



HAL
open science

ETUDE PAR SIMULATION DE DYNAMIQUE MOLECULAIRE DE LA VARIABILITE CONFORMATIONNELLE DU DIMERE DE LA SEQUENCE SL1 DU GENOME DE VIH-1

Samia Aci

► **To cite this version:**

Samia Aci. ETUDE PAR SIMULATION DE DYNAMIQUE MOLECULAIRE DE LA VARIABILITE CONFORMATIONNELLE DU DIMERE DE LA SEQUENCE SL1 DU GENOME DE VIH-1. domain_other. Université d'Orléans, 2004. Français. NNT : . tel-00008151v3

HAL Id: tel-00008151

<https://theses.hal.science/tel-00008151v3>

Submitted on 10 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.



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

Discipline :
Biophysique Moléculaire

PAR
Samia ACI

**Etude par simulation de dynamique moléculaire de
la variabilité conformationnelle du dimère de la
séquence SL1 du génome de VIH-1.**

Soutenue le 25 novembre 2004

MEMBRES DU JURY :

| | | |
|--------------------|----------------------------------|--------------------|
| M. M. GHOMI | Professeur, Université Paris 6 | Rapporteur |
| M. R. LAVERY | Directeur de Recherche, CNRS | Rapporteur |
| M. L. MORIN-ALLORY | Professeur, Université d'Orléans | Examineur |
| M. J. PAOLETTI | Professeur, Université d'Orléans | Président |
| M. D. GENEST | Directeur de Recherche, CNRS | Directeur de Thèse |

REMERCIEMENTS

Je remercie le Pr Paul Vigny et le Dr Jean-Claude Beloeil pour m'avoir accueillie au Centre de Biophysique Moléculaire.

J'adresse de vifs remerciements au Pr Mahmoud Ghomi et au Dr Richard Lavery pour avoir accepté de juger et de corriger mon manuscrit de thèse.

Je tiens à remercier le Pr Luc Morin--Allory et le Pr Jacques Paoletti pour avoir accepté de faire partie de mon jury de thèse.

Je remercie chaleureusement l'ensemble de l'équipe « Simulation numérique, modélisation et dynamique moléculaire » pour m'avoir accueillie et pour la cordiale ambiance qui règne au sein de l'équipe ; le Dr Daniel Genest, mon directeur de recherche, pour m'avoir encadrée et pour sa confiance tout au long des ces travaux, le Dr Monique Genest, directeur de l'équipe, et le Dr Norbert Garnier pour leurs conseils et leurs encouragements, Alain Boyer, pour toute l'aide technique apportée au cours de ces recherches, et Pierre Aller, doctorant de la même promotion, pour avoir arpenté avec moi le périlleux et sinueux chemin du doctorat. Merci également aux différents stagiaires qui ont fait partie de l'équipe pour avoir participé à cette sympathique ambiance : David, Geoffrey, Dorothée, Sébastien, Florent, Loïs, Valérie, Oumarou, Mathieu et Sonia.

Je remercie l'équipe de Jacques Paoletti pour les expériences de biochimie menées en collaboration avec mes travaux ainsi que l'équipe de Gérard Lancelot pour les coordonnées des structures RMN.

Je remercie la Région Centre pour le financement de ma thèse.

Merci à Edouard pour tant et tant de choses...

Merci à Nathalie et Véronique pour les heures de discussion en pause-café, de tram épargnées et de franches rigolades.

Je remercie l'ensemble des étudiants du CBM, présent et passé, pour les loisirs, les soirées, leurs soutiens : Stéphanie, Natacha, Pedro, Céline, Fabien, Laure, Karine, Eric, Vanessa, Mélanie, Claire, Michaël, Sarah, Florian, Guillaume, Christine, Gaëlle-Anne, Erwan, Simon, Patrick, Philippe, Anne, Hélène, Elodie, François,...

Merci aux *Compagnons du Crépuscule*, dont les flammes ardentes ont su m'inspirer...

Merci à mes amis, Casper, Fred, Lucie, Jules, Nadia et Ben.

Merci à Ghislaine, David et Karine, Julien et Adeline.

Merci à toute ma famille, à mes parents Djaffar et Cécile, à mes sœurs Nouara, Ourida et Nadia, à Grégory, à Zaya et Ismaël et à Chloé (tiens bon !).

SOMMAIRE

| | |
|---|-----------|
| ABREVIATIONS | 8 |
| I. INTRODUCTION | 11 |
| II. GENERALITES | 13 |
| II.a. Le VIH | 13 |
| II.a.i. Taxonomie et caractéristiques rétrovirales | 13 |
| II.a.ii. Aspects structuraux des particules virales | 15 |
| → <i>Protéines de l'enveloppe</i> | 16 |
| → <i>Protéines structurales</i> | 16 |
| → <i>Protéines enzymatiques</i> | 17 |
| → <i>Protéines régulatrices</i> | 17 |
| II.a.iii. Cycle rétroviral de VIH-1 | 17 |
| → <i>Entrée du virus</i> | 18 |
| → <i>Transcription inverse</i> | 19 |
| → <i>Intégration de l'ADN proviral au génome de la cellule-hôte</i> | 19 |
| → <i>Transcription et traduction de l'ADN proviral</i> | 20 |
| → <i>Assemblage et encapsidation</i> | 21 |
| → <i>Bourgeonnement et maturation</i> | 21 |
| II.a.iv. Organisation du génome rétroviral | 22 |
| II.a.v. La dimérisation de l'ARN génomique de VIH-1 | 23 |
| → <i>Un génome diploïde</i> | 23 |
| → <i>Implication de la dimérisation dans des étapes-clé du cycle rétroviral</i> | 25 |
| II.b. Le dimère de l'ARN génomique de VIH-1 | 27 |
| II.b.i. Deux formes de dimères | 27 |
| II.b.ii. Etudes expérimentales et numériques de la tige-boucle SL1 | 29 |
| → <i>Dimère étendu « stable »</i> | 29 |
| → <i>Kissing-complex « instable »</i> | 31 |
| II.b.iii. Les simulation de dynamique moléculaire de la tige-boucle SL1 | 37 |
| → <i>Mécanismes de transition</i> | 41 |
| → <i>La protéine de nucléocapside NCp7</i> | 44 |
| II.c. But du travail de Thèse | 45 |
| III. MATERIEL ET METHODES | 48 |
| III.a. Rappels sur l'ARN | 48 |
| III.a.i. Structure de l'ARN | 48 |
| III.a.ii. Paramètres structuraux des acides nucléiques | 50 |
| III.b. La simulation de dynamique moléculaire de polymères | 52 |
| III.b.i. Principe de la dynamique moléculaire libre | 52 |

| | | |
|---------------|--|------------|
| → | <i>Champs de Force</i> | 53 |
| → | <i>Minimisation d'énergie</i> | 55 |
| → | <i>Algorithmes d'intégration</i> | 57 |
| → | <i>Ensemble thermodynamique</i> | 59 |
| → | <i>Conditions périodiques</i> | 59 |
| → | <i>Particle Mesh Ewald (PME)</i> | 62 |
| → | <i>Limitation de la dynamique moléculaire</i> | 64 |
| → | <i>Le RMSD</i> | 64 |
| III.b.ii. | <i>La dynamique moléculaire sous contraintes</i> | 64 |
| → | <i>Contraintes de distance entre deux atomes, d'angle ou d'angle de torsion</i> | 64 |
| → | <i>Dynamique moléculaire dirigée</i> | 65 |
| III.b.iii. | <i>Protocole des dynamiques moléculaires</i> | 67 |
| → | <i>Structures initiales</i> | 67 |
| → | <i>Topologie</i> | 67 |
| → | <i>Minimisation</i> | 68 |
| → | <i>Chauffage</i> | 69 |
| → | <i>Production</i> | 69 |
| III.b.iv. | <i>Analyses des trajectoires</i> | 69 |
| → | <i>Liaisons hydrogène</i> | 69 |
| → | <i>Carte de flexibilité</i> | 70 |
| → | <i>Calcul de l'énergie libre de transition (ΔG) par la méthode WHAM</i> | 71 |
| → | <i>Calcul de l'entropie et de l'enthalpie</i> | 73 |
| III.c. | Les logiciels utilisés | 73 |
| III.c.i. | <i>Le logiciel AMBER</i> | 73 |
| → | <i>Modèles pour l'eau et les ions</i> | 74 |
| → | <i>Structure du programme</i> | 74 |
| III.c.ii. | <i>Logiciel CURVES</i> | 75 |
| III.c.iii. | <i>Visualisation : VMD</i> | 76 |
| IV. | RESULTATS | 78 |
| IV.a. | Etude du dimère instable de SL1 | 78 |
| IV.a.i. | <i>Etude de la structure RMN de Mujeeb et al. avec la séquence SL1Δ</i> | 78 |
| IV.a.ii. | <i>Etude de la structure RMN de Mujeeb et al. avec la séquence SL1 sauvage</i> | 82 |
| I.a.i. | <i>Etude des structures cristallographiques de Ennifar et al. à partir de la séquence SL1. Influence des ions magnésiums</i> | 86 |
| → | <i>Etude du dimère instable de SL1_{Lai}</i> | 86 |
| → | <i>Etude du dimère instable de SL1_{Mal}</i> | 89 |
| I.a.ii. | <i>Etude de la structure RMN de Kieken à partir de la séquence SL1*</i> | 92 |
| → | <i>Etude d'une structure caractéristique de la famille 1</i> | 93 |
| → | <i>Etude d'une structure caractéristique de la famille 2</i> | 95 |
| I.a.iii. | <i>Discussion-conclusion</i> | 99 |
| IV.b. | Etude du dimère stable de SL1 | 106 |

| | | |
|--------------|---|------------|
| IV.b.i. | Etude de la structure RMN de Girard et al. à partir de la séquence SL1 _{Lai} | 107 |
| IV.b.ii. | Etude de la structure cristallographique de Ennifar et al. à partir de la séquence SL1 _{Ma1} | 113 |
| IV.b.iii. | Discussion-conclusion | 121 |
| IV.c. | Thermodynamique de fermeture de la paire de base G7-C17 à la jonction de la tige et de la boucle | 125 |
| IV.c.i. | Détermination des grandeurs thermodynamiques de transition | 126 |
| IV.c.ii. | Discussion-conclusion | 133 |
| IV.d. | Transition « kissing-complex » → dimère étendu: application de la dynamique dirigée sur un changement conformationnel d'importance | 135 |
| IV.d.i. | Objectif du travail et choix de la méthode | 135 |
| IV.d.ii. | Glossaire des dénominations utilisées dans ces expériences et particularités d'analyse | 136 |
| | → <i>Dénomination des structures initiales et finales utilisées</i> | 136 |
| | → <i>La température</i> | 137 |
| | → <i>Paramètres spécifiques de la TMD</i> | 137 |
| | → <i>Spécificité des analyses</i> | 137 |
| IV.d.iii. | Transition par TMD | 138 |
| IV.d.iv. | TMD ⁻¹ : mise en évidence d'une famille d'intermédiaires | 142 |
| IV.d.v. | Transition complète par combinaison des méthodes de dynamique dirigée inverse et directe | 153 |
| IV.d.vi. | Discussion-conclusion | 160 |
| V. | CONCLUSION GENERALE | 166 |
| VI. | PERSPECTIVES | 170 |
| VII. | PUBLICATIONS ET COMMUNICATIONS | 172 |
| VII.a. | Publications | 172 |
| VII.b. | Communication Orale | 172 |
| VII.c. | Communications par affiche | 172 |
| VIII. | BIBLIOGRAPHIE | 135 |

ABBREVIATIONS

ABREVIATIONS

| | |
|------------------|---|
| ADN | Acide désoxyribonucléique |
| ARN | Acide ribonucléique |
| ARN _t | ARN de transfert |
| ARN _m | ARN messenger |
| Ca | Capside |
| DIS | Dimer Initiation Site (site d'initiation à la dimérisation) |
| DE | Dimère Étendu |
| ΔG | Variation d'énergie libre |
| ΔH | Variation d'enthalpie |
| ΔS | Variation d'entropie |
| DLS | Dimer Linkage Site (site de liaison du dimère) |
| Env | Enveloppe |
| Gag | Group specific antigen |
| gp | glycoprotéine |
| HSQC | Heteronuclear Single Quantum Coherence (cohérence hétéronucléaire à 1 quantum) |
| IN | Intégrase |
| KC | Kissing-Complex (dimère boucle-boucle) |
| LTR | Long Terminal Repeat |
| MA | Matrice |
| NC | Nucléocapside |
| NCp | protéine de Nucléocapside |
| Nef | Negative regulatory factor (|
| NOE | Nuclear Overhauser Effect (effet Overhauser nucléaire) |
| NOESY | Nuclear Overhauser Effect Spectroscopy (spectroscopie par effet Overhauser nucléaire) |
| ns | nanosecondes |
| PBS | Primary Binding Site (site primaire de liaison) |
| PDB | Protein Brookhaven Databank (banque de données de protéines de Brookhaven) |
| Pol | Polymérase |
| PR | Protéase |
| Pr | précurseur protéique |
| ps | picosecondes |
| Ψ | séquence d'encapsidation de l'ARN génomique de VIH |

| | |
|-------------------|---|
| Rev | Regulator of virion expression (régulateur de l'expression du virion) |
| ρ_0 | valeur cible du RMSD lors d'une TMD/TMD ⁻¹ |
| ρ | valeur réelle du RMSD lors d'une TMD/TMD ⁻¹ |
| RMN | Résonance Magnétique Nucléaire |
| RMSD | Root Mean Square Deviation (déviatiion par la méthode du moindre carré) |
| RBE | Rev Binding Element (Elément de liaison de Rev) |
| RT | Transcriptase inverse |
| RRE | Rev Response Element (Elément de réponse de Rev) |
| SD | Splice Donnor (donneur d'épissure) |
| SIDA | Syndrome de l'immunodéficience acquise |
| //36 | famille de structures à 36 Å de rmsd du KC en forme de deux brins parallèles allongés |
| SU | Surface |
| TAR | TransActivation Responsive (amplificateur de la transcription) |
| Tat | TransActivator of Transcription (activateur de transcription) |
| TM | TransMembranaire |
| TMD | dynamique moléculaire dirigée directe |
| TMD ⁻¹ | dynamique moléculaire dirigée inverse |
| UU22 | famille de structures à 22 Å de rmsd du KC en forme de double U |
| Vif | Virion Infectivity Factor (Facteur d'infectivité du virion) |
| VIH-1 | Virus de l'immunodéficience humaine de type 1 |
| VIH-2 | Virus de l'immunodéficience humaine de type 2 |
| Vpr | Viral Protein R (protéine virale R) |
| Vpu | Viral Protein U (protéine virale U) |
| WHAM | Weighted Histogram Analysis Method (méthode d'analyse des histogrammes pondérés) |

I . INTRODUCTION

I . INTRODUCTION

Le syndrome de l'immunodéficience acquise ou SIDA a fait son apparition en 1981 aux Etats-Unis. En 20 ans, ce fléau s'est étendu aux régions les plus éloignées du globe, entraînant un nombre croissant de contaminations et de décès.

Depuis la découverte en 1983 du virus VIH, l'agent responsable du SIDA, par l'équipe du professeur Montagnier, des scientifiques du monde entier se sont penchés sur ce virus. Malgré des avancées scientifiques considérables telles que l'élaboration de moyens de diagnostic peu coûteux vers le milieu des années 80, le séquençage intégral du génome du VIH moins de 15 ans plus tard et la mise au point d'une thérapie anti-rétrovirale efficace en 1995, l'épidémie a continué de se propager, touchant quelques 5 millions de personnes au cours de la seule année 2002 et en tuant près de 3 millions cette même année selon l'OMS. Le continent le plus touché actuellement est l'Afrique sub-saharienne. Plus de 70% aujourd'hui des cas de porteurs du VIH y vivent et ce continent a supporté la grande majorité des décès imputables au virus du SIDA, soit 28 millions pour l'ensemble du monde fin 2002.

Le développement des multi-thérapies a permis d'augmenter l'espérance de vie des patients atteints de la maladie du SIDA. Cependant, la grande faculté de mutation du virus entraîne le développement de résistance aux médicaments. De plus les nombreux effets secondaires liés à ces thérapies se révèlent très durs à supporter pour le malade.

La recherche de nouvelles cibles thérapeutiques inhibitrices de l'activité du virus reste donc d'une grande importance dans la lutte contre le SIDA. Notre étude porte sur l'étude du mécanisme de la dimérisation des deux molécules d'ARN constituant le génome de VIH, qui s'avère être une étape clé du cycle de réplication du virus dans l'organisme. Cette étape est une cible potentielle pour de nouvelles thérapies.

II . GENERALITES

II. GENERALITES

II.a.Le VIH

II.a.i.Taxonomie et caractéristiques rétrovirales

Les rétrovirus, de la famille des *Retroviridae*, infectent principalement les vertébrés. Certains, dits endogènes, se sont finalement intégrés au génome de l'hôte et sont ainsi transmis héréditairement mais la majorité d'entre eux est exogène, leur transmission est donc réalisée par contagion entre individus distincts.

Au cours de leur évolution les rétrovirus ont divergé les uns des autres, présentant ainsi la capacité de provoquer des pathologies diverses. Ils conservent cependant des caractéristiques structurales et fonctionnelles communes permettant de les regrouper dans une même famille :

- l'information génétique est portée par un ARN simple-brin au lieu d'un ADN double-brin. Cette caractéristique est à l'origine de leur nom de rétrovirus, le préfix *rétro* se référant à la direction inhabituelle de transcription ARN → ADN incluse dans le cycle de réplication du virus. Ce passage du génome sous forme d'ADN s'effectue grâce à l'action de la transcriptase inverse, une ADN polymérase ARN-ADN dépendante.
- le génome est toujours diploïde, constitué de deux molécules identiques d'ARN génomique.
- leur morphologie est similaire. La particule virale est issue du bourgeonnement de la cellule hôte et présente un diamètre compris entre 70 et 130 nm.

Les rétrovirus se différencient en groupes selon leurs propriétés biologiques et biochimiques.

Le VIH est l'agent responsable du syndrome de l'immunodéficience humaine (SIDA) et appartient au groupe des *Lentivirus*. Ces virus se caractérisent par leur capacité à provoquer des maladies chroniques à évolution lente dont les signes cliniques ne se manifestent que plusieurs mois, voire plusieurs années après l'infection. Le virus du SIDA est réputé, en plus de sa virulence, pour sa faculté à muter. Deux lignées principales se distinguent actuellement, VIH-1 et VIH-2, la lignée de VIH-1 se subdivisant elle-même en trois

groupes : M, O et N. Ce sont les descriptions de variants de VIH-1 très divergents qui ont conduit à diviser l'ensemble des souches de VIH-1 en groupes. La première subdivision en deux groupes des souches de VIH-1 est intervenue en 1994 avec les groupes M et O (Charneau et al. 1994), M pour Majeur et O pour Outlier, un troisième groupe a ensuite été identifié en 1998 et nommé N pour non M/non O (Simon et al. 1998). Les VIH-1 groupe M sont responsables de la pandémie du SIDA. Au moins dix sous-types de ce groupe, nommés de A à J, ont été découverts à ce jour. La différence de séquence nucléotidique du gène env entre les sous-types est de 20% à 30% en général.

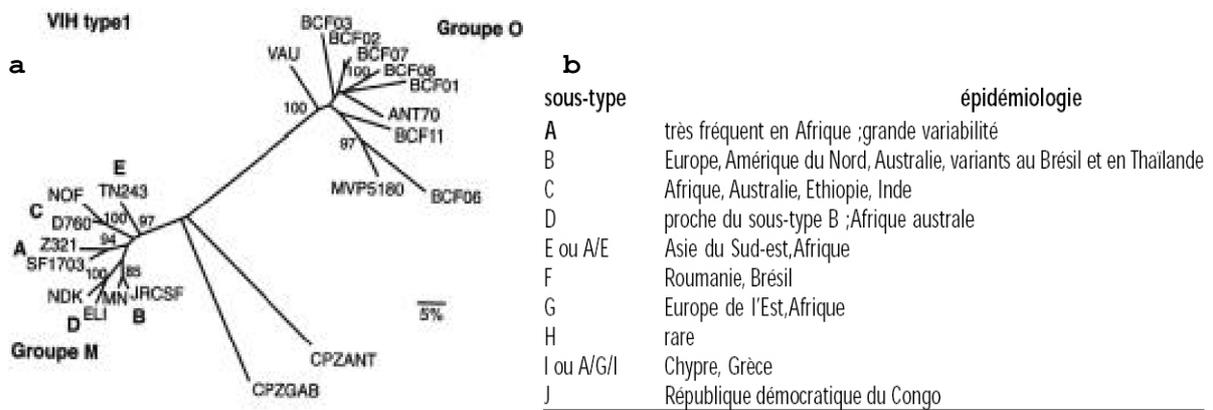


Figure 1 : a) Analyse phylogénétique des virus des groupes M et O ; b) Caractéristiques épidémiologiques pour les sous-types de VIH-1 groupe M (Brun-Vezinet et al. 1999).

Cette grande variabilité de VIH est liée aux erreurs d'incorporation de nucléotides qu'effectue la transcriptase inverse pendant la transcription de l'ARN viral en ADN (McGrath et al. 2001; Quinones-Mateu and Arts 1999; Termin 1993). Ce taux d'erreur est de 1 à 10 mutations par génome et par cycle et reflète d'une manière générale la capacité du virus à s'adapter à son environnement (An and Telesnitsky 2002). Ceci explique les résistances que le virus développe à l'encontre des anti-rétroviraux ainsi que sa faculté à échapper aux réponses immunes de l'organisme.

Par la suite, nous nous attacherons plus particulièrement à développer les propriétés et spécificités de VIH-1 groupe M.

II.a.ii.Aspects structuraux des particules virales

La structure du virion de VIH-1, établie par microscopie électronique, est caractéristique du groupe des *lentivirus* auquel elle appartient. Ces particules présentent un diamètre de 100 à 130 nm. Elles sont constituées d'une enveloppe lipidique (la membrane externe) entourant une première capsidie icosaédrique (la matrice) contenant un noyau interne dense (la nucléocapsidie ou capsidie) (Figure 2).

La bicouche lipidique externe de VIH-1 contient des spicules composés par l'association de deux glycoprotéines : la glycoprotéine de surface SU gp120 contenant les déterminants viraux qui vont se fixer à la cellule hôte et la glycoprotéine transmembranaire TM gp41, liée à la gp120, responsable de la fusion de l'enveloppe avec la membrane cellulaire. La protéine MA p17 tapisse la face interne de l'enveloppe et constitue ainsi la matrice.

La capsidie est formée par la répétition de la protéine Ca p24 chaperonnée par la protéine de nucléocapsidie NCp. Ces deux protéines sont issues du gène Gag. La capsidie contient le génome viral et les enzymes. Le noyau nucléoprotéique comprend l'ARN génomique associé à des molécules d'ARNt servant d'amorce à la transcription inverse. Ces deux ARN sont eux-mêmes en association étroite avec des protéines NC (Darlix et al. 1990; Prats et al. 1990; Stewart et al. 1990) dont le rôle est de les réunir, de les protéger de l'activité des enzymes cellulaires et d'intervenir dans l'assemblage des virions.

La protéine mature Pol compose un système enzymatique complexe réunissant la protéase (PR), la transcriptase inverse (RT) et l'intégrase (IN). D'autres protéines telles Tat, Nef, Vif, Vpr et Vpu sont également présentes dans le virion de VIH-1.

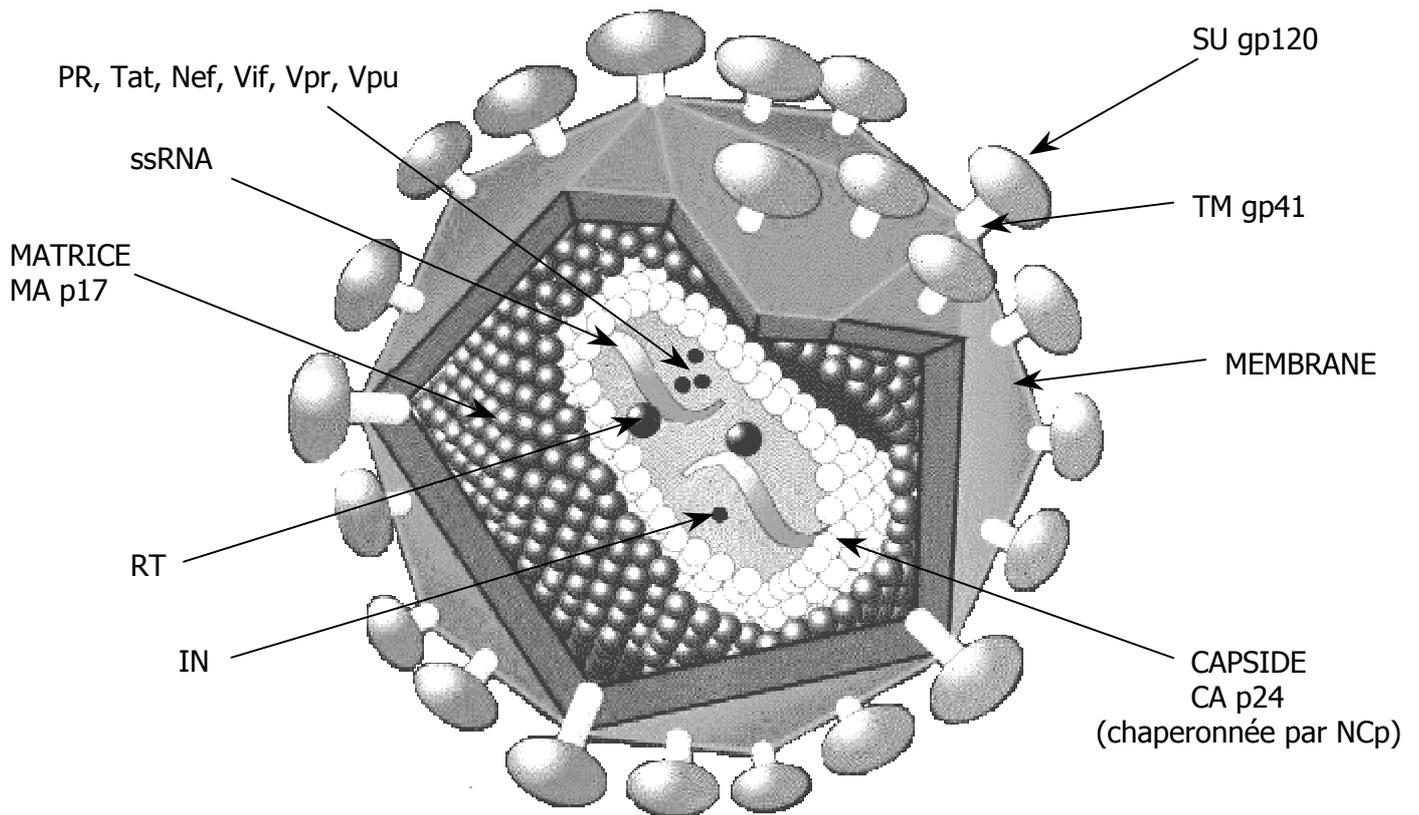


Figure 2 : Morphologie et structure de la particule virale de VIH-1

Un inventaire plus détaillé des fonctions des différentes protéines montrées dans la Figure 2 est reporté dans le Tableau 1, élaboré d'après (Pavlakis 1996), (Frankel and Young 1998).

| protéines virales | Fonctions |
|-------------------|-----------|
|-------------------|-----------|

→ Protéines de l'enveloppe

| | |
|----|---------------------------|
| TM | Protéine transmembranaire |
| SU | Protéine de surface |

→ Protéines structurales

| | |
|----|---|
| MA | Protéine de la matrice myristilée à son extrémité NH2 |
| CA | Protéine de la capside |
| NC | Protéine de nucléocapside, protéine chaperonne |

→ Protéines enzymatiques

| | |
|-----------|---|
| RT | Transcriptase inverse, responsable du passage ARN génomique en ADN proviral |
| IN | Intégrase, responsable de l'insertion de L'ADN proviral dans le patrimoine génétique de la cellule hôte |
| PR | Protéase, responsable de la maturation des précurseurs polyprotéiques |

→ Protéines régulatrices

| | |
|------------|--|
| Nef | (Negative Factor) augmente le taux de réplication et réduit le nombre de récepteurs CD4 à la surface de la cellule infectée |
| Rev | (Regulator of Virion expression) régule l'épissage des ARNm ainsi que leur transport à partir du noyau |
| Tat | (TransActIvator) activateur de la transcription de l'ADN proviral |
| Vif | (Virion Infectivity Factor) permet la formation des particules virales |
| Vpr | (Viral Protein R) intervient dans la translocation de L'ADN viral vers le noyau |
| Vpu | (Viral Protein U) augmente le nombre de virions relargués à la surface de la cellule infectée et diminue le niveau d'expression des récepteurs CD4 |

Tableau 1 : Description des différentes protéines impliquées dans le cycle rétroviral de VIH-1

II.a.iii. Cycle rétroviral de VIH-1

Le cycle répliatif commun à l'ensemble des rétrovirus est détaillé sur la Figure 3 (Cullen 1991; Young 2003).

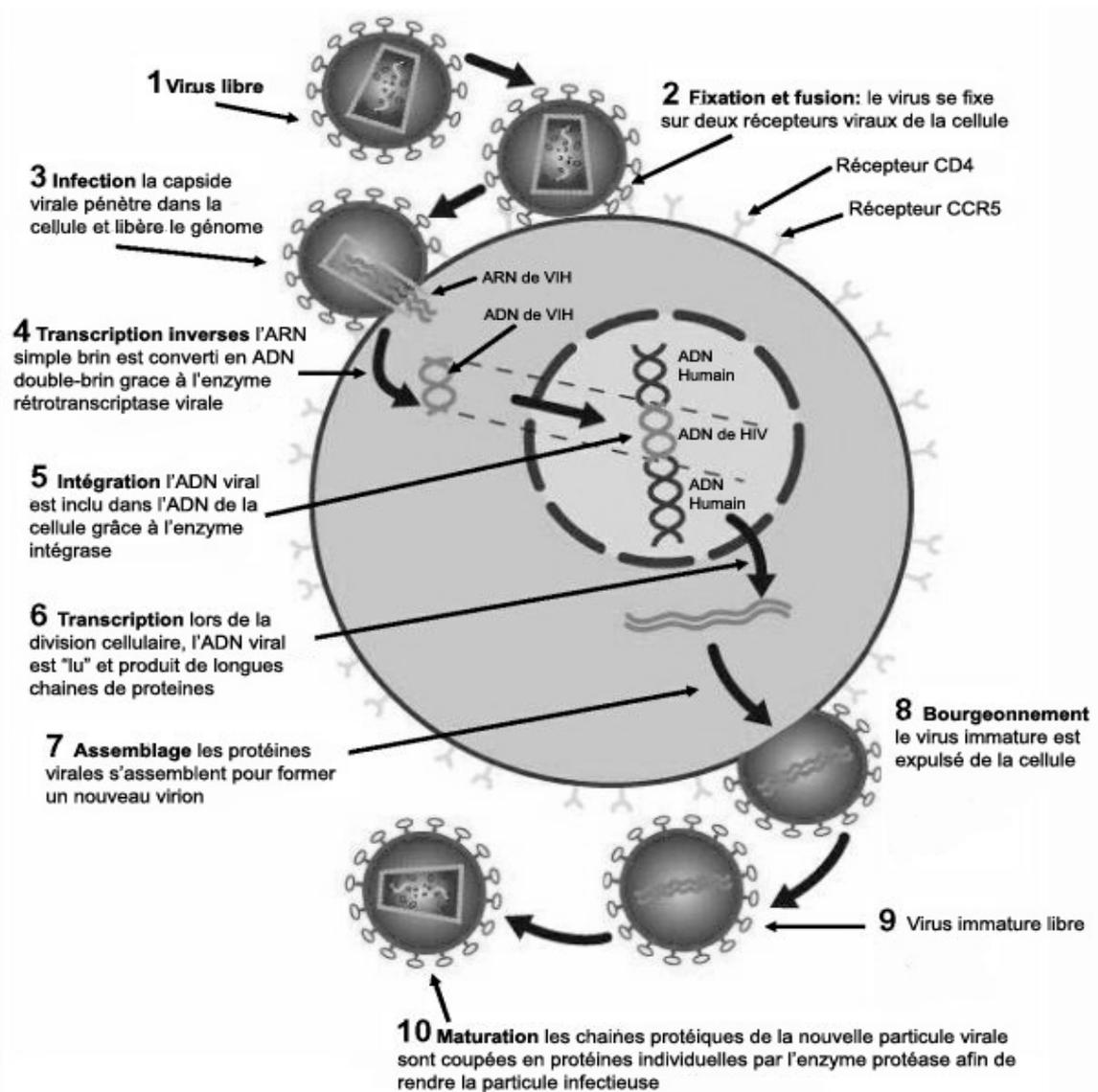


Figure 3 : Cycle de réplication de VIH-1

→ Entrée du virus

L'infection commence par l'adhésion de la particule virale à des récepteurs membranaires de la cellule cible. Chez VIH-1, les récepteurs spécifiques mis en jeu sont, d'une part la glycoprotéine virale d'enveloppe gp120 pour le virion, d'autre part des récepteurs glycoprotéiques extra-membranaires de la famille des immunoglobulines, les récepteurs CD4, pour la cellule-cible. Ces derniers sont principalement exprimés à la surface des lymphocytes T4, des macrophages, des cellules dendritiques des ganglions, de la rate et de l'épiderme (cellules de Langerhans) ainsi que sur les cellules microgliales du cerveau. Cette interaction met également en jeu des co-récepteurs, indispensables à gp41 pour lui permettre

d'exercer son activité fusionnante, dont deux types ont été identifiés à ce jour : la fusine, un récepteur transmembranaire couplé aux protéines G et des récepteurs de la β -chimiokines de type CCR5. Il en résulte une fusion des membranes permettant à la capsid virale de pénétrer dans le cytoplasme de la cellule. Une fois dans le cytoplasme, la capsid se désagrège et libère le génome (Grewe et al. 1990).

→ Transcription inverse

La transcription inverse de l'ARN génomique est initiée par la rétrotranscriptase virale. A l'issue de ce processus, l'ARN simple brin dimérique est converti en ADN double brin possédant à chaque extrémité des séquences LTR (Long Terminal Repeat) (Gilboa et al. 1979) qui, bien que de séquences identiques, jouent des rôles différents :

- en 5', le LTR est un promoteur puissant de la transcription
- en 3', le LTR fournit le signal de coupure qui précède la polyadénylation

L'ADN viral ainsi formé est ensuite transporté dans le noyau avec l'intégrase virale. Il se circularise et l'intégrase est fixée au niveau des LTR.

→ Intégration de l'ADN proviral au génome de la cellule-hôte

Le transport de l'ADN viral dans le noyau est possible grâce à son association avec un certain nombre de protéines virales, notamment la protéine de matrice MA qui contient un signal de localisation nucléaire autorisant ce transport. L'intégrase coupe ensuite les deux brins de l'ADN cellulaire pour introduire l'ADN viral. L'intégration semble pouvoir se faire dans de multiples sites de l'ADN cellulaire. Cette intégration dépend aussi de l'activation des cellules infectées : dans des cellules au repos, le processus est lent et incomplet.

Une fois intégré, l'activité transcriptionnelle du provirus est régulée par les facteurs de transcription de la cellule.

→ Transcription et traduction de l'ADN proviral

L'ARN polymérase II transcrit l'ADN proviral intégré au génome de la cellule hôte ce qui va conduire à la formation d'un ARN. Cet ARN va ensuite subir à l'intérieur du noyau plusieurs phénomènes d'excision-épissage générant plusieurs fragments d'ARN (ARN multi-épissés). Ces ARN vont être dirigés vers le cytoplasme pour y être traduits en protéines. Ces protéines correspondent aux protéines de régulation du cycle viral et notamment les protéines Nef, Rev et Tat. Les fonctions régulatrices de ces protéines vont varier :

- La protéine Nef est une protéine inhibitrice de la transcription. Elle a pour rôle de maintenir la latence virale. Son mode d'action est complexe : ancrée dans la membrane, elle va agir indirectement sur la séquence NRE (nef response element) située sur le LTR. C'est donc par des intermédiaires cellulaires qu'elle va provoquer l'inhibition de la transcription.
- Contrairement à la protéine Nef, la protéine Tat possède une propriété activatrice de la transcription. La protéine Tat se fixe sur la séquence TAR (Tat Responsive Element). Cette fixation a pour conséquence l'activation du processus de transcription.
- La protéine Rev a un rôle clé dans le cycle viral du VIH-1. Cette protéine se fixe sur le site RBE (Rev Binding Element) localisé dans la région RRE (Rev Response Element). Le recouvrement de cette région par les protéines Rev masque un site accepteur d'épissage ce qui aura pour conséquence de bloquer le phénomène d'excision-épissage des ARN transcrits et seuls des ARN non épissés ou mono épissés seront alors exportés vers le cytoplasme pour y être traduits.

La traduction des ARN pluri-épissés correspond donc à la première phase de l'expression génétique du VIH-1 caractérisée par la mise en place de protéines intervenant dans la régulation. La protéine Rev commande la transition de la phase de latence du VIH vers la phase de réplication en bloquant le phénomène d'épissage des ARN et permettant ainsi la production des différents éléments constitutifs des virions, à partir des ARN non et mono-épissés :

- les ARN_m non-épissés vont encoder pour les précurseurs Pr p55 (gag) et Pr p180 (gag-pol). Ces précurseurs sont des polyprotéines qui seront clivées ultérieurement. Le Pr p180, après clivage donnera la rétrotranscriptase, l'intégrase et une protéase virale. Le Pr p55 donnera des protéines constitutives de l'enveloppe virale, de la capsidie et de la nucléocapsidie. L'ARN non-épissé servira également de génome pour les nouveaux virions constitués.
- les ARN_m mono-épissés encodent pour le précurseur Pr gp160 qui, après maturation, donnera les glycoprotéines gp120 et gp41 de l'enveloppe virale.

Cette étape est schématisée Figure 4.

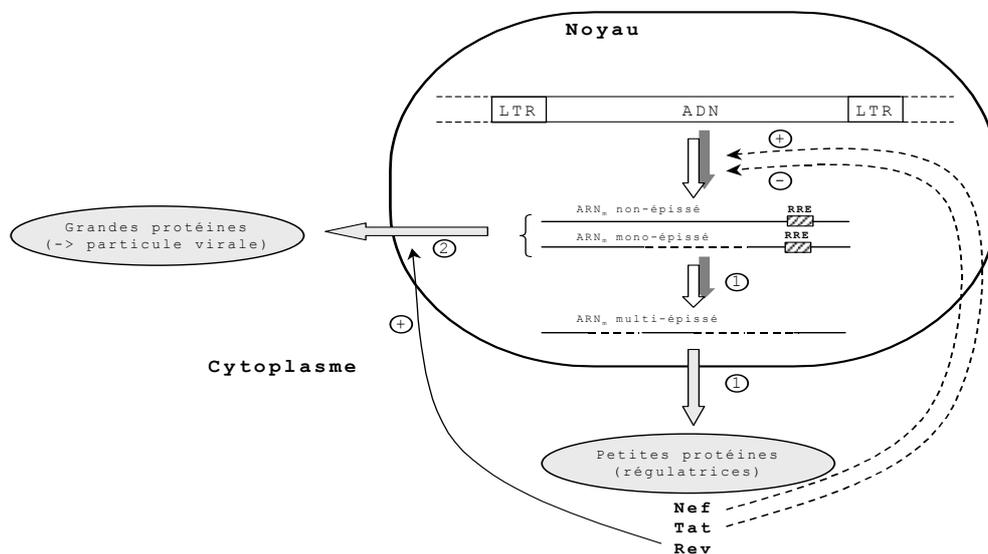


Figure 4 : expression et régulation du génome viral de VIH-1.

→ Assemblage et encapsidation

L'assemblage s'effectue au niveau de la membrane plasmique. L'encapsidation du génome rétroviral est assurée par le précurseur Pr55 dont le domaine Nc se fixe à la région d'encapsidation Ψ de l'ARN. Les éléments nécessaires au développement du virion sont également encapsidés ; par exemple dans le cas de VIH-1, l'ARN_t^{lys3} sera utilisée par la suite comme amorce à la rétrotranscription.

→ Bourgeonnement et maturation

L'agencement de la protéine Env, constituée des deux sous-unités TM et SU s'effectue pendant la période dite de bourgeonnement. Afin

d'acquérir ses qualités infectieuses, le virion subit une phase de maturation qui consiste en des coupures protéolytiques provoquant la libération des protéines de matrice, de capsid et de nucléocapside.

II.a.iv. Organisation du génome rétroviral

Le génome rétroviral de VIH-1 est codé par une molécule d'ARN génomique simple-brin d'environ 9.8kb de long (Figure 5)

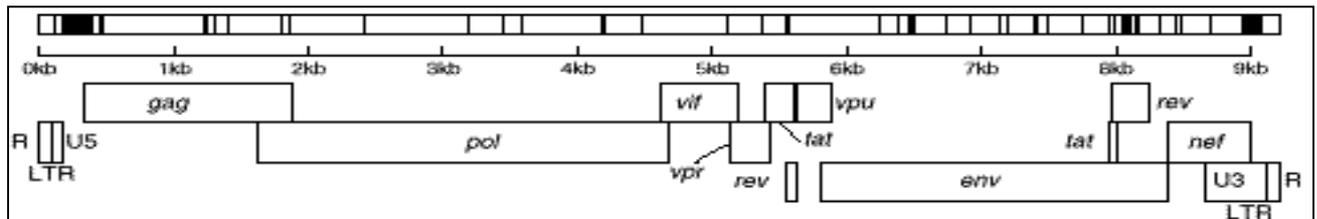


Figure 5 : Organisation générale du génome de VIH-1

La séquence provirale est encadrée par les deux régions non codantes identiques LTR constituées de séquences répétitives. Ces régions contiennent les signaux nécessaires à l'expression des protéines virales ainsi qu'à la synthèse et à l'intégration de l'ADN proviral dans la cellule hôte. Ces régions non codantes sont constituées par :

- une coiffe CAP en 5' et une queue de polyadénylation poly A en 3' comme dans tout ARN messenger ; ces éléments sont le fait de modifications transcriptionnelles générées par la machinerie cellulaire lors de la synthèse
- une région R (Redondant) présente à chacune des extrémités ; elle autorise le premier saut de brin lors de la rétrotranscription ; elle contient également la région TAR qui intervient via la protéine Tat dans l'expression de l'ADN proviral.
- une région U5; cette région très structurée contient un site de reconnaissance pour l'intégrase et semble participer à l'encapsidation de l'ARN viral.
- une région U3 dupliquée à l'autre extrémité de l'ADN proviral après la transcription inverse ; elle contient des signaux intervenant dans l'intégration de l'ADN proviral et dans la régulation de la transcription et un signal consensus de polyadénylation.

- une région PBS (Primary Binding Site), région où s'accomplit la fixation de l'ARN_t^{lys3} sur l'ARN génomique afin d'amorcer la transcription inverse.
- une région DLS (Dimer Linkage Site) située entre la région PBS et le gène Gag, contenant le site d'initiation à la dimérisation DIS (Dimer Initiation Site). Elle contient également le site d'épissage majeur SD (Splice Donor) et la séquence d'encapsidation Ψ de l'ARN génomique.

II.a.v. La dimérisation de l'ARN génomique de VIH-1

→ Un génome diploïde

L'une des caractéristiques communes à tous les rétrovirus est de posséder un génome sous forme diploïde, chaque particule virale contenant deux copies de l'ARN génomique positif (Canaani et al. 1973; Coffin 1979; Mangel et al. 1974; Stoltzfus and Snyder 1975). Il a été mis en évidence que ces deux brins d'ARN s'associent de façon non-covalente pour former des dimères. Grâce à la microscopie électronique, un site d'interaction ARN-ARN a été localisé à l'extrémité 5' terminale du génome (Bender et al. 1978; Bender and Davidson 1976; Kung et al. 1975; Kung et al. 1976; Murti et al. 1981; Rein 1994) (Figure 6).

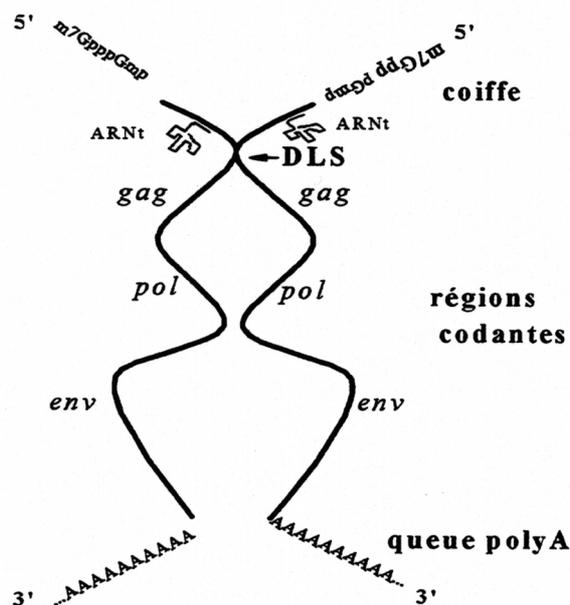


Figure 6 : Illustration du caractère diploïde du génome de VIH-1 et de son site de liaison dénommé DLS (Bender et al., 1976 ; Coffin, 1997).

Cette région a été nommée DLS (Dimer Linkage Site) et il a été montré qu'elle possède la faculté de dimériser spontanément en solution (Marquet et al. 1991; Sundquist and Heaphy 1993).

Des études sur cette région ont été effectuées pour les deux sérotypes les plus répandus chez VIH-1, les isolats Mal et Lai (correspondant respectivement aux sous-types A et B de VIH-1 groupe M). Ces recherches ont permis d'identifier une région de 206 nucléotides nommée domaine d'encapsidation Ψ comprenant le « major splice donor site » et le codon d'initiation de gag (Paillart et al. 1996d), ainsi qu'une région potentielle de structures secondaires complexes d'ARN (initialement observée par Harrison et Lever (1992)). Cette région de 110 nucléotides contient quatre tige-boucles indépendantes nommées SL1, SL2, SL3 et SL4.

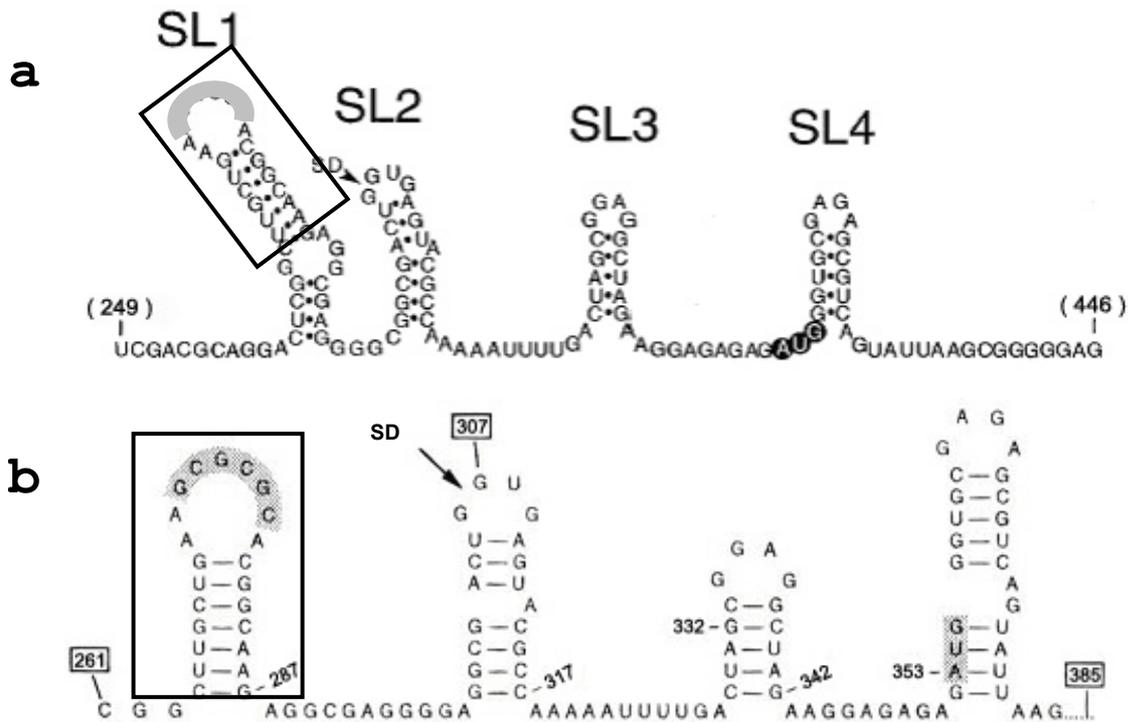


Figure 7 : Structure secondaire du domaine d'encapsidation Ψ de VIH-1_{Lai}. **a/** Modèle proposé par Clever et al. (1995), le DIS est noté en grisé; **b/** Modèle proposé par Berkhout et van Wamel (2000) d'après les données de Berkhout et al. (1996), la tige-boucle SL1 est simple, sans bulge.

Deux modèles de structure secondaire ont été proposés dans la littérature, le premier par Clever et al. (1995) montre une tige-boucle SL1 de 35 nucléotides comportant une boucle interne en milieu de tige, le second par Berkhout et van Wamel (1996; 2000) qui présente une tige-boucle simple de 23 nucléotides pour SL1.

Cependant (Takahashi et al. 2000) a montré l'importance, dans la séquence entière de 35 nucléotides de SL1, du bulge comme régulateur de la stabilité des tiges et guide du processus de dimérisation. Afin d'expliquer ce phénomène de dimérisation spontanée, d'autres études menées sur la souche VIH-1_{Mal} (Marquet et al. 1994; Skripkin et al. 1994) ont conduit à l'isolation d'une très courte séquence auto-complémentaire GUGCAC localisée en amont du site donneur d'épissage (désigné par SD sur la Figure 7). Cette même région a été décrite chez VIH-1_{Lai} par Laughréa et Jetté (1994) et Muriaux et al. (1995) comme responsable de l'initiation de la dimérisation, même à faible concentration ionique, avec une séquence un tant soit peu différente mais toujours auto-complémentaire GCGCGC. Cette région est hautement conservée, quoique des variations de séquence soient tolérées dans la boucle (Korber et al. 1997; St Louis et al. 1998), et n'est rencontrée qu'une fois sur tout le génome. L'utilisation de sondes chimiques et enzymatiques couplée aux résultats de repliement par minimisation ont permis par la suite de montrer que cette séquence se trouve au sommet d'une des quatre tige-boucle (SL1) du domaine ψ d'encapsidation (Clever et al. 1995) (voir Figure 7).

→ Implication de la dimérisation dans des étapes-clé du cycle rétroviral

• Encapsidation

Une des étapes du cycle dans laquelle la dimérisation semble le plus impliquée est l'étape d'encapsidation du génome viral. L'encapsidation est sélective et implique la reconnaissance de signaux sur l'ARN viral génomique, signaux absents sur les ARN_m viraux et cellulaires. Un domaine de l'ARN essentiel pour l'encapsidation a été identifié à proximité de l'extrémité 5' du génome viral (Aldovini and Young 1990; Clavel and Orenstein 1990; Hayashi et al. 1992; Lever et al. 1989). Cette région correspond au domaine ψ qui contient également la région d'initiation de la dimérisation. Les étapes de dimérisation et d'encapsidation de l'ARN viral génomique sont donc contrôlées par la même région. La suppression de l'une des quatre tige-boucle de la région évoquée ci-dessus entraîne des perturbations dans les processus de dimérisation comme d'encapsidation ; seule SL2 n'est pas nécessaire à l'encapsidation. En revanche, les tige-boucles SL1, SL3 et SL4

contribuent de manière équivalente au phénomène d'encapsidation (Clever and Parslow 1997; Harrison et al. 1998b; McBride and Panganiban 1996; McBride et al. 1997; Paillart et al. 1996b). D'autres études (Berkhout and van Wamel 1996; Laughrea et al. 1997; Paillart et al. 1996a) ont montré par ailleurs que des virions mutés au niveau de la séquence DIS provoquaient une réduction importante de la quantité d'ARN viral encapsidé. Ceci semble suggérer que les étapes de dimérisation et d'encapsidation sont liées entre elles et que cette séquence joue un rôle non négligeable dans le processus d'encapsidation.

Son implication dans d'autres étapes d'importance du cycle a également été suggérée (Paillart et al. 1996a).

- recombinaison génétique

Selon les travaux de Mikkelsen et al. (1996) et Haddrick et al. (1996), la nature dimérique de l'ARN génomique serait à l'origine de propriétés de recombinaison impliquées dans la grande variabilité rencontrée au sein de cette famille de virus (Coffin 1996a; Coffin 1996b; Haddrick et al. 1996). Or nous avons déjà évoqué plus haut l'importance fondamentale que cette grande variabilité du virus tenait dans sa capacité à résister aux traitements.

- traduction du gène gag

L'ARN génomique rétroviral 35S est un ARN_m pour la traduction des précurseurs Gag et Gag-Pol. Chez VIH-1, cette traduction est inhibée *in vitro* du fait de la dimérisation ; le DLS recouvre en effet une partie du site d'initiation de gag, en particulier une structure secondaire d'ARN située en 5' de la zone d'encapsidation ψ masque le codon AUG initiateur de la traduction du gène gag (Baudin et al. 1993; Bieth et al. 1990; Miele et al. 1996; Mougel et al. 1993; Rein 1994).

- transcription inverse

Shen et al. (2000) ont établi un lien entre dimérisation et transcription inverse. La tige-boucle DIS (nt 265-287) est située au sommet d'une tige-boucle interne nommée tige-boucle B (nt 260-264 et 288-294) (Laughrea et al. 1999) (voir Figure 7). La destruction de

cette tige-boucle B réduit d'environ 50% la dimérisation génomique et jusqu'à 85% de la synthèse d'ADN proviral, l'étape la plus touchée se trouvant être le transfert de brin de l'ADN lors de la transcription inverse, réduite d'approximativement 80%.

II.b.Le dimère de l'ARN génomique de VIH-1

Du fait de son rôle clé dans plusieurs des étapes du cycle de réplication, le mécanisme de la dimérisation a été étudié *in vitro* par de nombreuses équipes. Plusieurs modèles ont été proposés, aussi bien sur la nature des dimères impliqués dans ce processus que sur le mécanisme propre de ce processus.

II.b.i. Deux formes de dimères

Un premier modèle de dimère, issu des études menées par Marquet et al. (1991) et Sundquist et al. (1993), suggérait la formation d'une quadruple hélice d'ARN inter-brin stabilisée par des quartets de purines (Awang and Sen 1993). Ce modèle a été abandonné à la suite des études sur des fragments d'ARN mutés de Berkhout et al. (1993) et Muriaux et al. (1995) démontrant la non-implication de ces quartets dans la dimérisation, ce qui est par ailleurs confirmé par des expériences *in vivo* de Fu et al. (1994).

Peu après a été mise en évidence l'implication de la tige-boucle SL1 dans la dimérisation, et en particulier de la séquence auto-complémentaire à son sommet (Laughrea and Jette 1994; Marquet et al. 1994; Muriaux et al. 1995; Skripkin et al. 1994). Ces données ont suggéré un nouveau modèle pour la dimérisation, impliquant la reconnaissance boucle-boucle des ARN viraux lors de son initiation. Les études *in vitro* menées par Muriaux et al. (1996a; 1996b) sur la souche Lai de VIH-1 établissent la présence de deux dimères de stabilité différente, la formation d'un dimère dit « instable », correspondant à un « kissing-complex », se produisant à 37°C, celle d'un dimère dit « stable » ne s'opérant qu'aux alentours de 55°C. Le mécanisme de reconnaissance s'effectuerait alors de la façon suivante : les boucles auto-complémentaires situées au sommet de la tige-boucle SL1 s'associent par liaison Watson-Crick pour former un « kissing-complex » qui se réarrange en un dimère « stable » en double hélice lorsque la température augmente. (Figure 8)

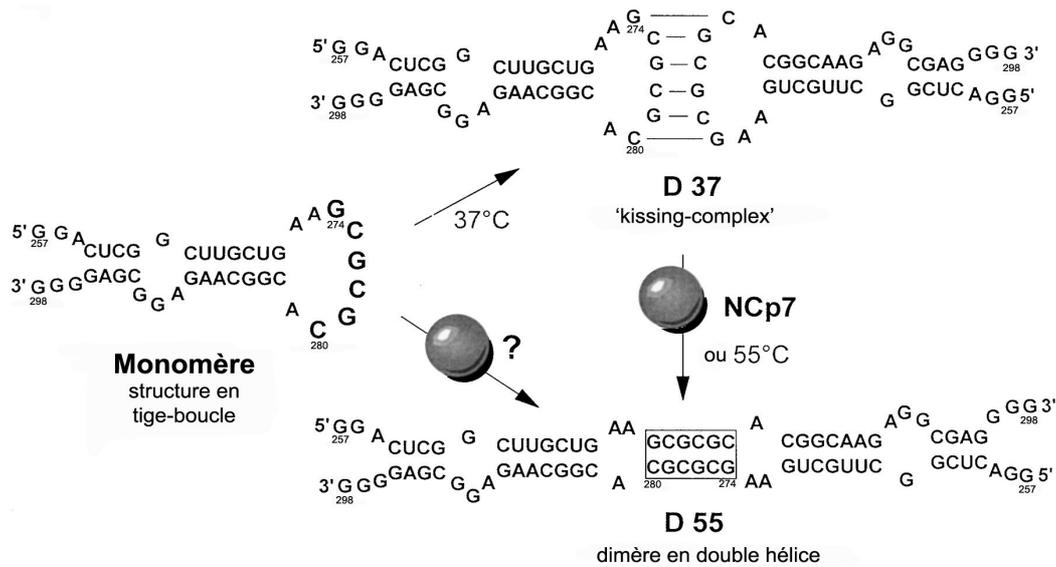


Figure 8 : Mécanisme de dimérisation *in vitro* de VIH-1_{Lai} et implication de la NCp7 dans ce processus (d'après Muriaux et al., 1996a)

Le modèle ainsi développé met en jeu un complexe boucle-boucle intermédiaire, imitant ainsi le processus de maturation de l'ARN dimérique intervenant *in vivo* et nécessaire à l'acquisition des propriétés infectieuses du virion (Fu et al. 1994). Ce modèle de dimérisation a déjà été trouvé chez d'autres rétrovirus, notamment MoMuLV par Girard et al. (1995) et Fu et al. (1993), et ASLV par Fossé et al. (1996) et Polge et al. (2000).

Des études thermodynamiques réalisées par spectroscopies RMN et UV sur un fragment court de 23 nucléotides de la tige-boucle SL1 de VIH-1_{Lai} (Theilleux-Delalande et al. 2000), incluant la séquence auto-complémentaire initiatrice du processus de dimérisation ont révélé que ce fragment possède toutes les propriétés requises pour la dimérisation, constituant ainsi un modèle utilisable à des fins de détermination structurale ainsi que pour l'étude du mécanisme de transition.

l'isolat Lai avec une mutation des quatre premières bases de chacune des extrémités. La structure déterminée n'a pas été publiée mais est en très forte adéquation avec la structure de Girard et al. (1999), et en particulier le motif en fermeture éclair des adénines dans les boucles internes se retrouve dans la structure de Mujeeb et al. (1999).

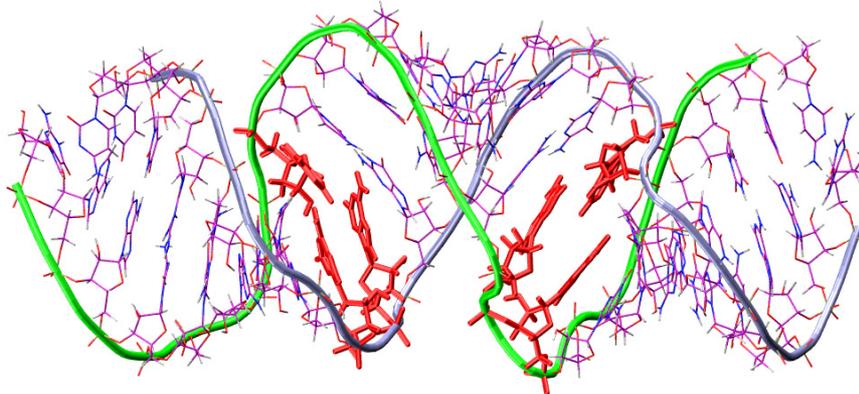


Figure 10 : Structure publiée par Girard et al. (1999) pour le dimère étendu de SL1_{Lai}. Sont représentées en bleu et en vert les brins 1 et 2 respectivement. Les adénines 272, 273 et 280 des boucles internes sont représentées en rouge.

La structure du dimère étendu de l'isolat Mal de SL1 a été résolue par cristallographie par Ennifar et al. (1999) (Figure 11). De même que pour l'isolat Lai, cette structure est formée de trois duplexes alignés sur un même axe et reliés par deux boucles internes. Les résidus constitutifs de ces cinq éléments sont les mêmes que pour la structure précédente en tenant compte des différences de séquence existant entre les isolats Mal et Lai. La principale divergence entre ces deux structures réside dans la position des purines au niveau des boucles internes. Dans cette structure, les résidus G273 et A280 forment des paires de bases inter-moléculaires non équivalentes. La paire de bases G273-A'280 est coudée de sorte qu'elle autorise la formation d'une liaison hydrogène G273-C281, tandis que la paire de base G'273-A280 est plane. Les résidus 272 présentent eux aussi une position très différente de celle de la structure RMN, ils sont rejetés vers l'extérieur de la structure. De plus, le résidu A'272 présente deux conformations possibles, l'une complètement tournée vers l'extérieur comme le résidu A272, l'autre « collée » contre la molécule.

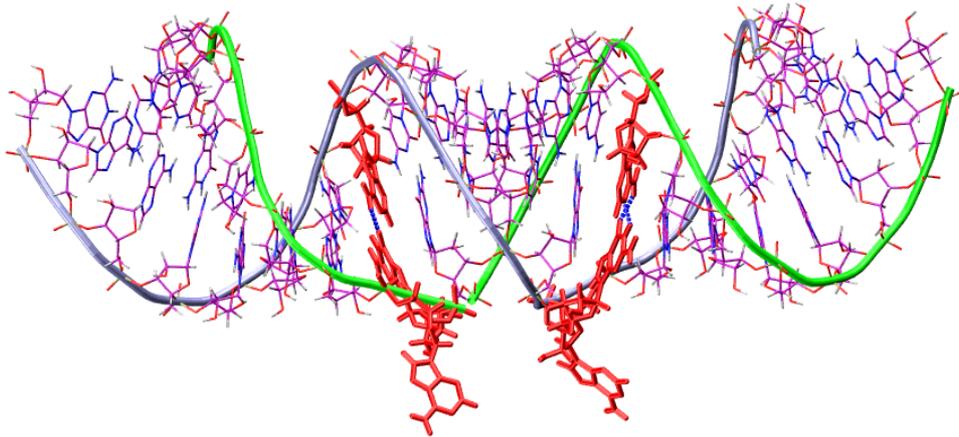


Figure 11 : Structure publiée par Ennifar et al.(1999) pour le dimère étendu de SL1_{Mal}, avec les brins 1 et 2 en bleu et vert respectivement, et les purines des boucles internes en rouge.

→ Kissing-complex « instable »

Si la forme générale du « kissing-complex » est globalement approuvée par tous (deux tige-boucles reliées par des liaisons hydrogène Watson-Crick au niveau des résidus de la séquence auto-complémentaire de leur boucle), la structure fine de ce dimère reste l'objet de nombreuses controverses. Des données structurales partielles parues dans la littérature sur le positionnement des résidus de ce complexe, notamment en ce qui concerne les résidus situés à la jonction entre tige et boucle, présentent de nombreuses contradictions entre elles et l'élucidation de la structure du «kissing-complex» par RMN et cristallographie n'a en rien concilié ces différents points de vue du fait des larges divergences qui existent entre les conformations publiées.

