui est difficilement envisageable pour un dimère étendu, nous laisse dans l'expectative. Là encore, il sera nécessaire d'effectuer une modélisation moléculaire avec les deux modèles de structures secondaires pour être totalement rigoureux, l'ensemble des informations structurales ne permettant pas de trancher entre les deux modèles.

#### **2.4 Caractérisation du squelette phosphodiester**

Le spectre phosphore de dSL1 a été effectué à la même température (27°C). Il est quasiment identique à celui observé pour le dimère instable (cf figure IV.10 haut). Il nous semble intéressant de comparer ce spectre phosphore à celui obtenu pour le fragment d'ARN SL1-23mer sous forme de dimère étendu (cf figure IV.10 bas). En effet, dans la structure du dimère étendu de SL1 le squelette phosphodiester est très régulier et l'ensemble de ces résonances phosphores apparaît dans une fenêtre étroite de déplacement chimique (cf figure IV.10 bas). Par conséquent, nous pouvons supposer que l'ensemble des conformations du squelette phosphodiester ne sera pas homogène pour le dimère stable de dSL1.

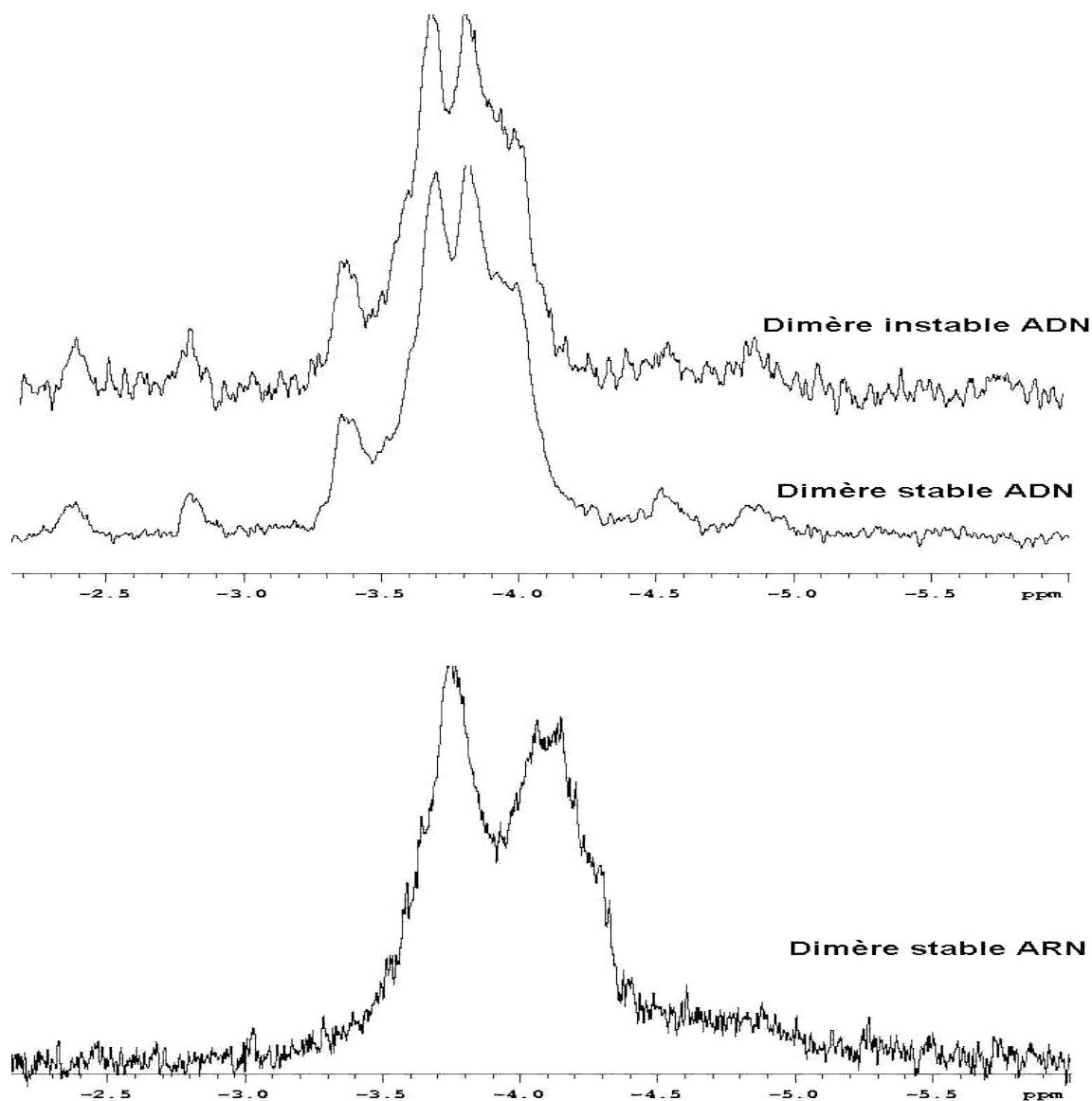


Figure IV.10 : Spectre unidimensionnelle du  $^{31}\text{P}$  à 27°C. En haut, superposition des deux spectres de dSL1 (dimère instable et stable). En bas, spectre du dimère stable de SL1-23mer.

### 3. Modélisation moléculaire

#### 3.1 Etablissement des contraintes

Les conformations des unités désoxy-ribores des résidus nous sont données par l'étude des spectres TOCSY et COSY-DQF. Du fait du recouvrement, la mesure des constantes de couplages H1'-H2' et H1'-H2'' n'est pas toujours possible. Cependant, l'observation continue sur le spectre TOCSY de ces corrélations montre que tous les résidus sont en

conformation sud. Enfin, cette conformation sud des sucres est confirmée par l'observation régulière d'un effet NOE faible entre les protons H6 ou H8 et leurs protons H1' intra-résidu, par rapport à l'effet NOE fort rencontré entre les protons H6 ou H8 et leurs protons H2'. Trois angles dièdres ( $\nu_1$ ,  $\nu_2$ , et  $\nu_3$ ) contraignent le plissement des cycles riboses dans cette modélisation, ils sont imposés avec une incertitude de 10°.

Comme nous l'avons vu précédemment, le spectre 1D phosphore est quasi identique à celui obtenu pour le dimère instable. Dès lors, les contraintes angulaires pour le squelette phosphodiester sont identiques à celles développées pour le dimère instable (chapitre III-3.1).

Les corrélations NOE issues des études dans D<sub>2</sub>O ont été intégrées à plusieurs temps de mélange (0, 50, 90, 150 et 300ms) à l'aide du programme RELAZ (Lancelot, et al., 1989). Les distances sont introduites dans un fichier de contraintes avec une incertitude de 15%. Les distances présentant des ambiguïtés intra ou inter-brin sont introduites en utilisant la fonction SUM (Nilges, 1993) pour ne pas décider d'une attribution *a priori*.

Les effets NOE observés dans l'eau légère et permettant de définir l'appariement de type Watson-Crick sont introduits sous forme de contraintes de deux distances pour chaque liaison hydrogène. Les autres effets NOE concernant les protons échangeables, ne sont pas intégrés mais introduits sous forme d'un intervalle d'inclusion ( $1,8 < d < 5,5 \text{ \AA}$ ) dans le but de s'affranchir des vitesses d'échanges avec le solvant.

Il existe un nombre important de contraintes rencontrées dans l'étude du dimère stable et également dans la structure du dimère instable. Plutôt que de montrer toutes les contraintes de distances, la figure IV.11 présente les différences rencontrées entre ces deux dimères.

Enfin, dans les premiers cycles de modélisation 20 contraintes de plans sont créées. Elles servent à maintenir les bases appartenant à un même appariement de type Watson-Crick dans le même plan : C<sub>1</sub>G<sub>23</sub>, T<sub>2</sub>A<sub>22</sub>, T<sub>3</sub>A<sub>21</sub>, G<sub>4</sub>C<sub>20</sub>, C<sub>5</sub>G<sub>19</sub>, T<sub>6</sub>G<sub>18</sub>, G<sub>10</sub>C<sub>15</sub>, G<sub>7</sub>C<sub>17</sub>, C<sub>11</sub>G<sub>14</sub>, G<sub>12</sub>C<sub>13</sub> et leurs symétriques. Le tableau IV.2 résume l'ensemble des contraintes expérimentales

employées lors de la modélisation.

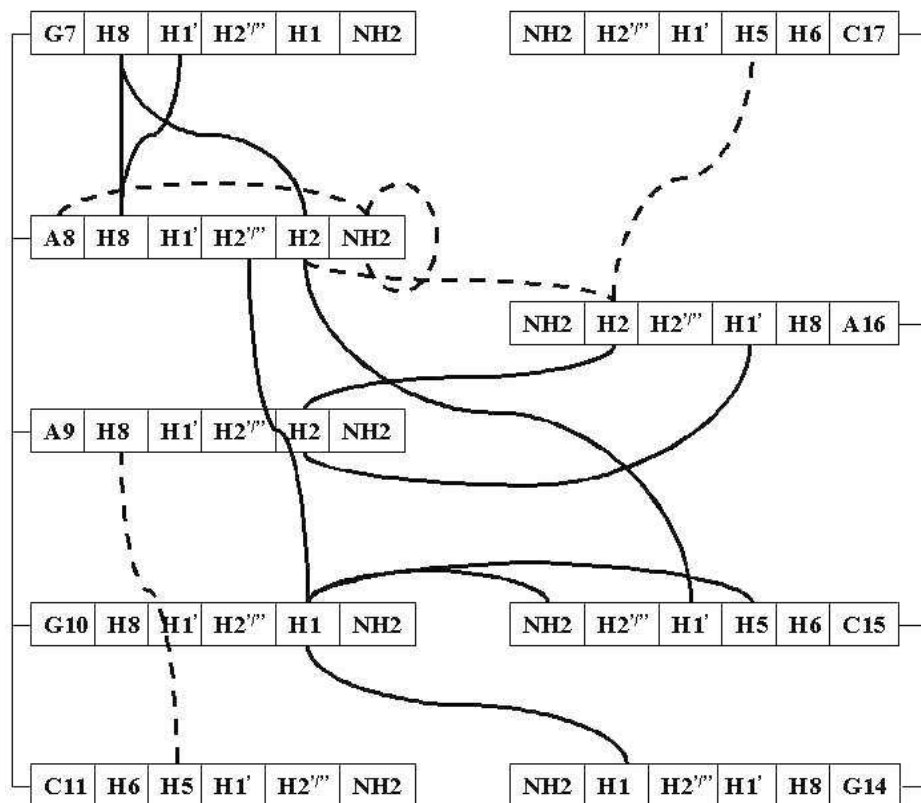


Figure IV.11 : Différences de contraintes de distances observées seulement dans le dimère stable (trait plein) ou seulement dans le dimère instable (pointillé).

| <u>Contraintes de distances</u>         |      |
|---|------|
| Issues de l'étude dans D <sub>2</sub> O |      |
| intra-résidu                            | 184  |
| inter-résidu                            | 320  |
| Issues de l'étude dans H <sub>2</sub> O |      |
| contraintes Watson-Crick                | 110  |
| protons échangeables                    | 70   |
| Total contraintes de distances          | 684  |
| NOE/résidu                              | 14,9 |
| <u>Contraintes d'angles</u>             |      |
| Plissement du sucre (v1, v2 et v3)      |      |
| backbone (α, β, γ, ε et ζ)              | 138  |
| Total contraintes d'angles              | 168  |
| Angles/résidu                           | 306  |
|   | 6,6  |
| Total contraintes                       | 942  |
| Contraintes/résidu                      | 21,5 |

Tableau IV.2 : résumé des différentes contraintes imposées dans la modélisation moléculaire du dimère stable de dSL1.

### 3.2 Résolution de la structure tridimensionnelle

Comme nous l'avons remarqué précédemment, il n'est pas possible à la simple vue des corrélations NOE, de discriminer une structure secondaire d'une autre. Par conséquent, les deux types de structures secondaires, dimère étendu et dimère boucle-boucle, sont testés sans aucun *a priori*.

Les structures aléatoires de départ sont les mêmes que celles effectuées pour le dimère instable. Ainsi, les différences que nous pouvons rencontrer entre les deux structures ne seront pas dépendantes du point de départ de la modélisation.

Les 100 structures aléatoires subissent le même protocole de modélisation que pour la détermination structurale du dimère instable.

Cependant, le modèle du dimère étendu ne montre aucune convergence. En effet, le nombre de violations NOE supérieures à  $1,0\text{\AA}$  est très important ( $>7$ ) dans la région des jonctions de tiges. La figure IV.12 représente les 6 meilleures structures obtenues pour ce type de dimère.

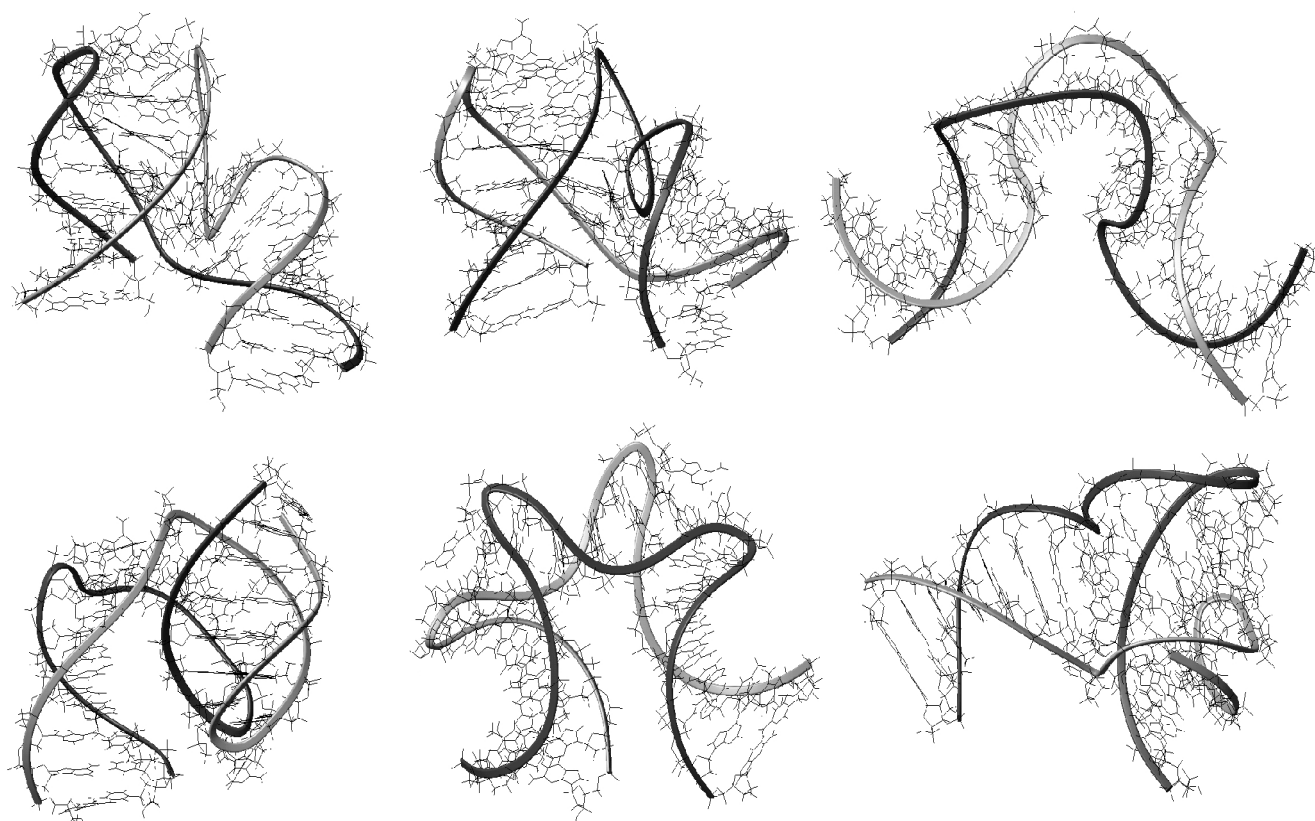


Figure IV.12 : Les six meilleures structures observées pour le modèle de duplexe étendu. Ces structures présentent bon nombre de violations NOE supérieures à 1,0Å.

