



**HAL**  
open science

# Modulation de l'interaction intégrase/ADN du VIH-1 par des dérivés des styrylquinoléines et par la modification de nucléotides

Sophie Barbe

► **To cite this version:**

Sophie Barbe. Modulation de l'interaction intégrase/ADN du VIH-1 par des dérivés des styrylquinoléines et par la modification de nucléotides. Biochimie [q-bio.BM]. École normale supérieure de Cachan - ENS Cachan, 2006. Français. NNT: . tel-00129455

**HAL Id: tel-00129455**

**<https://theses.hal.science/tel-00129455>**

Submitted on 7 Feb 2007

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

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

**THESE DE DOCTORAT  
DE L'ECOLE NORMALE SUPERIEURE DE CACHAN**

Présentée par

Mademoiselle Sophie Barbe

**pour obtenir le grade de**

**DOCTEUR DE L'ECOLE NORMALE SUPERIEURE DE CACHAN**

Domaine :  
Modélisation moléculaire

**Sujet de la thèse :**

**Modulation de l'interaction Intégrase/ADN du VIH-1 par des  
dérivés des Styrylquinoléines et par la modification de nucléotides**

Thèse présentée et soutenue à Cachan le 19 Septembre 2006 devant le jury composé de :

AUCLAIR Christian	Professeur	Président
GHERMANI Nour Eddine	Professeur	Rapporteur
AUFFINGER Pascal	Chargé de recherche	Rapporteur
MOUSCADET Jean-François	Directeur de recherche	Examineur
GOTTIKH Marina	Directeur de recherche	Examineur
LE BRET Marc	Directeur de recherche	Directeur de thèse

Laboratoire de Biotechnologies et Pharmacologie génétique Appliquée (LBPA)  
ENS CACHAN/CNRS/UMR 8113  
61, avenue du Président Wilson, 94235 CACHAN CEDEX (France)



## Remerciements

*Je tiens tout d'abord à adresser mes plus vifs remerciements au Dr. Marc LE BRET qui a encadré avec bienveillance mon travail de thèse. Je lui suis très reconnaissante de m'avoir fait profiter de la richesse et de la pertinence de ses connaissances scientifiques. Je suis également sensible à la confiance qu'il m'a témoignée.*

*J'aimerais ensuite exprimer ma profonde gratitude au Pr. Christian AUCLAIR pour m'avoir accueillie dans son laboratoire et donné les moyens nécessaires pour mener à bien cette thèse. Par ailleurs, je suis tout particulièrement sensible à l'honneur qu'il me fait de présider mon jury.*

*J'adresse mes sincères remerciements au Pr. Nour Eddine GHERMANI pour l'honneur qu'il me fait d'accepter d'être rapporteur de ce travail.*

*Que le Dr. Pascal AUFFINGER trouve ici le témoignage de ma profonde gratitude pour avoir accepté d'être rapporteur de cette thèse.*

*Je tiens à remercier le Dr. Jean-François MOUSCADET, le Dr. Marina GOTTIKH et le Dr. Eric DEPREZ avec lesquels j'ai réalisé des collaborations. Leur écoute et leurs conseils avisés m'ont été d'une aide précieuse tout au long de ce travail. Qu'ils trouvent ici les marques de ma reconnaissance et de mon respect. Merci enfin au Dr. Jean-François MOUSCADET et au Dr. Marina GOTTIKH de me faire l'honneur de participer au jury de cette thèse.*

*J'exprime toute mon amitié à Jérémie VENDÔME avec qui j'ai partagé mon quotidien au laboratoire. Ses qualités humaines et ses encouragements m'ont été d'un grand soutien pendant ces années.*

*Merci à l'ensemble des personnes du LBPA pour leur gentillesse et pour l'atmosphère de travail agréable qu'ils ont su créer au sein du laboratoire.*

*Enfin, je voudrais remercier ma famille qui m'apporte un grand soutien depuis plusieurs années. En espérant que le fruit de ce travail traduise en partie la récompense de tous leurs efforts et leur dévouement.*



*A Marine ma fille,  
A ma mère, à mon père,  
A tous ceux qui me sont chers...*



## Table des matières

<b>Table des abréviations.....</b>	<b>4</b>
<b>PRESENTATION GENERALE DU PROJET DE RECHERCHE.....</b>	<b>6</b>
<b>INTRODUCTION.....</b>	<b>9</b>
<b>Chapitre I : Le virus de l'immunodéficience humaine de type I.....</b>	<b>10</b>
<i>I- Caractéristiques générales du VIH.....</i>	<i>11</i>
<i>II- Organisation génétique et structurale du VIH-1.....</i>	<i>12</i>
A - Le génome viral.....	12
A-1- Séquences LTR.....	12
A-2- Régions codantes.....	13
B - La particule virale.....	14
<i>III- Cycle de réplication du VIH-1 et thérapies anti-VIH.....</i>	<i>14</i>
A - Les étapes du cycle réplicatif.....	14
A-1- Reconnaissance et pénétration dans la cellule hôte.....	15
A-2- Rétrotranscription du génome viral et intégration dans le génome de la cellule hôte.....	15
A-3- Transcription de l'ADN proviral et expression des protéines virales.....	16
A-4- Formation et maturation de nouvelles particules virales.....	17
B- Les thérapies anti-VIH actuelles.....	18
<b>Chapitre II : Le rôle et la structure de l'intégrase du VIH-1.....</b>	<b>19</b>
<i>I- Activités catalytiques de l'intégrase.....</i>	<i>19</i>
A - Le rôle de l'IN <i>in vivo</i> .....	19
B - Les tests <i>in vitro</i> de l'activité catalytique de l'IN.....	21
<i>II- Structure de différents fragments de l'intégrase et leurs fonctionnalités.....</i>	<i>23</i>
A- Le domaine N-terminal (résidus 1 à 50).....	25
A-1- Fonctionnalités du domaine N-terminal.....	25
A-2- Structure du domaine N-terminal.....	26
B- Le domaine central (résidus 50 à 212).....	28
B-1- Fonctionnalités du domaine central.....	28
B-2- Structure du domaine central.....	28
B-3- Comparaison avec des domaines centraux d'autres IN.....	32
B-4- Liaison d'ions métalliques divalents au site actif de l'IN.....	33
C - Le domaine C-terminal (résidus 212 à 288).....	35
C-1- Fonctionnalités du domaine C-terminal.....	35
C-2- Structure du domaine C-terminal.....	35
D - Les bidomaines de l'IN.....	37
D-1- Le domaine central lié au domaine N-terminal (résidus 1 à 212).....	37
D-2- Le domaine central lié au domaine C-terminal (résidus 50 à 288).....	38
<i>III- Oligomérisation de l'intégrase.....</i>	<i>39</i>





<b>Chapitre III : L'interaction de l'intégrase avec ses substrats et son mécanisme d'action .....</b>	<b>42</b>
<i>I- Généralités sur les interactions protéines/ADN .....</i>	42
<i>II- Interaction IN/ADN .....</i>	44
A - Etudes expérimentales de l'interaction IN/ADN .....	44
A-1- Effets des modifications de l'ADN viral sur l'activité catalytique de l'IN..	44
A-2- Expériences de complémentation fonctionnelle, de mutagenèse, d'empreinte protéique et de pontage.....	47
B - Exploration des sites de liaison de l'ADN sur l'IN par docking de dinucléotides .....	50
C - Modèles du complexe IN/ADN .....	51
<i>III- Modèles du mécanisme d'action de l'intégrase .....</i>	53
A - Les trois réactions sont de type SN2 et peuvent se réaliser au même endroit .....	53
B - Modèles atomiques .....	55
 <b>Chapitre IV : Les inhibiteurs de l'intégrase du VIH-1 .....</b>	<b>58</b>
<i>I- Peptides.....</i>	59
<i>II- Dinucléotides.....</i>	60
<i>III- Oligonucléotides .....</i>	60
<i>IV- Dicéto-acides et leurs dérivés .....</i>	61
<i>V- Polyphénols .....</i>	64
A- Des polyphénols naturels .....	64
A-1- Les Flavonoïdes.....	64
A-2- Les esters d'acide caféique .....	65
B- Des polyphénols synthétiques .....	66
B-1- Les dérivés dicaféoyliques.....	66
B-2- Les salicylhydrazides et mercaptosalicylhydrazides.....	67
B-3- Les dérivés des styrylquinoléines (SQ).....	68
B-4- D'autres dérivés quinoléiques.....	69
<i>VI- Des dérivés de l'acide naphthalènesulfonique.....</i>	69
<i>VII- Divers composés .....</i>	70
A- Les produits naturels .....	70
B- Les molécules synthétiques .....	70
 <b>RESULTATS ET DISCUSSION .....</b>	<b>72</b>
<b>PARTIE I.....</b>	<b>73</b>
<b>Mécanisme d'inhibition de l'IN du VIH-1 par des dérivés des Styrylquinoléines (SQ)</b>	
 <b>PARTIE II.....</b>	<b>90</b>
<b>Conformation et flexibilité de nucléosides différemment substitués en 2' et effets d'analogues de l'ADN viral sur l'activité de 3'-processing de l'IN</b>	
<i>Chapitre I : Détermination ab initio de la flexibilité de 2'-aminoribonucléosides et de 2'-aminoarabinonucléosides .....</i>	92



---

<i>Chapitre II : Etude ab initio de l'hydratation et de la flexibilité de l'uridine, la 2'-désoxyuridine et la 2'-méthoxyuridine</i> .....	125
<i>Chapitre III : Etude de l'interaction IN/ADN en utilisant des analogues de l'ADN viral</i> .....	160
<b>PARTIE III</b> .....	178
<b>Structure fine d'analogues de l'extrémité LTR U5 de l'ADN viral</b>	
<b>CONCLUSION</b> .....	205
<b>ANNEXE</b> .....	209
<b>Structure et conformations des doubles hélices d'acides nucléiques</b>	
<i>I- Les bases</i> .....	210
<i>II- Les sucres</i> .....	214
<i>III- Les nucléosides et les nucléotides</i> .....	216
<i>IV- Les chaînes polynucléotidiques</i> .....	218
<i>V- Corrélation entre les différents paramètres angulaires</i> .....	219
<i>VI- Les doubles hélices</i> .....	220
<b>REFERENCES BIBLIOGRAPHIQUES</b> .....	227



## Table des abréviations

<b>A</b>	
ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
ARNm	ARN messenger
ARNt	ARN de transfert
ASV	« Avian sarcoma virus »
<b>B</b>	
BAF	« Barrier-to-Autointegration Factor »
BIV	« Bovine immunodeficiency virus »
<b>C</b>	
CA	« Capsid protein »
CC <sub>50</sub>	Concentration de drogue réduisant de 50% la viabilité d'un stock de cellules infectées après 5 jours en présence de la molécule
CCR5	« Chemokine (C-C motif) receptor 5 »
CXCR4	« Chemokine (C-X-C motif) receptor 4 »
<b>D</b>	
DKA	Dérivés des dicéto-acides
<b>E</b>	
<i>env</i>	« envelope »
EC <sub>50</sub>	Concentration de drogue entraînant 50% de son effet maximum <i>ex vivo</i> .
<b>F</b>	
FDA	« Food and Drug Administration »
FIV	« Feline immunodeficiency virus »
<b>G</b>	
<i>gag</i>	« group-specific antigen »
<b>H</b>	
HAART	« Highly Active Antiretroviral Therapy »
HFV	« Human Foamy Virus »
HLA	« Human Leucocyte Antigens »
HMGA1	« High mobility group AT-hook 1 »
HSP60	« Heat Shock Protein 60 »
HTH	« Helix-Turn-Helix »
<b>I</b>	
IC <sub>50</sub>	Concentration de drogue nécessaire pour obtenir 50% d'inhibition
IN	Intégrase
INI1	« Integrase interactor-1 »
<b>K</b>	
Kd	Constante de dissociation
<b>L</b>	
LEDGF	« Lens Epithelium-Derived Growth Factor »
LTR	« Long terminal repeat »
<b>M</b>	
MA	« Matrix protein »
MLV	« Murine Leukaemia virus »
Mo-MuLV	« Murine Molloney Leukaemia Virus »



<b>N</b>	
NC	« Nucleocapsid protein »
NF-κB	« Nuclear Factor-Kappa Beta »
NLS	« Nuclear localization signal »
<b>P</b>	
PCR	« Polymerase Chain Reaction »
PDB	« Proteine Databank Brookhaven »
PIC	« Preintegration complex »
PML	« Protein promyelocytic leukaemia »
<i>pol</i>	« Polymerase »
PR	« Protease »
P-TEBb	« Positive Transcription Elongation Factor b »
<b>R</b>	
R	« Redundant » séquence redondante située aux deux extrémités de l'ADN viral
RMN	Résonance Magnétique Nucléaire
RNase H	Ribonucléase H
RuvC	« Holliday junction endodeoxyribonuclease »
RSV	« Rous sarcoma virus »
RT	«Reverse transcriptase »
RTC	« Retrotranscription complex »
<b>S</b>	
SIDA	Syndrôme d'immuno-déficience acquise
SIV	« Simian immunodeficiency virus »
SN2	Substitution Nucléophile d'ordre 2
SQ	Dérivés des styrylquinoléines
SU	Protéine de surface
<b>T</b>	
TAR	« Trans-acting responsive element »
Tat	« TransActivator of Transcription»
TM	Protéine transmembranaire
Tsg101	Potein of « Tumor susceptibility gene 101 »
<b>U</b>	
U3	« Unique sequence element at the 3' end of the viral RNA »
U5	« Unique sequence element at the 5' end of the viral RNA »
<b>V</b>	
Vif	« Viral infectivity factor »
VIH-1	Virus de l'immunodéficience humaine de type 1
VIH-2	Virus de l'immunodéficience humaine de type 2
Vpr	« Viral protein R »
Vpu	« Viral protein U »





# ***PRESENTATION***

---



## **PRESENTATION GENERALE DU PROJET DE RECHERCHE**

Les études biologiques réalisées ces 50 dernières années ont mis en évidence l'étroite relation entre la fonction des molécules biologiques, leur structure et leurs mouvements intramoléculaires.

La biologie structurale expérimentale, en particulier grâce aux techniques comme la cristallographie aux rayons X ou la résonance magnétique nucléaire, a permis d'accéder à la connaissance moléculaire voire atomique de structures biologiques. Cependant, ces techniques expérimentales ne permettent pas d'acquérir une description complète des molécules et de leur dynamique à une échelle atomique, souvent essentielle à la compréhension de certains processus biologiques. C'est ainsi que se sont développées des techniques *in silico* visant à fournir une telle description.

Aujourd'hui, parallèlement à l'accroissement considérable des données issues de la biologie structurale expérimentale, la modélisation moléculaire et la chimie computationnelle ont prouvé leur intérêt pour la compréhension du fonctionnement de molécules biologiques en relation avec leur structure.

Le projet de recherche dans lequel s'inscrit ce travail de thèse nécessite l'implication de ce champ disciplinaire théorique, parmi les autres disciplines mises à contribution. Ce programme de recherche porte sur l'étude biologique et pharmacologique d'une enzyme du VIH-1 : l'intégrase (IN) qui est responsable de l'insertion de l'ADN viral dans le génome de la cellule infectée par le virus. Dans l'introduction, nous présenterons le VIH-1, l'agent causal du SIDA et plus en détail l'enzyme virale, l'IN : son rôle, sa structure, son interaction avec ses substrats et ses inhibiteurs.

Les objectifs de ce programme de recherche sont de déterminer :

- le mécanisme d'action de molécules inhibitrices de l'IN
- la spécificité de l'interaction de l'IN avec son substrat ADN viral afin d'orienter la conception de drogues contre cette cible moléculaire.



Au sein de ce projet, le travail de thèse a consisté à :

- déterminer le mode de liaison d'inhibiteurs de l'IN (des dérivés des styrylquinoléines) et évaluer la stabilité de ces interactions
- prédire la structure et la flexibilité de nucléosides modifiés en 2' afin de comprendre les effets d'analogues de l'ADN viral sur l'activité de l'IN et donc sur l'interaction IN/ADN.
- étudier la structure de l'extrémité de l'ADN viral impliqué dans la spécificité de l'interaction avec l'IN et d'analogues de ce fragment nucléique.

L'ensemble de ces travaux nous a donc amenés à réaliser une étude plus fondamentale sur la conformation et la flexibilité d'analogues de nucléosides ainsi que sur la structure de doubles hélices d'ADN contenant des sucres modifiés en 2'.

Notre approche a été structurale et théorique (par des méthodes de modélisation moléculaire et de chimie computationnelle) et s'est effectuée en étroite collaboration avec des chimistes et des biologistes avec lesquels les échanges ont grandement dynamisé le déroulement de nos recherches.



# ***INTRODUCTION***

---

*Il s'agit dans cette partie de présenter le contexte biologique et pharmacologique du programme de recherche dans lequel s'inscrit ce travail de thèse afin de permettre au lecteur de mieux appréhender le sujet dans son ensemble.*





## **Chapitre I**

### **Le virus de l'immunodéficience humaine de type I**

---

En 1981, sont rapportés aux Etats-Unis (Hymes et al., 1981) les premiers cas de patients atteints du syndrome de l'immunodéficience acquise : le SIDA (Gottlieb et al., 1981). Rapidement, ce fléau s'est propagé jusqu'aux régions les plus éloignées du globe, entraînant un nombre croissant de contamination et de de décès.

Depuis la découverte en 1983 du VIH, l'agent causal du SIDA (Barre-Sinoussi et al., 1983; Gallo et al., 1983), la communauté scientifique a mis en place un nombre important de programmes de recherche pour lutter contre cette maladie. Malgré des progrès scientifiques considérables tels que l'élaboration de tests de diagnostic en 1985, le séquençage du génome viral moins de 15 ans plus tard et le développement de traitements antirétroviraux hautement actifs (HAART) à partir de 1996, l'épidémie de SIDA a continué à se propager. En 2005, elle a causé plus de 3 millions de décès. Parallèlement, environ 5 millions de personnes ont contracté le VIH cette même année, ce qui porte à environ 40 millions, le nombre total de malades dans le monde.

Le développement des multithérapies, HAART, qui combinent plusieurs types de molécules actives contre le VIH, a permis d'augmenter l'espérance de vie des patients. Ces traitements sont actuellement prescrits à vie. Cependant, leur efficacité est contrecarrée, à plus ou moins brève échéance, par l'apparition de souches mutantes du virus résistantes aux traitements. De plus, des effets secondaires à court et à long terme sont fréquents et l'interruption du traitement entraîne une recrudescence de la virulence de la maladie.

La recherche de nouveaux médicaments et de nouvelles cibles thérapeutiques reste donc d'une grande importance dans la lutte contre le SIDA.



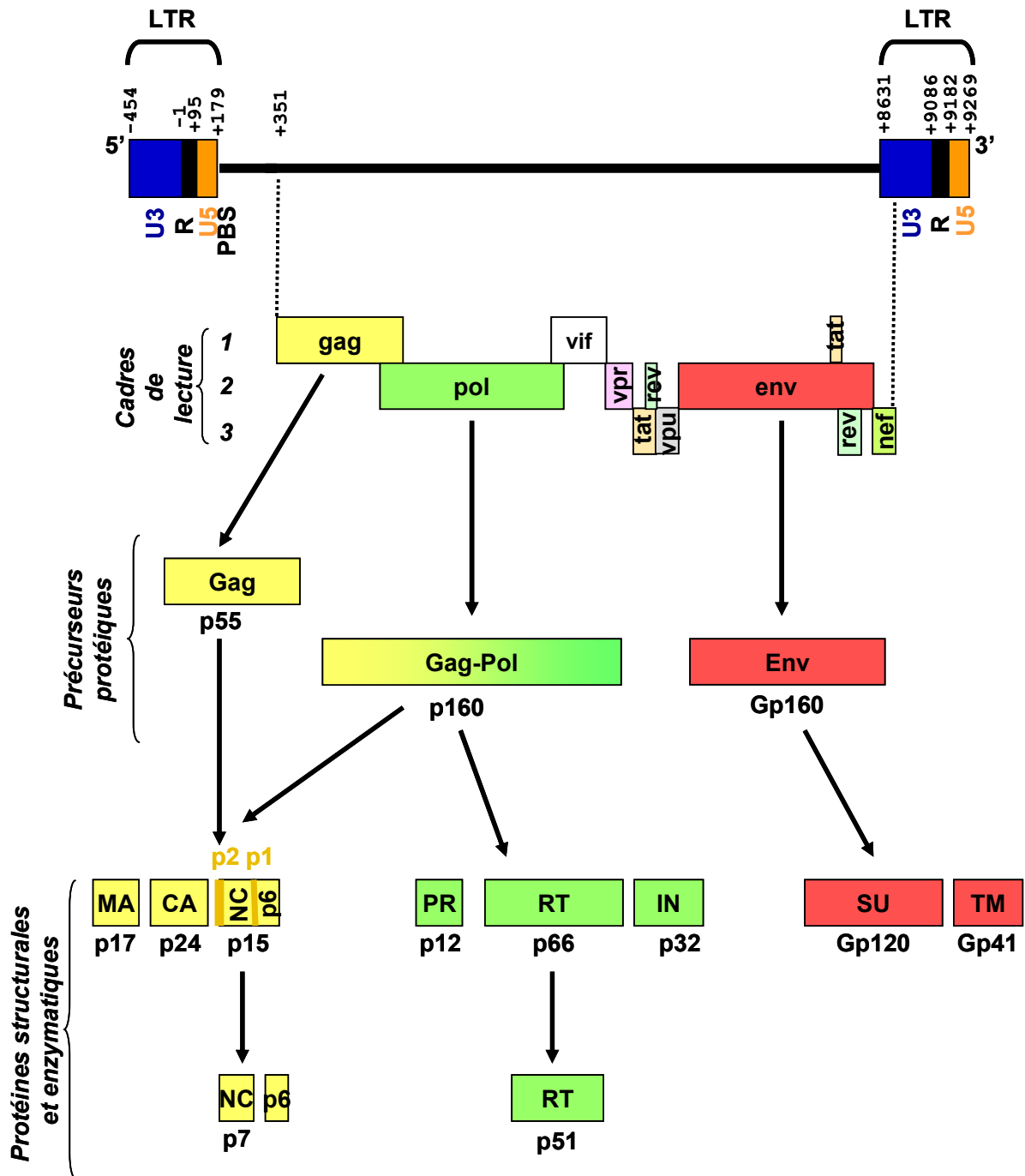
## **I- Caractéristiques générales du VIH**

Le VIH est un rétrovirus (famille des Retroviridae) du genre lentivirus. Le terme rétrovirus désigne un virus dont le génome est formé d'ARN simple brin. Afin de pouvoir infecter leurs cibles, les rétrovirus doivent rétrotranscrire l'ARN simple brin en ADN double brin. Cette opération est réalisée par une enzyme qui leur est spécifique : la rétrotranscriptase (RT). L'ADN produit peut alors être intégré au génome d'une cellule hôte sous l'action d'une autre enzyme virale, l'intégrase (IN).

Les rétrovirus du genre lentivirus infectent différentes espèces animales, par exemple les primates (VIH, SIV), les félins (FIV), les bovins (BIV)... Ils provoquent la mort des cellules infectées. Le VIH s'attaque presque exclusivement aux cellules humaines porteuses des récepteurs CD4 (cellules CD4+). Il s'agit principalement des lymphocytes T et des macrophages qui constituent les piliers du système de défense immunitaire de l'organisme. Il existe une phase de latence relativement longue entre l'infection et les premiers symptômes de la maladie (jusqu'à 11 ans). Ces symptômes sont corrélés à un état d'immunodépression aiguë, favorisant le développement de maladies auto-immunes et l'apparition de maladies opportunistes.

Il existe deux types de VIH : le VIH-1 (Barre-Sinoussi et al., 1983; Gallo et al., 1983) et le VIH-2 (Clavel et al., 1986) qui se subdivisent en plusieurs sous-types. Le VIH-1 est la forme la plus courante dans les pays industrialisés tandis que le VIH-2 sévit principalement en Afrique Subsaharienne. La variété dans le patrimoine génétique du VIH résulte des nombreuses erreurs d'incorporation de nucléotides qu'effectue la rétrotranscriptase pendant la transcription de l'ARN viral en ADN (McGrath et al., 2001). Cette enzyme, dépourvue de système de relecture, présente un taux d'erreurs de l'ordre de  $10^{-3}$  à  $10^{-4}$ , ce qui correspond à une ou deux mutations par cycle de réplication. Les mutations constamment produites émergent sous la pression de sélection introduite par le système immunitaire et les multithérapies. A cela, s'ajoutent des formes recombinantes, résultant d'une recombinaison génétique entre des virus de sous-types différents.

Des études phylogéniques ont permis d'établir les relations entre les différents lentivirus de primates (HIV sequence Compendium 2005). Ces comparaisons de séquences laissent supposer une transmission inter-espèces de différents singes à l'homme (Gao et al., 1999). Cette hypothèse était également suggérée par les premiers cas sporadiques de SIDA. On trouve les premières traces du VIH dans le plasma d'un adulte de la République démocratique



**Figure 1-** Organisation génétique de l'ADN proviral du VIH-1. L'ADN proviral est encadré de deux séquences LTR, composées des régions U3 (bleu), R (noir) et U5 (orange). Les cadres de lecture permettant d'exprimer les différents gènes sont notés. La partie inférieure de la figure montre le devenir des trois précurseurs polyprotéiques (Gag, Gag-Pol et Env), protéolysés en protéines structurales et enzymatiques.

du Congo mort en 1959 et dans les tissus d'un marin norvégien mort en 1974. Ce dernier aurait été infecté dans les années 1961-1965, lors de voyages sur les côtes d'Afrique occidentale. L'analyse d'échantillons de fèces de chimpanzés de la forêt tropicale camerounaise a identifié le chimpanzé Pan troglodyte comme le réservoir naturel du VIH-1 (Keele et al., 2006).

## **II- Organisation génétique et structurale du VIH-1**

### **A - Le génome viral**

L'ARN génomique du VIH-1 est diploïde ; il est composé de deux molécules homologues d'ARN simple brin de 9,2 kb liées de façon non covalente, à proximité de leur extrémité 5'. Cet ARN est rétrotranscrit par la RT en un ADN double brin linéaire qui s'intègre dans le génome de la cellule hôte : on parle alors d'ADN proviral.

#### **A-1- Séquences LTR**

L'ADN rétrotranscrit est bordé de part et d'autre d'une séquence LTR non codante (**figure 1**). Chaque LTR se subdivise en séquences U3, R et U5. Ces extrémités sont essentielles pour la transcription de l'ADN proviral. La séquence U3 (-454 à -1) contient :

- un promoteur (-78 à -1) dont la structure est celle d'un promoteur eucaryote qui fixe l'ARN polymérase II et ses cofacteurs,
- un domaine activateur (-104 à -79) et un domaine modulateur (-454 à -105) qui fixent des facteurs transcriptionnels cellulaires stimulant la transcription.

La séquence R (+1 à +95) inclut un domaine, TAR, impliqué également dans l'activation transcriptionnelle (Karn, 1999). La transcription de l'ADN proviral par la machinerie cellulaire débute en 5' de la séquence R située dans le LTR 5' et prend fin en 3' de la séquence R contenue dans la séquence LTR 3'.

Les LTR sont également essentiels pour l'intégration de l'ADN viral dans le génome de la cellule hôte. En effet, les séquences U3 et U5 possèdent des sites de reconnaissance pour la protéine virale, l'intégrase qui permet l'insertion de l'ADN viral dans le génome cellulaire.



## A-2- Régions codantes

L'ADN proviral comporte 3 gènes (*gag*, *pol* et *env*) qui codent des précurseurs polyprotéiques (**figure 1**). Le clivage de ces précurseurs par la protéase virale ou une protéase cellulaire libère les protéines structurales et enzymatiques du virus.

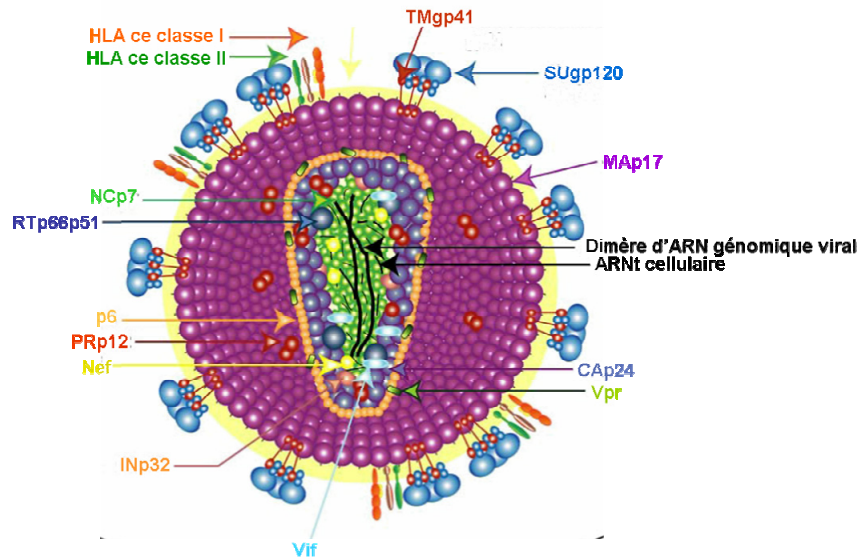
Le gène *gag* code un précurseur Gag de 55 kDa (Pr55<sup>Gag</sup>) myristylé. Cette polyprotéine est traduite à partir d'un ARNm non-épissé. Le clivage de ce précurseur par la protéase virale (PR), au cours de la maturation de la particule virale, libère les protéines structurales de matrice (MAp17), de nucléocapside (NCp7), de capsid (CAp24), p1, p2 et p6.

Le gène *pol* code une polyprotéine qui génère trois protéines enzymatiques : la protéase (PRp12), l'intégrase (INp32) et la rétrotranscriptase (RTp66/p51). PR est libérée par un mécanisme d'autocatalyse alors que l'IN et la RT sont obtenues grâce au clivage de la polyprotéine par PR. Le gène *pol* n'est jamais exprimé de manière indépendante mais sous forme d'une polyprotéine de fusion Gag-Pol (Pr160<sup>Gag-Pol</sup>). Ce précurseur est obtenu par un décalage -1 du cadre de lecture en fin de synthèse de Gag (Jacks et al., 1988). Cet évènement qui est déclenché par un motif d'ARN spécifique (Parkin et al., 1992) ne se produit que pour environ 5% des ARN viraux traduits ; il a été récemment décrit en détail (Namy et al., 2006). Par la suite, la protéase virale sépare les domaines Gag et Pol.