Les premières données sur la structure proprement dite de la tige-boucle SL1 ont été apportées par Paillart et al. (1997). Des expériences de mutagenèse dirigée couplées à une modélisation tridimensionnelle de la tige-boucle ont permis de mettre en relief deux points essentiels dans la conformation des résidus à la jonction tige-boucle : d'une part, les adénines en position 272 et 280 forment une paire de base *trans* entre le côté Watson-Crick de A272 et le côté Hoogsteen de A280 ; d'autre part, les résidus R273 forment un triplet de bases avec les paires centrales de la boucle G'276-C277 et G276-C'277 mais ce triplet de bases diffère selon qu'il s'agit de A273 ou de G273. Dans l'isolat Mal ou sous-type A,

le résidu G273 forme un triplet de bases avec la paire G'276-C277 tandis que l'atome N6 de l'adénine en 273 de l'isolat Lai (ou sous-type B) interagit avec les sites Hoogsteen de la paire G276-C'277.

Ces résultats sont corroborés par plusieurs autres travaux.

Une étude RMN effectuée sur un 19-mer de l'isolat Mal de la tige-boucle SL1 par Dardel et al. (1998) montre en effet l'existence d'une liaison non-canonique entre les résidus 6 et 14 du 19-mer, correspondant dans la séquence totale aux adénines 272 et 280. Il ressort de plus de ces expériences que les trois hélices du « kissing-complex », formées des tiges de chaque monomère et de l'hélice boucle-boucle sont alignées entre elles. En revanche, aucune donnée sur le positionnement de la guanine 273 n'a pu être extraite de ces expériences.

Jossinet et al. (1999) se sont attachés à montrer de façon plus précise, par des expériences chimiques, les interactions existant entre les résidus des boucles. Il découle de leurs expériences un comportement différent de R273 selon la nature du résidu impliqué. Dans le «kissing-complex» de sous-type A, les résidus G273 et G'273 forment des triplets de bases intermoléculaires avec les paires G276-C'277 et C277-G'276 du centre de l'hélice boucle-boucle, tandis que les adénines A273 et A'273 du «kissing-complex» de sous-type B forment des triplets de bases intramoléculaire avec les paires G276-C'277 et C277-G'276 respectivement. L'étude fine de l'interaction entre les adénines 272 et 280 révèle que deux liaisons hydrogène sont formées, impliquant les atomes N3 et H2 du résidu A272 et NH₂-6 et N7 du résidu A280.

Enfin des études par détermination de constantes de dissociation et substitution des bases dans la boucle effectuées par Lodmell et al. (2001) sont en parfait accord avec les observations détaillées ci-dessus. Il en ressort en effet que le résidu R273 est impliqué dans une liaison hydrogène non-canonique et qu'il existe un appariement entre les résidus A272 et A280.

Il est à noter qu'au cours de ces divers travaux, la jonction entre la tige et la boucle considérée ne comportait que les trois purines entourant la séquence auto-complémentaire, la tige étant ainsi constituée de sept paires de bases et l'hélice boucle-boucle de six paires de bases. De fait, les données RMN (Dardel et al. 1998)

révèlent un appariement entre les résidus G271 et C281 pour l'isolat Mal. Cependant les études de substitution de bases de Lodmell et al. (2001) suggèrent que, si cet appariement des résidus 271 et 281 est indispensable à la formation du dimère pour la tige-boucle SL1 du sous-type A, il ne semble pas qu'il en soit de même pour le sous-type B.

Cette dernière observation apparaît capitale lorsque l'on considère la structure résolue par spectroscopie RMN publiée par Mujeeb et al. (1998). Leur étude s'est portée sur la tige-boucle de 23 résidus correspondant à la séquence de l'isolat Lai muté au niveau de la séquence de la tige pour permettre une meilleure stabilité du « kissing-complex ». La structure ainsi déterminée du « kissing-complex » s'est révélée en contradiction avec plusieurs des points évoqués précédemment. La différence la plus flagrante réside dans le fait que les tiges du «kissing-complex» dans la structure résolue par spectroscopie RMN ne comportent pas 7 mais 6 paires de bases. En effet, les résidus G7 et C17 du 23-mer, correspondant aux résidus G271 et C281 dans la séquence totale, se révèlent être non-appariés, ce qui confère à cette structure une boucle de 11 nucléotides avec 6 de ces nucléotides engagés dans le complexe boucle-boucle. La jonction tige-boucle se trouve donc cette fois être composée de 5 résidus. Cette configuration inattendue est en particulier due à une conformation *syn* du résidu C17. D'autres divergences avec les données pré-établies ressortent de cette étude. Notamment, toujours au niveau de la jonction, la structure RMN révèle une liaison non-canonique inter-moléculaire entre les purines A9 et A'16 (respectivement A273 et A'280) au lieu de la liaison non-canonique intra-moléculaire A272-A280 évoquée jusqu'ici. La structure présente de plus un angle entre les axes de ses tiges compris entre 90° et 120°, ce qui contredit l'hypothèse de l'empilement coaxial des trois hélices avancée par Dardel et al. (1998). Enfin, les résidus A8 et C17 d'un monomère (respectivement A272 et C281) et le résidu A9 (A273) de l'autre se trouvent engagés dans un empilement stabilisant, hypothèse qui n'avait pas jusqu'alors été envisagée. (Figure 12)

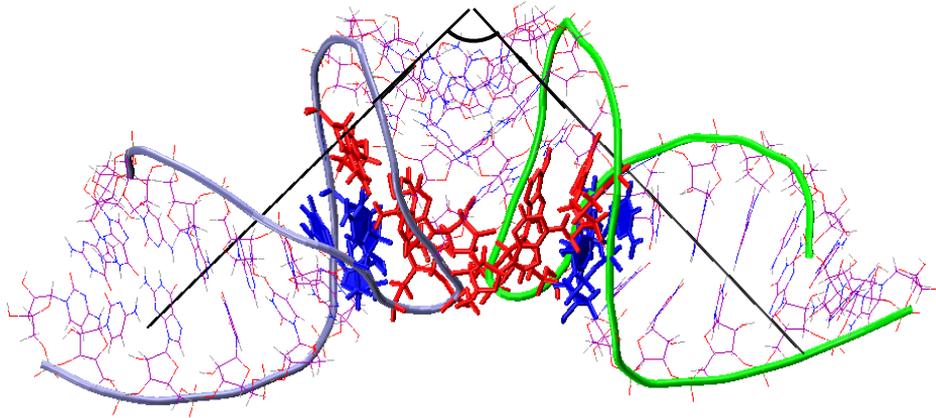


Figure 12 : Structure publiée par Mujeeb et al. (1998) pour le dimère en « kissing-complex » de SL1 Δ . Nous avons représenté respectivement en bleu et vert les monomères 1 et 2, en rouge les adénines à la jonction tige-boucle de chaque monomère et en bleu les résidus G271-C281 de chaque monomère. Les axes des tiges sont également représentés.

Une autre structure, cette fois résolue par diffraction de rayons X, a été publiée par Ennifar et al. (2001) (Figure 13). Elle se trouve être en désaccord sur plusieurs points, et avec la structure RMN, et avec les autres données expérimentales mentionnées précédemment. Les deux isolats Mal et Lai ont fait l'objet de cette investigation structurale et les structures déterminées pour ces deux dimères de SL1 sont très proches l'une de l'autre. Comme dans la structure RMN de Mujeeb et al. (1998), les sucres des résidus G271 sont en conformation C2'-endo et pour l'isolat Lai les résidus U270 et G282 forment une paire de bases Wobble. Par contre, les tiges contiennent 7 paires de bases alors que l'hélice boucle-boucle n'en contient que six. Ces trois hélices se trouvent empilées de façon coaxiale. Ces observations sont en accord avec les données RMN de Dardel et al. (1998), ainsi qu'avec les études expérimentales ne considérant que trois résidus dans la jonction tige-boucle. En revanche, la position de ces résidus diffère totalement des observations précédentes. Ces structures présentent en effet un positionnement des résidus A272 et R273 empilés ensemble et tournés vers l'extérieur de la structure, alors que les résidus 280 restent tournés vers l'intérieur des boucles et empilés dans l'alignement G279-A280-C281. Il n'y a donc dans ces structures aucune des liaisons hydrogène non-canoniques envisagées précédemment entre les purines des jonctions.

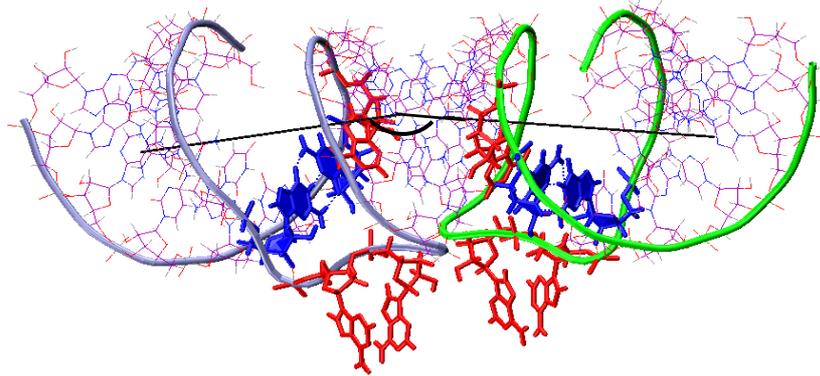


Figure 13 : Structure publiée par Ennifar et al. (2001) pour le dimère en « kissing-complex » de SL1_{Lai}. Nous avons représenté respectivement en bleu et vert les monomères 1 et 2, en rouge les adénines à la jonction tige-boucle de chaque monomère et en bleu les résidus G271-C281 de chaque monomère. Les axes de chacune des tiges sont également représentés.

Enfin, Kieken (2004) a déterminé la structure du « kissing-complex » à partir d'une séquence mutée du 23-mer de la tige-boucle SL1 de VIH-1_{Lai}. A cause de l'attribution équivoque des résidus A272 et A273, deux familles de structures sont proposées. Ces familles montrent à leur tour des résultats très différents des deux structures précédentes. Certes, comme pour la structure cristallographique, les tiges contiennent 7 paires de bases et l'interaction boucle-boucle est constituée de 6 paires de bases. Cependant contrairement à la structure cristallographique, l'ensemble des nucléotides de l'interaction boucle-boucle ne présente pas un agencement hélicoïdal mais un empilement de plans inclinés d'environ 50° vers la gauche par rapport à l'axe de l'hélice. Les deux familles se distinguent par les positions des nucléotides à la jonction de la tige et de la boucle. Dans les deux familles, le résidu A16, correspondant à A280 dans la séquence totale, se trouve empilé avec le résidu G'10 (soit le résidu G'274 de la séquence totale). Les deux adénines en 5' présentent par contre des conformations radicalement différentes selon la famille de structures décrite par Kieken (2004). Dans la première famille, les adénines 272 et 273 sont tournées vers l'intérieur de la boucle. L'adénine 272 possède un angle dièdre O3'-P-O5'-C5' d'environ -40°, donc proche de la valeur d'une hélice A qui est de -70°, et est partiellement empilée avec le résidu G271 avec lequel il forme par ailleurs une liaison hydrogène. L'adénine 273 peut, elle, former une

liaison hydrogène avec le résidu C281. L'angle entre les axes des tiges est d'environ 95° . (Figure 14)

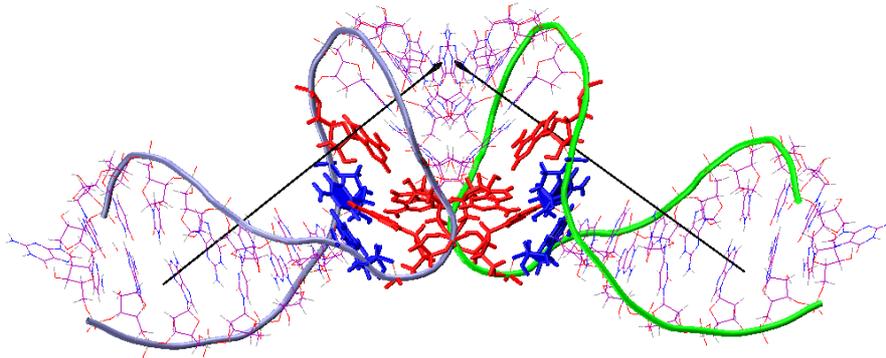


Figure 14 : Structure caractéristique de la famille 1 déterminée par Kieken (2004) pour le dimère en « kissing-complex » de SL1_{Lai}. Sont représentés respectivement en bleu et vert les monomères 1 et 2, en rouge les adénines à la jonction tige-boucle de chaque monomère et en bleu les résidus G271-C281 de chaque monomère. L'axe des tiges est également représenté.

Dans la deuxième famille, les adénines en 5' présentent des configurations opposées : le résidu A272 est tourné vers l'extérieur de la boucle et forme une liaison hydrogène avec le résidu C281 de l'autre monomère tandis que le résidu A273, tourné vers l'intérieur de la boucle, forme des liaisons hydrogène avec les résidus G274 et A280 du même monomère. Ce résidu A272 contient un angle dièdre O3'-P-O5'-C5' d'une valeur surprenante, d'environ 160° . L'angle entre les axes des tiges est légèrement supérieur à celui de l'autre famille, environ 110° . (Figure 15)

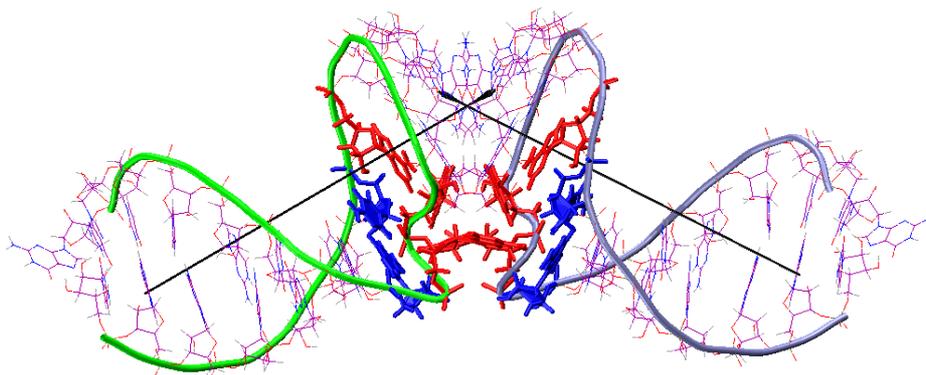


Figure 15 : Structure caractéristique de la famille 2 déterminée par Kieken (2004) pour le dimère en « kissing-complex » de SL1_{Lai}. Sont représentés respectivement en bleu et vert les monomères 1 et 2, en rouge les adénines à la jonction tige-boucle de chaque monomère et en bleu les résidus G271-C281 de chaque monomère. L'axe des tiges est également représenté.

Les divergences importantes existant entre les données structurales déterminées par expériences chimiques et biochimiques et les structures parues dans la littérature sur la tige-boucle SL1 de VIH-1 donnent à penser que la structure de la tige-boucle SL1 est très flexible et existe sous de nombreuses formes. Ces observations ont conduit différentes équipes, y compris la nôtre, à s'intéresser à la dynamique de cet ARN.

II.b.iii. Les simulations de dynamique moléculaire de la tige-boucle SL1

Les premières simulations de dynamique moléculaire sur la tige-boucle SL1 ont été publiées par Pattabiraman et al. (2002). Ces dynamiques ont été réalisées à partir de modèles de structures du «kissing-complex» de la tige-boucle SL1 obtenus en tenant compte des données expérimentales publiées par Paillart et al. (1997), Dardel et al. (1998) et Jossinet et al. (1999). Trois modèles ont été ainsi établis en vue d'étudier leur dynamique en solution, deux de ces modèles en utilisant la séquence SL1 de VIH-1_{Mal}, le troisième en utilisant la séquence de VIH-1_{Lai}.

Les trois modèles forment quasiment une double hélice A contenant cependant des imperfections du fait des trois purines à la jonction tige-boucle. Les deux modèles dénommés HIV-1 MAL et HIV-1 LAI contiennent les liaisons non-canoniques entre les adénines 272 et 280 décrites par Jossinet et al. (1999) ainsi que des triplets de base entre le résidu R273 et les atomes de la base du nucléotide G276 (plus ceux de son groupement phosphate en ce qui concerne HIV-1 LAI). L'autre modèle basé sur la séquence de VIH-1_{Mal} contient l'interaction non canonique entre les adénines 272 et 280 et les triplets de bases intermoléculaires décrits par Paillart et al. (1997). Il est référencé sous le nom de modèle de Westhof. Seuls les nucléotides n°268 à 284 ont été utilisés pour le calcul. Il est à noter que dans ces modèles, Pattabiraman et al. (2002) ont considéré les résidus G271 et C281 comme appariés.

Ces trois modèles ont été simulés en solution en présence de contre-ions sodium à une température de 300K en utilisant le programme académique AMBER-4.1 (Pearlman et al. 1995) sur une très courte durée de 400ps. Les simulations révèlent la formation d'empilements

entre les paires A272-A280 et G271-C281 ainsi qu'entre les paires A272-A280 et G'274-C279 contribuant à la stabilisation de la jonction tige-boucle. Cependant, les liaisons non-canoniques entre les résidus A272 et A280 semblent relativement instables et se rompent par intervalles au cours de la simulation. En revanche, le résidu R273 forme deux liaisons hydrogène, l'une avec G276, l'autre avec C277 qui semblent maintenir la boucle en conformation A. Forts de ces observations, Pattabiraman et al. (2002) ont envisagé l'hypothèse que le « kissing-complex » soit en fait pré-formé à la conversion en dimère étendu.

Les simulations numériques de Pattabiraman et al. (2002) présentent cependant une certaine faiblesse, principalement du fait que ces simulations sont basées sur des conformations issues de la modélisation moléculaire et non pas d'expérimentations structurales, ainsi que du fait que la durée de simulation reportée est trop courte pour voir se produire d'éventuels changements conformationnels importants.

Reblova et al. (2003) ont réalisé des expériences de simulations similaires mais en se basant sur les structures expérimentales parues dans la littérature et sur des temps de simulation beaucoup plus longs. Les structures des isolats Lai et Mal issues des expériences de diffraction de rayons X (Ennifar et al. 2001) en présence et en absence d'ions magnésium, ainsi que la structure de l'isolat Lai muté au niveau des tiges (Mujeeb et al. 1998) en présence seulement de contre-ions sodium, ont fait l'objet de ces investigations. Toutes ces structures ont été simulées en solution à la température de 300K dans un environnement électriquement neutre avec des temps de simulation variables, de 3 ns à 7.5 ns. Les résultats obtenus pour les structures cristallines varient en fonction de la séquence utilisée et de la présence ou non des ions magnésium. Pour l'isolat Mal, la structure cristalline présente une poche centrale contenant 3 ions magnésium. Au cours de la simulation en présence de ces ions, deux de ceux initialement présents dans la poche y restent à demeure pendant toute la durée de la simulation, deux autres ions magnésium se trouvant en permanence au bord de cette poche. Les purines 272 et 273 des jonctions subissent des fluctuations au cours de la simulation, de nature différente selon

le brin auquel elles appartiennent : les purines du brin A restent relativement proches de la structure initiale tandis que les purines du brin B tendent à rentrer vers l'intérieur de la structure. Cette différence de comportement peut s'expliquer par le fait que la structure cristallographique publiée dans la littérature est asymétrique, avec une conformation de G273 du brin 1 en *anti* alors que celle de G273 du brin 2 est en *syn*. La simulation de ce même isolat en absence d'ions magnésium montre une occupation permanente de la poche centrale par 3 ions Na⁺ et un comportement des purines de jonction plus stable, leur empilement se trouvant cependant partiellement réduit du fait d'une plus grande fluctuation des A272 que des G273. Les simulations de l'isolat Lai présentent de même une poche centrale et des fluctuations conformationnelles au niveau des purines de jonction. La simulation de cet isolat en absence d'ions magnésium révèle un comportement assez similaire à celui de l'isolat Mal, la poche étant occupée par trois ions sodium pendant la simulation et les purines des jonctions restant essentiellement rejetées vers l'extérieur de la structure. En revanche, le comportement de cet isolat en présence d'ions magnésium présente des divergences importantes avec celui de l'isolat Mal. En particulier, la poche centrale n'est jamais occupée par des ions magnésium, ni dans la structure initiale ni au cours de la simulation. De fait, cette poche subit une réduction de volume au cours de la dynamique qui la rend trop petite pour recevoir un complexe de magnésium hexahydraté. La réduction de cette poche survient à la suite d'un rapprochement des quatre adénines des jonctions entre elles, allant jusqu'à former une tige de quatre bases empilées. On observe toutefois la présence dans la poche de deux ou trois ions sodiums variables. Les conclusions de Reblova et al. (2003) sur les structures cristallines sont en faveur d'une bonne stabilité de ces structures en solution, meilleure selon eux que celle de la structure déterminée par spectroscopie RMN. Pour cette dernière, la simulation de 3 ns réalisée en présence d'ions sodium a révélé une augmentation de l'angle entre les axes des tiges, donc une linéarisation de la structure et la disparition de l'empilement A272/A'273/C281. Ils remarquent de plus que les résidus G271 et C281

deviennent coplanaires au cours de cette simulation sans toutefois observer d'appariement entre ces bases.

Ces résultats se trouvent en partie contredits par les simulations publiées par Beaurain et al. (2003), effectuées sur les deux structures existantes de l'isolat Lai, la structure cristalline (Ennifar et al. 2001) et la structure RMN (Mujeeb et al. 1998). Ces simulations ont été réalisées sur des temps plus longs, 5 ns pour la structure RMN, 10 ns pour la structure cristallographique. Les résultats obtenus lors de la simulation de la structure RMN sont relativement proches de ceux obtenus par Reblova et al. (2003). En effet, Beaurain et al. (2003) observent une décroissance de la courbure de la structure RMN et une tendance de G271 et C281 à devenir coplanaires. Ils constatent de surcroît que les valeurs des angles dièdres, très différents des valeurs d'une hélice A dans la structure initiale - pierre d'achoppement de nombreuses polémiques - reviennent rapidement au cours de la simulation vers des valeurs standard sans altérer la structure RMN de façon significative. Enfin, le résidu A273 ne subit pas de mouvements important bien que sa liaison non-canonique avec A'280 semble rompue. La paire de base A'273-A280 se trouve quant à elle engagée dans un empilement entre les résidus G271/A280 d'une part et C281/A'273 d'autre part. La simulation de la structure cristallographique en revanche présente des résultats assez différents. La structure reste globalement linéaire et coaxiale mais subit de plus grandes évolutions que la structure RMN. Les purines des jonctions, en particulier, se tournent vers l'intérieur de la structure jusqu'à se trouver « enfouies » à l'intérieur, entièrement en ce qui concerne les purines du monomère 1, partiellement pour celles du monomère 2. De plus, des désappariements se produisent au cours de la simulation, notamment l'ouverture de la paire G271-C281 du monomère 1 ainsi que de la paire G273-C'279 de l'hélice boucle-boucle. L'ouverture de cette dernière paire de bases semble nécessaire au repositionnement du résidu A280 de sorte qu'il puisse former une liaison hydrogène avec le résidu A272 « enfoui » dans la structure. Enfin, on observe pour les deux résidus G271 un changement de conformation du sucre, de C2'-endo en C3'-endo.

Ces observations ont conduit Beaurain et al. (2003) à penser que la structure RMN, pourtant plus irrégulière au premier abord, serait plus stable en solution que la structure cristallographique, bien que celle-ci soit très proche d'une hélice A régulière. Enfin, Beaurain et al. (2003) soumettent l'hypothèse que ces deux structures représenteraient en réalité deux états transitoires différents du « kissing-complex ».

→ Mécanismes de transition

Si les acteurs de la transition commencent à être relativement connus, on sait encore réellement peu de choses à l'heure actuelle sur le mécanisme proprement dit de la transition. Différents mécanismes ont été avancés à ce jour.

Le premier modèle, proposé par (Theilleux-Delalande et al. 2000), consiste en une fusion concertée des deux tiges engagées dans le « kissing-complex », suivie d'un réappariement inter-brin dans le but de reconstituer la double hélice du dimère « stable ». Les brins ainsi dissociés pourraient ensuite se réappairer mais de façon inter-moléculaire. Tout au long de ce processus, les résidus des séquences auto-complémentaires des boucles restent appariés. (Figure 16)

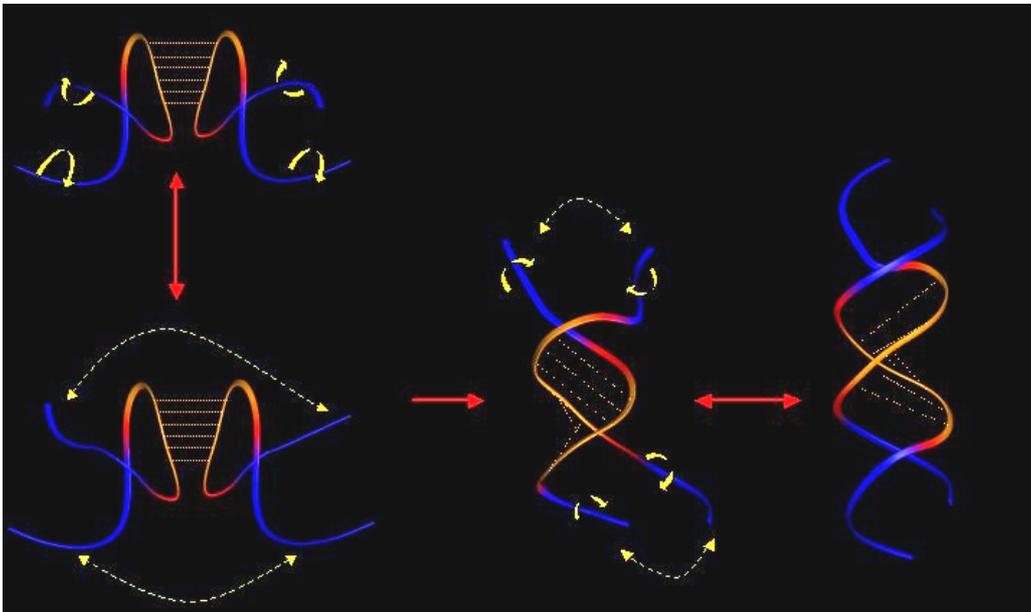


Figure 16 : Mécanisme de transition entre le dimère instable et le dimère stable, impliquant une fusion concertée des tiges suivie d'un réappariement inter-brin, suggéré par Theilleux-Delalande et al. (2000)

D'autres modèles de transition ont été récemment proposés, au vu des nouvelles données structurales qui ont été publiées. Notamment, la résolution par diffraction de rayons X de la structure du « kissing-complex » pour les isolats Lai et Mal a conduit Ennifar et al. (2001) à proposer un modèle de substitution au mécanisme de transition. Selon les auteurs, le passage du complexe boucle-boucle au duplexe pourrait s'expliquer par un mécanisme de clivage suivi d'une religation croisée, grâce à une double trans-estérification. Cette observation est basée sur le fait que, dans leur structure, les deux résidus R273 de la tige-boucle SL1 sont très proches l'un de l'autre et en conformation C2'-endo. Les groupements 2'-OH de ces résidus sont ainsi orientés de manière favorable à une attaque du groupement phosphate situé en 3', permettant une coupure du brin d'ARN par trans-estérification conduisant à la formation, d'un côté d'un cycle à cinq, de l'autre d'un groupement 5'-OH. Il s'ensuivrait un mouvement des deux résidus R273 aboutissant à une religation croisée par une nouvelle double trans-estérification, cette fois entre le phosphate du cycle d'un brin et le 5'-OH du deuxième brin, afin de former le dimère étendu. Ce mécanisme ne nécessite l'ouverture des paires de bases ni des tiges, ni de l'interaction boucle-boucle. (Figure 17)

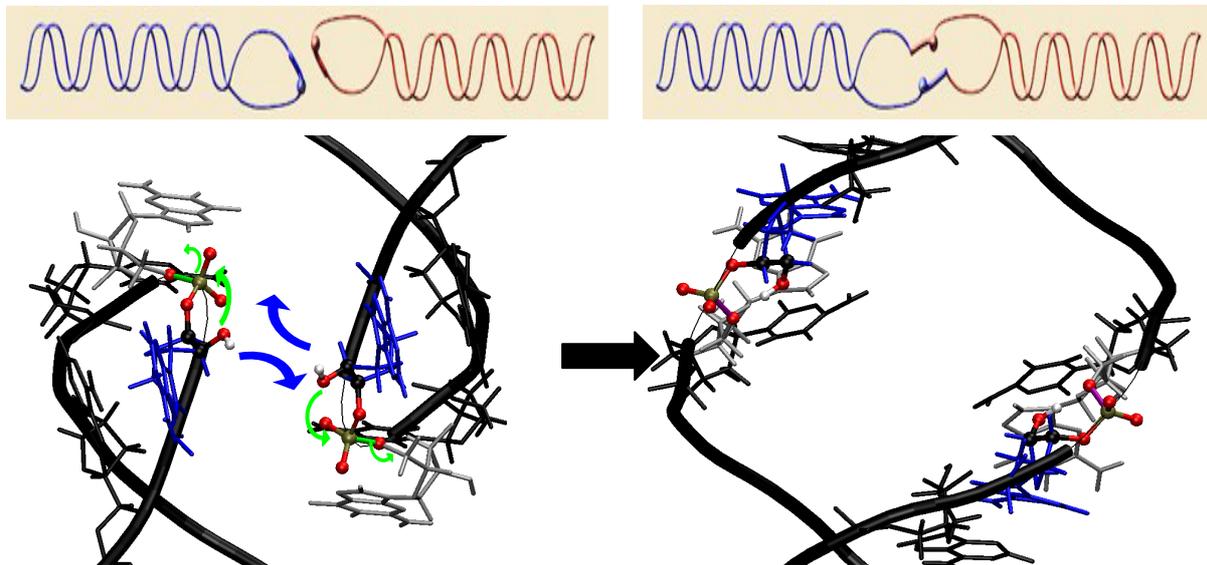


Figure 17 : Modèle de transition entre le dimère instable et le dimère stable suggéré par Ennifar et al. (2001), impliquant un mécanisme de coupure-religation du squelette phospho-diester.

Des cations de magnésium, présents en abondance dans les structures cristallographiques, pourraient intervenir comme catalyseurs dans ces réactions et certains d'entre eux sont d'ailleurs observés à proximité des atomes concernés dans les données structurales.

Un autre mécanisme comportant un processus de coupure-religation mais empruntant une voie quelque peu différente a été envisagé par Pattabiraman et al. (2002) sur la base d'études de modélisation moléculaire. Le mécanisme proposé diffère pour les souches Mal et Lai et repose sur des observations structurales de leurs modèles. En ce qui concerne la souche Mal, le groupement 2'-OH du résidu R273 aurait la capacité d'attaquer son propre groupement phosphate en 3', provoquant ainsi la coupure du brin d'ARN, et comme précédemment il y aurait re-formation d'une liaison phospho-diester mais inter-brins entre le groupement 5'-OH du résidu 274 d'un brin et le phosphore du résidu 273 de l'autre brin. Le résidu 273 du modèle de tige-boucle proposé pour la souche Lai ne possède pas cette capacité car son groupement 2'-OH n'est pas orienté de manière favorable. Le mécanisme de transition envisagé est donc légèrement différent. Il consiste en une attaque du groupement 2'-OH du résidu 272 d'un brin sur le groupement phosphate du résidu 273 de l'autre brin. La rupture de la liaison P-O3' conduit à la formation d'un groupement 3'-OH libre d'attaquer le phosphore en 5' sur l'autre brin afin de reformer une liaison phospho-diester, ce qui entraîne l'échange de brins. Ce modèle de transition n'est cependant proposé qu'à titre d'hypothèse et n'est appuyé par aucune autre donnée expérimentale que les observations des structures obtenues par mécanique moléculaire.

Enfin, un mécanisme totalement différent a été envisagé par Beaurain et al. (2003) sur la base d'observations du comportement des structures de « kissing-complex » de Mujeeb et al. (1998) et de Ennifar et al. (2001) lors de simulations de dynamique moléculaire, ainsi que de leurs divergences structurales en particulier concernant le nombre de paires de bases contenues dans les tiges et dans l'interaction boucle-boucle pour chacune de ces structures. Ces paires de bases sont au nombre de sept dans les tiges et de six dans les boucles chez Ennifar et al. (2001) alors qu'elles sont de six dans les tiges et de huit dans les boucles chez Mujeeb et al. (1998)

- si l'on tient compte des paires A273-A280 inter-moléculaires. Dans le mécanisme imaginé par Beaurain et al. (2003), la formation du dimère étendu à partir du « kissing-complex » pourrait se faire par extension de l'hélice boucle-boucle résidu par résidu. Les deux structures de « kissing-complex » ne représenteraient alors que deux états transitoires différents de la conversion « kissing-complex » -> dimère étendu.

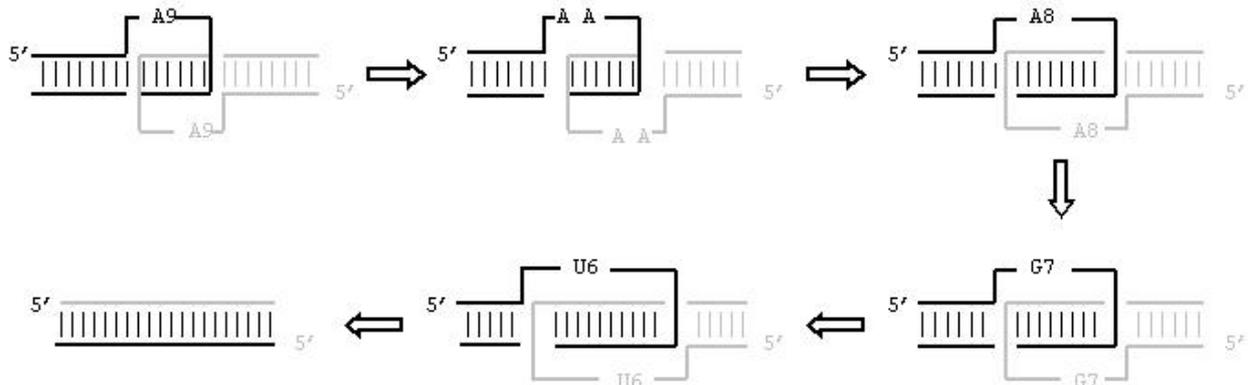


Figure 18 : Modèle de transition entre le dimère instable et le dimère stable suggéré par Beaurain et al. (2003) et consistant en une élongation de l'hélice boucle-boucle du dimère instable.

A l'heure actuelle, les données expérimentales publiées restent insuffisantes pour permettre de trancher entre ces différents mécanismes.

Un autre élément d'importance dans le mécanisme de transition *in vivo* est la présence d'un catalyseur à cette réaction, en l'occurrence la protéine de nucléocapside NCp7, dont le rôle primordial dans le phénomène de dimérisation a fait l'objet de nombreuses études.

→ La protéine de nucléocapside NCp7

Les études sur la dimérisation de l'ARN génomique et sur les cofacteurs associés ont montré l'implication du domaine NC de Pr55 dans ce processus (Berkowitz et al. 1993; Darlix et al. 1990; Dorfman et al. 1994; Fu et al. 1994). La NCp7 est issue du clivage par la protéase du précurseur poly-protéique Pr55gag. Sa séquence active n'est pas encore clairement identifiée mais comprend deux motifs CCHC (CCHC = C-X2-C-X4-H-X4-C ; C=Cys, H=His, X=ac. am. quelconque). Ce motif est présent en simple ou double exemplaire

chez toutes les protéines de nucléocapside rétrovirales (Summers et al. 1992). Ces régions s'organisent autour d'un atome de zinc en un complexe de coordination connu sous le nom de motif en « doigt de zinc ». (Figure 19)

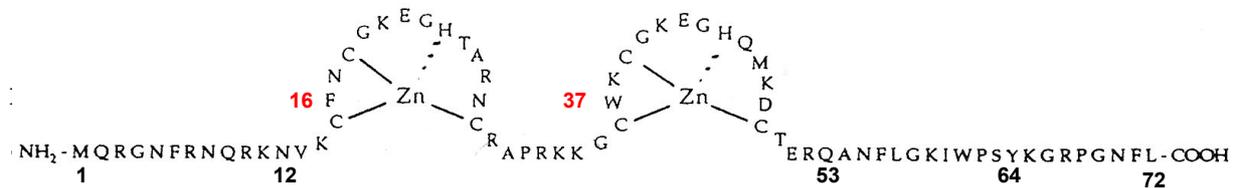


Figure 19 : Structure secondaire en « doigt de zinc » de la protéine de nucléocapside NCp7

Les études de la NCp7 en tant que catalyseur de la dimérisation ont suggéré que son rôle d'activateur du processus était dû à une interaction avec le complexe boucle-boucle d'ARN conduisant à la transformation de ce dimère instable en dimère stable (Muriaux et al. 1996a; Sakaguchi et al. 1993). Elle apparaît donc comme indispensable à la maturation du virus dans les conditions physiologiques (Damgaard et al. 1998; De Rocquigny et al. 1993; Takahashi et al. 2001; Takahashi et al. 2000; Yuan et al. 2003). Rist et Marino. (2002) ont montré de surcroît que la NCp7 accélérât l'échange de brin entre les tiges sans rompre les interactions boucle-boucle et que son action était optimale dans la proportion de deux NCp7 pour un « kissing-complex » d'ARN.

II.c. But du travail de Thèse

La dimérisation apparaît comme une des étapes clé du cycle rétroviral. En particulier, la tige-boucle SL1 pourrait constituer une nouvelle cible pour la conception d'inhibiteurs. L'élucidation des structures des complexes impliqués dans cette étape ainsi que du mécanisme de conversion du « kissing-complex » en dimère étendu peut donc représenter une grande avancée dans la lutte contre ce virus. Plusieurs structures de dimères sont parues dans la littérature mais ces structures présentent de grandes divergences entre elles, en particulier pour le « kissing-complex ». Or, la conception de médicaments assistée par ordinateur se base sur les connaissances structurales de la molécule. Il est donc nécessaire d'étudier soigneusement la stabilité des différentes structures publiées afin

de déterminer lesquelles de ces conformations les dimères de SL1 sont les plus susceptibles d'adopter en solution.

Un autre élément d'importance à étudier est le chemin conformationnel emprunté par la molécule pour réaliser la transition entre le « kissing-complex » et le dimère étendu afin de déterminer la région de la tige-boucle SL1 sur laquelle il serait plus judicieux de concentrer nos efforts. Les différents mécanismes envisagés jusque là ne nous permettent en effet pas de privilégier une région d'attaque, l'un de ces mécanismes envisageant une fonte des tiges avec conservation de l'interaction boucle-boucle lors qu'un autre prévoit une coupure du squelette phospho-diester entre la jonction et la boucle sans qu'aucune modification n'intervienne sur les tiges.

L'outil d'investigation que nous avons choisi d'utiliser est la simulation de dynamique moléculaire. Cette technique permet en effet d'explorer l'espace conformationnel d'une molécule et est souvent utilisée comme outil de raffinement des structures déterminées expérimentalement. De plus les nouvelles techniques développées permettent d'étudier le chemin conformationnel d'une transition.

III. MATERIEL ET METHODES

III. MATERIEL ET METHODES

III.a. Rappels sur l'ARN

III.a.i. Structure de l'ARN

L'acide ribonucléique (ARN) est un polymère similaire à l'ADN, aussi bien en terme structural qu'en terme fonctionnel, à ceci près que l'ose de l'ARN est un ribose alors que celui de l'ADN est un désoxyribose. Ses éléments constitutifs sont les quatre bases adénine (A), cytosine (C), uracile (U) et guanine (G) qui sont reliées au sucre ribose. La chaîne d'ARN se construit par des liaisons phospho-diester entre les sucres successifs par l'intermédiaire des groupements phosphates (Figure 20).

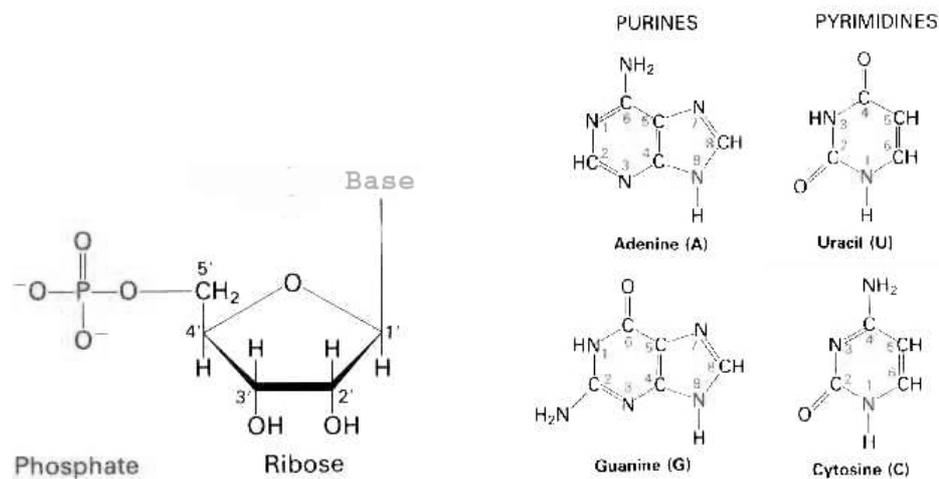


Figure 20 : Organisation d'un nucléotide d'ARN : le groupement phosphate, le sucre (un ribose dans le cas de l'ARN) et la base. Les 4 bases impliquées dans les ARN sont détaillées à droite.

Le caractère monocaténaire de l'ARN lui confère la possibilité d'adopter des structures secondaires et tertiaires conditionnées par l'agencement des nucléotides de sa propre séquence.

Les interactions participant à la structure secondaire des ARN sont les interactions électrostatiques, les interactions d'empilement, les interactions soluté-solvant et les liaisons hydrogène. Il existe plusieurs types de liaisons hydrogène entre les bases d'un ARN (Saenger 1984). Les liaisons les plus courantes sont de type Watson-Crick (liaisons canoniques) qui permettent de former trois liaisons hydrogène entre les résidus G et C ou deux liaisons hydrogène entre les résidus A et U. On trouve aussi assez

communément des paires G-U de type Wobble. Plus rares mais rencontrées de temps à autres dans les structures d'ARN sont les paires de base de type Hoogsteen. (Figure 21)

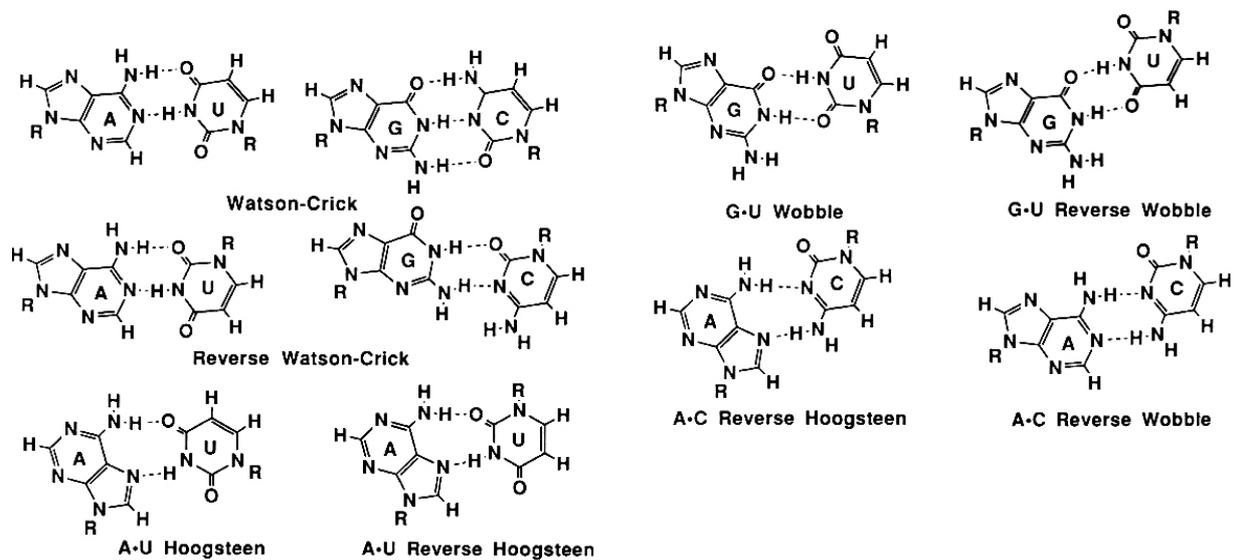


Figure 21 : Les différents types de liaisons hydrogène pouvant intervenir entre les bases d'un ARN d'après Saenger (1984).

A l'instar de l'ADN ces appariements peuvent donner naissance à des doubles hélices, notamment lorsque l'ARN constitue le support de l'information génétique, comme chez certains virus. Néanmoins, la présence en 2' du ribose du groupement hydroxyle défavorise la conformation B de l'hélice au profit d'une conformation de type A. De plus, comme mentionné ci-dessus, les brins qui constituent ces doubles hélices peuvent appartenir à la même molécule d'ARN. L'enchaînement de régions appariées et non-appariées conduit à des arrangements présentant un grand nombre de motifs secondaires, dont le plus fréquemment rencontré reste le motif dit en tige-boucle (communément appelé « épingle à cheveux »). Il est constitué d'une région en double hélice terminée par des nucléotides non-appariés formant une boucle dont les rôles sont variés, par exemple stabilisation ou interaction avec de l'ARN ou des protéines. Bien d'autres motifs de structure secondaire existent chez l'ARN, détaillés Figure 22 (Burckard et al. 1999). Les interactions à courtes ou longues distances de ces motifs entre eux forment la structure tridimensionnelle dite structure tertiaire (Figure 22).

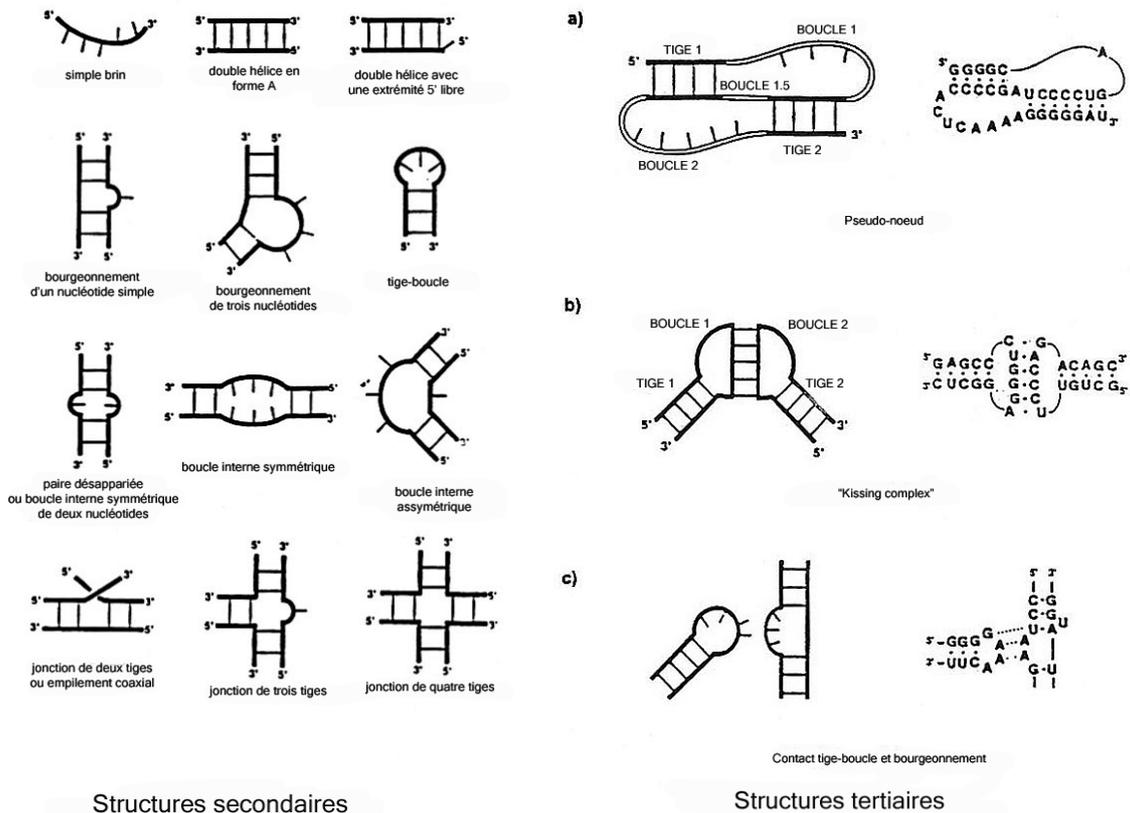


Figure 22 : Descriptif des structures secondaires et tertiaires rencontrées dans les ARN d'après Burckard et al. (1999).

III.a.ii. Paramètres structuraux des acides nucléiques

Il a été déterminé lors de la convention de Cambridge de 1988 (EMBOworkshop 1989; Lavery and Sklenar 1988; Sarma 1988) un ensemble de paramètres afin de décrire de façon univoque la conformation d'un acide nucléique. Les positions relatives des paires de bases sont décrites par des paramètres hélicoïdaux. La position de la base ou de la paire de base par rapport à l'axe de la double hélice est décrite par des paramètres de translation et de rotations nommés *Xdisp*, *Ydisp*, *inclination* et *tip*. (Figure 23)

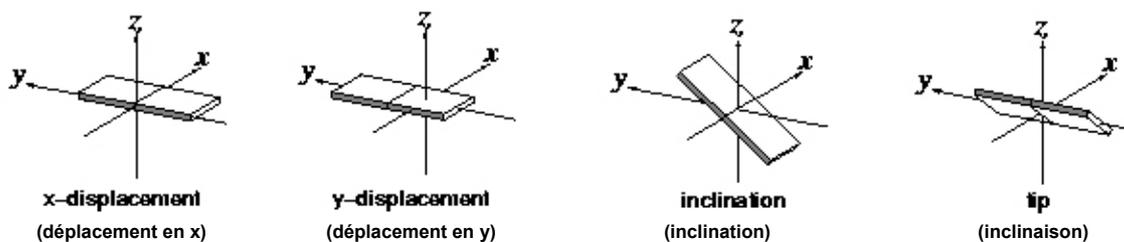


Figure 23 : Définition des paramètres intra-base ou paire de bases

Les positions relatives des deux bases d'une même paire sont évaluées grâce aux paramètres *shear*, *stretch*, *stagger*, *propeller*, *twist*, *buckle* et *opening*. (Figure 24)

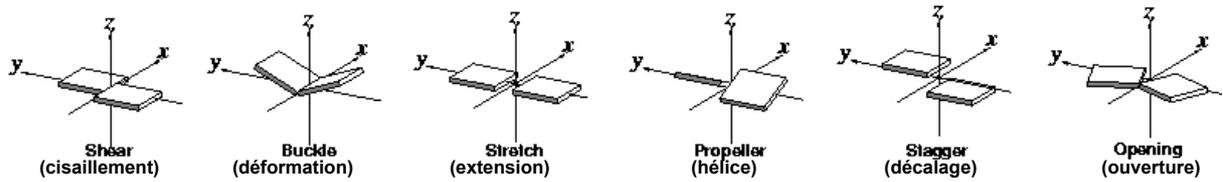


Figure 24 : Définition des paramètres inter-bases d'une même paire de bases

Enfin, l'organisation des paires de bases successives se détermine par les paramètres de *twist*, *roll*, *tilt*, *rise*, *slide* et *shift*. (Figure 25)

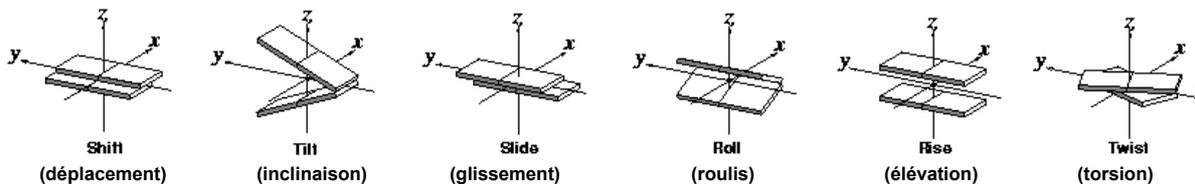


Figure 25 : Définition des paramètres inter-paires de bases.

Des paramètres d'angles dièdres permettent de mesurer la conformation d'un sucre (angles ν_1 à ν_5) ainsi que du squelette phospho-diester (angles α , β , γ , δ , ϵ , ζ). L'orientation du sucre par rapport à la base adjointe est décrite par l'angle dièdre χ . (Figure 26).

Du fait que le sucre est composé d'un cycle à 5 atomes, les 5 angles caractérisés précédemment (ν_1 à ν_5) sont liés les uns aux autres. Le plissement du sucre est en fait entièrement déterminé par la connaissance de deux paramètres du sucre, la phase P et l'amplitude A , reliés aux angles de torsion endocycliques par l'équation de pseudo-rotation :

$$\nu_i = A \cos[P + 144^\circ(i - 2)] \quad \text{où } i=1 \rightarrow 5 \quad (24)$$

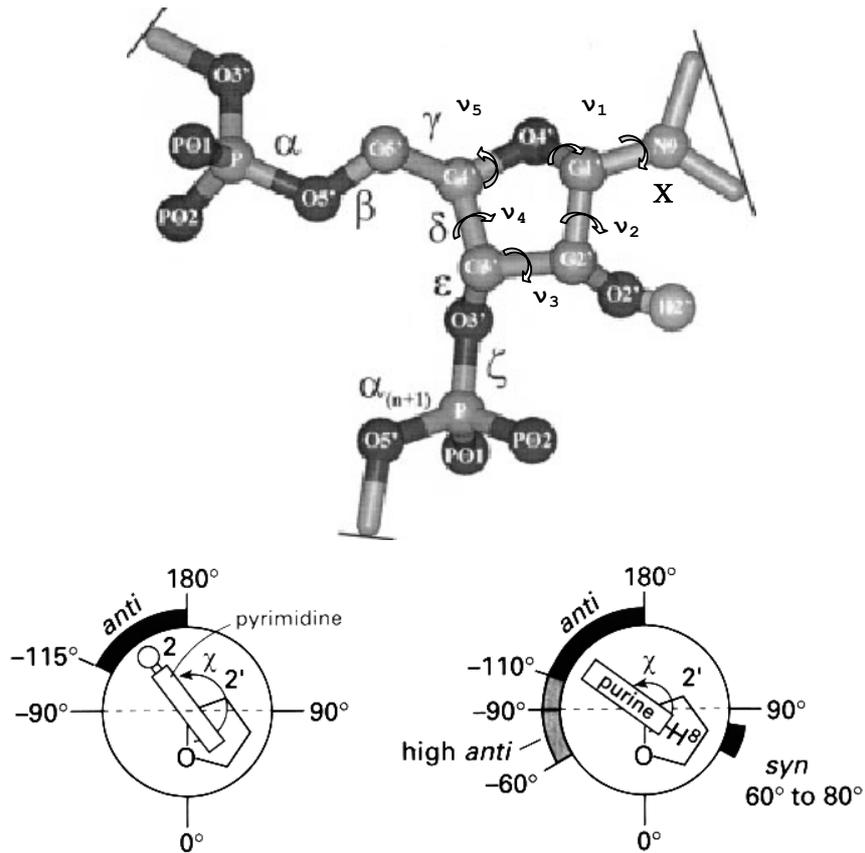


Figure 26 : Définition des angles dièdres caractéristiques de la conformation du sucre (angles ν_1 à ν_5 , X) ainsi que ceux caractéristiques de la conformation du squelette phospho-diester (angles α , β , γ , δ , ϵ , ζ)

III.b.La simulation de dynamique moléculaire de polymères

III.b.i. Principe de la dynamique moléculaire libre

La dynamique moléculaire consiste à étudier la trajectoire d'une molécule en lui appliquant les lois de la mécanique classique newtonienne. Une molécule est considérée comme un assemblage d'atomes dont les positions spatiales varient au cours du temps. Chaque atome est considéré comme une masse ponctuelle dont le mouvement est déterminé par l'ensemble des forces qui s'exercent sur elle. Il en résulte un mouvement atomique correspondant à des oscillations autour d'un minimum d'énergie ou au passage d'un de ces minima à un autre. La dynamique moléculaire possède ainsi la faculté d'extraire la molécule d'un minimum local.

Cette technique repose sur l'intégration des équations de mouvements de Newton pour chaque atome i :

$$\vec{F}_i(t) = m_i \cdot \vec{a}_i(t) \quad (1)$$

où m_i = masse de l'atome i

\vec{a}_i = accélération de l'atome i

\vec{F}_i = force s'exerçant sur l'atome i

L'accélération étant dérivée des coordonnées $\vec{r}_i(t)$ d'un atome i à un instant t , on en déduit que :

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

Si $\vec{r}_1, \dots, \vec{r}_n$ sont les coordonnées des différents atomes à un instant t , on peut déterminer la force \vec{F}_i qui s'exerce à cet instant sur un atome i par dérivation par rapport à ses coordonnées spatiales de la fonction énergie potentielle $E(\vec{r}_1, \dots, \vec{r}_n)$:

$$\vec{F}_i(t) = -\frac{d}{d\vec{r}_i} E(r_1, \dots, r_n) \quad (3)$$

Pour obtenir la trajectoire des atomes en fonction du temps, on va réaliser une intégration numérique sur un court intervalle de temps des équations de mouvements.

Ces équations de mouvements sont déterministes. Il faut, pour réaliser une simulation de dynamique moléculaire, partir d'une structure raisonnable du système (expérimentale ou modèle) que l'on veut étudier et on déterminera les coordonnées et vitesses initiales en procédant ainsi :

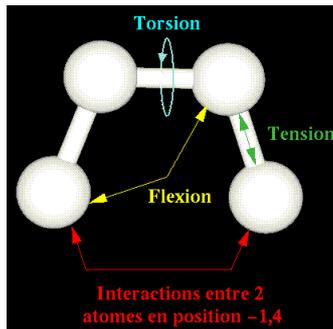
- Une minimisation d'énergie du système fournira le jeu de coordonnées de départ de la dynamique. Cette étape de minimisation nécessaire permet d'éviter qu'une énergie mécanique trop forte ne détériore l'assemblage moléculaire dès le début de la production de dynamique moléculaire.
- Les vitesses de départ des atomes sont attribuées aléatoirement à partir d'une distribution gaussienne de Maxwell-Boltzman correspondant à la température T .

→ Champs de Force

Un champ de force est une équation modélisant l'énergie potentielle d'un ensemble d'atomes comportant un certain nombre de paramètres. Il décrit les interactions entre atomes liés et non-liés. Apparus

dans les années 1970, ces champs de force sont toujours en cours d'évolution afin d'obtenir les simulations numériques les plus réalistes possibles. Les interactions entre atomes se divisent en deux catégories :

- les interactions entre atomes liés qui correspondent aux énergies de liaison, d'angle de valence et de torsion des angles dièdres.
- les interactions entre atomes non-liés soit les interactions de type Van der Waals, électrostatique et liaisons hydrogène.

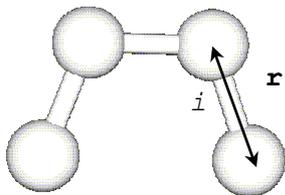


Le champ de force résultant est la somme de ces différents termes :

$$E_F = E_l + E_\theta + E_\tau + E_{vdw} + E_{\epsilon_1} + E_{HB}$$

Les termes de cette équation sont calculés comme suit :

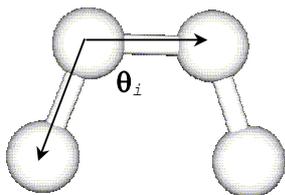
- **Energie d'élongation des liaisons (« Stretching ») :**



$$E_l = \frac{1}{2} \sum_{i=1}^n k_{r,i} (r_i - r_i^0)^2 \quad (4)$$

où r_i = longueur instantanée de la liaison i
 r_i^0 = longueur de la liaison i à l'équilibre
 $k_{r,i}$ = constante de force de la liaison i
 n = nombre de liaisons dans la molécule

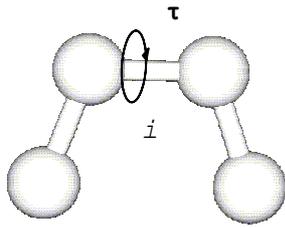
- **Energie de déformation des angles (« Bending »)**



$$E_\theta = \frac{1}{2} \sum_i^n k_{\theta,i} (\theta_i - \theta_i^0)^2 \quad (5)$$

où θ_i = angle de liaison entre 3 atomes
 θ_i^0 = angle de liaison entre 3 atomes à l'équilibre
 $k_{\theta,i}$ = constante de force de flexion
 n = nombre d'angles dans la molécule

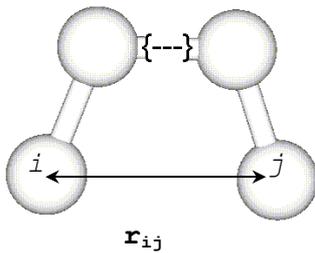
- **Energie de torsion des angles dièdres (« Torsion »)**



$$E_{\tau} = \frac{1}{2} \sum_{i=1}^n A_{i,m} [1 + \cos(m\tau_i - \phi)] \quad (6)$$

où τ_i =angle de torsion
 Φ =phase
 $A_{i,n}$ =constante de force de courbure
 m =périodicité de $A_{i,m}$
 n =nombre d'angles dièdres

- **Interactions de Van der Waals**

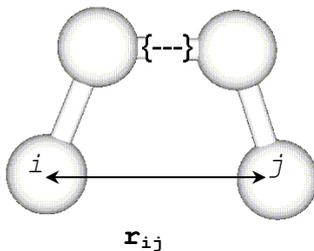


$$E_{\text{vdw}} = \sum_{i,j=1}^n \left[A_{ij} \left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - B_{ij} \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 \right] \quad (7)$$

où A_{ij}, B_{ij} =constantes de Van der Waals
 r_{ij} =distance entre 2 atomes i et j non-liés
 r_{ij}^0 =somme des rayons de Van der Waals de i et j
 n =nombre de paires d'atomes non-liés

On décrit ici la fonction de Lennard-Jones ou potentiel 6-12. Il existe d'autres formulations de l'interaction de Van der Waals.

- **Interactions électrostatiques**



$$E_{e1} = \sum_{ij} \frac{q_i q_j}{4\pi\epsilon_{ij} r_{ij}} \quad (8)$$

où q =charge partielle de l'atome
 $\frac{1}{4\pi\epsilon_{ij}}$ =permittivité électrique du milieu
 r_{ij} =distance entre les 2 atomes i et j
 n =nombre de paires d'atomes non-liés

→ Minimisation d'énergie

La minimisation permet d'optimiser la géométrie de l'assemblage atomique jusqu'à l'obtention d'un minimum local.

Pour une molécule de N atomes, la fonction à minimiser comprend $3N$ variables. A partir du jeu de coordonnées initiales, on recherche le jeu de coordonnées cartésiennes qui réduit à son minimum la somme de toutes les interactions énergétiques. Les méthodes couramment utilisées reposent sur les dérivées première et seconde de l'énergie potentielle :

- « Steepest descent » ou méthode de la plus grande pente

L'énergie E_0 initiale est tout d'abord calculée. Les atomes sont ensuite déplacés individuellement de $d\vec{r}$ et on calcule la nouvelle énergie E_1 . On obtient le gradient d'énergie suivant :

$$\vec{\nabla}E = \frac{E_1 - E_0}{dr} \cdot \frac{d\vec{r}}{|d\vec{r}|}$$

- si $\vec{\nabla}E < 0$, l'énergie diminue avec $d\vec{r}$: on garde la direction
- si $\vec{\nabla}E > 0$, l'énergie augmente avec $d\vec{r}$: la direction n'est pas conservée

On cherche alors dans la direction $\vec{\nabla}E$ un point qui minimise E . Ce point sert ensuite de point de départ à une réitération de ces étapes.

Dans cette méthode, on peut définir un pas de déplacement α , c'est à dire un changement de variables (coordonnées, angle de rotation autour d'une liaison, etc...). Comme on cherche à minimiser l'énergie, on fait évoluer α pour atteindre plus rapidement ce minimum : si l'énergie diminue, on va augmenter α car on a plus de chances, dans cette direction, d'aller vers un minimum. Au contraire, on diminuera α pour explorer des changements plus petits dans le cas d'une énergie croissante.

Cet algorithme suit donc la direction imposée par les forces interatomiques dominantes. On recherche la plus grande pente au cours de laquelle l'énergie décroît le plus rapidement, d'où son nom. C'est une méthode rapide dans les premiers cycles et donc très efficace pour la suppression des mauvais contacts ou des problèmes stéréochimiques qui peuvent exister dans la structure de départ mais dont la convergence finale est très lente.