Comme on peut le constater sur la figure IV.11, ces structures sont très différentes les unes des autres et ont des énergies totales beaucoup plus élevées que le modèle en "kissing-complex". Cette hypothèse est donc abandonnée, le dimère stable est donc également un dimère boucle-boucle.

13 structures convergent suivant le modèle de type kissing-complex. Elles présentent : aucune violation de distances supérieure à 0,5Å, aucune violation d'angles dièdres supérieure à 5° et elles montrent des différences de déplacements chimiques calculés et expérimentaux inférieures à 0,25ppm. Ce dernier critère de sélection est expliqué dans le chapitre suivant. Le tableau IV.3 résume l'ensemble des termes énergétiques et les RMSD obtenus pour ces 13 structures. La figure IV.13 montre les différentes superpositions.

### Précision des structures

|                                 |              |
|---------------------------------|--------------|
| Nombre de structures            | 13           |
| Nombre de violations NOE>0,5Å   | 0            |
| Nombre de violations NOE>0,3Å   | 2 ±1         |
| Nombre de violations dièdre>5°  | 0            |
|                                 |              |
| RMSD <sup>1</sup> global        | 1,71 ±0,80 Å |
| RMSD tige1 (C1-G7,C17'-G23')    | 0,46 ±0,22 Å |
| RMSD tige2 (C1'-G7',C17-G23)    | 0,44 ±0,18 Å |
| RMSD tige3 (G10-C15 ,G10'-C15') | 0,19 ±0,10 Å |
| RMSD adénines de jonction       | 0,52 ±0,22 Å |

### Energie

|                                  |             |                        |
|----------------------------------|-------------|------------------------|
| Energie totale                   | -708,40 ±42 | kcal.mol <sup>-1</sup> |
| Energie géométrique <sup>2</sup> | +282,65 ±43 | kcal.mol <sup>-1</sup> |
| Energie van der Waals            | -595,48 ±20 | kcal.mol <sup>-1</sup> |
| Energie électrostatique          | -394,85 ±23 | 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 torsions, angles impropres, distances NOE et angles dièdres.

Tableau IV.3 : Résumé des énergies et des RMSD pour dSL1 sous forme de dimère stable.

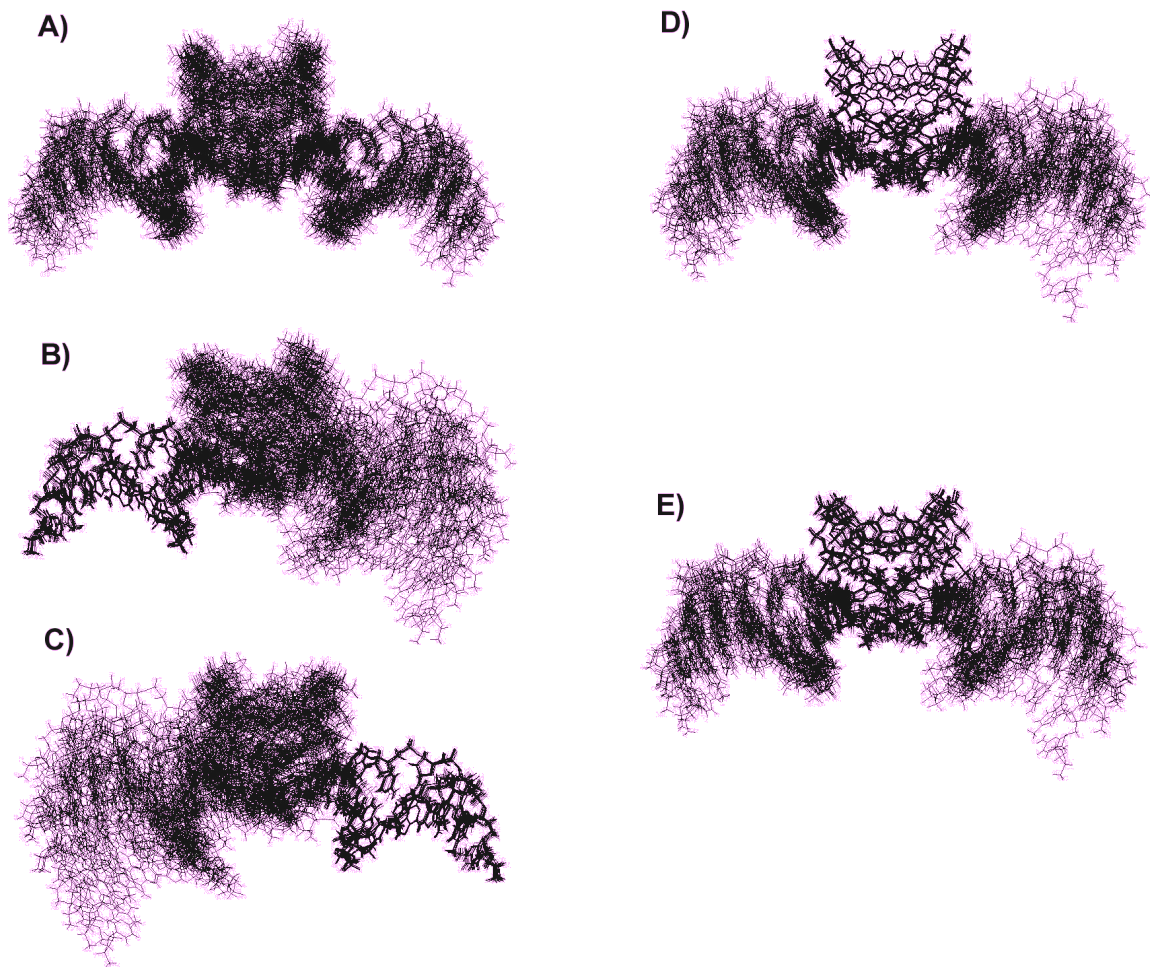


Figure IV.13 : Superpositions des 13 structures du dimère stable : A) superposition globale ; B) ,C) et D) superposition sur les tiges 1,2 et centrale ; E) superposition sur les adénines de jonctions.

### 3.3 Analyse du déplacement chimique en retour

L'interprétation structurale dans les acides nucléiques du déplacement chimique est délicate et comporte de nombreux paramètres difficilement quantifiables géométriquement. Le déplacement chimique dépend de la densité électronique présente aux alentours du noyau considéré et, dès lors, il existe une corrélation profonde entre le déplacement chimique et la structure tridimensionnelle d'une macromolécule.

Dans le cas des protéines, il a été montré (Wishart, et al., 1992) que le déplacement chimique de l'hydrogène alpha renseigne directement sur la structure secondaire (hélice  $\alpha$ ,



feuille  $\beta$ ) de la protéine. De plus, dans le logiciel de modélisation moléculaire CNS (Brunger, et al., 1998) un potentiel de contrainte utilisant le déplacement chimique du proton (Kuszewski, et al., 1995a; Kuszewski, et al., 1996), ou du carbone 13 (Kuszewski, et al., 1995b), peut être utilisé au même titre que l'effet NOE ou les constantes de couplages.

Dans le cas des acides nucléiques il n'existe pas de potentiel de contrainte qui utilise le déplacement chimique comme paramètre structural. Un essai prometteur de détermination de l'angle glycosidique à partir du déplacement chimique du  $^{13}\text{C}$  a été mis en œuvre il y a quelques années (Greene, et al., 1995) mais son application reste très ponctuelle. Il est néanmoins possible de calculer le déplacement chimique théorique à partir d'une structure tridimensionnelle (Adamo and Barone, 1998; Case, 1998; Case, 2000; de Dios, et al., 1999; Giessner-Prettre and Pullman, 1987; Helgaker, et al., 1999; Scheurer, et al., 1999; Wilson, et al., 1999).

Il y a quelques années, le logiciel NUCHEMICS (Wijmenga, et al., 1997) prenant en compte l'équation générale du calcul de déplacement chimique a été développé. Il fut testé sur des ADN canoniques et de petites tiges-boucles d'ADN. Le calcul fait intervenir les paramètres semi-empiriques développés par Giessner-Pietre (Giessner-Prettre and Pullman, 1987). L'équation générale du calcul du déplacement chimique pour un noyau 1 dans un oligonucléotide comportant N atomes est définie par l'équation 1.

$$\delta_1 = \delta_{1,\text{intrin}} + \delta_{1,\text{ICA}} + \sum_{j=1}^N [\delta_{1,\text{jrc}} + \delta_{1,\text{jma}} + \delta_{1,\text{jE}} + \delta_{1,\text{jCT}}] \quad (1)$$

Le premier terme de l'équation représente le déplacement chimique de l'hydrogène 1 dans le fragment A en l'absence d'autres nucléotides, ce terme est purement empirique. Le second correspond à la contribution due aux changements conformationnels (liaisons, angles, angles de torsion) du résidu A en l'absence d'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 : le premier qui correspond à l'effet de courant de cycle ; son influence est essentielle dans les acides nucléiques. Le second paramètre représente l'anisotropie magnétique locale, due par

exemple à une distribution asymétrique des électrons dans un atome voisin au noyau 1. Alors que les deux paramètres précédents représentent l'anisotropie magnétique, le troisième terme rend compte de la polarisation issue du déplacement des électrons le long des liaisons chimiques voisinant le noyau. Enfin, le dernier paramètre représente le transfert de charge lors d'une liaison chimique.

Le calcul, à l'aide du logiciel NUCHEMICS, du déplacement chimique en retour est très rapide (temps CPU 2s sur SGI IRIX-6.5 à 300MHz). Il nécessite un fichier de coordonnées et un fichier de paramètres intégrant la topologie et les charges des atomes. Ce dernier fichier est construit à partir des paramètres d'entrée de CNS. Pour chaque structure jugée acceptable au niveau des contraintes RMN (aucune violation de distance NOE supérieure à 0,5Å et aucune violation d'angle dièdre supérieure à 5°), on calcule les déplacements chimiques théoriques. Ceux-ci sont automatiquement comparés aux déplacements chimiques expérimentaux, et les violations sont enregistrées. Une structure est jugée totalement acceptable quand elle ne montre aucune violation supérieure à 0,25ppm.

La forte correspondance entre les déplacements chimiques calculés et expérimentaux permet d'apporter une preuve supplémentaire de la véracité de la structure résultante. Ce point est crucial dans notre cas où les séquences étudiées peuvent présenter plusieurs structures possibles suivant leur mode de préparation. Enfin, il est remarquable de noter que le calcul de déplacement chimique en retour fonctionne particulièrement bien, y compris pour des structures éloignées des structures hélicoïdales. Ainsi, il rend compte, pour nos deux structures de kissing-complex de dSL1, des différences de déplacements chimiques que nous avons observées.

De plus, il est utile pour confirmer des motifs rencontrés dans les structures convergentes. Par exemple, un motif d'empilement de base est difficilement vérifiable par la seule mesure des effets NOE, la comparaison avec les déplacements chimiques théoriques et expérimentaux, qui sont principalement influencés par l'effet de courant de cycle, permet de confirmer expérimentalement ce type de motif stabilisant.

Les structures jugées non acceptables, au vu des violations de déplacements chimiques,

sont réintroduites dans le protocole de modélisation moléculaire. On effectue alors pour elles un nouveau recuit simulé. Le protocole est représenté sur la figure IV.14. La méthode d'implémentation du calcul du déplacement chimique permet une sélection plus drastique des structures jugées convergentes. Il est notamment intéressant pour les régions ayant une forte dynamique car il permet de s'approcher plus rigoureusement de la structure moyenne. La figure IV.15 présente un exemple de cet apport en terme de précision qu'est le calcul en retour du déplacement chimique.

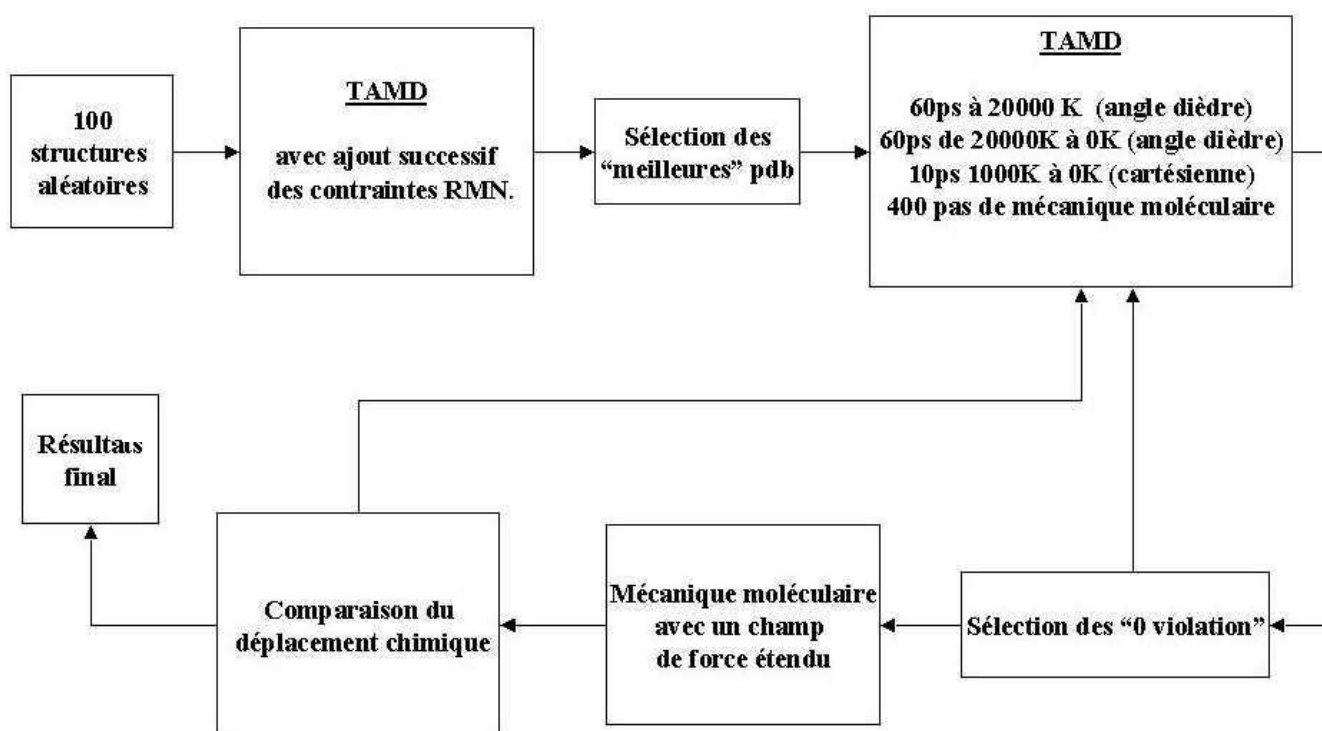


Figure IV.14 : Protocole de modélisation moléculaire incluant le calcul du déplacement chimique en retour.