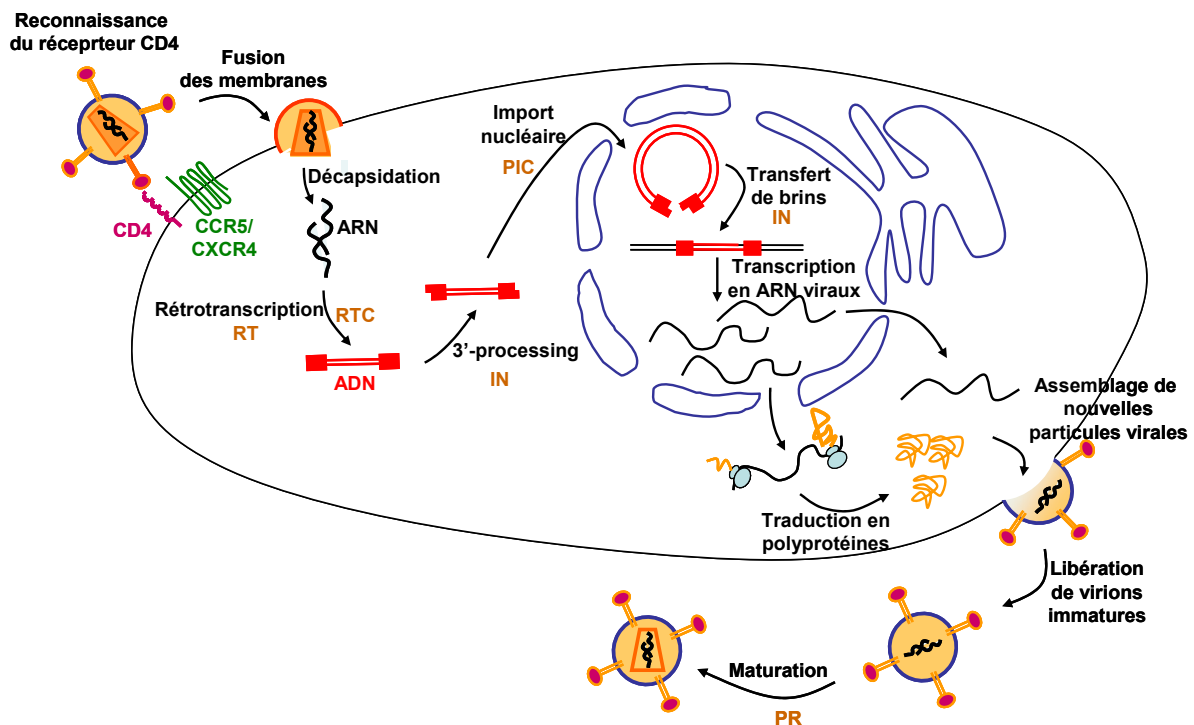
Le gène *env* code un précurseur Env de 160 kDa (Pr160<sup>Env</sup>) glycosylé (addition de 25 à 30 chaînes d'hydrates de carbone). Cette polyprotéine est exprimée à partir d'un ARNm mono-épissé. Le clivage de ce précurseur, au cours de son transport vers la membrane plasmique par une protéase cellulaire, libère la protéine de surface (SUgp120) et la protéine transmembranaire (TMgp41).

En plus de ces trois gènes communs à tous les rétrovirus, le génome du VIH-1 possède plusieurs autres gènes spécifiques des rétrovirus complexes, codant des protéines régulatrices et auxiliaires (**figure 1**). Les gènes *tat* et *rev* codent respectivement une protéine régulatrice de 14 kDa, Tat et une protéine régulatrice de 18 kDa, Rev. Les gènes *nef*, *vif*, *vpr* et *vpu* codent respectivement les protéines auxiliaires suivantes : Nef (25 kDa myristylée), Vif (23 kDa), Vpr (15 kDa) et Vpu (16 kDa).





**Figure 2-** Structure de la particule virale du VIH-1



**Figure 3-** Cycle de réplication du VIH-1. Après l'absorption du virus sur la membrane cellulaire par fixation sur le CD4 et l'un des deux co-récepteurs (CCR5/CXCR4), le virus pénètre dans la cellule par fusion entre l'enveloppe virale et la membrane cellulaire. Le génome viral sous forme de deux molécules d'ARN est transcrit par la RT en ADN. Une étape préparatoire de cet ADN, le 3'-processing est ensuite catalysée par l'IN. Puis l'ADN migre via un complexe de pré-intégration (PIC) dans le noyau où il est intégré définitivement dans l'ADN cellulaire par l'IN. Après la transcription et la traduction, les ARN viraux et les protéines viraux s'assemblent pour former une particule virale qui bourgeonne à la membrane cellulaire. Ces particules deviennent infectieuses après maturation par la PR qui clive les précurseurs polyprotéiques en protéines structurales et en enzymes.

## **B - La particule virale**

La particule virale mature et infectieuse se présente sous une forme sphérique dont le diamètre varie entre 110 et 130 nm ; elle a une masse d'environ  $2.5 \cdot 10^8$  Da (**figure 2**).

Son enveloppe externe est constituée d'une bicouche lipidique riche en cholestérol, associée à des protéines d'origine cellulaire (telles que les molécules HLA de classe I et II (Ott, 2002)) et des glycoprotéines virales de surface (SUgp120) et transmembranaires (TMgp41). Ces deux glycoprotéines forment des trimères à la surface de la particule virale : chaque monomère étant constitué d'une sous-unité SU et d'une sous-unité TM associées de façon non covalente

Sous l'enveloppe virale, la répétition de la protéine de matrice (MAp17) myristylée à son extrémité N-terminale forme la partie externe du cœur viral. Cette protéine est sous forme monomérique, séquestrant le groupe myristique (Tang et al., 2004). Son hydrophobicité l'amène à interagir avec la bicouche lipidique.

A l'intérieur de la matrice se trouve la capsid virale de forme conique, construite par l'assemblage de protéines de capsid (CAp24). Elle renferme les deux brins d'ARN qui constituent le génome viral, associés à une protéine de nucléocapsid (NCp7). Elle contient également les enzymes virales indispensables à la réplication (RT, PR et IN), d'autres protéines virales (p6, p2, p1, Vif, Nef et Vpr) ainsi que des protéines cellulaires (actine, cyclophiline A...) et des ARNt cellulaires.

## **III- Cycle de réplication du VIH-1 et thérapies anti-VIH**

### **A - Les étapes du cycle répliatif**

Le VIH infecte une cellule hôte dont il détourne les fonctions en vue de produire de nouveaux virions. Le cycle répliatif du VIH-1 (**figure 3**) comprend plusieurs étapes principales : la reconnaissance de la cellule hôte, la pénétration du virus au sein de la cellule, la rétrotranscription du génome viral en un ADN double brin, l'intégration de cet ADN dans le génome cellulaire, la transcription et la traduction des gènes viraux , l'assemblage et la libération de virions immatures et enfin la maturation de ces nouvelles particules.



### A-1- Reconnaissance et pénétration dans la cellule hôte

La première étape du cycle répliatif du VIH-1 consiste en une interaction peu spécifique entre une boucle (V3) de la glycoprotéine de surface, SUgp12, chargée positivement (Kwong et al., 1998) et le sulfate d'héparan chargé négativement et présent à la surface de nombreuses lignées cellulaires (Mondor et al., 1998; Ugolini et al., 1999). Le VIH-1 s'attaque cependant, de manière quasiment exclusive, à des lignées cellulaires porteuses du récepteur CD4 (Klatzmann et al., 1984). Cette reconnaissance spécifique est également l'œuvre de SU et implique un nombre limité de résidus conservés (Kowalski et al., 1987). Un grand nombre de contacts entre le récepteur CD4 et SU sont par contre réalisés grâce au squelette peptidique de la glycoprotéine (Kwong et al., 1998).

Cette interaction entre SU et le récepteur CD4 induit chez la glycoprotéine une série de changements conformationnels qui lui permettent d'interagir avec les récepteurs aux chimiokines. Le récepteur CXCR4 permet l'entrée des souches virales responsables de l'infection des lymphocytes T (ces souches sont alors dites X4 ou T-tropiques). Le récepteur CCR5 est employé par les souches virales s'attaquant aux macrophages (souches dites R5 ou macrophages-tropiques). Les deux récepteurs peuvent être utilisés par des souches dites à double tropisme (R5/X4). Des individus homozygotes  $\Delta 32$ -CCR5, possédant un gène CCR5 déficient, présentent une résistance innée contre une infection par des souches virales R5 qui sont le plus fréquemment transmises (Liu et al., 1996; Martinson et al., 1997).

Les interactions entre SU, le récepteur CD4 et les récepteurs aux chimiokines induisent un autre changement de conformation, cette fois chez la glycoprotéine TMgp41 qui est le médiateur majeur de l'entrée du virus (Chan and Kim, 1998). L'ouverture du trimère TM expose trois domaines de fusion peptidiques lipophiles qui s'incrudent dans la membrane cellulaire. Ils provoquent la fusion des membranes du virus et de la cellule hôte et permettent ainsi à la capsid virale de pénétrer dans la cellule.

### A-2- Rétrotranscription du génome viral et intégration dans le génome de la cellule hôte

La capsid virale libérée dans la cellule hôte est déstructurée. Un complexe viral nucléoprotéique appelé complexe de rétrotranscription (RTC) est ainsi libéré dans le cytoplasme de la cellule. Il est composé non seulement de la rétrotranscriptase et du génome



viral mais aussi d'autres protéines virales (IN, MA, NC et Vpr), de plusieurs protéines de la cellule hôte et d'un ARNt utilisé comme amorce lors de la rétrotranscription.

La rétrotranscription qui permet la synthèse de l'ADN double brin à partir de l'ARN viral génomique, se produit principalement dans le cytoplasme de la cellule même si elle peut être initiée dans les virions libres (Ventura et al., 1999). Elle fait appel aux activités ADN polymérase et RNase H de la RT. D'autres protéines virales et cellulaires seraient impliquées dans la régulation de ce processus, notamment les protéines MA (Kiernan et al., 1998), Nef (Aiken and Trono, 1995), Vif (Goncalves et al., 1996), la NC (Li et al., 1996), l'IN (Wu et al., 1999; Zhu et al., 2004), Tat (Harrich et al., 1997), la cyclophiline A (Thali et al., 1994), Vpr (Stark and Hay, 1998) et la topoisomérase I (Takahashi et al., 1995).

L'ADN double brin rétrotranscrit est ensuite intégré dans le génome de la cellule hôte. Ce processus se fait en plusieurs étapes. Tout d'abord une phase préparatoire de l'ADN viral, appelée 3'-processing, se produit dans le cytoplasme de la cellule sous l'action de l'IN. Puis l'ADN viral est transporté dans le noyau de la cellule, via un complexe de pré-intégration (PIC) où il est alors intégré dans un chromosome. Cette dernière étape appelée transfert de brins est également catalysée par l'IN. (Ces différentes étapes d'intégration sont décrites dans le chapitre II.)

### A-3- Transcription de l'ADN proviral et expression des protéines virales

L'ADN viral, intégré dans le génome de la cellule hôte, n'est pas forcément transcrit immédiatement. Plusieurs facteurs conditionnent son expression tels que :

- l'état d'activation de la cellule hôte (stress, différenciation, activation immunitaire...)
- le site d'intégration de l'ADN viral. Un provirus incorporé dans une zone de la chromatine où la transcription est réprimée, tend à se montrer moins productif.
- la présence de la protéine virale Tat en quantité suffisante. Cette protéine se lie à la séquence TAR de l'ARN viral (Kao et al., 1987) et recrute un complexe (P-TEBb) (Wei et al., 1998) qui phosphoryle alors l'ARN polymérase II, ce qui augmente le taux de transcription de l'ADN proviral (Price, 2000).

Une fois initiée, la transcription peut conduire à plus de 30 ARNm viraux à partir d'un seul transcrit, grâce à un mécanisme d'épissage alternatif. Les sites donneurs et accepteurs d'épissage permettent l'obtention de trois catégories d'ARN :

- l'ARNm non-épissé : codant Gag et Pol et/ou étant encapsidé
- les ARNm mono-épissés : codant Env, Vif, Vpr et Vpu



- les ARN multi-épissés : codant Tat, Rev et Nef

Les ARNm multi-épissés sont exportés du noyau de la cellule hôte. Dans le cytoplasme, ils sont traduits en protéines régulatrices Tat, Rev et Nef grâce à la machinerie cellulaire. Les gènes *tat*, *rev* et *nef* sont les premiers à être exprimés en protéines et constituent donc les gènes précoces du VIH. Les ARNm mono et non-épissés, contrairement aux ARNm multi-épissés, nécessitent l'implication de la protéine Rev pour l'export nucléaire (Strebel, 2003). Les gènes de ces ARNm ne peuvent donc être exprimés qu'en présence d'une quantité suffisante de Rev : ils constituent les gènes tardifs du VIH. Des ARNm non-épissés ne sont pas traduits en protéines ; ils constitueront le matériel génétique de futurs virions.

#### A-4- Formation et maturation de nouvelles particules virales

L'assemblage et le bourgeonnement de nouvelles particules virales ont lieu au niveau de la membrane cellulaire. Les polyprotéines Gag et Gag-Pol dirigent les étapes d'assemblage et permettent de concentrer les précurseurs viraux au site de bourgeonnement (Freed, 1998; Freed et al., 1994; Ono et al., 1997; Ono et al., 2000; Scarlata and Carter, 2003). Ces polyprotéines possèdent la propriété de s'assembler au niveau de régions riches en cholestérol et en sphingolipides, appelées radeaux lipidiques (Nguyen and Hildreth, 2000; Ono and Freed, 2001) qui servent de plates-formes au recrutement et à l'assemblage des protéines virales.

Le bourgeonnement des particules virales dépend d'un processus d'ubiquitination et de la protéine Tsg101. Environ 2-5% des précurseurs Gag, retrouvés au site de bourgeonnement, sont mono-ubiquitinés à leur région p6 (Ott et al., 1998). La protéine Tsg101 interagit avec un motif de Gag (PTAP) (VerPlank et al., 2001) et reconnaît l'ubiquitine. Le rôle de Tsg101 consiste à contrôler la machinerie vésiculaire de la cellule (les endosomes) dans leur bourgeonnement. Il semble donc que la protéine Tsg101 soit responsable de l'organisation du bourgeonnement du VIH par un mécanisme qui implique la mono-ubiquitination.

La maturation débute durant le bourgeonnement et se poursuit une fois la particule virale formée. Cette étape est essentielle à l'infectivité virale. La PR virale joue un rôle clé dans ce processus. Le dimère de précurseurs Gag-Pol, formé durant l'assemblage, s'autoclive pour libérer une PR dimérique mature. La PR fonctionnelle peut alors cliver les précurseurs Gag et Gag-Pol pour libérer les protéines structurales et enzymatiques fonctionnelles. Elle a plusieurs sites de clivage sur un même précurseur Gag et chacun de ces clivages est réalisé dans un ordre bien précis qui est le suivant :  $\text{Gag}^{\text{p2}}/\text{Gag}^{\text{NC}} > \text{Gag}^{\text{NC}}/\text{Gag}^{\text{p6}} > \text{Gag}^{\text{MA}}/\text{Gag}^{\text{CA}} >$





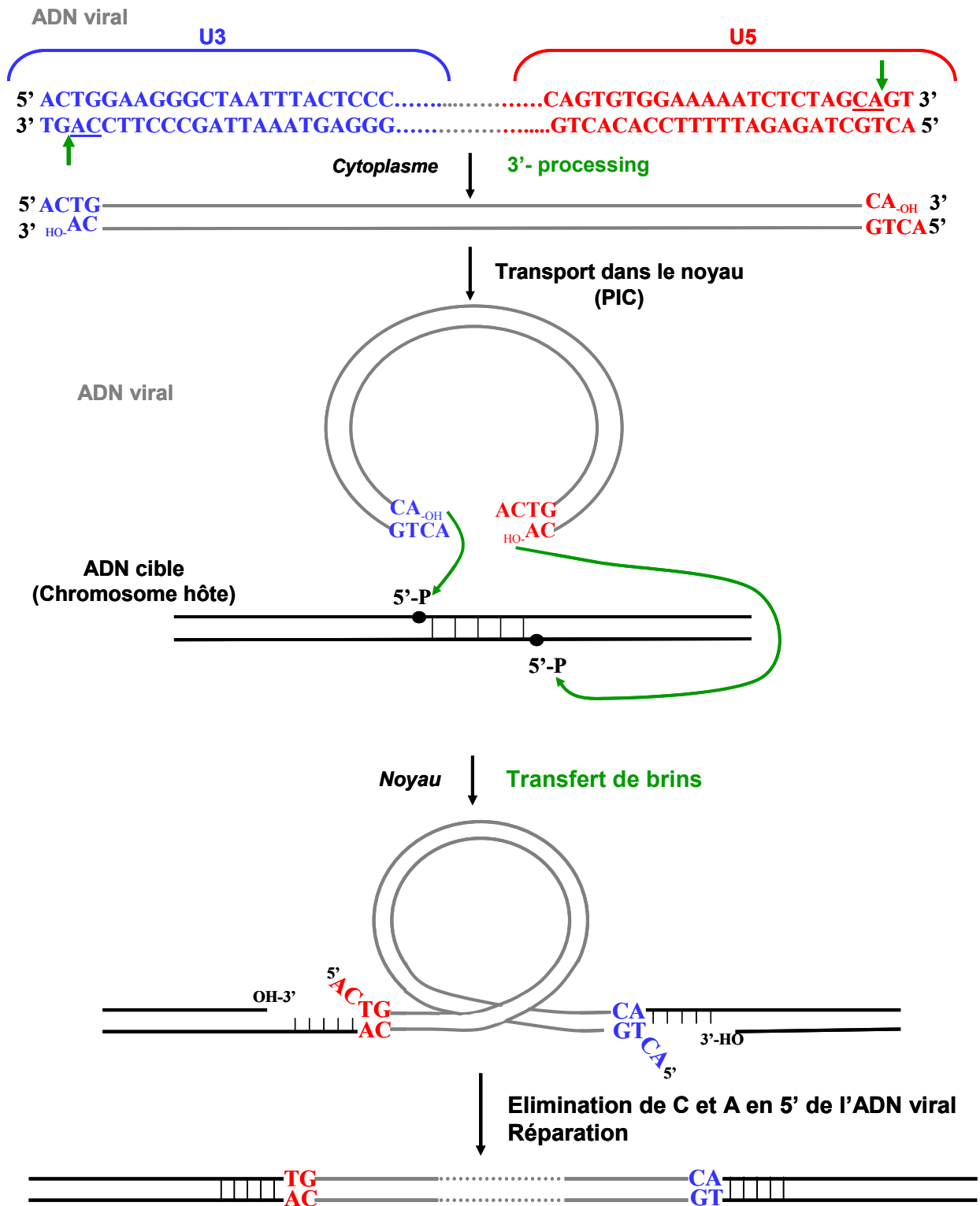
Gag<sup>CA</sup>/Gag<sup>p2</sup>. La formation de la capside en cône et la condensation de l'ARN par la NC du virus sont les dernières étapes atteintes au cours de la maturation par le clivage de Gag<sup>CA</sup>/Gag<sup>p2</sup>. Ce processus de maturation est essentiel à l'infectivité de la particule virale.

### **B- Les thérapies anti-VIH actuelles**

Les médicaments anti-VIH actuels, les antirétroviraux, ciblent trois étapes du cycle répliatif du VIH : la rétrotranscription, la maturation et la fusion. Vingt et un antirétroviraux divisés en quatre classes, sont approuvés par la FDA : neuf inhibiteurs nucléosidiques de la RT, trois inhibiteurs non-nucléosidiques de la RT, huit inhibiteurs de la PR et un inhibiteur de fusion.

Les traitements antirétroviraux hautement actifs (HAART) qui constituent le standard de la thérapie anti-VIH actuelle, utilisent des combinaisons de ces antirétroviraux.

Cependant, quelles que soient les combinaisons médicamenteuses, on observe à plus ou moins brève échéance l'apparition de virus résistants aux antirétroviraux. La résistance est liée à la présence, dans la séquence du virus, de mutations qui modifient la structure de la protéine ciblée. Ces mutations générées en raison de l'absence d'activité de relecture de la RT émergent sous la pression de sélection imposée par la présence de l'inhibiteur. Les antirétroviraux deviennent alors moins actifs car les mutations sélectionnées se traduisent par une perte d'affinité du produit pour sa cible. De plus, de nombreux effets secondaires liés à ces médicaments sont très durs à supporter pour le patient. Aussi de nouvelles molécules et approches thérapeutiques se révèlent nécessaires ; actuellement, elles font l'objet de multiples travaux de recherche. Une de ces approches consiste à cibler la troisième enzyme virale, l'IN pour laquelle aucun inhibiteur n'est disponible dans la pharmacopée actuelle alors qu'elle joue un rôle essentiel dans le cycle répliatif.



**Figure 4-** Représentation schématique de l'activité catalytique de l'IN du VIH-1 in vivo : processus d'intégration de l'ADN viral du VIH-1 dans l'ADN cible (chromosome de la cellule hôte).

## Chapitre II

### Le rôle et la structure de l'intégrase du VIH-1

---

#### **I- Activités catalytiques de l'intégrase**

Bien que l'IN puisse agir comme un cofacteur pour la transcription inverse, le rôle principal de l'IN est de participer à l'insertion de l'ADN viral à l'intérieur du génome des cellules infectées (Tsurutani et al., 2000; Zhu et al., 2004). Cette insertion ou intégration est requise pour la transcription du génome viral, la production de protéines virales et donc pour la réplication virale (Brown et al., 1998). Dans le cycle réplcatif du virus, l'intégration est un processus qui se produit une fois seulement. Pour pleinement jouer son rôle, l'IN n'a pas besoin d'être une enzyme efficace.

#### **A - Le rôle de l'IN *in vivo***

Un dimère d'IN se fixe à chacune des extrémités LTR et réalise l'intégration en deux temps. La première étape, le 3'-processing, se passe dans le cytoplasme. Elle consiste en un clivage endonucléotidique des deux extrémités 3' de l'ADN viral double brin (**figure 4**). A chaque extrémité de l'ADN viral, ce clivage se produit précisément en 3' d'un dinucléotide CA très conservé et entraîne la libération du dinucléotide GT. Cette réaction génère donc un ADN viral terminé à chaque extrémité 3' par un dinucléotide CA 3'-hydroxylé qui est l'intermédiaire réactionnel requis pour l'étape suivante, le transfert de brins.

Après le 3'-processing, l'IN doit passer dans le noyau. Elle reste liée à l'ADN viral sous la forme d'un complexe multimérique qui pontre les deux extrémités de l'ADN, à l'intérieur d'une particule intracellulaire appelée complexe de pré-intégration (PIC). Le PIC formé durant et après le 3'-processing puis transporté dans le noyau de la cellule infectée contient des protéines virales et cellulaires en plus du complexe IN/ADN. Les protéines virales (RT,



MA, NC et Vpr) peuvent contribuer au transport du PIC, à travers l'enveloppe nucléaire. Des protéines cellulaires assemblées à l'intérieur du PIC peuvent lier l'IN et stimuler ses activités enzymatiques. Parmi ces protéines, on trouve : INI1, la première protéine de liaison à l'IN découverte (Yung et al., 2004), LEDGF(p75) (Cherepanov et al., 2005a; Cherepanov et al., 2003; Cherepanov et al., 2005b), la protéine du développement ectodermique embryonnaire (Violot et al., 2003) et HSP60 (Parissi et al., 2001). La protéine PML est également retrouvée à l'intérieur du PIC (Turelli et al., 2001). Deux protéines cellulaires, HMGA1 et BAF, régulent l'intégration en se liant directement à l'ADN. HMGA1 stimule l'activité de l'IN (Farnet and Bushman, 1997; Hindmarsh et al., 1999). BAF stimule l'intégration intermoléculaire et empêche l'auto-intégration. Cette protéine intervient également dans la formation du PIC (Chen and Engelman, 1998; Harris and Engelman, 2000; Lin and Engelman, 2003). Contrairement à d'autres lentivirus tels que les oncorétrovirus Mo-MuLV et RSV qui exigent une détérioration de l'enveloppe nucléaire mitotique pour accéder aux chromosomes des cellules infectées, les PIC du VIH-1 sont capables de traverser l'enveloppe nucléaire. La karyophilie des PIC permet au VIH de se répliquer dans des cellules non prolifératives comme les macrophages (Bukrinsky et al., 1992). Des signaux de type NLS ont été identifiés sur la protéine MA et l'IN du VIH-1 et deux séquences de fixation aux pores nucléaires ont été trouvées au niveau de la protéine Vpr (Vodicka, 2001). Par ailleurs, la présence d'une séquence ADN flap centrale au niveau de l'ADN viral semble importante pour le transport nucléaire du PIC du VIH-1. En effet, un ADN viral sans la séquence ADN flap se retrouve accumulé au voisinage de la membrane nucléaire (Zennou et al., 2000). Le PIC pourrait donc passer par des pores nucléaires appelés nucléosporines, via un transport actif du complexe. Cependant, le mécanisme exact du transport dans le noyau n'est pas défini (Sherman and Greene, 2002).

Une fois dans le noyau, l'IN catalyse l'insertion de l'ADN viral dans un chromosome hôte. Cette réaction, appelée le transfert de brins, consiste à joindre de façon concertée les deux extrémités 3'-OH de l'ADN viral, générées par le 3'-processing, à deux 5'-phosphates situés sur des brins opposés et séparés par 5 paires de bases dans l'ADN du chromosome hôte (ADN cible) (**figure 4**). Pour cela, chacune des extrémités 3'-OH de l'ADN viral réalise une attaque nucléophile de deux liaisons phosphodiester sur chacun des brins de l'ADN cible, en deux sites séparés par 5 paires de bases. A cause de cette concertation, les deux dimères d'IN sont vraisemblablement assemblés en tétramère. Pour que l'intégration soit complète, les deux dinucléotides non appariés CA à chaque extrémité 5' de l'ADN viral sont éliminés et l'espace partant du 3'-OH de l'ADN cible jusqu'au 5'-phosphate de l'ADN viral est réparé



vraisemblablement par les enzymes de réparation de la cellule (Daniel et al., 2004; Yoder and Bushman, 2000). Le génome viral entièrement intégré est alors appelé provirus. Contrairement au 3'-processing qui nécessite une séquence stricte d'ADN viral (**figure 4**), il existe peu de sélectivité du site d'intégration dans l'ADN cible (Carteau et al., 1998; Schroder et al., 2002). Néanmoins, l'intégration se fait préférentiellement à l'intérieur de gènes transcrits dans le cas du VIH-1 (Schroder et al., 2002) alors que pour MLV, elle se fait plutôt dans des régions de départ de transcription (Wu et al., 2003). Cette différence pourrait contribuer à la plus grande oncogénéicité de MLV par rapport à celle du VIH. Il est plausible que des facteurs cellulaires soient impliqués dans la sélection des sites d'intégration, à l'intérieur des gènes transcrits.

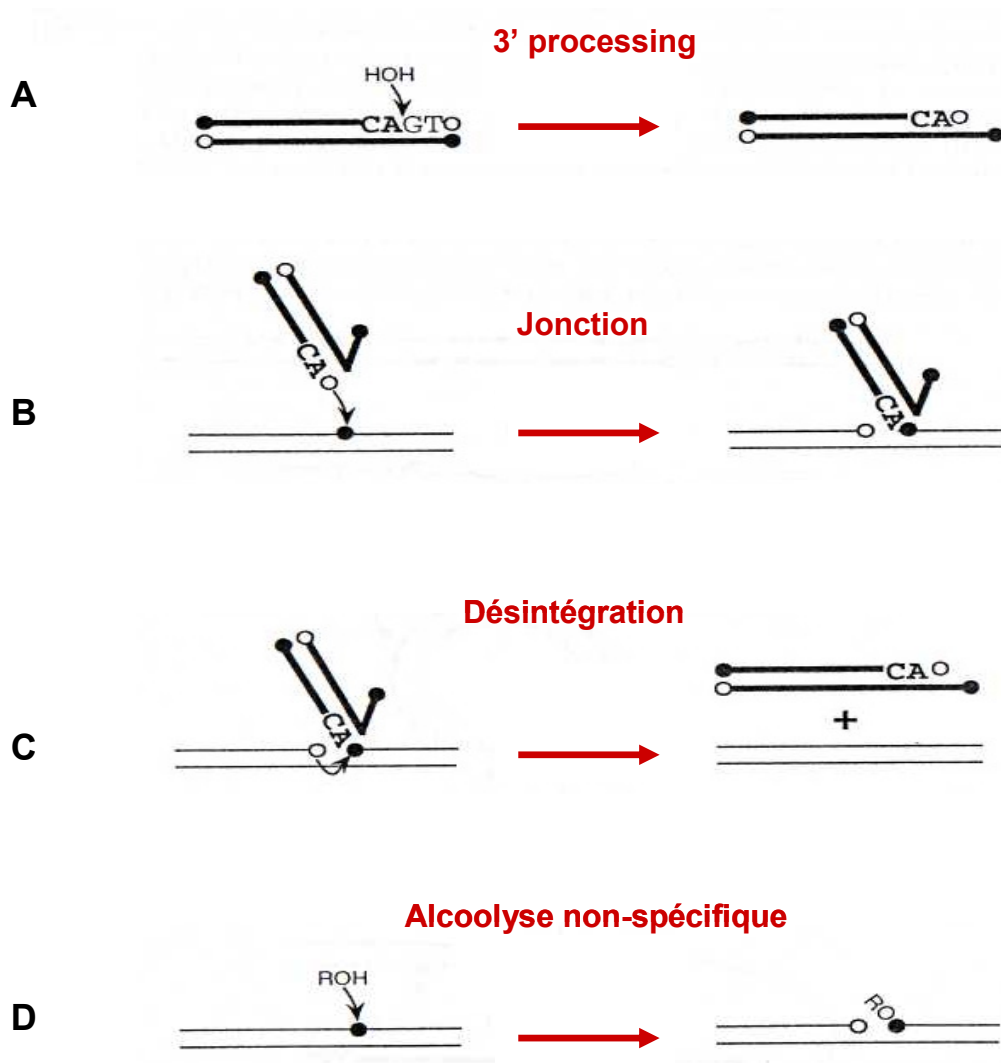
### **B - Les tests *in vitro* de l'activité catalytique de l'IN**

Des tests *in vitro* ont été développés pour caractériser les réactions de 3'-processing et de transfert de brins ainsi que pour évaluer les inhibiteurs potentiels de l'IN. Ces tests utilisent :

- des oligonucléotides doubles brins mimant les extrémités LTR de l'ADN viral et l'ADN cible et appelés respectivement oligonucléotide viral et oligonucléotide cible.
- des IN recombinantes ou des PIC extraits de cellules infectées par le VIH-1 (le plus souvent, l'IN recombinante commence par une méthionine au lieu d'une phénylalanine et peut comporter un tag d'histidines).
- des concentrations de 4nM pour les molécules de duplexes d'ADN et de 200nM pour l'IN. Cela correspond à 50 molécules d'IN pour un ADN double brin dans un volume de  $0.4 \mu\text{m}^3$ . L'ADN, bien que plus concentré qu'*in vivo*, reste en dessous de la constante de dissociation ( $K_d = 30\text{nM}$ ). Sa concentration de 4nM ne doit pas être dépassée pour que les bandes d'électrophorèse restent bien distinctes. Afin que le complexe IN-ADN se forme, il faut augmenter la concentration d'IN. Mais à forte concentration, l'IN s'agrège et l'optimum se situe vers 50 IN par ADN. Ce rapport de 50 coïncide fortuitement avec la valeur *in vivo*.
- des ions métalliques divalents ( $\text{Mg}^{2+}$  ou  $\text{Mn}^{2+}$ ) requis pour l'assemblage de complexes stables IN/ADN-viral (Ellison and Brown, 1994; Vink et al., 1994; Wolfe et al., 1996), le 3'-processing et le transfert de brins (Hazuda et al., 1997).

$\text{Mg}^{2+}$  est un million de fois plus abondant que  $\text{Mn}^{2+}$  dans les cellules ; il est donc considéré comme le cofacteur utilisé *in vivo*. Cependant, *in vitro*,  $\text{Mn}^{2+}$  est fréquemment utilisé car l'activité de l'IN est généralement plus robuste en sa présence. Beaucoup d'inhibiteurs de l'IN





**Figure 5-** Représentation schématique des réactions catalysées par l'IN du VIH-1 in vitro. Les 5'-phosphates et les 3'-OH sont respectivement représentés par des cercles pleins et vides. Les oligonucléotides viraux et cibles sont respectivement schématisés par des traits épais et fins. **A-** 3'-processing à une extrémité de l'oligonucléotide viral. Le groupe OH d'une molécule d'eau attaque le Phosphate de la liaison phosphodiester située en 3' du dinucléotide CA. **B-** Jonction de l'extrémité 3'-OH de l'oligonucléotide viral à un 5'-Phosphate interne d'un oligonucléotide cible. **C-** Désintégration. Le 3'-OH de l'ADN cible attaque le Phosphate de la liaison phosphodiester, liant le CA de l'oligonucléotide viral à l'oligonucléotide cible. **D-** Alcoolyse non spécifique. L'IN utilise divers petits alcools comme des nucléophiles qui coupent et se lient à n'importe quel site interne d'oligonucléotides non viraux.

ont été identifiés en présence de  $Mn^{2+}$  mais certains d'entre eux se sont révélés inactifs en présence de  $Mg^{2+}$ .