- Méthode du gradient conjugué

On utilise le même procédé que ci-dessus si ce n'est que les étapes précédentes sont prises en compte afin d'ajuster de façon plus fine le pas de déplacement et la direction imposée. Le pas est ajusté à chaque cycle pour obtenir la meilleure diminution d'énergie. L'intérêt de cet algorithme est d'éviter un comportement oscillatoire autour du minimum et d'accélérer la convergence.

- Méthode Newton-Raphson

Cette technique fait intervenir la dérivée seconde de l'énergie potentielle. On cherche le jeu de coordonnées \vec{r} qui minimise l'énergie, donc qui annule sa dérivée : $\vec{\nabla}E = \frac{dE(\vec{r})}{d\vec{r}} = 0$. Il convient donc tout d'abord de déterminer cette dérivée ∇E . Pour cela, on va utiliser un développement de Taylor du premier ordre, attendu que pour réaliser une minimisation on dispose de la structure initiale donc d'un jeu de coordonnées \vec{r}_0 :

$$\vec{\nabla}E = \frac{dE(\vec{r})}{d\vec{r}} = \left. \frac{dE(\vec{r})}{d\vec{r}} \right|_{\vec{r}=\vec{r}_0} + \left. \frac{d^2E(\vec{r})}{d\vec{r}^2} \right|_{\vec{r}=\vec{r}_0} \quad (9)$$

Dans cette équation, $\left. \frac{dE(\vec{r})}{d\vec{r}} \right|_{\vec{r}=\vec{r}_0}$ est un vecteur et $\left. \frac{d^2E(\vec{r})}{d\vec{r}^2} \right|_{\vec{r}=\vec{r}_0}$ est un tenseur d'ordre 2. La recherche de la valeur \vec{r}_{\min} , et plus généralement de la valeur \vec{r}_{p+1} en fonction de \vec{r}_p dans le mécanisme itératif, conduit à la résolution d'un système linéaire à chaque itération. Cette méthode est performante et peut converger vers le minimum global si le choix du \vec{r}_0 est judicieux mais il est nécessaire pour l'appliquer de pouvoir calculer les dérivées première et seconde de E en chaque point r_i .

→ Algorithmes d'intégration

L'algorithme de Verlet (Verlet 1967) permet de calculer les vitesses et les positions des atomes à un temps $t+\delta t$ lorsque l'on connaît ces dernières au temps t et à condition que l'on choisisse un pas δt suffisamment petit. On procède ainsi par subdivision de la trajectoire en une série d'états discrets séparés par un intervalle de temps δt très court qui définit alors le pas d'intégration des équations précédentes.

Sur cet intervalle de temps δt , d'après la relation (2) en lui appliquant un développement limité de Taylor, les positions des atomes sont calculées d'après les équations suivantes :

$$\vec{r}_i(t + \delta t) = \vec{r}_i(t) + \vec{v}_i(t)\delta t + \frac{\vec{F}_i(t)}{2m_i}(\delta t)^2 + \frac{d^3\vec{r}_i(t)}{dt^3}(\delta t)^3 + O((\delta t)^4) \quad (10)$$

$$\vec{r}_i(t - \delta t) = \vec{r}_i(t) + \vec{v}_i(t)\delta t + \frac{\vec{F}_i(t)}{2m_i}(\delta t)^2 - \frac{d^3\vec{r}_i(t)}{dt^3}(\delta t)^3 + O((\delta t)^4) \quad (11)$$

d'où les nouvelles valeurs des positions déterminées par :

$$\vec{r}_i(t + \delta t) + \vec{r}_i(t - \delta t) = 2\vec{r}_i(t) + \frac{\vec{F}_i(t)}{m_i}(\delta t)^2 + O((\delta t)^4) \quad (12)$$

On remarque que la vitesse n'intervient pas dans cet algorithme. Elle peut être obtenue comme suit :

$$\vec{v}_i(t) = \frac{\vec{r}_i(t + \delta t) - \vec{r}_i(t - \delta t)}{2\delta t} \quad (13)$$

Un algorithme apparenté à Verlet est connu sous le nom d'algorithme « leap-frog » ou algorithme « saute-mouton ». Il se différencie de Verlet par le fait qu'on utilise cette fois et les positions et les vitesses pour réaliser le calcul, les vitesses étant calculées pour des intervalles de temps demi-entiers et les positions pour des intervalles de temps entiers. Si on a défini les vitesses pour les temps $t + \frac{\delta t}{2}$ et $t - \frac{\delta t}{2}$:

$$\vec{v}_i\left(t + \frac{\delta t}{2}\right) = \frac{\vec{r}_i(t + \delta t) - \vec{r}_i(t)}{\delta t} \quad \text{et} \quad \vec{v}_i\left(t - \frac{\delta t}{2}\right) = \frac{\vec{r}_i(t) - \vec{r}_i(t - \delta t)}{\delta t} \quad (14)$$

on obtient alors :

$$\vec{r}_i(t + \delta t) = \vec{r}_i(t) + \vec{v}_i\left(t + \frac{\delta t}{2}\right) \cdot \delta t \quad \text{et} \quad \vec{r}_i(t - \delta t) = \vec{r}_i(t) - \vec{v}_i\left(t - \frac{\delta t}{2}\right) \cdot \delta t \quad (15)$$

d'où :

$$\vec{v}_i\left(t + \frac{\delta t}{2}\right) = \vec{v}_i\left(t - \frac{\delta t}{2}\right) + \frac{\vec{F}_i(t)}{m_i} \cdot \delta t \quad (16)$$

Par itérations de la procédure, on parvient à déterminer les positions et vitesses des atomes aux temps t de la trajectoire.

Le pas d'intégration n'est pas choisi de façon anodine. Il doit être choisi petit par rapport à la période du mouvement vibrationnel de plus haute fréquence (mouvement le plus rapide), en l'occurrence celui de la liaison C-H qui est de 10 femtosecondes. On choisit un pas environ 10 fois inférieur à cette période, les simulations de dynamique moléculaire sont donc généralement réalisées avec un pas d'intégration de 1 fs. Ce pas peut être augmenté à 2 fs lorsque l'on utilise l'algorithme de SHAKE (Ryckaert et al. 1977; Van Gunsteren

and Berendsen 1977) durant les simulations qui a pour fonction de figer les liaisons X-H, où X est un atome lourd quelconque.

→ Ensemble thermodynamique

Historiquement, l'ensemble microcanonique NVE (nombre de particules, volume et énergie totale constants) est l'ensemble naturel de la dynamique moléculaire d'un système conservatif, car l'énergie est conservée au cours du temps. Cependant, cet ensemble se trouve relativement éloigné des conditions expérimentales d'un système réel. Les simulations ont donc été par la suite réalisées dans différents ensembles statistiques (canonique, isobare-isotherme, etc...) selon les conditions dans lesquelles on veut la réaliser et afin de se rapprocher au mieux des observations expérimentales.

→ Conditions périodiques

Les simulations de dynamique moléculaire en solution sont effectuées dans des boîtes de solvants de différentes formes (parallélépipédique, octaédrique, etc...) détentrices de conditions aux limites telles que :

- lorsque l'oxygène d'une molécule d'eau sort d'une des faces de la boîte, une copie de cette molécule d'eau est repositionnée du côté opposé de la boîte avec la même vitesse
- de même, lorsqu'un atome de la molécule ou un ion sort d'un côté de la boîte, une copie de cet atome est placée dans la boîte, et ce afin que le nombre d'atomes soit constant dans la boîte et le système électriquement neutre

Afin d'éviter les effets de bord dans le calcul des interactions entre atomes non-liés et les calculs soluté-solvant, la boîte de simulation est dupliquée dans toutes les directions de sorte que, si on a par exemple un système atomique dans une boîte d'eau parallélépipédique, on aura en fait le système décrit dans la Figure 27.

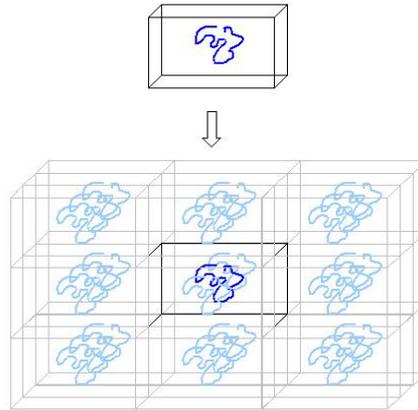


Figure 27 : Conditions périodiques utilisées en simulation. La molécule initiale (en bleu foncé) est immergée dans une boîte d'eau parallélépipédique (en noir). Pour effectuer le calcul des interactions entre atomes non-liés, cette boîte est dupliquée à l'infini dans toutes les directions (les duplicata sont représentés avec une molécule bleue claire dans une boîte d'eau grise)

Les boîtes sont en pratique copiées à l'infini dans toutes les directions mais par souci de clarté, seule la première couche est présentée ici.

Au cours de nos expériences, nous avons utilisé deux sortes de boîtes de simulation : des boîtes de type tétraédrique et des boîtes de forme octaédrique tronquée telles que l'on peut voir sur la Figure 28.

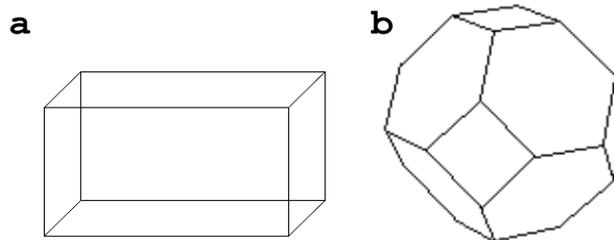


Figure 28 : Exemple des deux types de boîte d'eau que nous avons utilisés **a)** boîte d'eau de type tétraèdre et **b)** boîte d'eau de type octaèdre tronqué.

On utilise classiquement la boîte tétraédrique pour faire des simulations de dynamique moléculaire. Cette boîte possède cependant un gros inconvénient. Lors d'une simulation de dynamique moléculaire, la molécule peut se déformer ou même simplement tourner dans la boîte de telle manière que ses deux extrémités se retrouvent trop près des bords voire sortent de la boîte. La molécule risque ainsi de rentrer en interaction avec les copies d'elle-même

présentes dans les boîtes adjacentes ce qui risque de créer des interactions artefactuelles.

Pour une dynamique sans changement de structure, on peut pallier ce problème en contraignant la position d'une liaison judicieusement choisie (proche du centre de la boîte et parallèle au plus grand axe de la molécule), mais dans le cas d'un changement conformationnel important, cette solution est inefficace. La boîte octaédrique tronquée semble bien mieux adaptée à ce genre de cas. En effet, cette boîte combine l'avantage d'une géométrie pseudo-sphérique avec des conditions aux limites périodiques. Par contre, elle demande beaucoup plus de ressources, car pour un même soluté, la boîte octaédrique contient bien plus de molécules de solvant que la boîte tétraédrique. C'est pourquoi nous n'avons pas pu utiliser les boîtes octaédriques avant l'extension de nos moyens de calculs.

La solution provisoire que nous avons adoptée a donc été, lorsque la géométrie de la boîte ne convenait plus à la structure de la molécule, de stopper la simulation de dynamique moléculaire afin de réintroduire la molécule dans une boîte de géométrie plus conforme à sa structure. La Figure 29 illustre ce cas de figure.

Afin de parfaire encore le calcul des interactions entre particules non-liées dans des conditions périodiques aux bornes, l'algorithme Particle Mesh Ewald ou PME (Darden et al. 1997; Darden et al. 1993; Essmann 1995; Toukmaji 1996) basée sur la somme d'Ewald (Ewald 1921) est utilisé au cours de nos simulations.

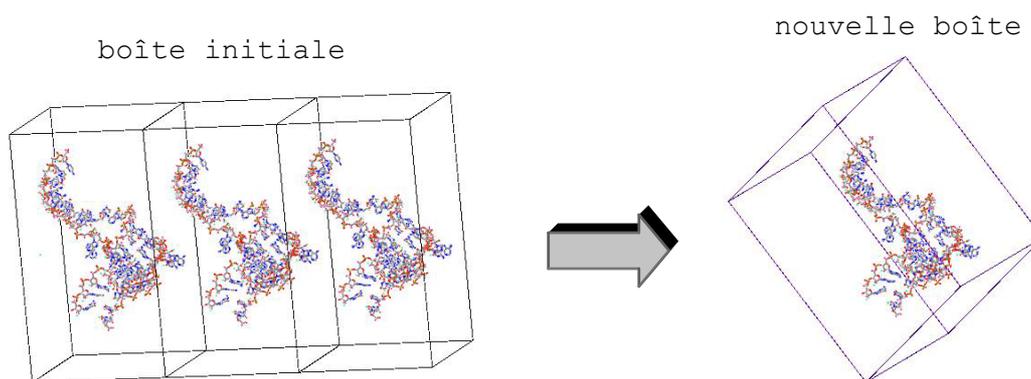


Figure 29 : Illustration d'un changement structural de la molécule incompatible avec la géométrie de sa boîte. On peut voir à gauche de la figure que la molécule a pris une forme telle qu'elle risque de se trouver à proximité des deux bords opposés de la boîte, et ainsi d'entrer en interaction avec elle-même. Grâce au changement de boîte, la molécule se trouve à nouveau à une distance minimale de 10Å par rapport à n'importe lequel des bords de la boîte.

→ Particle Mesh Ewald (PME)

La technique de la somme d'Ewald (Ewald 1921) a pour finalité de calculer de façon sûre les interactions électrostatiques entre atomes dans un environnement soumis à des conditions périodiques aux bornes. On considère pour cela un ensemble de N particules possédant des charges de sorte que la charge totale du système soit nulle :

$$\sum_i q_i = 0 \quad (17)$$

Ces N particules sont situées dans une boîte cubique de longueur L avec des conditions aux limites périodiques. A une particule i de coordonnées \vec{r}_i dans la boîte initiale correspond une infinité d'images situées dans les boîtes dupliquées, de coordonnées $\vec{r}_i + \vec{n}L$ où \vec{n} est le vecteur de composantes (n_x, n_y, n_z) entières. L'énergie totale du système s'exprime alors par :

$$E_{\text{tot}} = \frac{1}{2} \sum_n^* \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|\vec{r}_j - \vec{r}_i + \vec{n}L|} \quad (18)$$

Le symbole * indique que la somme est faite sur toutes les boîtes et toutes les particules, hormis sur $j=i$ quand $n=0$.

Pour des potentiels à courte portée (décroissance rapide en $1/r^n$, $n \geq 3$), l'interaction entre deux particules i et j est, avec une très bonne approximation, calculée comme l'interaction entre la particule i et l'image de j la plus proche de i (convention d'image minimale). C'est à dire que pour une distance r_{ij} supérieure à une valeur r_{cut} prédéfinie, l'interaction ne sera pas calculée car l'énergie est considérée comme suffisamment faible. Cette image ne se trouve pas nécessairement dans la boîte initiale. Cette approximation reste cependant insuffisante pour les interactions de portée longue du fait que la décroissance d'énergie entre les deux particules s'effectue trop lentement pour pouvoir se limiter à cette plus proche image. Afin de remédier à ce problème, Ewald a séparé l'équation (18) en trois parties :

- la somme directe, ou somme sur l'espace réel
- la somme réciproque, ou somme sur l'espace imaginaire
- la somme de correction qui est une constante

La somme directe est obtenue en écrantant chaque particule avec une distribution de charges (le plus souvent gaussienne) de même intensité mais de signes opposés à celle de la particule. Elle correspond au calcul des interactions à courte portée. La somme réciproque, qui prend en charge les interactions à longue portée, fait intervenir une distribution de charges symétrique à la précédente dont la contribution sera calculée dans l'espace réciproque du réseau cubique. La convergence de la somme dans cet espace réciproque dépend de la forme (plus ou moins diffuse) de la distribution de charges. La somme de correction, elle, correspond à l'interaction de chaque distribution de charges q_j avec la charge ponctuelle située au centre de la gaussienne. Les trois énergies issues de cette séparation se présentent sous les formes :

$$E_{\text{directe}} = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \sum_n^* q_i q_j \frac{\text{erfc}(\alpha |r_j - r_i + nL|)}{|r_j - r_i + nL|} \quad (19)$$

$$E_{\text{réciproque}} = \frac{1}{2\pi L^3} \sum_{i=1}^N \sum_{j=1}^N \sum_{k \neq 0} q_i q_j \left(\frac{4\pi^2}{k^2} \right) \exp\left(\frac{-k^2}{4\alpha} \right) \cos(k(r_j - r_i)) \quad (20)$$

$$E_{\text{correction}} = -\frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^N q_i^2 \quad (21)$$

où $n = (n_x, n_y, n_z)$

L'énergie d'interaction devient donc :

$$E_{\text{tot}} = E_{\text{directe}} + E_{\text{réciproque}} + E_{\text{correction}}$$

α est un paramètre qui dépend du système étudié, de la rapidité et de la précision voulues pour le calcul de la somme d'Ewald.

La fonction erfc ou complementary error function est définie par :

$$\text{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (22)$$

L'algorithme PME est une variante efficace de la somme d'Ewald. Les variantes se retrouvent dans le calcul des énergies directe et réciproque :

- on utilise des coupures de l'espace réel pour calculer la somme directe. Supposons une coupure sphérique de rayon R_{cutoff} , chaque particule i n'aura des interactions qu'avec les particules situées à l'intérieur de la sphère de centre i et de rayon R_{cutoff} .

- la somme réciproque est approximée par des convolutions utilisant des transformations rapides de Fourier sur une grille où les charges sont interpolées aux points de la grille.

→ Limitation de la dynamique moléculaire

La dynamique représente une évolution temporelle des mouvements moléculaires et permet d'étudier les propriétés dépendantes du temps telles la diffusion ou le repliement moléculaire. Elle ne permet pas en revanche d'étudier des réactions chimiques telles que les transitions électroniques, les phénomènes de transport de protons ou d'électrons ou les formations ou ruptures de liaisons covalentes.

→ Le RMSD

Un paramètre permettant de mesurer la distance existant entre deux structures tridimensionnelles S_1 et S_2 d'une molécule de N atomes est le *Root-Mean Square Deviation* (déviatoin quadratique moyenne) ou RMSD. Il est calculé à partir de l'équation suivante :

$$\text{RMSD} = \left[\frac{1}{N} \sum_{i=1}^N \|\vec{r}_{1i} - \vec{r}_{2i}\|^2 \right]^{\frac{1}{2}} \quad (25)$$

où $\vec{r}_{1i} = (x, y, z)$ de l'atome i dans la structure S_1

$\vec{r}_{2i} = (x, y, z)$ de l'atome i dans la structure S_2

Pour nos analyses le RMSD entre deux structures est calculé à l'aide d'un programme développé au laboratoire, FITRMSD.

III.b.ii. La dynamique moléculaire sous contraintes

Il est possible dans le cadre de certaines études de biaiser la simulation de dynamique moléculaire par l'ajout dans le champ de force d'une énergie de contrainte. Ces potentiels peuvent être de différentes formes selon le type de contraintes que l'on veut appliquer à la molécule pendant la simulation

→ Contraintes de distance entre deux atomes, d'angle ou d'angle de torsion

Cette opération consiste à rajouter comme énergie de contrainte au champ de force une énergie constituée par deux potentiels semi-harmoniques. Entre les valeurs r_2 et r_3 aucune contrainte n'est appliquée. En deçà et au-delà des valeurs r_1 et r_4 , des potentiels de

contrainte linéaires sont appliqués. Le potentiel de contrainte harmonique n'est appliqué qu'entre les valeurs $[r_1, r_2]$ et $[r_3, r_4]$ (Figure 30).

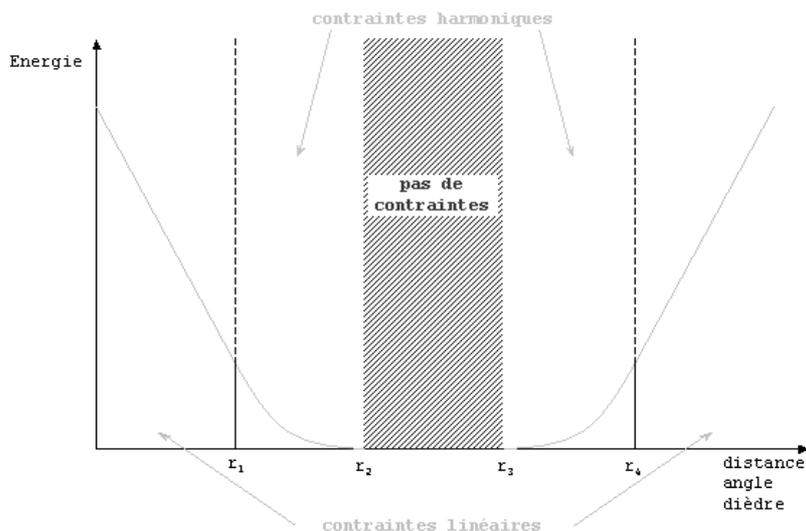


Figure 30 : Représentation en fonction de la coordonnée de réaction choisie de la forme de l'énergie de contrainte imposée au système.

→ Dynamique moléculaire dirigée

La *Targeted Molecular Dynamic* (dynamique moléculaire dirigée) ou TMD (Schlitter et al. 1994) a été récemment implantée dans AMBER (AMBER 7.0). Cette méthode impose un nouveau type de contrainte dont le but est de forcer la structure simulée à se superposer progressivement par la méthode du RMSD à une structure de référence. Elle consiste à ajouter à la fonction d'énergie un terme supplémentaire de la forme suivante :

$$E_{\text{TMD}} = \frac{1}{2} \cdot k_{\text{TMD}} \cdot N \cdot (\rho - \rho_0)^2 \quad (23)$$

où k_{TMD} = constante de force du potentiel harmonique
 N = nombre d'atomes sur lesquels est calculé le RMSD
 ρ = RMSD à l'instant t par rapport à la structure de référence
 ρ_0 = RMSD demandé par rapport à la structure de référence

On peut utiliser la dynamique dirigée de deux façons :

a) Directe

On dispose dans ce cas d'une structure initiale S_i et d'une structure de référence $S_{\text{réf}}$ vers laquelle on veut faire converger notre structure initiale. Soit R la valeur du rmsd entre les structures S_i et $S_{\text{réf}}$ et ρ la valeur du rmsd de la structure « en cours » par

rapport à la structure de référence. L'opération va consister en n simulations successives biaisées par le potentiel harmonique de contrainte dont le but est de maintenir la valeur du paramètre ρ proche d'une valeur de rmsd demandée (ρ_0). Chacune des simulations utilise comme jeu de coordonnées initiales la dernière conformation de la simulation précédente exceptée la première dont la structure de départ est S_i . Pour chaque nouvelle simulation, le paramètre ρ_0 est diminué de $\delta\rho$, de la valeur R pour la première simulation à la valeur 0 pour la dernière (Figure 31-a).

b) Inverse

Dans le cas de la TMD inverse, la même structure joue le rôle à la fois de structure initiale et de référence. Le paramètre RMSD est alors augmenté à chacune des simulations effectuées du pas δRMSD , de la valeur 0 pour la première simulation à une valeur R que l'on choisira de façon judicieuse. Le but de ce type de TMD n'est donc pas de faire converger une structure vers une autre, mais de contraindre une structure à s'éloigner de sa conformation initiale sans lui imposer une direction particulière afin de voir le chemin qu'elle choisira de prendre. (Figure 31-b)

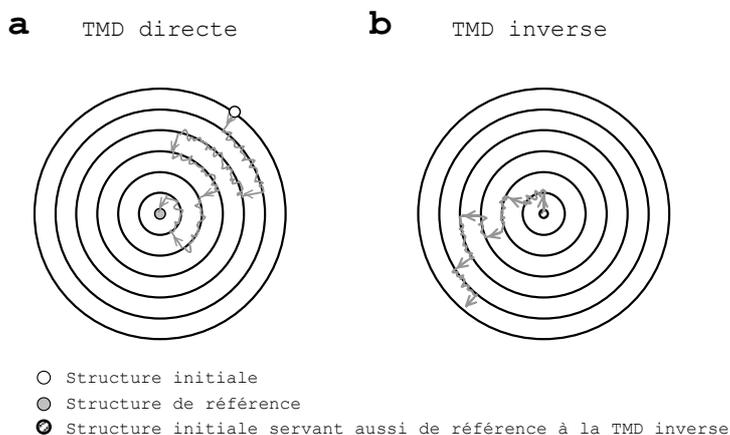


Figure 31 : Chemin emprunté par la molécule lors d'une TMD **a)** directe : la structure initiale et la structure cible sont représentées par des points vert et rouge respectivement. Le rayon du plus grand cercle correspond au rmsd entre les structures initiale et cible. A chaque nouvelle simulation, on diminue ρ_0 d'une valeur $\delta\rho$, et la molécule est équilibrée pendant un temps dt de sorte que son rmsd ρ reste proche de la valeur ρ_0 demandé. Cette opération est répétée jusqu'à atteindre un rmsd 0 avec la structure cible. **b)** inverse : dans ce cas, la structure initiale est également la structure de référence de la molécule. Le ρ_0 est progressivement augmenté d'une valeur de 0 à la valeur R voulue par incrément de δRMSD successifs. A chacun de ces paliers, la molécule est équilibrée pendant un temps dt de sorte que son rmsd reste proche de la valeur intermédiaire demandée.

III.b.iii. Protocole des dynamiques moléculaires→ Structures initiales

Les structures de départ utilisées pour réaliser ces simulations sont des structures expérimentales du dimère en « kissing-complex » et du dimère étendu de la tige-boucle SL1. Cinq structures du « kissing-complex » ont fait l'objet de simulations : une structure RMN sur la tige-boucle SL1 Δ de VIH-1_{Lai} (tige-boucle SL1 mutée au niveau des cinq premières paires de bases de la tige) (Mujeeb et al. 1998; n°PDB: 1BAU), deux structures cristallographiques sur la tige-boucle SL1 de VIH-1_{Lai} et VIH-1_{Mal} (Ennifar et al. 2001; n°PDB: 1JJM et 1JJN pour VIH-1_{Lai} et VIH-1_{Mal} respectivement), deux structures RMN sur la tige-boucle SL1* (tige-boucle SL1 de VIH-1_{Lai} mutée sur la première paire de bases de la tige possédant un nucléotide supplémentaire G₀ non-apparié) (Kieken 2004; communication personnelle).

| | |
|--------------------------|---|
| SL1 VIH-1 _{Mal} | d(C ₁ UUGCUGAGGUGCACACAGCAAG ₂₃) ₂ |
| SL1 VIH-1 _{Lai} | d(C ₁ UUGCUGAAGCGCGCACGGCAAG ₂₃) ₂ |
| SL1 Δ | d(G ₁ GCAAUGAAGCGCGCACGUUGCC ₂₃) ₂ |
| SL1* | d(G ₀ G ₁ UUGCUGAAGCGCGCACGGCAAC ₂₃) ₂ |

Tableau 2 : Sont répertoriées dans ce tableau les quatre séquences de SL1 utilisées pour déterminer des structures expérimentales de dimère. Les deux premières séquences représentent les séquences sauvages des isolats Mal et Lai respectivement, les deux dernières séquences sont des séquences mutées de SL1_{Lai}. Seule la dernière séquence (SL1*) est un 24-mer, les trois autres étant des 23-mer.

Deux structures du dimère étendu ont également été à l'origine de simulations : une structure RMN de la séquence SL1 de VIH-1_{Lai} (Girard et al. 1999; n°PDB: 1JUI) et une structure cristallographique de la séquence SL1 de VIH-1_{Mal} (Ennifar et al. 1999; n°PDB: 462D). Les différentes séquences de ces structures sont résumées dans le Tableau 2.

→ Topologie

Toutes nos simulations sont réalisées en présence d'ions, sodium ou sodium et magnésium selon les simulations, et d'eau explicite. Les

ions sodium sont rajoutés soit avec le module LEaP, soit avec un programme développé au laboratoire, `GENERE_ION`. Ce programme distribue aléatoirement les ions sur une grille de $0.5 \text{ \AA} \times 0.5 \text{ \AA} \times 0.5 \text{ \AA}$ de façon à ce que chacun des ions soit au moins à 4 \AA de tout autre ion ainsi que de tout atome de la molécule, chaque phosphore de la molécule étant associé à un ion situé à une distance maximale de 7 \AA . Dans le cas des simulations en présence d'ions magnésium, les coordonnées des ions magnésiums sont celles issues du fichier PDB. Les ions magnésium sont présents dans les structures déterminées par cristallographie. Afin d'avoir un système électriquement neutre, on place des ions sodium de la même façon que précédemment puis on supprime les ions sodium en surplus les plus proches des ions magnésium. Le système est ensuite immergé dans une boîte d'eau, de géométrie et de taille variables suivant les simulations. Les boîtes peuvent être parallélépipédiques ou octaédriques tronquées et la distance minimale entre n'importe quel atome du système et le bord de la boîte varie selon les simulations. La plus petite de ces distances utilisée pour une simulation est de 8.4 \AA afin de s'assurer qu'il y ait au moins trois couches d'eau autour de la molécule, et peut aller jusqu'à 12 \AA . Nous obtenons ainsi des systèmes à simuler possédant de 20000 à 50000 atomes.

→ Minimisation

Une fois le système construit, il subit une étape de minimisation afin de permettre une organisation réaliste de l'eau autour de la molécule ainsi qu'une optimisation de la géométrie de la structure de départ. Cette minimisation se déroule en trois étapes :

1^{ère} étape : on applique des contraintes de position sur les atomes de la molécule et sur les ions et on minimise l'énergie des molécules d'eau pendant 1000 étapes, les 100 premières utilisant l'algorithme de la plus grande pente, le reste des étapes étant effectué avec l'algorithme des gradients conjugués. La minimisation est réalisée à volume constant.

2^{ème} étape : cette fois les atomes de la molécule et les ions sont laissés libres de minimiser leur énergie tandis que les molécules d'eau sont assignées à leurs positions par

des contraintes. Les autres conditions de minimisation sont les mêmes que ci-dessus.

3^{ème} étape : tous les atomes sont laissés libres et on réalise 10000 étapes de minimisation dont les 100 premières avec l'algorithme de la plus grande pente.

→ Chauffage

Le chauffage est exécuté dans un ensemble NVT (nombre de particule, volume et température constants) par palier de 50 K entre 0 K et 300 K et, le cas échéant, par paliers de 30 K entre 300 K et 360 K. La plupart de nos dynamiques sont réalisées à l'une de ces deux températures. Les atomes de la molécule et les ions sont assignés à leurs positions par un potentiel harmonique de contrainte d'une constante de force $k=100$ kcal/mol afin de ne pas perturber le système pendant le chauffage. Les vitesses sont distribuées aléatoirement à chaque palier de chauffage lequel est suivi d'une période d'équilibration de 2 ps. Lorsque l'on a atteint la température T de la simulation, les contraintes sur les ions puis sur les atomes de la molécule sont progressivement relâchées sur des étapes de temps de 20 ps et une dernière étape d'équilibration de 20 ps à la température T est effectuée sur le système sans contraintes.

→ Production

Les productions sont réalisées dans un ensemble NPT (nombre de particules, pression et température constants) et en utilisant les algorithmes PME pour le calcul des forces électrostatiques et SHAKE sur les liaisons contenant un hydrogène afin d'autoriser un pas d'intégration de 0.002 ps.

III.b.iv. Analyses des trajectoires

→ Liaisons hydrogène

Nous analysons la probabilité de présence des liaisons hydrogène sur une période donnée à l'aide du module CARNAL de AMBER. Le programme procède comme suit : il calcule le pourcentage des configurations pour lesquelles la liaison hydrogène existe par rapport à l'ensemble des configurations stockées dans la trajectoire. On considère que la

liaison hydrogène existe entre un donneur D et un accepteur A lorsque la distance entre ces deux atomes n'excède pas 3.5 Å et lorsque l'angle D-H-A est compris entre 135° et 180° (Figure 32) :

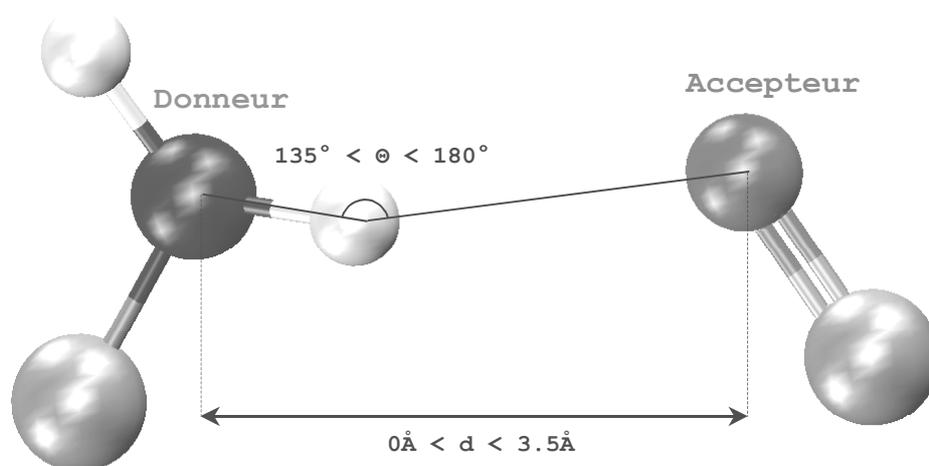


Figure 32 : Définition d'une liaison hydrogène. La distance entre D et A est comprise entre 0 Å et 3.5 Å et l'angle D-H...A est compris entre 135° et 180°.

→ Carte de flexibilité

Une méthode développée par Gaudin et al. (1997) permet de cartographier les régions flexibles et rigides d'une molécule à partir de l'ensemble ou d'une partie de la trajectoire. Sur ces cartes sont représentés les fluctuations moyennes des distances inter-atomiques calculées par la méthode des moindres carrés. Sur les côtés de la carte sont reportés les numéros des atomes tels qu'ils existent dans la topologie de la molécule. On utilise pour ce faire une topologie dans laquelle les atomes d'hydrogène de la molécule ont généralement été exclus. On détermine ensuite des valeurs de coupures σ_1 et σ_2 de sorte que les fluctuations inférieures à σ_1 sont représentées en noir sur la carte et correspondent aux zones rigides de la molécule, les fluctuations comprises entre σ_1 et σ_2 correspondent à des zones semi-rigides représentées en gris sur la carte et les fluctuations supérieures à σ_2 désignent les zones flexibles, laissées en blanc sur la carte (Figure 33).

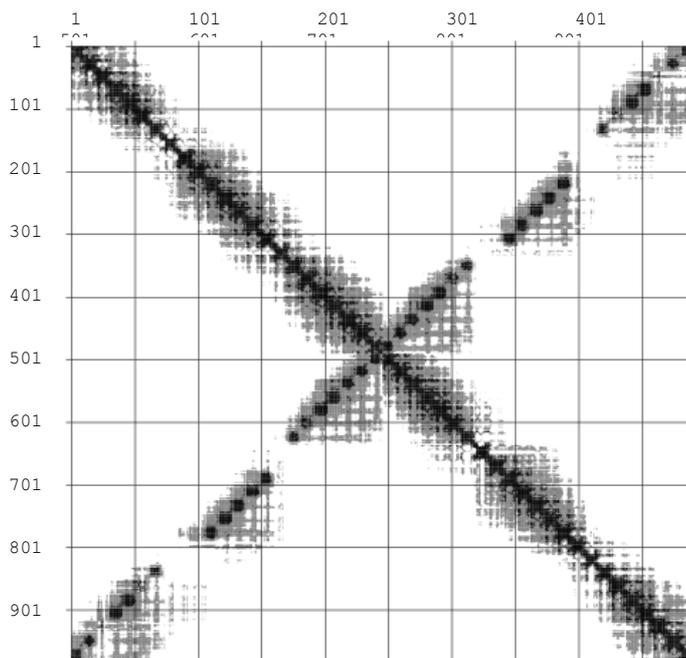


Figure 33 : La carte de flexibilité d'une molécule représente les fluctuations standards σ des distances inter-atomiques. Les numéros reportés sur le côté de la carte correspondent aux numéros des atomes. Deux valeurs de coupures σ_1 et σ_2 ont été déterminées pour réaliser cette carte. Les points noirs représentent les valeurs de $\sigma \leq \sigma_1$, les points gris les valeurs de σ telles que $\sigma_1 < \sigma \leq \sigma_2$. Les valeurs de $\sigma > \sigma_2$ sont laissées en blanc.

→ Calcul de l'énergie libre de transition (ΔG) par la méthode WHAM

WHAM ou Weighted Histogram Analysis Method (Kumar et al. 1992; Roux 1995) est une technique permettant d'estimer le potentiel de la force moyenne en fonction d'une ou plusieurs coordonnées de réaction x . Pour cela on réalise une série de n simulations sous contraintes. Pour chacune de ces simulations i , on applique un potentiel de contrainte $U_i(x)$ dont le but est de maintenir x proche d'une valeur ρ_i prédéfinie. On détermine ensuite la probabilité $P_{0i}(x)$ pour que la coordonnée de réaction ait la valeur x dans la simulation i . On peut démontrer que cette probabilité en absence de contraintes serait:

$$P_0(x) = P_{0i}(x) \left(\frac{Z_{0i}}{Z_0} \right) \exp \left\{ \frac{U_i(x)}{kT} \right\} \quad (27)$$

où Z_{0i} et Z_0 sont les fonctions de partitions de P_{0i} et P_0 respectivement
 k est la constante de Boltzmann
 T est la température

Comme $P_0(\mathbf{x})$ devrait être indépendant de i , on l'approxime par une combinaison linéaire des n probabilités P_{0i} dont les coefficients $\omega_i(\mathbf{x})$ répondent aux conditions suivantes : $\sum_{i=1}^N \omega_i(\mathbf{x}) = 1$ et ils minimisent $\langle P_0^2(\mathbf{x}) \rangle - \langle P_0(\mathbf{x}) \rangle^2$. La probabilité sans contrainte s'exprime alors par la formule suivante :

$$P_0(\mathbf{x}) = \sum_{i=1}^n \omega_i(\mathbf{x}) P_{0i}(\mathbf{x}) \left(\frac{Z_{0i}}{Z_0} \right) \exp \left\{ \frac{U_i(\mathbf{x})}{kT} \right\} \quad (28)$$

Ces nouvelles conditions mènent à l'expression:

$$P_0(\mathbf{x}) = \frac{\sum_{i=1}^n N_i P_{0i}(\mathbf{x})}{\sum_{i=1}^n N_i \left(\frac{Z_{0i}}{Z_0} \right) \exp \left\{ \frac{U_i(\mathbf{x})}{kT} \right\}} \quad (29)$$

avec

$$\frac{Z_0}{Z_{0i}} = \int P_0(\mathbf{x}) \exp \left\{ -\frac{U_i(\mathbf{x})}{kT} \right\} d\mathbf{x} \quad (30)$$

où N_i est le nombre de configurations utilisées pour déterminer $P_{0i}(\mathbf{x})$.

Les équations (29) et (30) sont résolues par une procédure itérative jusqu'à ce que la variation des $\frac{Z_0}{Z_{0i}}$ soit suffisamment faible. La condition de validité de l'approche WHAM est que les différentes probabilités adjacentes P_{0i} doivent être suffisamment proches pour se recouvrir. Le profil d'énergie libre est obtenu à partir de $P_0(\mathbf{x})$ d'après l'équation :

$$\Delta G = -kT \log \{ P_0(\mathbf{x}) \} \quad (31)$$

Pour notre étude, la coordonnée de réaction x que nous avons utilisée est le RMSD_w . Il s'agit d'une variante du RMSD décrit précédemment, où chaque terme de la somme est pondéré par la masse de l'atome correspondant :

$$\text{RMSD}_w = \left[\frac{1}{N} \sum_{i=1}^N m_i \|r_{1i} - r_{2i}\|^2 \right]^{\frac{1}{2}} \quad (26)$$

Les calculs de $RMSD_w$ sont effectués avec le module CARNAL du programme AMBER. Les analyses WHAM ont été réalisées à l'aide du logiciel WHAM écrit par A. Grossfield

(<http://dasher.wustl.edu/alan/>).

→ Calcul de l'entropie et de l'enthalpie

Entre deux états initial et final, on calcule les variations d'enthalpie et d'entropie à l'aide des équations thermodynamiques suivantes :

$$\Delta H = \Delta E + \Delta(PV) \quad (32)$$

$$\Delta G = \Delta H - T\Delta S \quad (33)$$

où E est l'énergie interne du système

P, V et T sont respectivement la pression, le volume et la température du système

Du fait que nos simulations sont réalisées dans l'ensemble NPT, donc à pression constante, et que les variations de volume de la boîte d'eau sont très faibles au cours d'une simulation, on fait l'approximation $\Delta(PV)=0$ d'où :

$$\Delta H = \Delta E \quad (34)$$

$$\Delta S = \frac{1}{T}(\Delta E - \Delta G) \quad (35)$$

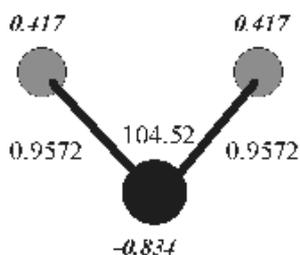
III.c. Les logiciels utilisés

III.c.i. Le logiciel AMBER

AMBER ou "Assisted Model Building with Energy Refinement" (Pearlman et al. 1995; Weiner and Kollman 1981) est un programme académique de simulation de dynamique moléculaire optimisé dans l'étude des acides nucléiques et des protéines. Il propose plusieurs champs de force tels OPLS (Jorgensen and Tirado-Rives 1988), parm94.dat (Cornell et al. 1995), parm98.dat (Cheatham et al. 1999), etc... selon les besoins de la simulation. Dans le cadre de simulations d'acides nucléiques en solvant explicite, qui est le cas qui nous préoccupe, le champ de force le mieux adapté est parm98. Nous avons utilisé les versions 5.0 (Case et al. 1997), 6.0 (Case et al. 1999) et 7.0 (Case et al. 2002) du logiciel AMBER pour réaliser nos simulations de dynamique moléculaire.

→ Modèles pour l'eau et les ions

Différents modèles d'eau explicite sont également implémentés dans le programme AMBER et notamment le modèle TIP3P (Jorgensen et al. 1983) qui est le modèle le plus couramment utilisé dans les simulations de dynamique moléculaire en eau explicite. Ce modèle est dit « à 3 points » car la molécule d'eau ainsi représentée comprend 3 sites d'interactions centrés sur les noyaux atomiques de l'oxygène et des protons.



Chaque site possède une charge partielle afin de simuler les interactions coulombiennes intermoléculaires (-0.834 e pour l'oxygène et 0.417 e pour chacun des hydrogènes). L'oxygène est de plus paramétré pour satisfaire à la fonction de Lennard-Jones. Ce modèle est rigide,

avec des distances fixes pour les liaisons O-H (0.96 Å) et un angle H-O-H de 104.5°.

Les ions peuvent également être représentés explicitement au cours des simulations. Ils sont modélisés comme une charge ponctuelle disposant de paramètres de Lennard-Jones ajustés de façon à reproduire les énergies libres de solvation et la distribution radiale avec le modèle TIP3P (Aqvist 1990).

→ Structure du programme

Le programme AMBER est composé de plusieurs modules, chacun d'eux jouant un rôle bien spécifique. (Figure 34)

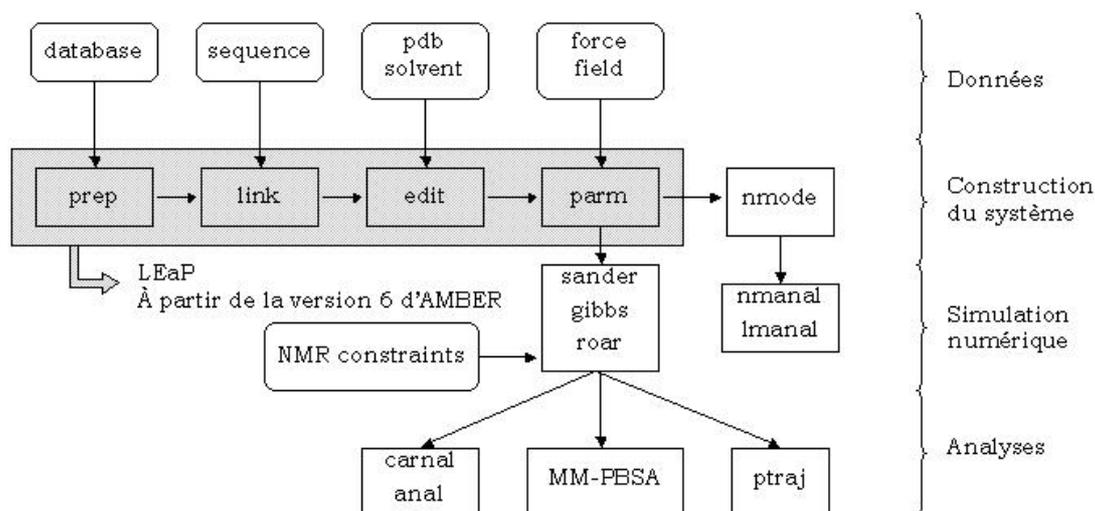


Figure 34 : Organigramme des modules constituant le logiciel AMBER.

Les modules les plus couramment utilisés au cours de nos simulations sont :

- le module **LEaP** : il permet, à partir d'un fichier de coordonnées au format PDB et du choix d'un champ de force de créer la topologie du système que l'on va simuler.
- le module **sander** : c'est le programme principal de minimisation d'énergie et de dynamique moléculaire d'AMBER.
- le module **carnal** : il permet d'analyser les données structurales des trajectoires produites par la simulation.

III.c.ii. Logiciel CURVES

CURVES (Lavery and Sklenar 1988) est un algorithme de calcul des paramètres hélicoïdaux développé au laboratoire de Biochimie Théorique de l'IBPC qui respecte le choix des noms et les signes de convention déterminés lors de la convention de Cambridge évoquée au chapitre précédent.

Ce programme est utilisable sur tout segment irrégulier d'oligonucléotides dont la géométrie globale respecte grossièrement celle d'une double-hélice hélicoïdale.

Ce programme fonctionne par détermination d'un axe hélicoïdal unique par rapport auquel sont calculés les paramètres hélicoïdaux. Cet axe est obtenu par minimisation d'une fonction qui tient compte des irrégularités géométriques à la fois intra- et inter-bases ainsi que des axes locaux pouvant être associés à chaque paire de bases.

Pour cela, deux systèmes d'axe sont définis :

- un système d'axe de référence ($E, \vec{j}, \vec{k}, \vec{l}$) pour localiser la base dans l'espace (Figure 35-a)
- un système d'axe hélicoïdal local ($P, \vec{v}, \vec{w}, \vec{u}$) (Figure 35-b)

L'origine du référentiel E correspond à la position de l'axe hélicoïdal dans le plan de la base. Le point P est choisi de sorte que le vecteur joignant les points E et P soit toujours perpendiculaire au vecteur \vec{u} .

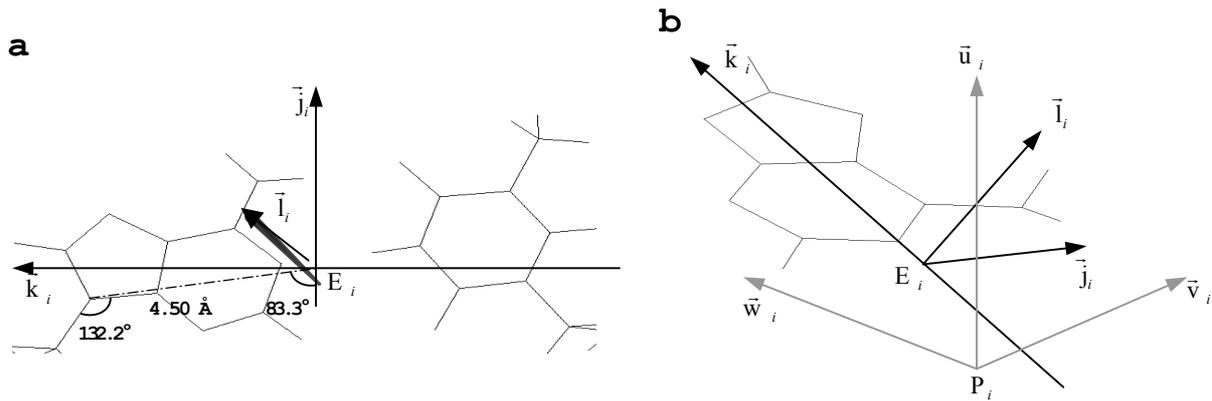


Figure 35 : Définition des axes hélicoïdaux locaux par CURVES **a)** système d'axe de référence localisant la base dans l'espace **b)** système d'axe hélicoïdal local de la base dans l'espace

Les paramètres calculés par ce programme sont divisés en plusieurs catégories :

- Paramètres de l'axe ou de l'axe par rapport aux bases ou aux paires de bases
position des origines E et P des systèmes d'axe
paramètres x -displacement, y -displacement, inclination, tip
- Paramètres intra-paires de bases
paramètres shear, buckle, stretch, propeller, stagger, opening
- Paramètres inter-bases ou paires de bases (globaux et locaux)
paramètres shift, tilt, slide, roll, rise, twist
- Paramètres de courbure et du squelette phospho-diester
angle et direction de courbure de l'axe
angles dièdres : Phase, amplitude, pucker, angles $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \chi$
- Paramètres de l'hélice

III.c.iii. Visualisation : VMD

Le logiciel Visual Molecular Dynamics ou VMD (Humphrey et al. 1996) est un programme de visualisation graphique, qui permet de représenter en 3D et même d'animer des macromolécules biologiques. Ce programme lit aussi bien les fichiers de coordonnées issus de la PDB que les fichiers trajectoires générés par le programme AMBER. Ce programme a été développé au sein du « Theoretical and Computational Biophysics Group » à l'université de l'Illinois (<http://www.ks.uiuc.edu/Research/vmd/>).

IV . RESULTATS

IV. RESULTATS

IV.a. Etude du dimère instable de SL1

La diversité des structures expérimentales du « kissing-complex » parues dans la littérature pour la tige-boucle SL1 nous a amenés à envisager comme première étape de ce travail de thèse l'étude de la stabilité de ces structures en solution.

Pour toutes les analyses des trajectoires obtenues nous avons découpé le « kissing-complex » en cinq groupes de résidus décrits Figure 36. La numérotation des résidus employée au cours de cette partie y figure également. Les résidus 1 à 23 correspondent aux résidus 265 à 287 de la molécule entière (voir §II.b.ii, Figure 9).

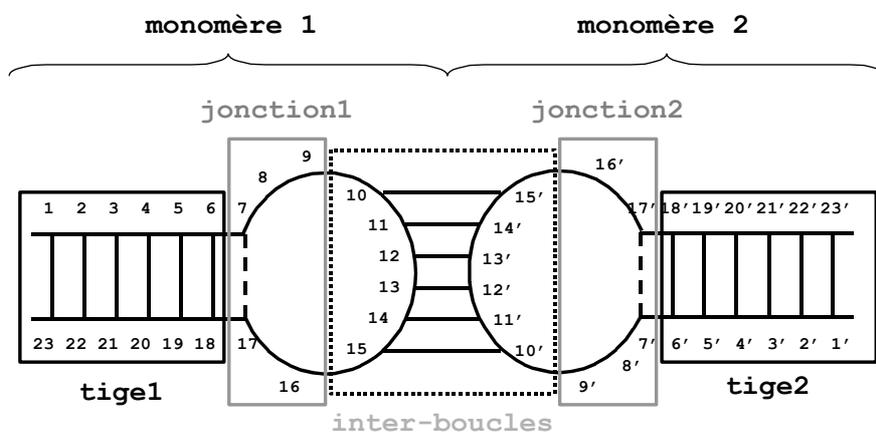


Figure 36 : Définition des cinq groupes de résidus utilisés lors des analyses sur les simulations du dimère instable. On appellera tige 1 et tige 2 les groupes composés des résidus {1→6 ; 18→23} et {1'→6' ; 18'→23'} respectivement, jonction1 et jonction2 les groupes composés des résidus {7→9 ; 16, 17} et {7'→9' ; 16', 17'} respectivement et inter-boucle le groupe composé des résidus {10→15 ; 10'→15'}

IV.a.i. Etude de la structure RMN de Mujeeb et al. avec la séquence SL1A

Nous avons effectué une dynamique moléculaire de 20 nanosecondes à 300K à partir de la structure RMN publiée par Mujeeb et al. (1998) en utilisant le logiciel AMBER5.0. La molécule a été simulée en présence de 44 contre-ions Na^+ dans une boîte d'eau parallélépipédique contenant 6987 molécules d'eau. On a de plus imposé une contrainte de position sur la liaison C3'-C4' du résidu U19 afin de limiter la rotation de la molécule dans la boîte (voir *Les conditions périodiques §III.b.i*).

Tout au long de la simulation, la structure reste en forme de « kissing-complex », c'est à dire que les appariements dans les tiges mais également entre les boucles sont maintenus de bout en bout. On observe cependant des réarrangements structuraux importants sur la molécule, notamment au niveau des jonctions. Le rmsd du dimère entier se stabilise autour de 4 Å par rapport à la structure initiale à l'exception d'un pic de 7 Å à environ 10 ns de simulation.

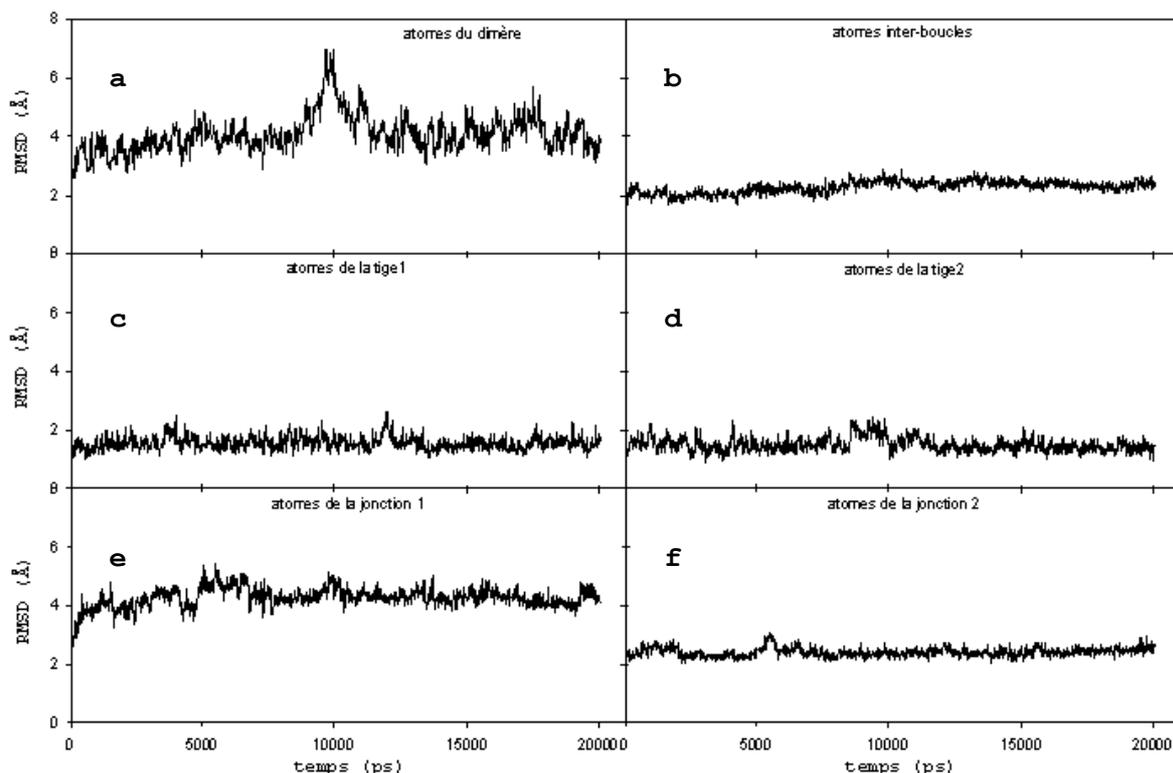


Figure 37 : Evolution du rmsd au cours de la simulation du dimère instable de Mujeeb et al. (1998) par rapport à la structure expérimentale sur les atomes : du dimère complet (a), du groupe inter-boucle (b), de la tige 1 (c), de la tige 2 (d), de la jonction 1 (e) et de la jonction 2 (f)

Ce pic est dû à une ouverture temporaire de l'angle entre les axes des tiges ainsi que représenté sur la Figure 38. Pendant le reste de la simulation, cet angle oscille entre 90° et 130° avec une valeur moyenne en fin de simulation de l'ordre de 110° . Le changement structural principal est observé au niveau de la jonction 1, dont le rmsd oscille entre 4 Å et 5 Å par rapport à la structure RMN, et ce dès les premières centaines de picosecondes de simulation. Une observation plus fine de cette jonction montre un réarrangement important au niveau des résidus G7 et C17.

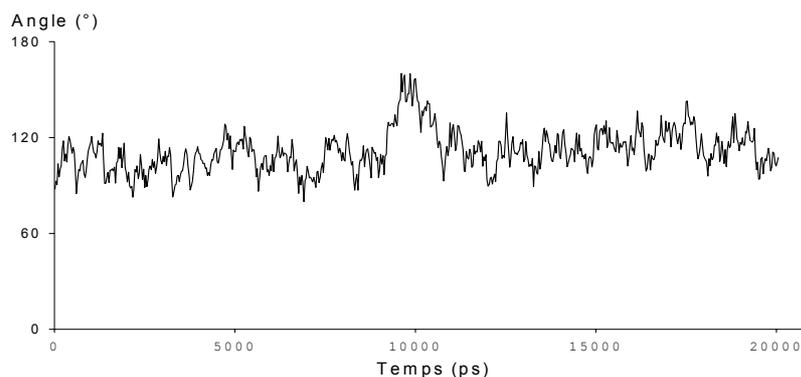


Figure 38 : Evolution de l'angle entre les axes des tiges au cours de la simulation du dimère instable de Mujeeb et al. (1998).

Ces résidus présentent dans la structure RMN une conformation particulière que nous avons déjà évoquée au paragraphe §II.b.ii, à savoir que le sucre du résidu G7 est en conformation C2'-endo et que les deux résidus G7 et C17 sont désappariés. Il s'avère qu'au cours de notre simulation nous avons remarqué l'enchaînement des phénomènes suivant : 1) le passage du résidu G7 de la conformation C2'-endo à la conformation C3'-endo ; 2) la formation simultanée de trois liaisons hydrogène canoniques entre les résidus G7 et C17, ces deux événements intervenant à 1 ns d'intervalle l'un de l'autre. La Figure 39 illustre ce phénomène.

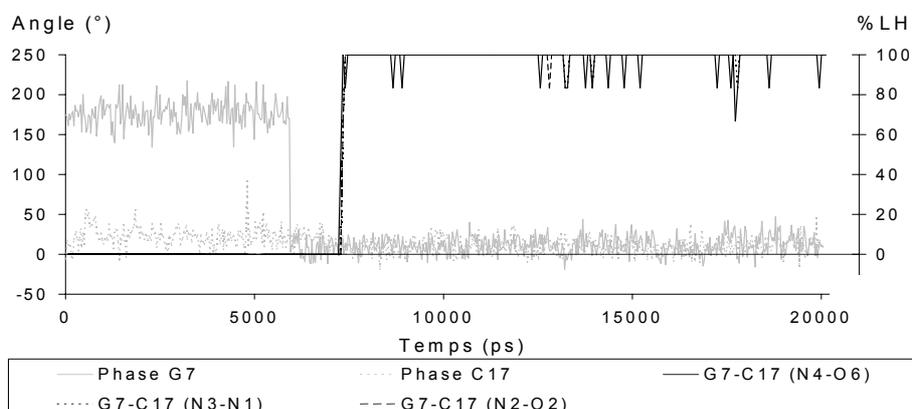


Figure 39 : Mise en parallèle de l'évolution de la conformation des sucres et de l'évolution du pourcentage d'existence des liaisons hydrogène entre les résidus G7 et C17 du dimère instable de Mujeeb et al. (1998). Si l'angle de phase est compris entre 0° et 36° , le sucre est en conformation C3'-endo (conformation du sucre dans une hélice A idéale). Si il est compris entre 144° et 180° , le sucre est alors en conformation C2'-endo (conformation du sucre dans une hélice B idéale.)

On constate de surcroît que cette paire de bases reste stable pendant les 13 ns de simulation suivantes. Ce résultat est confirmé

par la carte de flexibilité présentée Figure 40 et calculée sur les 10 dernières nanosecondes de la trajectoire. Les résidus corrélés entre eux apparaissent comme des taches noires sur la carte. Ainsi, on voit bien que les six résidus des boucles des monomères 1 et 2, mis en évidence dans l'ellipse grise, sont corrélés entre eux, ce qui nous conforte dans l'idée que les appariements sont maintenus au cours de la dynamique. On voit pour chacun des monomères sept taches de corrélations entre les résidus, visualisées par une ellipse bleue pour le monomère 1 et par une ellipse rouge pour le monomère 2. Cependant, si le monomère 1 présente de belles taches de corrélation et notamment une tache très nette entre G7 et C17, traduisant ainsi une forte corrélation entre ces deux résidus, le monomère 2, lui, ne présente que 5 taches bien dessinées pour les cinq premières paires de bases. La sixième paire, soit la paire U6'-G18', présente un net affaiblissement et la corrélation entre G7' et C17' se révèle quasi inexistante.

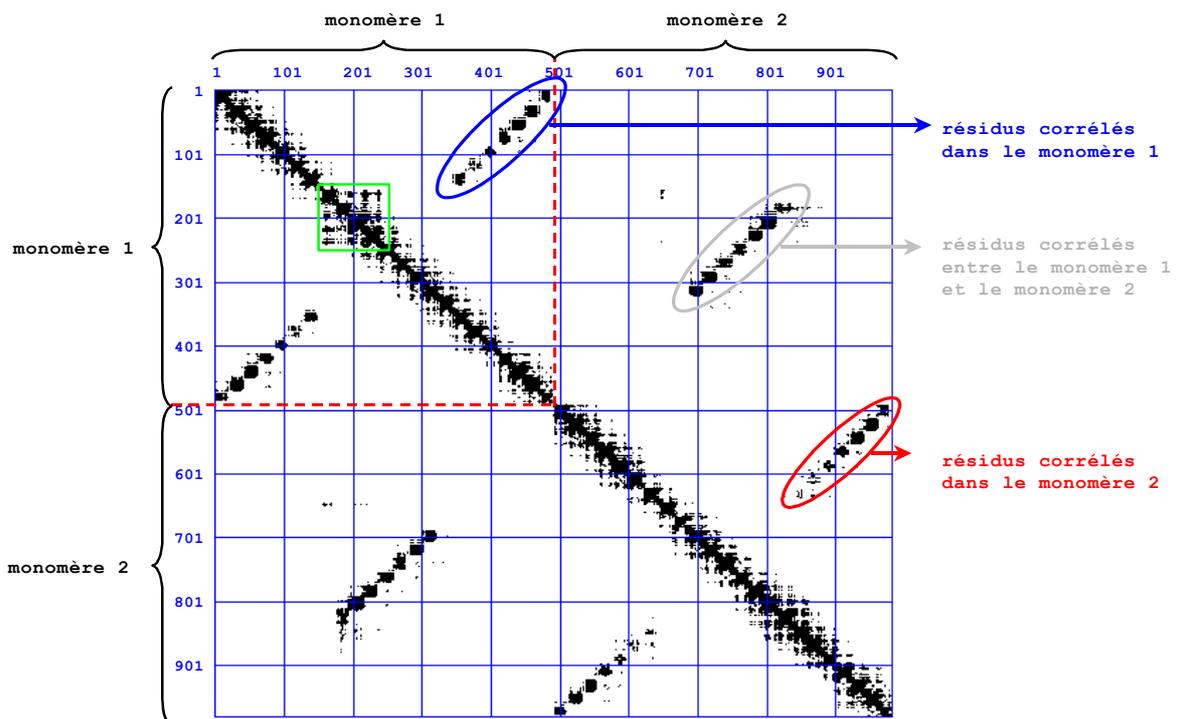


Figure 40 : Carte de flexibilité réalisée sur les 10 dernières nanosecondes de simulation de la structure de Mujeeb et al. (1998) avec une valeur de coupure $\sigma=0.3$ Å. Les zones noires correspondent à des fluctuations standard des distances inter-atomiques inférieures à σ , les valeurs supérieures étant laissées en blanc. Les numéros des atomes sont reportés sur les côtés de la carte.