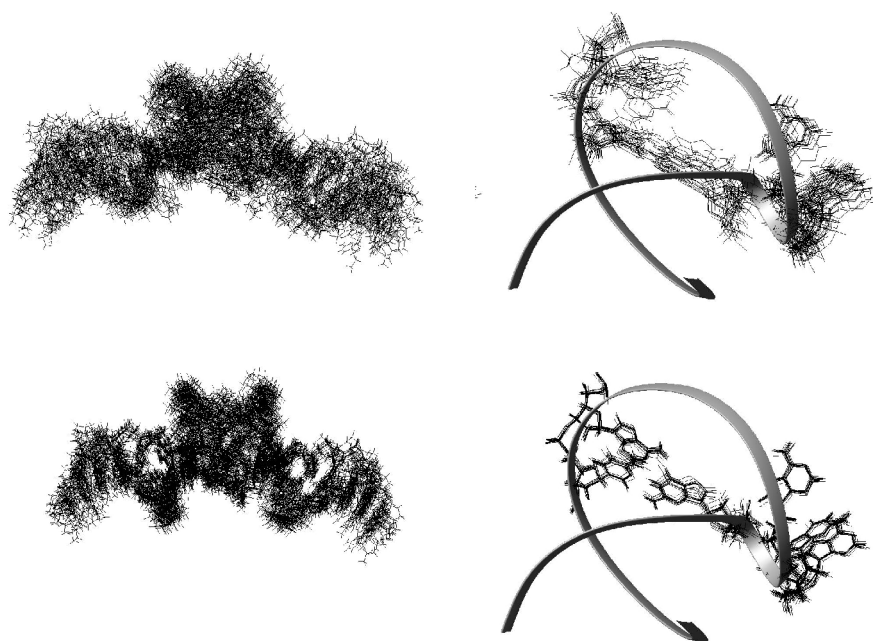


Figure IV.15 : Structure globale (gauche) et agrandissement de la région des adénines de jonction (droite) avant (haut) et après (bas) sélection à l'aide du déplacement chimique.

En haut de cette figure est représentée la superposition de 15 structures du dimère stable qui ont convergé. Elles présentent un RMSD global de  $4,55 \pm 1,72 \text{ \AA}$  et un RMSD sur les bases de jonctions de  $1,71 \pm 0,74 \text{ \AA}$ . Ces structures comportent des écarts de déplacement chimique supérieurs à  $0,25 \text{ ppm}$  et sont donc réintroduites dans le protocole de modélisation moléculaire (cf figure IV.13). En bas de cette figure sont présentées les 13 structures finales qui respectent les déplacements chimiques calculés. Elles observent une bien meilleure superposition et un RMSD global et sur les bases de jonction de :  $1,78 \pm 0,80 \text{ \AA}$  et  $0,50 \pm 0,19 \text{ \AA}$ .

Cette dernière figure montre comment la sélection supplémentaire par rapport aux déplacements chimiques du proton autorise une meilleure appréhension et une meilleure précision de la structure moyenne.

## 4. Analyses structurales

### 4.1 Organisation globale

La structure du dimère stable dSL1 est également une structure sous forme de kissing-complex. La figure IV.16 représente la structure moyenne de ce dimère. Comme dans le

dimère instable, on repère trois zones distinctes en appariement de type Watson-Crick. Deux de ces zones sont identiques car symétriques l'une à l'autre. Elles comprennent les résidus  $C_1$  à  $G_7$  et  $C_{17}$  à  $G_{23}$  de chaque monomère. Les paramètres hélicoïdaux sont élucidés à l'aide du logiciel CURVES (Lavery and Sklenar, 1988; Lavery and Sklenar, 1989; Lavery and Sklenar, 1990; Ravishanker, et al., 1989). Ils montrent des valeurs classiquement rencontrées pour un ADN de type B. Ces deux hélices présentent chacune sept appariements Watson-Crick, ce qui en fait des zones de forte stabilité.

La tige centrale, dans ce dimère, présente 6 paires de bases :  $G_{10}C_{15}'$  ;  $C_{11}G_{14}'$  ;  $G_{12}C_{13}'$  ;  $C_{13}G_{12}'$  ;  $G_{14}C_{11}'$  et  $C_{15}G_{10}'$ . Son organisation n'est pas hélicoïdale, elle représente des plans inclinés d'environ  $15^\circ$  vers la gauche par rapport à l'axe de la tige (cf figure IV.17).

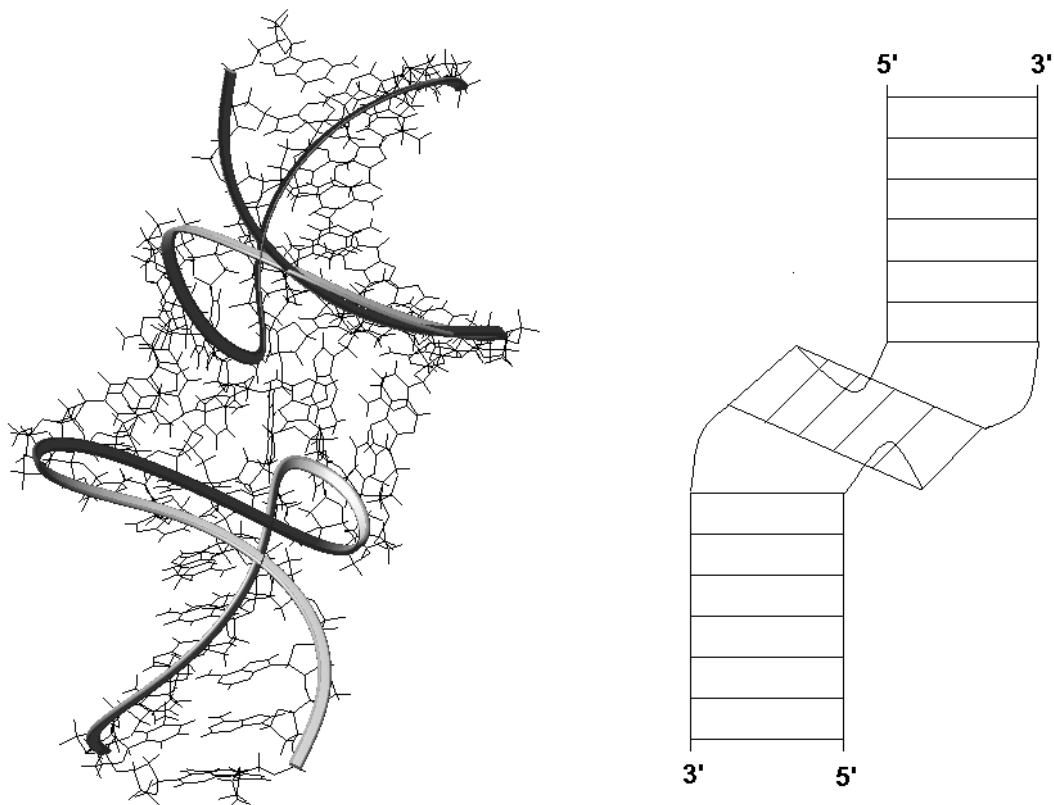


Figure IV.16 : Structure globale du dimère stable de dSL1. La figure de gauche représente la structure moyenne et celle de droite montre schématiquement l'organisation des appariements et du squelette phosphodiester.

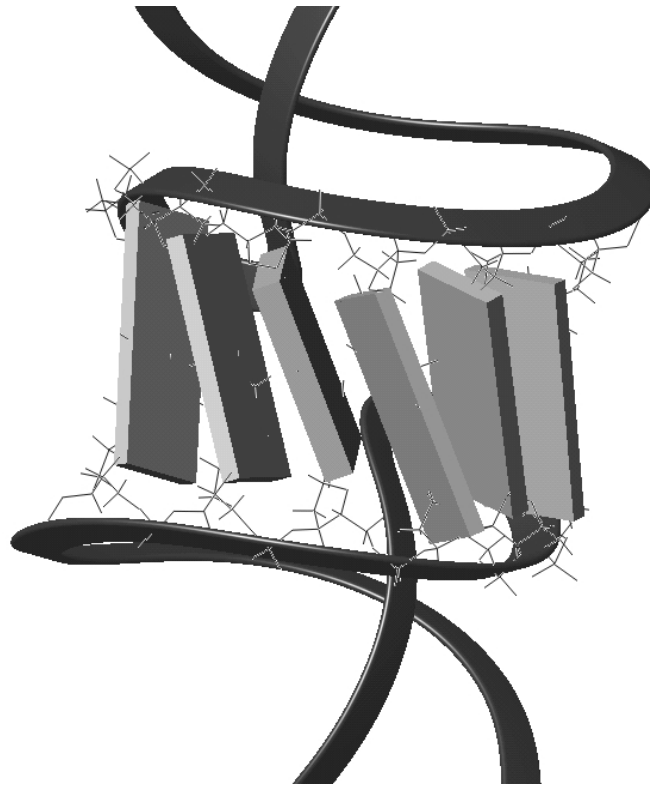


Figure IV.17 : Organisation des plans des paires de bases de la tige centrale.

Ces plans des paires de bases sont enroulés autour de l'interaction boucle-boucle en suivant un demi-cercle imposé par le squelette phosphodiester de la macromolécule. La structure résultante est régulière et ne présente pas de distorsion importante. Le fait que, dans cette structure, l'appariement de l'interaction boucle-boucle soit complet explique les différences de stabilité des deux dimères.

Le squelette phosphodiester montre de fortes distorsions. Si nous regardons les valeurs des angles du squelette (cf figure IV.18) nous nous apercevons qu'elles se situent uniquement pour les résidus de jonction (8, 9, 10, 15 et 16) et sont assez régulières le long de la séquence. En fait, le backbone présente une boucle interne de trois résidus ( $A_8$ ,  $A_9$  et  $G_{10}$ ). La structure présente donc une organisation globale de son squelette fortement distordue, avec des angles ne présentant pas réellement d'abrupte variation assurant ainsi une certaine stabilité à l'édifice.

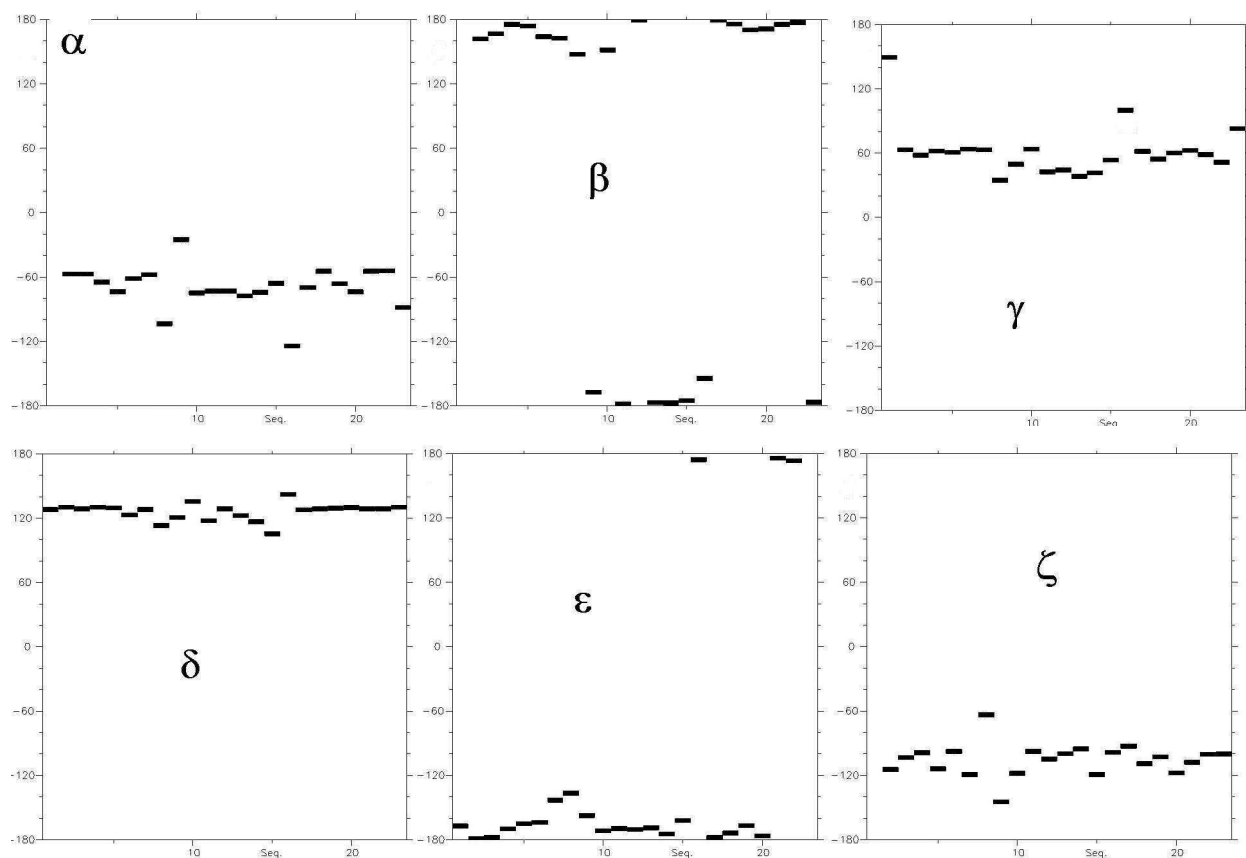


Figure IV.18 : Variation des angles du squelette phosphodiester en fonction de la séquence (abscisse)

## 4.2 Etude des motifs stabilisant la structure

Les résidus non-appariés, dans cette structure, sont au nombre de trois et concernent seulement les adénines de jonction  $A_8$ ,  $A_9$  et  $A_{16}$ . La figure IV.19 représente leur organisation. La paire de bases  $A_8A_8$  observée dans le dimère instable disparaît. Cette base est toujours placée sous l'interaction boucle-boucle. La base  $A_9$  est, là encore, en position d'empilement (cf figure IV.20) avec, cette fois, la paire de bases  $G_{10}C_{15}$  (au lieu de  $G_{14}C_{11}$  dans le dimère instable).

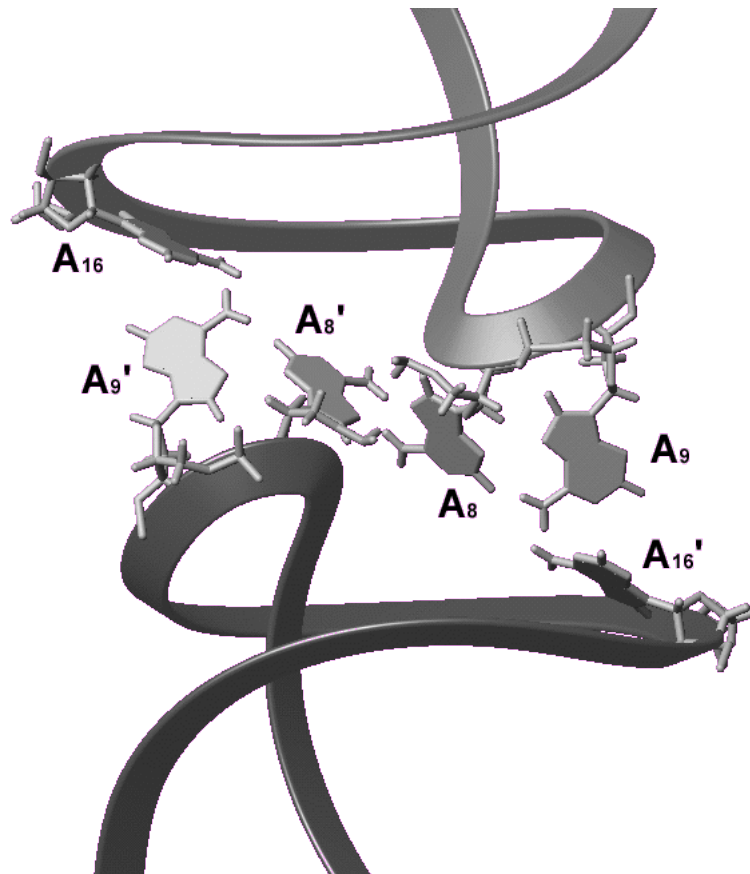


Figure IV.19 : Organisation des adénines de jonction dans le dimère stable de dSL1 (vue du dessous de l'interaction boucle-boucle).