Pour réaliser des tests rapides, un oligonucléotide contenant 10 à 20 paires de bases d'une seule extrémité LTR (U3 ou U5) est utilisé pour mimer l'ADN viral. L'IN recombinante purifiée effectue le clivage en 3' du CA de l'oligonucléotide viral, formant ainsi un oligonucléotide plus court de deux nucléotides (**figure 5-A**). Une molécule d'eau activée est généralement le nucléophile de cette réaction. Mais en présence de  $Mn^{2+}$ , des alcools courts (glycérol et 1,2-propanediol), certains acides aminés (L-sérine et L-thréonine) ou l'extrémité 3' de l'oligonucléotide viral peuvent également servir de nucléophiles à cette réaction (Katzman et al., 1991; Vink et al., 1991b). Après ce 3'-processing, l'IN peut joindre l'extrémité 3'-OH de l'oligonucléotide viral à un 5'-Phosphate interne d'un oligonucléotide cible (**figure 5-B**). Il en résulte un produit d'intégration en forme d'Y, représentant l'oligonucléotide viral attaché en un site de l'oligonucléotide cible. Les deux réactions formant des oligonucléotides plus courts et plus longs peuvent être contrôlées par électrophorèse sur gel de polyacrylamide. Ces tests ont l'avantage d'être rapides, commodes et adaptables pour réaliser du criblage à haut débit d'inhibiteurs de l'IN (Marchand et al., 2001). Cependant, ces tests basés sur l'utilisation d'IN recombinantes purifiées identifient un nombre significatif de faux positifs dont les inhibitions de l'IN ne traduisent pas des activités antivirales.

En plus des réactions physiologiques décrites précédemment, l'IN recombinante possède deux autres fonctions *in vitro*. Elle peut catalyser la réaction inverse de celle joignant l'oligonucléotide viral à l'oligonucléotide cible et appelée la désintégration (Chow et al., 1992) (**figure 5-C**). Lorsque le produit d'intégration en forme d'Y est incubé avec l'IN recombinante, le 3'-OH de l'ADN cible attaque la liaison phosphodiester liant le CA de l'oligonucléotide viral à l'oligonucléotide cible ; cela provoque alors la libération des deux oligonucléotides. L'autre réaction catalysée par l'IN recombinante *in vitro* est l'alcoolyse non-spécifique (Katzman and Sudol, 1996b) (**figure 5-D**). Dans cette réaction, l'IN utilise divers alcools courts tels que la glycérine et le glycol d'éthylène en tant que nucléophiles qui clivent et se lient en n'importe quel site interne d'oligonucléotides non viraux. Bien que ces deux réactions ne soient pas physiologiques, elles ont montré leur utilité dans l'étude de la fonction de l'IN et des interactions IN/ADN (Donzella et al., 1993; Hickman et al., 1994; van Gent et al., 1992; Vink et al., 1993). De plus, la réaction de désintégration est également intéressante pour un criblage rapide de drogues puisque les produits apparaissent dans une seule bande et sont facilement quantifiables.



L'IN recombinante présente une forte affinité pour l'oligonucléotide viral et, sous des conditions physiologiques, un complexe IN/oligonucléotide-viral/métal stable se forme. Généralement, l'IN est préincubée avec l'ion métallique et la drogue puis l'oligonucléotide viral est ajouté. Cependant, des tests ont été réalisés en préincubant l'IN avec l'oligonucléotide viral et l'ion métallique avant l'ajout de l'inhibiteur potentiel. Les activités inhibitrices sont généralement plus faibles avec la deuxième procédure (Marchand et al., 2002). Un test a alors été développé pour identifier, par un criblage à haut débit, des inhibiteurs ayant pour cible le complexe IN/oligonucléotide-viral/métal. Cette approche a été utilisée pour cribler de nombreux composés.

La transférabilité limitée des résultats issus des tests utilisant des IN recombinantes purifiées a conduit au développement de tests utilisant des PIC partiellement purifiés à partir de cellules infectées par le VIH-1 (Farnet et al., 1996). L'intégration obtenue *in vitro*, à partir de ces PIC, est considérée plus proche de la réaction *in vivo*. En effet, *in vitro*, les PIC peuvent diriger la jonction de deux extrémités LTR virales avec l'oligonucléotide cible de façon coordonnée, formant un produit ressemblant à l'intermédiaire d'intégration non réparé, obtenu *in vivo*. Comme on pouvait s'y attendre, des inhibiteurs actifs dans les tests avec l'IN recombinante ne le sont pas dans les tests avec les PIC indiquant que dans l'environnement protégé du PIC, l'IN est inaccessible à beaucoup de molécules (Farnet et al., 1998; Hazuda et al., 2000). Cependant, les tests basés sur les PIC sont beaucoup plus difficiles à mettre en place et n'ont pas pu être optimisés pour réaliser du criblage à haut débit.

## **II- Structure de différents fragments de l'intégrase et leurs fonctionnalités**

L'IN du VIH-1 est une protéine de 32 kDa (288 acides aminés) appartenant à la superfamille des transférases polynucléotidiques, comprenant également les transposases MuA et Tn5, la RNase H et l'enzyme RuvC (Rice et al., 1996). Des études de digestion protéolytique, d'alignements de séquences et de complémentation fonctionnelle montrent que l'IN est formée de trois domaines distincts: le domaine N-terminal (résidus 1 à 50), le domaine central (résidus 50 à 212) et le domaine C-terminal (résidus 212 à 288) (**figure 6**) (Engelman and Craigie, 1992; Katzman and Sudol, 1995; Mumm and Grandgenett, 1991; van Gent et al., 1993). Les trois domaines sont nécessaires pour les activités de 3'-processing et de transfert de brins. En présence de  $Mn^{2+}$ , le domaine central de l'IN suffit pour catalyser la réaction de désintégration. En présence de  $Mg^{2+}$ , l'IN restreinte au domaine (50-212) est inactive (Laboulais et al., 2001) : la protéine doit être entière et la désintégration est moins bien catalysée qu'en présence de  $Mn^{2+}$ .







Les structures de chacun des trois domaines isolés ont été déterminées par cristallographie aux rayons X (Bujacz et al., 1996a; Dyda et al., 1994; Goldgur et al., 1998; Greenwald et al., 1999; Maignan et al., 1998) ou par RMN (Cai et al., 1997; Eijkelenboom et al., 1999; Lodi et al., 1995). On connaît également les structures du domaine central lié au domaine N-terminal (Wang et al., 2001) ou au domaine C-terminal (Chen et al., 2000a). Par contre, encore aujourd'hui, il n'existe pas de structure expérimentale de l'IN entière ni du complexe qu'elle forme avec l'ADN. En effet, plusieurs facteurs, en particulier la faible solubilité, la flexibilité et l'agrégation de la protéine, rendent difficile la détermination de la structure de l'IN entière.

### **A- Le domaine N-terminal (résidus 1 à 50)**

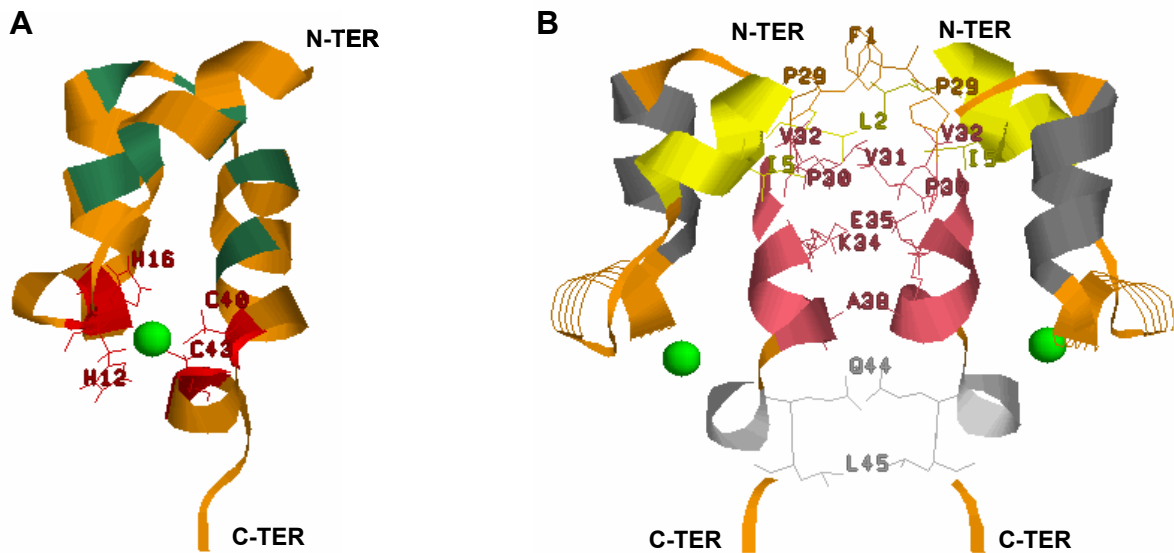
Le domaine N-terminal contient un motif His-Xaa3-His-Xaa20-30-Cys-Xaa2-Cys fortement conservé parmi les IN de tous les rétrovirus et des rétrotransposons eucaryotes. Dans l'IN du VIH-1, le motif (H12, H16, C40 et C43) lie  $Zn^{2+}$  avec une stoechiométrie d'un ion par monomère (Burke et al., 1992; Zheng et al., 1996).

#### **A-1- Fonctionnalités du domaine N-terminal**

Bien que le domaine N-terminal soit nécessaire pour les activités de 3'-processing et de transfert de brins de l'IN du VIH, son rôle exact dans ces réactions n'est pas compris. La délétion de ce domaine dans les IN de RSV (Bushman and Wang, 1994) et du virus Visna (Katzman and Sudol, 1995) n'a pas d'effet sur le 3'-processing. La délétion, dans l'IN de RSV, n'a également aucun effet sur le transfert de brins. Cependant, le virus ne peut pas se répliquer (Bushman and Wang, 1994; Katz et al., 1996). La substitution des histidines H12 et H16 du motif HHCC dans l'IN de RSV n'affaiblit pas significativement le 3'-processing et le transfert de brins (Khan et al., 1991). Par contre, pour l'IN du VIH-1, les deux histidines sont requises pour les deux activités (Engelman and Craigie, 1992). La mutation des deux cystéines dans le motif HHCC de l'IN du VIH-1 affecte également le 3'-processing et le transfert de brins (Engelman and Craigie, 1992). L'IN dépourvue du domaine N-terminal peut catalyser la réaction de désintégration, à l'exception de l'IN de HFV (Pahl and Flugel, 1995). La liaison de  $Zn^{2+}$  au domaine N-terminal joue un rôle important dans la multimérisation de l'IN entière et augmente l'activité catalytique de l'IN du VIH-1 (Deprez et al., 2000; Lee et al., 1997; Leh et al., 2000; van den Ent et al., 1999; Zheng et al., 1996).

D'autre part, le domaine N-terminal est la région de liaison préférentielle pour le facteur de transcription cellulaire du complexe de pré-intégration, INI1. Il augmente également l'affinité





**Figure 7-** Structure RMN du domaine N-terminal de l'IN du VIH-1 (code PDB :1WJA) (Cai et al., 1997). **A-** Un seul monomère en représentation ruban (conformation D). H12, H16, C40 et C43 sont représentés en fil de fer et colorés en rouge. La position de l'ion  $Zn^{2+}$  est montrée par une sphère verte. Les résidus formant le cœur hydrophobe stabilisant la partie supérieure de la structure sont colorés en vert foncé. **B-** Le dimère représenté en modèle ruban (conformation D). Les hélices  $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 3$  et  $\alpha 4$  sont respectivement colorées en jaune, gris foncé, rose et gris clair. Les ions  $Zn^{2+}$  sont représentés par des sphères vertes. Les résidus impliqués dans les interactions entre les deux sous-unités monomériques sont représentés en fil de fer. Les motifs tour-hélice connectant les hélices  $\alpha 1$  et  $\alpha 2$  sont représentés en modèle brins.

de l'interaction entre le facteur de transcription cellulaire du complexe de pré-intégration: LEDGF/p75 et le domaine central de l'IN.

### A-2- Structure du domaine N-terminal

La structure du domaine N-terminal isolé de l'IN du VIH-1 (code PDB : **1WJA**) a été déterminée par RMN (Cai et al., 1997). Ce domaine N-terminal a une structure dimérique (**figure 7-B**) et chaque monomère (**figure 7-A**) est formé de quatre hélices  $\alpha$  et d'un ion  $Zn^{2+}$  coordonné de manière tétraédrique au motif HHCC. Il existe un équilibre entre deux conformations qui diffèrent par le mode de fixation de l'ion  $Zn^{2+}$  à l'histidine H12. L'ion  $Zn^{2+}$  se lie au cycle imidazole de H12 par l'intermédiaire de l'azote  $N\delta_1$  dans la conformation D et par l'intermédiaire de l'azote  $N\epsilon_2$  dans la conformation E. Par contre, c'est toujours l'azote  $N\delta_1$  de H16 qui est lié à l'ion  $Zn^{2+}$ . Les cystéines C40 et C43 sont coordonnées à l'ion  $Zn^{2+}$  par l'intermédiaire du  $S\gamma$ . Les hélices  $\alpha_2$ ,  $\alpha_3$  et  $\alpha_4$  sont les mêmes dans les deux conformations E et D et comprennent respectivement les résidus 19 à 25, 30 à 39 et 41 à 45. Par contre, l'hélice  $\alpha_1$  est deux fois plus courte dans la conformation D que dans la conformation E. En effet, elle comprend seulement les résidus 2 à 8 dans la conformation D alors qu'elle s'étend du résidu 2 à 14 dans la conformation E. La boucle connectant les hélices  $\alpha_1$  et  $\alpha_2$  dans la conformation E est remplacée par quatre résidus formant un motif tour-hélice (résidus 14 à 17) dans la conformation D. Dans la conformation E, H12 est localisée dans l'hélice  $\alpha_1$  et enfouie à l'intérieur de la protéine. Par contre, dans la conformation D, H12 est localisée dans une boucle connectant l'hélice  $\alpha_1$  aux résidus en tour-hélice et est exposée au solvant (**figure 7-B**). Ainsi, dans la conformation E prédominante à basse température (au-dessous de 300 °K), l'hélice  $\alpha_1$  stable et continue est observée pour les résidus 2 à 14. Alors qu'à plus haute température (au-dessus de 300 °K) où la conformation D prédomine, l'hélice  $\alpha_1$  est partiellement fondue et les résidus 9 à 18 adoptent une conformation différente. Ceci n'est sans doute pas fonctionnellement approprié mais peut indiquer que d'autres parties de l'IN aident à stabiliser la structure du domaine N-terminal, dans le contexte de la protéine intacte.

La coordination de l'ion  $Zn^{2+}$  au motif HHCC stabilise la partie inférieure de la structure. La région supérieure de la structure est stabilisée par un cœur hydrophobe (**figure 7-A**) formé par l'ensemble des hélices  $\alpha_1$ ,  $\alpha_2$  et  $\alpha_3$  et comprenant : I5 et A8 de l'hélice  $\alpha_1$ , M22 de l'hélice



$\alpha 2$ , F26 et L28 localisées dans la boucle entre les hélices  $\alpha 2$  et  $\alpha 3$  et V32 et I36 de l'hélice  $\alpha 3$ .

L'interface du dimère est identique dans les conformations E et D et exclut une aire de surface accessible au solvant d'environ 550 Å<sup>2</sup> par sous-unité. Les deux sous-unités du dimère sont presque parallèles l'une par rapport à l'autre et des contacts intermoléculaires surviennent entre l'extrémité de l'hélice  $\alpha 1$  et les hélices  $\alpha 3$  et  $\alpha 4$  (**figure 7-B**). Toutes les interactions inter-sous-unitaires entre les hélices  $\alpha 1$  et  $\alpha 3$  sont hydrophobes. F1 d'une sous-unité interagit avec F1', P29' et V32' de la seconde sous-unité, L2 avec P29', P30' et V31' et I5 avec V31'. L'hélice  $\alpha 3$  d'une sous-unité fait environ un angle de 60° avec l'hélice  $\alpha 3$  de l'autre sous-unité et les interactions entre les deux hélices sont principalement hydrophobes. En effet, V31 et V32 interagissent avec V31' et V32' et A38 avec E35' et A38'. D'autre part, il y a un pont salin potentiel entre E35 et K34'. L'orientation de l'hélice  $\alpha 4$  d'une sous-unité par rapport à l'hélice  $\alpha 4$  de la seconde sous-unité est d'environ 110°. L45 de l'hélice  $\alpha 4$  d'une sous-unité interagit avec L45' de la seconde sous-unité. De plus, il y a une liaison hydrogène potentielle entre les groupes carboxyamides des Q44 des deux sous-unités.

Une caractéristique frappante du monomère N-terminal réside au fait qu'il présente une similarité structurale remarquable avec plusieurs motifs hélice-tour-hélice (HTH) de protéines de liaisons de l'ADN, incluant en particulier le répresseur Trp. Cependant, il n'y a pas d'identité de séquence significative entre le domaine N-terminal et ces protéines de liaison de l'ADN. De plus, aucune de ces protéines ne lie un ion Zn<sup>2+</sup>. Dans chaque cas de similarité structurale, les hélices  $\alpha 2$  et  $\alpha 3$  du domaine N-terminal correspondent respectivement à la première et à la seconde hélice du motif HTH de la protéine de liaison de l'ADN. Alors que la seconde hélice du motif HTH des protéines de liaison de l'ADN est utilisée pour reconnaître l'ADN, l'hélice  $\alpha 3$  du domaine N-terminal forme une partie de l'interface dimérique. Ainsi, cette hélice participe à des interactions protéine-protéine et pourrait ne pas se lier à l'ADN dans la structure dimérique. De plus, la distribution de charges sur cette hélice ne semble pas favorable pour la liaison de l'ADN. La similarité structurale du domaine N-terminal avec les HTH des protéines de liaison de l'ADN peut donc être purement fortuite et ne doit pas être prise comme une preuve que ce domaine de l'IN lie l'ADN.

La structure du domaine N-terminal lié au domaine central (résidus 1-212) de l'IN du VIH-1 triple mutante (W131D, F139D, F185K) a été déterminée par cristallographie aux rayons X (code PDB : **1K6Y**) (Wang et al., 2001). La structure monomérique cristallographique du

**Tableau 1-** Structures cristallographiques du domaine central de l'IN du VIH-1

	Code PDB	Rés. (Å)	Groupe d'espace	Tampon et Cation	1 <sup>ier</sup> et dernier résidus	Boucles de résidus absents	Autres mutations	Référence
Mutation F185K	1ITG	2.30	P3 <sub>1</sub> 21	cacodylate	55,209	141-153		(Dyda <i>et al.</i> , 1994)
	1B9D	1.70	P3 <sub>1</sub> 21	cacodylate	57,210	141-148 190-192		(Greenwald <i>et al.</i> , 1999)
	1B92	2.02	P3 <sub>1</sub> 21	cacodylate	56,210	141-148	G149A	(Greenwald <i>et al.</i> , 1999)
	1B9F	1.70	P3 <sub>1</sub> 21	cacodylate	56,210	189-192	G140A G149A	(Greenwald <i>et al.</i> , 1999)
	1BIZ	1.95	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	cacodylate	A:54,208 B:58,212	A:141-148 B:141-152	C56S	(Goldgur <i>et al.</i> , 1998)
	1BIS	1.95	P1	Hepes	A:56,209 B:56,209	A:143-150 B: présente	W131E	(Goldgur <i>et al.</i> , 1998)
	1BIU	2.50	C2	Hepes, Mg <sup>2+</sup>	A:56,209 B:56,209	A:141-147 B-C:141-148	W131E	(Goldgur <i>et al.</i> , 1998)
	1QS4*	2.10	C2	Hepes, Mg <sup>2+</sup>	A,B,C:56,209	A:141-144 B:141-142 C:141-143	W131E	(Goldgur <i>et al.</i> , 1999)
	1EXQ	1.60	P3 <sub>2</sub>	Citrate, Cd <sup>2+</sup>	A,B:56,209	A:144-150 B:142-150	C56S W131D F139D	(Chen <i>et al.</i> , 2000a)
	2B4J**	2.02	C2	Hepes	A :56-208 B :57-208	A:140-149 187-194 B:188-194		(Cherepanov <i>et al.</i> , 2005b)
Mutation F185H	2ITG	2.60	P3 <sub>1</sub> 21	cacodylate	51,210	présente		(Bujacz <i>et al.</i> , 1996a)
	1BHL	2.20	P3 <sub>1</sub> 21	cacodylate	57,207	138-153		(Maignan <i>et al.</i> , 1998)
	1BI4	2.46	C2	Hepes	A,B:57,208 C:50,209	A: 141-150 B: 140-149 C: présente		(Maignan <i>et al.</i> , 1998)
	1BL3	2.00	C2	Hepes, Mg <sup>2+</sup>	A:57,208 B:57,209 C:50,209	A: 141-150 B: 140-149 C: présente		(Maignan <i>et al.</i> , 1998)

Le pH du tampon cacodylate est de 6.5, celui du tampon hepes est de 7.0 (1BIZ à 1QS4) ou de 7.2 (2B4J) ou de 7.5 (1BI4 et 1BL3) et celui du tampon citrate est de 5.6. \*Structure du domaine central lié à un inhibiteur de l'IN : le 5CITEP \*\*Structure du domaine central lié à un facteur de transcription cellulaire du complexe de pré-intégration: LEDGF (Lens Epithelium-Derived Growth Factor, également connu comme p75).

domaine N-terminal est très similaire à la conformation E du domaine isolé, obtenue par RMN (**1WJA**) (Cai et al., 1997). L'ion  $Zn^{2+}$  se lie au cycle imidazole de H12 par l'intermédiaire de l'azote  $N_{\epsilon_2}$ . L'hélice  $\alpha_1$  est continue et stable sur 14 résidus et reliée à l'hélice  $\alpha_2$  par une boucle. Par contre, l'interface entre les deux sous-unités N-terminales, dans les structures RMN et cristallographique, est différente. Dans le dimère en solution, ce sont les interactions entre les hélices  $\alpha_3$  qui prédominent alors que dans la structure cristallographique, l'interface est beaucoup plus petite et comprend uniquement des contacts entre l'extrémité de l'hélice  $\alpha_1$  et l'hélice  $\alpha_3$ . Cependant, une partie de l'interface impliquant les résidus F1, I5, P29, V31, V32 et E35 est commune aux deux structures.

## **B- Le domaine central (résidus 50 à 212)**

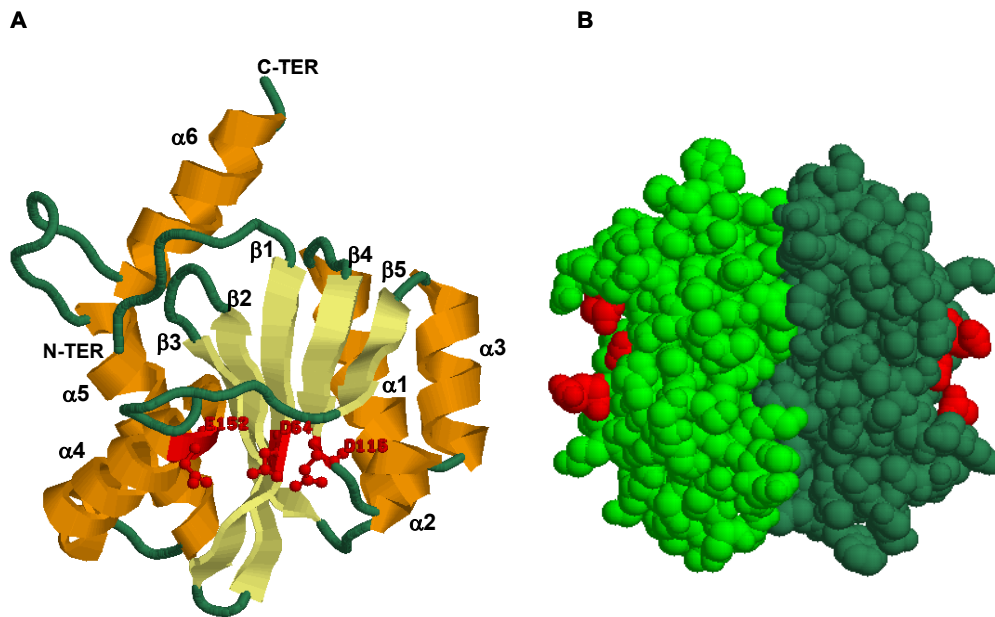
### **B-1- Fonctionnalités du domaine central**

Le domaine central, appelé également domaine catalytique, contient trois résidus acides essentiels pour l'activité catalytique qui forment le motif DDE très conservé dans toutes les IN et les rétrotransposases. La substitution de n'importe lequel des trois résidus de la triade catalytique (D64, D116, E152), dans l'IN du VIH-1, supprime les trois réactions normalement catalysées par l'IN (Drelich et al., 1992; Engelman and Craigie, 1992; Kulkosky et al., 1992; van Gent et al., 1993). Le domaine central isolé ne peut réaliser ni le 3'-processing ni le transfert de brins ; les deux domaines terminaux sont nécessaires pour ces réactions (Bushman et al., 1993; Drelich et al., 1992; Engelman and Craigie, 1992; Jonsson et al., 1996; Kulkosky et al., 1995; Schauer and Billich, 1992; Vink et al., 1993). Ce domaine est impliqué dans les interactions avec les substrats ADN. Il présente, en particulier, des contacts avec les extrémités LTR de l'ADN viral (Esposito and Craigie, 1998; Gerton and Brown, 1997; Heuer and Brown, 1997; Jenkins et al., 1997) (voir chapitre III).

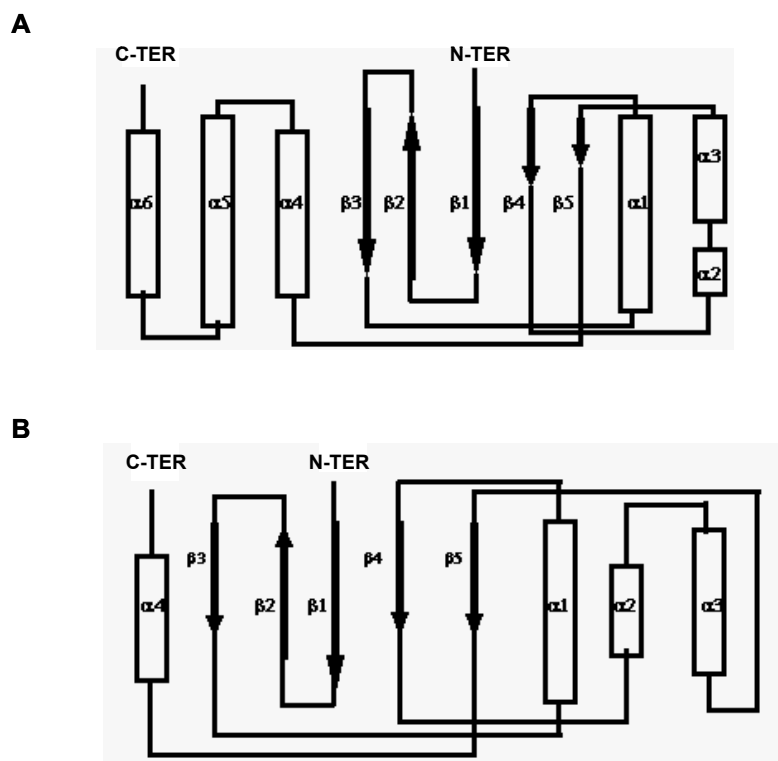
On peut également noter que le domaine central interagit avec un facteur de transcription cellulaire du complexe de pré-intégration: LEDGF(p75). Cette interaction est essentielle pour l'association de l'IN avec la chromatine et la stabilité de la protéine virale dans les cellules humaines (Llano et al., 2004; Maertens et al., 2003).

### **B-2- Structure du domaine central (tableau 1)**

La faible solubilité du domaine central ne permet pas de le cristalliser. Afin de préparer un domaine central plus soluble, les 29 résidus hydrophobes ont été substitués par A ou K. Le



**Figure 8-** Structure du domaine central de l'IN du VIH-1 déterminée par cristallographie aux rayons X (code PDB: **1BL3**) (Maignan et al., 1998). **A-** Un seul monomère (chaîne C, résidus 50 à 209) avec les feuillettes  $\beta$  colorés en jaune et les hélices  $\alpha$  en orange. Les résidus de la triade catalytique (D64, D116 et E152) colorés en rouge sont représentés en boules-bâtons. **B-** Le dimère (chaînes A et B) représenté en modèle compact avec un monomère coloré en vert clair et l'autre en vert foncé. Les résidus catalytiques sont colorés en rouge dans chaque monomère.



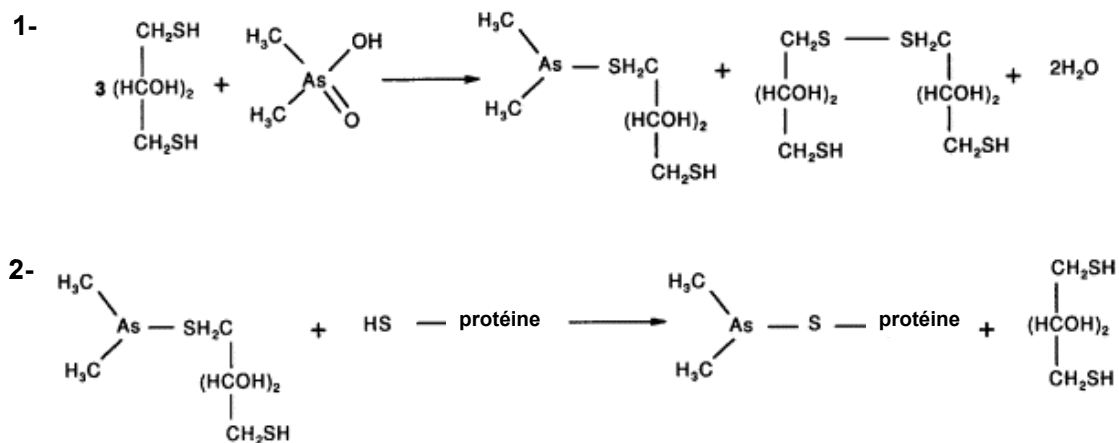
**Figure 9-** Topologie de la structure  $\alpha/\beta$  du domaine central de l'IN du VIH-1 (**A**) et du domaine RNase H de la transcriptase inverse du VIH-1 (**B**).

mutant le plus soluble, F185K, catalyse autant la réaction de désintégration que le domaine sauvage isolé (Jenkins et al., 1995). La protéine entière avec la mutation F185K peut effectuer, *in vitro*, les trois actions catalytiques de l'enzyme sauvage. Cependant, cette mutation est délétère à l'assemblage du virion, *in vivo* (Engelman et al., 1997; Jenkins et al., 1996). Par la suite, un autre mutant soluble a été mis en évidence. Il s'agit du F185H qui, à la pleine activité catalytique et contrairement au mutant F185K, ne perturbe pas l'assemblage du virion *in vivo* (Engelman et al., 1997). Actuellement, il existe plusieurs structures cristallographiques du domaine central de l'IN du VIH-1 qui ont été obtenues dans différentes conditions de cristallisation et en utilisant les mutants solubles F185H et F185K (**tableau 1**). A l'exception de quatre structures avec la mutation F185H, dans toutes les autres, la mutation F185K a été maintenue.