On remarque de surcroît une corrélation entre les résidus 8, 9, 10 et 11 du monomère 1 mise en évidence par le carré vert sur la Figure 40. En revanche, les résidus de la jonction 2 semblent très flexibles. On remarque sur la contre-diagonale une tache supplémentaire au niveau des interactions entre les deux monomères. Il semble qu'une interaction stable entre les résidus A9 et A16' soit ainsi mise en évidence. L'analyse des pourcentages d'existence des liaisons hydrogène sur l'ensemble de la trajectoire montre en effet l'existence d'une liaison A9(N6)-A16'(N3), présente dans la structure initiale et conservée tout au long de la simulation. D'autres liaisons hydrogène entre les adénines des jonctions sont ainsi visualisées au cours du temps de simulation, mais de façon plus intermittente. En particulier, les résidus 16' et 8' du second monomère sont reliés de façon discontinue tout au long de la simulation par une liaison hydrogène A16'(N6)-A8'(N1). Les résidus A16 et A9' sont également reliés par des liaisons hydrogène A16(N6)-A9'(N1) et A16(N1)-A9'(N6) jusqu'à environ 7-8 ns de simulation. La Figure 41 présente la structure obtenue en fin de dynamique.

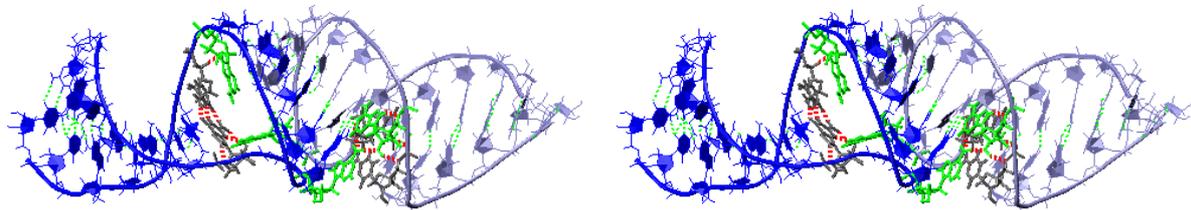


Figure 41 : Représentation en vue stéréo de la dernière conformation des 20 ns de simulation du dimère instable de Mujeeb et al. (1998). Les monomères 1 et 2 sont représentés en bleu et violet respectivement. Les résidus 7, 7', 17 et 17' sont représentés en gris. Les résidus 8, 8', 9, 9', 16 et 16' sont représentés en vert.

IV.a.ii. Etude de la structure RMN de Mujeeb et al. avec la séquence SL1 sauvage

Afin de vérifier l'influence de la séquence sur la stabilité de la structure en « kissing-complex » déterminée par RMN, nous avons utilisé le programme MUTATION développé au laboratoire pour muter la séquence SL1Δ en séquence SL1 sauvage à partir des coordonnées publiées par Mujeeb et al. (1998) dans la PDB. Nous avons ensuite réalisé une simulation de 6 nanosecondes à 300 K à partir de cette molécule en présence de 44 contre-ions Na⁺ dans une boîte

parallélépipédique contenant 7193 molécules d'eau. La molécule révèle au cours de cette simulation un comportement assez semblable à celui décrit au chapitre précédent.

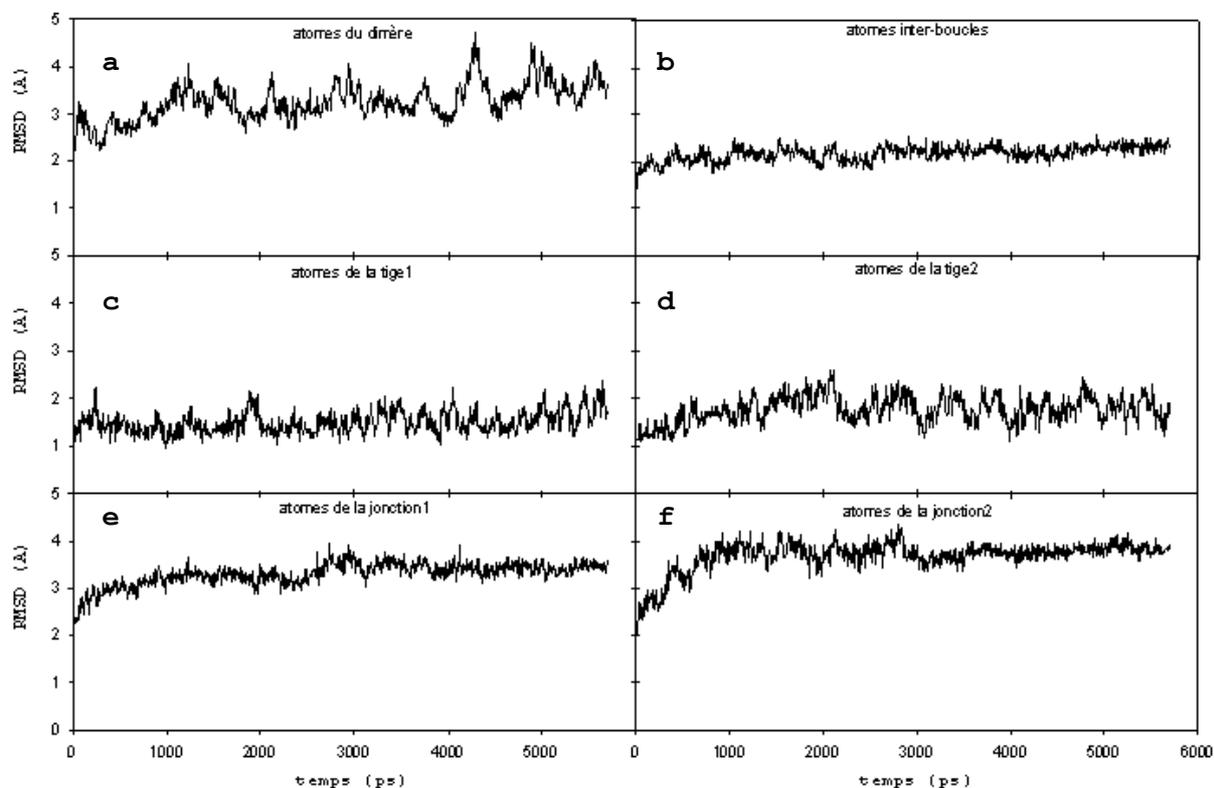


Figure 42 : Evolution du rmsd au cours de la simulation du dimère instable de Mujeeb et al. (1998) avec la séquence sauvage par rapport à la structure expérimentale sur les atomes : du dimère complet (a), du groupe inter-boucle (b), de la tige1 (c), de la tige 2 (d), de la jonction 1 (e) et de la jonction 2 (f)

Les courbes de RMSD représentées Figure 42 montre aussi une déviation structurale sur l'ensemble de la molécule de l'ordre 4-5 Å avec des changements conformationnels principalement localisés dans les jonctions. Contrairement à la simulation précédente, les deux jonctions présentent une divergence de structure importante avec la structure d'origine, et en particulier la jonction 2. En effet, on remarque pour cette molécule, mais cette fois sur le monomère 2, l'un des phénomènes observé précédemment, à savoir la formation simultanée des trois liaisons canoniques entre les résidus G7' et C17' (Figure 43).

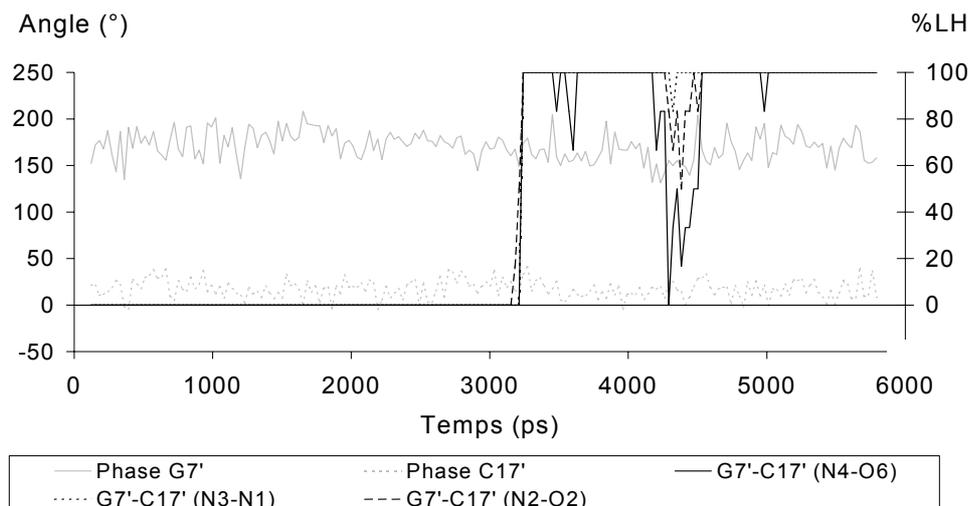


Figure 43 : Mise en parallèle de l'évolution de la conformation des sucres et de l'évolution du pourcentage d'existence des liaisons hydrogène entre les résidus G7 et C17 du dimère instable de Mujeeb et al. (1998) avec la séquence sauvage.

Cet appariement se produit cette fois sans qu'il y ait changement de conformation pour le sucre de G7'. La paire ainsi formée reste globalement stable sur tout le reste de la simulation bien que l'on remarque un affaiblissement autour de 4.3 ns de simulation.

La carte de flexibilité présentée Figure 44 a été réalisée sur la dernière nanoseconde de simulation avec une valeur de coupure $\sigma=0.3 \text{ \AA}$. On peut voir sur cette carte que le monomère 1 semble plus rigide que le monomère 2 et présente en particulier une zone de rigidité entre les résidus 5 à 10 et 16 à 20 (dans l'ellipse bleue). La première paire de base de sa tige semble en revanche assez flexible. Le monomère 2 présente sur sa contre-diagonale sept taches nettes, visualisées par l'ellipse rouge, correspondant aux sept paires de bases formées sur sa tige. Les interactions inter-boucles, montrées dans l'ellipse grise, révèlent, outre les six taches correspondant aux six paires de bases GC du dimère maintenues tout au long de la simulation, deux taches supplémentaires situées de part et d'autre de ce premier ensemble qui mettent en évidence des interactions stables existantes entre les résidus A9, A16' d'une part et A9', A16 d'autre part. L'étude des liaisons hydrogène montre que les liaisons A16(N3)-A9'(N6) et A9(N6)-A16'(N3), présentes dans la structure initiale, sont conservées au cours de la simulation.

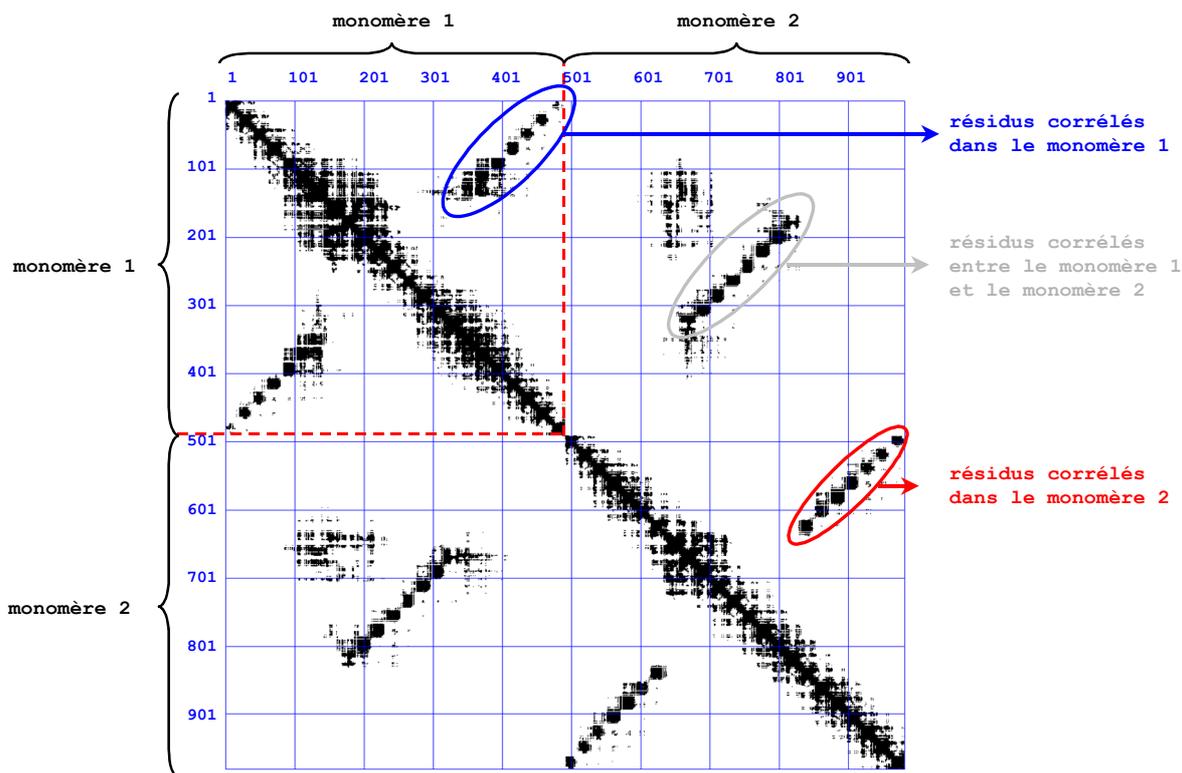


Figure 44 Carte de flexibilité réalisée sur la dernière nanoseconde de simulation de la structure de Mujeeb et al. (1998) avec la séquence sauvage avec une valeur de coupure $\sigma=0.3 \text{ \AA}$.

La Figure 45 présente l'évolution de l'angle entre les axes des tiges en fonction du temps. Cet angle fluctue entre 50° et 130° pendant la simulation, avec une moyenne de 110° sur la dernière nanoseconde de simulation.

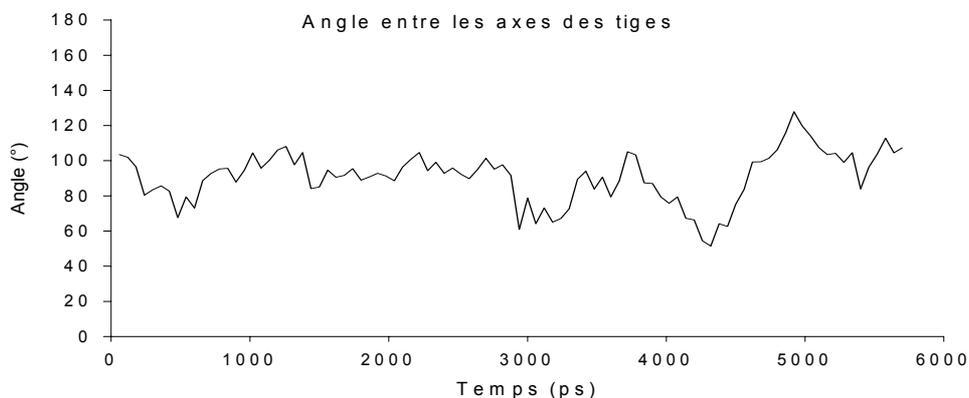


Figure 45 : Evolution de l'angle entre les axes des tiges pendant la simulation de la structure de Mujeeb et al. (1998) avec la séquence sauvage.

La structure obtenue en fin de dynamique est donnée Figure 46.

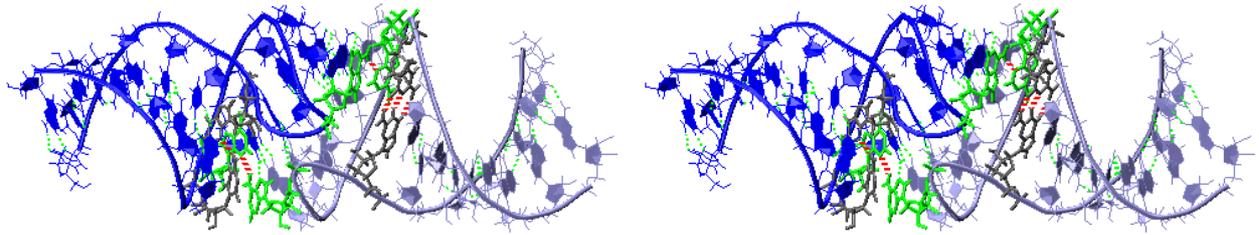


Figure 46 : Représentation en vue stéréo de la dernière structure obtenue lors de la simulation de la structure de Mujeeb et al. (1998) avec la séquence sauvage.

I.a.i. Etude des structures cristallographiques de Ennifar et al. à partir de la séquence SL1. Influence des ions magnésiums

L'influence des ions magnésium sur la stabilité du « kissing-complex » a fait l'objet de nombreuses controverses dans la littérature publiée sur le sujet, et ce plus particulièrement en ce qui concerne le dimère instable de la séquence SL1_{Mal}. En effet, du fait des deux paires A-U dans sa séquence auto-complémentaire GUGCAC le dimère de SL1_{Mal} forme une interaction boucle-boucle plus faible que celle de SL1_{Lai}, de séquence GCGGC, et l'intervention d'ions magnésium stabilisant pour cette interaction est une hypothèse qui a été plusieurs fois avancée (Ennifar et al. 2001; Jossinet et al. 1999; Marquet et al. 1991). Ces deux structures ayant été déterminées par cristallographie en présence d'ions magnésium, les fichiers de coordonnées publiés pour ces structures contiennent également les coordonnées des ions magnésium, ce qui nous a permis pour ces deux « kissing-complex » de réaliser des simulations de dynamique en absence et en présence d'ions magnésium.

→ Etude du dimère instable de SL1_{Lai}

Deux simulations ont été réalisées à partir de cette molécule, l'une de 5.5 nanosecondes en présence de 44 contre-ions Na⁺ dans une boîte d'eau parallélépipédique contenant 6513 molécules d'eau, l'autre de 1.8 nanosecondes en présence de 12 ions Mg²⁺ et de 20 ions Na⁺ dans une boîte parallélépipédique contenant 7517 molécules d'eau. Cette dernière dynamique a été effectuée en second afin de vérifier l'influence des ions magnésium sur la stabilité de cette structure de « kissing-complex » ; elle ne dure que 2 nanosecondes car nous

l'avons stoppée dès lors que nous avons vu apparaître les mêmes phénomènes que sur la première simulation. Les premières observations de ces trajectoires montrent un résultat surprenant : au cours de la simulation on remarque un déroulement des tiges suivi d'une ouverture des paires de bases des tiges, et ce pour les deux simulations. Parallèlement, les résidus 8 et 9, tournés vers l'extérieur dans la structure initiale, tendent à se tourner vers l'intérieur de la molécule. Sur la Figure 47 sont reportées les évolutions des rmsd en fonction du temps par rapport à la structure initiale. Celles-ci confirment que la molécule subit de grandes déviations au cours de la simulation, notamment au niveau des tiges du « kissing-complex ». On remarque en outre que le comportement des différentes parties de la molécule est similaire en présence comme en absence d'ions magnésium.

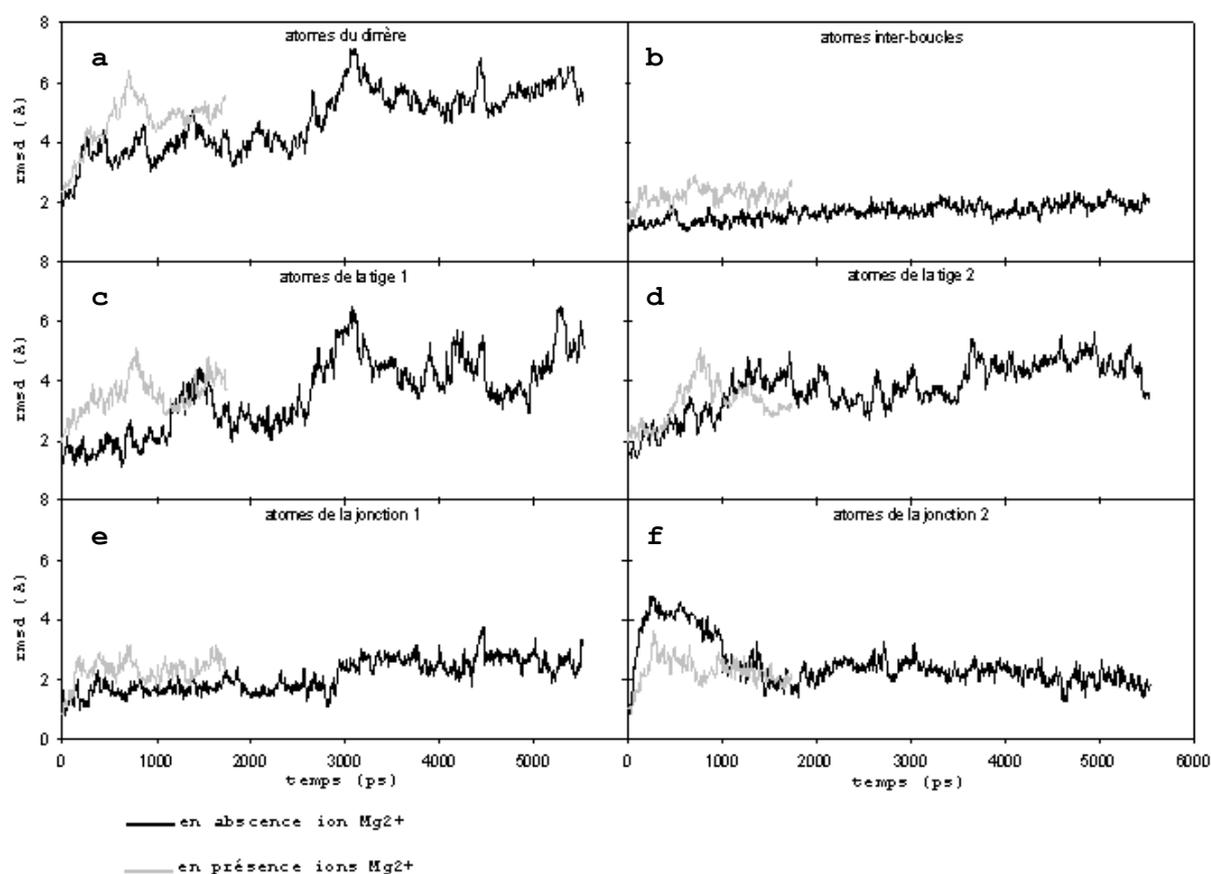


Figure 47 : Evolution du rmsd au cours de la simulation du dimère instable de SL1_{Lai} de Ennifar et al. (2001) en absence (en noire) et en présence (en gris) d'ions magnésium par rapport à la structure expérimentale sur les atomes : du dimère complet (a), du groupe inter-boucle (b), de la tige 1 (c), de la tige 2 (d), de la jonction 1 (e) et de la jonction 2 (f)

La Figure 48 présente le pourcentage d'existence moyen des liaisons hydrogène par paire de base au cours de la simulation. On peut constater que, si les paires de bases inter-boucles restent appariées pendant presque 100% du temps de simulation, ce n'est pas le cas pour beaucoup de paires de bases des tiges, en particulier pour les paires de bases des extrémités des tiges, ainsi que pour les paires formées par les résidus U6 et G18. En fin de simulation, en présence ou en absence d'ions magnésium, les trois premières paires de bases de la tige 1 et les deux premières de la tige 2 sont rompues, de même que les paires U6-G18 et U6'-G18' de la simulation en présence de magnésium et que la paire U6'-G18' de la simulation en absence de magnésium.

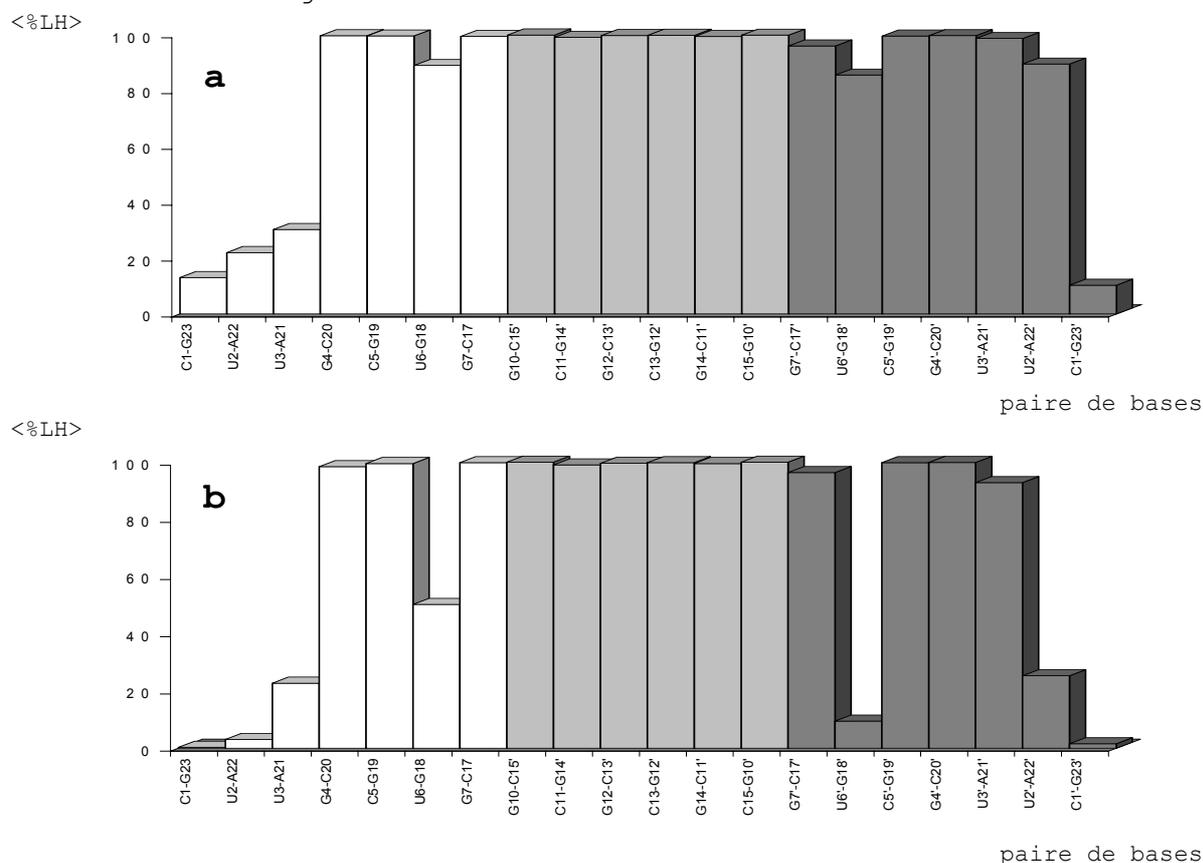


Figure 48 : Pourcentages moyens d'existence de l'appariement entre les paires de bases canoniques de la molécule sur l'ensemble de la simulation du dimère instable de SL1_{Lai} de Ennifar et al. (2001) **a)** en absence et **b)** en présence d'ions magnésium. En blanc sont représentées les paires de bases canoniques de la tige 1, en gris clair celles du groupe inter-boucles, en gris foncé celles de la tige 2.

Dans cette dernière simulation, la paire U6-G18, bien qu'encore présente en fin de simulation, fait montre d'une grande instabilité. La Figure 49 présente en vue stéréo les structures finales obtenues

pour les simulations de la structure du « kissing-complex » déterminée par Ennifar et al. (2001) respectivement en absence (en haut) et en présence (en bas) d'ions magnésiums. Cette figure permet de visualiser le déroulement des tiges avec ouverture des paires de bases des extrémités ainsi que le mouvement des adénines à la jonction des tiges-boucles partiellement tournées vers l'intérieur de la molécule.

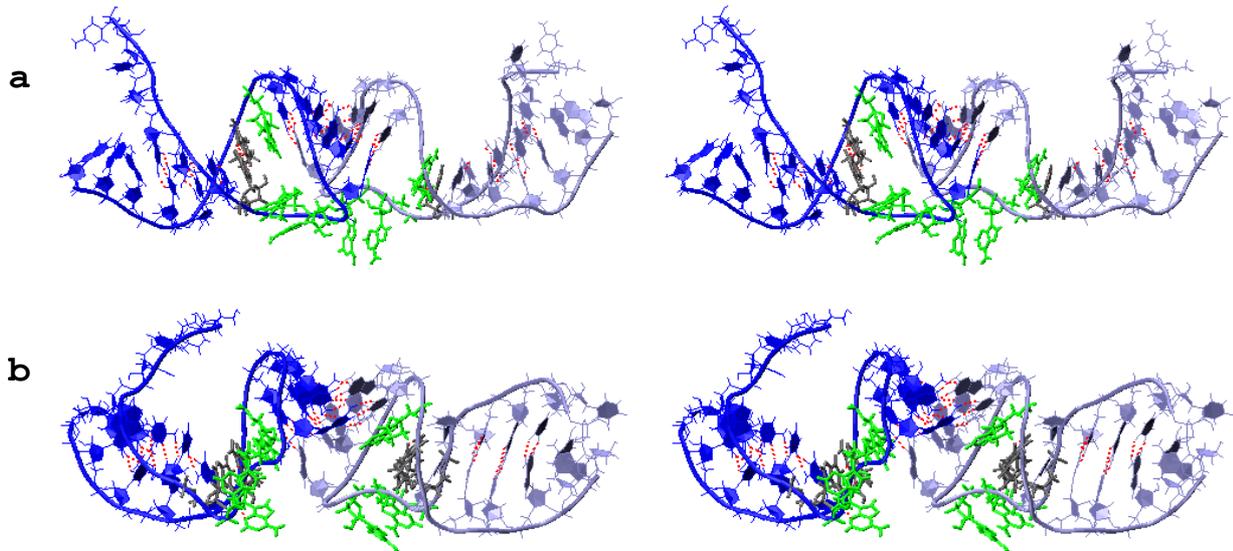


Figure 49 : Représentation en vue stéréo de la dernière configuration de la simulation du dimère instable de SL1_{Lai} de Ennifar et al. (2001) **a)** en absence et **b)** en présence d'ion magnésium.

→ Etude du dimère instable de SL1_{Ma1}

De même que pour l'isolat Lai, deux simulations ont été réalisées en absence et en présence d'ions magnésium. La simulation en présence d'ions magnésium requiert des conditions particulières du fait de la structure déterminée par cristallographie. En effet si l'on se réfère à l'article publié par Ennifar et al. (2001), le magnésium serait une condition essentielle à la stabilité de cette structure grâce à la formation, au niveau de l'interaction boucle-boucle, d'un complexe de magnésium hexahydraté dont les interactions avec les résidus des boucles seraient à même de pallier l'insuffisance de la séquence. La position de ce complexe dans la structure cristallographique est illustrée par la Figure 50. Nous avons donc pris soin, lorsque nous avons conservé les magnésiums du fichier PDB, de conserver également les six molécules d'eau de ce complexe.

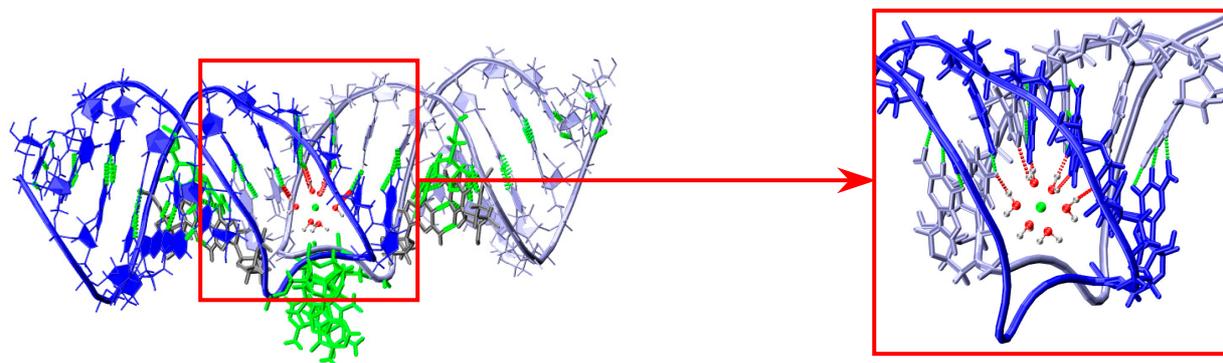


Figure 50 : Structure initiale du dimère instable de SL1_{Mal} de Ennifar et al. (2001). Le carré rouge met en évidence la présence au niveau de l'interaction boucle-boucle d'un complexe de magnésium hexahydraté.

Nous avons ainsi réalisé à partir de cette molécule une simulation de 5.7 nanosecondes avec 44 contre-ions Na⁺ dans une boîte d'eau parallélépipédique contenant 6918 molécule d'eau et une simulation de 1.7 nanosecondes avec 14 ions Mg²⁺ complétés par 16 ions Na⁺ dans une boîte d'eau parallélépipédique contenant 7451 molécules d'eau.

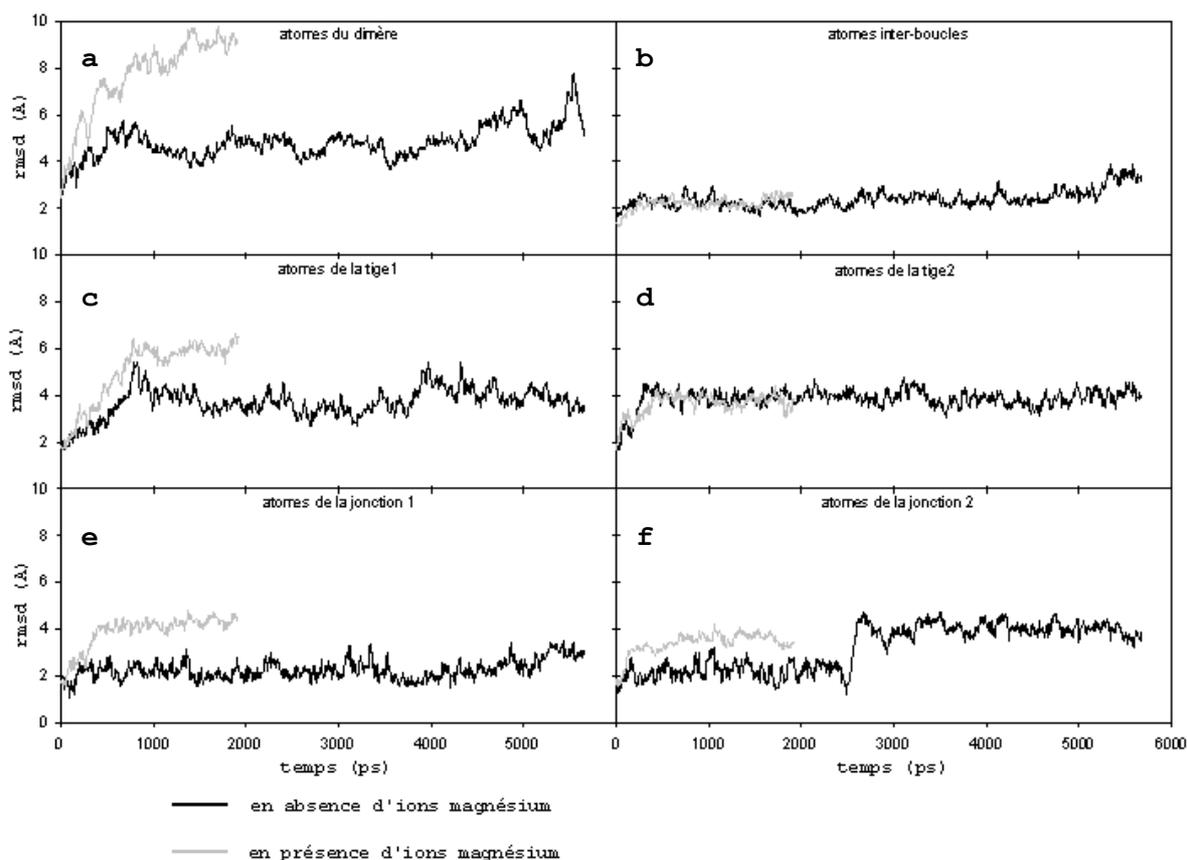


Figure 51 : Evolution du rmsd au cours de la simulation du dimère instable de SL1_{Mal} de Ennifar et al. (2001) en absence (en noire) et en présence (en gris) d'ions magnésium par rapport à la structure expérimentale sur les atomes : du dimère complet (a), du groupe inter-boucle (b), de la tige 1 (c), de la tige 2 (d), de la jonction 1 (e) et de la jonction 2 (f)

Ces simulations conduisent à des résultats similaires à ceux obtenus à partir de la séquence SL1_{Lai}. La Figure 51 reproduit les rmsd obtenus au cours de la simulation par rapport à la structure initiale.

On observe également pour cet isolat une évolution structurale très éloignée de la configuration d'origine avec une divergence principalement située dans les tiges des « kissing-complex » ainsi que, dans une moindre mesure, dans les jonctions entre tige et boucle. La hausse brusque de rmsd au niveau de la jonction 2 à 2500 ps de la simulation en absence d'ions magnésium correspond au passage des résidus A8' et G9' de la configuration rejetée vers l'extérieur à une configuration plus tournée vers l'intérieur de la molécule. Les configurations obtenues en fin de simulation sont reportées Figure 52 en haut pour la simulation en absence d'ions magnésium, en bas pour celle en présence d'ions magnésium.

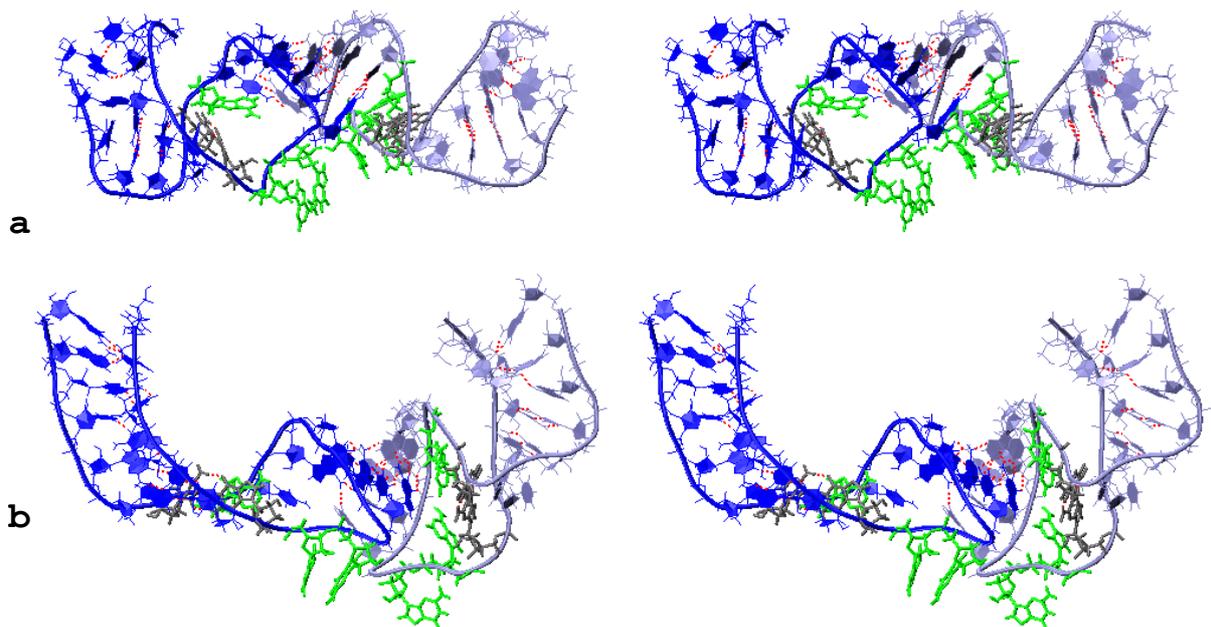


Figure 52 : Représentation en vue stéréo de la dernière configuration de la simulation du dimère instable de SL1_{Mal} de Ennifar et al. (2001) **a)** en absence et **b)** en présence d'ion magnésium.

Comme dans le paragraphe précédent, les simulations à partir des structures cristallographiques de l'isolat Mal conduisent à un déroulement des tiges suivi de l'ouverture des paires de bases intra-brins tandis que l'interaction boucle-boucle est maintenue. A noter cependant que cette interaction boucle-boucle présente un affaiblissement pendant la simulation, avec des ruptures momentanées

des liaisons hydrogènes entre les paires de bases, comme on peut le voir sur la Figure 53 décrivant les pourcentages d'existence moyens des liaisons hydrogène par paire de base au cours de la simulation, en particulier en ce qui concerne les liaisons hydrogène des paires A14-U11' et C15-G10'. Les trois paires de bases des extrémités des tiges sont ouvertes ou très instables en fin de simulation ainsi que les paires de bases U6-A18 et U6'-A18', en présence comme en absence de magnésium.

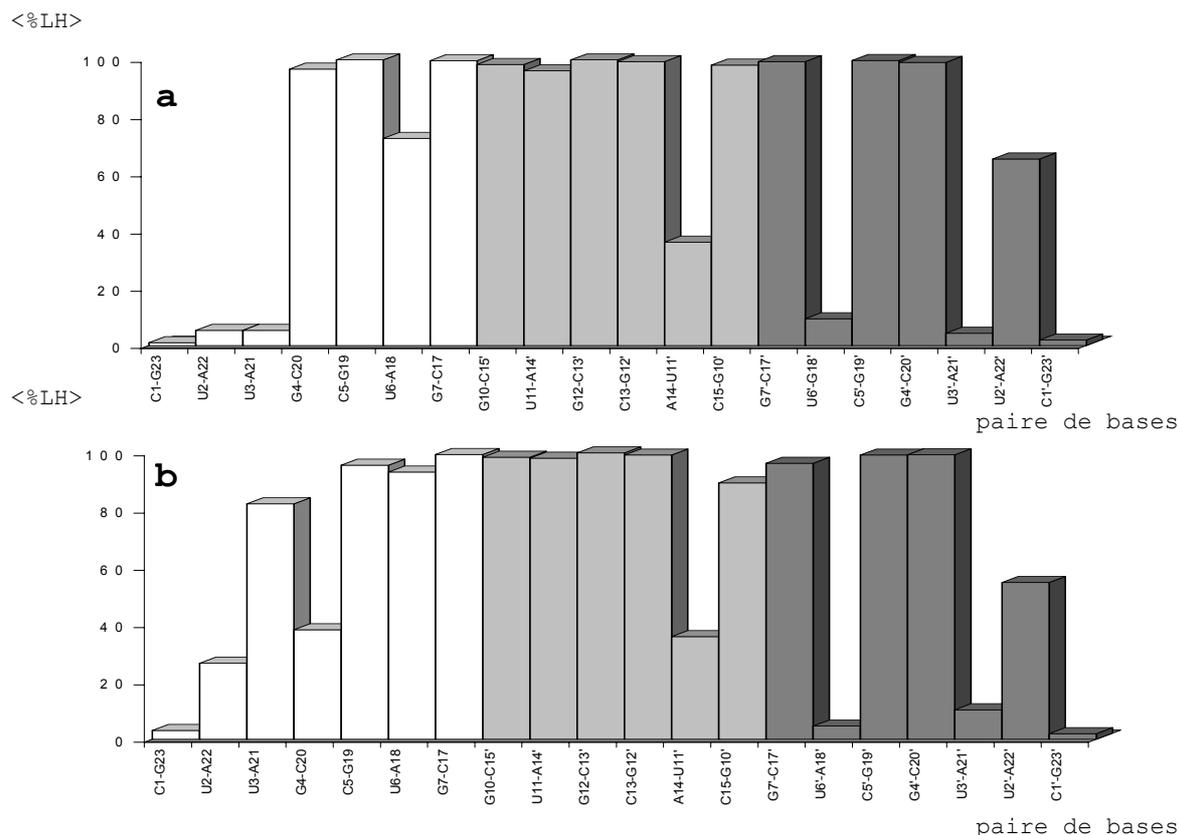


Figure 53 : Pourcentages moyens d'existence de l'appariement entre les paires de bases canoniques de la molécule sur l'ensemble de la simulation du dimère instable de SL1_{Mal} de Ennifar et al. (2001) **a**) en absence et **b**) en présence d'ions magnésium. blanc : Bptigel, gris clair : Bpinter-boucles, gris **foncé** : BPtige2.

I.a.ii. Etude de la structure RMN de Kieken à partir de la séquence SL1*

Kieken (2004) décrit deux familles de structures en « kissing-complex » qui diffèrent par la position des résidus à la jonction tige-boucle. De chacune de ces familles 1 et 2, nous avons extrait une structure caractéristique sur laquelle nous avons effectué 9.8 ns de simulation de dynamique moléculaire à 300 K.

La séquence SL1* étant un 24-mer et non plus un 23-mer comme les séquences étudiées précédemment, le résidu non apparié de l'extrémité 5' sera numéroté 0 (Figure 54).

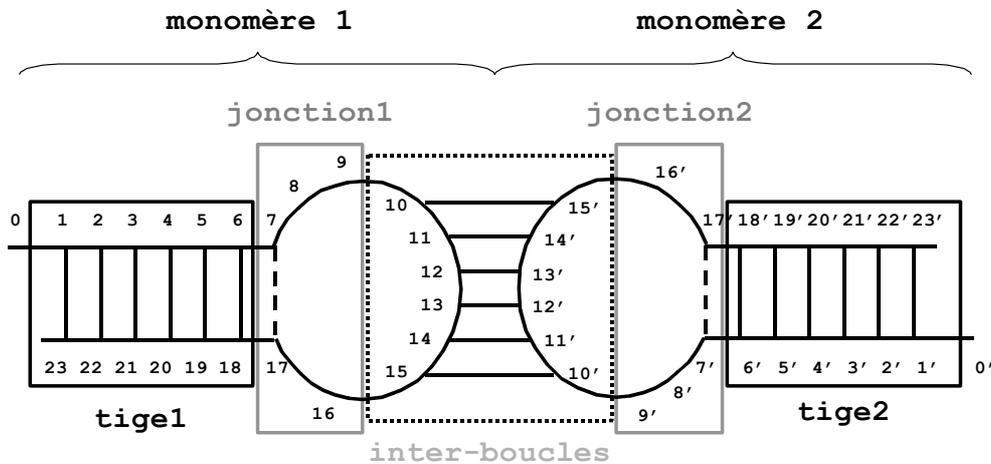


Figure 54 : Même groupes de définition que pour les séquences de 23-mer précédents (voir **Figure 36**). Le résidu isolé supplémentaire en début de brin a été dénommé 0 et 0' pour les monomères 1 et 2 respectivement.

→ Etude d'une structure caractéristique de la famille 1

Cette structure a été simulée à 300 K en présence de 46 contre-ions Na^+ dans une boîte d'eau parallélépipédique contenant 12638 molécules d'eau.

Au cours de la simulation, la structure conserve une structure en forme de « kissing-complex », c'est à dire que l'appariement des paires de bases dans les tiges ainsi que les interactions inter-boucles sont maintenues au cours du temps, comme l'illustre la Figure 55.

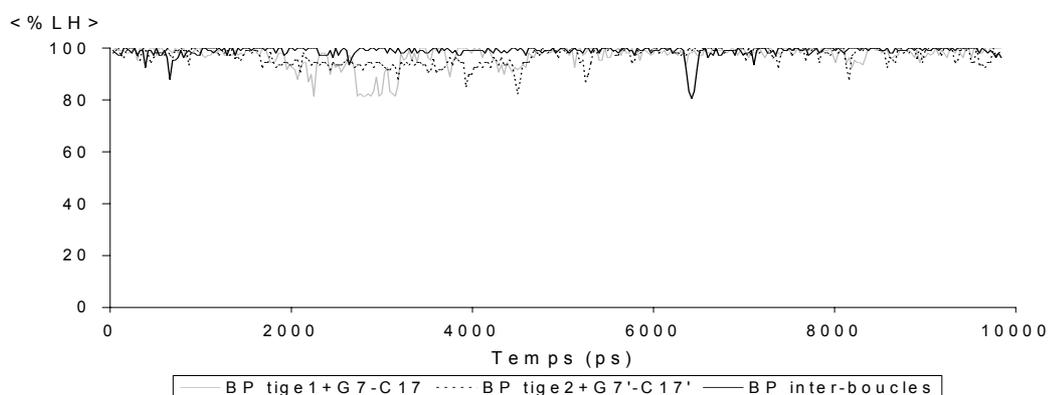


Figure 55 : Evolution du pourcentage moyen des liaisons hydrogène canoniques pour les résidus {1→7, 17→23} (en bleu), {1'→7', 17'→23'} (en violet) et {10→15, 10'→15'} (en noir) lors de la simulation de la structure de la famille 1 de Kieken (2004).

L'étude du rmsd en fonction du temps, reporté sur la Figure 56, montre cependant une légère déviation structurale de la molécule au cours du temps.

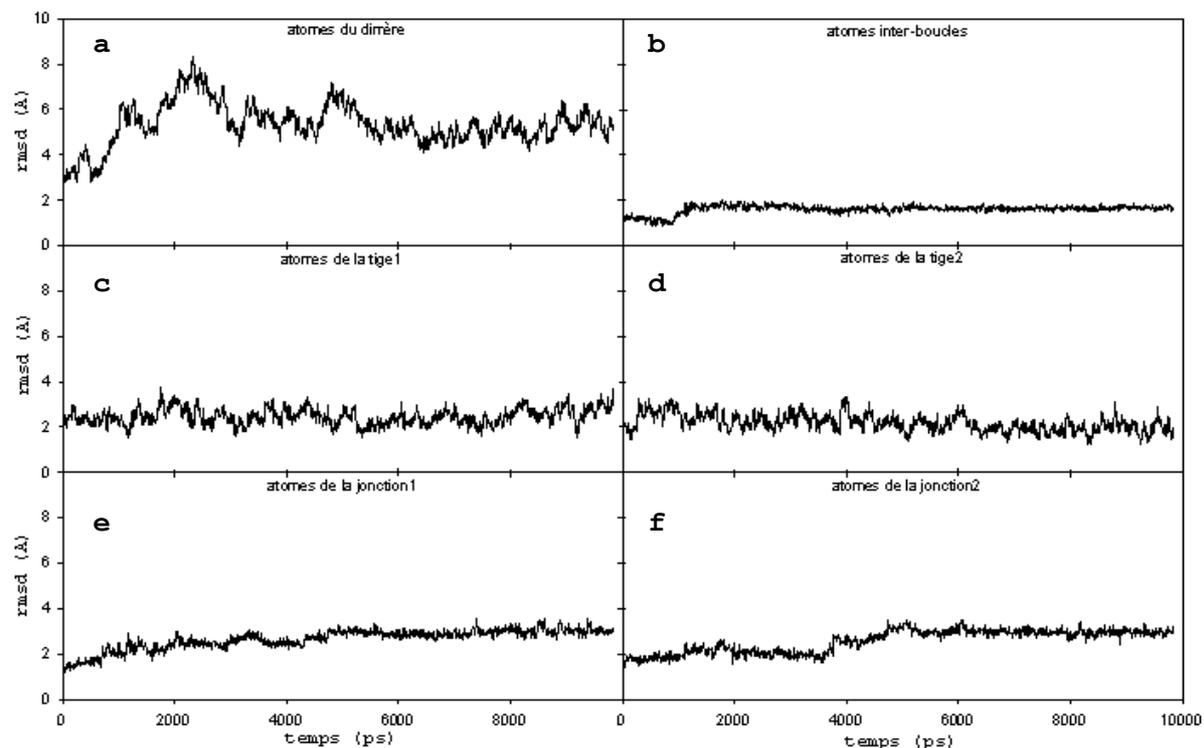


Figure 56 : Evolution du rmsd au cours de la simulation de la structure caractéristique de la famille 1 de Kieken (2004) par rapport à la structure expérimentale sur les atomes : du dimère complet (a), du groupe inter-boucle (b), de la tige 1 (c), de la tige 2 (d), de la jonction 1 (e) et de la jonction 2 (f)

Les plus grandes déviations structurales sont observées pour les jonctions avec un rmsd atteignant une valeur de 3 Å. Cependant cette valeur n'indique pas un grand changement conformationnel, résultat démenti par le rmsd global de la molécule atteignant jusqu'à une valeur de 8 Å aux alentours de 2.5 ns de simulation. Cette évolution conformationnelle est en réalité due à une ouverture de l'angle entre les axes des tiges au cours du temps. La Figure 57 présente l'évolution de cet angle au cours de la simulation, ainsi que la valeur de ce même angle dans la structure initiale. On voit que parti d'un angle entre les axes des tiges de l'ordre de 95°, celui-ci se stabilise aux alentours de 130° en fin de simulation.

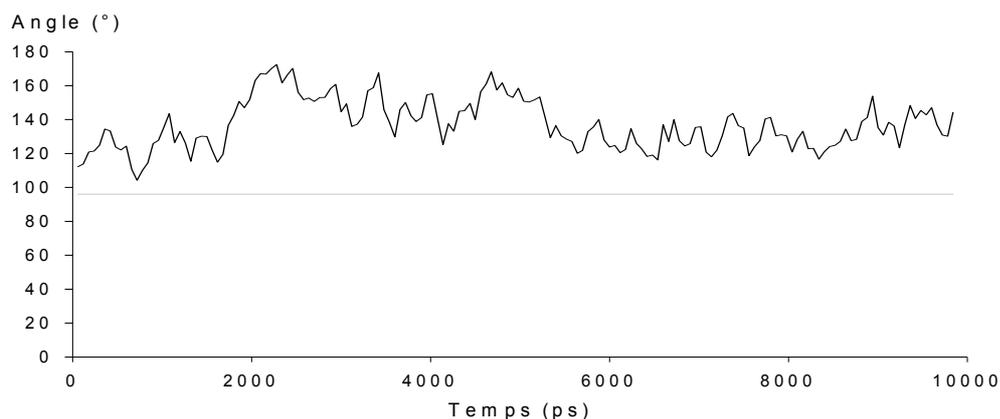


Figure 57 : Evolution de l'angle entre les axes des tiges au cours de la simulation de la structure caractéristique de la famille 1 de Kieken (2004).

On remarque sur ces deux figures la stabilisation, à partir de 5 ns de simulation, aussi bien des rmsd sur les différents groupes d'atomes par rapport à la structure initiale que de l'angle entre les axes des tiges.

La Figure 58 présente la structure moyenne obtenue sur la dernière nanoseconde de simulation, après stabilisation des valeurs de rmsd et d'angle. On observe sur cette figure que tiges et boucles ont conservé leurs appariements de paires de bases et que les adénines à la jonction ont conservé la configuration tournée vers l'intérieur de la molécule caractéristique de cette famille 1. Les appariements entre les adénines A9 et A16 fluctuent au cours de la simulation.

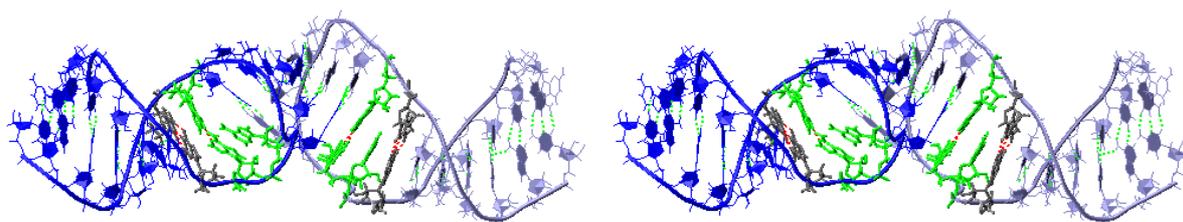


Figure 58 : Représentation en vue stéréo de la conformation finale de la simulation de la structure caractéristique de la famille 1 de Kieken (2004).

→ Etude d'une structure caractéristique de la famille 2

Nous avons effectué une simulation de dynamique moléculaire à partir de cette structure en présence de 46 contre-ions Na^+ à 300 K dans une boîte d'eau parallélépipédique contenant 11807 molécules d'eau.

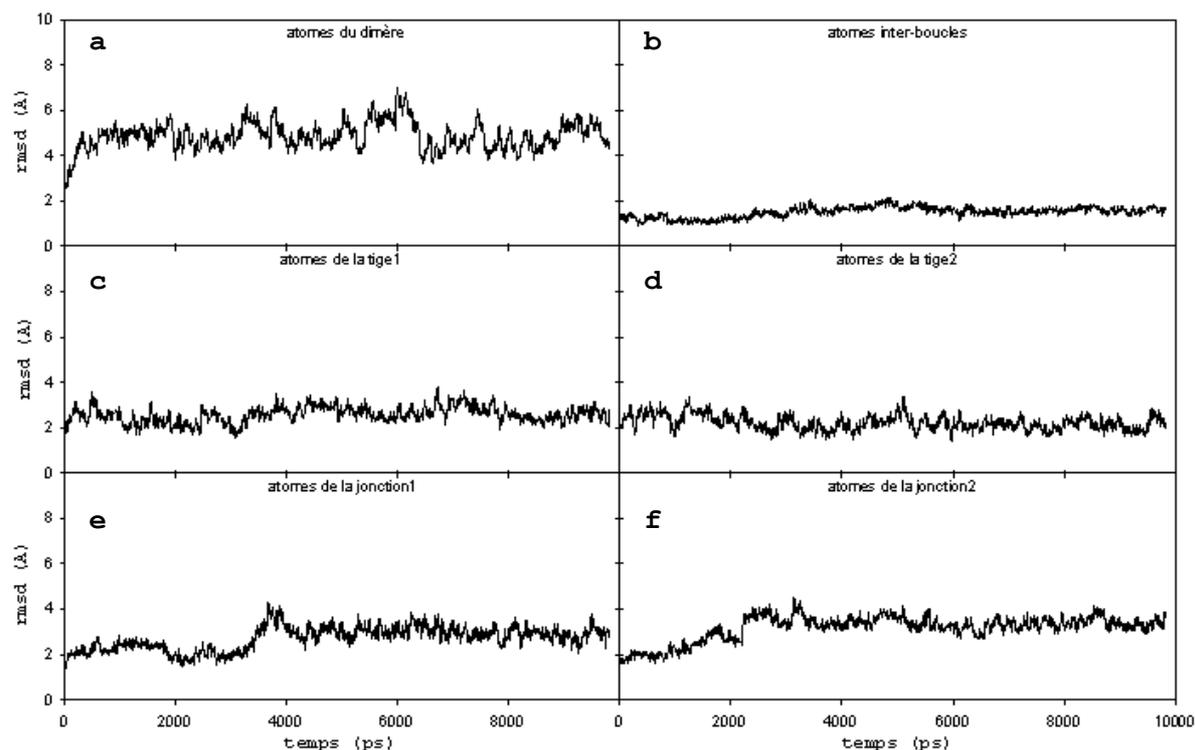


Figure 59 : Evolution du rmsd au cours de la simulation de la structure caractéristique de la famille 2 de Kieken (2004) par rapport à la structure expérimentale sur les atomes : du dimère complet (a), du groupe inter-boucle (b), de la tige 1 (c), de la tige 2 (d), de la jonction 1 (e) et de la jonction 2 (f)

La simulation présente globalement les mêmes résultats que précédemment, à savoir que la structure reste en forme de « kissing-complex » tout au long de la simulation. Cependant le comportement même de la molécule est différent dans cette famille sur plusieurs points et en particulier, la molécule semble beaucoup plus stable avec cette configuration de départ qu'avec la précédente.

En effet, on voit sur la Figure 59 qui présente l'évolution du rmsd de la molécule au cours du temps par rapport à sa structure initiale, que la molécule reste globalement beaucoup plus proche de sa structure d'origine, avec un rmsd maximal obtenu d'environ 6Å. De fait, l'angle entre les axes des tiges ne subit pas cette fois d'ouverture importante comme on peut le vérifier sur la Figure 60 présentant cet axe au cours du temps de simulation.

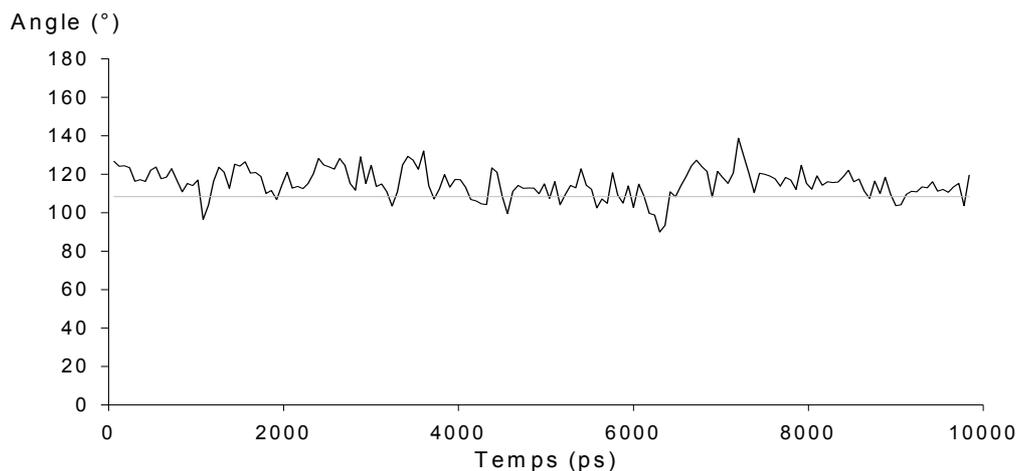


Figure 60 : Evolution de l'angle entre les axes des tiges au cours de la simulation de la structure caractéristique de la famille 2 de Kieken (2004).

La structure de départ présentait cette fois un angle de l'ordre de 110° entre les axes des tiges, et tout au long de la simulation, l'angle entre les axes des tiges fluctue autour de cette valeur. De même que pour la structure issue de la famille 1, la plus grande déviation structurale est obtenue pour les jonctions tige-boucle avec des valeurs de rmsd atteignant 3-4Å. L'une des raisons de cette déviation structurale des jonctions est la rupture d'une des paires G7-C17 à la jonction tige-boucle. La Figure 61 présente l'évolution des liaisons hydrogène entre les résidus G7 et C17 et de la phase des sucres de ces résidus pour les deux tiges. Pour la tige 1, les résidus G7 et C17 se désappariaient au bout de 2 ns de simulation et restent désappariés pendant les 7.8 ns de simulation suivant. Les trois liaisons hydrogène se rompent simultanément. On constate de surcroît que cette rupture de la paire de base est suivie à moins d'1 ns d'un changement de conformation du sucre de C17 qui passe de la conformation C3'-endo à une conformation C2'-endo. Aucun de ces deux phénomènes ne se produit sur la tige 2.

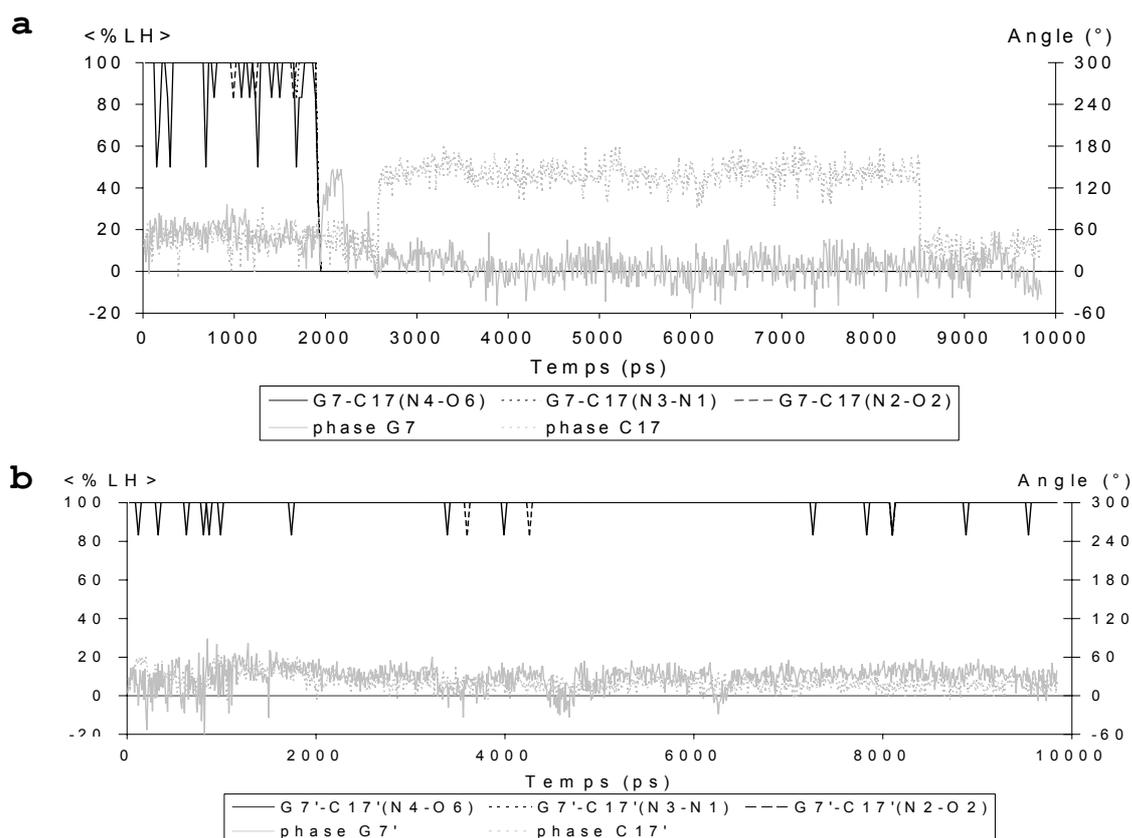


Figure 61 : Mise en parallèle de l'évolution de la conformation du sucre avec les pourcentages de présence des liaisons hydrogène Watson-Crick entre les résidus G7 et C17 du monomère 1 (a) et G7' et C17' du monomère 2 (b) de la simulation de la structure caractéristique de la famille 2 de Kieken (2004).