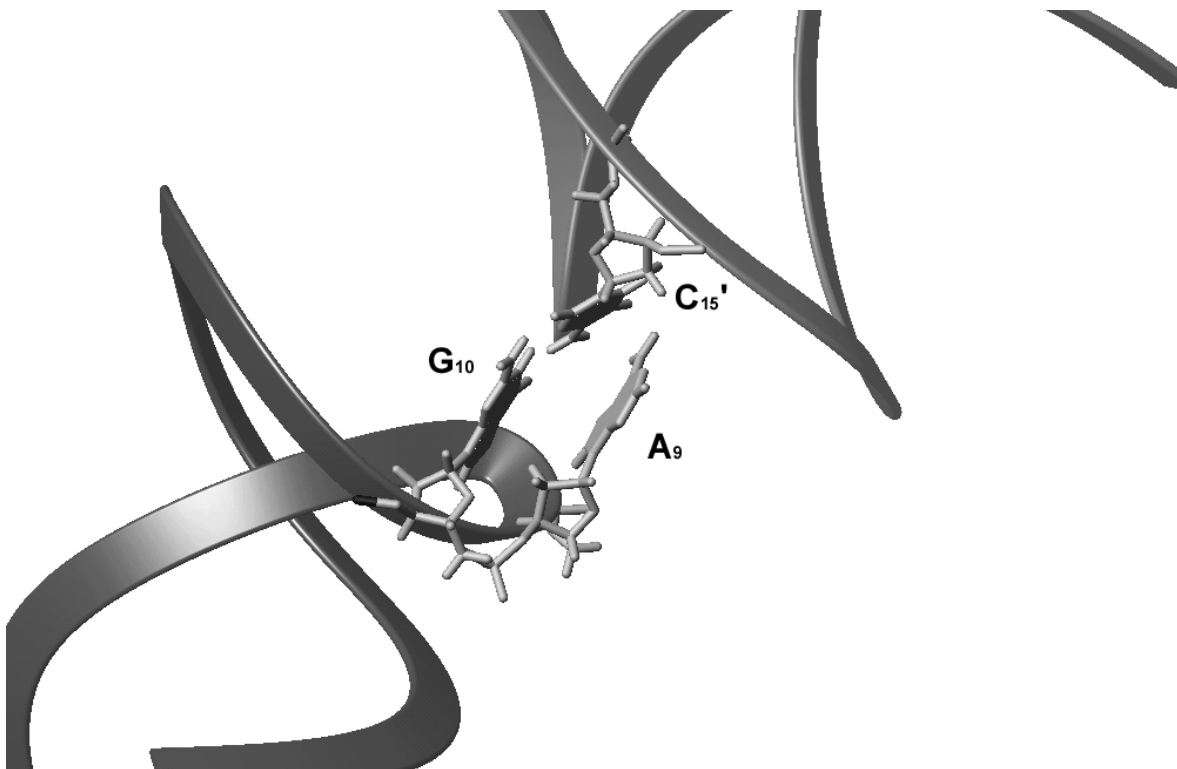


Figure IV.20 : Représentation du motif d'empilement impliquant la paire de bases inter-brin  $G_{10}C_{15}'$  et l'adénine  $A_9$ .



En plus de ce stacking, certaines liaisons hydrogène impliquant ces adénines de jonction apparaissent plus ou moins dans l'ensemble des structures convergentes. Les liaisons hydrogène ont été définies par une proximité inférieure à 2,4Å et un angle, entre l'atome donneur et accepteur de liaison hydrogène, inférieur à 35° (Billeter, et al., 1990). Le tableau IV.4 résume ces liaisons et leur conservation dans l'ensemble des structures convergentes. Toutes ces liaisons sont inter-brin, cependant il faut noter qu'aucune n'a été observée dans les expériences RMN faites dans l'eau légère. Ces liaisons sont donc fortement échangées avec le solvant.

| Hydrogène          | Accepteur            | Pourcentage |
|--------------------|----------------------|-------------|
| A <sub>8</sub> H61 | A <sub>8</sub> ' N6  | 53.8%       |
| A <sub>8</sub> H61 | A <sub>8</sub> ' N7  | 23.1%       |
| A <sub>9</sub> H62 | A <sub>16</sub> ' N1 | 30.8%       |
| A <sub>9</sub> H62 | A <sub>16</sub> ' N6 | 92.3%       |

Tableau IV.4 : Résumé et statistiques des liaisons hydrogène impliquant les adénines de jonction.

## 5. Comparaison avec différentes structures analogues

### 5.2 Comparaison avec le dimère instable

La première comparaison structurale que l'on peut faire concerne le dimère instable de dSL1. Bien évidemment, beaucoup de points communs existent entre ces deux conformères. Dans les deux cas, nous sommes en présence d'un dimère de type kissing-complex. Dès lors on peut comparer des zones quasiment identiques.

Au niveau des deux hélices intra-brin, elles comprennent les mêmes résidus (C<sub>1</sub> à G<sub>7</sub> et C<sub>17</sub> à G<sub>23</sub>) et sont organisées en hélices de type B. La figure IV.21 représente la superposition de ces tiges. Elles sont quasiment identiques et présentent donc une stabilité voisine. La différence entre ces deux structures ne se situe pas au niveau de ces zones.

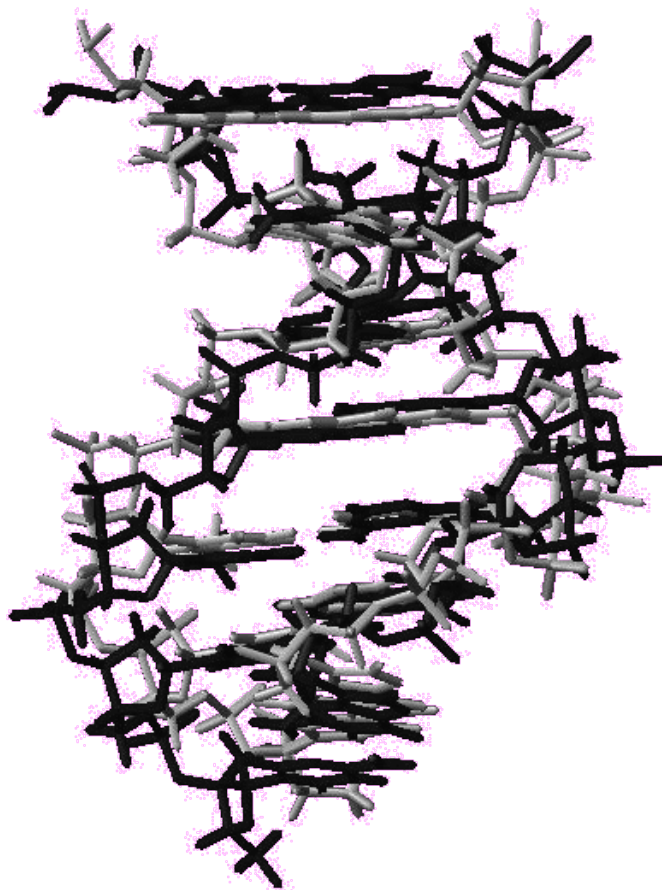


Figure IV.21 : Superposition des deux tiges moyennes intra-brin, en noir la tige du dimère stable et en gris celle du dimère instable.

La zone de l'interaction boucle-boucle n'est pas identique dans les deux structures. Dans la structure du dimère instable, la tige centrale comprend 4 paires de bases concernant les résidus de  $C_{11}$  à  $G_{14}$  et  $C_{11}'$  à  $G_{14}'$ , tandis que dans la structure du dimère stable l'interaction boucle-boucle est plus large puisqu'elle comprend 6 paires de bases ( $G_{10}$  à  $C_{15}$  et  $G_{10}'$  à  $C_{15}'$ ). Le fait d'obtenir deux paires de bases supplémentaires dans le dimère thermodynamique explique certainement cette différence de stabilité entre les deux dimères.

La différence du nombre de paire de bases de la tige centrale explique également le changement d'organisation globale du squelette phosphodiester. Dans le dimère instable, le squelette phosphodiester présente d'abruptes variations angulaires avec une paire de bases non canonique  $A_8A_8$  qui vient cadencasser cette interaction et apporter un caractère métastable à ce conformère. De plus, les axes des deux tiges sont alignés.

Dans le dimère stable, le squelette phosphodiester présente une boucle symétrique englobant les résidus  $A_8$ ,  $A_9$  et  $G_{10}$ . Cette boucle se forme avec de légères variations des angles du squelette phosphodiester. Cette régularité permet de conserver une bonne stabilité du dimère tout en autorisant la présence de deux paires de bases supplémentaires. Dans l'organisation globale du conformère qui en découle les axes des deux tiges sont parallèles mais pas confondus, en fait la boucle décale les deux tiges. La figure IV.22 schématise ces deux organisations.

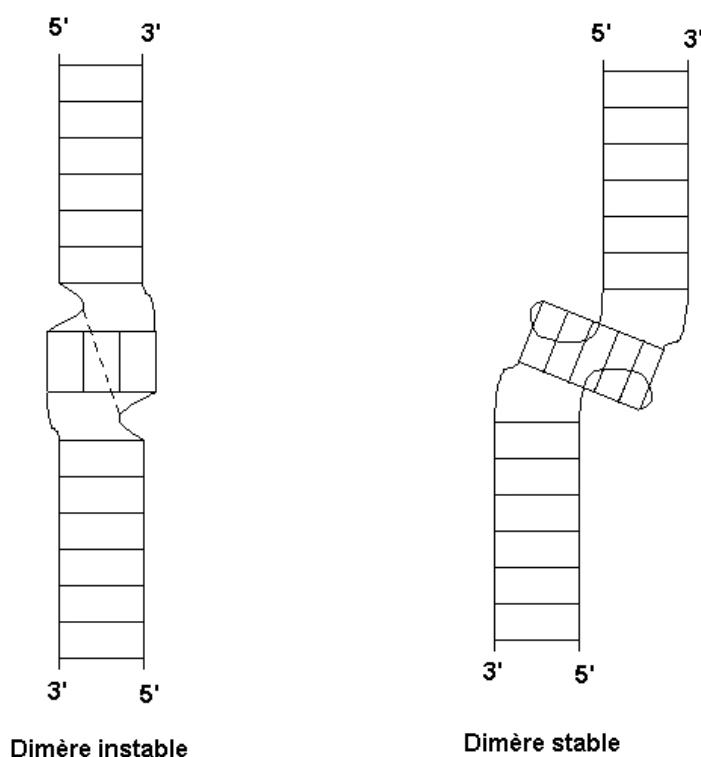


Figure IV.22 : Schéma représentant l'organisation des deux structures en kissing-complex du dimère dSL1. Dans le dimère cinétique, à gauche, les axes des deux tiges sont alignés tandis que dans le dimère instable ils sont parallèles.

## 5.2 Comparaison avec les kissing-complex d'ARN

Il importe de comparer la structure du dimère stable avec différentes autres structures de kissing-complex d'ARN. La première structure à envisager concerne celle du dimère instable SL1, mutée au niveau de la tige, élucidée par l'équipe du professeur James (Mujeeb, et al., 1998) et dont nous avons déjà décrit la structure dans le chapitre III-6.2 de la première partie. Cette structure présente l'analogie intéressante de posséder la même

séquence et le même nombre de résidus au niveau de l'interaction boucle-boucle.

Dans cette structure l'adénine A<sub>8</sub> se situe sous la tige centrale et la base A<sub>9</sub> est expulsée pour s'empiler entre les bases A<sub>8</sub> et C<sub>17</sub> de l'autre brin. Ce motif induit une forte distorsion du squelette phosphodiester, avec notamment une zone de "close-contact" entre les sommets de ces boucles dont nous avons déjà discuté (cf figure III.19 du chapitre III-6.2 de la première partie).

Dans notre structure, la base A<sub>8</sub> se situe là aussi sous l'interaction boucle-boucle et la base A<sub>9</sub>, dirigée vers l'autre monomère, se trouve en situation de stacking avec la paire de bases inter-brin G<sub>10</sub>C<sub>15</sub>. Nous pensons que c'est l'appariement total de la séquence autoccomplémentaire de dSL1 qui induit une torsion du squelette phosphodiester.

Cette hypothèse est d'ailleurs confirmée par une autre étude structurale d'interaction boucle-boucle ARN. Il s'agit du kissing-complex RNA Ii/RNA Iii du plasmide ColE1 chez *Escherichia coli* (Lee and Crothers, 1998; Marino, et al., 1995). Dans cette structure la torsion de l'hélice a été particulièrement décrite (Marino, et al., 1995) et présente de forte analogie avec la notre. La figure IV.23 représente cette comparaison.

Dans leur cas, des expériences sur gels électrophorétiques confirment l'existence de cette distorsion du squelette phosphodiester (Marino, et al., 1995). Les auteurs concluent leur article en supposant que la torsion est nécessaire pour obtenir l'appariement total de l'interaction boucle-boucle et que le stacking des 7 paires de bases compense, énergétiquement, largement cette interaction.

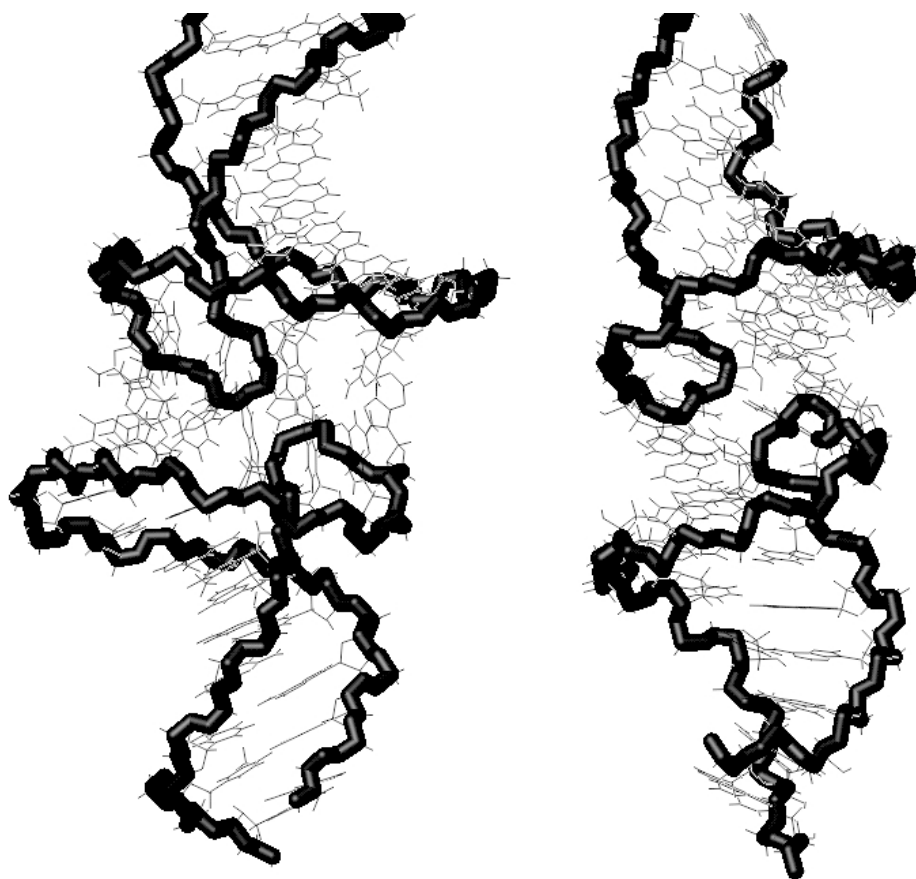


Figure IV.23 : Représentation des profils du squelette phosphodiester dans la structure du dimère stable dSL1 (à gauche) et dans la structure du kissing-complex RNA Ii/RNA Iii du plasmide ColE1.

## 6. Mécanisme de transition supposé entre les deux dimères

Une question importante reste en suspend quant à l'étude des deux dimères de dSL1 : pourquoi avons-nous, pour le dimère stable, une structure de type kissing-complex et non un dimère étendu?

Lors de la formation du dimère stable de dSL1, l'échantillon est chauffé à 85°C pendant 3 minutes, la concentration de l'ADN est de 2,7mM. La molécule s'organise sous la forme d'un simple brin désordonné. Le refroidissement est très lent (6 heures) de façon à obtenir le dimère thermodynamiquement le plus stable.

Le seul point qui justifie le fait que nous n'avons pas un dimère sous forme de double

hélice est la présence des adénines non appariées A<sub>8</sub>, A<sub>9</sub> et A<sub>16</sub>. En effet, quelle que soit la structure envisagée ces résidus jouent un rôle essentiel dans la stabilisation des structures.