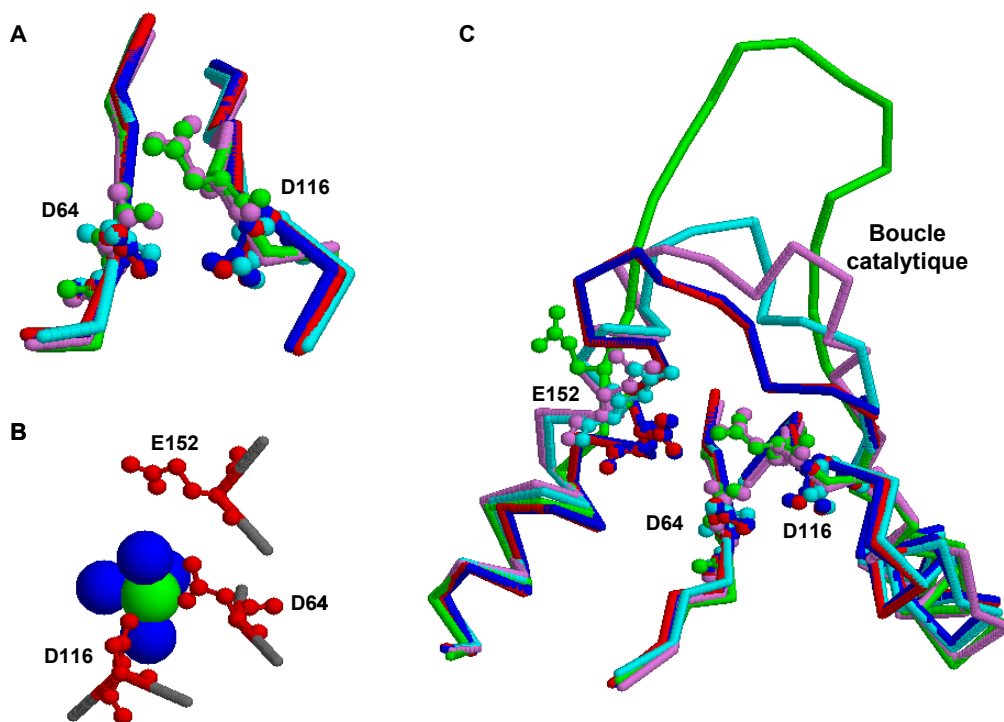
Sur le plan structural, le domaine central de l'IN du VIH-1 est remarquablement similaire à d'autres IN rétrovirales, aux transposases MuA et Tn5, à la RNase H et à la résolvasse RuvC (Bujacz et al., 1995; Chen et al., 2000b; Davies et al., 2000; Katayanagi et al., 1993; Rice and Mizuuchi, 1995; Yang et al., 2000). La similarité structurale avec la transposase MuA est particulièrement frappante, en considérant l'absence complète de similarité dans leur séquence d'acides aminés. Des études biochimiques antérieures ont démontré le lien entre les mécanismes moléculaires de la transposition de l'ADN et de l'intégration de l'ADN rétrovirale (Mizuuchi, 1992; Mizuuchi, 1997) : une similarité qui est reflétée dans la structure de leurs domaines catalytiques.

Dans la topologie d'ensemble, toutes les structures résolues du domaine central de l'IN du VIH-1 (**tableau 1**) sont semblables à la structure de la **figure 8-A**. Le domaine central a une structure  $\alpha/\beta$  qui consiste en un repliement de type RNase H, formé de 5 brins  $\beta$  et de six hélices  $\alpha$ . La **figure 9** montre la topologie de ce repliement dans le domaine central de l'IN du VIH-1 et dans le domaine RNaseH de la rétrotranscriptase du VIH-1. Les hélices  $\alpha 5$  et  $\alpha 6$ , observées dans le domaine central de l'IN du VIH-1, sont présentes seulement chez un autre membre de la superfamille des transférases polynucléotidiques : la transposase MuA. Les trois résidus catalytiques, D64, D116 et E152 définissent la position du site actif. Ce site est plutôt peu profond et localisé à la surface de la protéine. En considérant la grande taille des substrats de l'IN, cette caractéristique peut leur permettre d'accéder plus facilement aux résidus de la triade, facilitant ainsi les réactions catalytiques. Le domaine central peut former un dimère dans toutes les structures résolues. Ce dimère a une forme sphérique et chaque monomère représente un hémisphère (**figure 8-B**). Il possède également une large interface





**Figure 10-** Les deux étapes présumées de la réaction entre le cacodylate (le tampon utilisé pour la cristallisation) et les cystéines du domaine central de l'IN du VIH-1 (Maignan *et al.*, 1998). Pour la cristallisation, la solution protéique, contenant du DDT (dithiothreitol) est mélangée avec le tampon cacodylate (forme basique de l'acide cacodylique ou acide diméthylarsinique  $(\text{CH}_3)_3\text{AsO}(\text{OH})$ ). 1- Le cacodylate réagit avec le DDT formant spontanément un intermédiaire très réactif :  $(\text{CH}_3)_2\text{AsSR}$ . 2- Cet intermédiaire réagit à son tour avec les cystéines de la protéine par un processus d'échange de groupe thiol. Ce qui entraîne la fixation de  $(\text{CH}_3)_2\text{As}$ - sur les cystéines de la protéine.



**Figure 11-** A- Orientations des chaînes latérales des résidus catalytiques D64 et D116 du domaine central de l'IN du VIH-1 dans des structures cristallisées avec un tampon cacodylate (codes PDB: **2ITG** et **1B9F**) ou Hepes (codes PDB: **1BIS**, **1BI4** et **1BL3**) (Bujacz *et al.*, 1996a; Goldgur *et al.*, 1998; Maignan *et al.*, 1998; Greenwald *et al.*, 1999). Les structures, **2ITG**, **1B9F**, **1BIS**, **1BI4** et **1BL3** ont été superposées et colorées respectivement en vert, violet, turquoise, rouge et bleu. Les résidus D64 et D116 sont représentés en boules-bâtons. B- Liaison d'un ion  $\text{Mg}^{2+}$  au site actif du domaine central de l'IN du VIH-1 (code PDB : **1BL3**, monomère C) (Maignan *et al.*, 1998). Les trois résidus catalytiques sont colorés en rouge et représentés en boules-bâtons. L'ion  $\text{Mg}^{2+}$  représenté par une sphère verte est coordonné au site actif par l'intermédiaire de D64 et de D116. Quatre molécules d'eau représentées par des sphères bleues interagissent avec l'ion  $\text{Mg}^{2+}$ , complétant sa coordination octaédrique. C- Les conformations de la boucle catalytique dans plusieurs structures cristallisées du domaine central de l'IN du VIH-1, **2ITG**, **1B9F**, **1BIS**, **1BI4** et **1BL3**, colorées respectivement en vert, violet, turquoise, rouge et bleu. Les trois résidus catalytiques sont représentés en boules-bâtons.

d'exclusion du solvant, d'environ 1400 Å<sup>2</sup> par sous-unité. Dans le dimère, les deux sites actifs sont placés dans des sites opposés et sont séparés par environ 35 Å. Cet espacement ne permet pas d'expliquer l'intégration concertée de l'ADN viral à 5 paires de bases de distance dans l'ADN cible. En effet, pour un ADN de forme B canonique, une distance de 5 paires de bases correspond à environ 17 Å. L'ADN cible devrait alors être extrêmement étendu au-delà de la forme canonique pour que l'intégration puisse se faire de manière concertée en deux sites de l'ADN cible, séparés de 35 Å. En supposant que l'interface des dimères cristallins soit maintenue dans le multimère fonctionnel de l'IN, il apparaît évident qu'au moins un tétramère de l'enzyme soit nécessaire pour accomplir la réaction d'intégration concertée.

Bien que la topologie d'ensemble de toutes les structures cristallisées soit semblable, des différences apparaissent dans le détail. Ces différences se situent en particulier au niveau du site actif et d'une boucle située près de ce site, appelée boucle catalytique et comprenant les résidus 140 à 150.

Les structures cristallisées, en utilisant le cacodylate comme tampon, doivent être interprétées avec prudence, en raison de la réactivité du cacodylate avec les cystéines, C65 et C130. La **figure 10** montre les deux étapes présumées de la réaction entre le cacodylate et les résidus de cystéines (Maignan et al., 1998). Bien qu'il ait été démontré que le domaine central en présence de cacodylate et d'agent réducteur conserve toujours l'activité de désintégration (Molteni et al., 2001), la modification des cystéines par le cacodylate affecte la conformation de la protéine. En effet, la liaison du cacodylate entraîne une série de mouvements structuraux dans la région entourant le brin β5. Ceci se manifeste notamment par le déplacement des chaînes latérales des résidus catalytiques D64 et D116, déformant ainsi la géométrie du site actif (Greenwald et al., 1999). La cristallisation en l'absence de cacodylate a produit de nouvelles formes cristallines avec un site actif pouvant lier un ion Mg<sup>2+</sup> (Goldgur et al., 1999; Goldgur et al., 1998; Maignan et al., 1998). La **figure 11-A** montre la différence d'orientation des chaînes latérales des résidus catalytiques D64 et D116 entre des structures cristallisées, en présence de cacodylate et d'Hepes. Dans les structures cristallisées, en présence d'Hepes (codes PDB : **1BIU**, **1BL3**, **1QS4**) (Goldgur et al., 1998; Maignan et al., 1998), l'ion Mg<sup>2+</sup> requis pour l'activité catalytique peut se lier au site actif, entre les groupes carboxylates de D64 et D116, sans modifier sa structure. Deux (Goldgur et al., 1998) ou quatre (Maignan et al., 1998) molécules d'eau interagissent avec l'ion métallique, complétant ainsi sa coordination octaédrique (**figure 11-B**). Le troisième résidu catalytique E152 ne participe pas à la liaison de l'ion métallique.



Dans la première structure cristallisée du domaine central (code PDB: **1ITG**) (Dyda et al., 1994), le résidu catalytique E152 est absent, en raison du manque de densité électronique interprétable entre les résidus 140 et 154. Dans la seconde structure cristallisée (code PDB : **2ITG**) (Bujacz et al., 1996a), la région comprise entre les résidus 140 et 154 a pu être tracée et ainsi le résidu E152 est présent. Cependant, cette région correspond à une boucle très dépliée, loin du site actif (**figure 11-C**). De plus, le résidu E152 présent dans cette boucle tourne complètement le dos aux deux autres résidus de la triade catalytique. La conformation de la boucle est stabilisée par des contacts cristallins avec une molécule ne faisant pas partie du dimère. De plus, des facteurs de température élevés et une pauvre densité électronique indiquent que cette région est extrêmement flexible. Par la suite, dans la plupart des structures cristallisées, le résidu E152 appartient à l'hélice  $\alpha_4$  et sa chaîne latérale est davantage orientée vers le site actif. La boucle catalytique a pu être tracée dans quatre autres structures (codes PDB : **1B9F** ; **1BIS**, monomère B ; **1BI4**, monomère C ; et **1BL3**, monomère C) (Goldgur et al., 1998; Greenwald et al., 1999; Maignan et al., 1998). Dans ces structures, la boucle est moins dépliée que dans la structure **2ITG** (Bujacz et al., 1996a) mais adopte différentes conformations (**figure 11-C**). Dans les structures **1B9F** et **1BIS**, la boucle catalytique est au-dessus du site actif mais reste tout de même éloignée des résidus catalytiques. Par contre, dans les structures de Maignan et al. (**1BI4** et **1BL3**), en présence ou en absence de l'ion  $Mg^{2+}$ , la boucle se trouve juste au sommet du site actif, plus proche des résidus catalytiques. D'autre part, dans la structure **1BL3**, les résidus 145 à 148 forment un coude  $\beta$  de type I. La boucle catalytique est donc flexible et adopte une conformation appropriée, seulement en présence du substrat ADN. Des études de pontage et de mutagenèse montrent l'importance de cette région pour la liaison de l'ADN viral (Esposito and Craigie, 1998; Gerton et al., 1998; Heuer and Brown, 1997). Il a été également démontré que pour beaucoup d'enzymes, la mobilité d'une boucle protéique est essentielle à l'activité biologique (Clubb et al., 1996; Davies et al., 2000; Newman et al., 1995; Rice and Steitz, 1994; Weston et al., 1992). Pour l'IN du VIH-1, lorsque la boucle catalytique est rendue plus rigide grâce aux mutations G140A et G149A, dans la structure **1B9F**, l'activité catalytique est réduite sans détérioration significative de la liaison de l'ADN (Greenwald et al., 1999). Ceci indique que le mutant à boucle catalytique contrainte affecte principalement une étape située après la liaison de l'ADN. Il semblerait que la boucle, après cette liaison, soit impliquée dans des mouvements probablement importants dans le cycle catalytique (Greenwald et al., 1999).



Les structures du domaine central lié au domaine N-terminal (code PDB : **1K6Y**) (Wang et al., 2001) ou au domaine C-terminal (code PDB : **1EX4**) (Chen et al., 2000a) ont été déterminées par cristallographie aux rayons X. Au sein de ces bidomains, la structure du domaine central est similaire aux structures du domaine isolé mentionnées précédemment et l'interface dimérique reste la même.

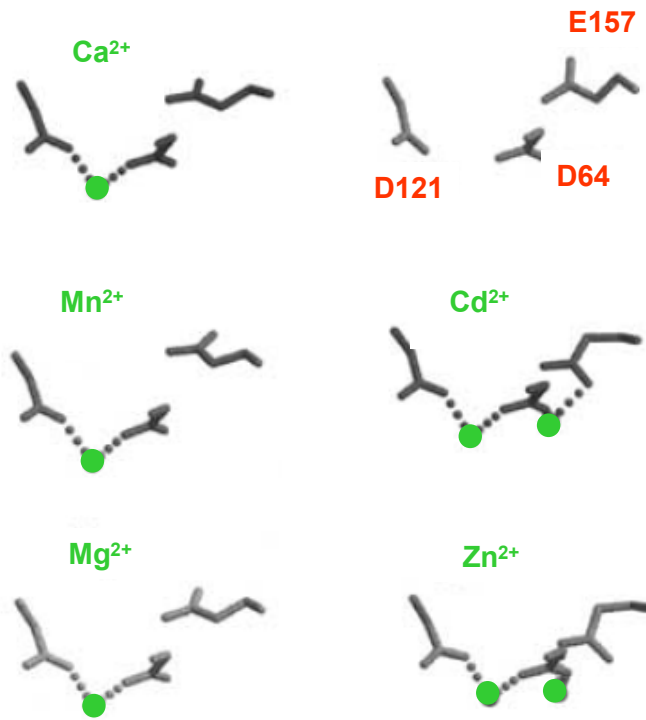
### B-3-Comparaison avec des domaines centraux d'autres IN

Les structures des domaines centraux de plusieurs autres IN rétrovirales ont été déterminées par cristallographie aux rayons X. Ces domaines sont structurellement similaires au domaine central de l'IN du VIH-1.

Le domaine central de l'IN d'ASV peut effectuer les réactions de processing et de désintégration. Cependant, la réaction de processing effectuée par le domaine isolé produit un trinucéotide à la place d'un dinucléotide (Kulkosky et al., 1995). Ce domaine central a été largement étudié (Bujacz et al., 1997; Bujacz et al., 1995; Bujacz et al., 1996b; Lubkowski et al., 1999; Lubkowski et al., 1998b). Sa structure est globalement très semblable à celle du domaine central de l'IN du VIH-1, excepté dans les régions désordonnées pour les deux structures. Le mode de dimérisation est semblable bien que non identique. L'aire de surface exclue au solvant, au niveau de l'interface du dimère, est de  $766 \text{ \AA}^2$  par monomère dans le cas d'ASV contre  $1395 \text{ \AA}^2$  dans le cas du VIH-1, ce qui est compatible avec une constante d'association inférieure (Bujacz et al., 1995). La comparaison des résidus du site actif montre une différence dans l'orientation de la chaîne latérale de D64. Lorsque la boucle catalytique est rendue plus rigide par une diminution du pH, l'activité catalytique est plus faible (Lubkowski et al., 1999), comme observé pour l'IN du VIH-1 (Greenwald et al., 1999).

La structure globale du domaine central de l'IN de SIV (Chen et al., 2000b) ressemble également à celle des domaines centraux des IN du VIH-1 et d'ASV. Dans les quatre structures de l'unité asymétrique, la chaîne latérale du résidu E152 adopte plusieurs conformations : elle peut tourner le dos ou au contraire être orientée vers les deux autres résidus catalytiques D64 et D116, selon la conformation de la boucle catalytique. Cette boucle est mal définie mais le résidu E152 peut être clairement vu dans les quatre structures. La position de la chaîne latérale de D64 varie également.

La structure cristallographique d'un bidomaine de l'IN de RSV (résidus 49-286) a également été déterminée (Yang et al., 2000). Le domaine central et l'interface dimérique sont semblables à ceux des IN du VIH-1, d'ASV et de SIV, mentionnés précédemment.



**Figure 12-** Coordination de bications par les résidus du site actif de domaines centraux de l'IN d'ASV (Bujacz et al., 1996b;Bujacz et al., 1997;Lubkowski et al., 1998a). Les bications sont représentés par des sphères vertes. La chaîne latérale d'E157 a la même conformation dans les structures sans bication ou liant un seul bication. Par contre, sa conformation est différente dans les structures liant deux bications.

Pour conclure, ces quatre domaines centraux rétroviraux sont structurellement très semblables; cela implique que ces IN partagent certainement le même mécanisme d'action. Les différences d'orientation des chaînes latérales des résidus du site actif et de conformation de la boucle catalytique sont dues en partie à l'effet cristallin mais aussi très probablement à l'absence du substrat ADN, dans les structures cristallographiques.

#### B-4- Liaison d'ions métalliques divalents au site actif de l'IN

Les structures cristallines de portions de l'IN du VIH-1 ne révèlent qu'un seul site de liaison pour un ion  $Mg^{2+}$ . Mais certains arguments suggèrent que le site actif de l'IN pourrait lier deux bications.

Des ions divalents ont été observés dans plusieurs structures cristallographiques des domaines centraux des IN du VIH-1 et d'ASV (Bujacz et al., 1997; Bujacz et al., 1996b; Chen et al., 2000a; Goldgur et al., 1999; Goldgur et al., 1998; Maignan et al., 1998).  $Mg^{2+}$  se lie en un seul site de l'IN du VIH-1, entre les carboxylates de D64 et D116.  $Mg^{2+}$ ,  $Mn^{2+}$  (qui permettent le 3'-processing et le transfert de brins) et  $Ca^{2+}$  (qui ne permet ni le 3'-processing, ni le transfert de brins) se lient de la même façon, en un seul site de l'IN d'ASV, situé entre D64 et D121 (équivalent du D116 de l'IN du VIH-1) (**figure 12**). Cependant,  $Zn^{2+}$  (qui permet seulement le 3'-processing) et  $Cd^{2+}$  (qui ne permet ni le 3'-processing, ni le transfert de brins) se lient en deux sites de l'IN d'ASV. Le second site implique les résidus catalytiques D64 et E157 (correspondant aux résidus D64 et E152 de l'IN du VIH-1) (**figure 12**) (Bujacz et al., 1997). La chaîne latérale d'E157 du site actif de l'IN d'ASV a la même conformation dans les structures sans ion divalent ou liant un seul ion. Par contre, sa conformation est différente dans les structures liant deux ions divalents; elle est alors orientée vers D64.

Dans le cas d'enzymes structurellement homologues à l'IN du VIH-1 telle que la RNase H, il a été montré qu'un seul  $Mg^{2+}$  se lie à la RNase H d'Escherichia Coli, malgré la concentration élevée de  $Mg^{2+}$  utilisée lors de la cocrystallisation (Katayanagi et al., 1993).  $Mg^{2+}$  se lie en un site similaire à celui de l'IN d'ASV. Par contre,  $Mn^{2+}$  se lie en deux sites de l'enzyme (Goedken and Marqusee, 2001). La cristallisation du domaine RNase H de la rétrotranscriptase du VIH-1, même avec une très faible concentration de  $Mn^{2+}$  (1 mM), aboutit également à deux sites de liaisons occupés (Davies et al., 1991). Ces sites sont similaires à ceux observés pour la RNase H d'Escherichia Coli.



**Tableau 2-** Distances (Å) Métal-Atome donneur de la protéine (Harding, 2001; Harding, 2002) (d'après la PDB août 2003)

	<b>O</b> eau	<b>O</b> Asp ou Glu monodentate	<b>O</b> Asp ou Glu bidentate	<b>O</b> Ser ou Thr	<b>O</b> CO peptidique	<b>N</b> His	<b>S</b> Cys	<b>Coordination</b> La plus fréquente    Autres valeurs	
<b>Ca</b>	2.39 <i>201</i>	2.34 <i>45</i>	2.47 <i>47</i>	2.43 <i>2</i>	2.33 <i>59</i>	(2.38) <i>1</i>	(2.56)	6	4,5,7,8
<b>Mg</b>	2.08 <i>211</i>	2.06 <i>31</i>	(2.16) <i>8</i>	2.11	(2.05) <i>5</i>	(2.05) <i>2</i>	(2.23)	6	3,4,5
<b>Mn</b>	2.21 <i>15</i>	2.15 <i>10</i>	(2.15) <i>11</i>	2.25	(2.14) <i>2</i>	2.17	2.35	5,6	4,7
<b>Fe</b>	2.09 <i>5</i>	2.01 <i>5</i>	(2.11)	2.13 <i>4</i>	(2.00) <i>5</i>	2.04 <i>5</i>	2.28 <i>157</i>	5,6	4
<b>Cu</b>	1.97 <i>2</i>	1.96	(2.06)	2.00	(1.95)	1.99	(2.17) <i>4</i>	3,4	5,6
<b>Zn</b>	2.09 <i>24</i>	2.00 <i>14</i>	2.16 <i>16</i>	2.14	(1.99) <i>4</i>	2.03 <i>4</i>	2.33 <i>48</i>	4	3,5,6
<b>Na</b>	2.42 <i>103</i>	2.40 <i>2</i>	(2.50) <i>4</i>	(2.45) <i>3</i>	2.40 <i>33</i>	(2.38) <i>33</i>	(2.56)	5,6	4,7
	2.84 <i>19</i>	2.82 <i>1</i>	(2.92) <i>2</i>	(2.87) <i>5</i>	2.78 <i>16</i>	(2.80) <i>16</i>	(2.98)	5,6	4,7,8

Les valeurs entre parenthèses sont déduites d'échantillons de trop peu d'individus pour être fiables. Erreurs attendues d'environ 0.1 Å pour Ca, Mg, Mn, Fe, Cu et Zn, de 0.3 Å pour Na et K. En italique, le nombre de structures cristallographiques de résolution  $d_{\min} < 1.6$  Å.

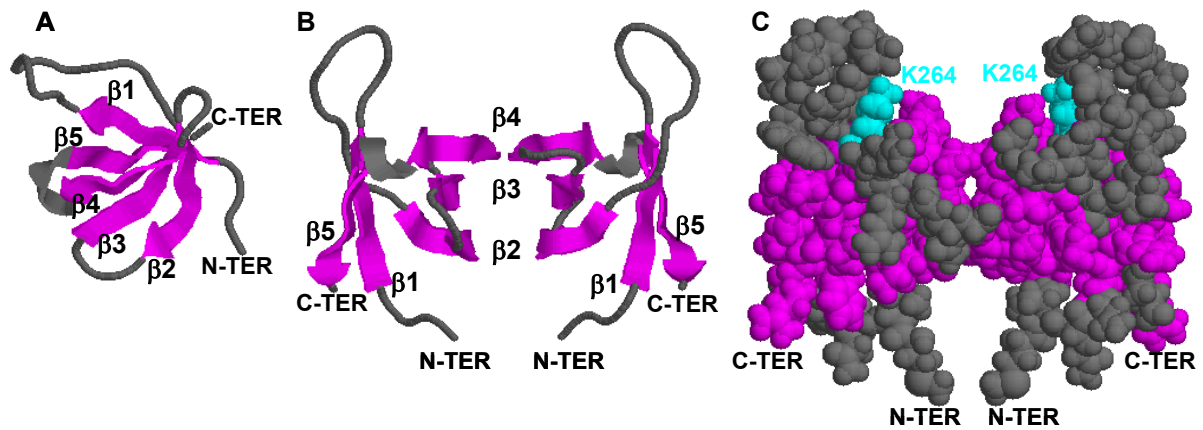
Lors d'une simulation de dynamique moléculaire de 2 ns du domaine central de l'IN du VIH-1, en présence de deux ions  $Mg^{2+}$ , deux modes vibrationnels ont été identifiés pour la boucle catalytique (Lins et al., 2000). Cette boucle semble être stabilisée lorsque le résidu E152 est complètement tourné vers le site actif. Ce comportement conformationnel est causé par l'introduction du second ion métallique divalent. Le site de liaison du second ion métallique n'a pu être localisé précisément. Ceci est en accord avec l'hypothèse qu'un second ion métallique pourrait être amené dans le site actif de l'IN par la liaison du substrat ADN.

Des inhibiteurs de l'IN peuvent interagir avec deux bications. La FZ41, un dérivé des styrylquinoléines, a une forme zwitterionique lui permettant d'interagir avec deux  $Mg^{2+}$  (Firley et al., 2006). Des dérivés des dicéto-acides peuvent également former des interactions favorables avec deux bications (Sechi et al., 2006).

Les bications  $Mg^{2+}$  et  $Mn^{2+}$  s'opposent souvent.  $Mg^{2+}$  est un ion dur de rayon relativement petit; il ne possède pas d'électrons d; son nuage électronique est peu déformable ; il interagit avec l'oxygène. Par contre,  $Mn^{2+}$  est un ion mou. Il est légèrement plus gros que  $Mg^{2+}$  mais surtout, il possède des électrons d et son nuage électronique peut se déformer et éventuellement former des liaisons covalentes; il interagit avec le soufre.

Gao et al. (2004) ont remplacé les résidus D64, D116 et E152 de l'IN par une cystéine et étudié les 3 activités des protéines D64C, D116C et E152C en fonction des bications  $Mg^{2+}$  et  $Mn^{2+}$ . Les protéines D64C, D116C et E152C deviennent inactives en présence de  $Mg^{2+}$ . Par contre, D116C et dans une moindre mesure D64C conservent un peu d'activité en transfert de brins et en 3'-processing, en présence de  $Mn^{2+}$ . E152C n'est jamais active en présence de  $Mn^{2+}$ . Seule D116C est active en désintégration. L'enzyme double mutante D64C, D116C est inactive. Ces données sont compatibles avec un modèle à un seul bication lié à D64 et D116. Comme la chaîne d'un glutamate est plus longue que celle d'une cystéine, le soufre de la cystéine pourrait être loin de la position des oxygènes du glutamate ce qui inactiverait E152C. Un deuxième bication n'est donc pas exclu.

En fait, quand on regarde le **tableau 2**, synthétisant les distances observées entre les ions et les protéines, on voit que le fer ou le zinc ont des rayons proches de ceux de  $Mg^{2+}$  ou de  $Mn^{2+}$ . Les partenaires sont cependant différents. Seuls  $Mn^{2+}$  et  $Mg^{2+}$  se lient préférentiellement aux groupes carboxyliques.  $Ca^{2+}$  est trop gros. Enfin la coordination de  $Zn^{2+}$  est trop éloignée de celle de  $Mg^{2+}$  et  $Mn^{2+}$ .



**Figure 13-** Structure RMN du domaine C-terminal de l'IN du VIH-1 (code PDB: **1IHV**) (Lodi et al., 1995). **A-** Un seul monomère en représentation « cartoons ». **B-** Le dimère en représentation « cartoons » avec l'interface formée par les brins  $\beta 2$ ,  $\beta 3$  et  $\beta 4$  des deux sous-unités **C-** Le dimère représenté en modèle compact. Les résidus K264 qui semblent importants pour la liaison de l'ADN, sont colorés en turquoise. Dans **A**, **B** et **C**, les brins  $\beta$  antiparallèles formant des tonneaux  $\beta$  sont colorés en violet.

## C - Le domaine C-terminal (résidus 212 à 288)

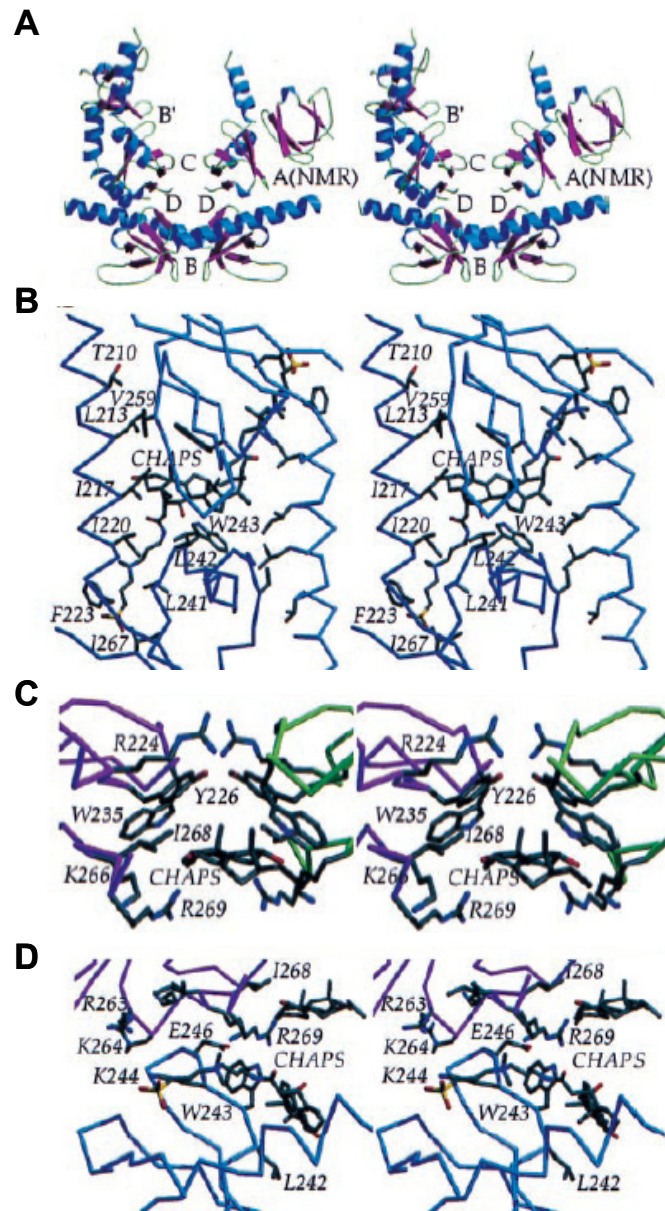
### C-1- Fonctionnalités du domaine C-terminal

Parmi les IN rétrovirales, le domaine C-terminal est considérablement moins conservé que les deux autres. La délétion de ce domaine supprime les activités de 3'-processing et de transfert de brins (Bushman et al., 1993; Bushman and Wang, 1994; Engelman et al., 1993; Jonsson et al., 1996; Pahl and Flugel, 1995; van Gent et al., 1993; Vink et al., 1993), excepté dans le cas de FIV (Shibagaki et al., 1997). Il a été montré que ce domaine possède une activité de liaison de l'ADN non-spécifique (Engelman et al., 1994; Khan et al., 1991; Lutzke et al., 1994; Vink et al., 1993; Woerner and Marcus-Sekura, 1993). Le résidu K264 semble important pour cette fonction (Lutzke et al., 1994). Il a été supposé, dans un premier temps, que le domaine C-terminal serait responsable de la liaison non-spécifique de l'ADN cible. Cependant, des études plus récentes ont montré que ce domaine est impliqué dans la liaison de l'ADN viral, au niveau de nucléotides un peu plus éloignés des extrémités LTR (Esposito and Craigie, 1998; Gao et al., 2001) (voir chapitre III).

D'autre part, ce domaine de l'IN lie, aussi bien que la rétrotranscriptase, la protéine cellulaire du développement ectodermique embryonnaire (Violot et al., 2003) et cette interaction semble être requise pour la rétrotranscription (Zhu et al., 2004).