Les adénines à la jonction des boucles présentent sensiblement le même comportement sur les deux tiges en dépit du fait que l'une des paires G7-C17 soit désappariée et pas l'autre. Leur configuration reste proche de celle de la structure d'origine. Les résidus A9 et A9' sont partiellement empilés ensemble à l'intérieur de l'espace créé par l'interaction boucle-boucle, ce qui correspond à leur position originelle. Ils forment au cours de la simulation plusieurs liaisons hydrogène temporaires avec les résidus des boucles. Les deux résidus A8 et A8' sont eux respectivement empilés avec les résidus G7' et G7, c'est à dire que chaque résidu A8 d'un brin est empilé avec le résidu G7 de l'autre brin, que celui-ci soit ou non apparié. Ceci diffère un peu de la conformation d'origine, où l'interaction entre le résidu A8 d'un brin et le résidu C17 de l'autre brin était établie par une liaison hydrogène et non par un empilement. Les résidus A16 et A16' présentent un comportement différent en fin de simulation : le résidu A16 est empilé avec le

résidu G15 de la séquence auto-complémentaire tandis que le résidu A16' tend à se tourner vers l'extérieur de la molécule. Cette configuration de la molécule et de la jonction en particulier est exposée sur la Figure 62.

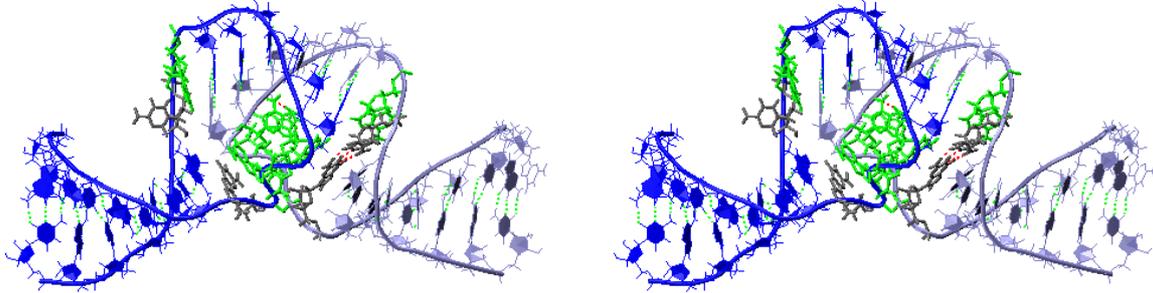


Figure 62 : Représentation en vue stéréo de la conformation finale de la simulation de la structure caractéristique de la famille 2 de Kieken (2004).

I.a.iii. Discussion-conclusion

Nous nous sommes efforcés dans cette partie de déterminer lesquelles des structures parues dans la littérature présentaient une forme stable de « kissing-complex » en solution à 300 K. Dans un souci de contrôle des résultats présentés, la simulation effectuée à partir de la structure RMN de Mujeeb et al. (1998) avec la séquence mutée, ainsi que les simulations réalisées à partir des deux structures cristallographiques de Ennifar et al. (2001), en présence comme en absence d'ions magnésium, ont été exécutées une seconde fois dans les mêmes conditions sur des temps plus courts. Les molécules ont adopté, lors de ces simulations, des comportements similaires à ceux exposés dans ce travail. De toutes les structures que nous avons simulées deux présentent un grand manque de stabilité en solution. Il s'agit des deux structures de « kissing-complex » déterminées par cristallographie et correspondant aux deux isolats Lai et Mal de la tige-boucle SL1. Le plus grand changement est la destruction des tiges. On remarque cependant des appariements inter-boucles plus forts pour la structure de l'isolat Lai que pour la structure de l'isolat Mal, en accord avec les résultats d'électrophorèse de Aci et al (2004a). Cette instabilité des tiges dans les structures cristallographiques est sans doute due à des interactions inter-dimère existant dans le cristal.

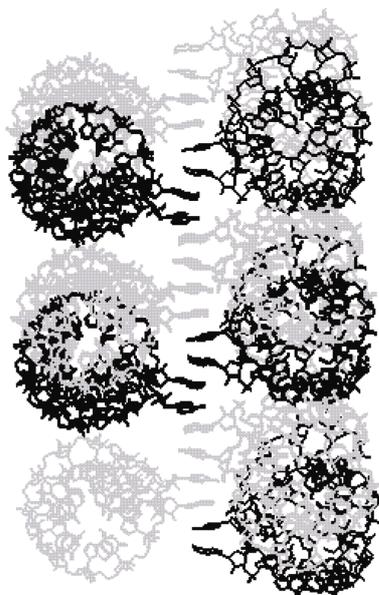


Figure 63 : Empilement structural dans la maille cristalline des dimères instables déterminés par Ennifar et al. (2001) pour les isolats Mal et Lai de SL1.

En effet, les purines 8 et 9 à la jonction tige-boucle en particulier présentent une conformation singulière qui s'explique lorsque l'on examine la maille cristalline (Figure 63). Ces résidus sont empilés et rejetés à l'extérieur de sorte qu'il reste une place suffisante, entre les deux paires de résidus d'un même dimère, pour intercaler une paire de résidus 8-9 d'un autre dimère. Cette configuration conduit dans la maille cristalline à la formation d'une longue chaîne de purines empilées qui stabilise la structure. De telles interactions n'existent pas en solution. En effet, il a été montré qu'aux concentrations RMN l'association se limitait à la formation de dimère (Kieken 2004). L'absence de ces interactions peut expliquer la déstabilisation de la structure cristallographique. Nous n'avons pas observé d'influence du magnésium à notre échelle de temps.

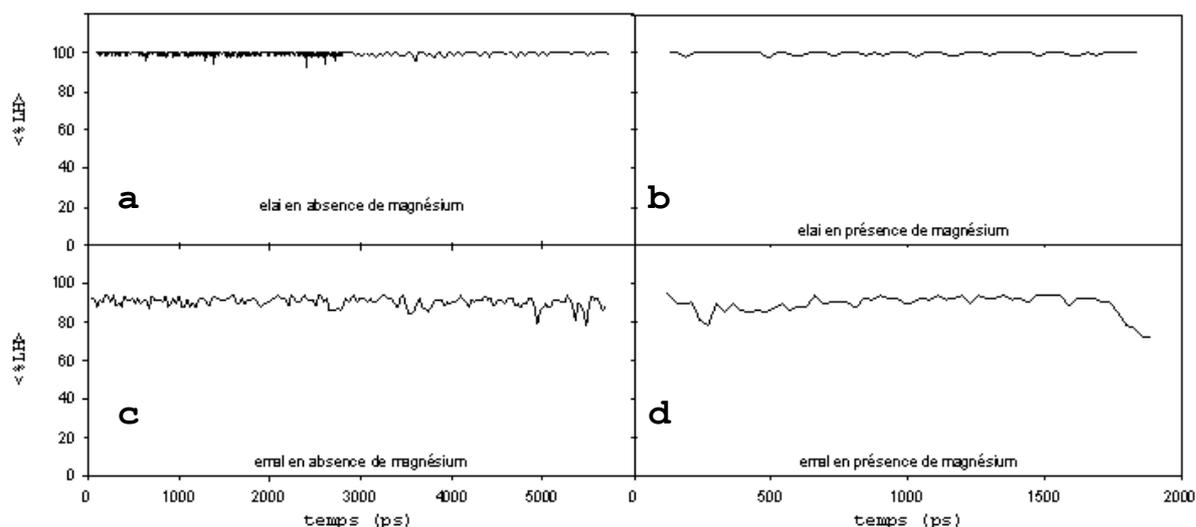


Figure 64 : Evolution des pourcentages d'existences des liaisons hydrogène Watson-Crick inter-boucle pendant le temps de simulation des dimères instables déterminées par Ennifar et al. (2001) : **a)** Lai en absence d'ions magnésium **b)** Lai en présence d'ions magnésium **c)** Mal en absence d'ions magnésium **d)** Mal en présence d'ions magnésium

Ainsi que le présente la Figure 64, l'interaction boucle-boucle plus faible de l'isolat Mal n'est ni compensée ni atténuée par la présence ou l'absence d'ions magnésium. Ceci contredit les résultats d'Ennifar et al. (2001) et de Jossinet et al. (1999) mais est en accord avec les expériences de Dardel et al. (1998) et de Aci et al. (2004a).

A l'opposé, les structures déterminées par RMN restent toutes plus stables que les structures cristallographiques. La conformation publiée par Mujeeb et al. (1998) reste globalement stable sur les 20 ns de simulation mais subit des réarrangements locaux importants au cours de cette période. Ces réarrangements sont principalement localisés au niveau de la jonction et ne peuvent être attribués à la mutation subie par la séquence du fait qu'une simulation effectuée sur cette même structure mais avec la séquence sauvage de la tige-boucle SL1_{Lai} conduit à des résultats analogues. La modification majeure est la formation spontanée et durable de la paire G7-C17 au bout de quelques nanosecondes de simulation, indépendamment de la séquence de la tige. Les axes des tiges conservent cependant leurs orientations. Cette paire de bases, qui correspond à la paire G271-C281 dans la séquence totale de l'ARN génomique a été sujette à plusieurs polémiques quant à son appariement. Il est intéressant de remarquer que l'on a observé le phénomène inverse avec la structure de la famille 2 de Kieken (2004). Contrairement à la structure de Mujeeb et al. (1998), les résidus G7 et C17 sont appariés dans cette structure initiale. Or on constate au bout de quelques nanosecondes de simulation, une rupture spontanée et durable des trois liaisons hydrogène Watson-Crick responsables de cet appariement. Ces observations nous incitent à penser que la paire de bases G7-C17 n'est pas stable en solution et que son passage entre les états ouvert et fermé forme une transition réversible capable d'intervenir sur une échelle de temps de quelques nanosecondes. Les études RMN menées par Kieken (2004) sur ces résidus sont en accords avec cette interprétation. En effet, elles ne mettent en évidence l'existence de liaisons entre les résidus G7 et C17 qu'au cours d'expériences HSQC en 2 dimensions et non par des expériences NOESY comme pour les autres paires G-C. Cela traduit une grande dynamique de la part de cette paire de bases avec sans doute des ruptures sporadiques de ses

liaisons hydrogène. Ces résultats sont également en bonne adéquation avec les expériences de Lodmell et al. (2001), suggérant que l'appariement de ces deux résidus n'était pas indispensable à la formation du « kissing-complex » dans le cas de l'isolat Lai. Nous avons par ailleurs montré par une étude thermodynamique que l'énergie libre au cours de la transition reste très faible (voir §IV.c).

Les trois structures déterminées pour le « kissing-complex » présentent ainsi, pour Ennifar et al. (2001) un appariement G7-C17 très fort, pour Mujeeb et al. (1998) des résidus G7 et C17 désappariés et pour Kieken (2004) un appariement G7-C17 relativement lâche. Ces différences traduisent l'existence d'une distribution d'états métastables pour la jonction de la tige-boucle. Il semble que piéger l'un ou l'autre de ces états dépende fortement des conditions expérimentales utilisées. Les mutations imposées à la séquence de la tige lors des expériences menées par Mujeeb et al. (1998) ont pu influencer sur la conformation adoptée par la jonction à la tige-boucle du dimère déterminé, de même que la forte concentration en ions magnésium nécessaire à la cristallisation dans le cas d'Ennifar et al. (2001).

Un autre foyer de divergences entre ces différentes structures est la position des purines à la jonction tige-boucle ainsi que les liaisons hydrogène qu'elles sont susceptibles de former avec leur environnement. D'autres modèles de « kissing-complex » ont été déterminés par modélisation à partir de données biochimiques (Jossinet et al. 1999; Paillart et al. 1997; Pattabiraman et al. 2002) dans lesquels le positionnement des purines a été longuement explicité. Dans tous ces modèles, les tiges et le duplexe inter-boucles sont alignés sur un seul axe dans une conformation proche d'une hélice A. Le modèle de Paillart et al. (1997) propose deux liaisons hydrogène intra-moléculaires entre les résidus A8 et A16 ($N6H(A8) \rightarrow N7(A16)$ et $N6H(A16) \rightarrow N1(A8)$) et deux liaisons inter-moléculaires entre le résidu G9 d'un monomère et le résidu G12 de l'autre ($N1H(G9) \rightarrow O6(G12)$ et $N2H(G9) \rightarrow N7(G12)$). Dans le modèle de Jossinet et al. (1999), les deux liaisons entre A8 et A16 du modèle précédent sont remplacées par $C2H(A8) \rightarrow N7(A16)$ et $N6H(A16) \rightarrow N3(A8)$. Le modèle publié par Pattabiraman et al. (2002) est inspiré de celui de

Paillart et al. (1997) mais les liaisons hydrogène entre les résidus 9 et 12 sont modifiées, pour l'isolat Lai par les liaisons N6H(A9)→O2(G12) et N4H(C13)→N1(A9) et pour l'isolat Mal par N2H(G9)→O6(G12) et N1H(G9)→N7(G12). Pourtant aucune de ces liaisons n'est observée dans les structures expérimentales parues dans la littérature. Pattabiraman et al. (2002) constatent du reste que, sur une courte simulation de 400 ps, la paire de bases formée entre A8 et A16 est instable. Sur l'ensemble de nos simulations, deux seulement présentent, mais de façon très éphémère, des appariements intramoléculaires A8-A16 : la simulation de la séquence de Mujeeb et al. (1998) de séquence SL1_{Lai}Δ avec une liaison N6H(A16')→N1(A8') et la simulation de la structure caractéristique de la famille 1 de Kieken (2004) avec les liaisons N6H(A16)→N7(A8), N6H(A8)→N1(A16), N6H(A16')→N7(A8') et N6H(A8')→N1(A16'). Des liaisons A9-A16 stables sont en revanche mises en évidence dans plusieurs de nos simulations, notamment la liaison inter-moléculaire N6H(A9)→N3(A16') dans la simulation à partir de la structure de Mujeeb et al. (1998) de séquence SL1_{Lai}Δ, les liaisons inter-moléculaires N6H(A9)→N3(A16') et N6H(A9')→N3(A16) dans la simulation à partir de la structure de Mujeeb et al. (1998) de séquence SL1_{Lai} et les liaisons intramoléculaires N6H(A16)→N7(A9), N6H(A9)→N1(A16), N6H(A16')→N7(A9') et N6H(A9')→N1(A16') dans la simulation à partir de la structure caractéristique de la famille 2 de Kieken (2004).

Reblova et al. (2003) et Beaurain et al. (2003) ont également réalisé des simulations de dynamique moléculaire à partir des structures RMN de Mujeeb et al. (1998) et cristallographique de Ennifar et al. (2001) et aboutissent à des conclusions opposées. Dans aucune de ces deux publications n'est donné ni un descriptif précis du comportement des tiges ni des liaisons hydrogène formées par les purines à la jonction tige-boucle. Toutes deux font néanmoins des observations similaires sur le comportement de la structure RMN de Mujeeb et al. (1998) qui se révèlent en bonne adéquation avec nos résultats. En effet il apparaît dans leurs deux simulations que la modification majeure intervenant dans cette structure est d'une part l'ouverture de l'angle entre les axes des tiges et d'autre part, le fait que les résidus G7 et C17 deviennent coplanaires. En revanche, aucune de ces deux équipes ne voit se

former l'appariement de G7 et C17. Dans les deux cas, les angles dièdres et les distances entre les atomes, dont certaines valeurs sont très éloignées de la normale dans la structure de Mujeeb et al. (1998), reprennent des valeurs standards dès les premières centaines de picosecondes de la simulation sans modifier la structure initiale de façon importante. Les résultats présentés sur les structures de Ennifar et al. (2001) sont en revanche très différents. Reblova et al. (2003) estiment les structures cristallographiques d'une grande stabilité. Les structures moyennes publiées sur leurs simulations montrent en effet pour les deux isolats des structures proches de la structure expérimentales, avec toutefois des positions différentes pour les purines à la jonction. Dans le cas de l'isolat Lai, les quatre adénines A8, A9, A8' et A9' forment par empilement les unes avec les autres une tige supplémentaire. Dans le cas de l'isolat Mal, les purines 8 et 9 restent partiellement tournées vers l'extérieur. Pour les deux structures, les adénines 16 sont empilées entre les résidus C15 et C17 et toutes les paires de bases des tiges restent appariées, sauf l'extrémité de la tige de l'un des monomères de l'isolat Lai. Peut-être est-ce là une amorce de la déstabilisation des tiges que nous avons observées avec la structure cristallographique, car dans nos simulations à partir des structures RMN les paires de bases des extrémités restent appariées même après plus de 20 ns de simulation. Les conclusions données par Beaurain et al. (2003) sont en meilleur accord avec nos résultats, la structure de Ennifar et al. (2001) leur paraissant moins stable en solution que la structure de Mujeeb et al. (1998). Ainsi ils observent au cours de leur simulation de la structure d'Ennifar et al. (2001) un enfouissement des adénines 8 et 9 à l'intérieur de la molécule, total pour le monomère 1 et partiel pour le monomère 2. Ils décrivent par contre une ouverture des paires de bases G10-C15' et G7'-C17', nécessaire au repositionnement des adénines, qui ne se produit pas dans nos simulations. Nous avons alors réalisé 3 simulations de vérification à partir de la structure cristallographique de Ennifar et al. (2001) de SL1_{Lai}. Dans l'une de ces simulations nous avons essayé de nous rapprocher au maximum des conditions décrites par Reblova et al. (2003) dans leur article. Nous observons une ouverture des deux premières paires de bases des

extrémités sur l'un des monomères dès 600 ps de simulation. Une autre simulation de 2 ns a été effectuée en gardant tous les atomes présents dans le fichier de coordonnées publié par Ennifar et al. (2001), c'est à dire que l'on a conservé non seulement la molécule et les atomes de magnésium mais également toutes les molécules d'eau contenues dans ce fichier. Au bout de 2 ns de simulation, les tiges se désenroulent et les deux premières paires de bases des tiges ainsi que la paire U6-G18 sont rompues. Les adénines 8 et 9 à la jonction tige-boucle ne sont plus que partiellement tournées vers l'extérieur. Une dernière simulation de 10 ns à 300 K a été exécutée en imposant des contraintes de distances sur toutes les liaisons hydrogène canoniques du « kissing-complex ». Au bout des 10 ns, les deux monomères présentent des configurations très différentes. Le monomère 1 est complètement déstructuré, sa tige est déroulée et la plupart des paires de bases sont rompues bien que les résidus restent proches du fait des contraintes imposées. Les adénines 8 et 9 sont tournées vers l'intérieur de la molécule. Le monomère 2 lui a gardé une conformation proche de la structure initiale, avec des résidus appariés dans les tiges et des adénines 8' et 9' à la jonction tournées vers l'extérieur. Il semble que la déstructuration de la tige 1 soit due aux contraintes générées par les adénines de la jonction pour se tourner vers l'intérieur de la molécule sans écarter les brins des tiges, les contraintes que nous avons imposées ne le permettant pas. Ceci nous confirme dans l'opinion que la position des adénines à la jonction dans la structure de Ennifar et al. (2001) impose des contraintes trop fortes à la molécule pour lui permettre d'être stable en solution.

De toutes les structures dont nous sommes partis pour effectuer ces simulations de dynamique moléculaire, la plus stable nous semble être la structure caractéristique de la famille 2 de Kieken (2004). En effet, elle conserve une structure en « kissing-complex » pendant toute la durée de la simulation avec des boucles bien appariées ainsi que les 6 premières paires de bases des tiges. De plus, l'ouverture de la paire de bases G7-C17 dans un monomère se produit sans qu'il y ait de réorganisation majeure au niveau des adénines à la jonction tige-boucle qui conservent une conformation très proche de la structure d'origine, ni de réorientation au niveau des axes

des tiges. Ce résultat conforte les conclusions de Kieken (2004), qui considère également la famille 2 comme la structure la plus probable du « kissing-complex » de SL1_{Lai}. Kieken a envisagé deux familles de structures pour la tige-boucle SL1_{Lai} car les données RMN ne permettaient pas de distinguer entre deux attributions différentes pour les protons H2 des adénines A8 et A9. Cependant la famille 2 lui paraît plus susceptible de représenter la structure finale du dimère instable, d'une part parce que cette famille ne présente aucune violation sur les angles dièdres ni sur les contraintes NOE alors que la famille 1 en présente toujours au minimum 4, d'autre part parce que les adénines des jonctions sont impliquées dans plus de liaisons hydrogène que celles de la famille 1. De plus, l'énergie totale de la famille 2 se révèle plus basse que celle de la famille 1.

IV.b.Etude du dimère stable de SL1

Une fois effectuée l'étude de la structure métastable du dimère de SL1, nous avons entrepris de contrôler la stabilité en solution des structures publiées sur le dimère stable de la séquence SL1.

Nous donnons sur la Figure 65 les 5 groupes d'atomes utilisés lors des analyses ainsi que la numérotation des résidus.

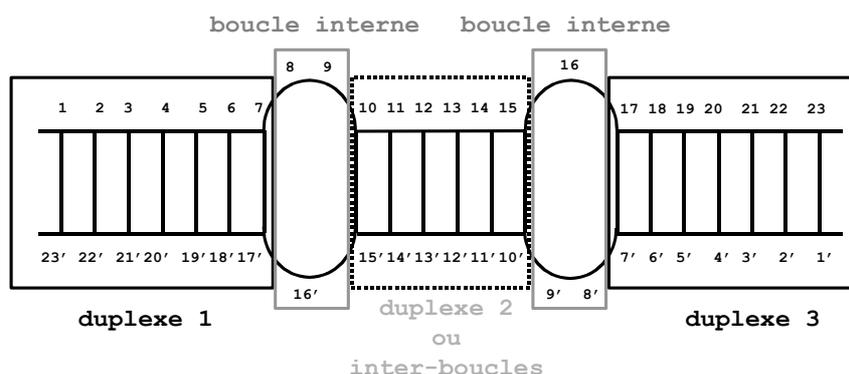


Figure 65 : Définition des cinq groupes de résidus utilisés lors des analyses sur les simulations du dimère stable. On appellera duplexe 1, duplexe 2 et duplexe 3 les groupes composés des résidus {1→6 ; 18'→23'}, {10→15 ; 10'→15'} et {18→23 ; 1'→6'} respectivement, boucle interne 1 et boucle interne 2 les groupes composés des résidus {8, 9, 16'} et {16, 8', 9'} respectivement.

IV.b.i. Etude de la structure RMN de Girard et al. à partir de la séquence SL1_{Lai}

Girard et al. (1999) ont déterminé par RMN la structure du dimère étendu pour la séquence SL1_{Lai} et publié une famille de 15 configurations correspondant à cette molécule. Afin de déterminer laquelle de ces structures nous allons utiliser comme structure de départ pour nos simulations, nous avons calculé la structure moyenne de cette famille puis mesuré les rmsd entre ces 15 structures et la structure moyenne. La Figure 66 présente les résultats de ces mesures.

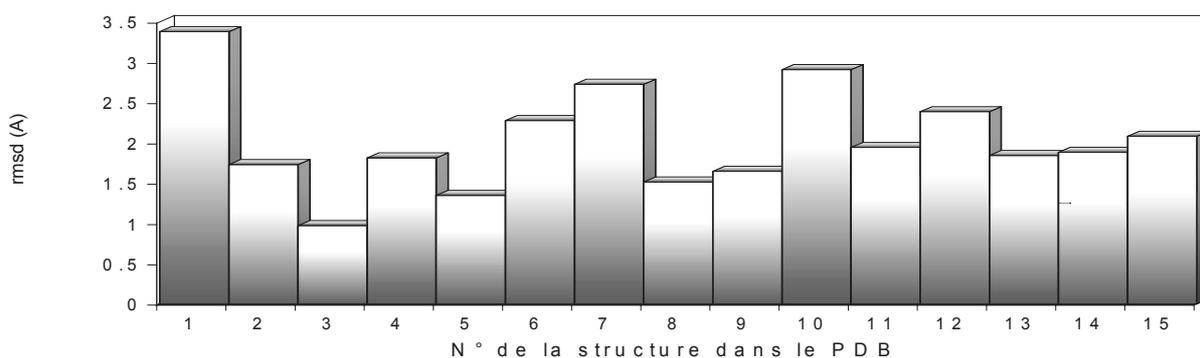


Figure 66 : Diagramme des valeurs de rmsd de chacune des structures de la famille déterminée pour le dimère étendu de SL1_{Lai} par Girard et al. (1999) par rapport à la structure moyenne de cette famille.

Nous avons choisi comme structure de départ de nos simulations la configuration la plus proche de la structure moyenne, à savoir la configuration n°3.

Nous avons réalisé une dynamique moléculaire de 8 ns à 300 K à partir de cette structure en présence de 44 contre-ions Na⁺ dans une boîte d'eau parallélépipédique contenant 6714 molécules d'eau.

Au cours de la simulation, le dimère conserve ses motifs structuraux principaux, c'est à dire les trois duplexes en forme A et les deux boucles internes. Cependant, ces motifs eux-mêmes subissent des modifications, en particulier le duplexe central (ou duplexe 2) ainsi que les deux boucles internes qui présentent des rmsd moyens sur la dernière nanoseconde de simulation de 3 Å pour le duplexe 2 et de 2.6 Å pour chacune des boucles internes. Les deux duplexes des extrémités restent très proches de la structure d'origine avec des rmsd moyens sur la dernière nanoseconde de simulation de 1.8 Å et

1.6 Å pour le duplexe 1 et le duplexe 3 respectivement. L'évolution de ces rmsd au cours du temps est représenté sur la Figure 67.

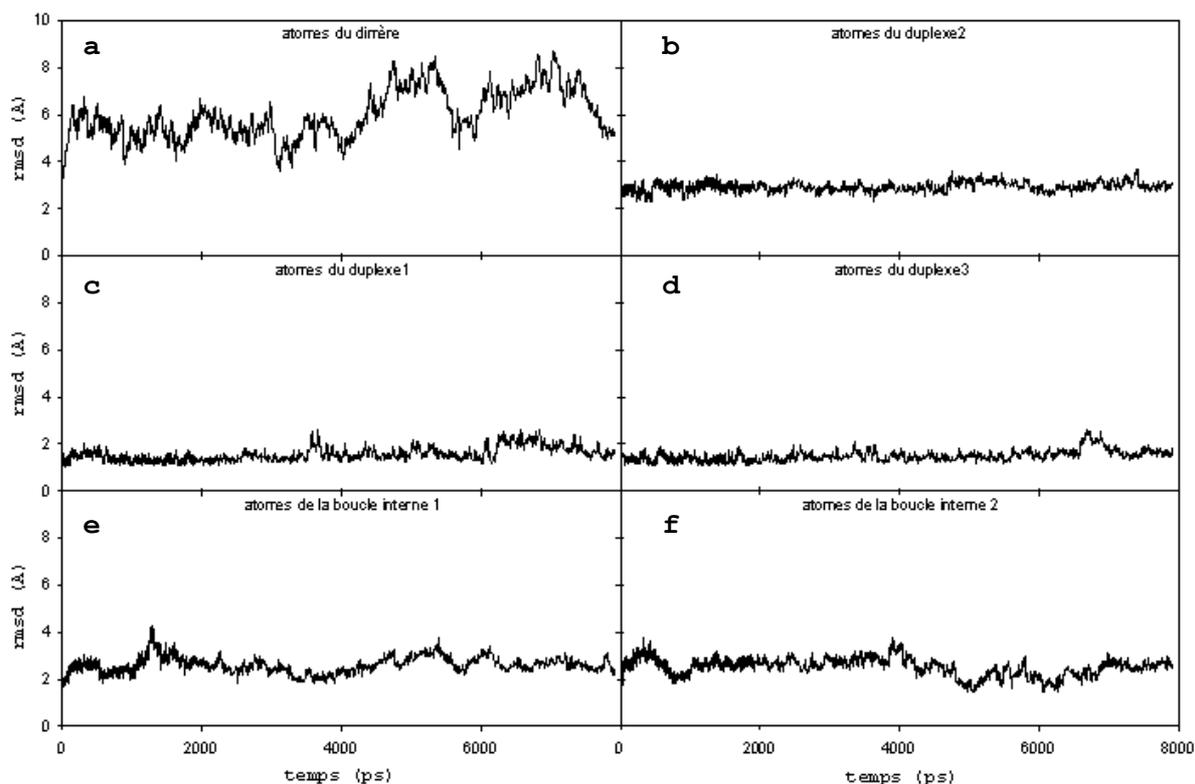


Figure 67 : Evolution du rmsd au cours de la simulation du dimère stable de Girard et al. (1999) par rapport à la structure expérimentale sur les atomes : du dimère complet (a), du duplexe 2 (b), du duplexe 1 (c), du duplexe 3 (d), de la boucle interne 1 (e) et de la boucle interne 2 (f)

| dimère | duplexe1 | boucle interne 1 | duplexe2 | boucle interne 2 | duplexe3 |
|--------|----------|------------------|----------|------------------|----------|
| 6.7 | 1.8 | 2.6 | 3.0 | 2.6 | 1.8 |

Tableau 3 : Valeurs moyennes des rmsd sur la dernière nanoseconde de simulation du dimère stable de Girard et al.

On remarque que les déviations structurales des éléments du dimère sont relativement faibles, avec des valeurs de rmsd maximum autour de 3 Å. Cependant la conformation globale du dimère évolue considérablement, jusqu'à atteindre des valeurs de rmsd de 7-8Å. Cette variation par rapport à la structure d'origine est essentiellement due à une courbure de la molécule au cours du temps. La Figure 68 montre l'évolution pendant la simulation de la courbure du dimère ainsi que la direction de cette courbure. La courbure calculée sur la structure de départ de la molécule est de 10°. On constate qu'au cours de la dynamique elle augmente jusqu'à atteindre

en moyenne sur la dernière nanoseconde de simulation une valeur de 90° environ.

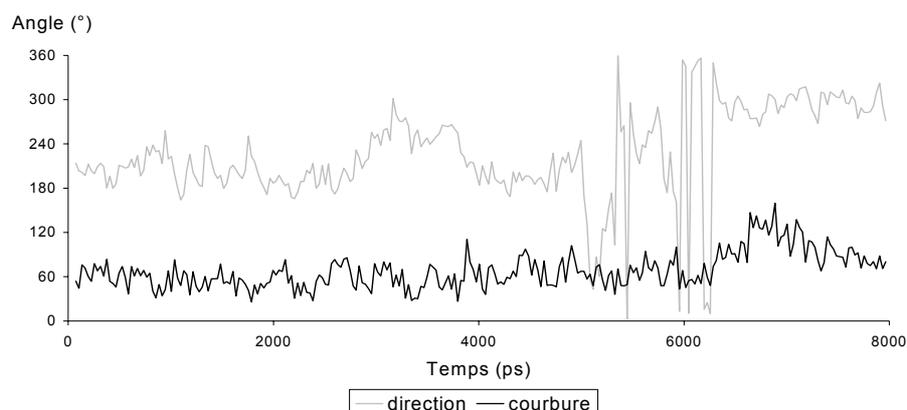


Figure 68 : Evolution de l'angle de courbure et de l'angle de direction de courbure au cours de la simulation du dimère stable de Girard et al. (1999).

De plus, la direction de courbure évolue aussi au cours du temps, partant de 60° dans la configuration d'origine, elle oscille autour de 300° sur la dernière nanoseconde de simulation. La Figure 69 présente la visualisation graphique des structures de début et de fin de dynamique avec leurs axes tels que CURVES les a calculés.

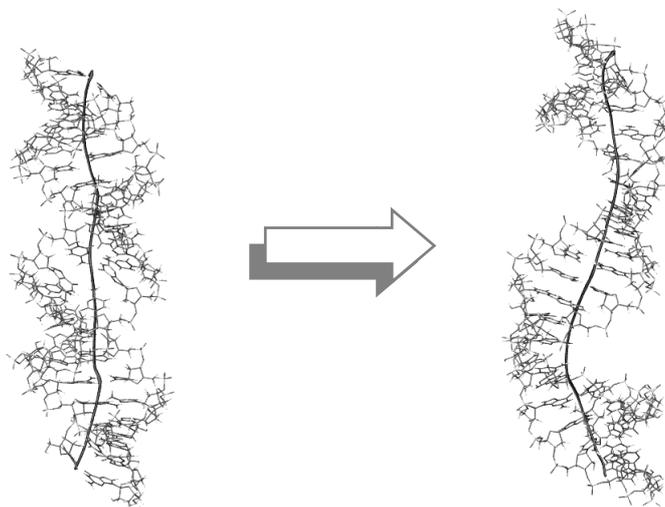


Figure 69 : Représentation des structures initiale et finale de la simulation du dimère de Girard et al. (1999) et de l'axe de ces molécules calculé par CURVES.

On peut constater sur cette figure que les zones de cassure de l'axe dans la structure finale de la simulation correspondent aux deux boucles internes du dimère étendu. Du fait de ces cassures, la molécule semble s'être globalement « étirée » le long de l'axe. Ces

déformations peuvent expliquer les valeurs importantes prises par le rmsd calculé sur tous les atomes du dimère étendu durant la simulation.

Le caractère flexible des boucles internes apparaît clairement sur la carte de flexibilité reproduite sur la Figure 70.

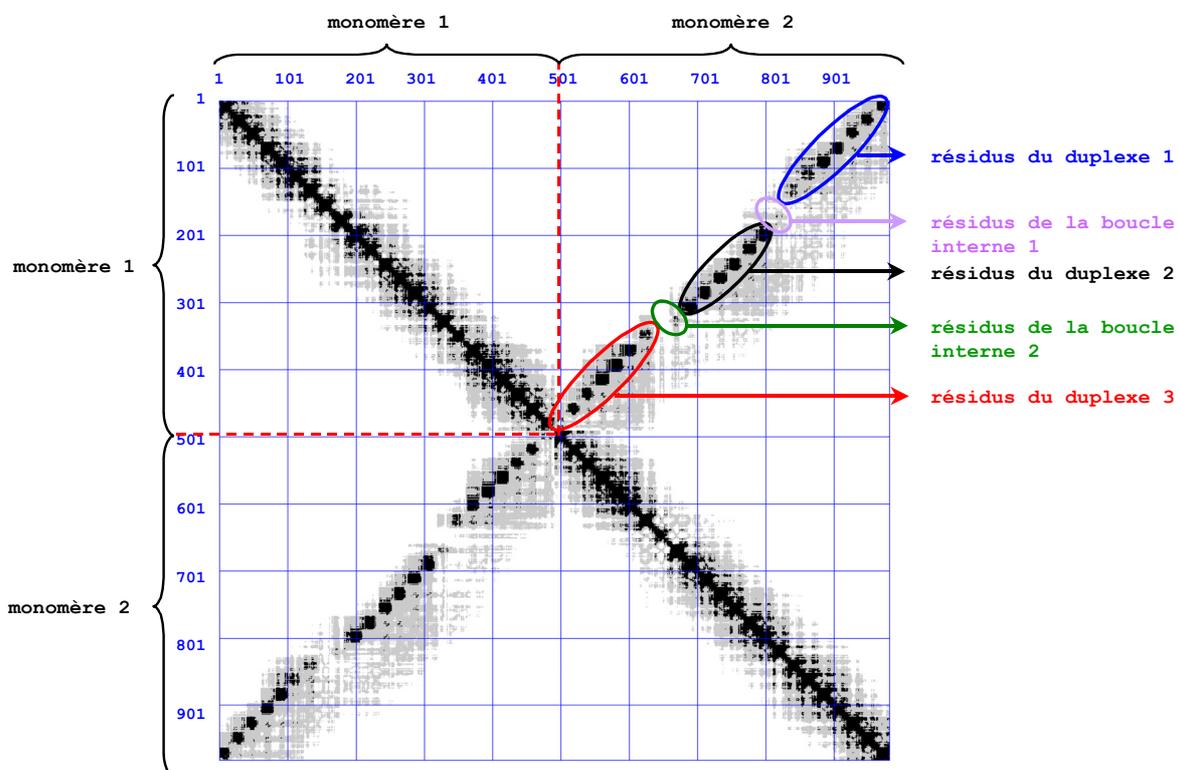


Figure 70 : Carte de flexibilité sur la dernière nanoseconde de simulation de la structure de Girard et al. (1999) avec des valeurs de coupure $\sigma_1=0.3 \text{ \AA}$ et $\sigma_2=0.6 \text{ \AA}$. Les zones noires correspondent à des fluctuations standards des distances inter-atomiques inférieures à σ_1 , les zones grises à des valeurs comprises entre σ_1 et σ_2 et les valeurs supérieures à σ_2 sont laissées en blanc. Les numéros des atomes sont reportés sur les côtés de la carte.

Les trois duplexes du dimère étendu sont nettement identifiables sur la contre-diagonale, séparés par des zones claires qui correspondent aux boucles internes. Un examen plus attentif de ces zones claires montre toutefois dans chacune une tache grise de semi-rigidité révélatrice de l'existence d'une corrélation entre les résidus A9 d'un monomère et A16 de l'autre. Les deux régions relatives aux deux duplexes des extrémités, en rouge et en bleue sur la Figure 70, présentent toutes deux sept taches de forte corrélation correspondant aux sept paires de bases de ces duplexes. On relève cependant pour chacun de ces deux duplexes une tache de corrélation

entre les résidus G7 d'un monomère et C17 de l'autre monomère moins nette que pour les autres paires de bases. Ces observations trouvent leur explication lorsque l'on étudie les évolutions des liaisons hydrogène entre les résidus. Sur la Figure 71 sont reportées en fonction du temps les moyennes des pourcentages d'existence des liaisons hydrogène des paires de bases canoniques de chaque duplexe.

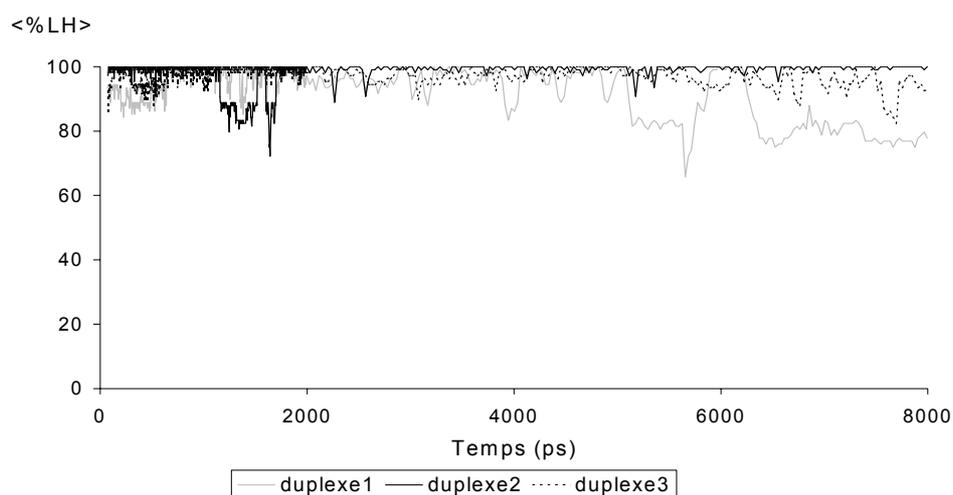


Figure 71 : Evolution du pourcentage moyen des liaisons hydrogène canoniques des trois duplexes au cours de la simulation du dimère de Girard et al. (1999).

On remarque pour le duplexe 1 et le duplexe 3 que les pourcentages moyens de liaisons hydrogène varient entre 80% et 100% alors que celles du duplexe 2 restent très proches des 100% pendant presque toute la durée de la simulation. Ces fluctuations dans les duplexes des extrémités sont dues à des ruptures momentanées des paires G7-C17' et G7'-C17, ainsi que l'expose la Figure 72.

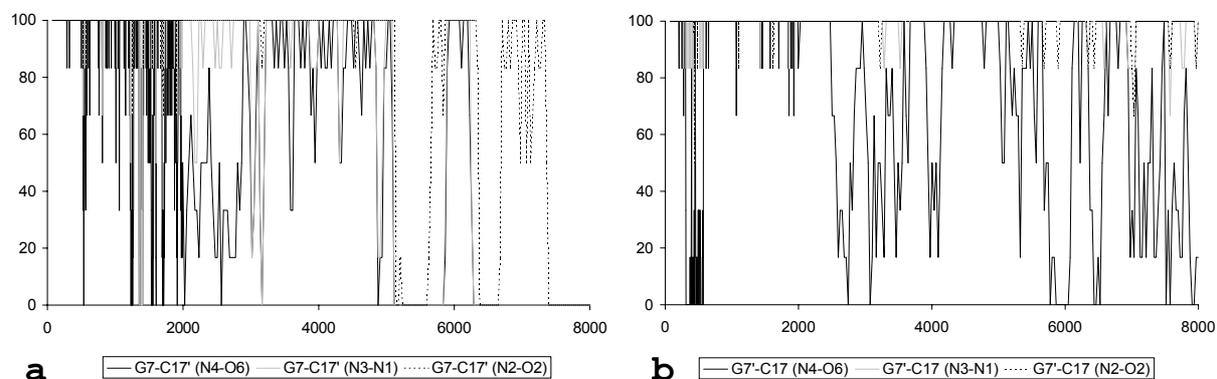


Figure 72 : Evolution du pourcentage d'existence de liaisons Watson-Crick entre les paires de bases G7-C17' (a) et G7'-C17 (b) pendant la simulation du dimère de Girard et al. (1999).

Cette séparation est en particulier visible sur la paire G7-C17' pour laquelle la rupture des trois liaisons hydrogène se fait simultanément deux fois en fin de simulation, alors qu'une seule de ces liaisons se rompt périodiquement dans la paire de base G7'-C17. Sur la Figure 73 sont reportées en fonction du temps les liaisons hydrogène formées en cours de simulation entre les résidus des boucles internes. Tout au long de la simulation, les résidus 8, 9 et 10 restent empilés ensemble, et ce dans les deux monomères.

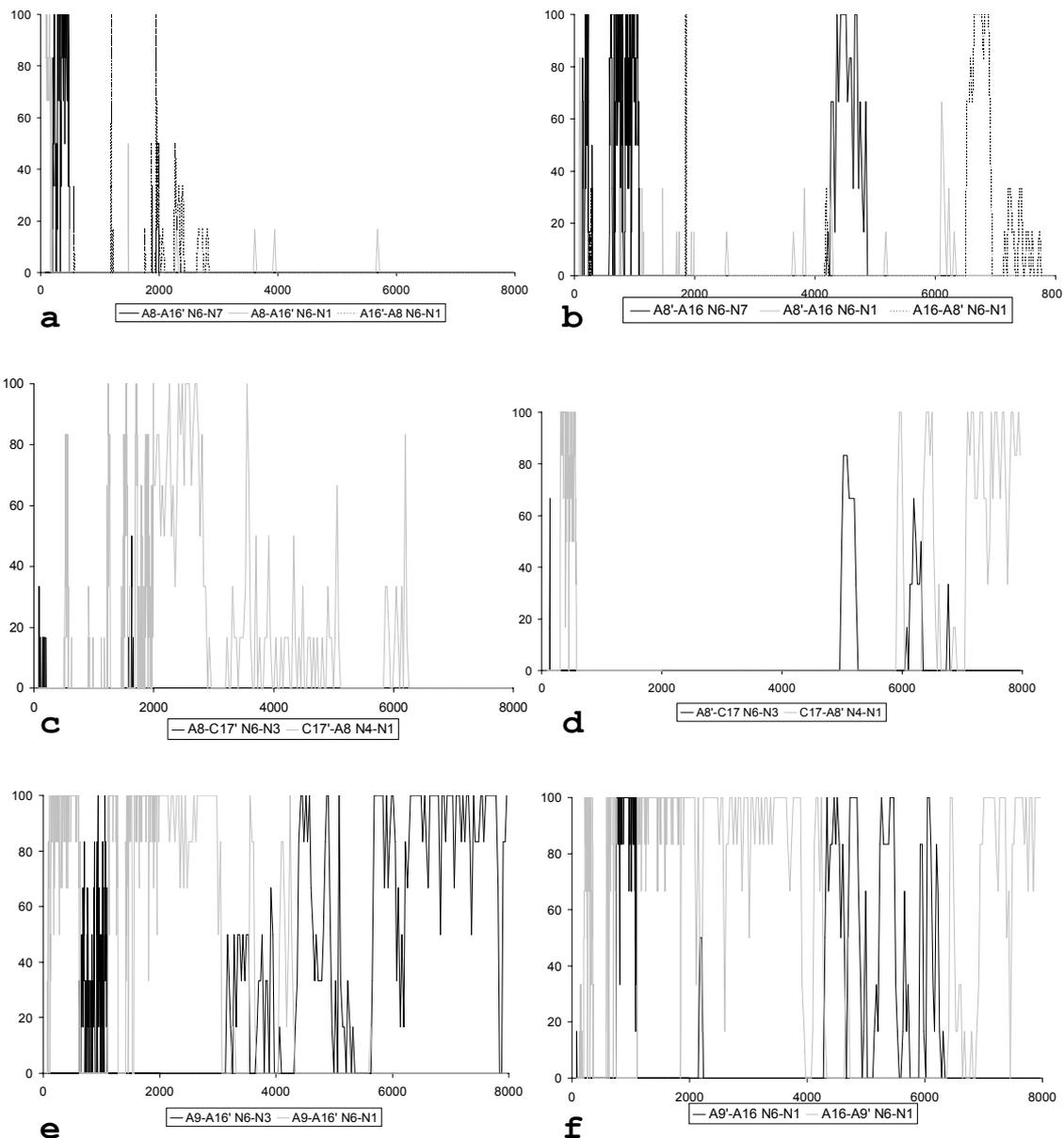


Figure 73 : Evolution pendant la simulation du dimère de Girard et al. (1999) des pourcentages d'existence des liaisons hydrogène dans les boucles internes entre les résidus : **a)** A8-A16' **b)** A8'-A16 **c)** A8-C17' **d)** A8'-C17 **e)** A9-A16' **f)** A9'-A16

Ces graphiques expliquent la corrélation constatée sur la carte de flexibilité entre les résidus A9 et A16' d'une part, A9' et A16 d'autre part. En effet ces résidus restent appariés tout au long de la simulation par deux liaisons hydrogène alternatives, les liaisons A9(N6)-A16'(N1) et A9(N6)-A16'(N3) pour la boucle interne 1 d'une part, A9'(N6)-A16(N1) et A16(N6)-A9(N1) pour la boucle interne 2 d'autre part. Les résidus 8 des deux monomères forment périodiquement des liaisons peu stables avec les résidus 16 et 17 du monomère opposé. La structure obtenue en fin de dynamique est présentée sur la Figure 74.

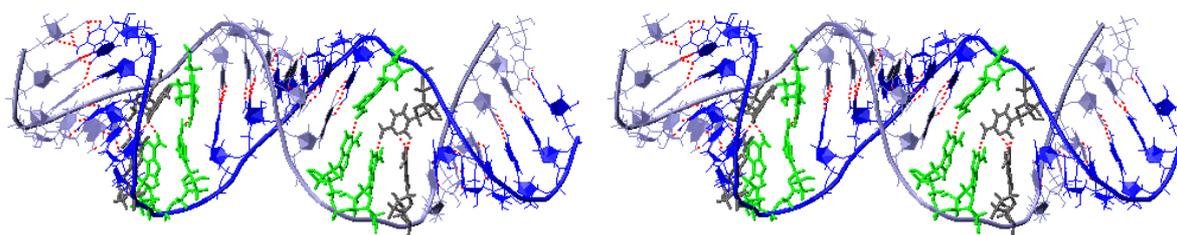


Figure 74 : Représentation en vue stéréo de la dernière structure de la simulation du dimère de Girard et al. (1999)

IV.b.ii. Etude de la structure cristallographique de Ennifar et al. à partir de la séquence SL1_{Mal}

Nous avons réalisé une simulation de dynamique moléculaire de 9 ns à partir de la structure publiée par Ennifar et al. (1999), en présence des 8 ions magnésium contenus dans le fichier de coordonnées d'origine complétés par 28 ions sodium, dans une boîte parallélépipédique contenant 7740 molécules d'eau.

La molécule reste relativement proche de sa structure initiale tout au long de la dynamique, avec un rmsd global sur l'ensemble des atomes qui évolue entre 1.5 Å et 4.6 Å (Figure 75). Les trois duplexes en particulier restent très proches de leur conformation d'origine en forme A et présentent ainsi des valeurs de rmsd inférieures à 2 Å pendant toute la simulation. Les valeurs moyennes de ces rmsd calculés sur la dernière nanoseconde de simulation sont reportées dans le Tableau 4.

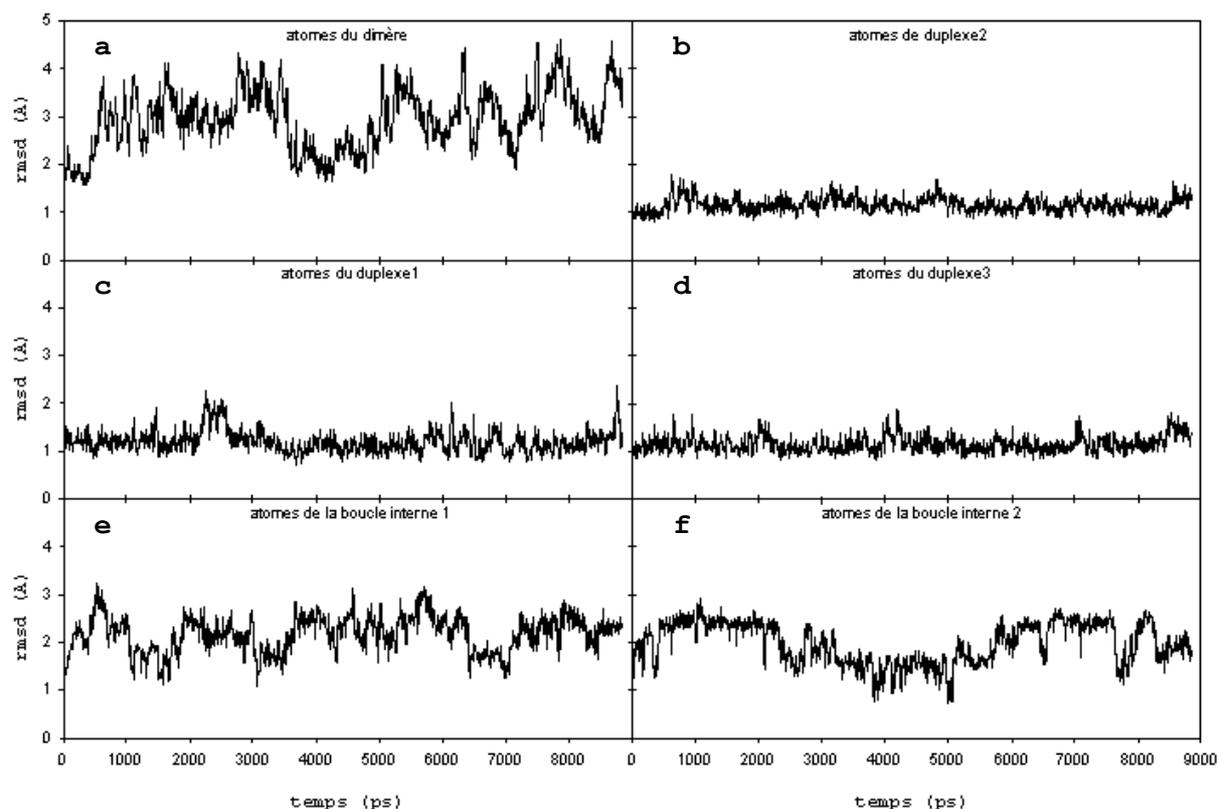


Figure 75 : Evolution du rmsd au cours de la simulation du dimère stable de Ennifar et al. (1999) par rapport à la structure expérimentale sur les atomes : du dimère complet (a), du duplexe 2 (b), du duplexe 1 (c), du duplexe 3 (d), de la boucle interne 1 (e) et de la boucle interne 2 (f)

| dimère | duplexe 1 | boucle interne 1 | duplexe 2 | boucle interne 2 | duplexe 3 |
|--------|-----------|------------------|-----------|------------------|-----------|
| 3.4 | 1.2 | 2.3 | 1.2 | 2.0 | 1.3 |

Tableau 4 : valeur moyenne du rmsd sur la dernière nanoseconde de simulation du dimère stable de Ennifar et al. (1999)

On constate que les plus grandes fluctuations sont obtenues au niveau des deux boucles internes du dimère. Encore peut-on voir sur la carte de flexibilité reproduite sur la Figure 76 que des trois résidus constitutifs de chacune des boucles internes, deux sont très fortement corrélés entre eux. Il s'agit des résidus G9 et A16' et A16 et G9'.

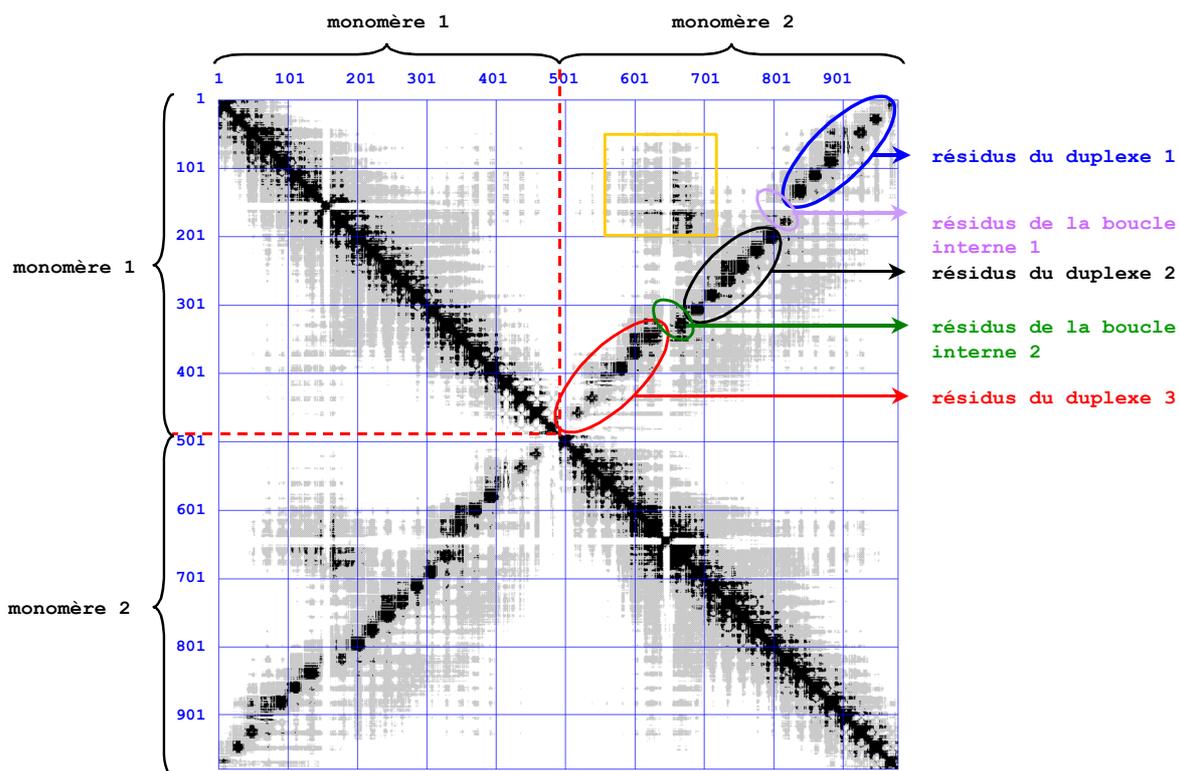


Figure 76 : Carte de flexibilité sur la dernière nanoseconde de simulation de la structure de Ennifar et al. (1999) avec des valeurs de coupure $\sigma_1=0.3 \text{ \AA}$ et $\sigma_2=0.6 \text{ \AA}$.

Seuls les résidus A8 et A8' présentent une très grande flexibilité responsable des modifications subies par les boucles internes au cours de la simulation. Ces résidus se distinguent très facilement sur la carte car ils se trouvent au centre d'une sorte de croix blanche sur chacun des monomères. Ces lignes blanches traduisent le fait que les deux résidus 8 et 8' ne sont corrélés avec aucun autre résidu. Cette « non-corrélation » est en particulier visible au centre de la région mise en relief par le carré jaune. Cette tache dévoile une zone de semi-rigidité impliquant les résidus 3 à 15 du monomère 1 et 4' à 10' du monomère 2, à l'exclusion des deux adénines 8 et 8', qui ne se révèlent pas plus corrélées entre elles qu'elles ne le sont avec les autres résidus de la molécule. La molécule présente une assez grande rigidité globale ainsi que le suggère le grand nombre de zones grises sur la carte. Le duplexe centrale du dimère montre six fortes taches de corrélation correspondant à ses six paires de bases. En revanche les

duplexes 1 et 3 arborent des taches moins nettes pour leurs trois paires de bases des extrémités du dimère.

De fait, lorsque l'on examine le pourcentage de présence des liaisons hydrogène canonique entre les paires de bases du dimère au cours du temps, on remarque que certaines de ces liaisons sont assez fragiles.

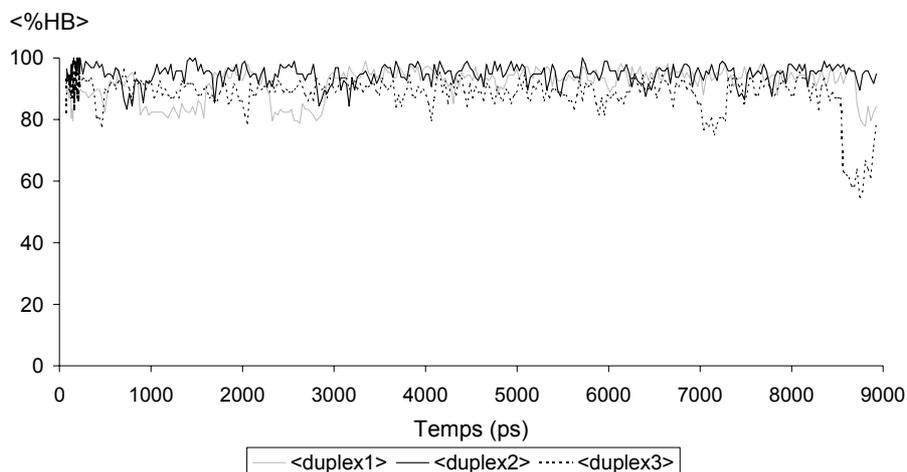


Figure 77 : Evolution du pourcentage moyen des liaisons hydrogène canoniques des trois duplexes pendant la simulation du dimère stable de Ennifar et al. (1999).

La Figure 77 retrace les moyennes de ces pourcentages de présence pour chacun des duplexes pendant la simulation. On voit sur cette figure que si pour le duplexe 2 les appariements entre les paires de bases restent proches de 100% avec quelques affaiblissements anecdotiques, ces mêmes pourcentages sont plutôt compris entre 80% et 90% pour les duplexes 1 et 3, en particulier en fin de simulation pour le duplexe 3. Cette faiblesse est due à des ruptures momentanées des paires C1-G23' et C1'-G23 des extrémités, ainsi qu'à des relâchements de courte durée de certaines liaisons hydrogène des paires U3-A21', G4-C20', U3'-A21, G4'-C20, U6'-A18 et G7'-C17. Les résidus des boucles internes présentent des appariements différents exposés Figure 78.

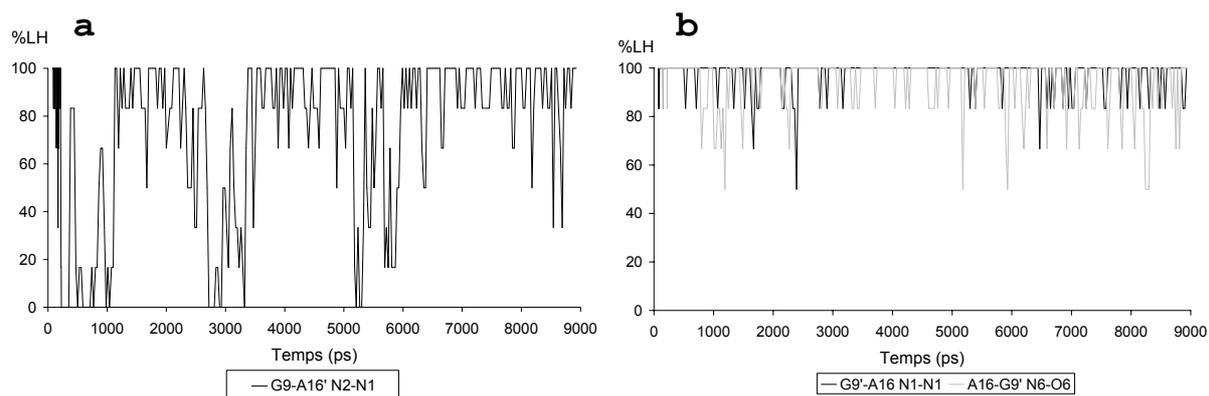


Figure 78 : Evolution du pourcentage d'existence des liaisons hydrogène dans les boucles internes entre les résidus G9-A16' (a) et G9'-A16 (b) pendant la simulation du dimère stable de Ennifar et al. (1999).

On observe que les deux liaisons formées entre les résidus G9' et A16 sont très stables alors que l'unique liaison existant entre les résidus G9 et A16' montre un comportement un peu plus chaotique. De plus, le résidu G9' contracte une liaison hydrogène stable avec le résidu C17 qui perdure pendant toute la simulation (Figure 79).

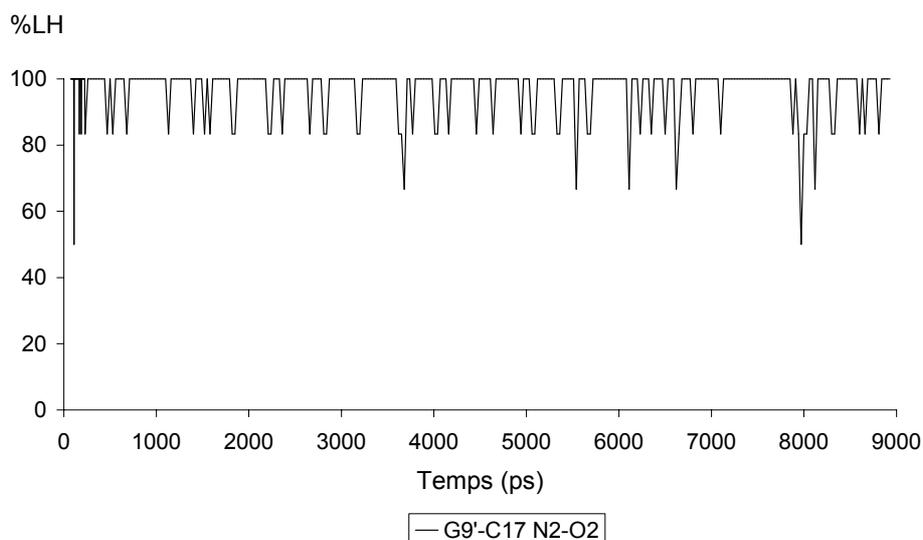


Figure 79 : Evolution au cours de la simulation du pourcentage d'existence de la liaison hydrogène G9' (N2)-C17 (O2) du dimère de Ennifar et al. (1999).