Dans la structure du dimère stable SL1 ces adénines s'organisent en fermeture éclair, minimisant la distorsion de l'hélice A (se référer au chapitre III-4.2 de la première partie). On pourrait s'attendre à une telle organisation pour l'ADN dSL1 mais il n'en est rien. La raison vient de la présence du groupement 2'OH présent uniquement dans les ARN. En effet, la présence d'un seul groupement hydroxyle sur le cycle ribose implique, dans des conditions de basse salinité, la formation d'hélice de type A (Ban, et al., 1994). Si on assimile une hélice à un cylindre, le rayon de celui-ci est beaucoup plus important dans le cas d'une hélice A que dans le cas d'une hélice B. Par conséquent, les résidus impliqués dans un motif en fermeture éclair pour une hélice A, se retrouveraient plus fortement encastrés dans une hélice B.

A notre connaissance, une seule étude en fermeture éclair existe pour l'ADN (Shepard, et al., 1998). La séquence étudiée par cette équipe est, d(GCGAAAGCT). La structure cristallographique montre un motif d'empilement de quatre résidus en fermeture éclair (cf figure IV.24). Néanmoins, ce motif montre de fortes torsions de l'hélice B avec notamment un déroulement de celle-ci. En solution, cette séquence apparaît, non pas comme un dimère, mais sous forme de tige-boucle (Hirao, et al., 1989; Hirao, et al., 1992; Tanikawa, et al., 1991).

Si nous supposons que notre dimère stable offre lui aussi un tel motif, il ne fait pas de doute que les paires de bases adjacentes seraient fortement déstabilisées voire brisées. Or, celles-ci sont essentielles pour la stabilité du dimère thermodynamique. La structure du dimère stable représente pour le fragment d'ADN dSL1 la seule méthode pour appairier toutes les paires de bases disponibles.

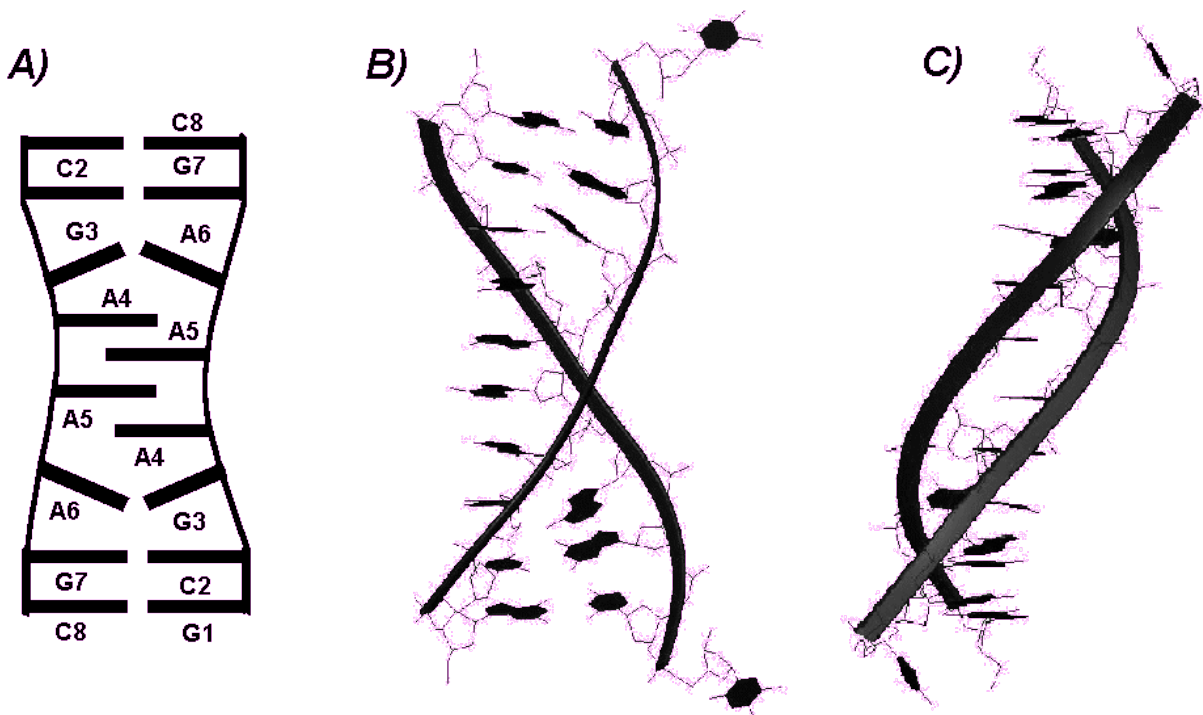


Figure IV.24 : Organisation du motif en fermeture éclair du fragment d(GCGAAAGCT). A) Schéma de l'organisation. B) Vue de face de la structure cristalline. C) Vue de profil de cette structure. Le motif en fermeture éclair implique dans le cas l'ADN un déroulement de l'hélice.

Par conséquent, nous supposons que le passage du dimère instable au dimère stable se fait par un réarrangement au niveau de l'interaction boucle-boucle. La paire de bases métastable  $A_8A_8$  se brise, ce qui permet les appariements supplémentaires des résidus  $G_{10}$  et  $C_{15}$  de chaque brin. Lors de cette évolution, les angles du squelette phosphodiester de la région précédant l'interaction boucle-boucle, retrouvent des valeurs "classiques" pour un ADN de type B tout en formant une boucle minimisant la distorsion de l'hélice.

## 7. Conclusions

Après l'étude RMN du dimère instable de dSL1 nous nous sommes intéressés à l'espèce thermodynamiquement la plus stable de ce fragment d'ADN. Celle-ci représente un dimère.

Les spectres RMN montrent des différences intéressantes confirmant l'existence d'une seconde structure pour dSL1. Tout d'abord, les études dans l'eau légère permettent

d'affirmer que tous les appariements possibles existent dans cette structure, ce qui n'est pas le cas du dimère instable. Les spectres dans l'eau lourde montrent des déplacements chimiques et des corrélations NOESY différents. Cependant, certaines caractéristiques structurales, observées dans le dimère instable et correspondant à de forts indices de l'existence d'un dimère de type kissing-complex, sont également observées pour le dimère stable. Enfin, le spectre du phosphore est différent de celui obtenu pour le dimère cinétique. Il présente un massif englobant la quasi-totalité des résonances des phosphores indiquant que le squelette phosphodiester ne présenterait que peu de variations tout au long de la séquence.

Comme précédemment, l'étape de modélisation moléculaire est cruciale car l'ensemble des renseignements structuraux ne nous permet pas de trancher sur un modèle de structure secondaire. Lors de la simulation, aucune structure correspondant à un dimère stable n'a convergé, ce qui n'est pas le cas du dimère stable. Nous avons utilisé, lors de ce calcul, une méthode de calcul en retour du déplacement chimique pour sélectionner et valider les structures convergentes. Ce protocole de calcul, qui est très facile à mettre en œuvre et à automatiser, permet de confirmer l'existence des deux structures de dimère pour dSL1 et de s'approcher plus rigoureusement de la structure moyenne.

La structure tridimensionnelle du dimère stable de dSL1 est un homodimère symétrique de type kissing-complex. Les zones hélicoïdales intra-brin sont de type ADN-B et quasi-identique à celles identifiées pour le dimère instable. Ces régions, stables, permettent de cadénasser l'interaction boucle-boucle. La zone centrale présente 6 appariements GC de type Watson-Crick, soit 2 appariements de plus que la structure du dimère instable, ce qui explique la différence de stabilité des deux dimères. Les adénines de jonction, qui restent les seuls résidus non-appariés de la séquence, observent sur l'ensemble des structures bon nombre de liaisons hydrogène inter-brin. Celles-ci ne sont pas confirmées par les études RMN dans l'eau légère, ce qui indique qu'elles doivent être fortement échangées avec le solvant.

La présence de deux appariements supplémentaires a une autre conséquence : le squelette phosphodiester doit compenser l'encombrement plus important de la tige



centrale en formant une boucle au niveau des résidus A<sub>8</sub>, A<sub>9</sub> et G<sub>10</sub>. Cette distorsion n'entraîne pas de variation importante du squelette phosphodiester préservant la stabilité de la région. Rajoutons, enfin, que cette observation est confirmée par la comparaison avec d'autres études RMN de kissing-complex ARN.

Il est remarquable de constater que, quelle que soit la salinité, on a toujours un dimère de type kissing-complex et jamais de type dimère étendu. La présence des résidus A<sub>8</sub>, A<sub>9</sub> et A<sub>16</sub>, qui sont les seules anomalies séquentielles, explique la formation de cette deuxième structure kissing-complex. En effet, pour obtenir un dimère linéaire, comme dans le cas de l'ARN, il faudrait observer un empilement de ces adénines. Une seule structure cristallographique d'ADN en fermeture éclair a été élucidée. Elle présente un fort déroulement laissant supposer que, dans notre cas, les bases adjacentes seraient affaiblies ce qui ne serait pas cohérent avec un dimère stable. La structure du dimère stable, avec sa distorsion de son squelette, présente la meilleure façon d'assembler toutes les paires de bases possibles de ce fragment.

## V. Conclusions de la seconde partie

Dans cette partie, nous nous sommes intéressés à l'analogue désoxyribose de la séquence ARN SL1-23mer de VIH-1<sub>Lai</sub>. Le fragment d'ARN présente trois structures isolables : le monomère (lorsque l'on empêche la dimérisation spontanée par mutation), le dimère instable sous forme de kissing-complex et le dimère stable qui se présente sous l'apparence d'un dimère étendu.

Dans l'étude ADN du monomère, nous avons, par analogie au travail précédent, muté la base G<sub>12</sub> par une adénine. Par conséquent, l'appariement inter-brin est fortement inhibé. Malgré tous nos efforts, nous n'avons pas trouvé de conditions permettant de stabiliser la structure de cet échantillon qui s'apparente à un simple brin désordonné. Indéniablement, le groupement 2'OH, qui différencie l'ADN de l'ARN, joue un rôle essentiel dans la stabilisation de la tige-boucle. Le groupement hydroxyle du cycle ribose est impliqué dans de nombreuses liaisons hydrogène qui stabilisent les hélices A d'ARN. De plus, l'étude de dynamique moléculaire avec solvant explicite effectuée pour le monomère SL1 a montré bon nombre de ces liaisons y compris dans la région de la boucle. Dans ce fragment dSL1\* d'ADN, la disparition de ces liaisons est préjudiciable à la stabilisation de l'édifice tridimensionnel. La tige n'étant pas assez stable pour compenser la forte dynamique de la boucle.

Le fragment dSL1 dimérise sous deux espèces distinctes, la première des deux structures représente le dimère cinétique et l'autre le dimère thermodynamique. Suivant le mode de préparation nous pouvons isoler un des deux types de dimères. L'étude RMN de l'espèce instable montre un dimère comportant des caractéristiques structurales rarement rencontrées dans les études d'ADN. La structure tridimensionnelle se présente sous la forme globale d'un homodimère symétrique de type kissing-complex. Le squelette phosphodiester est fortement tordu en aval et en amont de l'interaction boucle-boucle qui ne présente que 4 paires de bases GC. Enfin, la présence d'une paire de bases inter-brin non-canonique A<sub>8</sub>A<sub>8</sub> stabilisée par une liaison hydrogène avec le résidu A<sub>9</sub> qui s'empile

avec le résidu C<sub>15</sub> du brin opposé, offre une structure métastable du dimère. Cette structure représente, à notre connaissance, la première structure tridimensionnelle de type kissing-complex pour un ADN. Elle vient donc élargir le catalogue des structures secondaires possibles pour les oligonucléotides. La compréhension complète du rôle que prend le groupement 2'OH dans cette structure ne peut se faire qu'avec une simulation de dynamique newtonienne en présence d'eau. Celle-ci est en cours de réalisation.

La dernière structure à élucider correspond au dimère stable de dSL1. Les spectres RMN montrent des différences intéressantes confirmant l'existence d'une seconde structure pour ce fragment. On pourrait s'attendre à observer un dimère de type étendu pour cette structure mais il n'en est rien. Le dimère thermodynamique de dSL1 représente, là encore, un homodimère symétrique de type kissing-complex. La présence supplémentaire de deux paires de bases GC au niveau de l'interaction boucle-boucle explique à elle seule la différence de stabilité entre les deux dimères. Elle a une autre conséquence : le squelette phosphodiester doit minimiser l'encombrement des 6 appariements centraux en formant une distorsion du squelette phosphodiester des résidus A<sub>8</sub>, A<sub>9</sub> et G<sub>10</sub>. Les adénines de jonction, qui restent les seuls résidus non-appariés, forment des liaisons hydrogène inter monomères stabilisant l'édifice tridimensionnel. Le mécanisme de transition entre les deux structures de dimère de dSL1 reste le dernier point à élucider. Il est en cours de réalisation dans notre équipe.

# *Conclusions générales*

## Conclusions générales

La dimérisation des ARN rétroviraux est impliquée dans la régulation des étapes du cycle viral. Les travaux de Muriaux *et al* (Muriaux, et al., 1995) et Laughrea *et al* (Laughrea and Jette, 1994) ont démontré la participation de la tige-boucle SL1 dans l'initiation à la dimérisation de l'ARN génomique. La séquence SL1-23mer, plus courte, conserve l'ensemble des propriétés de SL1 (Theilleux-Delalande, 1998) à savoir : le monomère peut dimériser sous deux formes, l'une instable et l'autre stable ; le dimère instable évoluant vers la forme stable sous l'influence de la température ou de l'interaction avec la Ncp7.

Dans un premier temps, le but de l'étude était de décrire en détail les trois structures SL1 d'ARN impliquées dans le phénomène d'initiation à la dimérisation : le monomère, le dimère stable et le dimère instable.

Le fragment SL1 dimérise spontanément. Pour étudier le monomère, nous l'avons donc modifié en mutant la base G<sub>12</sub> en A<sub>12</sub>. La structure globale s'apparente à une épingle à cheveux. La tige, composée de 7 paires de bases s'organisant en hélice A, permet de stabiliser la boucle. Les adénines A<sub>8</sub>, A<sub>9</sub> et A<sub>16</sub>, situées à l'entrée et à la fin de la boucle, montrent un relatif degré de structuration et forment un empilement avec leurs résidus voisins en 3'. Ce motif est situé perpendiculairement à l'axe de l'hélice. Par conséquent, les résidus auto-complémentaires GCGCGC de la boucle sont dirigés vers l'extérieur et exposés au solvant. Ce dernier point avait déjà été supposé lors d'une étude d'inhibiteur d'oligonucléotides anti-dimérisation (Lodmell, et al., 1998) qui concluait sur le fait que le monomère était potentiellement la cible la plus intéressante pour une stratégie anti-sens anti-dimérisation.

Le dimère stable adopte une conformation étendue reposant sur la mise en jeu d'interactions intermoléculaires tout au long de la séquence. Cette structure développe deux "bulges" symétriques impliquant les adénines centrales A<sub>8</sub>, A<sub>9</sub> et A<sub>16</sub> non appariées. Ce motif, organisé en fermeture éclair, minimise les distorsions du squelette occasionnées par le mésappariement local. De plus, il favorise l'empilement des bases nucléiques, ce qui permet un gain de stabilité supporté par les interactions de "stacking".