### C-2- Structure du domaine C-terminal

Le domaine C-terminal entier (résidus 212 à 288) a tendance à s'agréger en solution. Une portion légèrement plus petite du domaine C-terminal (résidus 219 à 270) se comporte mieux en solution et a donc été utilisée pour les études structurales par RMN. Deux groupes indépendants ont déterminé la structure de cette portion du domaine C-terminal par RMN (Eijkelenboom et al., 1999; Lodi et al., 1995). Les deux structures RMN (codes PDB : **1IHV** et **1QMC**) sont étroitement en accord l'une avec l'autre. Le domaine C-terminal se présente sous la forme d'un dimère symétrique. Chaque monomère possède cinq brins  $\beta$  antiparallèles formant un tonneau  $\beta$  qui adopte un repliement de type SH3, observé dans plusieurs protéines impliquées dans la transduction du signal (**figure 13-A**). Le dimère est formé par l'empilement des brins  $\beta_2$ ,  $\beta_3$  et  $\beta_4$  des deux sous-unités (**figure 13-B**). Ces brins  $\beta$  d'une sous-unité sont antiparallèles aux brins  $\beta$  de l'autre sous-unité. L'interface du dimère est stabilisée par des contacts hydrophobes entre plusieurs résidus et exclut une aire de surface



**Figure 14-** Interactions entre domaines C-terminaux de l'IN du VIH-1. **A-** Les domaines C-terminaux appartenant aux dimères de bidomains (domaine central plus domaine C-terminal) de l'IN du VIH-1 (Chen et al., 2000). Les interfaces entre ces domaines C-terminaux forment quatre interfaces : B, B', C et D qui diffèrent de l'interface A observée dans la structure RMN du dimère de domaines C-terminaux (Lodi *et al.*, 1995). Les brins  $\beta$ , les hélices  $\alpha$  et les boucles sont respectivement colorés en violet, bleu et vert. **B-** Interface B dans une orientation tournée de  $90^\circ$  par rapport à celle de l'image A. **C-** Interface C. **D-** Interface D. Les interfaces C et D se trouvent dans des orientations identiques à celle de l'image A. Tiré de (Chen *et al.*, 2000a)

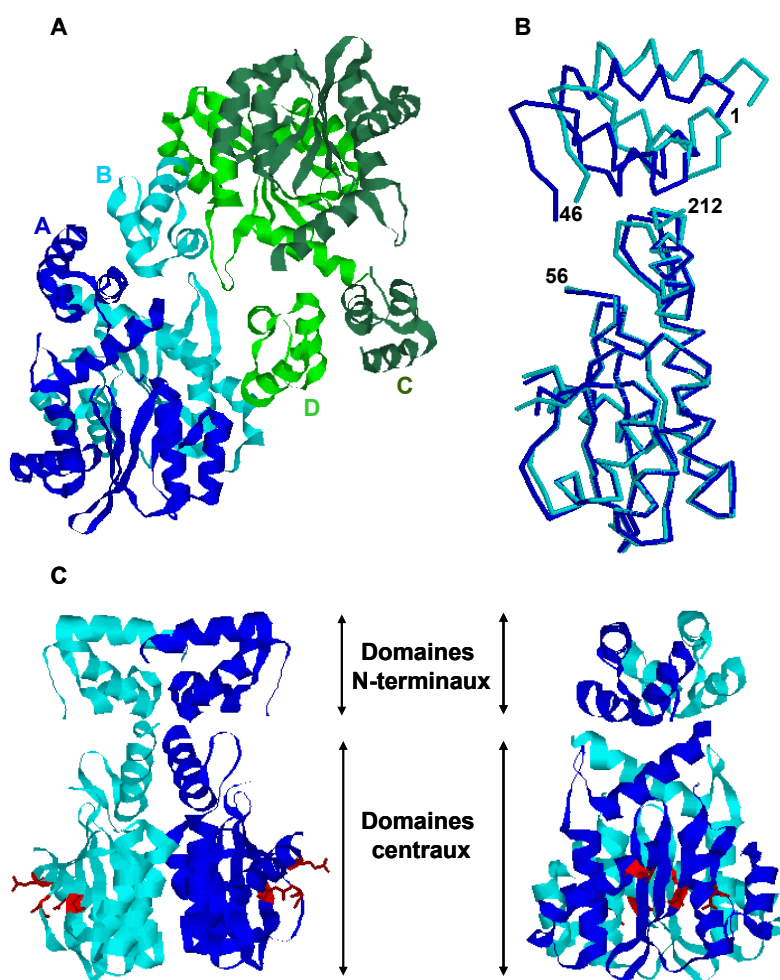
accessible au solvant, d'environ 360 Å<sup>2</sup> par sous-unité. La dimérisation forme une gouttière semblable à une selle contenant plusieurs résidus positivement chargés dont K264 (**figure 13-C**).

La structure du domaine C-terminal lié au domaine central a été déterminée par cristallographie aux rayons X (code PDB : **1EX4**) (Chen et al., 2000a). Au sein de ce bidomaine, la structure monomérique du domaine C-terminal est similaire à celles observées par RMN. Les résidus 212 à 218 du domaine C-terminal, absents dans les études RMN, appartiennent à l'hélice  $\alpha_6$  liant le domaine C-terminal au domaine central, dans la structure cristallographique. Au sein du cristal, la grande majorité des contacts dimère-dimère se font par des interactions entre les domaines C-terminaux. Ces interactions forment quatre interfaces différentes : B, B', C et D (**figure 14-A**). Bien que ces quatre interfaces soient différentes de l'interface du dimère RMN, celle qui en reste la plus proche est l'interface B qui implique les brins  $\beta_2$ ,  $\beta_3$  et  $\beta_4$  de deux dimères (**figure 14-B**). Cependant, ces brins  $\beta$  d'une molécule sont parallèles à ceux de l'autre molécule, contrairement à l'interface RMN où ils sont antiparallèles. De plus, au niveau de l'interface, une molécule de détergent CHAPS est en contact avec les résidus L242 et W243. Des interactions hydrophobes et des liaisons hydrogènes sont également observées entre les brins  $\beta_2$ ,  $\beta_3$  et  $\beta_4$  d'une molécule et l'hélice  $\alpha_6$  de l'autre molécule. Ces interactions sont encore plus nombreuses au niveau de l'interface B' qu'elles ne le sont au niveau de l'interface B. Les deux autres interfaces dimériques observées dans le cristal sont encore plus différentes de celle du dimère RMN. L'interface C implique un empilement antiparallèle des brins non contigus  $\beta_1$ ,  $\beta_2$  et  $\beta_5$  de deux dimères (**figure 14-C**). Une molécule CHAPS est également coincée entre les deux surfaces hydrophobes et le seul contact direct entre les deux feuillets est une liaison hydrogène entre R224N $\eta$ 1 et l'oxygène du groupe carbonyle de W235. L'interface D ne repose pas sur un axe de symétrie ; elle est formée par l'association bord à bord de la boucle entre les brins  $\beta_2$  et  $\beta_3$  (résidus 243-246) d'un dimère et le brin  $\beta_5$  (résidus 265-269) d'un autre dimère (**figure 14-D**). Une molécule CHAPS est également présente au niveau de cette interface. Ces résultats impliquent que l'interaction entre les monomères C-terminaux peut varier considérablement et il n'est pas du tout clair de savoir ce que serait le mode préféré de dimérisation pour le domaine C-terminal dans la protéine entière.

**Tableau 3-** Structures cristallographiques de bidomains de l'IN du VIH-1: domaine central lié au domaine N-terminal (**1K6Y**) ou au domaine C-terminal (**1EX4**).

Code PDB	Rés. (Å)	Groupe d'espace	Tampon	1 <sup>er</sup> et dernier résidus	Boucles de résidus absents	Mutations	Référence
1K6Y	2.40	P4 <sub>3</sub> 2 <sub>1</sub> 2	acétate	A:1,210 B:1,212 C:1,209 D:1,211	A-B-C:142-144 D:140-152	W131D F139D F185K	(Wang <i>et al.</i> , 2001)
1EX4	2.80	P3 <sub>1</sub> 2	citrate	A:56,270 B:55,270	A:142-144 B:138-149	C56S W131D F139D F185K C280S	(Chen <i>et al.</i> , 2000a)

Les pH des tampons acétate et citrate sont respectivement de 4.6 et 5.6.



**Figure 15-** Structure cristallographique du bidomaine (domaine N-terminal plus domaine central) de l'IN du VIH-1 (code PDB : **1K6Y**) (Wang *et al.*, 2001). **A-** Le tétramère formé par les quatre monomères (A, B, C et D) de l'unité asymétrique et représenté en modèle ruban. Les monomères A, B, C et D sont respectivement colorés en bleu, turquoise, vert foncé et vert clair. **B-** L'orientation du domaine N-terminal par rapport au domaine C-terminal diffère par un angle de 15° entre les deux sous-unités A (bleu) et B (turquoise), comme le montre la superposition des squelettes des domaines centraux des deux sous-unités. **C-** Deux vues orthogonales du dimère AB représentées en modèle ruban. Les résidus de la triade catalytique (D64, D116 et E152) sont colorés en rouge et représentés en boules-bâtons.

Bien que le domaine C-terminal ne présente pas d'homologie de séquence significative avec les autres IN rétrovirales, sa structure monomérique est similaire à celles des domaines C-terminaux des bidomaines (domaine central plus domaine C-terminal) des IN de RSV et de SIV (Chen et al., 2000a; Yang et al., 2000).

## **D - Les bidomaines de l'IN**

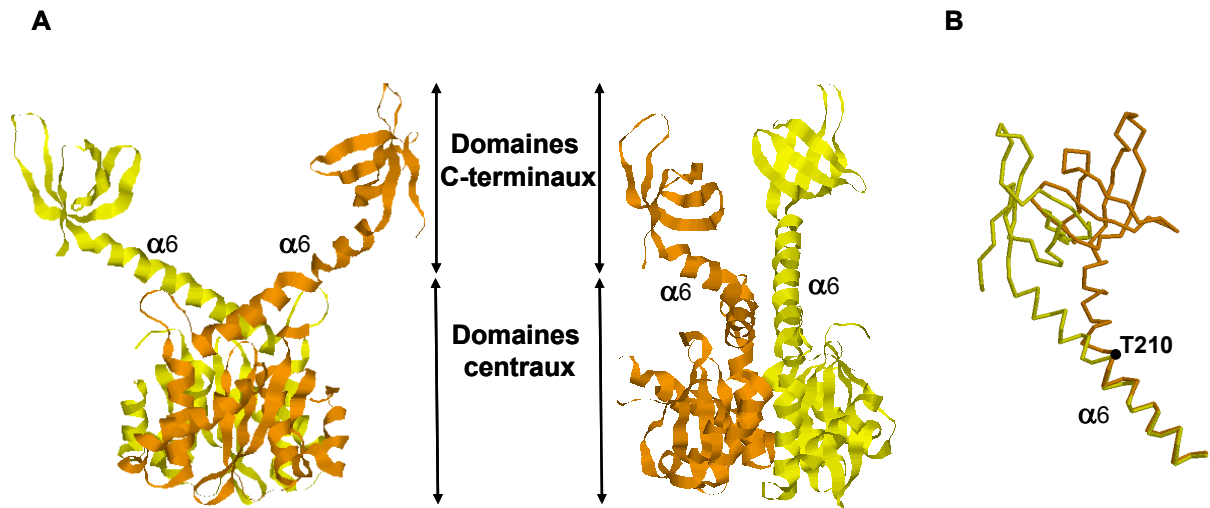
Les structures du domaine central lié au domaine N-terminal (code PDB : **1K6Y**) (Wang et al., 2001) ou au domaine C-terminal (code PDB : **1EX4**) (Chen et al., 2000a) ont été déterminées par cristallographie aux rayons X (**tableau 3**). La comparaison des structures de chacun des domaines formant les bidomaines avec celles des domaines isolés a été présentée précédemment. Il s'agit, maintenant, de montrer l'organisation spatiale entre les deux domaines liés et les arrangements des bidomaines en multimères.

### **D-1- Le domaine central lié au domaine N-terminal (résidus 1 à 212)**

Les cristaux du bidomaine (domaine central lié au domaine N-terminal) de l'IN du VIH-1 (code PDB : **1K6Y**, **tableau 3**) (Wang et al., 2001) contiennent quatre monomères (A, B, C et D) par unité asymétrique (**figure 15-A**). La région joignant le domaine central au domaine N-terminal (résidus 47 à 55) est désordonnée dans les quatre structures. Cependant, il semble n'y avoir aucune ambiguïté dans l'attribution des domaines N-terminaux aux domaines centraux appropriés. Les quatre monomères forment deux dimères (AB et CD) presque identiques dans l'unité asymétrique (**figure 15-C**). Dans chacun des deux dimères, l'angle entre les domaines N-terminaux et centraux diffère de 15 degrés entre les deux sous-unités (**figure 15-B**). Comme cette différence d'orientation est conservée dans les deux dimères AB et CD cristallographiquement indépendants, elle n'est vraisemblablement pas un artefact de l'effet cristallin mais doit plutôt être fonctionnellement appropriée. Les interfaces des dimères AB et CD sont très grandes. Elles recouvrent une surface moléculaire supérieure à  $1800 \text{ \AA}^2$  et sont constituées de résidus principalement polaires comme par exemple K14, Q18, Q44, K160, Q168, K186 et K188.

Les deux dimères sont reliés par un axe d'ordre 2 non cristallographique pour former un tétramère (**figure 15-A**). Les domaines N-terminaux et centraux sont en contact. Ils partagent une interface plutôt hydrophile ayant une aire totale de  $900 \text{ \AA}^2$ . Ce contact survient entre des





**Figure 16-** Structure cristallographique du bidomaine (domaine central plus domaine C-terminal) de l'IN du VIH-1 (code PDB : **1EX4**) (Chen *et al.*, 2000a). **A-** Deux vues orthogonales du dimère AB représentées en modèle ruban. Les monomères A et B sont respectivement colorés en orange et jaune. **B-** Le coude au niveau du résidu T210 de l'hélice  $\alpha 6$  crée une rotation d'environ 90° d'un domaine C-terminal par rapport à l'autre, comme le montre la superposition des résidus 195 à 210.

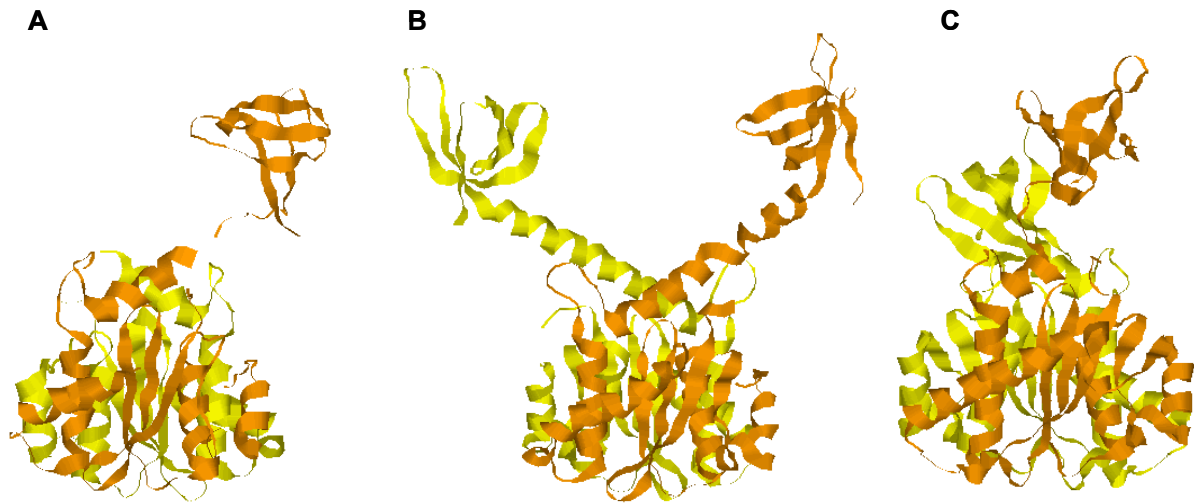
domaines de dimères séparés dans l'unité asymétrique : spécifiquement entre les résidus 13 à 26 du domaine N-terminal d'un dimère et les résidus 150 à 196 du domaine central de l'autre dimère. Il est tentant de considérer ce tétramère comme physiologiquement approprié. Cependant, comme le notent les auteurs (Wang et al., 2001), les expériences de filtration sur gel, à des concentrations salines physiologiques, montrent que le bidomaine (résidus 1 à 212) est dimérique (Jenkins et al., 1996). Ce qui soulève la question de savoir si ce tétramère peut être fonctionnellement approprié.

#### D-2- Le domaine central lié au domaine C-terminal (résidus 50 à 288)

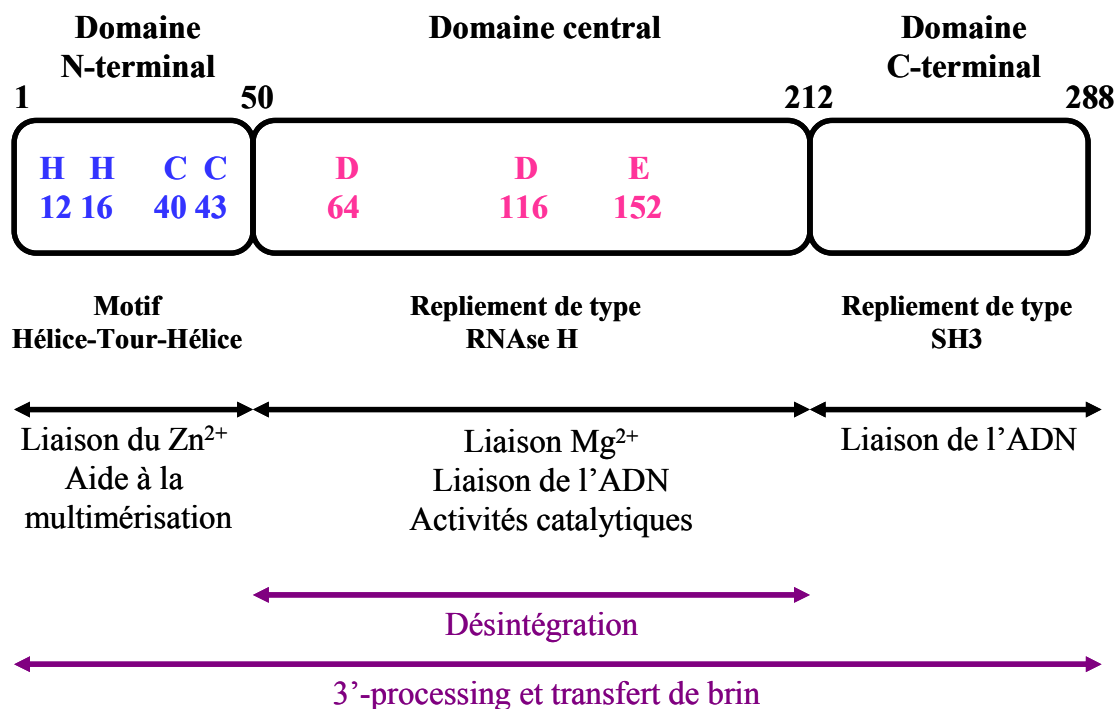
La structure du bidomaine (domaine central lié au domaine C-terminal) de l'IN du VIH-1 (code PDB : **1EX4**, **tableau 3**) (Chen et al., 2000a) est un dimère semblable à toutes les structures des domaines isolés. Ce dimère implique des interactions entre les domaines centraux des deux monomères formant une interface similaire à celle observée dans les dimères de domaines centraux isolés. Les domaines centraux sont reliés aux domaines C-terminaux par les résidus 195 à 220 de l'hélice  $\alpha_6$ . L'orientation divergente des hélices  $\alpha_6$ , à l'intérieur du dimère, place les centres des deux domaines C-terminaux, à environ 55 Å de distance, imposant une forme en « Y » à la structure dimérique (**figure 16-A**).

Un coude dans l'hélice  $\alpha_6$  du monomère B, au niveau du résidu T210, introduit une rotation d'environ 90° du domaine C-terminal du monomère B par rapport à celui du monomère A (**figure 16-A-B**). L'identification d'un site de clivage protéolytique, près du résidu T210, suggère que le coude de l'hélice  $\alpha_6$  peut représenter une jonction flexible en solution, permettant un rôle dynamique du domaine C-terminal au cours du processus catalytique.

Les structures des bidomaines (domaine central plus domaine C-terminal) des IN de RSV et de SIV ont été résolues par cristallographie aux rayons X (Chen et al., 2000b; Yang et al., 2000). Dans la structure de SIV (code PDB : **1C6V**), un seul domaine C-terminal, parmi les quatre existants, a pu être tracé dans l'unité asymétrique ; le lien entre ce domaine et le domaine central est désordonné. Une des quatre connections possibles entre le domaine C-terminal et les domaines centraux a tout d'abord été choisie car les deux domaines sont en contact mais l'organisation diffère alors énormément de celle du bidomaine du VIH-1. Par contre, si le domaine C-terminal peut se connecter au domaine central d'un dimère différent de celui suggéré précédemment, l'organisation spatiale du bidomaine est alors similaire à



**Figure 17-** Structure cristallographique du bidomaine (domaine central plus domaine C-terminal) de trois INs homologues: **A-** de l'IN de SIV (code PDB : **1C6V**) (Chen et al., 2000b), **B-** de l'IN du VIH-1 (code PDB : **1EX4**) (Chen *et al.*, 2000a) **C-** de l'IN de RSV (code PDB : **1C0M**) (Yang et al., 2000). Dans chaque dimère, une sous-unité est colorée en orange et l'autre en jaune. Les domaines C-terminaux de la sous-unité orange des trois structures sont positionnés de façon similaire par rapport au dimère central.



**Figure 18-** Les trois domaines de l'IN du VIH-1. Les résidus conservés importants sont indiqués à l'intérieur de chaque domaine. Les rôles putatifs sont mentionnés. Les types de repliement ou les motifs observés par les études de cristallographie et de RMN sont notés. Les domaines nécessaires pour chacune des réactions catalysées par l'IN sont mentionnés.

celle du bidomaine du VIH-1 (**figure 17-A**). Dans les deux structures cristallographiques du bidomaine de RSV (codes PDB : **1C0M** et **1C1A**), les domaines C-terminaux forment des dimères asymétriques et leur orientation par rapport au domaine central place l'un d'entre eux beaucoup plus près de ce dernier (**figure 17-C**). Ainsi, un domaine C-terminal n'entre pas en contact avec le domaine central auquel il est lié ; il a, alors, une position similaire à celle d'un domaine C-terminal du dimère du VIH-1. Par contre, le second domaine C-terminal du dimère de RSV entre en contact avec le domaine central auquel il est lié, ce qui peut être un artéfact dû à l'effet cristallin. Le segment joignant les domaines centraux et C-terminaux est formé par une hélice  $\alpha$  mais beaucoup plus petite (un tour) que l'hélice  $\alpha_6$  du bidomaine du VIH-1 (sept tours) et encadrée par deux boucles.

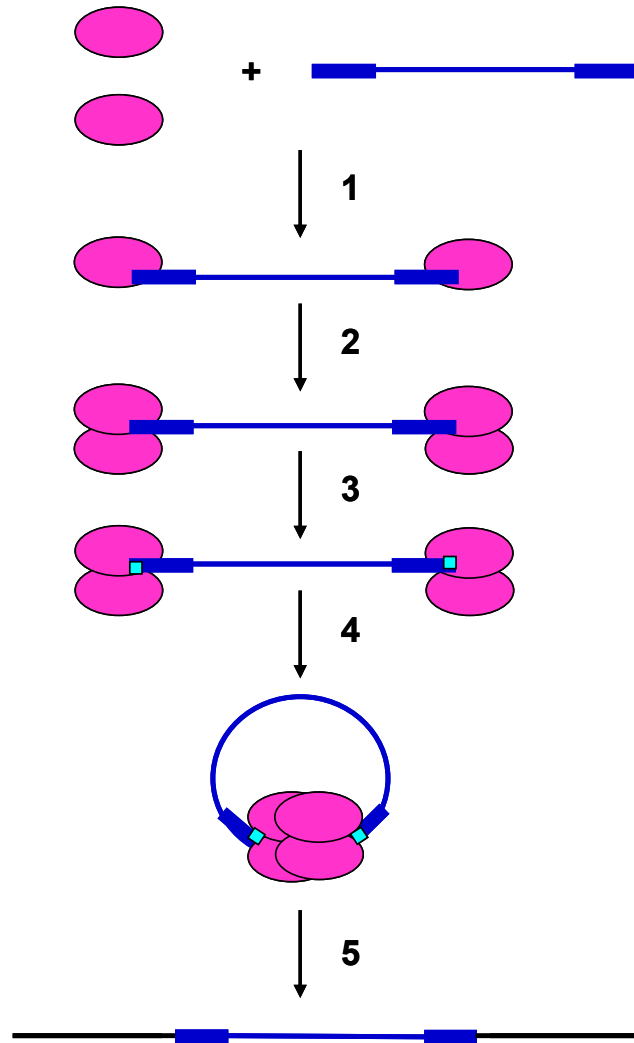
Un seul domaine C-terminal est positionné par rapport au dimère central, de façon similaire dans les trois structures (montré en orange dans la **figure 17**).

La connexion entre les domaines centraux et C-terminaux est flexible et diffère en longueur ; l'orientation et la conformation dans les bidomaines des trois IN sont homologues. Ceci soutient la notion de flexibilité fonctionnelle du segment de jonction entre les deux domaines, évoquée précédemment.

Cette analyse structurale et fonctionnelle de différents fragments de l'IN du VIH-1 peut être résumée par le schéma ci-contre (**figure 18**).

### **III- Oligomérisation de l'intégrase**

De nombreuses études indiquent que l'IN fonctionne sous la forme d'un multimère. Cela a été montré, en particulier, par le fait que certaines combinaisons de mutants de l'IN qui sont individuellement inactifs peuvent retrouver de l'activité *in vitro* (Ellison et al., 1995; Engelman et al., 1993; van den Ent et al., 1999). Ces travaux indiquent que la forme active de l'IN est au moins un dimère. Différents multimères de l'IN (monomères, dimères, tétramères, octamères...) sont présents en équilibre, en solution. L'état d'oligomérisation de l'IN en solution dépend de plusieurs facteurs tels que la présence de ses substrats ADN, la concentration en protéine, l'interaction avec des cations et la présence de détergents. (Deprez et al., 2000; Deprez et al., 2001; Lee et al., 1997; Leh et al., 2000). Il a été montré que  $Zn^{2+}$  stimule l'oligomérisation (Deprez et al., 2000; Zheng et al., 1996) et augmente l'activité  $Mg^{2+}$  dépendante de l'IN (Asante-Appiah and Skalka, 1999; Lee et al., 1997; Leh et al., 2000;



**Figure 19-** Représentation schématique des états oligomériques de l'IN du VIH-1 au cours du processus catalytique. L'IN est présente en solution dans un équilibre entre différentes formes oligomériques. Un monomère se lie à chaque extrémité LTR de l'ADN viral (1), permettant la dimérisation correcte de l'IN (2). Le 3'-processing de chaque extrémité peut être réalisé par les dimères (3). Les deux dimères peuvent interagir ensemble (4) et former un tétramère prêt à intégrer l'ADN viral dans l'ADN cible (5). Chaque monomère de l'IN est représenté par un ovale rose. L'ADN viral est coloré en bleu avec les LTR représentés par des rectangles. Les petits carrés turquoise indiquent les extrémités virales clivées au cours du 3' processing. D'après (Faure et al., 2005).

Zheng et al., 1996), certainement via sa liaison au motif HHCC du domaine N-terminal de l'IN. L'activité de l'IN  $Mg^{2+}$ -dépendante est directement reliée à l'état d'oligomérisation de la protéine (Leh et al., 2000) ; ce qui n'est pas le cas de l'activité de l'IN  $Mn^{2+}$ -dépendante. En effet, l'IN préparée de façon traditionnelle en présence de détergent est monomérique aux concentrations inférieures au micromolaire et active seulement en présence de  $Mn^{2+}$  (un bication beaucoup moins présent *in vivo* que  $Mg^{2+}$ ). Par contre, l'IN préparée en l'absence de détergent et en présence de  $Zn^{2+}$  est tétramérique et active en présence de  $Mg^{2+}$  et de  $Mn^{2+}$ . Un lien a été établi entre la capacité de l'IN de lier l'ADN et l'oligomérisation (Vercammen et al., 2002). Des complexes stables sont obtenus entre le substrat d'ADN portant une extrémité LTR et des monomères ou des dimères de l'IN, à la température physiologique (Deprez et al., 2001). Cependant, cette étude n'exclut pas la possibilité d'une organisation d'ordre plus élevée, en présence de l'ADN viral entier, avec les deux extrémités LTR.

L'étude de l'activité de différentes formes oligomériques de l'IN, stabilisé chimiquement par pontage, a apporté des résultats importants concernant l'état d'oligomérisation de l'IN, au cours des étapes catalytiques (Faure et al., 2005). Un dimère est impliqué dans la catalyse du 3'-processing et de l'intégration d'une extrémité LTR de l'ADN viral. Par contre, pour réaliser l'intégration des deux extrémités de l'ADN viral, un tétramère au moins est nécessaire. Ceci est cohérent avec l'analyse structurale du dimère de domaines centraux, montrant qu'en raison de la distance trop grande entre les deux sites actifs du dimère, un tétramère au moins est nécessaire pour l'intégration concertée. De plus, le tétramère ponté de l'IN apporte un pourcentage de produits intégrés à 5 paires de bases plus élevé que l'IN non pontée. Ces données suggèrent fortement que le tétramère est la forme minimale active impliquée dans l'intégration provirale nucléaire. Cependant, de façon surprenante, le tétramère ponté n'est pas capable de réaliser l'intégration d'une seule extrémité LTR de l'ADN viral, comme peut le faire la forme dimérique pontée. Ainsi, il se peut que le dimère se lie seulement à une extrémité LTR alors que la structure du tétramère permet seulement la liaison de deux extrémités LTR, proches l'une de l'autre. Ceci suggère alors que la structure du dimère isolé et celle du dimère impliqué dans la forme tétramérique soient différentes. Le tétramère ponté nécessaire pour l'intégration des deux extrémités LTR n'est pas actif pour le 3'-processing, contrairement à la forme dimérique pontée : des formes oligomériques différentes sont donc impliquées dans les deux étapes catalytiques (**figure 19**) (Faure et al., 2005). Les monomères se lient à chaque LTR, permettant la dimérisation de l'IN. Les deux dimères peuvent effectuer le 3'-processing des deux extrémités virales puis interagir ensemble pour former un tétramère capable d'intégrer les deux LTR dans l'ADN cible. Ce modèle est



en accord avec différentes approches (Deprez et al., 2001; Podtelezhnikov et al., 2003). Des études de mutagenèse montrent également l'implication de différents multimères de l'IN dans les deux étapes catalytiques et aboutissent à des conclusions similaires (Calmels et al., 2004). Une formation séquentielle de différents complexes, accompagnée par des changements conformationnels (Asante-Appiah et al., 1998; Espeseth et al., 2000; Hazuda et al., 2000) semble donc nécessaire pour réaliser les deux étapes catalytiques. Après la réaction de 3'-processing, les complexes sont plus stables. Ils contribuent probablement à l'assemblage correct et évitent la réversibilité de la réaction (Gao et al., 2001).

L'association oligomérique de l'IN *in vivo* est probablement gouvernée par des protéines cellulaires. La protéine chaperonne cellulaire HSP60 est impliquée dans la stabilisation de la structure active de l'IN (Parissi et al., 2001). La protéine cellulaire LEDGF(p75) qui interagit fortement avec l'IN et qui a été trouvée associée avec l'IN du VIH-1 dans des cellules humaines, peut être impliquée dans la formation de multimères stables de l'IN (Cherepanov et al., 2003; Cherepanov et al., 2005b). De telles protéines, absentes *in vitro*, pourraient stabiliser une organisation tétramérique de l'enzyme, dans le processus cellulaire d'intégration.