Les adénines 8 et 8' ne forment aucune liaison hydrogène durable à cause de leur position rejetée vers l'extérieur. Elles oscillent pendant toute la simulation entre une position totalement rejetée à l'extérieur et une position plus proche de la molécule. Ce phénomène est illustré par la Figure 80.

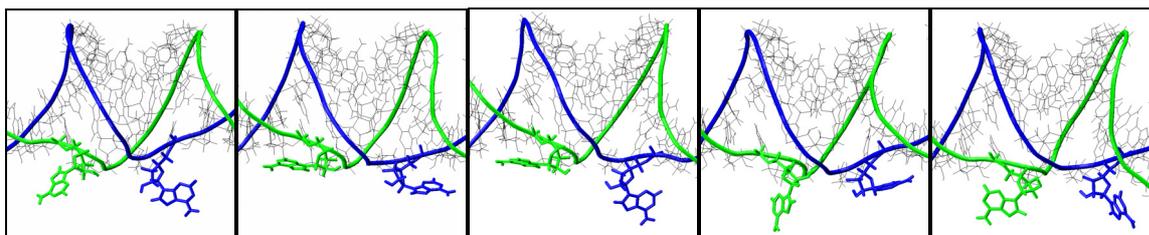


Figure 80 : Images instantanées des configurations prises par les résidus A8 (en bleu) et A8' (en vert) au cours de la simulation du dimère de Ennifar et al. (1999).

Les valeurs de courbure de la molécule au cours du temps sont retracées sur la Figure 81. La mesure de cette courbure sur la structure initiale donne une valeur d'environ 10° . Sur la dernière nanoseconde de simulation, cette courbure est en moyenne de 30° et elle reste comprise entre 2° et 60° pendant toute la simulation.

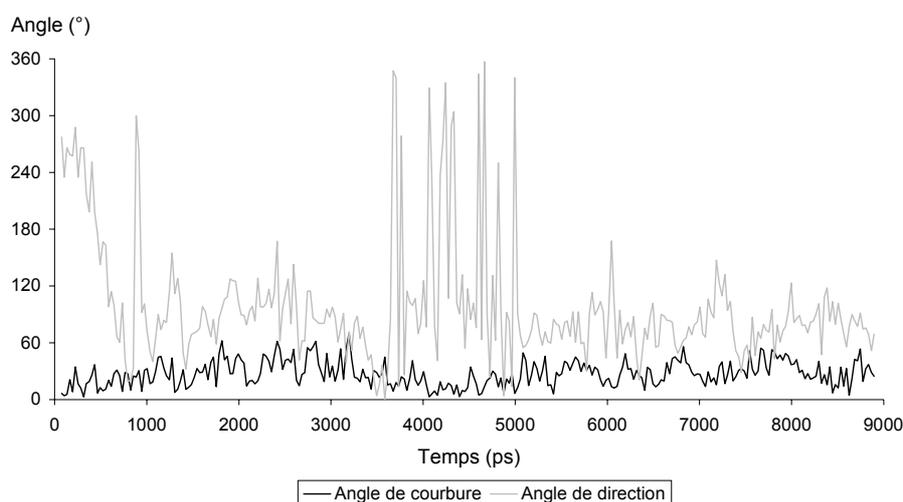


Figure 81 : Evolution de l'angle de courbure et de l'angle de direction de courbure pendant la simulation du dimère stable de Ennifar et al. (1999).

A la différence de la structure de Girard et al. (1999), la structure de Ennifar et al. (1999) a été simulée en présence d'ions magnésium. Nous avons étudié le comportement de ces ions au cours de la dynamique. Ennifar et al. (1999) ont identifié 8 sites de cations divalents dans la structure du dimère étendu de SL1_{Mal}, détaillés dans le Tableau 5. Nous avons conservé dans la structure initiale de notre simulation les coordonnées originelles des 8 ions magnésium mais pas les molécules d'eau qui leur avaient été associées.

En première observation, ce tableau nous montre que les sites de magnésium sont globalement bien conservés au cours de la dynamique,

exception faite du site β . Ce dernier, initialement en interaction avec le résidu G7, forme dès le début un complexe avec les molécules d'eau proches et se trouve en interaction avec plusieurs résidus par la suite. Tous les autres sites restent en interactions fortes avec les résidus auxquels ils étaient initialement associés. De plus, on constate que tous ces sites, déshydratés en majorité dans la structure minimisée, se réhydratent au cours de la simulation.

| Site Mg ²⁺ | Système publié par Ennifar et al. (1999) | | Système de départ de notre simulation (minimisée) | | Système final de notre simulation | |
|-----------------------|--|--------------------|---|--------------------|-----------------------------------|--------------|
| | Ligand | Distance (Å) | Ligand | Distance (Å) | Ligand | Distance (Å) |
| α | - | - | O4 U2 | 3.4 | - | - |
| | O4 U3 | 2.9 | O4 U3 | 1.9 | O4 U3 | 2.0 |
| | O6 G4 | 2.8 | O6 G4 | 1.9 | O6 G4 | 2.0 |
| | N7 G4 | 4.0 | N7 G4 | 4.3 | N7 G4 | 4.3 |
| | - | - | - | - | N4 C5 | 4.1 |
| | O6 G19' | 4.6 | O6 G19' | 4.3 | O6 G19' | 3.9 |
| | N4 C20' | 4.3 | N4 C20' | 2.7 | N4 C20' | 4.7 |
| | - | - | N6 A21' | 2.1 | N6 A21' | 4.3 |
| | - | - | N6 A224 | 4.2 | - | - |
| | - | - | 1 H ₂ O | 1.9 | 4 H ₂ O | 1.9 à 2.0 |
| α' | - | - | O6 G19 | 4.4 | O6 G19 | 4.1 |
| | - | - | N4 C20 | 4.0 | N4 C20 | 4.6 |
| | - | - | N6 A21 | 4.1 | N6 A21 | 4.5 |
| | - | - | O4 U2' | 4.5 | O4 U2' | 4.9 |
| | - | - | O4 U3' | 1.9 | O4 U3' | 2.0 |
| | O6 G4' | 4.4 | O6 G4' | 3.7 | O6 G4' | 4.0 |
| | N7 G4' | 4.7 | N7 G4' | 1.8 | N7 G4' | 2.1 |
| | - | - | N4 C5' | 4.6 | N4 C5' | 4.4 |
| | - | - | 2 H ₂ O | 1.9 | 4 H ₂ O | 1.9 à 2.0 |
| | 5 H ₂ O | 2.0 à 2.4 | - | - | - | - |
| β | - | - | - | - | OP C5 | 4.0 |
| | - | - | - | - | OP U6 | 4.0 |
| | - | - | OP A14' | 3.2 | - | - |
| | - | - | OP C15' | 4.3 | - | - |
| | O6 G7 | 4.2 | - | - | - | - |
| | N7 G7 | 4.2 | - | - | - | - |
| 4 H ₂ O | 2.0 à 2.4 | 5 H ₂ O | 1.9 à 2.0 | 6 H ₂ O | 2.0 | |
| β' | - | - | N4 C17 | 3.6 | N4 C17 | 4.1 |
| | - | - | N6 A18 | 3.6 | N6 A18 | 4.5 |
| | - | - | N4 C5' | 3.9 | - | - |
| | - | - | O4 U6' | 1.8 | O4 U6' | 1.9 |
| | O6 G7' | 3.9 | O6 G7' | 2.0 | O6 G7' | 3.9 |
| | N7 G7' | 4.6 | N7 G7' | 3.8 | N7 G7' | 2.0 |
| | 6 H ₂ O | 2.0 à 2.4 | 2 H ₂ O | 1.9 | 4 H ₂ O | 1.9 à 2.0 |
| | - | - | - | - | - | - |
| γ | OP A8 | 2.4 | OP A8 | 1.8 | OP A8 | 1.9 |
| | OP G9 | 4.5 | - | - | OP G9 | 4.4 |
| | - | - | O6 G9 | 2.0 | O6 G9 | 2.1 |
| | N7 G9 | 2.2 | N7 G9 | 2.4 | N7 G9 | 4.1 |
| | - | - | O6 G10 | 4.8 | - | - |
| | N7 G10 | 4.4 | N7 G10 | 4.8 | - | - |
| 4 H ₂ O | 2.1 à 2.4 | 2 H ₂ O | 1.9 | 4 H ₂ O | 2.0 à 2.1 | |
| γ' | OP A8' | 2.0 | OP A8' | 1.8 | OP A8' | 1.8 |
| | OP G9' | 4.6 | OP G9' | 3.6 | OP G9' | 4.6 |
| | - | - | O6 G9' | 2.0 | O6 G9' | 3.8 |
| | N7 G9' | 2.3 | N7 G9' | 3.5 | N7 G9' | 2.1 |
| | N7 G10' | 4.7 | N7 G10' | 4.2 | N7 G10' | 4.4 |
| | 4 H ₂ O | 1.9 à 2.5 | 1 H ₂ O | 1.9 | 4 H ₂ O | 1.9 à 2.1 |
| δ | OP G9 | 2.1 | OP G9 | 1.8 | OP G9 | 1.8 |
| | OP G9' | 2.1 | OP G9' | 1.8 | OP G9' | 1.9 |
| | - | - | OP G10' | 1.8 | OP G10' | 1.9 |
| | 4 H ₂ O | 2.1 à 2.5 | - | - | 3 H ₂ O | 2.0 à 2.1 |
| ϵ | - | - | - | - | O4 U11 | 4.1 |
| | O6 G12 | 3.1 | O6 G12 | 1.9 | O6 G12 | 3.6 |
| | N4 C13 | 3.9 | N4 C13 | 2.0 | N4 C13 | 4.3 |
| | - | - | N6 A14 | 3.9 | - | - |
| | - | - | O4 U11' | 3.1 | O4 U11' | 3.9 |
| | O6 G12' | 3.1 | O6 G12' | 1.8 | O6 G12' | 2.1 |
| | N4 C13' | 3.7 | N4 C13' | 3.9 | N4 C13' | 4.4 |
| | 1 H ₂ O | 1.9 | 1 H ₂ O | 1.9 | 5 H ₂ O | 1.9 à 2.0 |

Tableau 5 : Listing des interactions formées dans les 8 sites d'ion magnésium décrit par Ennifar et al. (1999) dans la structure publiée par cette équipe, dans la structure minimisée et dans la structure finale de notre simulation.

Cette déshydratation des sites en début de simulation est responsable des désappariement ponctuels que nous avons pu observer au niveau des paires U3-A21', G4-C20', U3'-A21, G4'-C20, U6'-A18 et G7'-C17. Ces liaisons se trouvent en effet par instant en compétition avec l'interaction générée par l'ion magnésium du site, et en particulier lorsqu'il y a peu de molécules d'eau dans son environnement, l'ion magnésium tend à interagir avec les groupements carbonyles et amines des résidus adjacents. Les 6 sites α , α' , β , β' , γ et γ' contribuent ainsi à la rigidité des tiges et des jonctions du dimère de SL1_{Mal}. En revanche, la séparation des 8 sites en deux groupes décrite par Ennifar et al. (1999) dans leurs travaux n'a plus cours en fin de simulation. Cette équipe distingue en effet les sites α , α' , β , β' et ε , pleinement hydratés et donc sans interaction directe avec l'ARN, des sites γ , γ' et δ qui, eux, présentent des coordinations directes avec les atomes de la molécule. A la fin de notre simulation, seul le site β possède un magnésium pleinement hydraté ne présentant aucune coordination directe avec l'ARN. Tous les autres sites possèdent au moins une interaction directe avec l'un des atomes du dimère de SL1_{Mal}. Les deux motifs uniques δ et ε (Figure 82), situés sur l'axe de symétrie de la molécule, sont très bien conservés au cours de la dynamique, en particulier le motif δ reliant les deux boucles internes du dimère par une interaction directe pontée OP(G9)-Mg²⁺-OP(G9'), motif décrit par Ennifar et al. (1999) sous le nom de « magnesium clamp ». Ces deux sites d'interactions peuvent en particulier expliquer la faible variation de courbure du dimère au cours du temps, ainsi que le fait que les boucles apparaissent comme des zones plutôt rigides sur la carte de flexibilité (zones grisées, voir Figure 76).

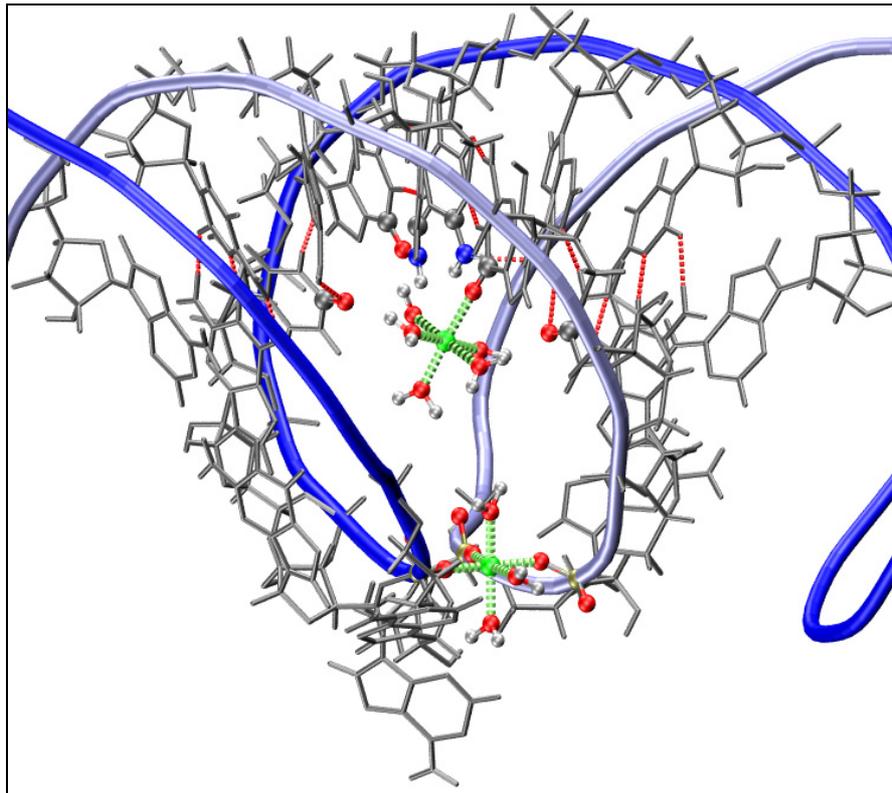


Figure 82 : Représentation des sites d'ion magnésium δ et ϵ dans la dernière configuration de la simulation du dimère stable de Ennifar et al. (1999) Le monomère 1 est représenté en bleu, le monomère 2 en violet.

La structure finale est représentée en vue stéréo sur la Figure 83. Les adénines 8 et 8' ont adopté sur cette conformation les deux types de configuration évoqués auparavant.

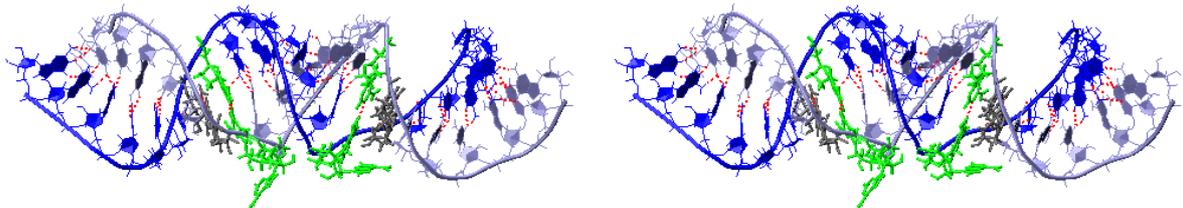


Figure 83 : Représentation en vue stéréo de la dernière structure obtenue au cours de la simulation du dimère de Ennifar et al. (1999).

IV.b.iii. Discussion-conclusion

Nous nous sommes dans ce chapitre attachés à étudier la stabilité en solution à 300 K des structures déterminées expérimentalement pour le dimère étendu de la séquence SL1 de VIH-1. Il convient toutefois de remarquer que, contrairement aux structures publiées pour le « kissing-complex », les structures parues pour le dimère étendu

sont en relativement bon accord l'une avec l'autre. Nous avons représenté sur la Figure 84 les valeurs du rmsd calculé sur le squelette phospho-diester pour les 15 structures publiées par Girard et al. (1999) par rapport soit à la conformation moyenne de cette famille de structures, soit à la structure publiée par Ennifar et al. (1999).

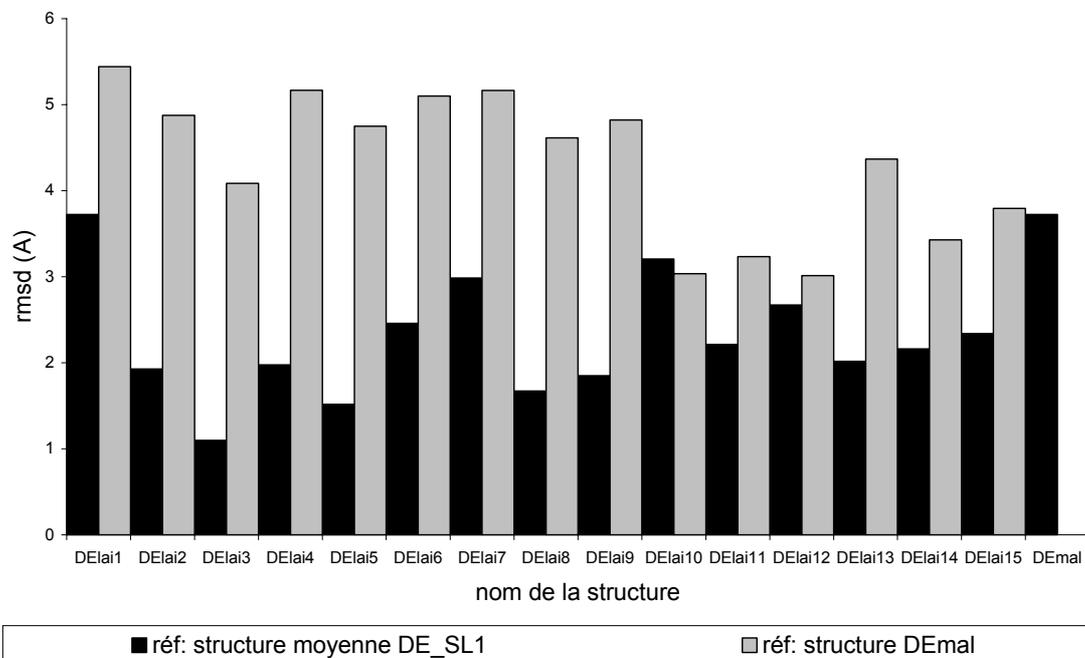


Figure 84 : Diagramme des rmsd calculé sur les atomes du squelette phospho-diester pour les 15 structures de la famille déterminée par Girard et al. (1999) par rapport soit à la structure moyenne de cette famille (bleu) soit à la structure de dimère déterminée par Ennifar et al. (1999) (gris)

Nous avons également reporté dans le Tableau 6 les valeurs moyennes de ces rmsd ainsi que les moyennes correspondant aux 5 groupes constitutifs de ce dimère.

| moyenne des rmsd calculés sur les 15 structures RMN | du dimère | du duplexe1 | de la boucle interne 1 | du duplexe2 | de la boucle interne 2 | du duplexe3 |
|---|-----------|-------------|------------------------|-------------|------------------------|-------------|
| par rapport à la conformation moyenne | 2.3 | 0.6 | 0.7 | 1.1 | 0.7 | 0.6 |
| par rapport à la structure d'Ennifar et al. | 4.3 | 1.7 | 3.4 | 3.1 | 3.5 | 1.6 |

Tableau 6 : Valeur moyenne sur l'ensemble de la famille de Girard et al. (1999) du rmsd calculé sur les atomes du squelette phospho-diester par rapport à la structure moyenne de la famille ou du dimère déterminé par Ennifar et al. (1999).

La structure d'Ennifar et al. (1999) diffère certes de la famille déterminée par Girard et al. (1999), notamment au niveau du duplexe centrale et des deux boucles internes, mais dans des proportions relativement faibles. Les valeurs de rmsd calculées sur ces 5 groupements pour la structure d'Ennifar et al. (1999) par rapport à la famille déterminée par Girard et al. (1999) restent du même ordre que les évolutions du rmsd de ces mêmes groupements au cours de la simulation du dimère étendu de l'isolat Lai. De plus, le rmsd calculé pour le dimère d'Ennifar et al. (1999) par rapport à la structure moyenne de la famille déterminée par Girard et al. (1999) est de 3.7 Å, or la conformation Delail (voir Figure 84), qui s'avère être la plus éloignée de cette structure moyenne, présente exactement la même valeur de rmsd. Les dimères étendus publiés dans la littérature présentent donc des structures plus comparables entre elles que les structures de « kissing-complex » évoquées au chapitre précédent.

Au cours de ces deux simulations, les structures se révèlent relativement stables avec une structure globale proche de celle d'une double hélice en forme A. En particulier ces deux simulations présentent peu de modifications au niveau de leurs duplexes 1 et 3 et ces mêmes duplexes se trouvent être très semblables d'un isolat à l'autre dans les structures initiales. On remarque cependant pour les deux molécules une flexibilité au niveau des boucles internes en accord avec l'étude RMN de Mujeeb et al. (1999), plus accentuée et occasionnant une courbure plus prononcée dans le cas de la structure de Girard et al. (1999) que dans celui de la structure d'Ennifar et al. (1999). Il semble que les paires G7-C17' et G7'-C17, qui présentent des ruptures périodiques, confèrent une grande souplesse aux boucles internes et par extension à la molécule toute entière. Cet affaiblissement des liaisons hydrogène des paires G7-C17' et G7'-C17 de l'isolat Lai par rapport à un appariement G-C canonique a aussi été observé par RMN (Girard 2000). Ce phénomène induit ainsi une courbure globale du dimère qui n'existe pas dans la structure d'origine et permet en outre de maintenir les résidus A8 à l'intérieur des boucles internes en formant des liaisons épisodiques entre ces résidus et les résidus C17. L'aptitude de ces paires de

bases G7-C17' et G7'-C17 à se rompre de façon périodique est due à une particularité de séquence de l'isolat Lai. En effet, ceci n'est possible que, d'une part parce que la paire de base précédente dans la tige est une paire G-U, d'autre part parce que la paire G7-C17 est située à une extrémité de la tige, la boucle occasionnant des contraintes plus faibles sur cette paire de bases. Le cas de la structure d'Ennifar et al. (1999) est différent, étant donné que l'isolat Mal de la séquence SL1 présente en amont de la paire G7-C17 non plus une paire G-U mais une paire A-U. Cette différence au niveau de la séquence est sans doute la cause de la plus grande solidité des paires G7-C17 de l'isolat Mal par rapport à l'isolat Lai.

La différence structurale principale existant entre ces deux dimères est la position des adénines 8, rejetées vers l'extérieur dans la structure de Ennifar et al. (1999) et tournées vers l'intérieur dans celle de Girard et al. (1999). Ennifar et al. (1999) suggèrent l'existence d'un équilibre entre une conformation intérieure et une conformation extérieure en ce qui concerne ces adénines. Ces auteurs émettent l'hypothèse que cet équilibre serait magnésium dépendant, ce qui expliquerait que la structure isolée par diffraction de rayons X en présence de magnésium se trouve en conformation extérieure au niveau de la boucle interne, alors que la structure déterminée par RMN en absence de magnésium présente une conformation intérieure à ce niveau. Cette conformation extérieure des boucles internes serait en particulier due aux sites de magnésium δ , γ et γ' ce qui est en accord avec les résultats de notre simulation (Figure 85). En particulier, les deux sites γ restent en interaction avec les groupements phosphate des purines 8 et 9 pendant la simulation, restreignant ainsi la liberté de mouvement de ces résidus.

Comme nous l'avons dit précédemment, les adénines 8 adoptent plusieurs conformations au cours de la simulation de dynamique moléculaire, formant une espèce de balancier entre la position tournée vers l'extérieur et une position plus proche des tiges. Cependant, du fait de l'occupation des sites γ ainsi que de la fermeture entre les deux tiges provoquée par le site δ , il est

impossible à ces adénines 8 d'adopter une conformation intérieure au cours de notre simulation.

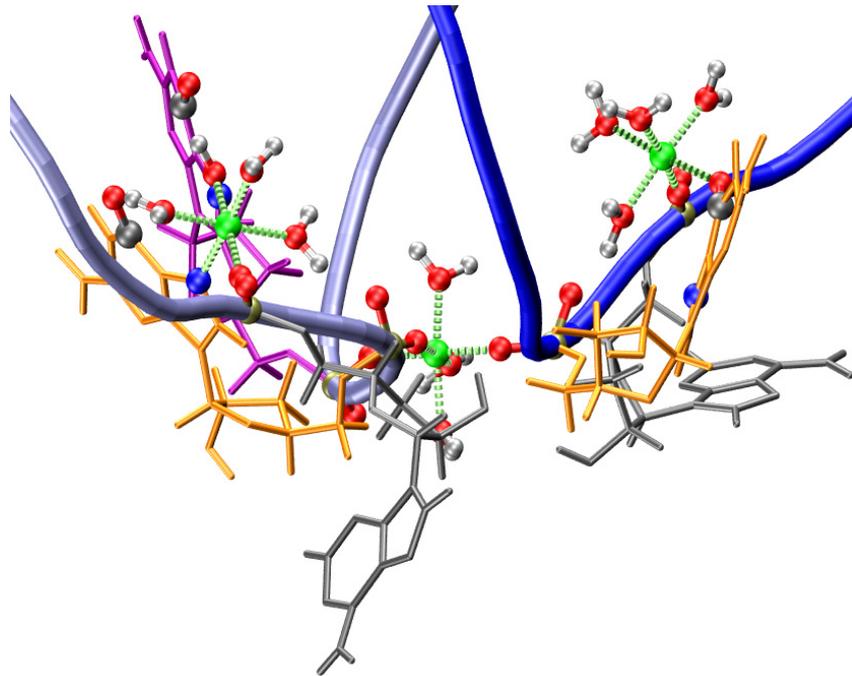


Figure 85 : Représentation des sites d'ion magnésium δ , γ et γ' pour la dernière configuration de la simulation du dimère de Ennifar et al. (1999). Les monomères 1 et 2 sont représentés en bleu et violet respectivement, les ions magnésium sont en vert, les résidus 8 et 8', 9 et 9' et 10' sont représentés en gris, orange et rose respectivement.

IV.c. Thermodynamique de fermeture de la paire de base G7-C17 à la jonction de la tige et de la boucle

La structure de Mujeeb et al. (1998) présente ce résultat inattendu d'une tige de six paires de bases seulement, les résidus G7 et C17 se trouvant désappariés et toutes les tentatives pour forcer leur appariement conduisant systématiquement à un échec. Cependant, une simulation moléculaire de 20 ns effectuée à partir de cette molécule conduit à une structure asymétrique, les résidus G7 et C17 de l'un des monomères s'étant spontanément appariés au bout de 7 ns de simulation. De plus, une simulation de 6 ns effectuée à partir de cette même structure mais avec la séquence sauvage de SL1_{Lai} conduit au même résultat d'une conformation asymétrique contenant sept paires de bases dans l'une de ses tiges alors que l'autre n'en renferme que six. Au vu de tels résultats, nous avons décidé d'utiliser conjointement la technique de simulation de

dynamique moléculaire dirigée (TMD) et la méthode Weighted Histogram Analysis (WHAM) afin de déterminer l'énergie libre de transition permettant le passage de la paire G7-C17 de l'état ouvert à l'état fermé.

IV.c.i. Détermination des grandeurs thermodynamiques de transition

Toutes les simulations présentées dans cette partie sont réalisées à 300 K, en présence de 44 contre-ions Na⁺ dans une boîte parallélépipédique contenant 6987 molécules d'eau. Nous avons utilisé comme structure de départ pour ces simulations la conformation finale issue de la simulation de dynamique moléculaire à 300 K de la structure de Mujeeb et al. (1998) (§IV.a.i). Cette conformation est dissymétrique avec un monomère 1 composé d'une tige de 7 paires de bases et d'une boucle de 9 résidus tandis que le monomère 2 possède une tige de 6 paires de bases et une boucle de 11 résidus. Afin de composer la structure qui servira par la suite de cible à nos TMD, nous avons dupliqué le monomère 1 que nous avons ensuite superposé au monomère 2. Les coordonnées de la paire G7-C17 ainsi obtenues constitueront la structure cible pour une simulation. Ces conformations sont représentées Figure 86.

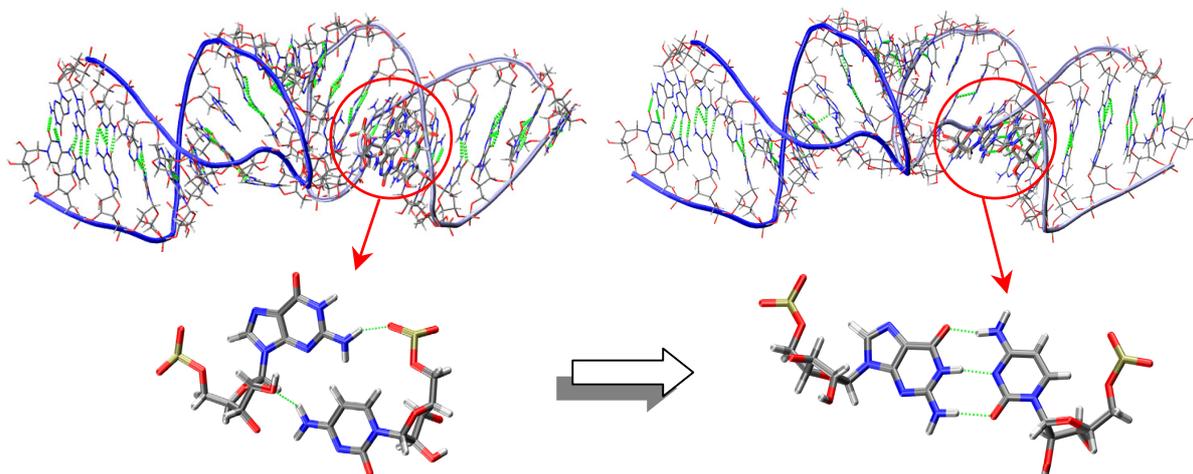


Figure 86 : Représentation des structures initiale et cible de la transition ainsi que des appariements entre les résidus 7 et 17 du second monomère pour chacune de ces structures.

Le rmsd entre les paires G7-C17 de chacune de ces structures est de 3.1 Å. Nous avons effectué à partir de cette structure initiale une TMD constituée de 37 simulations en faisant varier ρ_0 d'abord de

3.1 Å à 0 Å puis de 3.2 Å à 3.6 Å par pas de 0.1 Å. Chacune des 37 simulations a été réalisée sur 500 ps avec une constante de force $k=2$ kcal/mol.Å². Les conformations ayant été stockées toutes les picosecondes, nous disposons d'un échantillonnage de 500 conformères pour chaque valeur de ρ_0 .

La méthode WHAM nous a permis d'obtenir, à partir de ces échantillonnages, un profil d'énergie libre pour la transition entre les états ouvert et fermé de la paire G7-C17. Nous avons exécuté ce calcul en utilisant divers sous-ensembles des 500 conformères de chaque simulation. Les différents profils d'énergie libre obtenus sont présentés Figure 87.

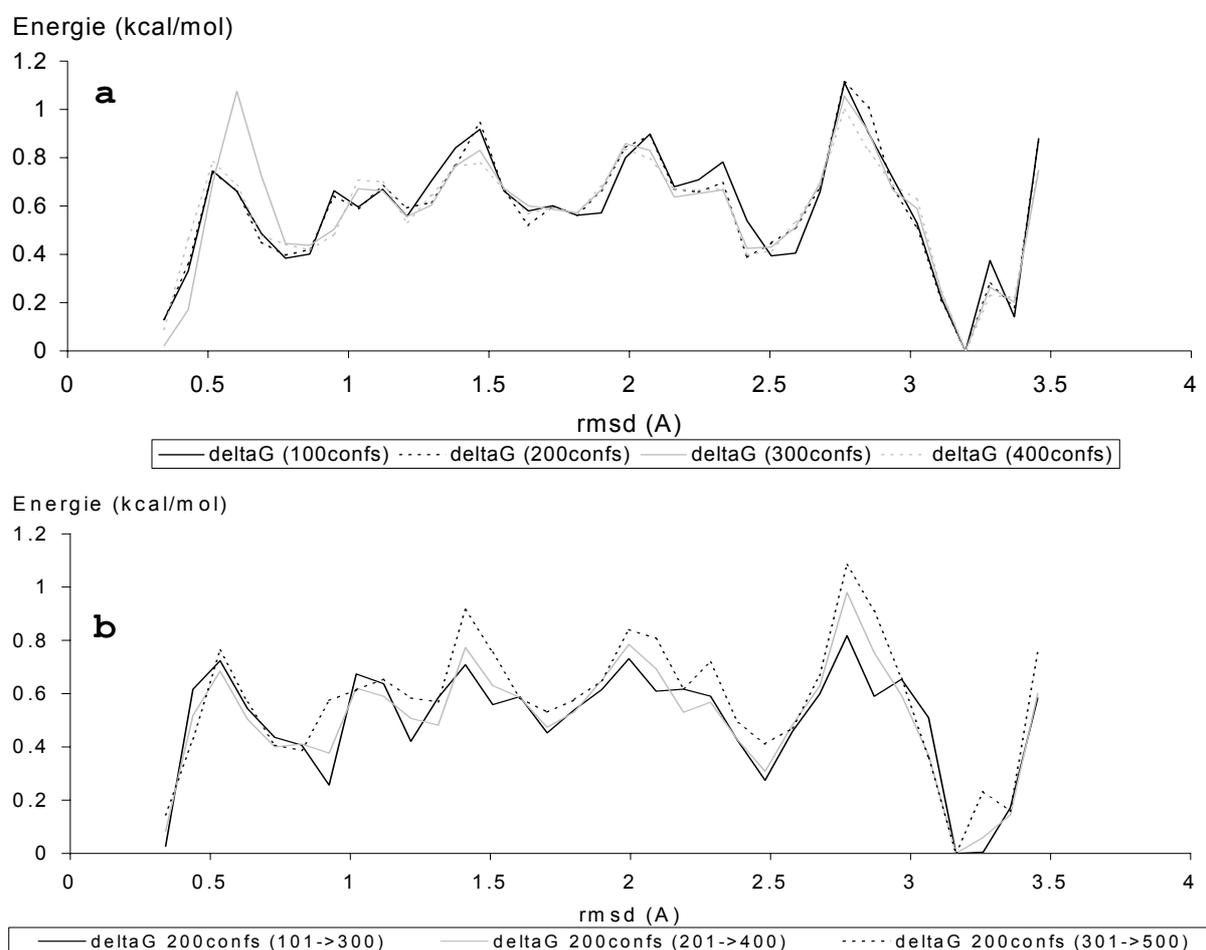


Figure 87 : Profil du ΔG de la fermeture de la paire de base G7-C17 en fonction du rmsd entre la conformation instantanée de cette paire de résidus et sa conformation fermée cible. Les différents sous-ensembles de configurations utilisés pour obtenir ces ΔG à partir de WHAM sont soit de tailles différentes (a) soit de tailles identiques mais extraits à des temps différents de simulation (b).

Nous constatons sur ces graphiques que l'énergie libre ΔG entre l'état apparié et l'état désapparié est nulle, et que les barrières d'énergie les plus élevées à franchir pour passer d'un état à l'autre ne dépassent pas une valeur de l'ordre de 1 à $2RT$, où R est la constante des gaz parfaits et T la température absolue. Ces valeurs sont bien plus faibles que la variation d'énergie libre intervenant dans le processus d'ouverture des paires de bases dans une double hélice d'ADN qui est de l'ordre de $10-20RT$ (Banavali and MacKerell 2002; Giudice and Lavery 2003a; Giudice et al. 2003b; Varnai and Lavery 2002).

Ce résultat ne dépend pas de la taille de l'échantillonnage choisi pour appliquer la méthode WHAM. La Figure 87-a présente les courbes obtenues en considérant comme échantillonnages les ensembles constitués respectivement par les 100, 200, 300 et 400 dernières configurations de la trajectoire. Le même type de profil est également obtenu en utilisant des échantillonnages contenant le même nombre de configurations mais extraites à des temps différents des trajectoires (Figure 87-b). Le résultat ne dépend donc pas non plus du temps d'équilibration précédent l'échantillonnage. Pour toutes ces courbes, la déviation par la méthode des moindres carrés est toujours inférieure à 0.025 kcal/mol. Il est intéressant de noter que les courbes présentent toutes un minimum aux environs de 3.1 Å de rmsd, ce qui correspond à notre état ouvert initial. Ce minimum est dû à la présence, dans cette conformation ouverte, de deux liaisons hydrogène stabilisantes entre les résidus G7 et C17, les liaisons G7(N2)-C17(O2P) et G7(O2')-C17(N4) (Figure 86).

Nous avons également calculé les pourcentages de liaisons hydrogène canoniques existantes ainsi que les évolutions de l'angle de plissement du sucre au cours de ces simulations. Ces valeurs sont reportées sur la Figure 88 superposées au profil de l'énergie libre de transition. Cette figure montre que l'appariement se produit par formation simultanée des trois liaisons hydrogène Watson-Crick entre G7 et C17 et que cet appariement précède le passage du sucre de G7 de la conformation C2'-endo à la conformation C3'-endo. Ces courbes montrent aussi que nous avons bien, dans notre état final, des résidus G7 et C17 appariés possédant des sucres en conformation C3'-endo.

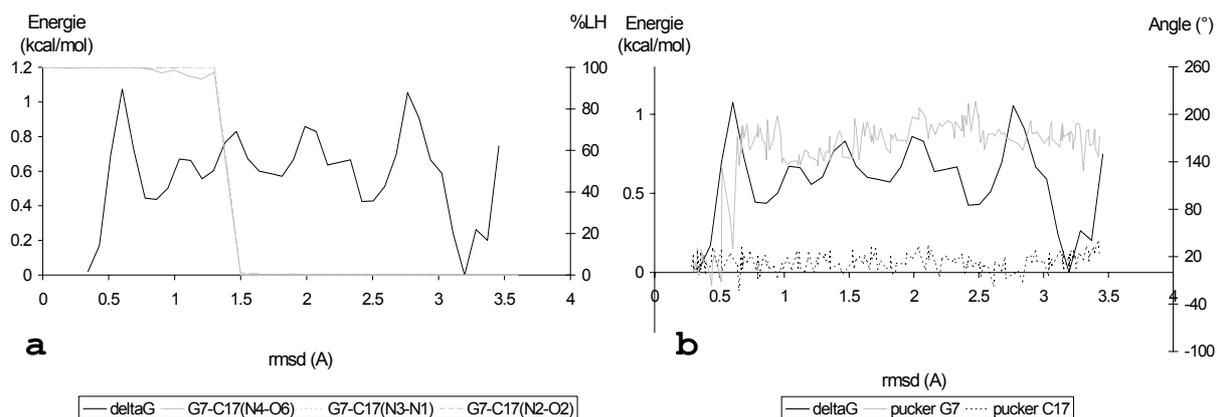


Figure 88 : Superposition du profil de ΔG de fermeture de la paire de bases G7-C17 obtenu sur l'ensemble de 300 configurations avec **a)** le pourcentage de liaisons Watson-Crick entre G7 et C17 et **b)** l'évolution de l'angle de phase de G7 et C17 en fonction du rmsd entre la configuration instantanée et fermée de G7-C17.

Les calculs d'enthalpie et d'entropie libre ont également fait l'objet de notre investigation. Pour réaliser ces mesures, nous nous sommes basés sur les équations suivantes de la thermodynamique classique :

$$\Delta H = \Delta E + \Delta(PV) \quad (36)$$

$$\Delta G = \Delta H - T\Delta S \quad (37)$$

Du fait que nos simulations sont réalisées à pression constante (ensemble NPT) dans une boîte d'eau dont la variation de volume est très faible, nous pouvons, avec une bonne adéquation, approximer ΔH à la différence d'énergie totale entre les deux états ΔE . Pour obtenir ces énergies totales, nous avons réalisé une équilibration d'1.1 ns à 300 K avant et après la fermeture de la paire de bases G7-C17 en stockant les valeurs d'énergie toutes les picosecondes. Les histogrammes normalisés des énergies obtenues au cours de ces étapes d'équilibration sont reportés Figure 89 avec une résolution de 1 kcal/mol.

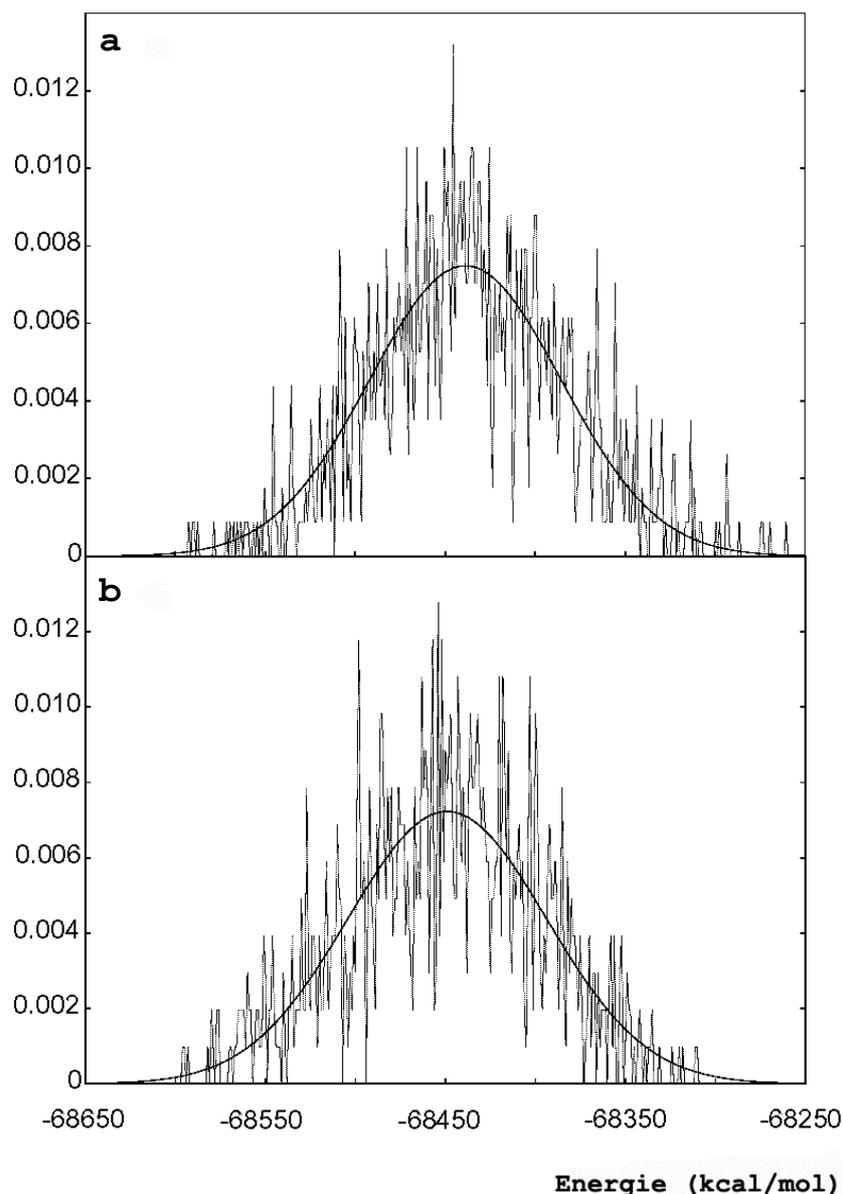


Figure 89 : Histogrammes normalisés de l'énergie totale obtenue lors des simulations sans contraintes des états **a)** ouvert et **b)** fermé de la paire G7-C17 avec une résolution de 1 kcal/mol. Les meilleures distributions normales obtenues à partir de ces histogrammes sont également représentées.

Les courbes de Gauss décrivant le mieux ces histogrammes sont obtenus par la méthode de Marquardt (1963). Les valeurs moyennes des énergies totales sont de -68438 kcal/mol et -68449 kcal/mol pour les états désappariés et appariés respectivement. Nous obtenons ainsi la valeur $\Delta H = -11$ kcal/mol et selon l'équation (37), nous en déduisons la valeur $\Delta S = -0.037$ kcal/mol.K. Ces valeurs ainsi que d'autres données thermodynamiques extraites de la littérature sont regroupées dans le Tableau 7.

| Paire de bases | Source | Système étudié | ΔG | ΔH | ΔS |
|--|--------------------------|--|------------|------------|------------|
| ajout de G-C après G-U | cette étude | « kissing-complex » de SL1 _{Lai} | 0 | -11 | -0.037 |
| ajout de G-C après G-U | (Cantor and Schimmel) | duplexe d'ARN | 0 | - | - |
| ajout de G-U après G-C | (Cantor and Schimmel) | duplexe d'ARN | 0 | - | - |
| fermeture par G-C d'une boucle de 9 nucléotides | (Bloomfield et al.) | tige-boucle d'ARN | +5 | - | - |

Tableau 7 : Paramètres thermodynamiques de formation d'une paire de bases. Les valeurs données par Cantor et Schimmel (1980) et par Blommfiel et al. (1974) ont été déterminées dans les conditions standard, notre étude dans les conditions suivantes : T=300 K, [Na+]≈0.33 M. ΔG et ΔH sont donnés en kcal/mol, ΔS en kcal/mol.K.

Nous avons voulu replacer cette étude de fermeture d'une paire de base à la jonction d'une tige-boucle dans un contexte plus général. La décomposition en deux chemins possibles pour effectuer cette transition entre les états ouvert et fermé de la paire de bases considérée est exposée sur la Figure 90. Le premier est celui que nous avons utilisé pour effectuer cette étude. Le second consiste, partant de la tige-boucle initiale avec une boucle de 11 nucléotides, à considérer tout d'abord la fusion de la tige à l'exception de la paire U6-G18 qui clôt la boucle. L'étape suivante décrit l'ouverture de cette paire de bases, transformant la molécule en un ARN linéaire. Les étapes suivantes détaillent la formation de la tige-boucle finale possédant une boucle de 9 nucléotides selon le processus suivant, tout d'abord la formation de cette boucle de 9 nucléotides grâce à la fermeture de la paire G-C, puis la formation de la deuxième paire de base G-U de cette tige, enfin l'élongation de cette tige jusqu'à atteindre les deux extrémités du brin d'ARN.

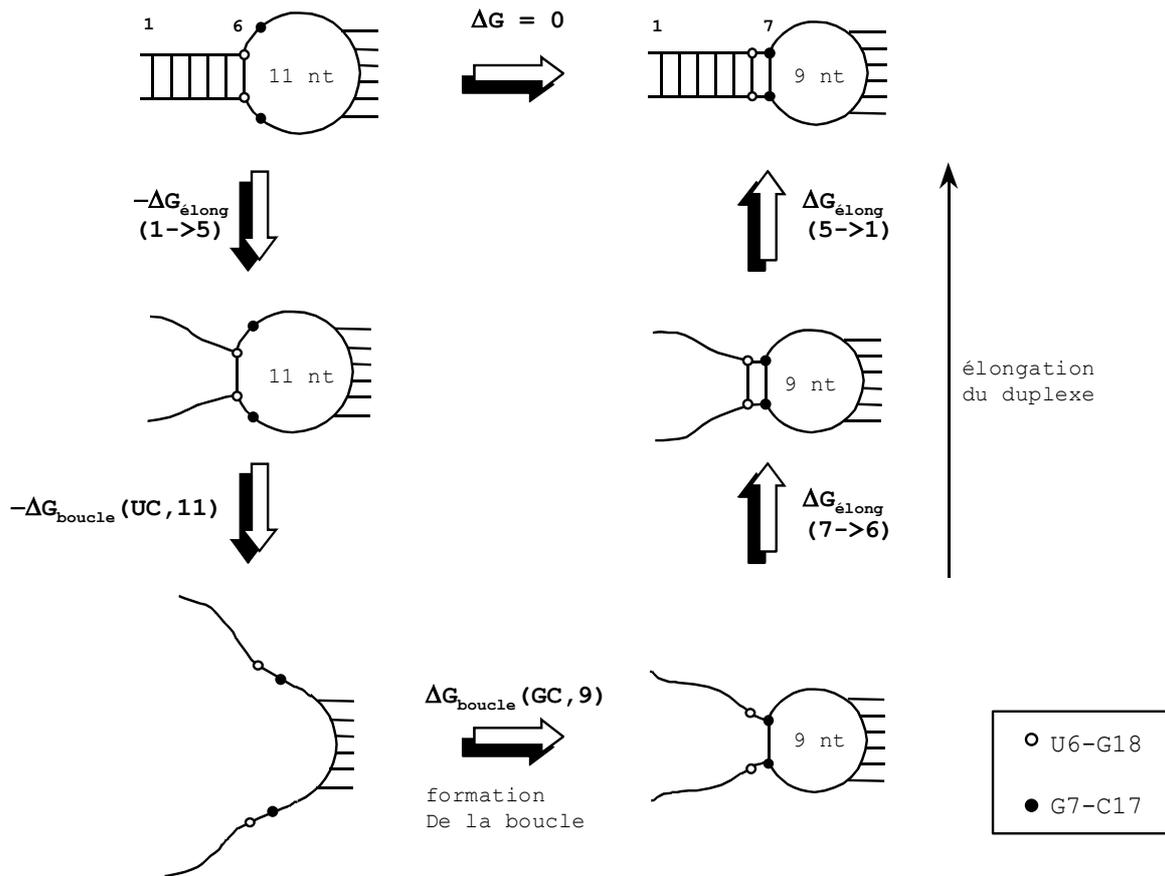


Figure 90 : Représentation de deux chemins possibles pour la transition entre les états ouvert et fermé de la paire de bases G7-C17 d'un monomère du dimère. On appelle $\Delta G_{\text{élong}}$ l'énergie libre d'élongation d'un duplexe et $\Delta G_{\text{boucle}}(XY, n)$ l'énergie de formation d'une boucle de n nucléotides fermée par une paire X-Y. Seul un des monomères du dimère est représenté sur cette figure, l'autre monomère possédant déjà une paire de bases G7-C17 formée.

L'énergie libre étant une fonction d'état thermodynamique, elle est indépendante du chemin emprunté. On peut donc écrire la relation suivante :

$$\begin{aligned} \Delta G &= -\Delta G_{\text{élong}}(1 \rightarrow 5) - \Delta G_{\text{boucle}}(\text{UC}, 11) + \Delta G_{\text{boucle}}(\text{GC}, 9) + \Delta G_{\text{élong}}(7 \rightarrow 6) + \Delta G_{\text{élong}}(5 \rightarrow 1) \\ &= -\Delta G_{\text{boucle}}(\text{UC}, 11) + \Delta G_{\text{boucle}}(\text{GC}, 9) + \Delta G_{\text{élong}}(7 \rightarrow 6) \\ &= 0 \end{aligned} \quad (38)$$

où ΔG est la différence d'énergie libre entre les états ouvert et fermé de la paire de base G-C, $\Delta G_{\text{boucle}}(\text{UC}, 11)$ et $\Delta G_{\text{boucle}}(\text{GC}, 9)$ sont les énergies libres de formation d'une boucle de 11 nt fermée par une paire U-G et d'une boucle de 9 nt fermée par une paire G-C respectivement à partir d'un ARN linéaire et $\Delta G_{\text{élong}}$ est l'énergie associée à la formation des paires de bases de la tige à partir

d'une ou de plusieurs paires de bases déjà existantes. Si l'on se réfère aux valeurs reportées dans le Tableau 7, on obtient la relation suivante : $\Delta G_{\text{boucle}}(\text{UC},11) = \Delta G_{\text{boucle}}(\text{GC},9)$.

IV.c.ii. Discussion-conclusion

Nous avons montré au cours de cette étude que la transition entre les états ouvert et fermé de la paire de bases G7-C17 à la jonction tige-boucle peut se faire assez facilement dans la mesure où elle ne demande pas une grande énergie d'activation. Les fluctuations thermiques de l'énergie d'interaction soluté-solvant peuvent suffire à la provoquer. Cette hypothèse est confirmée par nos simulations de dynamique moléculaire de la structure de Mujeeb et al. (1998), dans lesquelles nous remarquons un appariement systématique et spontané d'une paire de bases G7-C17 et par notre simulation de la structure caractéristique de la famille 2 de Kieken (2004) où une paire G7-C17 préexistante est rompue. Ces résultats sont en accord avec les résultats de Lodmell et al. (2001) et de Kieken (2004) suggérant une grande flexibilité de cette paire de base. Du fait que la différence d'énergie libre entre ces deux états soit approximativement nulle ils sont équiprobables. Ceci peut expliquer que les différentes structures parues dans la littérature présentent des différences radicales au niveau de cette paire de bases.

Bien sûr les paramètres thermodynamiques déterminés ici ne s'appliquent qu'au système étudié. Il faut donc tenir compte du fait que l'on a une structure en « kissing-complex » avec une séquence particulière d'ARN, ainsi que du fait que sur l'un des monomères, la paire de bases G7-C17 est déjà appariée. Cependant les résultats que nous obtenons sont en bon accord avec les valeurs expérimentales obtenues sur des ARN ou des ADN. En effet, il a été déterminé que la valeur de ΔG pour ajouter une paire de bases quelconque à une double hélice d'ARN terminée par une paire de bases G-U est de 0 kcal/mol (Cantor and Schimmel 1980). Bien sûr notre paire de bases G-U ne fait pas partie d'une double hélice d'ARN mais clôt la tige d'une tige-boucle. Cela semble suggérer que bien que les boucles soient appariées entre elles par 6 paires de bases, elles se révèlent suffisamment flexibles pour permettre l'appariement de la paire G-C à la jonction tige-boucle. Cette hypothèse est soutenue par la très

faible valeur de ΔS traduisant ainsi une légère déstabilisation lors de la fermeture de G-C, compensée par une faible enthalpie de stabilisation. Nous déduisons de la relation de la thermodynamique classique $\Delta G = -R \cdot T \cdot \log K$ que la constante d'équilibre K de cette transition est de 1. Il semble donc qu'en l'absence de contraintes extérieures ces deux résidus ont autant de chance de se trouver en état apparié que désapparié, ce qui revient à considérer que cette paire de bases est formée la moitié du temps et désappariée l'autre moitié.

Il convient cependant d'étudier de plus près l'incertitude sur le calcul de ΔH . L'observation des distributions normales d'énergie révèle une largeur à mi-hauteur pour chacune de ces courbes de l'ordre de 55 kcal/mol, ce qui semble élevé comparé à la valeur de ΔH de -11 kcal/mol. Il est pourtant possible de mesurer cette incertitude. Pour les calculs effectués précédemment, nous avons utilisé les 1100 conformations générées par les simulations de dynamique moléculaire dans l'état initial et dans l'état final avec une résolution de 1 kcal/mol sur les histogrammes. Si l'on change cette résolution, à 2 ou 5 kcal/mol par exemple, ou si l'on divise l'échantillonnage de 1100 configurations en deux sous-ensembles de 550 configurations chacun, nous obtenons des valeurs moyennes pour ΔH comprises entre -10 kcal/mol et -12 kcal/mol. La position du maximum de chaque distribution gaussienne est déterminée avec une incertitude de 2 kcal/mol dans le processus de Marquardt (1963), ce qui conduit à 4 kcal/mol pour un couple de gaussiennes. Si l'on tient compte de l'ensemble des valeurs ΔH que nous avons obtenues, nous pouvons donc estimer que $\Delta H = -11 \pm 3$ kcal.mol⁻¹ et $\Delta S = -0.037 \pm 0.009$ kcal.mol⁻¹K⁻¹. Les résultats que nous présentons pour ΔH et ΔS sont du même ordre de grandeur que les valeurs publiées sur les appariements des différentes paires de bases dans les ADN (Cantor and Schimmel 1980).

Cette étude nous permet également de mesurer la valeur de $\Delta G(UG,11)$ c'est à dire l'énergie de formation d'une boucle d'ARN de 11 nucléotides fermée par une paire de bases Wobble G-U, paire de base souvent retrouvée dans les structures d'ARN et contribuant de plus au repliement ou aux fonctions des ARN (Lim 1994; Strobel et

al. 1994). Nous montrons dans cette partie que $\Delta G(UG,11)$ a la même valeur que $\Delta G(GC,9)$ qui est l'énergie de formation d'une boucle de 9 nucléotides close par une paire G-C. Cette valeur a été déterminée de +5 kcal/mol (Tableau 7). Ces résultats nous conduisent à penser que la perte de stabilité induite par une paire U-G comparativement à une paire G-C est dans ce cas compensée par l'accroissement de l'entropie de la boucle, du fait que l'on passe de 9 nucléotides dans la boucle à 11.

IV.d.Transition « kissing-complex » → dimère étendu: application de la dynamique dirigée sur un changement conformationnel d'importance

IV.d.i. Objectif du travail et choix de la méthode

Une fois achevée l'analyse de la stabilité des structures des deux formes dimériques que peut adopter la séquence SL1, notre but est d'étudier le chemin conformationnel qu'emprunte le dimère pour effectuer la transition « kissing-complex » → dimère étendu. Pour cela, nous avons résolu d'utiliser les techniques particulières de simulation de dynamique moléculaire que sont la dynamique moléculaire dirigée (TMD) et la dynamique moléculaire dirigée inverse (TMD⁻¹).

Le problème majeur d'une telle étude réside en effet dans le temps que prendrait la simulation si l'on envisageait de l'effectuer en dynamique moléculaire sans contraintes. Malgré l'élaboration d'algorithmes de plus en plus performants et l'évolution des moyens de calcul, les simulations de dynamique moléculaire ne permettent encore d'appréhender le comportement d'un système de cette taille que sur des intervalles de temps de quelques dizaines de nanosecondes au plus. Or des réactions telles que cette transition se produisent vraisemblablement sur des périodes de temps de l'ordre de la centaine de millisecondes, voire de la seconde que nous ne sommes pas à même de simuler sans contraintes. Il nous faut donc utiliser des procédés de simulation numérique particuliers afin d'accélérer le processus de transition.

Un autre problème concerne le choix de coordonnées de réaction pour contraindre une transition faisant intervenir tant de degrés de liberté. Un des avantages de la TMD est qu'elle n'impose de contraintes que sur le RMSD entre la structure active et la structure de référence. Tous les autres paramètres sont laissés libres d'évoluer à leur gré. Il est capital pour que nos simulations conduisent à un résultat plausible que les contraintes imposées sur le système soient les moins fortes possibles.

IV.d.ii. Glossaire des dénominations utilisées dans ces expériences et particularités d'analyse

→ Dénomination des structures initiales et finales utilisées

Au cours de ce travail nous nommerons KC tout ce qui se rapporte au « kissing-complex » et DE tout ce qui se rapporte au 'dimère étendu'.

Trois structures initiales du KC sont utilisées au cours de ces expériences, les deux structures de « kissing-complex » déterminées par Ennifar et al. (2001) ainsi que la structure RMN publiée par Mujeeb et al. (1998) sur la séquence mutée. Par commodité, nous dénommerons X-ray_KC_{Mal} et X-ray_KC_{Lai} les structures cristallographiques issues de séquences SL1_{Mal} et SL1_{Lai} respectivement et NMR_KC_{LaiΔ} et NMR_KC_{Lai} la structure déterminée par Mujeeb et al. (1998) avec la séquence mutée SL1Δ et avec la séquence sauvage SL1_{Lai} respectivement. Les coordonnées de cette dernière structure ont été obtenues à partir du fichier PDB du KC isolé par Mujeeb et al. (1998) et la mutation de séquence a été effectuée à l'aide d'un programme développé au laboratoire nommé MUTATION. Trois structures de DE ont également été utilisées à titre de structure finale dans ces expériences. L'une de ces structures est la structure cristallographique publiée par Ennifar et al. (1999) à partir de la séquence SL1_{Mal} que nous nommerons X-ray_DE_{Mal}, les deux autres sont issues de l'ensemble de 15 structures du DE publié par Girard et al. (1999) et déterminées par RMN à partir de la séquence SL1_{Lai}. Les structures nommées NMR_DE_{Lai_1} et NMR_DE_{Lai_3} correspondent respectivement à la structure la plus « éloignée » et à la structure la plus proche de la structure moyenne de l'ensemble, le numéro de ces structures se référant à leurs positions dans le fichier PDB.

→ La température

La plupart des TMD et TMD⁻¹ sont effectuées à 360K, d'une part car une température plus élevée permet en dynamique d'accélérer les processus de modification structurale, d'autre part du fait que la transition « kissing-complex » → dimère étendu s'accomplit par accroissement de température *in vitro*.

→ Paramètres spécifiques de la TMD

On désigne par δp la variation de RMSD imposée au système pendant un temps Δt donné. Dans certaines simulations, on effectue des étapes d'équilibration plus longues à certaines valeurs de rmsd.

Au cours de ces expériences les contraintes ne sont imposées que sur les atomes de phosphore du squelette phospho-diester de l'ARN, afin de restreindre le minimum de degrés de liberté de la molécule. La séquence SL1 étant un 23-mer, nous avons 44 phosphores. Pour certaines de ces simulations, on a exclu du nombre de phosphores contraints les trois premières paires de bases des tiges du KC afin de diminuer encore les contraintes imposées aux systèmes. Ce groupe est dénommé 34P par la suite, l'autre 44P. Dans une seule de nos simulations, nous avons considéré tous les phosphores de la molécule exception faite des 12 phosphores des boucles ; c'est le groupe 32P.

→ Spécificité des analyses

Les rmsd ne sont calculés que sur le nombre de phosphores contraints (32P, 34P ou 44P). Nous noterons toujours entre parenthèses la structure par rapport à laquelle le rmsd est calculé. Pour simplifier la description des évolutions structurales de la molécule entre KC et DE, on découpe la molécule en groupes, différents selon la structure, et présentés sur la Figure 91.

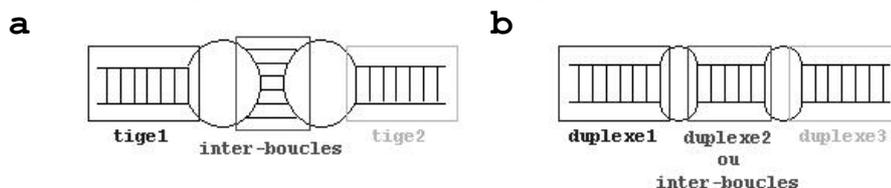


Figure 91 : Division des dimères en trois groupes distincts : a) pour le « kissing-complex » on appelle tige1 les résidus {1→7, 17→23}, inter-boucles les résidus {10→5, 10'→15'} et tige2 les résidus {1'→7', 17'→23'} b) pour le dimère étendu, on appelle duplexe 1 les résidus {1→7, 17'→23'}, inter-boucles ou duplexe 2 les résidus {10→5, 10'→15'} et duplexe 3 les résidus {17→23, 1'→7'}

Par souci de clarté ne sont présentées sur les graphiques de liaisons hydrogène que les moyennes sur l'ensemble des liaisons hydrogène canoniques présentes dans ces groupes.

IV.d.iii. Transition par TMD

La première approche que nous avons envisagée pour mener à bien ce travail a été d'utiliser la TMD entre le dimère instable et le dimère stable.

Notre première étape a été de déterminer des valeurs raisonnables pour les paramètres entrant en jeu dans la TMD ainsi que nous l'avons mentionné dans le §III.b.ii et tels qu'ils apparaissent dans la formule (23), à savoir la valeur de la constante de force k_{TMD} , la vitesse de variation de rmsd(DE) ($\delta\rho/\Delta t$) et les éventuelles étapes d'équilibration. Il est à noter que dans les expériences numériques parues dans la littérature et utilisant cette méthode, aucune à notre connaissance ne fait état d'un changement conformationnel de cette importance (Apostolakis et al. 1999; Diaz et al. 1997; Ferrara 2000a; Kruger et al. 2001; Ma and Karplus 1997; Mendieta et al. 2004; Roche and Field 1999; Schlitter et al. 1994). La transition « kissing-complex » → dimère étendu présente en effet un rmsd moyen entre les structures initiale et finale d'environ 20-22 Å et fait intervenir un réarrangement dans l'appariement des brins. En nous basant sur le travail de Ferrara et al. (2000a), nous avons réalisé nos premières TMD en utilisant les valeurs de paramètres suivants :

- structure initiale : X-ray_KC_{Mal}; ce choix s'est imposé du fait que l'on ne disposait à ce moment là comme structure de DE que de la structure cristallographique de l'isolat Mal
- structure finale : X-ray_DE_{mal}
- rmsd entre les structures initiale et finale de 20 Å
- groupe de contraintes 44P
- $k_{TMD}=25 \text{ kcal/mol.Å}^2$
- $\delta\rho=0.2 \text{ Å}$ pour $\Delta t=3 \text{ ps}$. Sachant que l'on a décidé de faire évoluer rmsd(DE) de sa valeur initiale de 20 Å à 2 Å, on va réaliser ainsi une TMD de 270 ps.