Différentes études ont mis en évidence que le dimère instable adopte une conformation en

“kissing-complex” où deux tiges-boucles d’ARN SL1 interagissent *via* leur boucle respective. L’équipe de T. James a décrit l’organisation d’un complexe boucle-boucle issu d’une séquence mutée de SL1-23mer (Mujeeb, et al., 1998). Cette structure repose sur le développement d’une interaction de stacking intermoléculaire impliquant les résidus A<sub>8</sub>, A<sub>9</sub> et C<sub>17</sub>. Cette organisation conduit à la rupture de l’appariement G<sub>7</sub>C<sub>17</sub>. Or, cette dernière paire de bases a été plusieurs fois identifiée dans le kissing-complex de SL1-23mer (Girard, 2000; Takahashi, et al., 2000; Theilleux-Delalande, et al., 2000). De plus, l’étude approfondie de la structure révèle de nombreuses proximités entre atomes inférieures aux rayons de van der Waals et notamment entre les atomes O1P, chargés négativement, des résidus A<sub>9</sub> de chaque brin. Sur la base de ces observations, nous avons décidé de poursuivre notre étude structurale du dimère instable. Afin de faciliter l’attribution des raies de résonance et d’augmenter le nombre de données structurales expérimentales, nous étudions un échantillon d’ARN marqué <sup>15</sup>N/<sup>13</sup>C. L’investigation structurale est en cours.

Un des points intéressants que nous avons voulu élucider est le rôle qu’apportait le groupement 2’OH dans ces différentes structures d’ARN. Pour cela, nous avons utilisé une séquence analogue à celle de SL1 mais totalement désoxyribose : dSL1. Nous nous sommes alors intéressés aux trois structures que pouvait potentiellement prendre cette séquence.

Dans la même optique que le travail précédent, nous avons muté la base G<sub>12</sub> en A<sub>12</sub> pour étudier le monomère de dSL1. Malgré tous nos efforts, nous n’avons pas trouvé de conditions permettant de stabiliser la structure de cet échantillon qui s’apparente à un simple brin désordonné. Par conséquent, nous pouvons affirmer sans équivoque que le groupement 2’OH, qui différencie l’ADN de l’ARN, joue un rôle essentiel dans la stabilisation de la tige-boucle. Ce groupement est impliqué dans de nombreuses liaisons hydrogène et sa disparition est préjudiciable à la stabilisation de l’édifice tridimensionnel, la tige n’étant pas assez stable pour compenser la forte dynamique de la boucle.

Le fragment dSL1 dimérise sous deux conformations distinctes et isolables. La première structure que nous avons étudiée représente le dimère cinétique de cette séquence. Sa séquence adopte la forme globale d’un homodimère symétrique de type kissing-complex. Le squelette phosphodiester est fortement tordu en aval et en amont de l’interaction boucle-boucle qui ne comporte que 4 paires de bases GC. La présence d’une paire de bases inter-brin non-canonique A<sub>8</sub>A<sub>8</sub>, stabilisée par une liaison hydrogène avec le résidu A<sub>9</sub> s’empilant avec le résidu C<sub>15</sub> du brin opposé, offre un aspect métastable à la structure de ce dimère. Cette structure est la première structure de kissing-complex

rencontrée pour un ADN. Elle vient donc élargir le catalogue des structures possibles de l'ADN. Nous envisageons d'effectuer une simulation de dynamique newtonienne en présence d'eau pour élucider le rôle que prend le groupement 2'OH dans cette structure.

La dernière structure que prend le fragment dSL1 est celle du dimère stable. Étonnamment, celle-ci représente encore un kissing-complex. La présence de deux paires de bases supplémentaires au niveau de l'interaction boucle-boucle explique la différence de stabilité entre les deux dimères. Elle a aussi une autre conséquence : le squelette phosphodiester forme une boucle symétrique au niveau des résidus A<sub>8</sub>, A<sub>9</sub> et G<sub>10</sub>. Cette distorsion minimise l'encombrement des 6 appariements centraux. Nous n'avons malheureusement pas eu le temps d'étudier le mécanisme de transition entre ces deux dimères, ce dernier point étant en cours de réalisation au laboratoire.

Il est intéressant de noter que pour le virus HIV-1<sub>Lai</sub>, la transition entre le dimère instable (de type kissing-complex) et le dimère stable (de type étendu), est supposée correspondre à la maturation du rétrovirus (Muriaux, et al., 1996). De nombreuses équipes cherchent à inhiber ce processus de maturation en agissant sur l'ARN ou sur la protéine de nucléocapside NCp7. Le fragment dSL1 présente la particularité de transiter d'un kissing-complex vers un autre kissing-complex. Il représente donc le but espéré pour le fragment ARN et peut servir de modèle d'étude. La première perspective de cette étude consiste à élucider le mécanisme de transition des deux dimères. De plus, l'étude de l'interaction de dSL1 avec la protéine NCp7 peut apporter des éléments intéressants sur le mécanisme de transition et les structures résultantes. Enfin, une étude du fragment ADN-ARN de dSL1/SL1 peut révéler des caractéristiques biologiques et structurales très intéressantes pour la dimérisation. C'est notamment l'optique qui a été choisie pour inhiber le complexe TAR/TAT du virus VIH-1 (Boiziau, et al., 1999; Collin, et al., 2000).

D'une façon générale l'ensemble des structures que nous avons élucidé constitue une base importante dans des stratégies de "drug-design" impliquant le développement d'inhibiteurs susceptibles d'interférer avec les sites d'interactions ARN/ARN ou ARN/NCp7.

# *Matériels et méthodes*



# Matériels et méthodes

## **Monomère SL1 :**

Les études RMN ont été réalisées à 500MHz sur un spectromètre AMX BRUKER. 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 (pH=7,0). La modélisation moléculaire c'est faite sur une station calcul Silicon Graphics (IRIX 6.5 à 300MHz et 128Mo de RAM) avec le logiciel XPLOR version 3.851.

## **Dimère stable SL1 :**

La dénaturation des ARN est suivie par UV à 258 nm sur un spectromètre VARIAN *Cary 100 scan* couplé à un régulateur de température *Cary* à effet Peltier. Les études RMN ont été réalisées à 500MHz sur un spectromètre AMX BRUKER. 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). La modélisation moléculaire a été faite sur une station calcul Silicon Graphics (IRIX 6.5 à 300MHz et 128Mo de RAM) avec les logiciels DYANA et XPLOR version 3.851.

## **Monomère dSL1 :**

Les études RMN ont été réalisées à 500MHz sur un spectromètre AMX BRUKER et à 600MHz sur un spectromètre VARIAN *INOVA Unity*. Les spectres dans H<sub>2</sub>O et D<sub>2</sub>O ont été obtenus à partir d'un échantillon de 600µL d'ADN à 2,3mM. Les études se sont faites pour différentes valeurs de pH (4,8 ; 5,5 ; 6,3 ; 6,8 ; 7,2) avec ajout de différents sels (NaCl ; MgCl<sub>2</sub> ; NaClO<sub>4</sub> ; KCl) à différentes concentrations (10, 50, 100 et 200 mM).

## **Dimère instable et dimère stable dSL1 :**

La dénaturation des ADN est suivie par UV à 256 nm sur un spectromètre VARIAN *Cary 100 scan* couplé à un régulateur de température *Cary* à effet Peltier. La dénaturation est suivie à la vitesse de 0,5° par minute. Les études RMN ont été réalisées à 600MHz sur un spectromètre

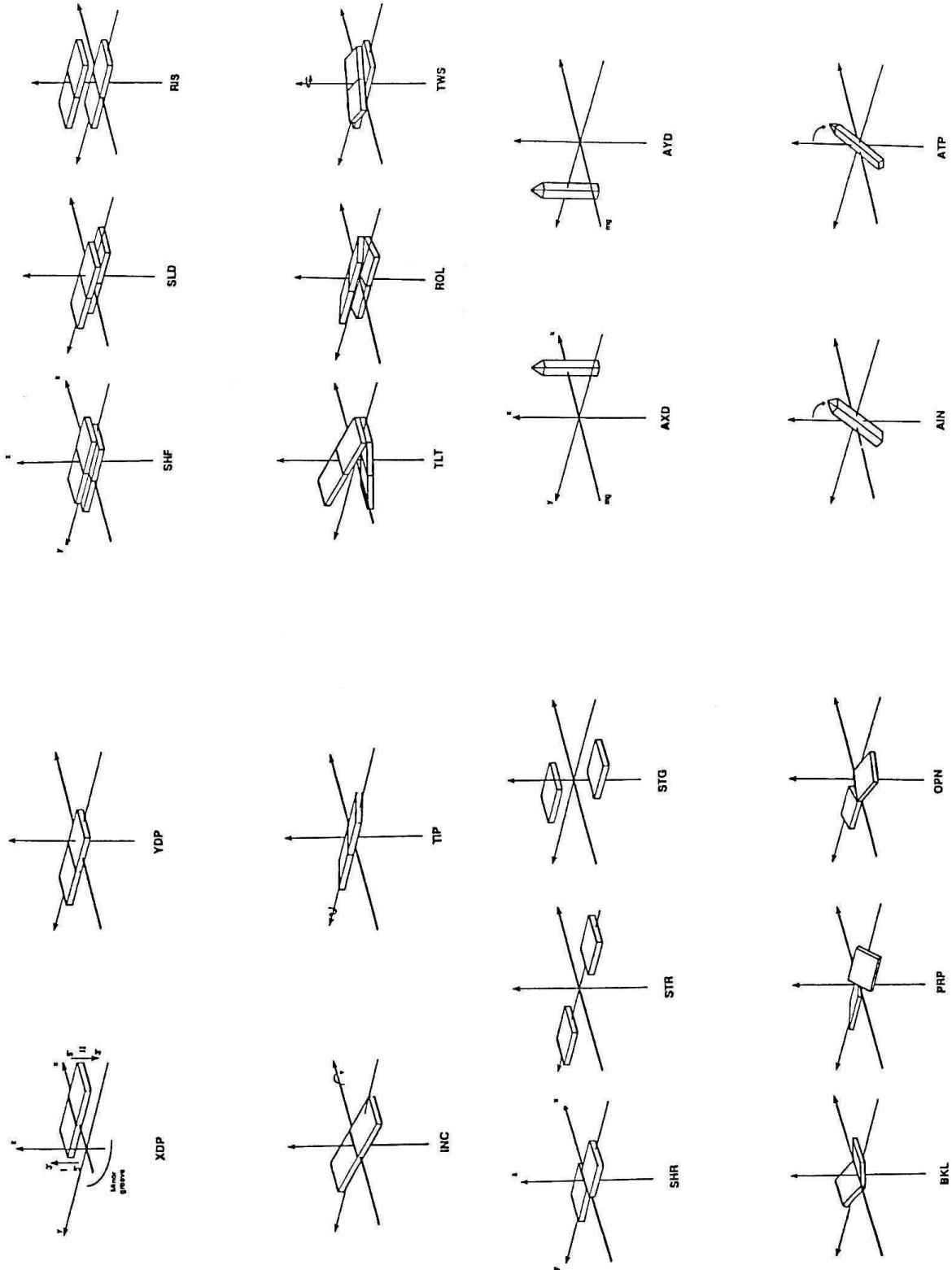
VARIAN *INOVA Unity*. Les spectres dans H<sub>2</sub>O et D<sub>2</sub>O ont été obtenus à partir d'un échantillon de 600µL d'ADN à 2,7mM en présence de tampon phosphate (10mM, pH 5.7). Pour le dimère stable, l'échantillon a été testé avec différentes concentrations de NaCl (50, 100 et 150mM). Les modélisations moléculaires ont été réalisées sur une station de calcul linux bi-processeurs (à 450MHz et 512Mo de RAM) à l'aide du logiciel CNS version 1.0.

Les analyses et représentations des différentes structures ont été réalisées à l'aide des logiciels CURVES (version 5.0) et MOLMOL (version 2K.1).

# *Annexes, abréviations*

# Annexe 1

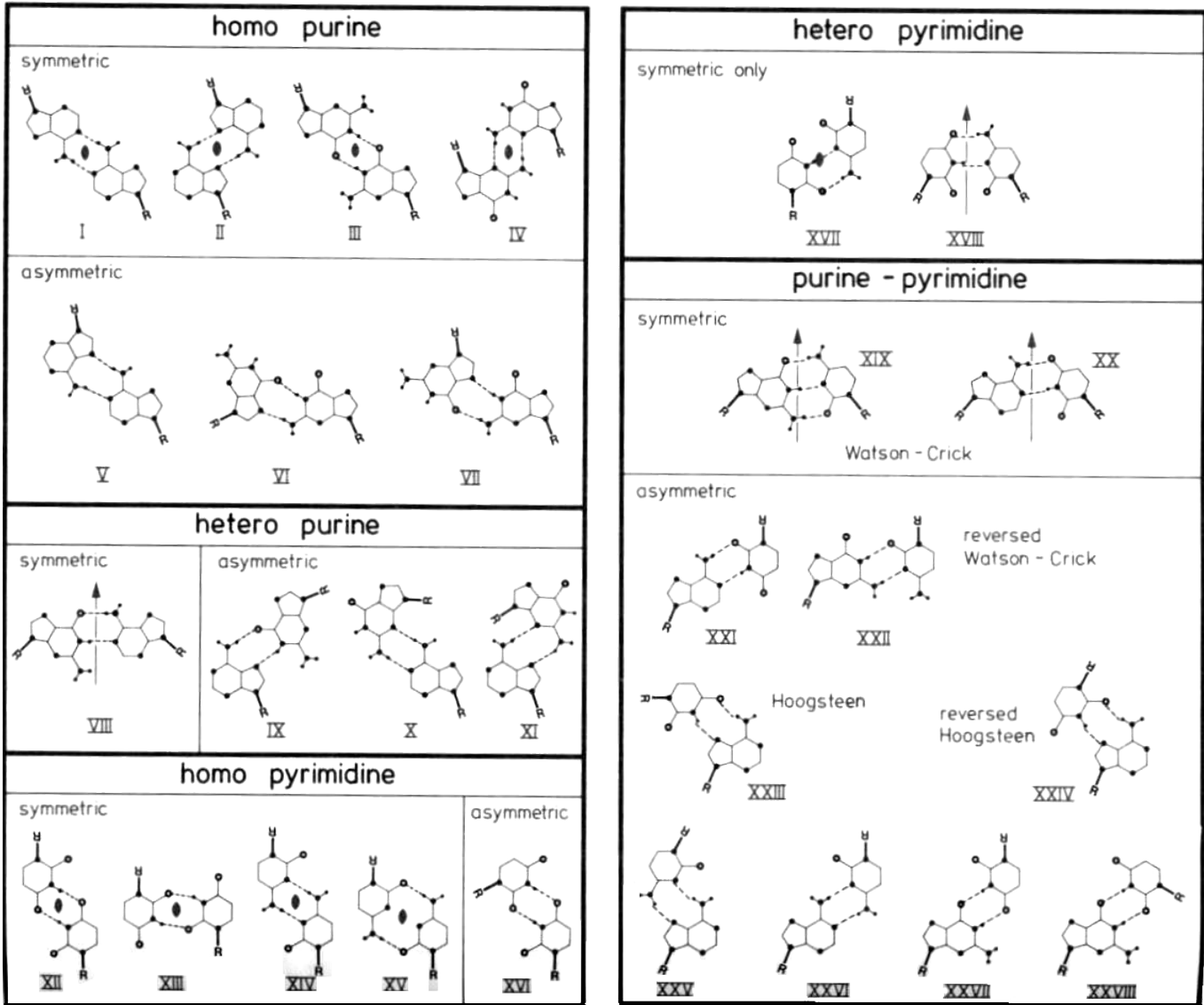
Définition des paramètres d'hélice :



d'après Lavery et Sklenar (Lavery and Sklenar, 1988).

## Annexe 2

Définition des appariements possibles dans les acides nucléiques :



D'après Saenger (Saenger, 1984).