## **Chapitre III**

### **L'interaction de l'intégrase avec ses substrats et son mécanisme d'action**

---

Afin de catalyser les réactions de 3'-processing et de transfert de brins, l'IN doit se lier à l'ADN viral et à l'ADN cible. L'IN interagit avec ces deux molécules d'ADN de deux façons différentes : la liaison à l'ADN cible ne dépend pas directement de la séquence nucléotidique alors que l'interaction avec les extrémités LTR de l'ADN viral est un processus séquence-spécifique. Néanmoins, les séquences des LTR U3 et U5 ne sont pas exactement identiques.

#### **I- Généralités sur les interactions protéines/ADN**

L'interaction séquence-spécifique entre protéines et acides nucléiques est un processus complexe pouvant impliquer une reconnaissance directe à travers des liaisons spécifiques entre les deux molécules mais aussi une reconnaissance indirecte dépendant de la structure, de la forme et de la flexibilité des molécules mises en jeu. Seule la connaissance précise à l'échelle atomique des structures tridimensionnelles des molécules, de leur flexibilité et des termes énergétiques mis en jeu, peut permettre de comprendre la nature ainsi que le mécanisme de ces interactions. On a pu observer, sur de nombreux complexes, que la fixation d'une protéine induit des déformations de la molécule d'ADN. Bien que certaines de ces déformations soient directement induites par la protéine, il est aujourd'hui reconnu que la séquence des bases de l'ADN, elle-même, influence la structure et la flexibilité de la double hélice. Ainsi, certaines déformations sont préexistantes dans la molécule d'ADN et accentuées par la liaison de la protéine. Initialement perçue comme une molécule régulière et rigide, la double hélice ADN présente en fait une forte hétérogénéité structurale et dynamique en fonction de sa séquence en bases et de son environnement. Ces propriétés jouent un rôle



majeur dans les processus de reconnaissance impliquant les protéines qui ciblent la double hélice. Il faut, par conséquent, prendre en compte l'ensemble des caractéristiques dynamiques intrinsèques à la molécule d'ADN pour pouvoir réellement comprendre comment une protéine peut reconnaître et interagir avec un site spécifique de la double hélice.

La courbure intrinsèque de l'ADN participe activement à sa reconnaissance et à son interaction avec les protéines (Flatters and Lavery, 1998; Flatters et al., 1997; Wu and Crothers, 1984). Des courtes successions d'adénine « A-tracts » favorisent la courbure (Bostock-Smith et al., 1999; Burkhoff and Tullius, 1987; McConnell and Beveridge, 2001). La courbure est intimement liée aux discontinuités structurales.

Le plissement du sucre des nucléotides joue un rôle très important dans les complexes ADN/protéine. Des transitions Sud  $\rightarrow$  Nord de certains nucléotides sont souvent observées dans les complexes protéine/ADN (Berman et al., 1992). Ces transformations locales de la forme B vers la forme A de l'ADN entraînent un élargissement et un aplatissement du petit sillon de l'ADN, permettant d'étendre les interactions hydrophobes avec les protéines dans le petit sillon (Tolstorukov et al., 2004). Par contre, le grand sillon devient plus étroit et plus profond, ce qui favorise les interactions avec les hélices  $\alpha$  des protéines. La plupart des acides aminés polaires interagissent avec les nucléotides dont le sucre est en Sud alors que les acides aminés hydrophiles interagissent avec les nucléotides dont le sucre est en Nord. La majorité des complexes protéines/ADN où surviennent des transformations B  $\rightarrow$  A, implique des enzymes qui coupent ou forment la liaison phosphodiester O3'-P (Lu et al., 2000). Ces transitions conformationnelles exposent l'atome O3' qui est habituellement enfoui à l'intérieur du squelette et protégé des attaques enzymatiques. Presque tous les ADN polymérasés caractérisés structurellement induisent une transition B  $\rightarrow$  A des pas immédiatement adjacents au site d'incorporation du nucléotide. Par exemple, deux des sept pas de l'ADN liés dans le site actif de la *Taq* polymérase ont une forme A et trois autres sont partiellement distordus vers une forme A (Eom et al., 1996). Des caractéristiques structurales de forme A sont observées dans les substrats ADN des endonucléases. Le coude du substrat ADN d'EcoRV présente des caractéristiques structurales d'une forme A (Lu et al., 2000). Deux fragments de l'ADN liés à l'endonucléase I-PoI de *Physarum polycephalum* présentent un comportement typique de l'ADN A (Flick et al., 1998; Galburt et al., 1999). Il existe beaucoup d'autres exemples de ce type de déformation (Lu et al., 2000). Le coût énergétique de ce type de transition B  $\rightarrow$  A est séquence dépendante (Minchenkova *et al.*,

**Tableau 4-** Effets sur le 3'-processing de substitutions de bases sur le brin clivé de l'extrémité LTR U5 de l'ADN viral du VIH-1 (Esposito *et al.*, 1998)

			Substitutions de bases	Taux initial de 3'-processing	
				Mg <sup>2+</sup>	Mn <sup>2+</sup>
Extrémité du LTR U5 de l'ADN viral du VIH-1 : brin clivé lors du 3'-processing	1	T <sub>3'</sub>	A	1,00	0,69
			C	0,08	0,90
			G	0,50	0,85
	2	G	A	0,11	0,23
			C	0,08	0,54
			T	1,50	1,04
	3	A	C	0,00	0,00
			G	0,00	0,00
			T	0,00	0,00
	4	C	A	0,00	0,08
			G	0,04	0,04
			T	0,00	0,02
	5	G	A	0,70	0,92
			C	0,72	0,85
			T	0,23	0,15
	6	A	C	0,98	0,70
			G	0,08	0,35
			T	0,50	0,84
	7	T	A	0,85	0,96
			C	1,15	1,00
			G	1,00	0,92
	8	C	A	0,85	0,96
			G	1,10	1,20
			T	1,12	0,92
	9	T	A	0,96	0,96
			C	0,88	1,00
			G	0,80	0,95
	10	C	A	1,06	1,15
			G	0,15	0,35
			T	1,20	1,11
	11	T	A	0,96	0,90
			C	0,08	0,42
			G	0,70	1,04
	12	A	C	0,23	0,85
			G	0,40	1,04
			T	0,35	0,92
	13	5'A	C	0,30	0,83
			G	0,08	0,92
			T	0,60	0,92

Le taux initial de 3'-processing avec l'oligonucléotide sauvage est considéré égal à 1

1986; Tolstorukov *et al.*, 2001). Il est donc possible que ces transitions augmentent la sélectivité de la liaison ADN/protéine.

Des interconversions  $B_I \rightarrow B_{II}$  de certains nucléotides jouent également un rôle important dans les interactions ADN/protéine (Winger *et al.*, 1998; Tisne *et al.*, 1998) et elles sont aussi séquence dépendante. La conformation  $B_I$  place le phosphate dans une position rigoureusement symétrique par rapport aux deux sillons alors que la conformation  $B_{II}$  déplace le phosphate dans le petit sillon. L'importance de ces transitions a notamment été montrée pour la liaison du facteur de transcription NF- $\kappa$ B avec son site ADN (Tisne *et al.*, 1998). La transition  $B_I \rightarrow B_{II}$  permet d'attirer et de lier le facteur de transcription. Une fois le complexe formé, les contacts entre la protéine et l'ADN induisent le retour à la conformation  $B_I$  d'énergie plus favorable. La transition  $B_I \rightarrow B_{II}$  est la conséquence d'un processus de désemplètement de paires de bases adjacentes et elle est couplée à une migration de l'eau du phosphate ionique vers l'oxygène du sucre (Winger *et al.*, 1998). Des ions à haute densité de charge tel que  $Mg^{2+}$  favorisent les transitions  $B_I \rightarrow B_{II}$  (van Dam and Levitt, 2000).

## **II- Interaction IN/ADN**

L'interaction de l'IN avec ses substrats ADN est très mal connue. Etant donné qu'il n'existe pas de structure expérimentale de l'IN entière ni du complexe qu'elle forme avec ses substrats, les sites de liaisons IN/ADN ne sont pas définis : seuls des contacts ont été déterminés ou des régions de liaisons ont été prédites. Ainsi, on ne connaît pas le mécanisme de reconnaissance impliqué dans l'interaction séquence-spécifique de l'IN avec l'ADN viral.

### **A - Etudes expérimentales de l'interaction IN/ADN**

#### **A-1- Effets des modifications de l'ADN viral sur l'activité catalytique de l'IN**

Les oligonucléotides reproduisant les extrémités LTR (U5 ou U3) de l'ADN viral ont subi diverses modifications au niveau et près du site de clivage :

- 1) substitution de bases par d'autres bases (A, T, G, C ou U) au niveau du brin clivé ou non clivé de l'oligonucléotide mimant l'extrémité LTR U5 ou U3 (Esposito and Craigie, 1998; Katzman and Sudol, 1996a; LaFemina *et al.*, 1991; Mazumder and Pommier, 1995; Reicin *et al.*, 1995; Sherman *et al.*, 1992; van den Ent *et al.*, 1994; Vink *et al.*, 1991b; Yoshinaga and Fujiwara, 1995) (**tableau 4**).

**Tableau 5-** Effets sur le 3'-processing de substitutions de paires de bases au niveau de l'extrémité LTR U5 de l'ADN viral du VIH-1 (Scottoline *et al.*, 1997)

		Substitutions de paires de bases	Taux initial de 3'-processing Mn <sup>2+</sup>	
Extrémité double brins du LTR U5 de l'ADN viral du VIH-1	1	T <sub>3'</sub> A <sub>5'</sub>	G/C	0,3
			G/G	1,1
			C/C	1,8
			C/G	0,3
	2	G C	A/T	0,2
			A/A	1,7
			T/T	1,4
			T/A	0,3
	3	A T	C/G	0,08
			C/A	0,8
			T/G	1,0
			T/A	0,06
	4	C G	G/C	0,01
			G/A	0,01
			T/C	0,1
			T/A	0,03
5	G C	A/T	0,8	
		A/G	0,1	
		C/T	0,5	
		C/G	1,0	
6	A <sub>5'</sub> T <sub>3'</sub>	G/C	0,4	
		G/G	0,4	
		C/C	0,5	
		C/G	0,5	

Le taux initial de 3'-processing avec l'oligonucléotide sauvage est considéré égal à 1

**Tableau 7-** Effets sur le 3'-processing de modifications de bases sur le brin clivé et non clivé de l'extrémité LTR U5 de l'ADN viral du VIH-1 (Wang *et al.*, 1999)

Extrémité du LTR U5 de l'ADN viral du VIH-1			Taux initial de 3'-processing Mn <sup>2+</sup>
Brin clivé			
5 5'-G	4 C	3 A <sub>3'</sub>	
		6-MeA	1,00
		7-deazaA	0,30
		2,6-(NH <sub>2</sub> ) <sub>2</sub> Pur	0,10
	5-MetC		0,50
6-MeA			0,90
7-deazaA			0,67
2,6-(NH <sub>2</sub> ) <sub>2</sub> Pur			0,76
A			0,60
Inosine			0,92
6-MeG			0,33
Brin non clivé			Taux de 3'-processing Mn <sup>2+</sup>
5 3'-C	4 G	3 T <sub>5'</sub>	
		O <sup>4</sup> -MeT	3,00
	Inosine		0,06
	6-MeG		0,04

Le taux initial de 3'-processing avec l'oligonucléotide sauvage est considéré égal à 1

**Tableau 6-** Effets sur le 3'-processing et le transfert de brins de remplacement de bases par des sites abasiques au niveau du brin clivé de l'extrémité LTR U5 de l'ADN viral du VIH-1 (Mazumder *et al.*, 1995).

Positions du brin clivé du LTR U5 remplacées par un site abasique	Taux initial de 3'-processing Mn <sup>2+</sup>	Taux initial de transfert de brins Mn <sup>2+</sup>
3 A <sub>3'</sub>	1	0
4 C	0	0
5 G	0	0
9 T	0,7	0,5
13 A <sub>5'</sub>	0,9	0,8

Les taux initiaux de 3'-processing et de transfert de brins avec l'oligonucléotide sauvage sont considérés égaux à 1

- 2) substitution de paires de bases de l'oligonucléotide reproduisant l'extrémité LTR U5 par des paires de bases complémentaires ou non (A/T, T/A, T/T, A/A, C/G, G/G, G/G, C/C) (Scottoline et al., 1997; Yoshinaga and Fujiwara, 1995) (**tableau 5**).
- 3) remplacement de nucléotides par des sites abasiques sur le brin clivé de l'oligonucléotide reproduisant l'extrémité LTR U5 (Mazumder and Pommier, 1995; van den Ent et al., 1994) (**tableau 6**).
- 4) ajout de nucléotides à l'extrémité 3' du brin clivé du LTR U5 (Scottoline et al., 1997; van den Ent et al., 1994)
- 5) modification des bases du brin clivé et non clivé de l'oligonucléotide reproduisant l'extrémité LTR U5 par :
  - ajout d'un groupe méthyle sur l'oxygène O4 de la thymine (O<sup>4</sup>-MéthyleThymine), l'oxygène O6 de la guanine (6-MéthyleGuanine), l'azote N6 de l'adénine (6-MéthyleAdénine) et le carbone C5 de la cytosine (5-MéthyleCytosine) (Wang et al., 1999) (**tableau 7**).
  - ajout d'un groupe amino sur l'azote N2 de l'adénine (2,6-diaminopurine) (Wang *et al.*, 1999) (**tableau 7**).
  - remplacement de l'azote N7 de l'adénine par un carbone (7-DeazaAdénine) (Wang *et al.*, 1999) (**tableau 7**).
  - ajout d'une chaîne de grand encombrement stérique sur le N2 de la guanine (Johnson et al., 2004).
- 6) remplacement de nucléotides par des Inosines sur le brin clivé ou non clivé de l'oligonucléotide reproduisant l'extrémité LTR U5 (Wang et al., 1999) (**tableau 7**).

Les effets de ces modifications sur le 3'-processing et dans certains cas sur le transfert de brins et la désintégration, ont été examinés.

De nombreuses études montrent l'importance du dinucléotide C4-A3 du brin clivé pour les réactions de 3'-processing et de transfert de brins (Esposito and Craigie, 1998; LaFemina et al., 1991; Sherman et al., 1992; Vink et al., 1991a). Cependant, certaines modifications au niveau de la troisième paire de bases n'ont pas d'effet sur le 3'-processing (Mazumder and Pommier, 1995; Scottoline et al., 1997). Par exemple, les substitutions par T/G et C/A n'affectent pas le 3'-processing (**tableau 5**). De même, la substitution d'A3 par la désoxyuridine ou par un site abasique (**tableau 6**) n'inhibe pas la réaction (Mazumder and Pommier, 1995).

Plusieurs études, réalisées en présence de Mn<sup>2+</sup>, indiquent que les nucléosides les plus importants pour l'intégration se situent dans la région allant de la 3<sup>ème</sup> à la 9<sup>ème</sup> base du brin





clivé (Balakrishnan and Jonsson, 1997; Katzman and Sudol, 1996a; LaFemina et al., 1991; Sherman et al., 1992; van den Ent et al., 1994; Vink et al., 1991b; Yoshinaga and Fujiwara, 1995). Cependant, certaines substitutions plus éloignées de l'extrémité du LTR U5 (positions 10,11,12 et 13) montrent un effet inhibiteur sur le 3'-processing (**tableau 4**) (Esposito and Craigie, 1998). Des substitutions des bases des positions 10 et 11 sur le brin clivé diminuent le taux de 3'-processing en présence de  $Mn^{2+}$  mais l'effet est plus important en présence de  $Mg^{2+}$ . Au niveau des positions 12 et 13, des substitutions affectent également le 3'-processing mais uniquement en présence de  $Mg^{2+}$ .

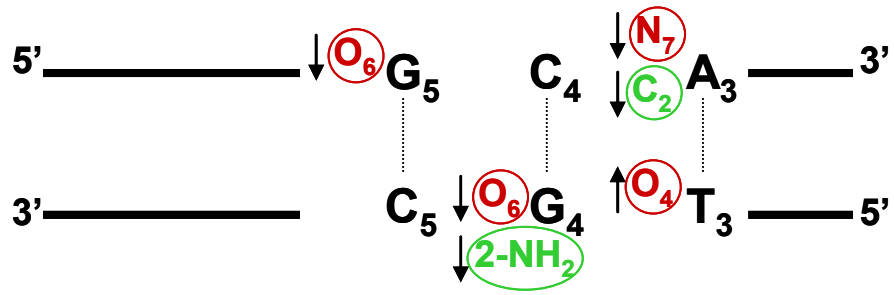
La comparaison des résultats obtenus en présence de  $Mn^{2+}$  et de  $Mg^{2+}$  pour les substitutions des bases 1 à 13 du brin clivé (**tableau 4**) suggère que le site actif de l'IN puisse être moins rigide en présence de  $Mn^{2+}$ . Cela permettrait alors une sélection plus large de nucléophiles et réduirait potentiellement la spécificité de liaison de l'ADN lors de la réaction de 3'-processing. Cette hypothèse aiderait à expliquer la raison pour laquelle l'activité de désintégration, en présence de  $Mn^{2+}$ , ne semble pas exiger une spécificité de séquence, au-delà du dinucléotide C4-A3 (Balakrishnan and Jonsson, 1997; Chow et al., 1992).

Certaines modifications en positions 5 et 6 inhibent considérablement le 3'-processing, en présence de  $Mn^{2+}$  et de  $Mg^{2+}$  (**tableaux 4,5 et 6**) (Esposito and Craigie, 1998; Mazumder and Pommier, 1995; Scottoline et al., 1997). Par contre, les modifications en positions 7, 8 et 9 ne montrent pas d'effet majeur sur la réaction (**tableaux 4 et 6**) (Esposito and Craigie, 1998; Mazumder and Pommier, 1995).

L'ajout de nucléotides, à l'extrémité 3' du brin clivé inhibe le 3'-processing (van den Ent et al., 1994).

Les substitutions par des paires de bases complémentaires, au niveau des positions 1, 2 et 3, affectent le 3'-processing (**tableau 5**) (Scottoline et al., 1997). Par contre, les substitutions des mêmes positions par des paires de bases non-complémentaires n'ont pas d'effet (position 3) ou stimulent (positions 1 et 2) le 3'-processing. Des effets favorables de substitutions par des paires de bases non-complémentaires ne sont pas observés pour les positions 4, 5 et 6. Au contraire, le 3'-processing est inhibé. Ainsi, des substrats contenant des paires de bases non complémentaires à l'extrémité du LTR U5 (positions 1 et 2) sont mieux clivés que le substrat sauvage, que se soit en présence de  $Mg^{2+}$  ou de  $Mn^{2+}$ . Des modifications de l'appariement normal Watson-Crick des deux premières paires de bases semblent donc favorables au 3'-processing.

Est-ce que l'ADN interagit avec l'IN par son petit ou son grand sillon ? Pour répondre à cette question, Wang et al. (1999) ont modifié les groupes donneurs ou accepteurs de liaisons



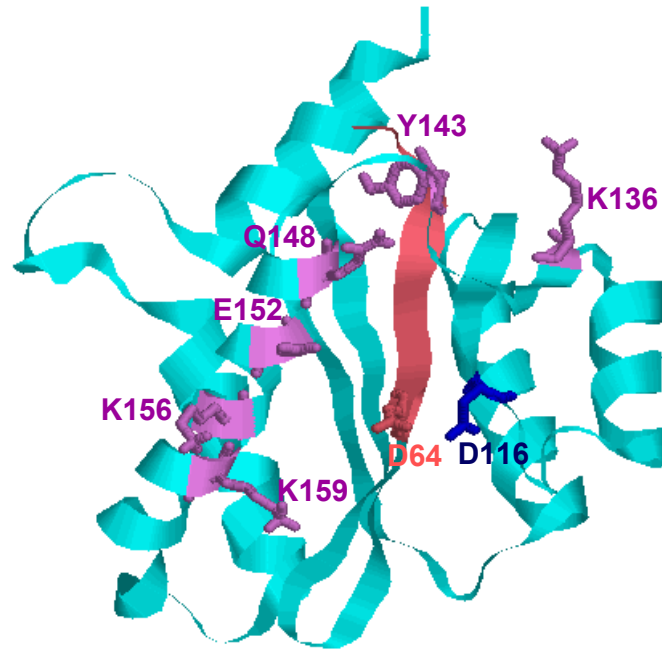
**Figure 20-** Des sites critiques pour les interactions de l'IN avec l'extrémité U5 de l'ADN viral du VIH-1, dans le grand sillon (en rouge) et dans le petit sillon (en vert) (Wang et al., 1999). Les flèches indiquent si une augmentation ou une diminution du 3'-processing a été observée.

hydrogènes dans le petit et le grand sillon de l'ADN (**tableau 7**) (Wang et al., 1999). Les sites d'interactions identifiés sont résumés dans la **figure 20**. Au niveau d'A3, N7 (grand sillon) et C2 (petit sillon) pourraient être impliqués dans l'interaction avec l'IN. Dans le brin non clivé, le remplacement de T3 par la O<sup>4</sup>-Méthylethymine stimule le 3'-processing. Ceci suggère que le groupe méthyle (grand sillon) augmente les contacts hydrophobes avec l'IN ou réduit l'interaction de la paire de bases A3-T3, au bénéfice de l'IN. Au niveau de la position 4, les modifications du brin non clivé inhibent considérablement le 3'-processing. L'O6 (grand sillon) et le 2-NH<sub>2</sub> (petit sillon) de G4 pourraient donc être impliqués dans des liaisons hydrogènes. L'implication du groupe 2-NH<sub>2</sub> de G4 dans des interactions avec l'IN a également été suggéré par d'autres auteurs (Johnson et al., 2004). Les modifications à la position 5 sur le brin clivé affectent moins le 3'-processing. Seule l'O6-MéthyleGuanosine diminue de 3,7 fois le 3'-processing, suggérant que l'O6 (grand sillon) de G5 pourrait être impliquée dans une liaison hydrogène. Globalement, les perturbations au niveau du petit sillon défavorisent davantage la réaction de 3'-processing que celles ayant lieu dans le grand sillon.

Les auteurs cités pensent en liaisons hydrogènes ou en stabilité de l'ADN, notamment en stabilité des appariements des bases complémentaires. Cependant, il peut être difficile de distinguer si la modification de l'ADN crée ou annihile une liaison hydrogène entre l'ADN et l'IN ou bien si elle entraîne un changement de la conformation de l'ADN qui modifie sa fixation sur l'IN.

#### A-2- Expériences de complémentation fonctionnelle, de mutagenèse, d'empreinte protéique et de pontage

Le domaine central de l'IN est la seule partie de l'enzyme capable de réaliser le transfert polynucléotidique. On a supposé alors que ce domaine possédait les déterminants principaux de la spécificité de la liaison aux extrémités LTR de l'ADN viral. Des expériences de pontage et de complémentation fonctionnelle ont montré que le domaine central interagit avec les extrémités LTR de l'ADN viral (Ellison and Brown, 1994; Esposito and Craigie, 1998; Gao et al., 2001; Heuer and Brown, 1997; Heuer and Brown, 1998; Yoshinaga et al., 1994). Des contacts IN/ADN ont pu être mis en évidence. Le résidu K159, présent dans toutes les IN rétrovirales et dans beaucoup de transposases bactériennes, peut interagir avec l'azote N7 d'A3 (Jenkins et al., 1997). Le résidu catalytique E152 peut également être impliqué dans la



**Figure 21-** Les résidus et portions du domaine central de l'IN du VIH-1 qui ont été mis en évidence pour être impliqués dans la liaison de l'ADN viral par des expériences de pontage. Un monomère du domaine central de l'IN (code PDB : 1BIS) (Goldgur *et al.*, 1998) est représenté en modèle ruban et coloré en turquoise. Les résidus colorés en violet sont impliqués dans la liaison de l'extrémité LTR U5 de l'ADN viral. Des résidus appartenant à la portion (51-64), colorée en rose, peuvent également être en contact avec l'ADN viral. Le résidu catalytique D116 est coloré en bleu.

liaison avec A3 (Gerton et al., 1998). Deux autres résidus lysines, K156 et K136, sont importants pour la liaison de l'ADN viral ; K136 peut interagir avec G5 (Jenkins et al., 1997). Des contacts avec le brin non clivé ont également été identifiés. Les résidus Q148 et Y143 peuvent être en contact avec A1 (Esposito and Craigie, 1998). Des études plus récentes indiquent que Q148 peut être lié à C2 lorsque l'ADN viral est clivé; ce résidu serait alors impliqué dans le positionnement correct de cet ADN, avant le transfert de brins (Johnson et al., 2006a). Des résidus de la portion 51 à 64 du domaine central peuvent interagir avec C2 (Esposito and Craigie, 1998). La **figure 21** montre, dans le domaine central de l'IN, la localisation des résidus qui interagiraient avec l'ADN viral. Tous ces résidus du domaine central seraient donc responsables de la reconnaissance séquence-spécifique des derniers nucléotides, à l'extrémité des LTR de l'ADN viral.

Le domaine C-terminal est impliqué dans la liaison de l'ADN non-spécifique car des mutations de ce domaine peuvent abolir la liaison (Lutzke and Plasterk, 1998; Lutzke et al., 1994). La mutation de K264 a un effet dramatique sur la liaison non spécifique de l'ADN par le domaine C-terminal isolé. De plus, la réaction de 3'-processing, catalysée par l'IN entière contenant la mutation K264E, est affectée significativement. Ce résidu est bien conservé parmi les IN rétrovirales. L'implication de R262 et L234, dans la liaison de l'ADN, est également suggérée. Les mutations de S230 et R263 affectent aussi l'activité catalytique et la liaison de l'ADN. Le domaine C-terminal ayant un rôle dans la liaison de l'ADN non-spécifique, il a tout d'abord été proposé que ce domaine soit responsable de la liaison de l'ADN cible. Mais des travaux plus récents montrent que ce domaine interagit avec l'ADN viral, au niveau de nucléotides éloignés de 7 paires de bases de l'extrémité des LTR (Esposito and Craigie, 1998; Heuer and Brown, 1997; Jenkins et al., 1997). Des expériences de pontage indiquent que :

- des résidus du domaine C-terminal (région 247-270) peuvent interagir avec T7 du brin clivé (Esposito and Craigie, 1998).
- le résidu E246 peut être localisé près de l'atome C2' d'A7 du brin non clivé (Gao et al., 2001).

Ces résultats expliquent pourquoi des mutations du domaine C-terminal affectent la réaction de 3'-processing.

Par ailleurs, un rôle important du domaine central dans la liaison de l'ADN cible a été montré. Des IN chimériques dans lesquelles le domaine central a été remplacé par celui d'une autre IN rétrovirale, ont été construites; les domaines terminaux sont restés inchangés (Katzman and



Sudol, 1995; Shibagaki et al., 1997). Dans tous les cas, le modèle d'intégration obtenu correspond au modèle attendu avec l'IN dont le domaine central est présent dans la protéine chimérique. Par exemple, une protéine chimérique avec le domaine central de l'IN du VIH-1 et les domaines terminaux de l'IN de FIV présente un modèle d'intégration du VIH-1 et inversement. De plus, il a été montré que le domaine central interagit avec l'ADN cible flanquant le site d'intégration (Gerton and Brown, 1997). Le résidu S119 pourrait être impliqué dans la liaison de l'ADN cible (Harper et al., 2001).

En ce qui concerne le domaine N-terminal, des études de pontage suggèrent qu'il pourrait être tout près de l'ADN cible, au niveau du site d'intégration (Heuer and Brown, 1997). Des expériences de complémentation fonctionnelle avec les IN du VIH-1 et du virus Visna (Engelman et al., 1993) qui partagent environ 31% d'identité de séquences, suggèrent que le domaine N-terminal n'est pas impliqué dans la détermination de la spécificité de substrats pour le 3'-processing et le transfert de brins (Katzman and Sudol, 1995). En particulier, les 26 premiers résidus des IN du VIH-1 et de RSV qui comprennent les deux histidines du motif HHCC, ne sont pas requis pour la liaison de l'ADN (Mazumder et al., 1996b; Mumm and Grandgenett, 1991).

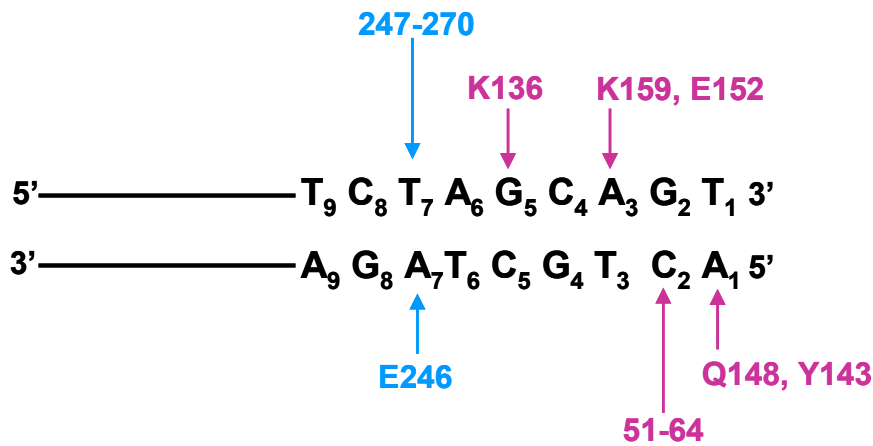
Les données de complémentation fonctionnelle et de pontage indiquent aussi que le domaine N-terminal agit en trans du domaine central contenant un site actif fonctionnel (Ellison et al., 1995; Engelman et al., 1993; Heuer and Brown, 1998; Jonsson et al., 1996) tandis que le domaine C-terminal peut être en cis (Engelman et al., 1993) ou en trans (Engelman et al., 1993; Gao et al., 2001; Heuer and Brown, 1998; van Gent et al., 1993).

Des expériences d'empreintes protéiques (Dirac and Kjems, 2001) montrent que la sensibilité à l'attaque protéolytique est affectée par l'ADN, au niveau des résidus :

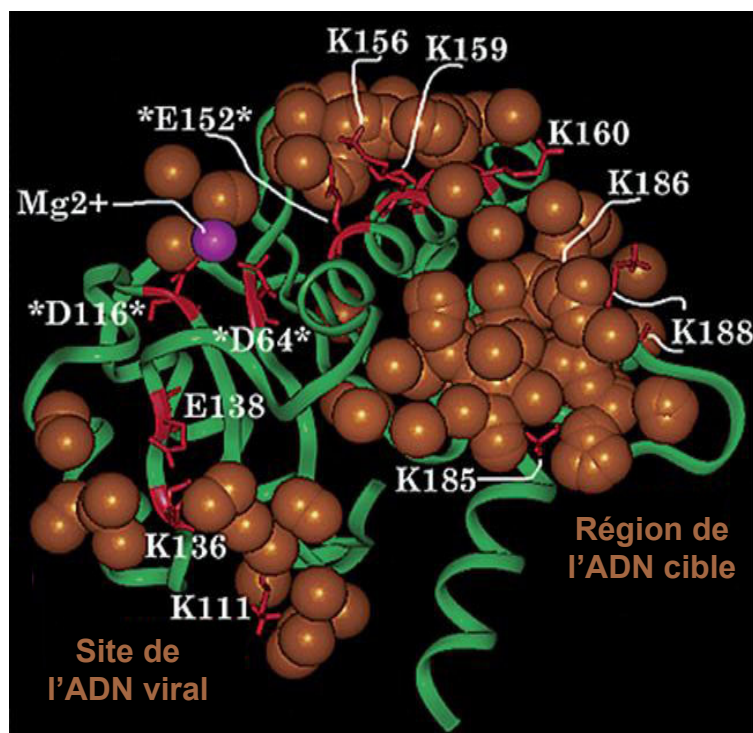
- K111, K136, E138, K156, K159, K160, K185, K186 et K188 du domaine central
- K215, E246, K258 et K273 du domaine C-terminal
- K34 du domaine N-terminal.