Cette première tentative s'est soldée par un échec. En effet les premières simulations présentent bien un $\text{rmsd}(\text{DE})$ calculé (ρ) proche de celui demandé (ρ_0) et on voit une ouverture des paires de bases des tiges avec rapprochement des différents brins, mais on a rapidement rupture des paires de bases inter-boucles ainsi qu'un enroulement des brins sur eux-mêmes allant jusqu'à générer des nœuds dans la structure. Afin de vérifier l'influence des différents paramètres sur ces résultats de TMD, nous avons effectué dix simulations utilisant des valeurs différentes pour les paramètres et partant des diverses structures expérimentales du KC. Toutes ont conduit à la formation de nœuds dans la structure générant de trop grandes énergies de contrainte pour permettre la bonne continuation de la simulation. A titre d'exemple, je présente ici les résultats de deux de ces simulations.

Ces deux TMD ont été réalisées à partir de X-ray_KC_{Lai}. Cette dernière structure semble en effet plus simple à déstabiliser car une simple production de dynamique moléculaire à 300 K provoque une destruction du KC (ouverture des paires de bases des tiges) tout en conservant le caractère dimérique de l'ARN (préservation des paires de bases boucle-boucle). Les différents paramètres utilisés dans ces simulations sont répertoriés dans le Tableau 8.

| | TMD1 | TMD2 |
|---|-------------------------|------------------------------|
| Structure initiale | X-ray_KC _{Lai} | X-ray_KC _{Lai} |
| Structure finale | NMR_DE _{Lai_1} | NMR_DE _{Lai_3} |
| k_{TMD} (kcal/mol.Å ²) | 5 | 25 |
| $\delta\rho/\Delta t$ | 0.05Å/2ps | 0.1Å/2.5ps |
| équilibérations intermédiaires | - | 100ps / 16, 12, 8, 4 et 0Å * |
| Evolution de RMSD demandée (ρ_0) | de 21Å à 1Å | de 20Å à 0Å |

Tableau 8 : Paramètres utilisés pour effectuer deux TMD directes entre le « kissing-complex » et le dimère étendu.

* Pendant la période de temps indiquée, ρ_0 est maintenu constant aux valeurs indiquées (en Å)

La Figure 92 présente les évolutions temporelles des $\text{rmsd}(\text{DE})$ réels ($\rho(\text{DE})$) calculés au cours de la simulation et des $\text{rmsd}(\text{DE})$ demandés

($\rho_0(\text{DE})$). On peut remarquer que $\rho(\text{DE})$ et $\rho_0(\text{DE})$ restent relativement proches jusqu'aux alentours de 6.5-7 Å.

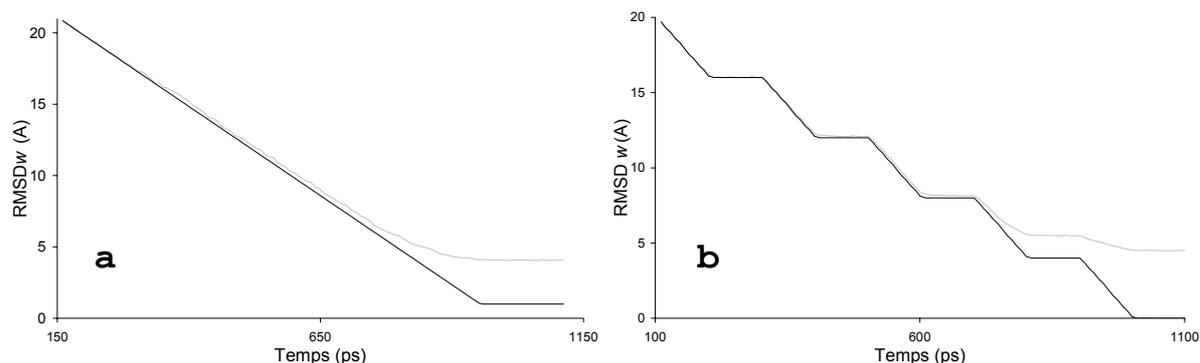


Figure 92 : Evolution en fonction du temps de $\rho(\text{DE})$ (en gris) et $\rho_0(\text{DE})$ (en noir) par rapport à la structure cible pour les TMD directes 1 (a) et 2 (b).

En deçà de ces valeurs, la molécule ne parvient plus à se rapprocher de la structure cible.

Les valeurs des énergies de contrainte sont reproduites en fonction du $\rho_0(\text{DE})$ sur la Figure 93. On peut constater sur ces graphiques que les valeurs de ces énergies restent globalement en dessous de 50 kcal/mol pendant la première partie de la dynamique moléculaire dirigée puis augmentent rapidement de façon très importante aux alentours de 7-8 Å. Ces valeurs traduisent la présence dans cette région conformationnelle d'une barrière de potentiel impossible à franchir.

Les 'snapshots' des conformations de la molécule aux alentours de ces valeurs de $\rho_0(\text{DE})$ révèlent en effet la formation de « nœuds » dans le squelette phospho-diester bloquant la transition conformationnelle du dimère. Ces nœuds perdurent jusqu'à la fin de la simulation, donnant à la molécule une conformation déstructurée telle que l'on peut le voir sur la Figure 94. On remarque également sur la trajectoire de la seconde dynamique dirigée un phénomène surprenant : la force de contrainte aux alentours de 8 Å est telle qu'elle entraîne la rupture du squelette phospho-diester afin de laisser passer un brin à travers l'autre. Ce phénomène est montré sur la Figure 95.

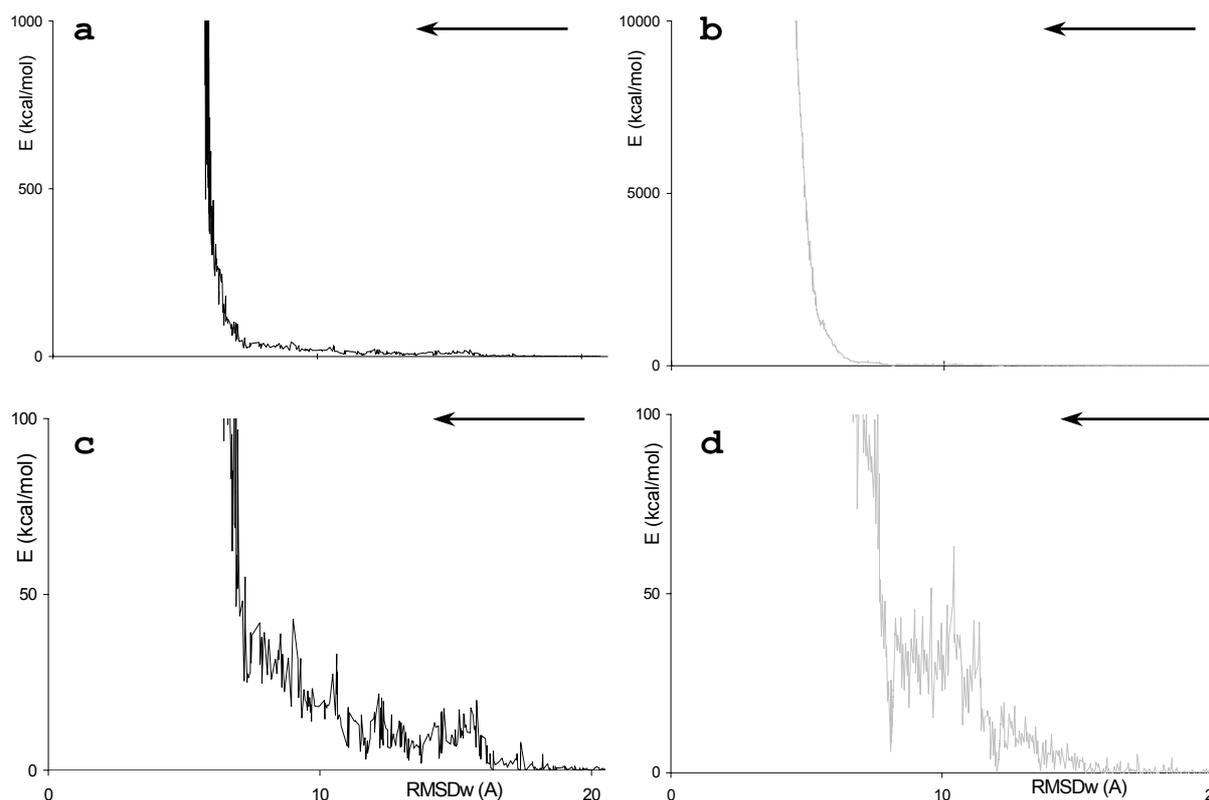


Figure 93 : Evolution de l'énergie de contrainte subie par la molécule en fonction de $\rho_0(DE)$ pendant la TMD1 (a et c) et la TMD2 (b et d). Ces énergies sont représentées à des échelles différentes (a et b de 0 à 1000 kcal/mol ; c et d de 0 à 100 kcal/mol) afin de mettre en évidence les valeurs de $\rho_0(DE)$ pour lesquelles les courbes augmentent brusquement. La flèche indique le sens de la TMD.

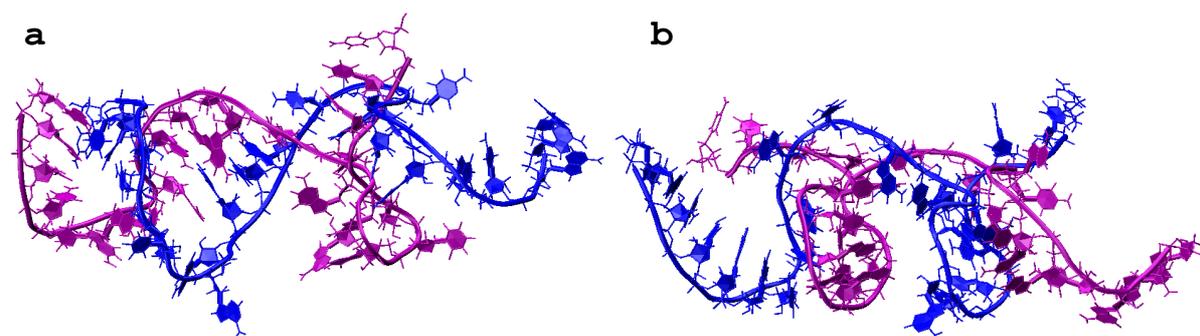


Figure 94 : Structure finale obtenue pour la TMD1 (a) et pour la TMD2 (b). Les monomères 1 et 2 sont représentés en bleu et violet respectivement.

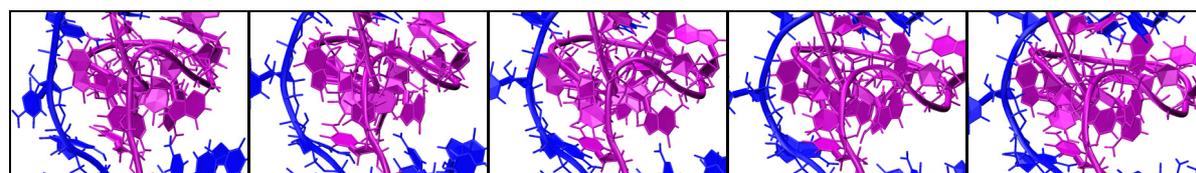


Figure 95 : Snapshots réalisés sur la TMD2 montrant la coupure d'un squelette du monomère 2 (violet) au cours de la simulation.

On a effectué parallèlement l'étude des pourcentages de présence des liaisons hydrogène, dont les courbes sont présentées sur la Figure 96. Pour les deux simulations, nous observons une ouverture complète des tiges avant d'avoir atteint un $\rho_0(\text{DE})$ de 10 Å et cette ouverture perdure jusqu'à la fin des simulations.

Le comportement des boucles diffère pour les deux simulations ; la simulation TMD1 conserve 3 paires de bases fermées sur les 6 originelles, les paires de bases G10-C38, C11-G37 et G12-C36. En revanche, les boucles se désappariaient totalement au cours de la TMD, sans doute du fait, comme expliqué ci-dessus, qu'au cours de la dynamique on a un phénomène de passage au travers du brin 2 de la molécule.

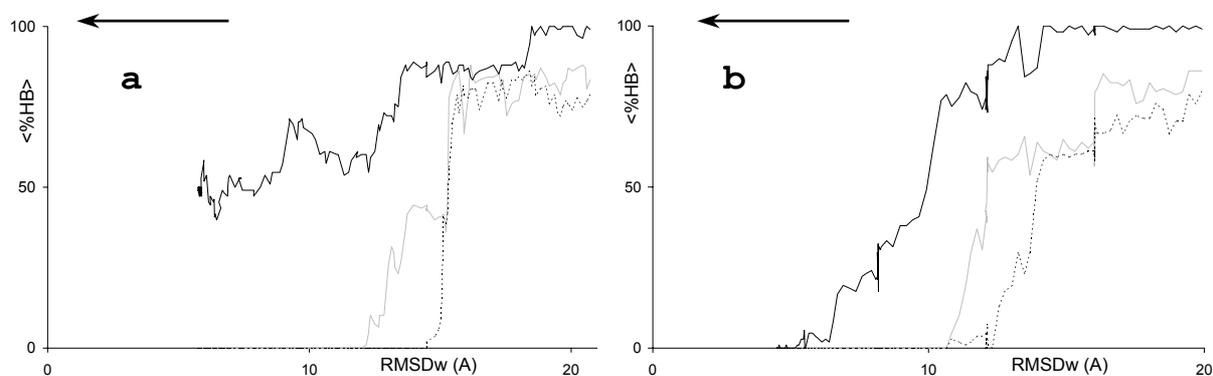


Figure 96 : Evolution des pourcentages moyens des liaisons hydrogène canoniques de la tige 1 (gris), de la tige 2 (pointillés) et inter-boucles (noir) pour les TMD 1 (a) et 2 (b). La flèche indique le sens de la TMD.

La TMD directe pour la transition entre les structures en KC et en DE conduit donc systématiquement à un échec. Il a de ce fait fallu envisager une autre méthode pour étudier la transition conformationnelle du dimère de la tige-boucle SL1.

IV.d.iv.TMD⁻¹ : mise en évidence d'une famille d'intermédiaires

Face à la difficulté rencontrée par le KC pour effectuer la transition par la méthode TMD, nous avons envisagé le problème inverse : au lieu de contraindre la molécule à se rapprocher de la structure en DE, nous allons la contraindre à « s'éloigner » de la structure en KC sans imposer de direction privilégiée. Les rmsd utilisés dans cette étude sont donc relatifs à la structure en KC.

Nous avons ainsi réalisé 20 simulations de TMD^{-1} dans des conditions différentes ; les conditions utilisées lors de ces expériences sont répertoriées dans le Tableau 9.

| Simulations de dynamique dirigée inverse | | | | | | | | | Prolongations | | |
|--|-------------------------|--------------------|------------------------------|--------------|-----------|-----|-----|------------|---------------|------------------------------|------------|
| Nom de la TMD^{-1} | Structure initiale | Boîte d'eau | $\delta\rho/\Delta t$ (Å/ps) | équil (ps/Å) | k_{TMD} | NPC | T | RMSD final | Boîte d'eau | $\delta\rho/\Delta t$ (Å/ps) | RMSD final |
| elaiI | X-ray_KC _{Lai} | tétra 6513/7287 | 0.2/3 0.13/3 | 90/18 | 25 | 44P | 360 | 22Å | octa 16369 | 0.13/3 | 40Å |
| elaiII | X-ray_KC _{Lai} | tétra 6513 | 0.2/3 0.13/3 | - | 25 | 44P | 360 | 22Å | octa 15844 | 0.13/3 | 36Å |
| elaiIII | X-ray_KC _{Lai} | tétra 6513 | 0.2/3 0.13/3 | - | 25 | 44P | 360 | 22Å | octa 15708 | 0.13/3 | 36Å |
| elaiIV | X-ray_KC _{Lai} | tétra 6513 | 0.17/3 0.1/3 | 90/15 | 25 | 44P | 360 | 18Å | octa 16118 | 0.13/3 | 36Å |
| elaiV | X-ray_KC _{Lai} | tétra 6513 | 0.2/3 0.13/3 | - | 5 | 44P | 360 | 22Å | octa 16075 | 0.13/3 | 36Å |
| elaiVI | X-ray_KC _{Lai} | tétra 6513 | 0.05/2.5 0.03/2.5 | - | 5 | 44P | 360 | 22Å | - | - | - |
| elaiVII | X-ray_KC _{Lai} | tétra 6513 | 0.2/3 0.13/3 | - | 100 | 44P | 360 | 22Å | - | - | - |
| elaiVIII | X-ray_KC _{Lai} | tétra 6513 | 0.2/3 0.13/3 | - | 25 | 44P | 300 | 22Å | - | - | - |
| elaiIX | X-ray_KC _{Lai} | tétra 6513 | 0.5/25 0.75/25 | - | 0.05 | 44P | 360 | 36Å | - | - | - |
| elaiX | X-ray_KC _{Lai} | octa 20935 | 0.5/25 | - | 0.05 | 44P | 360 | 36Å | - | - | - |
| mjI | NMR_KC _{LaiA} | tétra 6331/6987 | 0.2/3 0.13/3 | 90/18, 22 | 25 | 44P | 360 | 26Å | - | - | - |
| mjII | NMR_KC _{LaiA} | tétra 6331. | 0.2/3 0.13/3 | 90/22, 28 | 25 | 44P | 360 | 28Å | - | - | - |
| mjIII | NMR_KC _{LaiA} | tétra 6331 | 0.2/3 0.13/3 | 90p/22 | 25 | 44P | 360 | 28Å | - | - | - |
| mjIV | NMR_KC _{LaiA} | tétra 6987 | 0.2/5 | - | 25 | 44P | 360 | 20Å | octa 15166 | 0.2/5 | 36Å |
| mjV | NMR_KC _{LaiA} | tétra 6987 | 0.2/10 0.1/5 | - | 25 | 44P | 360 | 20Å | octa 14980 | 0.1/5 | 36Å |
| mjlaiI | NMR_KC _{Lai} | octa 15242 | 0.2/5 | - | 25 | 44P | 360 | 20Å | octa 14980 | 0.2/5 | 36Å |
| mjlaiII | NMR_KC _{Lai} | tétra 7193 | 0.2/5 | - | 25 | 44P | 360 | 20Å | octa 16400 | 0.2/5 | 36Å |
| mjlaiIII | NMR_KC _{Lai} | octa 15242 | 0.5/25 | - | 25 | 44P | 360 | 36Å | - | - | - |
| mjlaiIV | NMR_KC _{Lai} | octa 15242 | 36/2000 (chgt continu) | - | 20 | 44P | 360 | 36Å | - | - | - |
| mjlaiV | NMR_KC _{Lai} | octa 15242 | 0.5/25 | - | 25 | 32P | 360 | 36Å | - | - | - |

Tableau 9 : Conditions de simulation pour les 20 TMD^{-1} . Nous désignons par « tétra » ou « octa » les boîtes respectivement tétraédriques et octaédriques, les chiffres suivant ces désignations étant le nombre de molécules d'eau contenues dans la boîte. Dans le cas d'un changement de boîte tétraédrique, les chiffres sont séparés par un slash. La température T est donnée en K, la constante k_{TMD} en kcal/mol.Å². NPC désigne le nombre de phosphores contraints

L'une des idées que nous avons à l'esprit en réalisant des TMD^{-1} à partir du KC était que la molécule pourrait éventuellement prendre d'elle-même directement le chemin la menant vers le DE. C'est pourquoi, bien que cette probabilité paraisse de prime abord très faible, nombre des simulations de TMD^{-1} que nous avons d'abord réalisées ont été arrêtées à $\rho_0(KC)$ final = 22 Å, cette valeur étant

à peu près identique au rmsd entre les structures en KC et les structures en DE. Au vu des résultats que nous avons obtenus nous avons prolongé certaines de ces TMD⁻¹ au-delà de 22 Å.

La première prolongation devait être réalisée jusqu'à $\rho_0(KC) = 44\text{\AA}$, le double de la valeur utilisée précédemment, et n'a en pratique été réalisée que jusqu'à 40 Å. On a en effet constaté qu'à cette valeur les deux brins du dimère étaient totalement séparés. C'est pourquoi on a choisi une valeur finale inférieure pour les autres TMD, soit $\rho_0(KC) = 36\text{\AA}$.

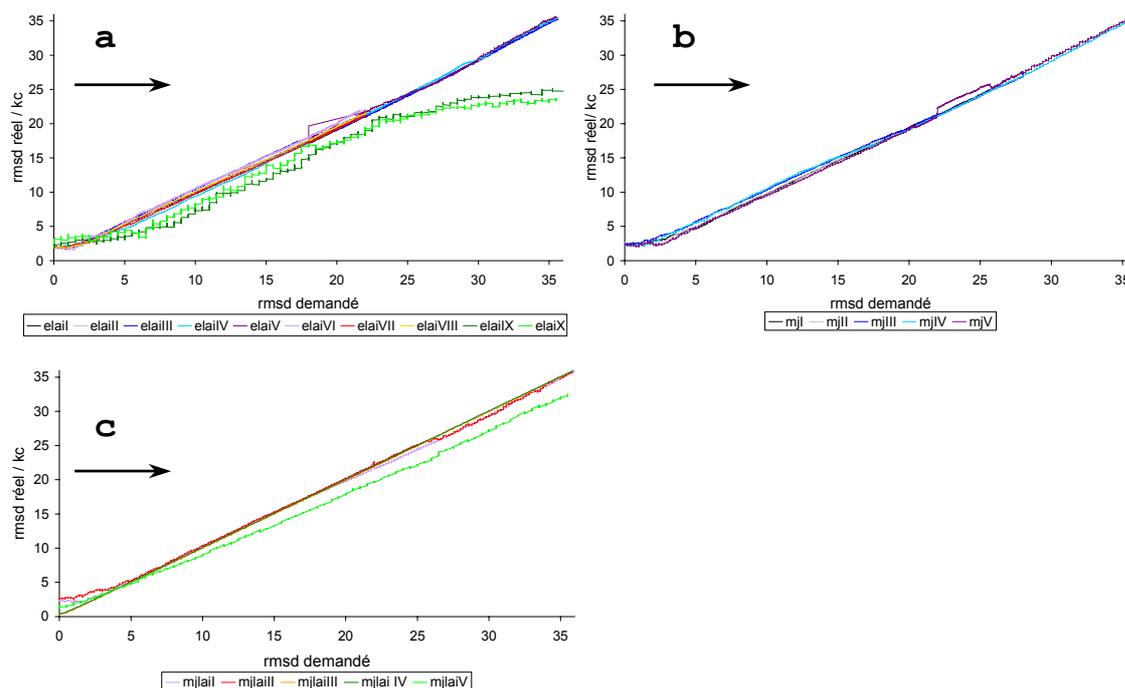


Figure 97 : Evolution des valeurs de $\rho(KC)$ en fonction $\rho_0(KC)$ pour les TMD⁻¹ à partir des structures initiales X-ray_KC_{Lai} (a), NMR_KC_{LaiΔ} (b) et NMR_KC_{Lai} (c). La flèche noire donne le sens de la simulation.

| TMD inverse | Coefficient de corrélation |
|-------------|----------------------------|
| elaiI | 0.9988 |
| elaiII | 0.9994 |
| elaiIII | 0.9990 |
| elaiIV | 0.9990 |
| elaiV | 0.9991 |
| elaiVI | 0.9996 |
| elaiVII | 0.9992 |
| elaiVIII | 0.9996 |
| elaiIX | 0.9882 |
| elaiX | 0.9810 |
| mjI | 0.9991 |
| mjII | 0.9992 |
| mjIII | 0.9997 |
| mjIV | 0.9992 |
| mjV | 0.9985 |
| mjlaiI | 0.9995 |
| mjlaiII | 0.9993 |
| mjlaiIII | 1.0000 |
| mjlaiIV | 1.0000 |
| mjlaiV | 0.9993 |

Tableau 10 : Coefficients de corrélation entre les valeurs de $\rho(KC)$ et $\rho_0(KC)$ pour les 20 TMD⁻¹

La Figure 97 représente la valeur $\rho(KC)$ calculé à chaque instant t en fonction de la valeur de $\rho_0(KC)$ à cet instant. On peut remarquer sur ces courbes que $\rho(KC)$ reste globalement bien corrélé avec $\rho_0(KC)$, sauf en ce qui concerne deux simulations, les simulations elaiIX et elaiX. On

voit sur le Tableau 10 qu'elles ont les coefficients de corrélation les plus faibles de l'ensemble des simulations et on peut constater que, si les valeurs de $\rho(KC)$ suivent relativement bien celles de $\rho_0(KC)$ jusqu'à environ 22 Å, elles stagnent par la suite aux environs de cette valeur. Ces deux simulations ayant été réalisées avec une constante de force de 0.05 kcal/mol.Å², il semble que celle-ci soit trop faible pour permettre le franchissement d'une barrière de potentiel située aux environs de $rmsd(KC)=26-28$ Å. Ce résultat est confirmé par l'étude des énergies de contraintes en fonction de la valeur de $\rho_0(KC)$, représentées sur la Figure 98.

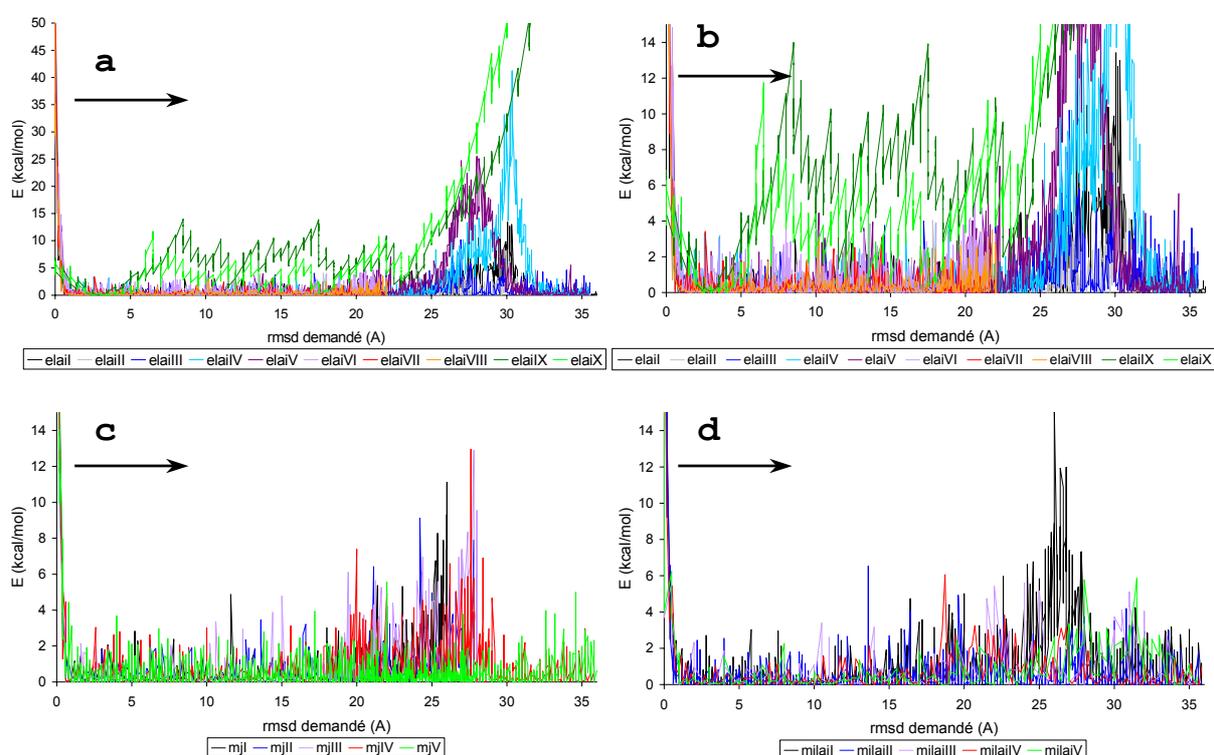


Figure 98 : Evolution des valeurs des énergies de contraintes en fonction du $\rho_0(KC)$ pour les TMD⁻¹ à partir de la structure initiale X-ray_KC_{Lai} (**a** et **b**), NMR_KC_{Lai} Δ (**c**) et NMR_KC_{Lai} (**d**). Les graphiques **a** et **b** présentent les mêmes courbes mais à des échelles différentes. La flèche noire donne le sens de la simulation.

Les deux premiers graphiques représentent l'évolution des énergies de contraintes en fonction de $\rho_0(KC)$ pour les dix premières simulations présentées dans le Tableau 9 et désignées par le nom générique « elai » mais à des échelles différentes. On voit nettement la présence entre 25 Å et 32 Å d'une barrière de potentiel, du fait d'une hausse brusque de la valeur de l'énergie de

contrainte dans cette région. Cette constatation est accentuée par l'observation des énergies de contrainte des simulations elaiIX et elaiX qui, bien que présentant des valeurs supérieures aux autres courbes dès le début des simulations, restent dans des valeurs d'énergies inférieures à 15 kcal/mol jusqu'aux environ de $\rho_0(KC)=22 \text{ \AA}$. A partir de ce point, les énergies de contrainte augmentent rapidement jusqu'à atteindre des valeurs très au-dessus de la normale, ce qui s'explique par l'existence de barrières de potentiel que la molécule ne peut franchir. Cette barrière est nettement moins visible si l'on se réfère à l'évolution des énergies de contrainte issues des simulations à partir de la structure RMN. En effet, on peut constater sur quelques-unes d'entre elles un faible pic autour de 25-28 \AA , encore que les valeurs restent très en dessous de 15 kcal/mol. \AA^2 . Il convient de noter toutefois que l'on ne part pas de la même structure de KC et que les valeurs de $\rho_0(KC)$ reportées en abscisse ne sont pas tout à fait comparables, c'est à dire qu'une structure à 22 \AA par rapport à la structure RMN peut très bien ne se trouver qu'à 18 \AA par rapport à la structure cristallographique ou au contraire à 25 \AA .

La Figure 100 présente, pour $\rho_0(KC)=22 \text{ \AA}$ et $\rho_0(KC)=36 \text{ \AA}$, les différentes structures obtenues au cours des 20 simulations de TMD^{-1} . La visualisation de ces structures montre déjà un comportement global différent selon que l'on considère l'ensemble des structures à 22 \AA ou l'ensemble de celles à 36 \AA . A 22 \AA , on voit que les structures ont toutes grossièrement la même forme, celle d'un double U plus ou moins déformé, les tiges étant déroulées et se faisant face, les boucles restant globalement appariées.

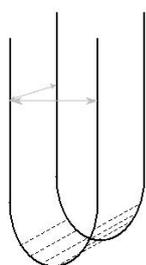
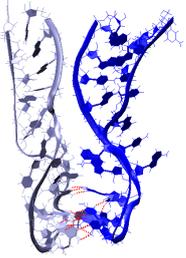
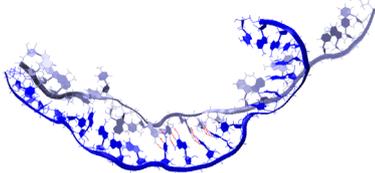
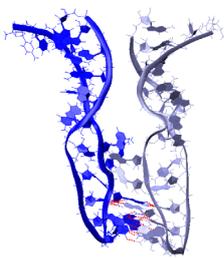
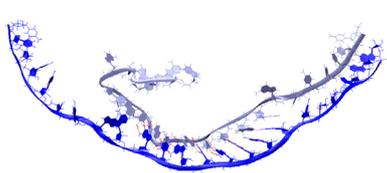
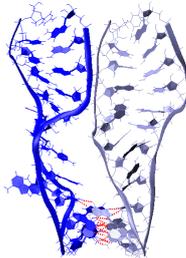
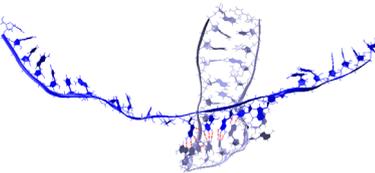
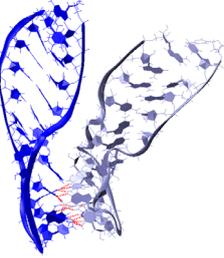
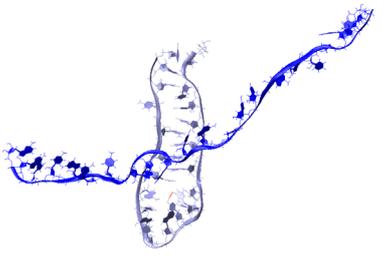
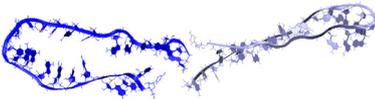
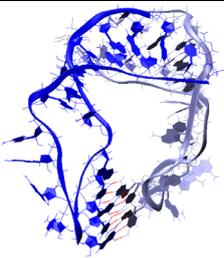
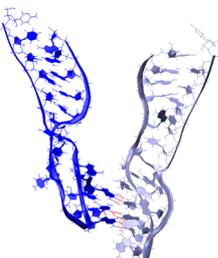
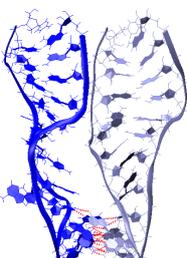
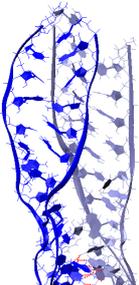


Figure 99 : Illustration de la structure en double U adoptée par les conformations des TMD^{-1} à $\rho_0(KC) = 22 \text{ \AA}$

Les distances intramoléculaires et intermoléculaires entre les deux branches des U sont à peu près équivalentes, comme illustré sur la Figure 99. Cette structure semble propice à un échange de liaisons hydrogène des brins intramoléculaires aux brins intermoléculaires.

| elaiI | | elaiII | |
|---|---|--|---|
|  22 / 21.01 |  36 / 35.39 |  22 / 21.21 |  36 / 35.33 |
| elaiIII | | elaiIV | |
|  22 / 20.96 |  36 / 35.15 |  22 / 21.44 |  36 / 35.28 |
| elaiV | | elaiVI | |
|  22 / 21.70 |  36 / 35.49 |  22 / 21.95 | - |
| elaiVII | elaiVIII | | elaiIX |
|  22 / 21.40 |  22 / 21.38 |  22 / 18.40 |  36 / 24.73 |

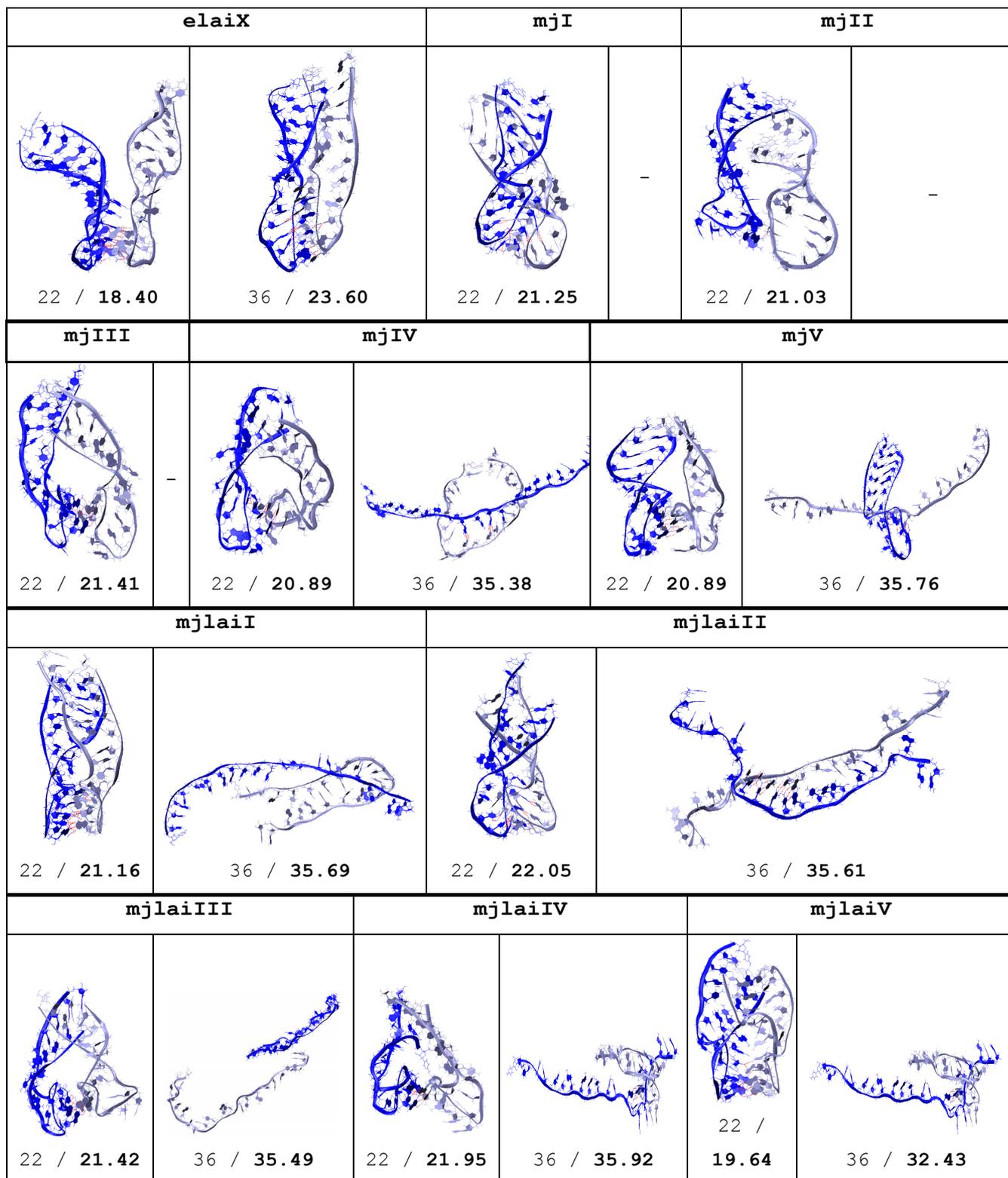


Figure 100 : Ensemble des structures obtenues à $\rho_0(KC)=22 \text{ \AA}$ et $\rho_0(KC)=36 \text{ \AA}$ pour les 20 TMD^{-1} . Les $\rho(KC)$ (en gras) et $\rho_0(KC)$ sont reporté en dessous de chacune des structures et donnés en \AA

Les structures obtenues pour $\rho_0(\text{KC})=36 \text{ \AA}$ divergent sensiblement entre elles ; quelques-unes d'entre elles nous ont cependant semblé montrer une évolution intéressante, du fait qu'elles forment deux brins antiparallèles allongés reliés entre eux par les liaisons hydrogène initialement présentes entre les paires de bases des boucles. Cette conformation paraît aussi susceptible de conduire à la formation par la suite d'une double hélice. Toutefois, cette structure n'est trouvée que pour trois des quatorze simulations menées jusqu'à cette valeur de $\rho_0(\text{KC})$. Les autres structures obtenues pour $\rho_0(\text{KC})=36 \text{ \AA}$ présentent des configurations beaucoup plus insolites, avec parfois un brin déroulé alors que l'autre reste en forme de U, des entortillements de brins occasionnant la rupture de certaines des liaisons hydrogène voire le désappariement total des paires de bases inter-boucle entraînant la dissociation du dimère en deux molécules distinctes.

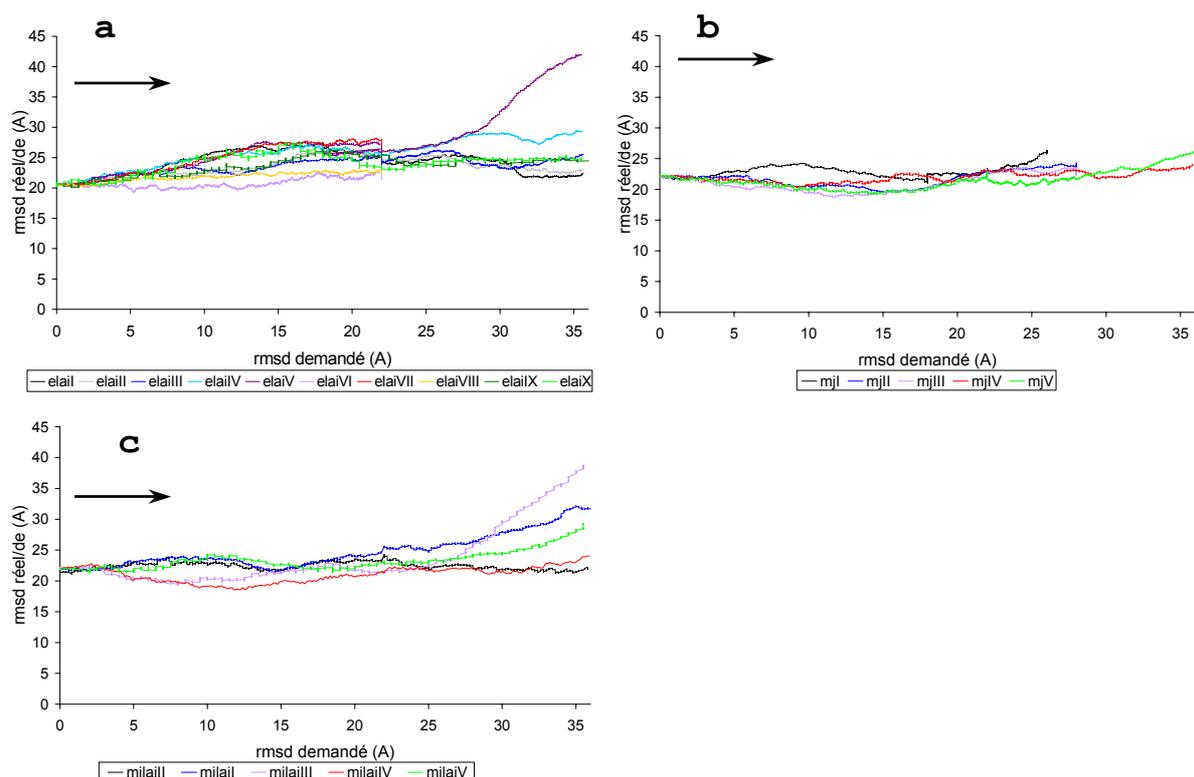


Figure 101 : Evolution du $\rho(\text{DE})$ en fonction du $\rho_0(\text{KC})$ pour les TMD⁻¹ à partir des structures initiales X-ray_{KC_{Lai}} (a), NMR_{KC_{LaiA}} (b) et NMR_{KC_{Lai}} (c). La flèche noire donne le sens de la simulation.

Toutes les structures observées sont assez dissemblables et du KC, et du DE. Afin d'obtenir des informations conformationnelles

supplémentaires, nous avons calculé l'évolution de $\rho(\text{DE})$ en fonction de $\rho_0(\text{KC})$ (Figure 101). On constate que dans la majorité des simulations, $\rho(\text{DE})$ reste compris entre 18 Å et 30 Å. Les structures obtenues par les TMD⁻¹ sont à peu près équidistantes de KC et de DE. Seules trois d'entre elles dépassent 30 Å en fin de simulation. On remarque que cette fois encore les divergences de structures interviennent plutôt au-delà de $\rho_0(\text{KC})=25$ Å.

La Figure 102 représente l'évolution du pourcentage d'existence des liaisons hydrogène base-base au cours des TMD⁻¹ en fonction de $\rho_0(\text{KC})$. On remarque un comportement différent pour les liaisons des tiges et pour celles impliquées dans les interactions inter-boucles : les simulations débutent avec un pourcentage de 85% pour les tiges alors qu'il est voisin de 100% pour les liaisons inter-boucles. Cela signifie que le chauffage du système à 360 K provoque des ruptures de liaisons dans les tiges et non dans les boucles.

Dans les simulations issues de X-ray_KC_{Lai}, les ruptures de liaisons hydrogène interviennent en général dans la première paire de base de la tige et dans la paire U6-G18 alors que pour les simulations issues de la structure RMN, les plus faibles pourcentages sont essentiellement dus, d'une part à des ruptures de liaisons hydrogène dans ces mêmes paires de bases, d'autre part au fait que les résidus G7 et C17 sont désappariés dans cette structure. En ce qui concerne les simulations de nom générique « elai », on voit que le pourcentage d'existence des liaisons inter-boucle reste supérieur à 50% pour toutes les simulations jusqu'à $\rho_0(\text{KC})=22$ Å.

Deux de ces simulations voient se produire un désappariement total entre les deux molécules du dimère lorsque la simulation est prolongée jusqu'à $\rho_0(\text{KC})=36$ Å, les simulations elaiIV et elaiV. On peut constater sur la Figure 98 qu'il s'agit des deux simulations dont l'énergie de contrainte a subi la plus forte augmentation entre 25 Å et 32 Å si l'on excepte les simulations avec une constante de force de 0.05 kcal/mol.Å². Il y a aussi rupture de plusieurs paires de bases inter-boucle à la fin de la simulation elaiIX. Toutes les autres simulations conservent un pourcentage d'existence des interactions inter-boucle supérieur à 50% tout au long de la simulation.

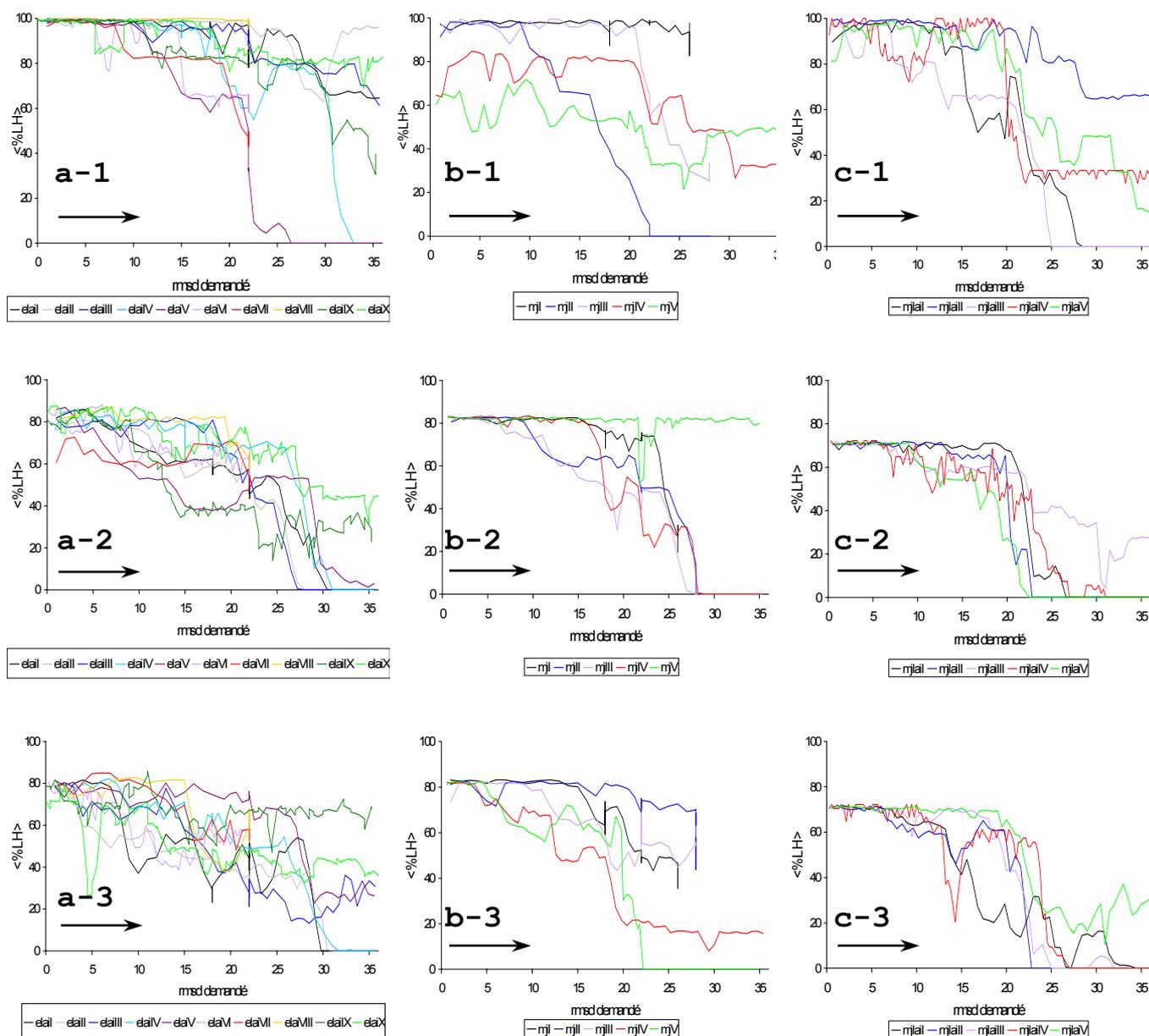


Figure 102 : Evolution des pourcentages d'existence des liaisons hydrogène canoniques des groupes inter-boucles (1), tige 1 (2) et tige 2 (3) en fonction du $\rho_0(KC)$ pour les TMD^{-1} à partir des structures initiales X-ray_KC_{Lai} (a), NMR_KC_{LaiΔ} (b) et NMR_KC_{Lai} (c). La flèche noire donne le sens de la simulation.

Les pourcentages moyens d'existence des liaisons hydrogène canoniques dans les tiges sont compris entre 35% et 70% à $\rho_0(KC)=22 \text{ \AA}$ aussi bien pour la tige 1 que pour la tige 2 ; en fin de simulation, les deux tiges sont complètement désappariées pour les simulations elaiI, elaiII et elaiIV ainsi que la tige 1 de elaiIII, tous les

autres pourcentages d'existence sont inférieurs à 50% sauf la tige 2 de elaiIX.

L'étude des courbes des simulations de nom générique « mj » et « mjlai » montre un comportement quelque peu différent. Pour des valeurs de $\rho_0(KC)$ comprises entre 0 Å et 22 Å, les simulations « mj » montrent des interactions inter-boucles plus lâches que précédemment, avec deux simulations présentant au départ un pourcentage d'existence inférieure à 80% et passant en dessous de 50% avant 22 Å. Les simulations « mjlai » gardent toutes des interactions inter-boucles supérieures à 50% d'existence jusqu'à $\rho_0(KC)=20$ Å et seule la simulation mjlaiIV passe en dessous de cette valeur à 22 Å. En revanche, pour la majorité des simulations « mj » et « mjlai », le pourcentage d'existence des liaisons hydrogène inter-boucles passe en dessous de 50% au-delà de 22 Å avec une dissociation totale de la forme dimérique pour les simulations mjII, mjlaiI et mjlaiIII. Les paires de bases des tiges présentent un comportement similaire à celui des simulations « elai », avec un pourcentage de liaisons hydrogène à $\rho_0(KC)=22$ Å compris entre 35% et 70% pour les simulations « mj », entre 5% et 60% pour les simulations « mjlai », et un désappariement plus important voir total intervenant entre 22 Å et 36 Å de rmsd. Seule la simulation mjIII présente une tige dont les liaisons hydrogène sont présentes à 80% en fin de simulation.

L'ensemble des résultats obtenus avec la méthode TMD⁻¹ suggère l'existence de structures intermédiaires en forme de double U facilement accessibles à partir de la structure KC puisque toutes nos simulations nous ont conduit à des structures de ce type sans avoir subi d'énergies de contraintes élevées. Elles sont également distantes de la structure en « kissing-complex » et de la structure en dimère étendu avec $\text{rmsd}(KC)=\text{rmsd}(DE)=22$ Å, ce qui bien sûr n'implique pas qu'elles soient identiques.

La comparaison des conformations obtenues à $\rho_0(KC)=22$ Å, montre qu'elles sont pour la plupart plus proches les unes des autres que de la conformation initiale du « kissing-complex » (rmsd entre 2 structures finales compris entre 5.9 Å et 20 Å). Ceci indique que les différentes simulations de TMD⁻¹ ont fait évoluer KC dans une même région de l'espace conformationnel, sinon les rmsd entre les

structures finales seraient compris entre 0 et 44 Å, puisque ces structures sont réparties sur une hyper-sphère de centre KC et de rayon 22 Å. Cette analyse confirme donc de manière quantitative que les structures obtenues à $\rho_0(KC)=22$ Å appartiennent bien grossièrement à une même famille comme le suggérait leur visualisation. La conformation obtenue lors de la simulation elaiI est la plus centrale, si on définit comme telle celle dont le rmsd moyen avec les autres conformations est le plus faible (son rmsd moyen est de 12.6 Å). La conformation la plus excentrée a été obtenue avec mjlaiIV (rmsd moyen=15.5 Å). Dans la suite de ce mémoire cette famille de structures sera nommée UU22. Nous nommerons d'autre part //36 l'ensemble des trois structures qui ont été obtenues avec $\rho_0(KC)=36$ Å et qui se présentent globalement sous la forme de 2 brins allongés antiparallèles.

IV.d.v. Transition complète par combinaison des méthodes de dynamique dirigée inverse et directe

Au vu des résultats précédents, nous pouvons imaginer deux possibilités pour effectuer la transition complète selon que l'on considère la famille //36 ou non ainsi que l'illustre la Figure 103, cette famille de structures n'ayant été observée que dans 27% de nos simulations. Nous avons réalisé 18 TMD à partir de 7 structures initiales issues des simulations de TMD⁻¹. Toutes ces simulations sont faites à une température de 360 K.

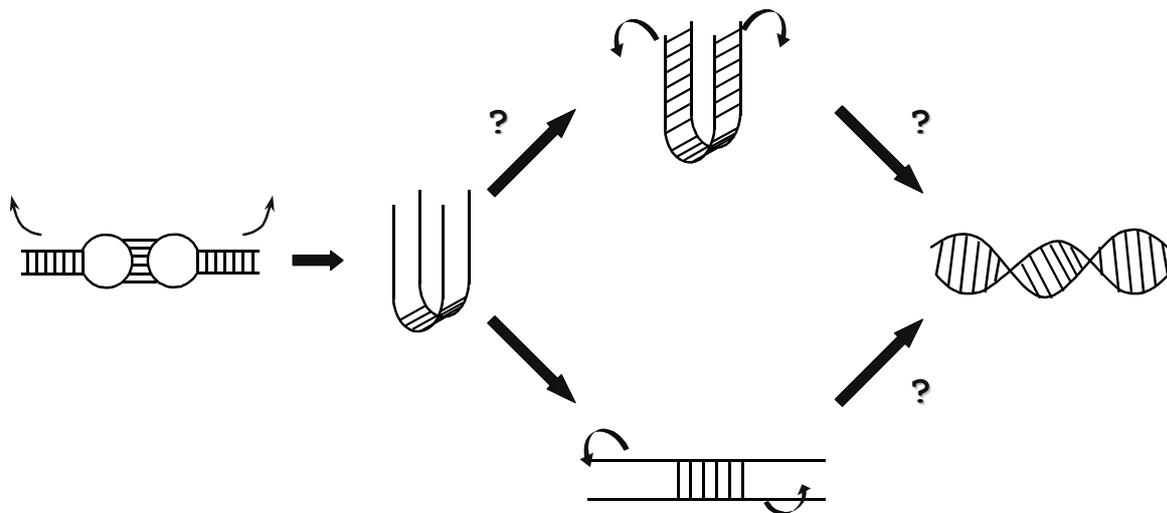


Figure 103 : Deux chemins possibles pour passer du « kissing-complex » au dimère étendu, selon que l'on passe ou non par la structure //36.

Deux de ces structures initiales appartiennent à la famille UU22, deux autres à la famille //36. Nous avons aussi testé trois structures ayant des rmsd(KC) respectivement de 15 Å, 18 Å et 26 Å. Pour certaines structures initiales les simulations ont été réalisées plusieurs fois en variant les paramètres de simulation afin d'éliminer de possibles artefacts. En particulier les 2 structures de dimère étendu NMR_DE_{Lai}_1 et NMR_DE_{Lai}_3 définies au début de ce chapitre ont été utilisées comme structure cible.

| TMD ⁻¹ initiale | Nom de la TMD directe | ρ_0 (KC) | S _c | Boîte d'eau | $\delta\rho/\Delta t$ (Å/ps) | équil (ps/Å) | k _{TMD} | NPC | ρ_0 (DE) | succès |
|-------------------------------|--------------------------|---------------|---------------------------------------|--------------------|---------------------------------|--------------------------|------------------|-----|---------------|--------|
| elaiI | elaiI_TMDI | 22 | NMR_DE _{Lai} _1 | tétra 7287/7170 | 0.2/3 | 90/12, 0 | 25 | 44P | 24 | + |
| | elaiI_TMDII | 22 | NMR_DE _{Lai} _1 | tétra 7287 | 0.13/3 | - | 25 | 34P | 20 | + |
| | elaiI_TMDIII | 36 | NMR_DE _{Lai} _3 | octa 16369 | 0.13/3 | - | 25 | 44P | 22 | - |
| | elaiI_TMDIV | 36 | NMR_DE _{Lai} _3 | octa 16369 | 0.5/25 | - | 25 | 44P | 22 | - |
| | elaiI_TMDV | 26 | NMR_DE _{Lai} _3 | octa 16369 | 0.5/25 | - | 25 | 44P | 24 | - |
| | elaiI_TMDVI | 36 | NMR_DE _{Lai} _3 | octa 16369 | 0.13/3 | 90/18, 12, 6 | 25 | 44P | 22 | - |
| | elaiI_TMDVII | 36 | NMR_DE _{Lai} _3 | octa 16369 | 0.5/25 | - | 25 | 34P | 22 | - |
| | elaiI_TMDVIII | 36 | NMR_DE _{Lai} _3 | octa 16369 | 0.5/25 | - | 5 | 44P | 22 | - |
| | elaiI_TMDIX | 36 | NMR_DE _{Lai} _3 | octa 16369 | 0.5/25 | - | 0.5 | 44P | 22 | - |
| | elaiI_TMDX | 36 | NMR_DE _{Lai} _3 | octa 16369 | 0.5/25 | - | 0.05 | 44P | 22 | - |
| | elaiI_TMDXI | 22 | NMR_DE _{Lai} _3 | octa 10554 | 0.2/3 | 90/12, 0 | 25 | 44P | 24 | + |
| | elaiI_TMDXII | 36 | NMR_DE _{Lai} _1 | octa 16369 | 0.13/3 | 90/18, 12, 6 | 25 | 44P | 22 | - |
| elaiIV | elaiIV_TMDI | 15 | NMR_DE _{Lai} _1 | tétra 6513 | 0.2/3 | 90/21, 15, 9 | 25 | 44P | 27 | - |
| | elaiIV_TMDII | 18 | NMR_DE _{Lai} _1 | tétra 6513/6910 | 0.2/3 0.1/3 | 90/21, 15, 9, 3, 0 | 25 | 44P | 27 | + |
| | elaiIV_TMDIII | 18 | NMR_DE _{Lai} _3 | tétra 6513 | 0.2/3 | - | 25 | 44P | 25 | - |
| | elaiIV_TMDIV | 18 | NMR_DE _{Lai} _3 | octa 10901 | 0.2/3 | 90/19, 13, 7, 1 | 25 | 44P | 25 | - |
| mjI | mjI_TMD | 22 | NMR_DE _{Lai} _1 muté SL1A | tétra 6331/6889 | 0.2/3 0.1/3 | 90/16, 10, 4, 1 | 25 | 44P | 22 | + |
| mjlaiII | mjlaiII_TMD | 36 | NMR_DE _{Lai} _1 | octa 16400 | 0.2/5 | - | 25 | 44P | 22 | - |

Tableau 11 : Tableau récapitulatif des conditions de simulations des 18 TMD réalisées à partir de TMD⁻¹. Les ρ_0 (KC) et ρ_0 (DE) sont donnés en Å, k_{TMD} en kcal/mol.Å². S_c correspond à la structure cible, NPC au nombre de phosphores contraints.

L'ensemble des conditions de simulation de ces 18 TMD est répertorié dans le Tableau 11.

Dans le cas de elaiI_TMDII, les contraintes sur les phosphores originellement exclus ont ensuite été rajoutées paire de bases par paire de bases en fin de simulation entrecoupées d'étapes d'équilibrations de 90 ps. Nous considérons que la structure de type dimère étendu est atteinte lorsque la valeur de $\rho(\text{DE})$ obtenue est inférieure à 2.5 Å.

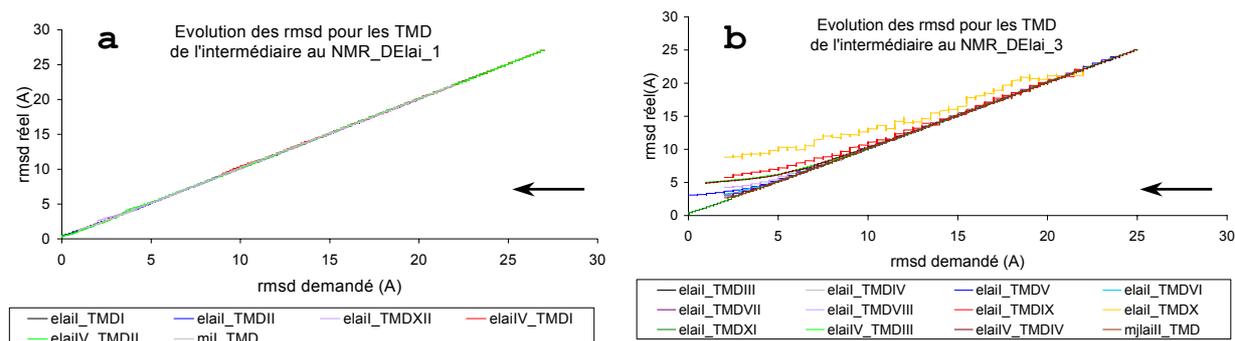


Figure 104 : Evolution du rmsd(DE) réel en fonction de $\rho_0(\text{DE})$ pour les TMD ayant pour structure cible NMR_DE_{Lai_1} (a) et NMR_DE_{Lai_3} (b). La flèche noire donne la direction de la simulation.

| Tmd Directe | Coefficient de corrélation |
|---------------|----------------------------|
| elaiI_TMDI | 0.9999 |
| elaiI_TMDII | 0.9999 |
| elaiI_TMDIII | 0.9995 |
| elaiI_TMDIV | 0.9997 |
| elaiI_TMDV | 0.9975 |
| elaiI_TMDVI | 0.9995 |
| elaiI_TMDVII | 0.9998 |
| elaiI_TMDVIII | 0.9986 |
| elaiI_TMDIX | 0.9955 |
| elaiI_TMDX | 0.9880 |
| elaiI_TMDXI | 0.9999 |
| elaiI_TMDXII | 0.9999 |
| elaiV_TMDI | 0.9999 |
| elaiV_TMDII | 0.9999 |
| elaiV_TMDIII | 0.9906 |
| elaiV_TMDIV | 0.9897 |
| mjl_TMD | 0.9999 |
| mjlaiI_TMD | 0.9999 |

Tableau 12 : Coefficients de corrélation entre les valeurs de $\rho(\text{DE})$ et les valeurs de $\rho_0(\text{DE})$ pour les 18 TMD directes.

La Figure 104 présente tout d'abord les courbes de $\rho(\text{DE})$ en fonction du $\rho_0(\text{DE})$. Les coefficients de corrélation de ces différentes courbes sont également reportés dans le Tableau 12. Il y a globalement une bonne corrélation entre les valeurs de $\rho(\text{DE})$ et $\rho_0(\text{DE})$ au cours de ces simulations. La seule TMD qui se démarque nettement des autres correspond à une simulation réalisée avec une constante de force de 0.05 kcal/mol.Å². Les plus faibles coefficients de corrélation sont obtenus pour les simulations utilisant des constantes de force de 0.5 et 0.05 kcal/mol.Å² ainsi que pour trois TMD ne partant pas de structures UU22 ou //36, deux d'entre elles

partant d'une structure avec $\rho_0(\text{KC})=18 \text{ \AA}$, la troisième d'une structure avec $\rho_0(\text{KC})=26 \text{ \AA}$. On peut remarquer de surcroît qu'en dessous de $\rho_0(\text{DE})=4-5 \text{ \AA}$, nombre de simulations semblent éprouver des difficultés à se rapprocher de la structure en dimère étendu, la courbe formant alors un palier.