## Abréviations

|                |  |
|----------------|--|
| 1D             | Unidimensionnel  |
| 2D             | Bidimensionnel   |
| A              | Adénine  |
| ADN            | Acide désoxyribonucléique  |
| ARN            | Acide ribonucléique  |
| C              | Cytosine   |
| COSY           | “correlation spectroscopy”   |
| CSA            | Anisotropie de déplacement chimique                                |
| DLS            | “dimer linkage structure”  |
| G              | Guanine  |
| HSQC           | “heteronuclear single quantum coherence”                           |
| MoIMuLV        | virus de la leucémie murine de Moloney                             |
| NC             | Protéine de nucléocapside  |
| NOESY          | “nuclear Overhauser effect spectroscopy”                           |
| PPM            | partie par million   |
| RMN            | Résonance magnétique nucléaire                                     |
| RMSD           | “root main square deviation”                                       |
| T              | Thymine  |
| TAR            | “trans activating responce element” ou élément de trans-activation |
| $\tau_c$       | Temps de corrélation   |
| T <sub>m</sub> | Température de demi-dénaturation                                   |
| TOCSY          | “total correlation spectroscopy”                                   |
| TRIS           | Tris(hydroxyméthyl)aminométhane                                    |
| TROSY          | “transverse optimized spectroscopy”                                |
| U              | Uracile  |
| UV             | Ultra violet   |
| V dW           | Van der Waals  |
| VIH            | Virus de l'immunodéficiencie humaine                               |

# *Bibliographie*

## A

- Abragam, A. (1961). Principles of nuclear magnetism, Carendon Press.
- Adamo, C. and Barone, V. (1998). Toward chemical accuracy in the computation of NMR shieldings : the PBE0 model. *Chem Phys Lett* **298**, 113-9.
- Aldovini, A. and Young, R. A. (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.
- Allain, F. H., Gubser, C. C., Howe, P. W., Nagai, K., Neuhaus, D. and Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature* **380**, 646-50.
- Allemand, J. F., Bensimon, D., Lavery, R. and Croquette, V. (1998). Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proc. Natl. Acad. Sci. USA* **95**, 14152-7.
- Altschul, Stephen, F., Thomas, L. M., Shäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402.
- Amir-Aslani, A., Mauffret, O., Sourgen, F., Neplaz, S., Maroun, R. G., Lescot, E. and Femandjian, S. (1996). Hairpins in a DNA site for topoisomerase II studied by combined NMR and energy minimization methods. *J Mol Biol* **263**, 776-88.
- Anderson, J. A., Teufel, R. J., 2nd, Yin, P. D. and Hu, W. S. (1998). Correlated template-switching events during minus-strand DNA synthesis: a mechanism for high negative interference during retroviral recombination. *J Virol* **72**, 1186-94.
- Astell, C. R., Chow, M. B. and Ward, D. C. (1985). Sequence analysis of the termini of virion and replicative forms of minute virus of mice DNA suggests a modified rolling hairpin model for autonomous parvovirus DNA replication. *J Virol* **54**, 171-7.
- Awang, G. and Sen, D. (1993). Mode of dimerization of HIV-1 genomic RNA. *Biochemistry* **32**, 11453-7.

## B

- Ban, C., Ramakrishnan, B. and Sundaralingam, M. (1994). A single 2'-hydroxyl group converts B-DNA to A-DNA. Crystal structure of the DNA-RNA chimeric decamer duplex d(CCGGC)r(G)d(CCGG) with a novel intermolecular G-C base-paired quadruplet. *J Mol Biol* **236**, 275-85.
- Baudin, F., Marquet, R., Isel, C., Darlix, J. L., Ehresmann, B. and 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. and 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.

Berkhout, B., Essink, B. B. and Schoneveld, I. (1993). In vitro dimerization of HIV-2 leader RNA in the absence of PuGGAPuA motifs. *Faseb J* **7**, 181-7.

Bieth, E., Gabus, C. and Darlix, J. L. (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.

Billeter, M., Schaumann, T. H., Braun, W. and Wüthrich, K. (1990). Restrained energy refinement with two different algorithms and force fields of the structure of the  $\alpha$ -Amylase inhibitor tendamistat determined by NMR in solution. *Biopolymers* **29**, 695-706.

Boiziau, C., Dausse, E., Ludmila, Y. and Toulmé, J. J. (1999). DNA aptamers selected against the HIV-1 trans-activation-responsive RNA element form RNA-DNA kissing complexes. *J Biol Chem* **274**, 12730-7.

Bonnet-Mathoniere, B., Girard, P. M., Muriaux, D. and 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, P. N., Pappalardo, L., Kerwood, D. J. and Pelczer, I. (1997). NMR-based structure determination for unlabelled RNA and DNA. *Adv Biophys Chem* **6**, 173-216.

Braun, W. and Go, N. (1985). Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. *J Mol Biol* **186**, 611-26.

Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983). CHARMM : a program for macromolecular energy minimization and dynamics calculations. *J Comp Chem* **4**, 187-217.

Brünger, A. T. (1992). *X-PLOR : a system for X-ray crystallography and NMR*. Yale University Press.

Brünger, A. T., Adams, P. D., Clore, G. M., De Lano, W. L., Gros, P., Grosse-Kuntze, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998). Crystallography & NMR system : A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-21.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R.

W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-21.

Brünger, A. T. and Nilges, M. (1993). Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy. *Q Rev Biophys* **26**, 49-125.

## C

Case, D. A. (1998). The use of chemical shifts and their anisotropies in biomolecular structure determination. *Curr Opin Struct Biol* **8**, 624-30.

Case, D. A. (2000). Interpretation of chemical shifts and coupling constants in macromolecules. *Curr Opin Struct Biol* **10**, 197-203.

Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatnam, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Ferguson, D. M., Radmer, R. J., Seibel, G. L., Singh, U. C., Weiner, P. K. and Kollman, P. A. (1997). *AMBER 5*. University of California.

Chang, K. Y. and 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, K. Y. and 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.

Cheatham, T. E., III and Kollman, P. A. (1997). Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid complex. *J. Am. Chem. Soc.* **119**, 4805-4825.

Clavel, F. and Orenstein, J. M. (1990). A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J Virol* **64**, 5230-4.

Clever, J., Sasseti, C. and Parslow, T. G. (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, J. L., Wong, M. L. and Parslow, T. G. (1996). Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J Virol* **70**, 5902-8.

Clore, G. M., Starich, M. R. and Gronenborn, A. M. (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, J. M. (1984). RNA tumor viruses in Weiss, R., Teich, N., Varmus, H. and Coffin, J. (Eds), Cold Spring Harbor, pp. 261-368.
- Collin, D., Heijenoort, C. V., Boiziau, C., Toulmé, J. J. and 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.
- Connolly, J. C., Kirkham, L. A. and Leach, D. R. (1998). The SbcCD nuclease of *Escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proc Natl Acad Sci USA* **7**, 7969-74.
- Cromsigt, J. A. M. T. C., Schleucher, J., Kidd-Ljubggren, K. and Wijmenga, S. S. (2000). Synthesis of specifically deuterated nucleotides for NMR studies on RNA. *J Biomol Struct Dyn* **11**, 211-19.

## D

- Dam, E., Pleij, K. and Draper, D. (1992). Structural and functional aspects of RNA pseudoknots. *Biochemistry* **31**, 11665-76.
- Dardel, F., Marquet, R., Ehresmann, C., Ehresmann, B. and Blanquet, S. (1998). Solution studies of the dimerization initiation site of HIV-1 genomic RNA. *Nucleic Acids Res* **26**, 3567-71.
- Darlix, J. L., Gabus, C., Nugeyre, M. T., Clavel, F. and 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.
- de Dios, A. C., Roach, J. L. and Walling, A. E. (1999). The NMR chemical shift : local geometry effects in J.C., F. and de Dios, A. C. (Eds), *Modelling NMR chemical shifts. Gaining insights into structure and environment*, pp. 220-239.
- De Rocquigny, H., Gabus, C., Vincent, A., Fournie-Zaluski, M. C., Roques, B. and Darlix, J. L. (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.
- Delaglio, F., Kontaxis, G. and Bax, A. (2000). Protein structure determination using molecular fragment replacement and NMR dipolar coupling. *J. Am. Chem. Soc.* **122**, 2142-3.
- Delepierre, M., Igolen, J. and Roques, B. P. (1988). Study of the bisintercalation of the antitumor drug ditercalinium by <sup>31</sup>P-NMR. *Biopolymers* **27**, 957-68.
- Dieckmann, T., Suzuki, E., Nakamura, G. K. and Feigon, J. (1996). Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *Rna* **2**, 628-40.

Dornberger, U., Hillisch, A., Gollmick, F. A., Fritzsche, H. and Diekmann, S. (1999). Solution structure of a five-adenine bulge loop within a DNA duplex. *Biochemistry* **38**, 12860-8.

## E

Egli, M., Portmann, S. and Usman, N. (1996). RNA hydration: a detailed look. *Biochemistry* **35**, 8489-94.

El Amri, C., Mauffret, O., Monnot, M., Tevanian, G., Lescot, E., Porumb, H. and Femandjian, S. (1999). A DNA hairpin with a single residue loop closed by a strongly distorted watson-crick G.C base-pair. *J Mol Biol* **294**, 427-42.

Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann, B., Ehresmann, C. and 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.

## F

Frankel, A. D. and Young, J. A. (1998). HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* **67**, 1-25.

## G

Gehring, K., Leroy, J. L. and Guéron, M. (1993). A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* **363**, 561-5.

Gesteland, R. F., Cech, T. R. and Atkins, J. F. (1998). *The RNA world, second edition*. CSHL press.

Giessner-Prettre, C. and 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 d'ARN SL1-23mer initiant la dimérisation chez VIH-1Lai. *Thèse de doctorat, Université d'Orléans*.

Girard, F., Barbault, F., Gouyette, C., Huynh-Dinh, T., Paoletti, J. and 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, P. M., de Rocquigny, H., Roques, B. P. and 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, L. A. and

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.

Gorenstein, D. G. (1984). Phosphorus 31 NMR, Principles and applications. *Academic Press, New York*.

Greene, K. L., Wang, Y. and Live, D. (1995). Influence of the glycosidic torsion angle on <sup>13</sup>C and <sup>15</sup>N shifts in guanosine nucleotides: investigations of G-tetrad models with alternating syn and anti bases. *J Biomol NMR* **5**, 333-8.

Gubser, C. C. and Varani, G. (1996). Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3'-untranslated region and interaction with the U1A protein. *Biochemistry* **35**, 2253-67.

Güntert, P. (1998). Structure calculation of biological macromolecules from NMR data. *Q Rev Biophys* **31**, 145-237.

Güntert, P., Braun, W. and Wuthrich, K. (1991a). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J Mol Biol* **217**, 517-30.

Güntert, P., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. and Wuthrich, K. (1991b). Structure determination of the Antp (C39----S) homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. *J Mol Biol* **217**, 531-40.

Gyi, J. I., Lane, A. N., Conn, G. L. and Brown, T. (1998). Solution structures of DNA.RNA hybrids with purine-rich and pyrimidine-rich strands: comparison with the homologous DNA and RNA duplexes. *Biochemistry* **37**, 73-80.

## H

Hall, K., Cruz, P., Tinoco, I., Jr., Jovin, T. M. and van de Sande, J. H. (1984). 'Z-RNA'--a left-handed RNA double helix. *Nature* **311**, 584-6.

Havel, T. F. and Wuthrich, K. (1985). An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *J Mol Biol* **182**, 281-94.

Helene, C. (1991). The anti-gene strategy: control of gene expression by triplex-forming-oligonucleotides. *Anticancer Drug Des* **6**, 569-84.

Helgaker, T., Jaszunski, M. and Ruud, K. (1999). Ab initio methods for the calculation of NMR shielding and indirect spin-spin coupling constants. *Chem Rev* **99**, 293-352.

Hermann, T. and Patel, D. J. (2000). RNA bulges as architectural and recognition motifs. *Structure Fold Des* **8**, R47-54.

Hirao, I., Nishimura, Y., Naraoka, T., Watanabe, K. and Miura, K. (1989). Extraordinary stable structure of short single-stranded DNA fragments containing a specific base sequence : d(GCGAAAGC). *Nucleic Acids Res* **17**, 2223-31.

Hirao, I., Nishimura, Y., Tagawa, Y., Watanabe, K. and Miura, K. (1992). Extraordinarily stable mini-hairpins : electropherical and thermal properties of the various sequence variants d(GCGAAAGC) and their effect on DNA sequencing. *Nucleic Acids Res* **20**, 3891-6.

## I

Ippolito, J. A. and Steitz, T. A. (1998). A 1.3-Å resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation. *Proc Natl Acad Sci U S A* **95**, 9819-24.

## I

James, T. L. (1994a). Assessment of quality of derived macromolecular structures. *Methods Enzymol* **239**, 416-39.

James, T. L. (1994b). Computational strategies pertinent to NMR solution structure determination. *Curr Op Stru Biol* **4**, 275-84.

Jiang, F., Kumar, R. A., Jones, R. A. and Patel, D. J. (1996). Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature* **382**, 183-6.

Jones, J. S., Allan, R. W. and Temin, H. M. (1994). One retroviral RNA is sufficient for synthesis of viral DNA. *J Virol* **68**, 207-16.

Joshua-Tor, L., Frolow, F., Appella, E., Hope, H., Rabinovich, D. and Sussman, J. L. (1992). Three-dimensional structures of bulge-containing DNA fragments. *J Mol Biol* **225**, 397-431.

Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. and Pardi, A. (1996). A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol* **264**, 968-80.

## K

Kalnik, M. W., Norman, D. G., Swann, P. F. and Patel, D. J. (1989). Conformation of adenosine bulge-containing deoxytridecanucleotide duplexes in solution. Extra adenosine stacks into duplex independent of flanking sequence and temperature. *J Biol Chem* **264**, 3702-12.

- Karplus, M. (1959). Contact electron-spin coupling of nuclear magnetic moments. *J Chem Phys* **30**, 11-5.
- Katz, R. A., Gravuer, K. and Skalka, A. M. (1998). A preferred target DNA structure for retroviral integrase in vitro. *J Biol Chem* **273**, 24190-5.
- Kemper, B. and Janz, E. (1976). Function of gene 49 of bacteriophage T4. I. Isolation and biochemical characterization of very fast-sedimenting DNA. *J Virol* **18**, 992-9.
- Kim, C. H. and 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.
- Kim, S. G., Lin, L. J. and Reid, B. R. (1992). Determination of nucleic acid backbone conformation by <sup>1</sup>H NMR. *Biochemistry* **31**, 3564-74.
- Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H., Seeman, N. C. and Rich, A. (1974). Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* **185**, 435-40.
- Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983). Optimization by simulated annealing. *science* **220**, 671-80.
- Kolk, M. H., van der Graaf, M., Wijmenga, S. S., Pleij, C. W., Heus, H. A. and Hilbers, C. W. (1998). NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA. *Science* **280**, 434-8.
- Kung, H. J., Bailey, J. M., Davidson, N., Nicolson, M. O. and McAllister, R. M. (1975). Structure, subunit composition, and molecular weight of RD-114 RNA. *J Virol* **16**, 397-411.
- Kuszewski, J., Gronenborn, A. M. and Clore, G. M. (1995a). The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J Magn Reson B* **107**, 293-7.
- Kuszewski, J., Gronenborn, A. M. and Clore, G. M. (1996). A potential involving multiple proton chemical-shift restraints for nonstereospecifically assigned methyl and methylene protons. *J Magn Reson B* **112**, 79-81.
- Kuszewski, J., Qin, J., Gronenborn, A. M. and Clore, G. M. (1995b). The impact of direct refinement against <sup>13</sup>C alpha and <sup>13</sup>C beta chemical shifts on protein structure determination by NMR. *J Magn Reson B* **106**, 92-6.

## L

- Lacroix, L., Liénard, H., Labourier, E., Djavaheri-Mergny, M., Lacoste, J., Leffers, H., Tazi, J., Hélène, C. and Mergny, J.-L. (2000). Identification of two human nuclear proteins

that recognise the cytosine-rich strand of human telomers in vitro. *Nucleic Acids Res* **28**, 1564-75.

Lancelot, G., Guesnet, J. L. and Vovelle, F. (1989). Solution structure of the parallel-stranded duplex oligonucleotide alpha-d(TCTAAAC)-beta-d(AGATTIG) via complete relaxation matrix analysis of the NOE effects and molecular mechanics calculations. *Biochemistry* **28**, 7871-8.

Laughrea, M. and 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. and 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. and 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.

Lavery, R. and 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. and Sklenar, H. (1989). Defining the structure of irregular nucleic acids: conventions and principles. *J Biomol Struct Dyn* **6**, 655-67.