L'ensemble de ces travaux met en évidence que l'interaction de l'IN avec ses substrats ne se fait pas selon un modèle dans lequel des domaines individuels sont initialement responsables de la liaison de l'ADN viral ou cible. Il semblerait plutôt qu'un réseau complexe d'interaction intervienne. Néanmoins, d'après ces études, le domaine central serait responsable de la reconnaissance séquence-spécifique de l'ADN viral et serait impliqué dans la liaison de l'ADN cible. Le domaine C-terminal, quant à lui, serait responsable de la liaison de nucléotides situés à plus de 6 paires de bases de l'extrémité LTR de l'ADN viral. L'interaction du domaine C-terminal avec ces nucléotides pourrait avoir un rôle dans





**Figure 22-** Les contacts IN/ADN-viral mis en évidence par pontage. Les résidus ou portions du domaine central sont colorés en violet et ceux du domaine C-terminal en bleu.



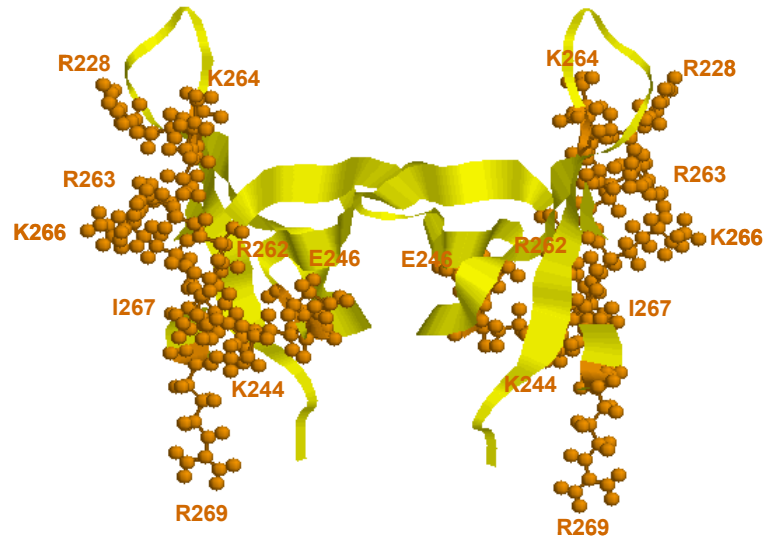
**Figure 23-** Sites de liaison de l'ADN viral et cible sur le domaine central de l'IN du VIH-1 (code PDB : 1QS4, chaîne A) (Goldgur *et al.*, 1999) proposés par les études de docking de dinucléotides (Perryman and McCammon, 2002). Le domaine central est représenté en modèle ruban et coloré en vert. Les positions des phosphates des dinucléotides dockés sont montrées par des sphères colorées en marron. Les résidus qui ont été identifiés comme étant protégés par l'ADN de l'attaque protéolytique (Dirac and Kjemis, 2001) sont colorés en rouge et leurs noms sont mentionnés. Les résidus de la triade catalytique sont également montrés en rouge et leurs noms sont entourés d'un astérisque.

l'orientation et le positionnement correct de l'ADN viral afin que les extrémités LTR puissent être reconnues par les résidus du domaine central et que la catalyse se produise. En effet, l'IN dépourvue du domaine C-terminal ne peut pas catalyser les réactions de 3'-processing et de transfert de brins ; cela est certainement dû au fait qu'elle ne puisse pas positionner et orienter correctement les extrémités LTR au site actif. Par contre, lorsque l'ADN est correctement pré-positionné comme dans un substrat de désintégration, la catalyse se produit sans les domaines terminaux. Les contacts entre l'ADN viral et le domaine central et C-terminal de l'IN décrits au-dessus sont résumés dans la **figure 22**.

### **B - Exploration des sites de liaison de l'ADN sur l'IN par docking de dinucléotides**

Les sites de liaisons possibles de l'ADN ont été explorés par docking de dinucléotides entièrement flexibles, sur le domaine central monomérique (code PDB : 1QS4, chaîne A) (Perryman and McCammon, 2002) et sur le domaine C-terminal dimérique de l'IN du VIH-1 (code PDB : 1QMC) (Zhu et al., 2005).

La majorité des dinucléotides dockés sur le domaine central de l'IN (Perryman and McCammon, 2002) se trouvent près de résidus qui ont été identifiés comme étant protégés par l'ADN de l'attaque protéolytique (Dirac and Kjems, 2001). Beaucoup d'entre eux sont même en contact avec ces résidus, supposés être impliqués dans la liaison de l'ADN. Ainsi, les résultats de docking sur le domaine central sont en accord avec les études d'empreinte protéique réalisées sur l'IN entière. Les dinucléotides dockés sont répartis dans deux régions du domaine central. La **figure 23** montre les positions des phosphates des dinucléotides dockés sur le domaine central. La région de droite dans la **figure 23** présente un grand nombre de sites différents mais chevauchants qui peuvent interagir favorablement avec des groupes phosphates. Cette région peut donc être impliquée dans la liaison de phosphates ayant de nombreuses positions différentes. Cela suggère que différents ADN doubles brins devraient être capables d'interagir dans cette région. Les auteurs ont alors proposé que cette région soit plutôt impliquée dans les interactions avec l'ADN non guidées par la spécificité de séquence et donc avec l'ADN cible. La seconde région, à gauche dans la **figure 23**, présente un nombre de sites pouvant interagir avec des groupes phosphates beaucoup plus réduits. Ainsi, moins de séquences d'ADN devraient vraisemblablement être capables de se déformer convenablement afin d'interagir avec cette région. La distribution des phosphates des dinucléotides, dockés dans cette région, est similaire à l'organisation des positions des phosphates d'un ADN double brin de forme B. Les auteurs ont alors proposé que cette région



**Figure 24-** Sites de liaison de l'ADN viral sur le dimère de domaines C-terminaux (code PDB : 1QMC) (Eijkelenboom et al., 1999) proposés par les études de docking de dinucléotides (Zhu et al., 2005). Le dimère est représenté en modèle ruban et coloré en jaune. Les résidus impliqués dans les liaisons hydrogènes avec les dinucléotides sont représentés en boules-bâtons et colorés en orange. Ces résidus montrent la localisation des sites de liaison de l'ADN viral à la surface externe du dimère.

**Tableau 8-** Les modèles du complexe IN/ADN

Modèles (références)	IN*	ADN (oligonucléotides de quelques pb)	Bications dans le site actif	Disponibilités des coordonnées
(Heuer <i>et al.</i> , 1998)	Tétramère du bidomaine (DC + DCT)	2 ADN viraux U5 ADN cible	Non	Oui, dans le matériel supplémentaire
(Chen <i>et al.</i> , 2000a)	Dimère du bidomaine (DC + DCT)	1 ADN viral U5	Non	Non
(Gao <i>et al.</i> , 2001)	Tétramère de l'IN entière	2 ADN viraux U5 ADN cible	Non	Oui, sur demande
(Wang <i>et al.</i> , 2001)	Tétramère de l'IN entière	Séquence d'ADN différente (ADN substrat de la Transposase Tn5)	Non	Non
(Podtelezhnikov <i>et al.</i> , 2003)	Tétramère de l'IN entière	2 ADN viraux U5 ADN cible	Non	Oui, sur demande
(Adesokan <i>et al.</i> , 2004)	DC	1 ADN viral U5	2 Mg <sup>2+</sup>	Non
(De Luca <i>et al.</i> , 2003; De Luca <i>et al.</i> , 2005)	Dimère de l'IN entière	1 ADN viral U5	2 Mg <sup>2+</sup>	Oui, dans la PDB, code 1WKN
(Karki <i>et al.</i> , 2004)	Tétramère de l'IN entière	1 ADN viral U5 et 1 ADN viral U3 ADN cible	1 Mg <sup>2+</sup> et 2 Mg <sup>2+</sup>	Non
(Wang <i>et al.</i> , 2005)	Tétramère de l'IN entière	1 ADN viral U5	1 Mg <sup>2+</sup> et 2 Mg <sup>2+</sup>	Non
(Wielens <i>et al.</i> , 2005)	Tétramère de l'IN entière	1 ADN viral U5 et 1 ADN viral U3 ADN cible	1 Mg <sup>2+</sup>	Oui, dans la PDB, code 1ZA9
(Chen <i>et al.</i> , 2006)	Tétramère de l'IN entière	2 ADN viraux U5	1 Mg <sup>2+</sup>	Bientôt dans la PDB, code 2G3L

\* Les sigles DC et DCT signifient Domaine Central et Domaine C-Terminal.

devrait plutôt être impliquée dans des interactions séquence-spécifique et donc avec l'extrémité LTR de l'ADN viral.

Concernant le dimère C-terminal, les résultats de docking montrent une distribution symétrique des dinucléotides sur les surfaces externes du dimère (Zhu et al., 2005). La distribution des atomes de phosphates des dinucléotides, à la surface périphérique des deux domaines, ressemble quelque peu à l'organisation des phosphates au sein d'un ADN double brin. Le site de liaison de l'ADN ne serait donc pas localisé dans la gouttière entre les deux domaines C-terminaux, comme il a été suggéré par d'autres auteurs (Eijkelenboom et al., 1999). La liaison de l'ADN aurait lieu au niveau des surfaces périphériques du dimère. Un tel modèle de liaison est soutenu par d'autres travaux (De Luca et al., 2003; Dirac and Kjems, 2001). L'orientation suggérée pour l'ADN est compatible avec le modèle obtenu par docking de l'ADN viral sur l'IN entière construite par modélisation (voir partie sur les modèles IN/ADN) (De Luca et al., 2003). Les dinucléotides forment des liaisons hydrogènes avec les résidus du domaine C-terminal : R228, K244, E246, R262, R263, K264, K266, I267 et R269 (**figure 24**). La plupart de ces résidus exposent des charges positives à la surface, fournissant une région de liaison favorable aux dinucléotides chargés négativement. Plusieurs de ces résidus ont été mis en évidence pour être impliqués dans la liaison de l'ADN : E246, R262, R263 et K264 (Lutzke and Plasterk, 1998; Lutzke et al., 1994). Il a aussi été montré qu'E246 et des résidus de la région 247 à 270 peuvent interagir avec l'ADN viral, comme nous l'avons vu précédemment (Esposito and Craigie, 1998; Gao et al., 2001). Ainsi, les deux régions situées à la surface périphérique du dimère seraient des sites de liaison de l'ADN viral.

### **C - Modèles du complexe IN/ADN**

Les seules représentations de l'interaction IN/ADN actuellement disponibles sont des modèles (**tableau 8**) qui proposent principalement :

- des assemblages des trois domaines de l'IN pour former une protéine entière
- des arrangements de plusieurs monomères entre eux pour former un dimère ou un tétramère
- des modes de liaison de l'ADN viral et de l'ADN cible

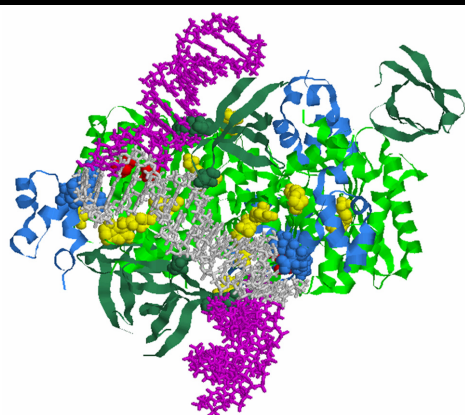
Les différentes méthodes utilisées pour construire ces modèles se basent généralement sur :

- les structures cristallographiques et RMN des différents fragments de l'IN
- différentes données biochimiques disponibles (données de pontage, d'empreinte protéique, d'anisotropie de fluorescence résolue en temps ...).

**Tableau 10-** Modèle du complexe IN(tétramère)/ADN(viral et cible)

Modèle	IN	ADN	Construction du complexe IN/ADN	
			Données utilisées	Les étapes de la construction
(Podtelezhnikov <i>et al.</i> , 2003)	Tétramère  Pas de Mg <sup>2+</sup> au site actif	2 ADN viraux U5 (13 pb) clivé Forme B  ADN cible (17 pb)	<p><b>1-</b> Données d'anisotropie de fluorescence résolue en temps (TFA) de l'IN et ses oligomères (Deprez <i>et al.</i>, 2000; Leh <i>et al.</i>, 2000)</p> <p><b>2-</b> Structures de domaines individuels <b>DC*</b>: 1BIS (chaîne B) <b>DCT*</b>: 1IHV (1 chaîne) <b>DNT*</b>: 1WJA (1 chaîne)</p> <p><b>3-</b> Données de pontage, d'empreinte protéique et autres données biochimiques</p> <p>Programmes : HYDROPRO, InsightII</p>	<p><b>1-Monomère</b> - Arrangement des trois domaines entre eux afin qu'ils soient en accord avec les données de TFA - Affinement de l'orientation des domaines en fonction de données de pontage et d'empreinte protéique. → Monomère compact (Distances DNT-DC=16 Å, DC-DCT=20 Å) différent de la structure 1EX4 (DC-DCT=42 Å) et des modèles construits à partir de cette structure.</p> <p><b>2- Dimère</b> - Superposition de 2 modèles monomériques avec la structure cristallographique de DC. → Dimère en accord avec les valeurs de TFA (Distance DC-DC= 45 Å)</p> <p><b>3- Tétramère</b> Arrangement de 2 dimères entre eux et placement des substrats de telle sorte que : - le transfert de brins puisse se faire sans perturber l'ADN cible → Distance DC-DC ~20-25 Å - la symétrie de la réaction concertée soit respectée → axe de symétrie d'ordre 2 du tétramère - le modèle soit en accord avec des données de pontage → Tétramère satisfaisant les valeurs de TFA</p>

**Modèle du complexe IN(tétramère)/ADN (viral et cible) (Podtelezhnikov *et al.*, 2003)**



Le tétramère est représenté en modèle ruban. Les DC sont colorés en vert clair. Les résidus D64 et D116 sont montrés en modèle compact et colorés en rouge. Les DCT sont colorés en vert foncé. Les résidus S230 et R263 qui sont respectivement en contact direct avec l'ADN cible et l'ADN viral sont représentés en modèle compact. Les DNT sont colorés en bleu. Les résidus F1, L2 et D3 qui sont en contact direct avec l'ADN cible, sont représentés en modèle compact. Les fragments d'ADN cible et viraux sont respectivement montrés en gris et violet. Les résidus montrés en modèle compact et colorés en jaune sont les résidus qui ont été décrits comme étant en contact avec l'ADN viral d'après des études de pontage.

- Les 2 DC actifs interagissent avec l'ADN viral et l'ADN cible au niveau des sites d'intégration. Les résidus catalytiques sont près des sites d'intégration. D64 et D116 sont proches d'A3 de l'ADN viral. Par contre E152 est en contact direct avec l'ADN cible.
- 2DCT se lie à l'ADN cible entre les sites d'intégration (contact direct de S230). 1 DCT interagit avec chaque ADN viral (contact direct R263-T6)
- 2 DNT se lie à l'ADN cible en dehors du segment compris entre les deux sites d'intégration. Les pb 2 et 3 de l'ADN cible, en aval du site d'intégration, sont près des résidus (1-12) (contacts directs de F1, L2 et D3). Les DNT n'interagissent pas avec l'ADN viral.
- L'interface entre les deux dimères implique des interactions entre 2 DNT de monomères inactifs. La plus grande aire de surface de l'interface tétramérique est formée par des interactions entre les DC. On peut noter que F185 d'un DC actif est près de la région (164-168) du DC actif opposé.

**Comparaison avec des données expérimentales**

Contacts IN-ADN brin clivé (études de pontage) (résidus en jaune)	Modèle	Contacts IN-ADN brin non clivé (études de pontage) (résidus en jaune)	Modèle
K159-A3	Oui	Q148-A1 ou Q148-C2	Non
E152-A3	Non	Y143-A1	Non
K136-G5	Non	(51-64)-C2	Non
(247-270)-T7	Non (T6)	E246-A7	Oui
ADN viral se lie en trans**	Oui	5 pb entre les deux sites d'intégration	Non

\* Les sigles DC, DCT et DNT signifient : Domaine Central, Domaine C-Terminal et Domaine N-Terminal

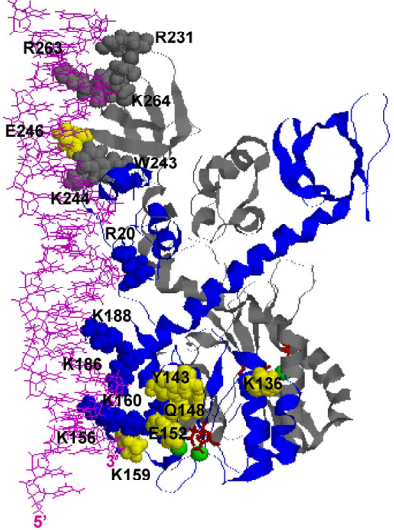
\*\*Les résidus soulignés ont été décrits comme étant impliqués dans la liaison de l'ADN (études expérimentales).

\*\*\* L'ADN viral se lie avec le DC d'un monomère actif et le DCT d'un monomère inactif.

Différents types de docking, des simulations de dynamique moléculaire, l'homologie avec la structure cristallographique Tn5 transposase/transposon ont été également utilisés pour générer des modèles.

Deux des modèles du tableau 8, un dimère et un tétramère sont présentés ci-dessous (tableaux 9 et 10). La méthode utilisée pour construire le modèle est décrite et une comparaison avec certains résultats expérimentaux est montrée.

**Tableau 9-** Modèle du complexe IN (dimère)/ADN (viral)

Modèle	IN	ADN	Construction de l'IN dimérique	Complexe IN/ADN
(De Luca <i>et al.</i> , 2003; De Luca <i>et al.</i> , 2005)	Dimère 2 Mg <sup>2+</sup> au site actif	ADN viral U5 (27 pb) clivé Forme B	<b>DC*</b> : 1QS4 (chaîne A et B) avec 1 Mg <sup>2+</sup> -région (141-144) construite par homologie avec 1BIS - Mutations F158K et W131E remplacées par résidus sauvages - Ajout d'un second Mg <sup>2+</sup> dans la même position relative que le second bication dans l'IN d'ASV (VSH) <b>DCT*</b> : 1EX4 - Superposition des DC de 1EX4 avec les DC du modèle - Liaison des DC avec les DCT (résidus 209 et 210) <b>DNT*</b> : 1K6Y - région (47-55) construite par homologie avec 1WJD - Superposition des DC du dimère AB avec les DC du modèle Des minimisations d'énergie aux différentes étapes de la construction	Docking Programme ESCHER modifié pour réaliser du docking protéine/ADN
<b>1WKN</b>				
<b>Modèle du complexe IN(dimère)/ADN(viral) après Docking (De Luca <i>et al.</i>, 2003)</b>				
			<p>Le dimère est représenté en modèle ruban avec les monomères A et B colorés en gris et bleu, respectivement. Les ions Mg<sup>2+</sup> sont montrés par des sphères vertes. Les résidus D64 et D116 sont colorés en rouge. L'ADN est colorée en violet. Les résidus impliqués dans les contacts IN-ADN sont représentés en modèle compact et ont la couleur de la sous-unité à laquelle ils appartiennent. Les résidus montrés en modèle compact et colorés en jaune sont les résidus qui ont été décrits comme étant en contact avec l'ADN viral, d'après des études de pontage.</p>	
Résidus impliqués dans des contacts directs IN-ADN **	DC (chaîne B)	K156, K159, K160, K186, K188 (bleu, modèle compact)		
	DCT (chaîne A)	S230, R231, W243, K244, R263, K264 (gris, modèle compact)		
	DNT (chaîne B)	R20 (bleu, modèle compact)		
Interactions hydrophobes	F1,L2,W19,A21,M22,A23,F26,P30,M50,M154,I191 (chaîne B)	W243, P261 (chaîne A)		
<b>Comparaison avec des données expérimentales</b>				
Contacts IN-ADN brin clivé (études de pontage) (résidus en jaune)	Modèle	Contacts IN-ADN brin non clivé (études de pontage) (résidus en jaune)	Modèle	
K159-A3	Oui	Q148-A1 ou Q148-C2	Non Non	
E152-A3	Non	Y143-A1	Non	
K136-G5	Non	(51-64)-C2	Non	
(247-270)-T7	Non	E246-A7	Non	
ADN viral se lie en trans***	Oui			
<b>Dynamique Moléculaire du complexe IN/ADN (De Luca <i>et al.</i>, 2005)</b>				
Le monomère B présente une transition conformationnelle avec un haut degré de flexibilité durant la simulation. Y143, en particulier, s'oriente complètement vers les résidus du site actif. Un tel changement conformationnel n'est pas observé dans le monomère A dont le DC n'est pas en contact avec l'ADN viral.				

\* Les sigles DC, DCT et DNT signifient : Domaine Central, Domaine C-Terminal et Domaine N-Terminal

\*\* les résidus soulignés ont été décrits comme étant impliqués dans la liaison de l'ADN (études expérimentales).

\*\*\* L'ADN viral se lie avec le DC du monomère actif et le DCT du monomère inactif.



Les deux modèles sont en accord, seulement avec certains résultats expérimentaux. Il en est de même pour tous les autres modèles. Les méthodes utilisées ne permettent pas d'accéder à des changements de conformation dus à une adaptation de l'IN et de l'ADN. Les fragments d'ADN sont rigides et de forme B ou A régulière alors que généralement la liaison de la protéine induit ou accentue des déformations de la double hélice. Ces modèles ne tiennent donc pas compte de la flexibilité des molécules qui est essentielle à la formation des complexes protéines/ADN.

### **III- Modèles du mécanisme d'action de l'intégrase**

Le processus d'intégration est unique, lent et relativement peu efficace. *In vitro*, les réactions catalysées par l'IN se font avec un temps de demi-vie apparent de l'ordre de l'heure. Il est donc probable que toute une série d'événements se produise, comme l'ajustement de l'ADN à l'IN, avant les événements purement chimiques, catalytiques que nous allons décrire ici.

#### **A - Les trois réactions sont de type SN2 et peuvent se réaliser au même endroit**

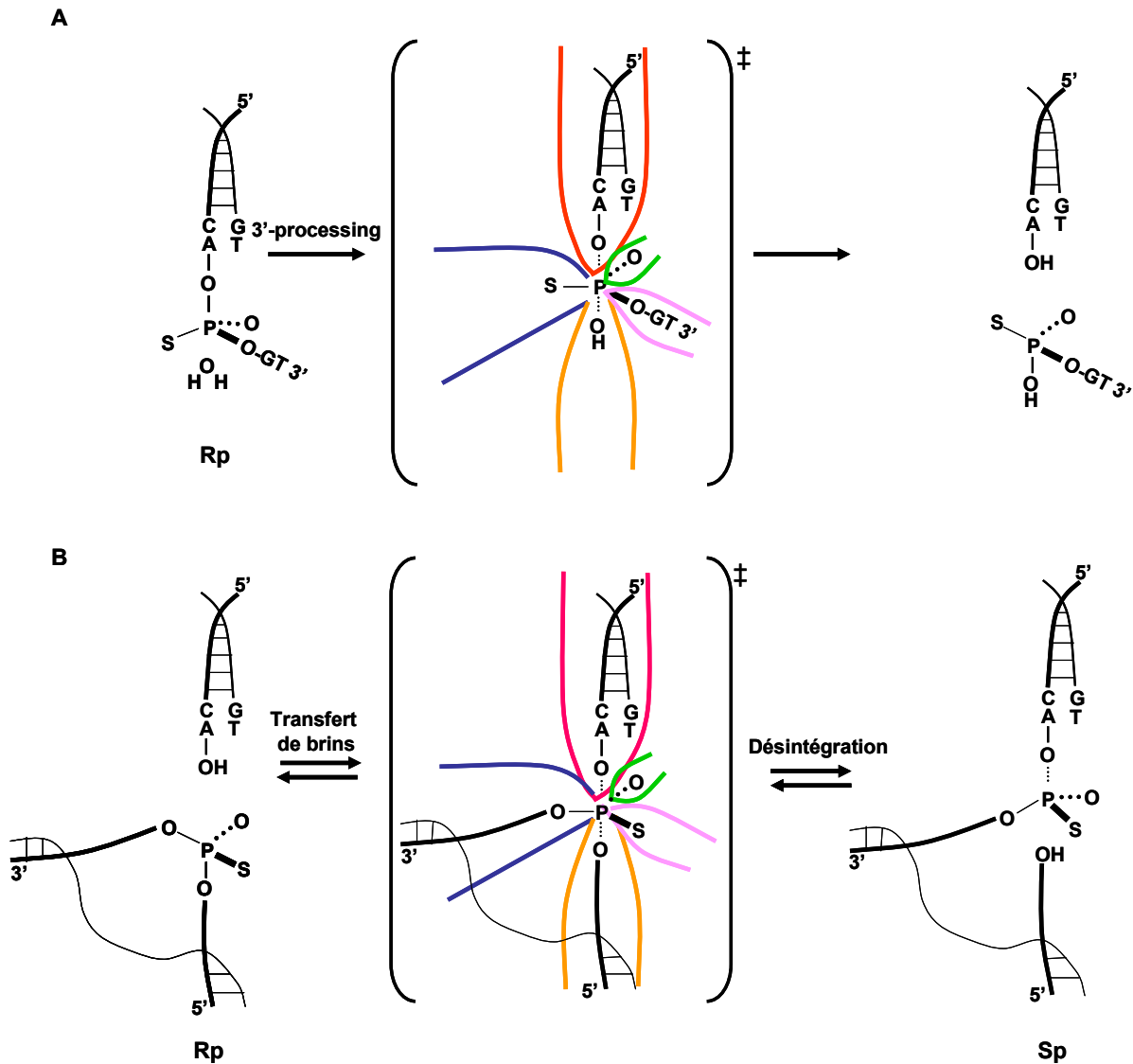
Gerton et al. (1999) ont montré que les trois réactions catalysées par IN sont des substitutions nucléophiles SN2 avec inversion optique du phosphate. D'autre part, ils proposent un modèle macroscopique où les trois réactions peuvent se réaliser au même endroit sur la protéine.

Le remplacement du phosphate par un phosphorothioate de forme énantiomère Rp ou Sp permet de distinguer le rôle de chacun des deux oxygènes libres du phosphate de la guanosine 2, O1P et O2P, dans les réactions de 3'-processing, de transfert de brins et de désintégration (Gerton et al., 1999). Les tests *in vitro*, en présence de  $Mn^{2+}$ , montrent que l'IN du VIH-1 forme des produits Sp, à partir de substrats Rp, au cours des réactions de 3'-processing et de transfert de brins et l'inverse, au cours de la réaction de désintégration. Chacune de ces réactions se produit probablement en une seule étape de transestérification. Par ailleurs, l'IN est inactive dans le 3'-processing d'un substrat Sp.

Gerton et al. (1999) ont proposé un modèle, en partant des trois suppositions suivantes:

1) L'IN utilise un seul site actif pour catalyser les trois réactions, comme suggéré par les études structurales et de mutagenèse (Bujacz et al., 1996a; Chen et al., 2000a; Drelich et al., 1992; Dyda et al., 1994; Engelman and Craigie, 1992; Goldgur et al., 1999; Goldgur et al., 1998; Greenwald et al., 1999; Kulkosky et al., 1992; Leavitt et al., 1993; Maignan et al., 1998; van Gent et al., 1992; Wang et al., 2001).





**Figure 25-** Modèle de l'arrangement des substrats dans le site actif de l'IN du VIH-1, au cours des réactions de 3'-processing (**A**), de transfert de brins (**B**) et de désintégration (**B**) (Gerton et al., 1999). La paire CA/TG de l'oligonucléotide viral est liée au même site, au cours des trois réactions. Ce site est en position apicale dans l'état de transition (coloré en rouge). L'autre position apicale dans l'état de transition (colorée en orange) est occupée par une molécule d'eau au cours de la réaction de 3'-processing et par le 3'-OH de l'oligonucléotide cible agissant respectivement comme le nucléophile ou le groupe partant, au cours des réactions de désintégration et de transfert de brins. Une position équatoriale dans l'état de transition est définie par un site qui lie l'oxygène non-estérifiée (coloré en vert) mais qui ne peut pas loger un atome de soufre. Une seconde position équatoriale (colorée en rose) est définie par un site de liaison qui fournit des interactions favorables avec l'extrémité 3'-GT de l'oligonucléotide viral durant le 3'-processing et qui peut loger l'atome de soufre du phosphorothioate durant le transfert de brins et la désintégration. La troisième position équatoriale (colorée en bleu) est définie par un site de liaison qui peut loger la partie de l'oligonucléotide cible s'étendant à partir de l'extrémité 3' durant le transfert de brins et la désintégration et un atome de soufre durant le 3'-processing. La stéréoisomérie (Rp ou Sp) des substrats de départ de chaque réaction est notée.

2) Certaines interactions spécifiques dans le site actif sont les mêmes pour les trois réactions. La paire CA/TG de l'oligonucléotide viral est cruciale dans les trois cas (Bushman and Craigie, 1991; Chow et al., 1992; Gerton and Brown, 1997; LaFemina et al., 1991; Leavitt et al., 1992; Sherman et al., 1992; Sherman and Fyfe, 1990; van den Ent et al., 1994). Il a alors été supposé que cette paire de dinucléotides est liée au même site spécifique, au cours de ces réactions (coloré en rouge dans la **figure 25**).

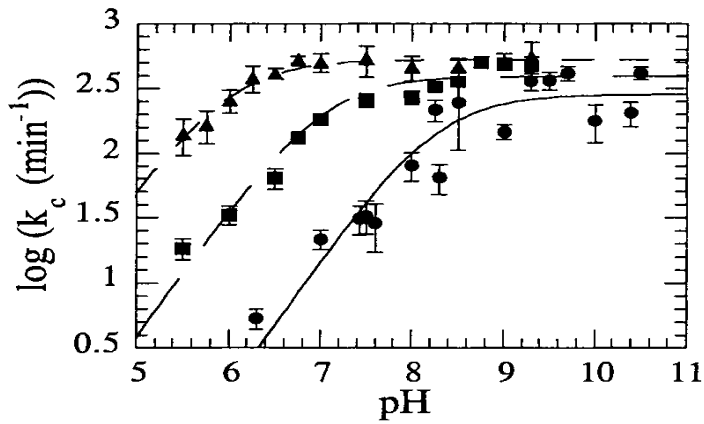
3) Dans l'état de transition, l'atome de phosphore adopte une géométrie bipyramidale trigonale et est lié à 5 atomes (Dennis and Westheimer, 1966; Knowles, 1980) : le nucléophile et le groupe partant occupent les positions apicales et les trois autres substituants sont dans les positions équatoriales. Le 3'-OH de l'adénosine de l'oligonucléotide viral (coloré en rouge dans la **figure 25**) étant soit le groupe partant soit le groupe attaquant, au cours des réactions catalytiques, il a été mis en position apicale pour les trois réactions. L'autre groupe partant ou attaquant de chaque réaction a alors été positionné au site apical opposé (coloré en orange dans la **figure 25**). Ce site est donc occupé par une molécule d'eau, au cours de la réaction de 3'-processing et par le 3'-OH de l'oligonucléotide cible agissant respectivement comme le nucléophile ou le groupe partant, au cours des réactions de désintégration et de transfert de brins.