Toutes les simulations partant de la famille UU22 aboutissent à la formation d'un dimère étendu sans rencontrer de difficultés majeures (elaiI_TMDI, elaiI_TMDII, elaiI_TMDXI et mjiI_TMD). On voit sur la Figure 105 que pour ces simulations, les énergies de contrainte restent inférieures à 15 kcal/mol jusqu'à atteindre une valeur de $\rho_0(\text{DE})$ inférieure à 1 \AA .

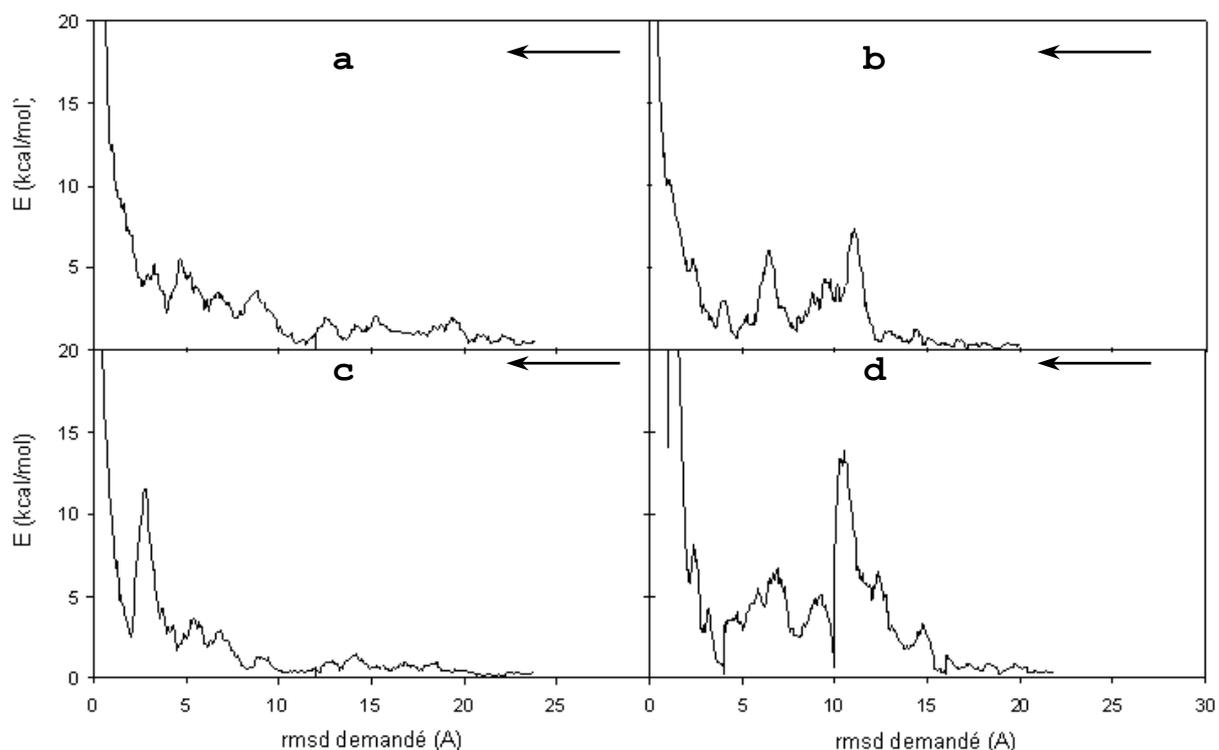


Figure 105 : Evolution de l'énergie de contrainte pour les quatre TMD partant d'une structure de la famille UU22 en fonction du $\rho_0(\text{DE})$: a) elaiI_TMDI b) elaiI_TMDII c) elaiI_TMDXI d) mjiI_TMD. La flèche noire donne la direction de la simulation.

Ceci indique que les barrières énergétiques rencontrées le long du chemin de transition UU22-DE sont aisément franchissables.

Au contraire aucune des 9 TMD réalisées à partir de la famille //36 n'aboutit à la formation du DE. L'énergie de contrainte au cours de ces simulations peut prendre des valeurs très élevées (de 100 à 8500 kcal/Mol) dans la gamme 10-4 \AA . Ceci résulte de la formation

systematique de nœuds, voire pour certaines TMD au passage d'un brin au travers de l'autre, phénomène déjà évoqué au §IV.d.iii. A titre d'exemple sur ce que l'on veut exprimer par la formation de « nœuds », la Figure 106 présente la structure finale obtenue pour la simulation elaiI_TMDIII.

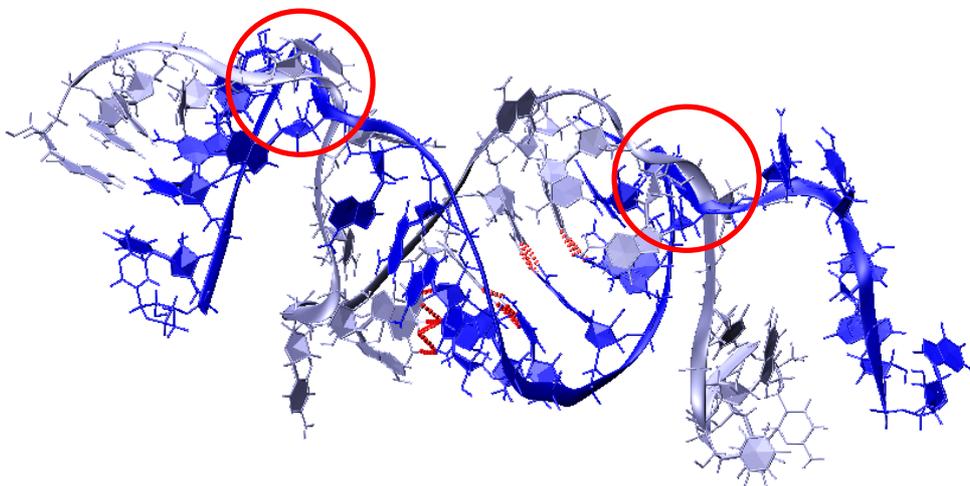


Figure 106 : Exemple d'une structure présentant des nœuds, la structure finale de la simulation elaiI_TMDIII. Les nœuds sont mis en évidence par les ronds rouges.

Sur les 5 TMD réalisées à partir de structures différentes de UU22 ou //36, seule une simulation (elaiIV_TMDII) sur les trois partant de $\rho_0(\text{KC})=18 \text{ \AA}$ aboutit au dimère étendu. Encore faut-il remarquer qu'aux environs de $\rho_0(\text{DE})=4-5 \text{ \AA}$, une barrière d'énergie a été difficile à franchir car on observe une énergie de contrainte atteignant 70 kcal/mol. Ce pic quelque peu élevé est dû au passage difficile du second brin qui a frôlé la création d'un nœud lors de l'enroulement du duplexe 2 au cours de la formation du dimère étendu, information obtenue par la visualisation de la trajectoire de la molécule sur cette période de temps.

Afin de suivre de manière plus fine l'évolution structurale subie par la molécule au cours des transitions réussies, nous avons calculé les pourcentages d'existence des liaisons hydrogène (Figure 107).

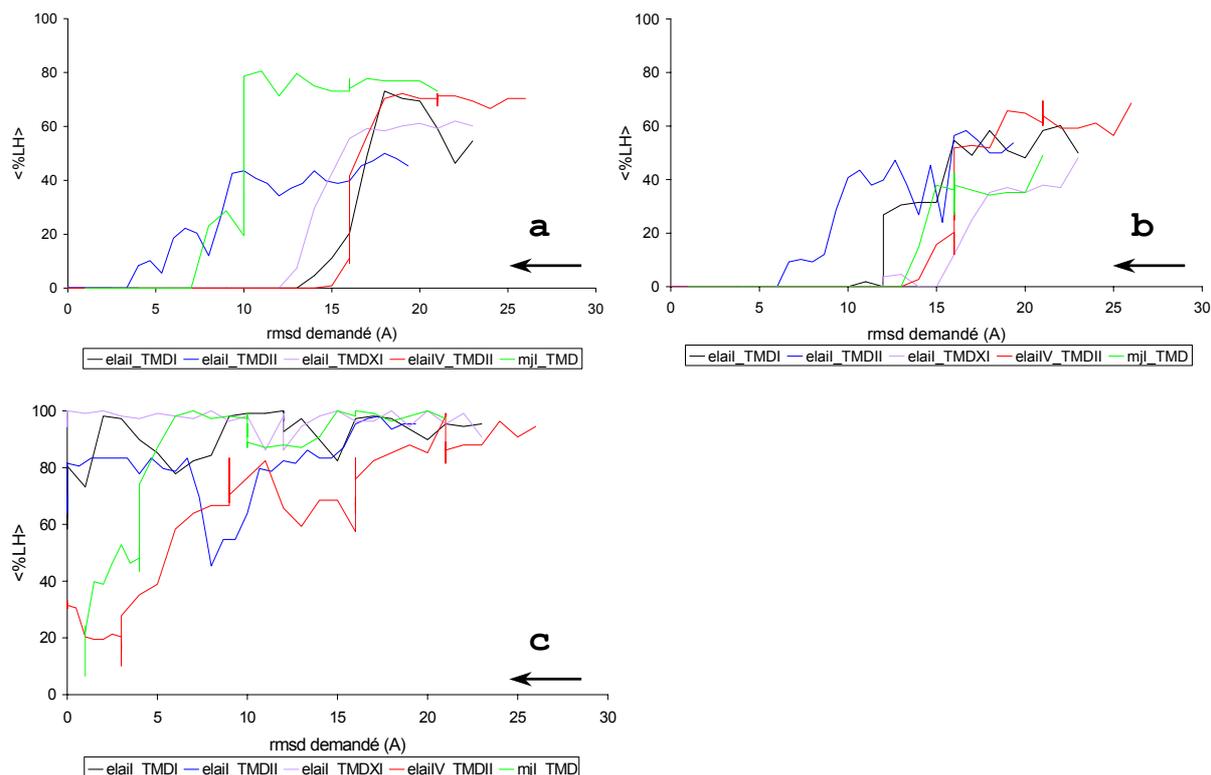


Figure 107 : Evolution des pourcentages d'existence des liaisons hydrogène canoniques dans les groupes tige 1 (a), tige 2 (b) et inter-boucles (c) pour les 5 TMD aboutissant à une structure de type dimère étendu. La flèche noire donne la direction de la simulation.

Les pourcentages d'existence des liaisons hydrogène sur les duplexes 1 et 3 du dimère étendu ne sont pas retranscrites ici car les liaisons hydrogène ne se forment pas dans ces TMD, sauf à la fin de la simulation elaiI_TMDII lorsque l'on rajoute les contraintes sur les phosphores omis en début de simulation et dans des pourcentages inférieurs à 25% pour les deux duplexes. On voit sur ces courbes qu'au départ des TMD les tiges sont appariées avec un pourcentage de liaisons hydrogène compris entre 40% et 80% et sont totalement désappariées lorsque l'on atteint $\rho_0(DE)=10$ Å sauf pour la tige 1 de elaiI_TMDII et mjl_TMD qui ne sont totalement désappariées qu'à 4 Å et 7 Å respectivement. Les liaisons inter-boucles quant à elles sont présentes à des pourcentages compris entre 90% et 100% dans les structures de départ. Les trois simulations elaiI_TMDI, elaiI_TMDII et elaiI_TMDXI conservent une structure dimérique solide au cours de la transition, avec des pourcentages d'existence de liaisons hydrogène inter-boucles compris entre 80% et 100% en fin de simulation malgré un affaiblissement localisé pour elaiI_TMDII entre

$\rho_0(\text{DE})=10 \text{ \AA}$ et $\rho_0(\text{DE})=6 \text{ \AA}$. Il est à noter que ces 3 TMD correspondent aux TMD réalisées à partir d'une structure UU22 sur la séquence native de SL1_{Lai}.

Les structures finales obtenues pour ces trois dynamiques sont présentées sur la Figure 108.

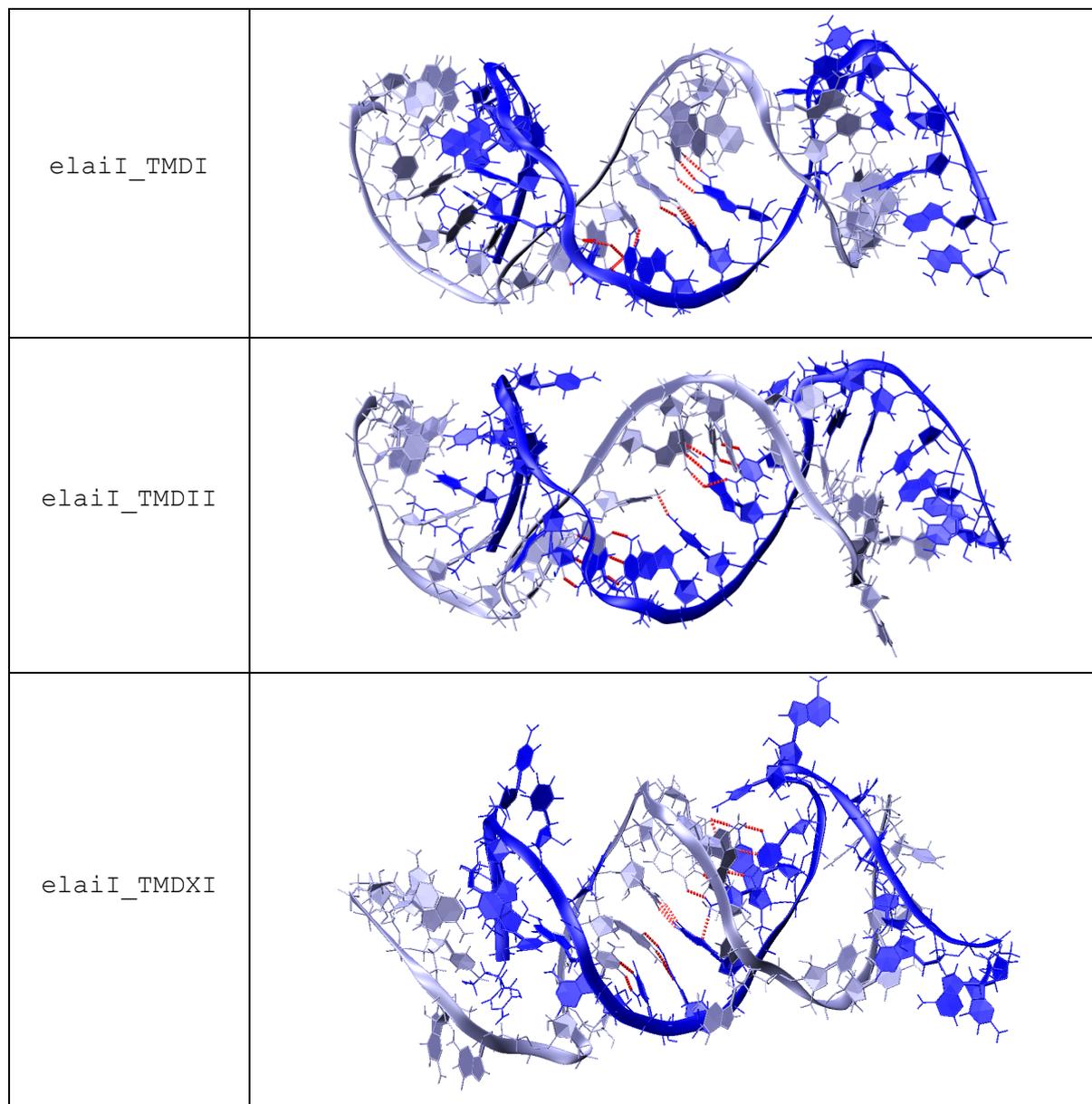


Figure 108 : Représentation des structures finales des TMD elaiI_TMDI, elaiI_TMDII et elaiI_TMDXI. Le monomère 1 est représenté en bleu, le monomère 2 en violet.

Les deux simulations restantes présentent un fort affaiblissement des interactions inter-boucles en fin de simulation, avec des pourcentages d'existence compris entre 5% et 30%. Pour la simulation elaiIV_TMDII, réalisée à partir d'une structure n'ayant pas encore

atteint l'état UU22, cet affaiblissement peut être dû aux contraintes imposées sur le squelette par la formation du pseudo-nœud et en effet l'affaiblissement des interactions inter-boucles correspond au pic observé pour l'énergie de contrainte. Quant à la simulation mjI_TMD il convient de se souvenir qu'il s'agit d'une séquence mutée de SL1 dans le but de favoriser l'équilibre de la réaction vers la formation du KC en stabilisant les tiges ; il est donc possible que le fait d'imposer la transition vers le DE génère des contraintes à l'origine de la rupture des liaisons hydrogène dans les boucles.

IV.d.vi. Discussion-conclusion

Dans cette étude nous avons tenté de déterminer un chemin réactionnel possible pour décrire la transition du dimère de SL1 depuis la structure KC relativement instable vers la structure étendue DE plus stable. Les simulations de TMD directe ont toutes échouées. Vingt simulations de TMD⁻¹ réalisées à partir de KC ont toutes conduit sans contraintes excessives vers une famille de structures (UU22) caractérisée par un rapprochement des 2 tiges ainsi qu'un déroulement de celles-ci, tout en conservant une grande partie des interactions boucle-boucle. Un éloignement supplémentaire de la structure KC semble poser plus de problèmes : non-convergence des simulations et valeurs élevées des énergies de contrainte, bien que 3 simulations aboutissent à la formation d'une famille de structures où les 2 brins sont alignés antiparallèlement tout en conservant l'interaction boucle-boucle (//36).

Des simulations de TMD directe à partir de UU22 et //36 montrent qu'un chemin réactionnel UU22 \rightarrow DE existe alors qu'aucune des simulations partant de //36 n'arrive à atteindre l'état DE. Cependant, le long du chemin réactionnel UU22 \rightarrow DE, notre étude ne permet pas de déterminer à quel endroit se forment les paires de bases caractéristiques de DE.

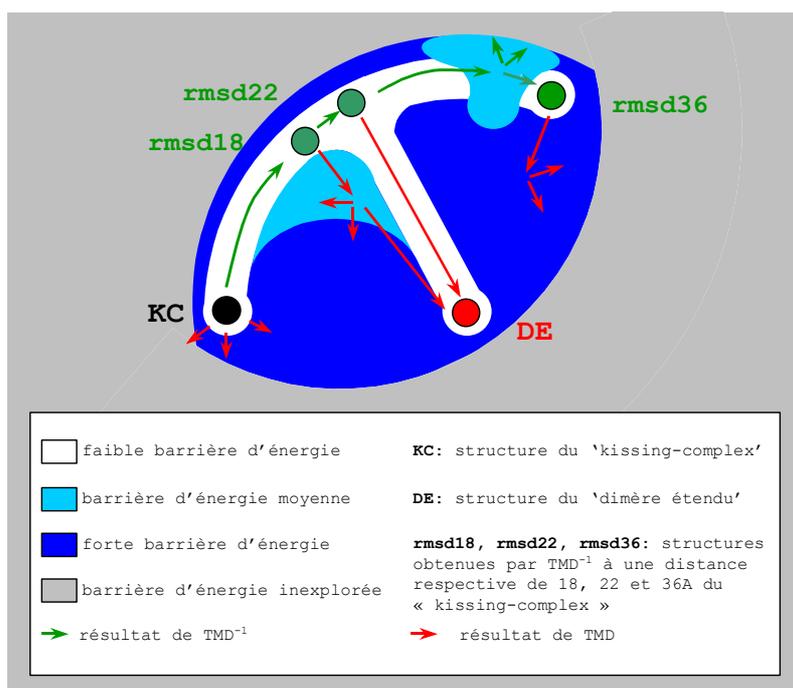


Figure 109 : Représentation schématique des barrières énergétiques rencontrées lors des expériences de transition « kissing-complex » → dimère étendu réalisées, en utilisant le rmsd comme coordonnée de réaction.

En effet, les contraintes imposées au cours de la TMD ne s'appliquent qu'aux atomes de phosphore d'une part, d'autre part il aurait certainement fallu augmenter considérablement le temps de ces simulations ainsi que baisser la température pour pouvoir observer la formation des liaisons hydrogène correspondantes. Nous obtenons donc au final une structure proche du dimère étendu par la forme de son squelette phospho-diester mais en aucun cas le dimère étendu réel. La Figure 109 résume l'ensemble de ces résultats. Une visualisation des changements conformationnels subis par la molécule au cours de ce chemin est présenté Figure 110.

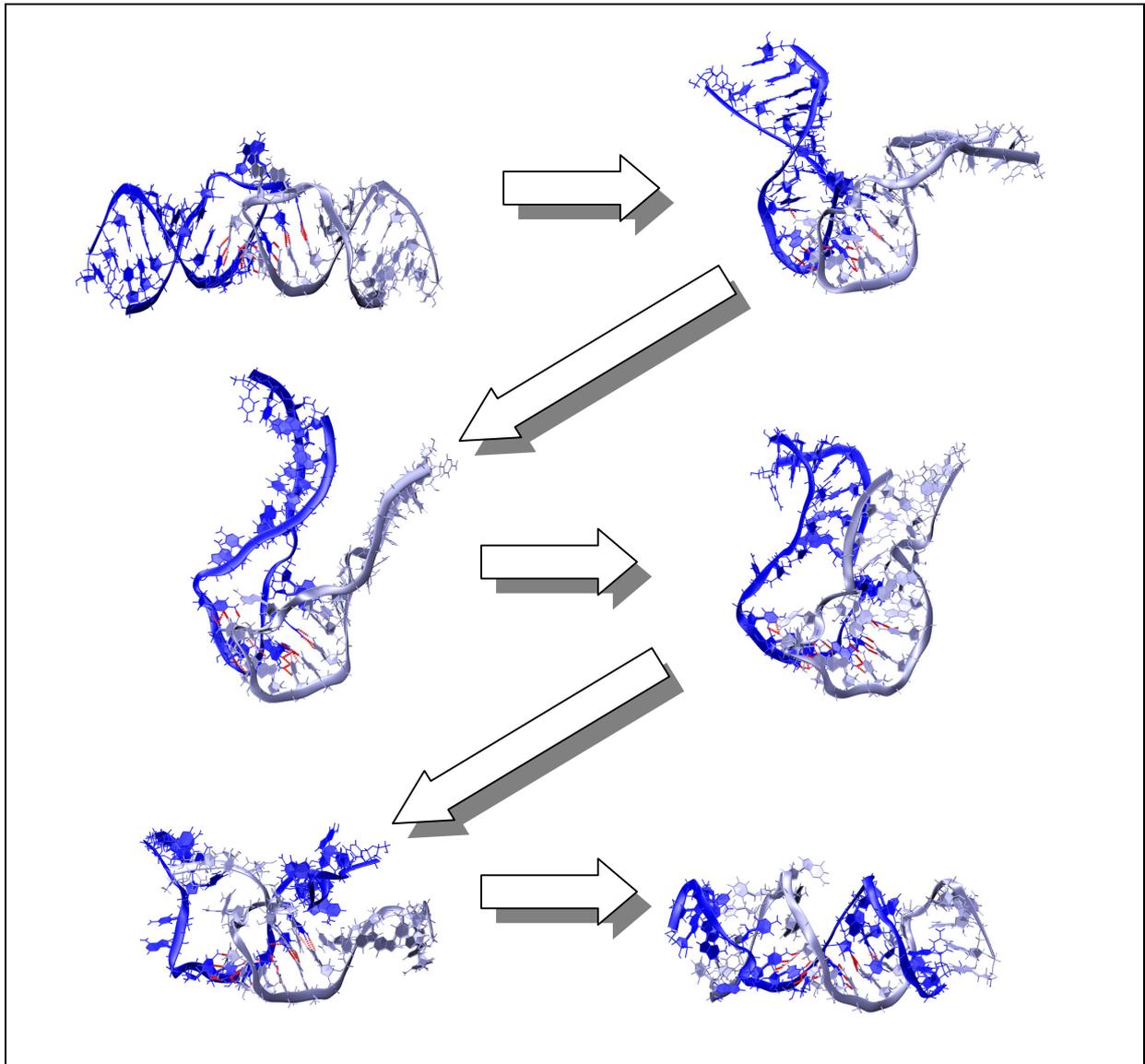


Figure 110 : Snapshot d'une trajectoire complète de la transition entre le « kissing-complex » et le dimère étendu. Le monomère 1 est représenté en bleu et le monomère 2 en violet.

Nous avons cherché à comprendre les raisons des échecs des simulations partant des structures //36. Pour cela nous avons varié les paramètres de contraintes, tels que la valeur de la constante de force ou le nombre d'atomes contraints. La Figure 111 présente en fonction du temps et par rapport à la valeur demandé l'évolution du rmsd(DE) pour les quatre simulations elaiI_TMDIV, elaiI_TMDVIII, elaiI_TMDIX et elaiI_TMDX dont tous les paramètres sont identiques exceptée la valeur de la constante de force.

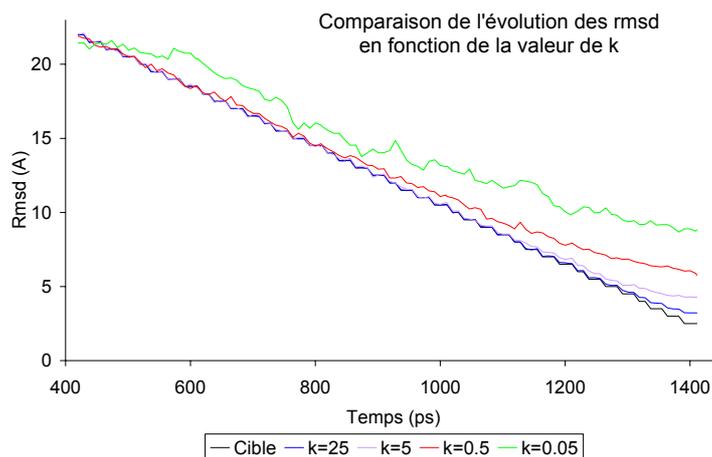


Figure 111 : Evolution au cours du temps du rmsd(DE) des TMD directes à partir d'une structure de la famille //36 en fonction de la valeur de la constante de force k_{TMD} utilisée lors de la simulation.

L'idée était de vérifier que l'on n'imposait pas, avec une constante de force de 25 kcal/mol.Å², de contraintes trop fortes pour le système, l'obligeant ainsi à former des nœuds. Nous pouvons remarquer qu'avec les deux dernières constantes de force, à savoir $k=0.5$ kcal/mol.Å² et $k=0.05$ mol.Å², le rmsd peine à rejoindre les valeurs demandées. Cependant, la molécule forme tout de même des nœuds au cours de la simulation, il ne semble donc pas que la valeur de la constante de force soit le facteur premier des échecs de ces simulations.

Un autre paramètre que nous nous sommes efforcés de tester est la vitesse de variation du RMSD ainsi que le temps d'équilibration de chaque palier. Les simulations elaiI_TMDIII, elaiI_TMDIV et elaiI_TMDVI ne diffèrent que par ces paramètres cependant on ne parvient pas à former le dimère étendu sans que les brins de la molécule ne s'entremêlent entre eux. Il en a été de même lorsque nous avons tenté d'éviter ces nœuds en n'imposant les contraintes que sur 34 atomes de phosphores, excluant ainsi les trois premières paires de bases de chacune des tiges du KC du groupe d'atomes contraints. Enfin nous avons essayé de partir d'une structure comprise entre les deux structures intermédiaires dans la simulation elaiI_TMDV qui a elle aussi abouti à un échec. Il semble donc que la condition première pour réaliser cette transition soit d'utiliser UU22 comme structure de départ de la TMD, ce qui semblerait confirmer notre impression première sur sa suprématie en tant que

structure intermédiaire dans la transition « kissing-complex » → dimère étendu.

Examinons maintenant si les résultats que nous avons obtenus sont dépendants de la méthode employée. En ce qui concerne la TMD directe, si la simulation aboutit sans faire intervenir des énergies de contraintes trop élevées, on peut être confiant que le chemin conformationnel suivi par la molécule au cours de la transition est un chemin possible bien que non obligatoirement unique. Par contre, si la simulation échoue, deux causes peuvent être invoquées : soit le chemin conformationnel est incompatible avec un rapprochement continu de la structure vers la structure cible mais nécessite un détour qui l'en éloigne temporairement, soit l'approche de la structure cible est réalisée trop rapidement de sorte qu'à la fin de chaque simulation intermédiaire durant la TMD la molécule n'est pas totalement relaxée. Bien que nous ayons testé différentes vitesses d'approche de la cible nous ne pouvons pas exclure cette dernière éventualité.

En ce qui concerne la TMD inverse, si des simulations multiples conduisent toutes au même résultat, on peut raisonnablement penser que le chemin emprunté correspond à une réalité. Ainsi l'absence d'une transition directe KC → DE avec rapprochement continu vers DE est confirmée par les 20 TMD inverses qui conduisent toutes vers UU22. Par contre, l'échec de la TMD directe à partir des structures //36 peut être dû soit à une approche trop rapide, soit au fait que, pour réaliser la transition, la structure doit s'éloigner temporairement de DE. Dans tous les cas, il semble néanmoins qu'au cours de la transition KC → DE le chemin conformationnel passe par une structure du type UU22 puisque pour arriver à //36 il a fallu passer par cette structure.

V. CONCLUSION GENERALE

V. CONCLUSION GENERALE

Le premier objectif de notre travail était de déterminer lesquelles, parmi les structures expérimentales du dimère de SL1 publiées dans la littérature, étaient stables en solution. De toutes les structures du « kissing-complex » publiées, nous trouvons que la plus stable en solution est la structure caractéristique de la famille 2 de Kieken (2004). Les structures de « kissing-complex » des isolats Lai et Mal déterminées par cristallographie (Ennifar et al. 2001) se révèlent très instables en solution. Il ne semble pas que cette déstructuration des « kissing-complex » issues de la cristallographie soit le fait de la méthode employée. En effet, toutes les structures expérimentales ont été simulées avec le même protocole, or les structures de « kissing-complex » issues d'expériences RMN restent toutes relativement stables, et la structure du dimère étendu déterminée par cristallographie (Ennifar et al. 1999) et simulée dans les mêmes conditions que les « kissing-complex » issues de la cristallographie, se montre très stable au cours de notre simulation. Nous pouvons noter sur l'ensemble des résultats obtenus sur les simulations des « kissing-complex » une grande flexibilité des jonctions tige-boucle. En ce qui concerne les dimères étendus, les deux structures publiées sont stables en solution, bien que la structure RMN semble plus flexible que la structure cristallographique.

Nous avons remarqué qu'une grande majorité des structures obtenues en fin de simulation est dissymétrique, en particulier, les structures présentant une ouverture de la paire G7-C17 sur l'un des monomères ne montrent jamais une ouverture de cette même paire sur l'autre monomère. Cette dissymétrie du « kissing-complex » est déjà observée dans les résultats expérimentaux de Ennifar et al. (2001). Nous avons relevé au cours de ces études des différences de comportement entre les isolats Lai et Mal. La première de ces différences concerne la paire G7-C17 qui se révèle très flexible dans les dimères de SL1 de l'isolat Lai qu'il s'agisse du « kissing-complex » ou du dimère étendu alors qu'elle est rigide dans les dimères de SL1 de l'isolat Mal. De plus, si l'on se réfère à l'ensemble des simulations réalisées sur le « kissing-complex » et

CONCLUSION GENERALE

sur le dimère étendu, nous pouvons remarquer d'une part que le complexe inter-boucle est moins stable pour l'isolat Mal que pour l'isolat Lai, d'autre part que le dimère étendu présente une plus grande rigidité pour l'isolat Mal que pour l'isolat Lai. Ces données semblent indiquer que l'équilibre $KC \leftrightarrow DE$ est plus décalé vers le dimère étendu pour l'isolat Mal. Il faut cependant noter que les simulations des deux dimères étendus Lai et Mal ont été réalisées dans des conditions différentes, les ions magnésium présents dans la simulation du dimère étendu Mal pouvant être responsables de la plus grande rigidité observée pour ce dimère en solution. Cependant, des expériences d'électrophorèse menées en parallèle sur les différents complexes de Mal et Lai en présence et en absence d'ion magnésium ont montré que l'isolat Lai peut former deux types de dimères, une forme instable dans les conditions d'obtention du « kissing-complex » et une forme stable dans les conditions d'obtention du dimère étendu, alors que Mal ne forme que le complexe stable quelles que soient les conditions d'obtention (Aci et al. 2004a). Ces observations sont indépendantes de la présence d'ions magnésium et sont aussi en faveur d'un déplacement de l'équilibre vers le DE pour Mal. Il convient cependant de rappeler que le « kissing-complex » de l'isolat Mal a pu être observé dans d'autres conditions expérimentales (Dardel et al. 1998; Ennifar et al. 2001).

L'autre partie de notre travail a consisté en l'étude du chemin conformationnel suivi par le dimère de SL1 pour passer de la forme instable à la forme stable. Nous avons utilisé pour cela les techniques de dynamique moléculaire dirigée (TMD) directe et inverse. Ces techniques sont relativement récentes et donc, en particulier la TMD inverse, encore assez peu utilisées. La TMD directe a jusqu'à maintenant été utilisée pour déterminer le chemin conformationnel lors d'une transition entre deux états bien définis d'une protéine (Apostolakis et al. 1999; Diaz et al. 1997; Kong et al. 2002; Kruger et al. 2001; Ma and Karplus 1997; Ma et al. 2000; Madhusoodanan and Lazaridis 2003; Mendieta et al. 2004; Roche and Field 1999; Schlitter et al. 1994; Wroblowski et al. 1997; Yang et al. 2002) ou pour en étudier le dépliement (Ferrara et al. 2000b; Ferrara et al. 2000c; Ferrara 2000a). En revanche la TMD inverse n'a été utilisée à notre connaissance que par Ferrara et al. (2000b;

CONCLUSION GENERALE

2000a) pour étudier le dépliement de l'inhibiteur de la chymotrypsin en l'absence de solvant explicite. La combinaison de ces deux techniques appliquée à une séquence d'acides aminés nous a permis de déterminer un chemin possible pour la transition KC ↔ DE du dimère de SL1, et a mis en évidence un intermédiaire réactionnel de cette transition. Nous soulignons que les simulations réalisées dans cette étude ne concernent qu'une partie de la tige-boucle SL1, et le chemin que nous observons pour la transition est vraisemblablement facilité par l'extrémité libre des tiges qui ont une longueur limitée. A ce propos il est bon de rappeler que dans la séquence entière de SL1, la tige-boucle de 23 nucléotides que nous avons étudiée est prolongée par une boucle interne dont l'importance a été mise en évidence (Takahashi et al. 2000). Bien que peu d'informations structurales sur cette boucle soient disponibles, on peut supposer en effet qu'elle possède une flexibilité intrinsèque bien supérieure à celle d'une structure en double hélice, laissant une relative liberté à la tige pour se désappairer.

Au cours du chemin réactionnel que nous proposons il est à noter que les interactions boucle-boucle restent très présentes (au moins en ce qui concerne la séquence sauvage). Ceci est en bon accord avec les résultats expérimentaux de Theilleux-Delalande et al. (2000) qui mettent en évidence que la transition KC → DE s'accompagne de la rupture des paires de bases des tiges alors que les paires de bases inter-brins restent appariées. Notre étude ne nous permet cependant pas d'éliminer les modèles proposés par Ennifar et al. (2001) et Pattabiraman et al. (2002). Ces deux modèles impliquent un mécanisme de type coupure du squelette phospho-diester suivi d'une religation croisée, permettant le passage du « kissing-complex » au dimère étendu sans ouverture des tiges des « kissing-complex ». Ce type de réaction chimique ne peut être abordé par le biais de la simulation de dynamique moléculaire. Cependant, les résultats présentés par ces auteurs sont en désaccord et avec l'étude de Windbichler et al. (2003), pour qui le mécanisme de transition entre ces deux dimères ne peut être de type coupure-religation, et avec l'étude RMN de Theilleux-Delalande et al. (2000).

VI . PERSPECTIVES

VI . PERSPECTIVES

Nous avons débuté au cours de ce travail une étude d'un chemin conformationnel possible pour effectuer la transition KC → DE pour le dimère de SL1. Cependant, il ne s'agit là que d'une étude préliminaire qui demande à être approfondie. En premier lieu, il serait intéressant d'effectuer une analyse énergétique du système au cours de la transition complète à l'aide d'un logiciel tel que DELPHI (Rocchia et al. 2001; Rocchia et al. 2002). Nous pourrions ensuite refaire cette même transition avec des étapes d'équilibration plus longues et des variations $\delta\rho$ beaucoup plus faibles afin de déterminer le profil d'énergie libre de la molécule au cours de la transition et d'évaluer ainsi les barrières énergétiques à franchir pour effectuer la transition, notamment en ce qui concerne l'intermédiaire réactionnel UU22 identifié.

Il faut aussi rappeler que lorsque nous avons commencé cette étude, nous ne disposions alors comme structures initiales que des structures de « kissing-complex » publiées par Mujeeb et al. (1998) et Ennifar et al. (2001). Entre temps, Kieken (2004) a déterminé pour le « kissing-complex » une autre structure, ou plus exactement deux familles susceptibles de représenter cette structure. Notre étude sur la stabilité des « kissing-complex » ayant révélée que la structure la plus stable en solution était la structure caractéristique de la famille 2 de Kieken (2004), il serait judicieux de reprendre l'étude à partir de cette structure pour refaire la transition KC → DE. Enfin, l'importance de la boucle interne dans les deux étapes de la dimérisation pour la séquence complète de la tige-boucle SL1 a été mise en évidence par Takahashi et al. (2000). Nous notons par ailleurs le fait que notre transition est grandement facilitée par le fait que nous n'utilisons que la séquence 23-mer de SL1, dont les tiges sont courtes et l'extrémité libre. Il pourrait être très instructif de réitérer notre protocole de transition à partir cette fois, non plus de la séquence 23-mer, mais 35-mer de SL1. Cela nous permettrait de vérifier l'existence de notre intermédiaire dans le cas d'une tige plus longue contenant une boucle interne, et préciserait le rôle de cette boucle dans le mécanisme de dimérisation.

VII. PUBLICATIONS ET COMMUNICATIONS

VII. PUBLICATIONS ET COMMUNICATIONS

VII.a. Publications

1. S Aci, L Gangneux, J Paoletti, D Genest: On the stability of different experimental dimeric structures of the SL1 sequence from the genomic RNA of HIV-1 in solution: a molecular dynamics simulation and electrophoresis study. *Biopolymers* 2004a, 74:177-88.
2. S Aci, J Ramstein, D Genest: Base pairing at the stem-loop junction in the SL1 kissing complex of HIV-1 RNA: a thermodynamic study probed by molecular dynamics simulation. *J Biomol Struct Dyn* 2004b, 21:833-40.
3. G De Vuyst, S Aci, D Genest, F Culard: Atypical recognition of particular DNA sequences by the archeal chromosomal MC1 protein. *soumis à Biochemistry*.

VII.b. Communication Orale

"Relative stabilities of the dimeric structures of Mal and Lai HIV-1 subtype", XIII^{ème} Colloque du **GGMM**, Cabourg (mai 2003)

VII.c. Communications par affiche

1. " On the stability of different experimental dimeric structures of the SL1 sequence from genomic HIV-1 RNA in solution: a molecular dynamics simulation study" 4th European Biophysics Congress, Alicante (juillet 2003)
2. « Stabilité des différentes structures dimériques expérimentales de la séquence SL1 de l'ARN génomique de VIH-1 en solution : étude par simulation de dynamique moléculaire » 4^{ème} rencontre de Figeac (octobre 2003)

VIII . BIBLIOGRAPHIE

VIII. BIBLIOGRAPHIE

- Aci, S., Gangneux, L., Paoletti, J., and Genest, D. (2004a). "On the stability of different experimental dimeric structures of the SL1 sequence from the genomic RNA of HIV-1 in solution: a molecular dynamics simulation and electrophoresis study." *Biopolymers*, 74(3), 177-88.
- 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(5), 1920-6.
- An, W., and Telesnitsky, A. (2002). "HIV-1 Genetic Recombination: Experimental Approches and Observations." *AIDS Rev.*, 4(4), 195-212.
- Apostolakis, J., Ferrara, P., and Caflisch, A. (1999). "Calculation of conformational transitions and barriers in solvated systems: Application to the alanine dipeptide in water." *J. Chem. Phys.*, 110(4), 2099-2108.
- Aqvist, J. (1990). "Ion-Water interaction Potentials Derived from Free Energy Perturbation Simulations." *J. Phys. Chem.*, 94, 8021-8024.
- Awang, G., and Sen, D. (1993). "Mode of dimerization of HIV-1 genomic RNA." *Biochemistry*, 32(42), 11453-7.
- Banavali, N. K., and MacKerell, A. D., Jr. (2002). "Free energy and structural pathways of base flipping in a DNA GCGC containing sequence." *J Mol Biol*, 319(1), 141-60.
- 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(2), 382-97.
- Beaurain, F., and Laguerre, M. (2003). "MD studies of the DIS/DIS kissing complex solution and x-ray structures." *Oligonucleotides*, 13(6), 501-14.
- Bender, W., Chien, Y. H., Chattopadhyay, S., Vogt, P. K., Gardner, M. B., and Davidson, N. (1978). "High-molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures." *J Virol*, 25(3), 888-96.
- 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(4), 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(1), 181-7.
- Berkhout, B., and van Wamel, J. L. (1996). "Role of the DIS hairpin in replication of human immunodeficiency virus type 1." *J Virol*, 70(10), 6723-32.
- Berkhout, B., and van Wamel, J. L. (2000). "The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure." *Rna*, 6(2), 282-95.
- Berkowitz, R. D., Luban, J., and Goff, S. P. (1993). "Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays." *J Virol*, 67(12), 7190-200.

BIBLIOGRAPHIE

- 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(1), 119-27.
- Bloomfield, V., Crothers, D., and Tinico, I. J. (1974). *Physical Chemistry of Nucleic Acids*, New York.
- Brun-Vezinet, F., Damond, F., and Simon, F. (1999). "[Variability of human immunodeficiency virus type 1]." *Bull Soc Pathol Exot*, 92(4), 261-3.
- Burckard, M. E., Turner, D. H., and Tinico, I. J. (1999). *The interactions that shape RNA structure*, Cold Spring Harbor Press, New York.
- Canaani, E., Helm, K. V., and Duesberg, P. (1973). "Evidence for 30-40S RNA as precursor of the 60-70S RNA of Rous sarcoma virus." *Proc Natl Acad Sci U S A*, 70(2), 401-5.
- Cantor, C. R., and Schimmel, P. R. (1980). *Biophysical Chemistry*, Freeman and Co., San Francisco.
- Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, I., 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, San Fransisco.
- Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, I., 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., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1999). "AMBER 6." University of California, San Fransisco.
- Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, I., T.E., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (2002). "AMBER 7." University of California, San Fransisco.
- Charneau, P., Borman, A. M., Quillent, C., Guetard, D., Chamaret, S., Cohen, J., Remy, G., Montagnier, L., and Clavel, F. (1994). "Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group." *Virology*, 205(1), 247-53.
- Cheatham, T. E., 3rd, Cieplak, P., and Kollman, P. A. (1999). "A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat." *J Biomol Struct Dyn*, 16(4), 845-62.
- Clavel, F., and Orenstein, J. M. (1990). "A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology." *J Virol*, 64(10), 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(4), 2101-9.
- Clever, J. L., and Parslow, T. G. (1997). "Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation." *J Virol*, 71(5), 3407-14.

BIBLIOGRAPHIE

- Coffin, J. M. (1979). "Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses." *J Gen Virol*, 42(1), 1-26.
- Coffin, J. M. (1996a). *Retroviridae: the viruses and their replication.*, Lipincott-Raven Publishers, Philadelphia.
- Coffin, J. M. (1996b). "The scoop on HIV mutations." *J Int Assoc Physicians AIDS Care*, 2(7), 45.
- Cornell, W. D., Cleplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules." *J Am Chem Soc*, 117(19), 5179-5197.
- Cullen, B. R. (1991). "Human immunodeficiency virus as a prototypic complex retrovirus." *J Virol*, 65(3), 1053-6.
- Damgaard, C. K., Dyhr-Mikkelsen, H., and Kjems, J. (1998). "Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions." *Nucleic Acids Res*, 26(16), 3667-76.
- 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(15), 3567-71.
- Darden, T., Toukmaji, A., and Pedersen, L. G. (1997). "Long-range electrostatic effects in biomolecular simulations." *J. Chim. Phys.*, 94, 1346-1364.
- Darden, T., York, D., and Pedersen, L. (1993). "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems." *J. Chem. Phys.*, 98(12), 10089-10092.
- 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(3), 689-99.
- De Rocquigny, H., Ficheux, D., Gabus, C., Allain, B., Fournie-Zaluski, M. C., Darlix, J. L., and Roques, B. P. (1993). "Two short basic sequences surrounding the zinc finger of nucleocapsid protein NCp10 of Moloney murine leukemia virus are critical for RNA annealing activity." *Nucleic Acids Res*, 21(4), 823-9.
- Diaz, J. F., Wroblowski, B., Schlitter, J., and Engelborghs, Y. (1997). "Calculation of pathways for the conformational transition between the GTP- and GDP-bound states of the Ha-ras-p21 protein: calculations with explicit solvent simulations and comparison with calculations in vacuum." *Proteins*, 28(3), 434-51.
- Dorfman, T., Mammano, F., Haseltine, W. A., and Gottlinger, H. G. (1994). "Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein." *J Virol*, 68(3), 1689-96.
- EMBOworkshop. (1989). "Definitions and nomenclature of nucleic acid structure parameters." *Embo J*, 8(1), 1-4.
- Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C., and Dumas, P. (2001). "Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site." *Nat Struct Biol*, 8(12), 1064-8.
- Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann, B., Ehresmann, C., and Dumas, P. (1999). "The crystal structure of

BIBLIOGRAPHIE

- the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges." *Structure Fold Des*, 7(11), 1439-49.
- Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L. (1995). "A smooth particle mesh Ewald method." *J. Chem. Phys.*, 103(19), 8577-8593.
- Ewald, P. P. (1921). "Die Berechnung optischer und elektrostatischer Gitterpotentiale." *Ann. Phys.*, 64, 253-287.
- Ferrara, P., Apostolakis, J., and Caflisch, A. (2000b). "Computer simulations of protein folding by targeted molecular dynamics." *Proteins*, 39(3), 252-60.
- Ferrara, P., Apostolakis, J., and Caflisch, A. (2000c). "Targeted Molecular Dynamics Simulations of Protein Folding." *J. Phys. Chem. B*, 104, 4511-4518.
- Ferrara, P., Apostolakis, J., Caflisch, A. (2000a). "Targeted Molecular Dynamics Simulations of Protein Unfolding." *J. Phys. Chem. B*, 104(18), 4511 -4518.
- Fosse, P., Motte, N., Roumier, A., Gabus, C., Muriaux, D., Darlix, J. L., and Paoletti, J. (1996). "A short autocomplementary sequence plays an essential role in avian sarcoma-leukosis virus RNA dimerization." *Biochemistry*, 35(51), 16601-9.
- Frankel, A. D., and Young, J. A. (1998). "HIV-1: fifteen proteins and an RNA." *Annual Review of Biochemistry*, 67, 1-25.
- Fu, W., Gorelick, R. J., and Rein, A. (1994). "Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions." *J Virol*, 68(8), 5013-8.
- Fu, W., and Rein, A. (1993). "Maturation of dimeric viral RNA of Moloney murine leukemia virus." *J Virol*, 67(9), 5443-9.
- Gaudin, F., Lancelot, G., and Genest, D. (1997). "Search for rigid sub domains in DNA from molecular dynamics simulations." *J Biomol Struct Dyn*, 15(2), 357-67.
- Gilboa, E., Mitra, S. W., Goff, S., and Baltimore, D. (1979). "A detailed model of reverse transcription and tests of crucial aspects." *Cell*, 18(1), 93-100.
- Girard, F. (2000). "Etude par RMN de la structure des dimères de la séquence SL1 initiant la dimérisation chez VIH-1Lai," Université d'Orléans, 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(6), 1145-57.
- Girard, P. M., Bonnet-Mathoniere, B., Muriaux, D., and Paoletti, J. (1995). "A short autocomplementary sequence in the 5' leader region is responsible for dimerization of MoMuLV genomic RNA." *Biochemistry*, 34(30), 9785-94.
- Giudice, E., and Lavery, R. (2003a). "Nucleic acid base pair dynamics: the impact of sequence and structure using free-energy calculations." *J Am Chem Soc*, 125(17), 4998-9.
- Giudice, E., Varnai, P., and Lavery, R. (2003b). "Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations." *Nucleic Acids Res*, 31(5), 1434-43.
- Grewe, C., Beck, A., and Gelderblom, H. R. (1990). "HIV: early virus-cell interactions." *J Acquir Immune Defic Syndr*, 3(10), 965-74.

BIBLIOGRAPHIE

- Haddrick, M., Lear, A. L., Cann, A. J., and Heaphy, S. (1996). "Evidence that a kissing loop structure facilitates genomic RNA dimerisation in HIV-1." *J Mol Biol*, 259(1), 58-68.
- Harrison, G. P., and Lever, A. M. (1992). "The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure." *J Virol*, 66(7), 4144-53.
- Harrison, G. P., Miele, G., Hunter, E., and Lever, A. M. (1998b). "Functional analysis of the core human immunodeficiency virus type 1 packaging signal in a permissive cell line." *J Virol*, 72(7), 5886-96.
- Hayashi, T., Shioda, T., Iwakura, Y., and Shibuta, H. (1992). "RNA packaging signal of human immunodeficiency virus type 1." *Virology*, 188(2), 590-9.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). "VMD - Visual Molecular Dynamics." *J. Molec. Graphics*, 14(33-38).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Kleon, M. L. (1983). "Comparison of simple potential functions for simulating liquid water." *J. Chem. Phys.*, 79, 926-935.
- Jorgensen, W. L., and Tirado-Rives, J. (1988). "The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin." *J. Am. Chem. Soc.*, 110(6), 1657-1666.
- Jossinet, F., Paillart, J. C., Westhof, E., Hermann, T., Skripkin, E., Lodmell, J. S., Ehresmann, C., Ehresmann, B., and Marquet, R. (1999). "Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding." *Rna*, 5(9), 1222-34.
- Kieken, F. (2004). "Etude structurale par RMN des changements conformationnels de la séquence SL1 de l'ARN de l'isolat VIH-1 Lai," Université d'Orléans, Orléans.
- Kong, Y., Shen, Y., Warth, T. E., and Ma, J. (2002). "Conformational pathways in the gating of Escherichia coli mechanosensitive channel." *Proc Natl Acad Sci U S A*, 99(9), 5999-6004.
- Korber, B., Hahn, B., Foley, B., Mellors, J. W., Leitner, T., Myers, G., Mc-Cutchan, F., and Kuiken, C. (1997). *Human retroviruses and AIDS. A compilation and analysis of nucleic acid and amino acid sequences.*, Theoretical Biology and Biophysics Group T-10, Los Alamos, New Mexico.
- Kruger, P., Verheyden, S., Declerck, P. J., and Engelborghs, Y. (2001). "Extending the capabilities of targeted molecular dynamics: simulation of a large conformational transition in plasminogen activator inhibitor 1." *Protein Sci*, 10(4), 798-808.
- Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. (1992). "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method." *J. Comput. Chem.*, 13, 1011-1021.
- 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(2), 397-411.
- Kung, H. J., Hu, S., Bender, W., Bailey, J. M., Davidson, N., Nicolson, M. O., and McAllister, R. M. (1976). "RD-114, baboon, and woolly monkey viral RNA's compared in size and structure." *Cell*, 7(4), 609-20.

BIBLIOGRAPHIE

- 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(45), 13464-74.
- Laughrea, M., Jette, L., Mak, J., Kleiman, L., Liang, C., and Wainberg, M. A. (1997). "Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization." *J Virol*, 71(5), 3397-406.
- Laughrea, M., Shen, N., Jette, L., and Wainberg, M. A. (1999). "Variant effects of non-native kissing-loop hairpin palindromes on HIV replication and HIV RNA dimerization: role of stem-loop B in HIV replication and HIV RNA dimerization." *Biochemistry*, 38(1), 226-34.
- 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(1), 63-91.
- 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(9), 4085-7.
- Lim, V. I. (1994). "Analysis of action of wobble nucleoside modifications on codon-anticodon pairing within the ribosome." *J Mol Biol*, 240(1), 8-19.
- Lodmell, J. S., Ehresmann, C., Ehresmann, B., and Marquet, R. (2001). "Structure and dimerization of HIV-1 kissing loop aptamers." *J Mol Biol*, 311(3), 475-90.
- Ma, J., and Karplus, M. (1997). "Molecular switch in signal transduction: reaction paths of the conformational changes in ras p21." *Proc Natl Acad Sci U S A*, 94(22), 11905-10.
- Ma, J., Sigler, P. B., Xu, Z., and Karplus, M. (2000). "A dynamic model for the allosteric mechanism of GroEL." *J Mol Biol*, 302(2), 303-13.
- Madhusoodanan, M., and Lazaridis, T. (2003). "Investigation of pathways for the low-pH conformational transition in influenza hemagglutinin." *Biophys J*, 84(3), 1926-39.
- Mangel, W. F., Delius, H., and Duesberg, P. H. (1974). "Structure and molecular weight of the 60-70S RNA and the 30-40S RNA of the Rous sarcoma virus." *Proc Natl Acad Sci U S A*, 71(11), 4541-5.
- Marquardt, D. W. (1963). "An Algorithm for Least-Squares Estimation of NonLinear Parameters." *J.Soc. Indust. Appl. Math*, 11(2), 431-441.
- 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(9), 2349-57.
- Marquet, R., Paillart, J. C., Skripkin, E., Ehresmann, C., and Ehresmann, B. (1994). "Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site." *Nucleic Acids Res*, 22(2), 145-51.
- McBride, M. S., and Panganiban, A. T. (1996). "The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures." *J Virol*, 70(5), 2963-73.

BIBLIOGRAPHIE

- McBride, M. S., Schwartz, M. D., and Panganiban, A. T. (1997). "Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation." *J Virol*, 71(6), 4544-54.
- McGrath, K. M., Hoffman, N. G., Resch, W., Nelson, J. A., and Swanstrom, R. (2001). "Using HIV-1 sequence variability to explore virus biology." *Virus Res*, 76(2), 137-60.
- Mendieta, J., Martin-Santamaria, S., Priego, E. M., Balzarini, J., Camarasa, M. J., Perez-Perez, M. J., and Gago, F. (2004). "Role of histidine-85 in the catalytic mechanism of thymidine phosphorylase as assessed by targeted molecular dynamics simulations and quantum mechanical calculations." *Biochemistry*, 43(2), 405-14.
- Miele, G., Mouland, A., Harrison, G. P., Cohen, E., and Lever, A. M. (1996). "The human immunodeficiency virus type 1 5' packaging signal structure affects translation but does not function as an internal ribosome entry site structure." *J Virol*, 70(2), 944-51.
- Mikkelsen, J. G., Lund, A. H., Kristensen, K. D., Duch, M., Sorensen, M. S., Jorgensen, P., and Pedersen, F. S. (1996). "A preferred region for recombinational patch repair in the 5' untranslated region of primer binding site-impaired murine leukemia virus vectors." *J Virol*, 70(3), 1439-47.
- 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(20), 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(6), 432-6.
- 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(3), 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(52), 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(15), 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(14), 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(1), 411-19.
- Paillart, J. C., Berthoux, L., Ottmann, M., Darlix, J. L., Marquet, R., Ehresmann, B., and Ehresmann, C. (1996a). "A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis." *J Virol*, 70(12), 8348-54.

BIBLIOGRAPHIE

- Paillart, J. C., Marquet, R., Skripkin, E., Ehresmann, C., and Ehresmann, B. (1996b). "Dimerization of retroviral genomic RNAs: structural and functional implications." *Biochimie*, 78(7), 639-53.
- Paillart, J. C., Skripkin, E., Ehresmann, B., Ehresmann, C., and Marquet, R. (1996d). "The use of chemical modification interference and inverse PCR mutagenesis to identify the dimerization initiation site of HIV-1 genomic RNA." *Pharm Acta Helv*, 71(1), 21-8.
- 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(1), 36-49.
- Pattabiraman, N., Martinez, H. M., and Shapiro, B. A. (2002). "Molecular modeling and dynamics studies of HIV-1 kissing loop structures." *J Biomol Struct Dyn*, 20(3), 397-412.
- Pavlakakis, G. (1996). "The molecular biology of HIV-1." In *AIDS: Diagnosis, Treatment and Prevention*, S. H. VT DeVita, SA Rosenberg, ed., Lippincott Raven, Philadelphia, 45-74.
- Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, I., T.E., DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995). "AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules." *Computer Physics Communications*, 91(1-3), 1-41.
- Polge, E., Darlix, J. L., Paoletti, J., and Fosse, P. (2000). "Characterization of loose and tight dimer forms of avian leukosis virus RNA." *J Mol Biol*, 300(1), 41-56.
- 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(2), 774-83.
- Quinones-Mateu, M. E., and Arts, E. J. (1999). "Recombination in HIV-1: Update and Implications." *AIDS Rev.*, 1(2), 89-100.
- Reblova, K., Spackova, N., Sponer, J. E., Koca, J., and Sponer, J. (2003). "Molecular dynamics simulations of RNA kissing-loop motifs reveal structural dynamics and formation of cation-binding pockets." *Nucleic Acids Res*, 31(23), 6942-52.
- Rein, A. (1994). "Retroviral RNA packaging: a review." *Arch Virol Suppl*, 9, 513-22.
- Rist, M. J., and Marino, J. P. (2002). "Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1." *Biochemistry*, 41(50), 14762-70.
- Rocchia, W., Alexov, E., and Honig, B. (2001). "Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions." *J Phys. Chem. B*, 105, 6507-6514.
- Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., and Honig, B. (2002). "Rapid Grid-based Construction of the Molecular Surface for both Molecules and Geometric Objects: Applications to the Finite Difference Poisson-Boltzmann Method." *J. Comp. Chem.*, 23(128-137).
- Roche, O., and Field, M. J. (1999). "Simulations of the T \leftrightarrow R conformational transition in aspartate transcarbamylase." *Protein Eng*, 12(4), 285-95.

BIBLIOGRAPHIE

- Roux, B. (1995). "The calculation of the potential of mean force using computer simulations." *Comp. Phys. Comm.*, 91, 275-282.
- Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977). "Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of N-alkanes." *J. Comput. Phys.*, 23, 327-341.
- Saenger, W. (1984). *Principles of nucleic acid structure*, Springer-Verlag New York Inc., New York.
- 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(11), 5219-23.
- Sarma, R. H. (1988). "New nomenclature for nucleic acid helix parameters Cambridge, UK, September 14, 1988." *J Biomol Struct Dyn*, 6(3), 391-5.
- Schlitter, J., Engels, M., and Kruger, P. (1994). "Targeted molecular dynamics: a new approach for searching pathways of conformational transitions." *J Mol Graph*, 12(2), 84-9.
- Shen, N., Jette, L., Liang, C., Wainberg, M. A., and Laughrea, M. (2000). "Impact of human immunodeficiency virus type 1 RNA dimerization on viral infectivity and of stem-loop B on RNA dimerization and reverse transcription and dissociation of dimerization from packaging." *J Virol*, 74(12), 5729-35.
- Simon, F., Mauclore, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barre-Sinoussi, F., and Brun-Vezinet, F. (1998). "Identification of a new human immunodeficiency virus type 1 distinct from group M and group O." *Nat Med*, 4(9), 1032-7.
- Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B., and Ehresmann, C. (1994). "Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro." *Proc Natl Acad Sci U S A*, 91(11), 4945-9.
- St Louis, D. C., Gotte, D., Sanders-Buell, E., Ritchey, D. W., Salminen, M. O., Carr, J. K., and McCutchan, F. E. (1998). "Infectious molecular clones with the nonhomologous dimer initiation sequences found in different subtypes of human immunodeficiency virus type 1 can recombine and initiate a spreading infection in vitro." *J Virol*, 72(5), 3991-8.
- Stewart, L., Schatz, G., and Vogt, V. M. (1990). "Properties of avian retrovirus particles defective in viral protease." *J Virol*, 64(10), 5076-92.
- Stoltzfus, C. M., and Snyder, P. N. (1975). "Structure of B77 sarcoma virus RNA: stabilization of RNA after packaging." *J Virol*, 16(5), 1161-70.
- Strobel, S. A., Cech, T. R., Usman, N., and Beigelman, L. (1994). "The 2,6-diaminopurine riboside.5-methylisocytidine wobble base pair: an isoenergetic substitution for the study of G.U pairs in RNA." *Biochemistry*, 33(46), 13824-35.
- Summers, M. F., Henderson, L. E., Chance, M. R., Bess, J. W., Jr., South, T. L., Blake, P. R., Sagi, I., Perez-Alvarado, G., Sowder, R. C., 3rd, Hare, D. R., and et al. (1992). "Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1." *Protein Sci*, 1(5), 563-74.

BIBLIOGRAPHIE

- Sundquist, W. I., and Heaphy, S. (1993). "Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA." *Proc Natl Acad Sci U S A*, 90(8), 3393-7.
- Takahashi, K., Baba, S., Koyanagi, Y., Yamamoto, N., Takaku, H., and Kawai, G. (2001). "Two basic regions of NCp7 are sufficient for conformational conversion of HIV-1 dimerization initiation site from kissing-loop dimer to extended-duplex dimer." *J Biol Chem*, 276(33), 31274-8.
- Takahashi, K. I., Baba, S., Chattopadhyay, P., Koyanagi, Y., Yamamoto, N., Takaku, H., and Kawai, G. (2000). "Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome." *Rna*, 6(1), 96-102.
- Termin, H. M. (1993). "Retrovirus variation and reverse transcription: Abnormal strand transfers result in retrovirus genetic variation." *Proc Natl Acad Sci U S A*, 90, 6900-6903.
- 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(9), 2711-9.
- Toukmaji, A., Paul, D., Board, J.A. (1996). "Distributed Particle-Mesh Ewald: A Parallel Ewald Summation Method." *PDPTA*, 33-43.
- Van Gunsteren, W. F., and Berendsen, H. J. C. (1977). "Algorithms for macromolecular dynamics and constraint dynamics." *Mol. Phys.*, 34, 1311-1327.
- Varnai, P., and Lavery, R. (2002). "Base flipping in DNA: pathways and energetics studied with molecular dynamic simulations." *J Am Chem Soc*, 124(25), 7272-3.
- Verlet, L. (1967). "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules." *Phys. Rev.*, 159(1), 98-103.
- Weiner, P. K., and Kollman, P. A. (1981). "AMBER: Assisted Model Building with Energy Refinement. A General Program for Modeling Molecules and Their Interactions." *J. Comp. Chem.*, 2, 287-303.
- Windbichler, N., Werner, M., and Schroeder, R. (2003). "Kissing complex-mediated dimerisation of HIV-1 RNA: coupling extended duplex formation to ribozyme cleavage." *Nucleic Acids Res*, 31(22), 6419-27.
- Wroblowski, B., Diaz, J. F., Schlitter, J., and Engelborghs, Y. (1997). "Modelling pathways of alpha-chymotrypsin activation and deactivation." *Protein Eng*, 10(10), 1163-74.
- Yang, L., Beard, W., Wilson, S., Roux, B., Broyde, S., and Schlick, T. (2002). "Local deformations revealed by dynamics simulations of DNA polymerase Beta with DNA mismatches at the primer terminus." *J Mol Biol*, 321(3), 459-78.
- Young, J. A. T. (2003). *Cellular Factors Involved in Early Steps of Retroviral Replication*, Springer.
- Yuan, Y., Kerwood, D. J., Paoletti, A. C., Shubsda, M. F., and Borer, P. N. (2003). "Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1 x 3 internal loop." *Biochemistry*, 42(18), 5259-69.