Lavery, R. and Sklenar, H. (1990). *A quantitative description of conformation of biological macromolecules*. Adenin Press.

Lavery, R., Zakrzewska, K. and Sklenar, H. (1995). JUMNA : junction minimisation of nucleic acids. *Comp Phys Comm* **91**, 255-62.

Lee, A. J. and Crothers, D. M. (1998). The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence. *Structure* **6**, 993-1005.

Leroy, J. L., Gehring, K., Kettani, A. and Guéron, M. (1993). Acid multimers of oligodeoxycytodine strands: stoichiometry, base-pair characterization, and proton exchange properties. *Biochemistry* **32**, 4634-40.

Lever, A., Gottlinger, H., Haseltine, W. and 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.

Liu, Y., Zhao, D., Altman, R. and Jardetzky, O. (1992). A systematic comparison of three structure determination methods from NMR data : dependence upon quality and quantity of data. *J Biomol NMR* **2**, 373-88.



Lodmell, S. J., Paillart, J. C., Mignot, D., Ehresmann, B., Ehresmann, C. and Marquet, R. (1998). Oligonucleotide-mediated inhibition of genomic RNA dimerization of HIV-1 strains Mal an Lai : a comparative analysis. *Antisense Nucleic Acid Drug Development* **8**, 517-29.

## M

Marino, J. P., Gregorian, R. S., Jr., Csankovszki, G. and Crothers, D. M. (1995). Bent helix formation between RNA hairpins with complementary loops. *Science* **268**, 1448-54.

Marion, D., Ikura, M., Tschudin, R. and 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**, 393-9.

Marquet, R., Baudin, F., Gabus, C., Darlix, J. L., Mougel, M., Ehresmann, C. and Ehresmann, B. (1991). Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and possible mechanism. *Nucleic Acids Res* **19**, 2349-57.

McCammon, J. A., Gelin, B. R. and Karplus, M. (1977). Dynamics of folded proteins. *Nature* **267**, 585-90.

Mizuno, H. and Sundaralingam, M. (1978). Stacking of Crick Wobble pair and Watson-Crick pair: stability rules of G-U pairs at ends of helical stems in tRNAs and the relation to codon-anticodon Wobble interaction. *Nucleic Acids Res* **5**, 4451-61.

Mougel, M., Tounekti, N., Darlix, J. L., Paoletti, J., Ehresmann, B. and 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, J. L., Billeci, T. M., James, T. L. and Parslow, T. G. (1998). Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat Struct Biol* **5**, 432-6.

Mujeeb, A., Kerwin, S. M., Kenyon, G. L. and James, T. L. (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, T. G., Zarrinpar, A., Das, C. and James, T. L. (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, B. P. and 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. and 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, P. M., Bonnet-Mathoniere, B. and Paoletti, J. (1995). Dimerization of HIV-1Lai 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, K. G., Bondurant, M. and 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.

## N

Narlikar, G. J., Khosla, M., Usman, N. and Herschlag, D. (1997). Quantitating tertiary binding energies of 2'OH groups on the P1 duplex of the *Tetrahymena* Ribozyme : intrinsic binding energy in an RNA enzyme. *Biochemistry* **36**, 2465-77.

Neidle, S. (1999). *Oxford handbook of nucleic acid structure*. Oxford science publications.

Nikonowicz, E. P., Kalurachchi, K. and DeJong, E. (1997a). Comparison of H5 and H8 relaxation rates of a 2H/13C/15N labeled RNA oligonucleotide with selective protonation at C5 and C8. *FEBS Lett* **415**, 109-13.

Nikonowicz, E. P., Michnicka, M., Kalurachchi, K. and DeJong, E. (1997b). Preparation and characterization of a uniformly 2 H/ 15 N-labeled RNA oligonucleotide for NMR studies. *Nucleic Acids Res* **25**, 1390-6.

Nilges, M. (1993). A calculation strategy for the structure determination of symmetric dimers by 1H-NMR. *Proteins* **17**, 297-309.

Nilges, M. (1996). Structure calculation from NMR data. *Curr Op Stru Biol* **6**, 617-23.

Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988a). Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. Circumventing problems associated with folding. *FEBS Lett* **239**, 129-36.

Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988b). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett* **229**, 317-24.

Nilsson, L. and Karplus, M. (1986). Empirical energy functions for energy minimization and dynamics of nucleic-acids. *J Comp Chem* **7**, 591-616.

## O

- Ohki, S. and Kainosho, M. (1999). [New aspects is stable-isotope-assisted NMR techniques: application to be structural analysis of larger complexes between proteins and nucleic acids]. *Tanpakushitsu Kakusan Koso* **44**, 574-81.
- Orita, M., Nishikawa, F., Shimayama, T., Taira, K., Endo, Y. and Nishikawa, S. (1993). High-resolution NMR study of a synthetic oligoribonucleotide with a tetranucleotide GAGA loop that is a substrate for the cytotoxic protein, ricin. *Nucleic Acids Res* **21**, 5670-8.

## P

- Paillart, J. C., Marquet, R., Skripkin, E., Ehresmann, C. and Ehresmann, B. (1996). Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie* **78**, 639-53.
- Paillart, J. C., Westhof, E., Ehresmann, C., Ehresmann, B. and 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.
- Pardi, A., Hare, D. R. and Wang, C. (1988). Determination of DNA structures by NMR and distance geometry techniques : A computer simulation. *Proc Natl Acad Sci USA* **85**, 8785-9.
- Pavlakis, G. N. (1996). The molecular biology of human immunodeficiency virus type 1 in De Vita, J. V. T., Hellman, S. and Rosenberg, S. A. (Eds), *AIDS: biology, diagnosis, treatment and prevention*, Lippincott-Raven.
- Perreault, J. P., Labuda, D., Usman, N., Yang, J. H. and Cedergren, R. (1991). Relationship between 2'-hydroxyls and magnesium binding in the hammerhead RNA domain: a model for ribozyme catalysis. *Biochemistry* **30**, 4020-5.
- Pervushin, K., Riek, R., Wider, G. and 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. and Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **2**, 661-5.
- Plateau, P. and Guéron, M. (1982). Exchangeable proton NMR without base-line distorsion, using new strong-pulse sequences. *J Am Chem Soc* **104**, 7310-11.
- Plavec, J., Thibaudeau, C. and Chattopadhyaya, J. (1994). How does the 2'-hydroxy group drive the pseudorotational equilibrium in nucleoside and nucleotide by tuning of

the 3'-gauche effect? *J Am Chem Soc* **116**, 6558-65560.

Pleij, C. W. (1995). Structure and function of RNA pseudoknots. *Genet Eng* **17**, 67-80.

Pley, H. W., Flaherty, K. M. and McKay, D. B. (1994). Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**, 111-3.

Powell, M. J. D. (1977). restart procedures for the conjugate gradient method. *Mathematical Programming* **12**, 241-54.

Prats, A. C., Roy, C., Wang, P. A., Erard, M., Housset, V., Gabus, C., Paoletti, C. and Darlix, J. L. (1990). cis elements and trans-acting factors involved in dimer formation of murine leukemia virus RNA. *J Virol* **64**, 774-83.

Prestegard, J. H., al-Hashimi, H. M. and Tolman, J. R. (2000). NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q Rev Biophys* **33**, 371-424.

Puglisi, E. V. and Puglisi, J. D. (1998). Nuclear magnetic resonance spectroscopy of RNA, *RNA structure and function*, Cold Spring Harbor Laboratory Press, pp. 117-46.

## R

Rahmouni, R. (1992). Z-DNA as a probe for localized supercoiling in vivo. *Molecular Microbiology* **6**, 569-72.

Ravishanker, G., Swaminathan, S., Beveridge, D. L., Lavery, R. and Sklenar, H. (1989). Conformational and helicoidal analysis of 30 PS of molecular dynamics on the d (CGCGAATTCGCG) double helix: "curves", dials and windows. *J Biomol Struct Dyn* **6**, 669-99.

Rein, A. (1994). Retroviral RNA packaging: a review. *Arch Virol Suppl* **9**, 513-22.

Rich, A. (1995). Speculation on the biological roles of left-handed Z-DNA. *Annals New York Acad. Sci.* **726**, 1-17.

Rich, A., Nordheim, A. and Wang, A. H. J. (1984). The chemistry and biology of left-handed Z-DNA. *Annual Review Biochem* **53**, 791-846.

Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. and Klug, A. (1974). Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **250**, 546-51.

Rosen, M. A., Shapiro, L. and Patel, D. J. (1992). Solution structure of a trinucleotide A-T-A bulge loop within a DNA duplex. *Biochemistry* **31**, 4015-26.

## S

- Saenger, W. (1984). *Principles of nucleic acid structure*. Springer.
- Sakaguchi, K., Zambrano, N., Baldwin, E. T., Shapiro, B. A., Erickson, J. W., Omichinski, J. G., Clore, G. M., Gronenborn, A. M. and Appella, E. (1993). Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. *Proc Natl Acad Sci U S A* **90**, 5219-23.
- SantaLucia, J., Jr., Kierzek, R. and Turner, D. H. (1992). Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science* **256**, 217-9.
- SantaLucia, J., Jr. and Turner, D. H. (1993). Structure of (rGGCGAGCC)<sub>2</sub> in solution from NMR and restrained molecular dynamics. *Biochemistry* **32**, 12612-23.
- Schaaf, S. G., Beck, J. and Nassal, M. (1999). A small 2'-OH and base-dependent recognition element downstream of the initiation site in the RNA encapsidation signal is essential for hepatitis B virus replication initiation. *J Biol Chem* **274**, 37787-94.
- Scheurer, C., Skrynnikoc, N., Lienin, S. F., Straus, S. K., Brüschweiler, R. and Ernst, R. R. (1999). Effects of dynamics and environment of <sup>15</sup>N chemical shielding anisotropy in proteins. A combination of density functional theory, molecular dynamics simulation, and NMR relaxation. *J Am Chem Soc* **121**, 4230-1.
- Shepard, W., Cruse, W. B., Fourme, R., de la Fortelle, E. and Prangé, T. (1998). A zipper-like duplex in DNA : the crystal structure of d(GCGAAAGCT) at 2.1Å resolution. *Structure* **6**, 849-61.
- Sippl, M. J. and Scheraga, H. A. (1986). Cayley-Menger coordinates. *Proc Natl Acad Sci U S A* **83**, 2283-7.
- Smith, D. E., Su, J. Y. and Jucker, F. M. (1997). Efficient enzymatic synthesis of <sup>13</sup>C,<sup>15</sup>N-labeled DNA for NMR studies. *J Biomol NMR* **10**, 245-53.
- Snoussi, K., Nonin-Lecomte, S. and Leroy, J. L. (2001). The RNA i-motif. *J Mol Biol* **309**, 139-53.
- Sobell, H. M. (1972). Molecular mechanism for genetic recombination. *Proc Natl Acad Sci U S A* **69**, 2483-7.
- Söll, D. (1993). Transfert RNA: an RNA for all seasons in Gesteland, R. F. and Atkins, J. F. (Eds), *The RNA World*, Cold Spring Harbor Laboratory Press, pp. 157-84.
- Stein, E. G., Rice, L. M. and Brunger, A. T. (1997). Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J Magn Reson* **124**, 154-64.

Sugimoto, N., Kierzek, R., Freier, S. M. and Turner, D. H. (1986). Energetics of internal GU mismatches in ribooligonucleotide helices. *Biochemistry* **25**, 5755-9.

Sundquist, W. I. and Heaphy, S. (1993). Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA [see comments]. *Proc Natl Acad Sci U S A* **90**, 3393-7.

Szewczak, A. A. and Moore, P. B. (1995). The sarcin/ricin loop, a modular RNA. *J Mol Biol* **247**, 81-98.

## T

Takahashi, K., Baba, S., Hayashi, Y., Koyanagi, Y., Yamamoto, N., Takaku, H. and Kawai, G. (2000). NMR analysis of intra- and inter-molecular stems in the dimerization initiation site of the HIV-1 genome. *J Biochem (Tokyo)* **127**, 681-6.

Tanikawa, J., Nishimura, Y., Hirao, I. and Miura, K. (1991). NMR spectroscopic study of single-stranded DNA fragments of d(CGCGCAAAGCCG) and d(CGCGCAAAGCCG). *Nucleic Acids Symp Ser* **25**, 47-8.

Theilleux-Delalande, V. (1998). Les interactions ARN/protéine de nucléocapside chez VIH-1 : leurs implications dans les relations structure-fonction in vitro. *Thèse de doctorat, Université René Descartes*.

Theilleux-Delalande, V., Girard, F., Huynh-Dinh, T., Lancelot, G. and 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.

Thibaudeau, C., Plavec, J., Garg, N., Pachikin, A. and Chattopadhyaya, J. (1994). How does the electronegativity of the substituent dictate the strength of the gauche effect? *J. Am. Chem. Soc.* **116**, 4038-4043.

Tjandra, N. and Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium [see comments] [published erratum appears in Science 1997 Dec 5;278(5344):1697]. *Science* **278**, 1111-4.

## U

Ulyanov, N. B., Ivanov, V. I., Minyat, E. E., Khomyakova, E. B., Petrova, M. V., Lesiak, K. and James, T. L. (1998). A pseudosquare knot structure of DNA in solution. *Biochemistry* **37**, 12715-26.

## V

Varani, G. (1995). Exceptionally stable nucleic acid hairpins. *Annu Rev Biophys Biomol Struct* **24**, 379-404.

Verlet, L. (1967). Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys Rev* **159**, 98-105.

## W

Waston, J. D. and Crick, F. H. C. (1953). The molecular structure of nucleic acids. A structure for deoxyribose nucleic acids. *Nature* **171**, 737-9.

Weiss, S., Hausl, G., Famulok, M. and 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.

Wiberg, K. B. (1977). Intramolecular interactions in molecular mechanics calculations. *Computers & Chemistry* **1**, 221-3.

Wijmenga, S. S., Kruithof, M. and Hilbers, C. W. (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. S., Mooren, M. M. W. and Hilbers, C. W. (1993). *NMR of macromolecules. A practical approach*. Oxford University Press.

Wijmenga, S. S. and van Buuren, B. N. M. (1998). The use of NMR methods for conformational studies of nucleic acids. *Prog Nucl Magn Reson Spec* **32**, 287-387.

Wilson, P. J., Amos, R. D. and Handy, N. C. (1999). Density function prediction for magnetizabilities and nuclear shielding constants. *Mol Phys* **97**, 757-68.

Wilson, W. D. and Jones, R. L. (1982). Interaction of actinomycin D, ethidium, quinacrine, daunorubicin, and tetralysine with DNA: <sup>31</sup>P NMR chemical shift and relaxation investigation. *Nucleic Acids Res* **10**, 1399-410.

Wimberly, B., Varani, G. and Tinoco, I., Jr. (1993). The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry* **32**, 1078-87.

Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647-51.

Woese, C. R., Winker, S. and Gutell, R. R. (1990). Architecture of ribosomal RNA: constraints on the sequence of 'tetra-loops'. *Proc Natl Acad Sci U S A* **87**, 8467-71.

## Y

Yamakage, S. I., Maltseva, T. V., Nilson, F. P., Foldesi, A. and Chattopadhyaya, J. (1993). Deuteriation of sugar protons simplify NMR assignments and structure determination of large oligonucleotide by the  $^1\text{H}$ -NMR window approach. *Nucleic Acids Res* **21**, 5005-11.

## Z

Zhang, X. L. and Patel, D. J. (1991). Solution structure of the luzopeptin-DNA complex. *Biochemistry* **30**, 4026-41.

Zimmerman, S. B. and Pfeiffer, B. H. (1979). A direct demonstration that the ethanol-induced transition of DNA is between the A and B forms: an X-ray diffraction study. *J Mol Biol* **135**, 1023-7.

Zimmermann, G. R., Jenison, R. D., Wick, C. L., Simorre, J. P. and Pardi, A. (1997). Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA [see comments]. *Nat Struct Biol* **4**, 644-9.