Après le placement des groupes en positions apicales, les occupants des sites équatoriaux ont été positionnés. La stéréospécificité de l'IN pour les trois réactions catalytiques suggère que dans chacune des réactions, l'IN interagisse avec un seul des atomes d'oxygène non-estérifié du phosphate. Dans l'énantiomère Rp, le soufre remplace l'atome O2P. Gerton et al. (1999) invoquent alors un argument d'encombrement stérique. Si Rp est un bon substrat, cela signifie que l'augmentation du rayon d'O2P en S ne gêne pas la réaction. Donc l'IN n'interagit pas avec O2P mais seulement avec O1P (vert dans la **figure 25**). Une fois les positions apicales et une des trois positions équatoriales spécifiées, l'arrangement des substituants restants a été dicté par la stéréospécificité observée dans les trois réactions (**figure 25**).

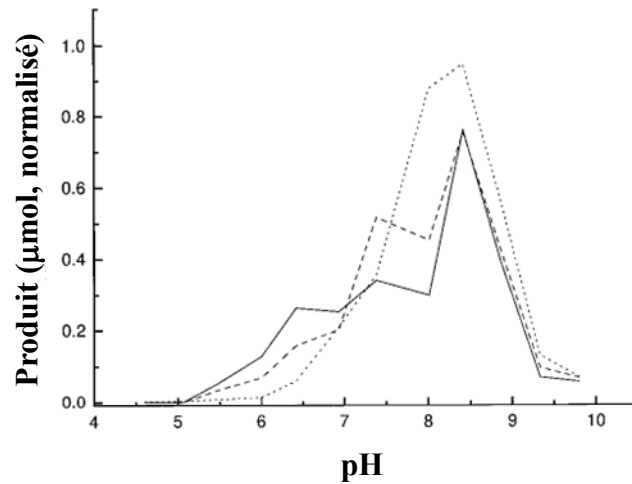
De façon surprenante, le modèle implique que le site occupé par le dinucléotide GT de l'extrémité 3' de l'oligonucléotide viral (coloré en rose dans la **figure 25**), au cours du 3'-processing, soit différent du site occupé par la partie de l'oligonucléotide cible s'étendant à partir de l'extrémité 3' (coloré en bleu dans la **figure 25**) au cours du transfert de brins ou de la désintégration. Il est donc possible que le site qui lie le dinucléotide GT 3'-terminal dans la réaction de 3'-processing, interagisse favorablement avec un ADN simple brin, dans une configuration que peut seulement adopter un ADN dont les paires de bases ne sont pas appariées. Ainsi, si un ADN en double hélice ne peut pas se loger dans ce site mais peut être

**Tableau 11- pK1**  
de quelques ions

Ion	pK1
Cu <sup>2+</sup>	8.00
La <sup>3+</sup>	9.06
Co <sup>2+</sup>	9.65
Mn <sup>2+</sup>	10.46
Mg <sup>2+</sup>	11.40



**Figure 26-** La vitesse de réaction en fonction du pH pour la T5 FlapEndonuclease (triangles  $\text{Co}^{2+}$ , carrés  $\text{Mn}^{2+}$ , cercles  $\text{Mg}^{2+}$ ) (Tock et al., 2003)



**Figure 27-** Activités endonucléasiques de l'IN d'ASV entière (-2 continu, -3 tirets) et du domaine central de l'IN d'ASV (-3 points) en présence de 10mM  $\text{Mn}^{2+}$  en fonction du pH (Lubkowski *et al.*, 1998a).

positionné dans un autre site (coloré en bleu dans la **figure 25**) alors, la partie de l'oligonucléotide cible s'étendant à partir de l'extrémité 3' serait contrainte d'occuper cette position.

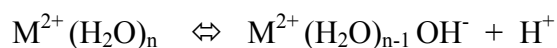
Macroscopiquement, la désintégration est l'inverse de la réaction de transfert de brins. De plus, la spécificité de substrat des deux réactions est homologue (Chow and Brown, 1994; Ellison et al., 1995; Gerton and Brown, 1997; Vincent et al., 1993). La préférence de l'IN pour des stéréoisomères opposés, lors des deux réactions, fournit un argument supplémentaire pour la relation de réciprocité entre ces deux réactions ainsi que pour un arrangement local identique de l'ADN viral et de l'ADN cible, dans le site actif.

Il est clair que ce modèle ignore les effets non purement stériques dus au remplacement d'un oxygène par un soufre. L'introduction du soufre modifie les charges coulombiennes à l'intérieur du phosphorothioate. L'introduction d'un phosphorothioate Sp dans un ADN a peu d'effet sur sa stabilité. Par contre, l'énantiomère Rp est déstabilisant (Jaroszewski et al., 1992). Or la déstabilisation de l'extrémité de l'ADN a été proposée comme favorisant le 3'-processing. D'autre part, l'énantiomère Rp élargit le petit sillon de l'ADN en forme B, ce qui peut favoriser les interactions entre l'ADN et la protéine. Enfin, Gerton et al. (1999) ont réalisé les réactions en présence de  $Mn^{2+}$ , un ion mou qui préfère interagir avec le soufre par ses électrons d.

## **B - Modèles atomiques**

Le véritable agent nucléophile est non pas une molécule d'eau mais l'ion  $OH^-$  dans le 3'-processing et l'extrémité 3' O- de l'adénine dans le transfert de brins.

Une des n molécules d'eau d'hydratation d'un bication  $M^{2+}$  peut former un ion hydroxyle et libérer un proton suivant la réaction :

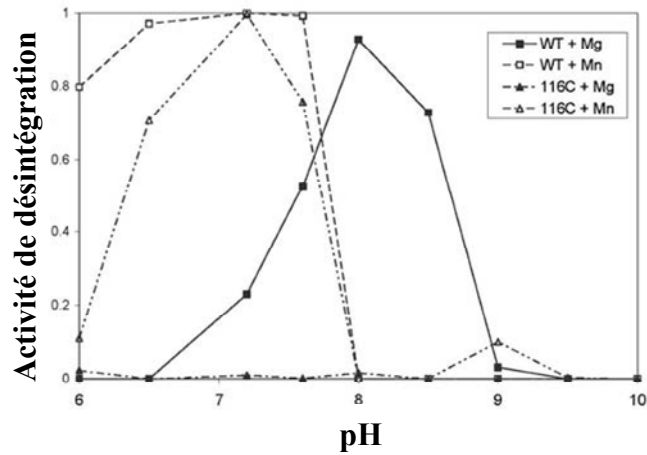


Le rapport de la forme hydroxylée sur la forme neutre suit la formule :

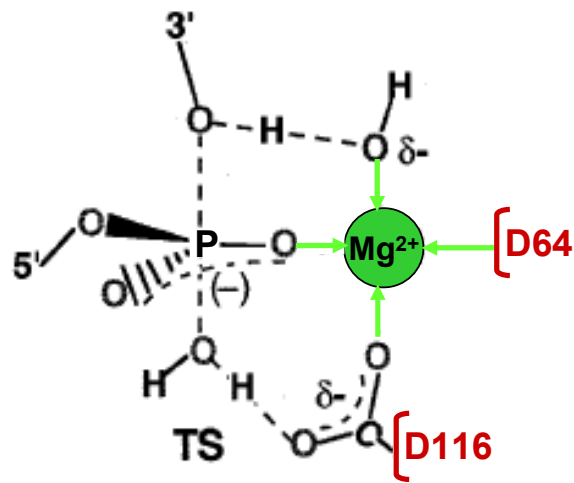
$$\log \left\{ \frac{[M^{2+}(H_2O)_{n-1}OH^-]}{[M^{2+}(H_2O)_n]} \right\} = pH - pK_1$$

La concentration de forme hydroxylée augmente avec le pH puis atteint un plateau (elle ne peut pas dépasser la concentration de bications). On voit sur **la figure 26** que la vitesse de coupure de la liaison phosphoester par la flapendonucléase dépend bien du pH et de la nature du bication (Tock et al., 2003), comme attendu suivant le **tableau 11**.

C'est sans doute ce qui se passe aussi pour l'IN d'ASV (**figure 27**). La chute de l'activité au-delà de 9 peut s'expliquer par la neutralisation des lysines et donc une baisse de l'affinité de



**Figure 28-** Activité de désintégration de l'IN du VIH-1 (Gao et al., 2004).



**Figure 29-** L'état de transition du mécanisme de la réaction de 3'-processing dans le site actif de l'IN du VIH-1 contenant un ion  $Mg^{2+}$  (Bernardi F. et al., 2002). Une molécule d'eau de la couche d'hydratation du  $Mg^{2+}$  fournit l'hydrogène pour former le H3T de l'extrémité 3'-OH du nucléotide partant. Une autre molécule d'eau n'appartenant pas à la couche d'hydratation, mais polarisée par le résidu D116, fournit l'OH attaquant.

l'ADN pour l'IN d'ASV. Mais le domaine d'activité de l'IN du VIH-1 ne s'étend pas aux pH basiques (**figure 28**), ce qui suggère que ce n'est pas la fixation de l'ADN qui est limitante. Il n'en reste pas moins qu'il existe un décalage d'environ une unité dans le pH optimum, en présence de  $Mn^{2+}$  et  $Mg^{2+}$  (Gao et al. J. Virol., 2004). Ceci montre que le rôle du bication consiste probablement à favoriser la coupure d'une molécule d'eau. On remarque qu'en présence de  $Mn^{2+}$ , le maximum d'activité de l'IN d'ASV se situe vers 8.5 (Lubkowski et al., 1998a) alors que celui de l'IN du VIH avoisine 7 (Gao et al., 2004) et celui de l'IN de Moloney 6.2 (Jonsson et al., 1993).

L'analyse des séquences des IN d'ASV, du VIH et de Moloney nous suggère que la lysine K114 chez ASV ou l'H114 du VIH-1 pourrait aussi jouer un rôle dans le mécanisme catalytique en donnant un proton. Ce proton pourrait provenir du solvant dans le cas de l'IN de Moloney où le résidu équivalent est une glycine.

Le résidu H114 de l'IN du VIH-1 doit donc avoir un rôle dans la catalyse.

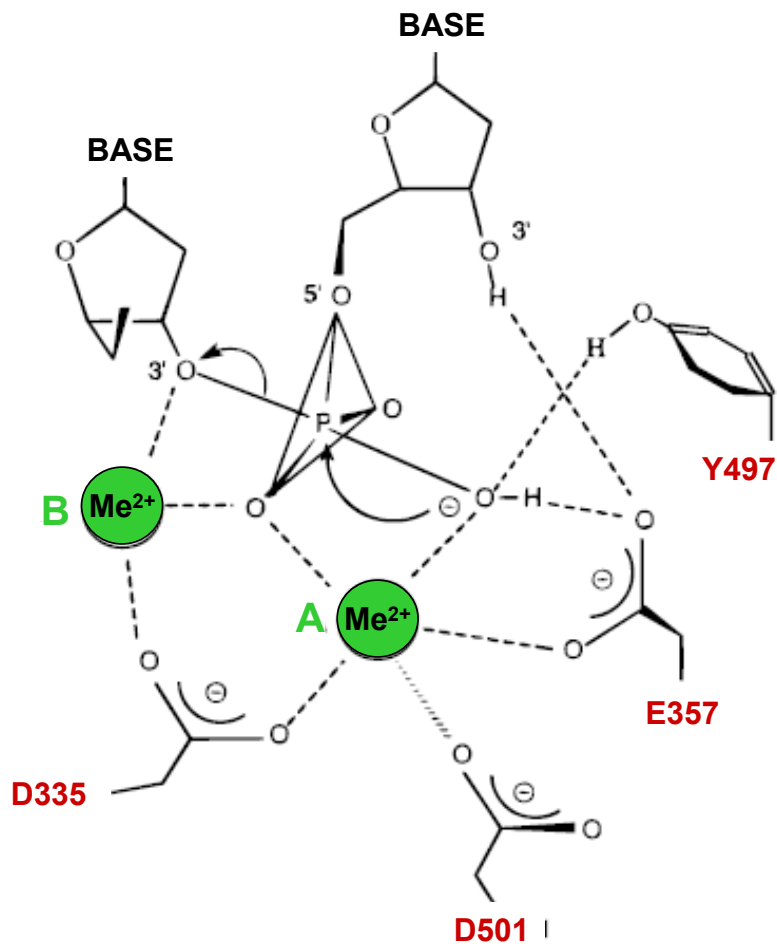
Van Gent et al. (1993) ont montré que la mutation H114S chez l'IN du VIH-2 ne modifie pas la fixation de l'ADN mais change la préférence du nucléophile attaquant (eau, glycérol, extrémité de l'ADN viral), en présence de  $Mn^{2+}$ .

Vink et al. (1994b) remarquent que la préférence du nucléophile chez FIV est différente de ce qu'elle est chez le VIH-1 ou le VIH-2. Ils attribuent ce fait au remplacement de Q115 chez FIV par H114 chez les VIH-1 et VIH-2.

Vers les pH acides, la chute de l'activité de l'IN peut être due à la faible ionisation de l'eau d'hydratation du bication mais aussi à la neutralisation des groupes carboxyliques du motif DDE. Si on remplace l'aspartate D116 (pK vers 4.4) par une cystéine (pK vers 8.5), la chute de l'activité vers les pH acides est plus rapide (**figure 28**). Cela suggère que D116 reçoit un proton.

Dans le mécanisme proposé par Bernardi et al. (2002), une molécule d'eau est polarisée par le groupement carboxyle de D116 et fournit l'ion  $OH^-$  attaquant (**figure 29**). Une autre molécule d'eau, faisant partie de la couche d'hydratation du bication  $Mg^{2+}$  unique, fournit le proton qui devient l'H3T de l'extrémité 3'OH de l'adénine. Des calculs théoriques au niveau B3LYP montrent que la barrière d'énergie de 70,3 kcal/mol s'abaisse à 33,5 kcal/mol, en présence de l'enzyme.

Un mécanisme impliquant deux bications a été proposé pour la réaction exonucléosique 3'-5', catalysée par le fragment de Klenow de l'ADN polymérase I d'Escherichia Coli (Beese and



**Figure 30-** L'état de transition du mécanisme enzymatique à deux ions métalliques divalents de la réaction exonucléasique 3'-5' catalysée par le fragment de Klenow de l'ADN polymérase I d'*Escherichia coli* (Beese and Steitz, 1991). L'ion A sert à activer la molécule d'eau fournissant le groupe OH<sup>-</sup> attaquant alors que les résidus Y497 et E357 servent à positionner correctement OH<sup>-</sup> vers l'atome de phosphore et la liaison P-O clivée. L'ion B stabilise l'angle de liaison O-P-O à 90° (l'angle entre l'oxygène apical et l'oxygène équatorial dans l'état de transition pentacoordonné). Les longueurs des liaisons P-O sont de 1,9 Å pour les oxygènes apicaux et de 1,5 Å pour les oxygènes méridiens.

Steitz, 1991). L'état de transition du mécanisme proposé est montré en **figure 30**. Un cation divalent (A) est coordonné à la protéine par les résidus D355, E357 et D501. Le phosphate de la liaison phosphodiester d'un dinucléotide qui est au centre de la réaction, interagit avec le cation divalent A, via son atome O1P. Le groupe OH<sup>-</sup> attaquant le phosphate interagit aussi avec le cation divalent A. Ainsi ce bication est pentacoordonné. Le groupe OH<sup>-</sup> attaquant forme également des liaisons hydrogènes avec E357 et Y497. Le bication A sert à activer la molécule d'eau fournissant le groupe OH<sup>-</sup> attaquant alors que les résidus Y497 et E357 servent à positionner correctement OH<sup>-</sup> vers l'atome de phosphore et la liaison P-O clivée. Dans la structure sauvage du fragment de Klenow contenant le produit de la réaction, un second bication (B) interagit avec l'oxygène O1P du phosphate et avec l'oxygène O3' de l'adénine. Il est également coordonné par D355 et trois molécules d'eau. La structure du complexe formé par l'oligonucléotide et le fragment de Klenow peut être déterminée seulement sur un complexe inactif. Les auteurs ont muté D424 en alanine. Le second cation divalent (B) est alors absent du cristal mais son absence n'empêche pas la fixation de l'ADN au fragment de Klenow. Dans le mécanisme de Beese et Steitz, une molécule d'eau est polarisée par le bication A et forme le groupe OH<sup>-</sup> attaquant. La liaison P-O3' est polarisée par le second bication B et facilement coupée. L'origine du proton du 3'-OH partant reste inconnue dans ce mécanisme.

Le mécanisme de Beese et Steitz (1991) est certainement plus efficace que celui de Bernardi et al. mais cela n'est pas un critère déterminant pour une enzyme à action unique et qui peut donc être « lente ». D'autre part, le résidu Y497 du fragment de Klenow interagit avec l'O2P du résidu partant. Le résidu Y143 de l'IN du VIH-1 ne joue certainement pas ce rôle mais interagit avec l'adénine de l'extrémité 5' du brin non clivé de l'ADN viral (Esposito and Craigie, 1998).





## Chapitre IV

### Les inhibiteurs de l'intégrase du VIH-1

---

L'intérêt thérapeutique du développement d'inhibiteurs de l'IN est évident depuis de nombreuses années, même avant la validation pharmacologique récente de l'IN (Hazuda et al., 2004b). En effet, l'IN est essentielle pour la réplication rétrovirale et l'absence d'une enzyme équivalente dans les cellules hôtes devrait amener à des inhibiteurs qui n'interféreraient pas avec les processus cellulaires normaux et qui auraient un index thérapeutique élevé. De plus, de nouveaux médicaments et de nouvelles cibles sont également nécessaires. Les inhibiteurs de l'IN pourraient devenir des additifs potentiels des HAART ou bien, une thérapie de secours pour des patients résistants aux médicaments anti-VIH actuellement disponibles.

Ces dernières années ayant été marquées par des avancées dans la recherche sur l'IN du VIH-1, le nombre de brevets concernant les inhibiteurs d'IN a fortement augmenté. Actuellement, quatre molécules sont en étude clinique.

L'identification et l'évaluation d'inhibiteurs de l'IN repose dans un premier temps sur des tests *in vitro* (voir chapitre II). Ensuite, il est vérifié que la molécule candidate issue de ces tests est active *ex vivo*, c'est-à-dire bloque la propagation du VIH-1 dans les cultures de cellules. Une fois l'activité antivirale démontrée, au moins quatre critères doivent être trouvés pour conclure que l'IN est bien la cible cellulaire de cette molécule.

1) Des expériences au cours desquelles la molécule candidate est ajoutée à différents temps après l'infection au VIH-1, doivent prouver l'efficacité de l'inhibiteur à un temps qui soit compatible avec la phase d'intégration (entre 4 et 16 heures après l'infection) (Pauwels et al., 1990).



2) Les cellules infectées, traitées avec l'inhibiteur, doivent montrer une accumulation de cercles à 2 LTR et une diminution de l'intégration du VIH-1, dans les chromosomes hôtes. Les cercles à 2 LTR résultent de l'accumulation d'ADN viral et de sa circularisation par des enzymes cellulaires.

3) Des mutations de l'IN doivent être trouvées dans les virus résistants à l'inhibiteur.

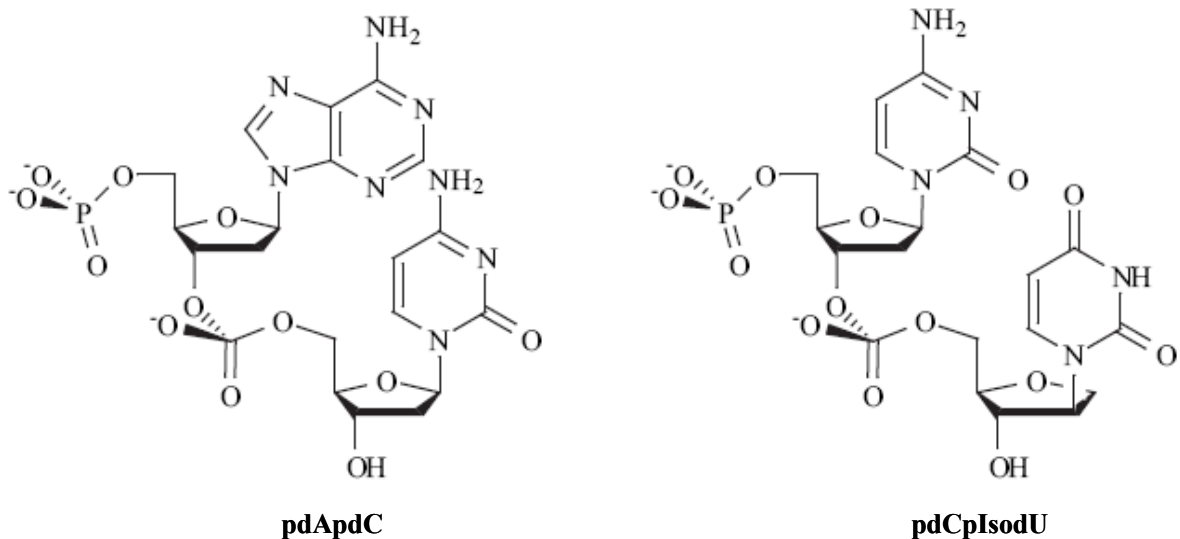
4) L'inhibiteur doit être inactif (ou beaucoup moins actif) dans des tests biochimiques contre les IN recombinantes comportant des mutations identifiées dans les virus résistants à cette molécule.

Ces dix dernières années, un grand nombre d'inhibiteurs de l'IN ont été décrits tels que des peptides, des dinucléotides, des oligonucléotides, des molécules naturelles et synthétiques appartenant à diverses classes structurales... Il s'agit ici de présenter quelques uns de ces composés afin d'avoir une vision générale mais non exhaustive de la recherche sur les inhibiteurs de l'IN. Nous insistons sur les progrès récents concernant la compréhension de la formation du complexe à l'IN et du mécanisme d'action. Les valeurs d'IC<sub>50</sub> (concentration de drogue nécessaire pour obtenir 50% d'inhibition), mentionnées dans cette partie, doivent être interprétées avec prudence car leur détermination dépend des conditions expérimentales.

## **I- Peptides**

Le premier inhibiteur peptidique de l'IN identifié est un hexamère (HCKFWW) (Puras Lutzke et al., 1995). Il agit sur l'IN avec une IC<sub>50</sub> d'environ 2 µM. La dimérisation de ce peptide, via un lien thioether, diminue d'au moins dix fois l'IC<sub>50</sub>. Plus récemment, en utilisant l'IN comme amorce et une bibliothèque génomique de levure comme source d'inhibiteurs, un peptide de 33-mer nommé I33 a été isolé (de Soultrait et al., 2002). Un dodécapeptide (YQLLIRMIYKNI) nommé EBR28, correspondant à la région N-terminale d'I33, inhibe le 3'-processing avec une IC<sub>50</sub> de 5 µM. Ce peptide adopte une conformation en hélice α avec des propriétés amphipathiques. Il interagit avec le domaine central de l'IN mais le mode de liaison n'est pas connu.

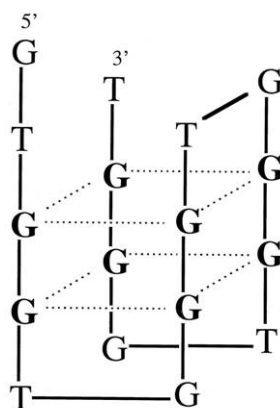
Une autre approche a consisté à synthétiser des peptides reproduisant la séquence en acides aminés de deux hélices α du domaine central de l'IN du VIH-1 : l'hélice α1 (INH1, résidus 93 à 107) et l'hélice α5 (INH5, résidus 167 à 187) (Maroun et al., 2001). Les hélices α1 et α5 forment l'interface dimérique entre deux sous-unités de domaines centraux. INH5 est le plus actif de ces deux peptides avec une valeur d'IC<sub>50</sub> de l'ordre de 0,085 µM pour le 3'-processing et de 0,06 µM pour le transfert de brins, en présence de Mg<sup>2+</sup>. Il interagit avec



**Figure 31-** Exemples de deux dinucléotides, naturel (à gauche) et synthétique (à droite), inhibiteurs de l'IN du VIH-1.

**Tableau 12-** Quelques inhibiteurs dinucléotidiques (naturels et synthétiques) de l'IN du VIH

	IC <sub>50</sub> (μM) 3'-processing	IC <sub>50</sub> (μM) Transfert de brins	Référence
pdApdC	6	3	(Mazumder <i>et al.</i> , 1997b)
pdApdT	7	7	(Mazumder <i>et al.</i> , 1997b)
pdCpdT	8	6	(Mazumder <i>et al.</i> , 1997b)
pdCpIsodU	7	6	(Taktakishvili <i>et al.</i> , 2001)
pIsodApdC	19	25	(Taktakishvili <i>et al.</i> , 2000; Taktakishvili <i>et al.</i> , 2001)



**Figure 32-** Repliement du G-quartet : AR177 (Zintevir)

l'hélice  $\alpha 1$  du domaine central de l'IN entière et induit la dissociation d'oligomères d'IN. En utilisant la même approche, un autre peptide interfacial dérivé de l'hélice  $\alpha 6$  a été trouvé actif contre la réaction de 3'-processing, catalysée par l'IN mutante (F185K/C280S), en présence de  $Mn^{2+}$  (Zhao et al., 2003).

## II- Dinucléotides

Des dinucléotides naturels et synthétiques ont été étudiés comme inhibiteurs potentiels de l'IN (Mazumder et al., 1997b; Taktakishvili et al., 2000; Taktakishvili et al., 2001). Quelques dinucléotides naturels comme pdApdC (**figure 31**) et synthétiques tel pdCpIsodU (**figure 31**), présentant un effet inhibiteur contre l'IN *in vitro*, sont mentionnés dans le **tableau 12**. Alors que les dinucléotides naturels sont instables, vis-à-vis des nucléases, les dinucléotides non-naturels sont totalement résistants au clivage par les exonucléases 5' et 3'. Les valeurs obtenues pour l'inhibition de la liaison de l'ADN et pour l'inhibition de l'activité catalytique de l'IN sont de la même ampleur, ce qui suggère que ces composés se lient au domaine central de l'IN et agissent de manière compétitive. D'autre part, il a été montré que l'inhibition de l'IN par les dinucléotides est métal-indépendant.

## III- Oligonucléotides

Des oligonucléotides formant des quartets de guanines (G-quartets) sont des inhibiteurs efficaces de l'IN du VIH-1 et de puissants antiviraux (**tableau 13**). Le G-quartet, AR177, (Zintevir) (**figure 32**) interagit avec gp120 et inhibe l'attachement viral *in vivo* (Cherepanov et al., 1997). Ce G-quartet est actuellement en phase I d'essai clinique. Le G-quartet T40214 entre dans les noyaux des cellules 3T3, CEMSS, MT4 et diminue la réplication du VIH-1 (**tableau 13**). Ceci suggère que l'IN pourrait être ciblée intracellulairement par les G-quartets (Jing et al., 2002). Cependant, l'inhibition de l'IN dans les cellules n'a pas encore été démontrée.

Des aptamères, 93del et 112del, dérivés d'inhibiteurs de la RNase H, sont également capables d'inhiber l'IN *in vitro* avec des valeurs d'IC<sub>50</sub> inférieures au micromolaire (**tableau 13**) (de Soultrait et al., 2002). Ces aptamères ont des séquences riches en guanine, leur permettant de former des G-quartets stabilisés par la présence d'ions potassium. La présence de ces ions améliore considérablement leur efficacité inhibitrice. Les deux aptamères sont de puissants

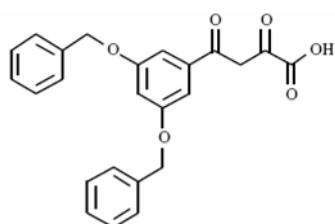
**Tableau 13-** Quelques inhibiteurs oligonucléotidiques de l'IN du VIH-1

Nom	Séquence	IC <sub>50</sub> (μM)	EC <sub>50</sub> (μM)	Référence
AR177 (Zintevir) (* phosphorothioate)	5'-G*TGGTGGGTGGGTGGG*T-3'	0,079 3'-P 0,049 TB	0,075	(Mazumder et al., 1996a)
T40214	5'-GGGCGGGCGGGCGGGC-3'	0,04 TB	0,92	(Jing <i>et al.</i> , 2000)
93del	5'-GGGGTGGGAGGAGGGT-3'	0,121 3'-P 0,166 TB	0,02	(de Soultrait <i>et al.</i> , 2002)
112del	5'-CGGGTGGGTGGGTGGT-3'	0,160 3'-P 0,108 TB	0,02	(de Soultrait <i>et al.</i> , 2002)
Oligonucléotide non-modifié	5'-AGAGATTTTCC-3'	Non inhibiteur	ND	(Brodin <i>et al.</i> , 2001)
Oligonucléotide contenant 6-oxycytosine (*)	5'-AGAGATTTTC*C*-3'	0,3 3'-P 0,3 TB	ND	(Brodin <i>et al.</i> , 2001)

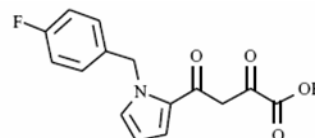
Les sigles 3'-P, TB et ND signifient respectivement 3'-processing, transfert de brins et non disponible.

IC<sub>50</sub> et EC<sub>50</sub> ont été déterminées par des tests utilisant une IN purifiée et des oligonucléotides ou des tests basés sur des cultures cellulaires et mesurant l'inhibition de l'IN ou de la réplication du VIH-1, respectivement.

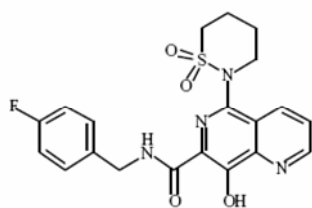
L'EC<sub>50</sub> est la concentration de la drogue entraînant 50% de son effet maximum *ex vivo*.



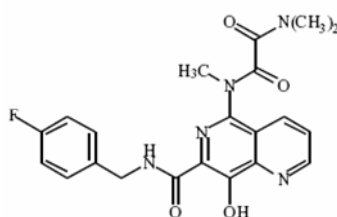
**L-708,906**



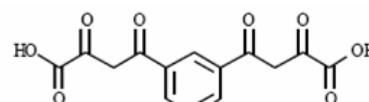
**L-731,988**



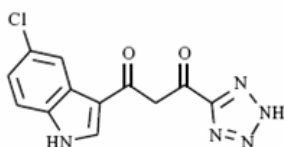
**L-870,810**



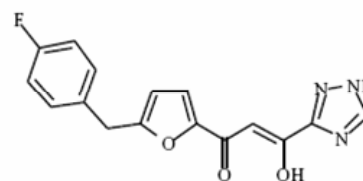
**L-870,812**



**DKA2**



**5CITEP**



**S-1360**

**Figure 33-** Structure chimique de dicéto-acides et de leurs dérivés.

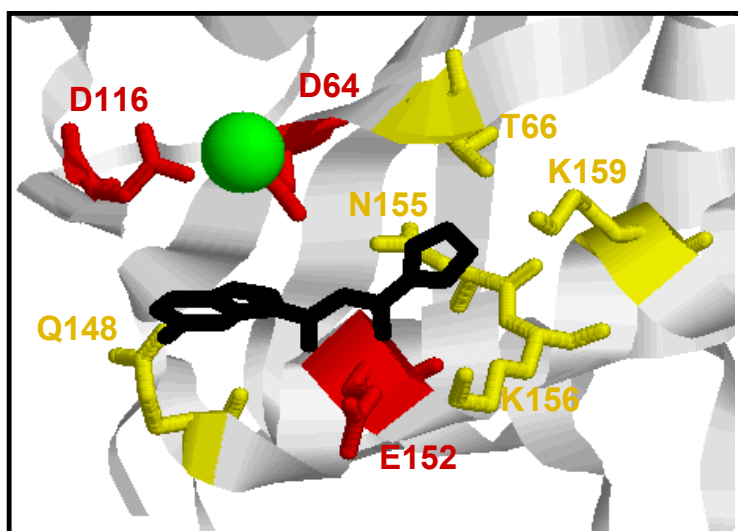
antiviraux dans les cellules P4 humaines (**tableau 13**) et ils sont dénués de cytotoxicité. Ces oligonucléotides n'inhibent pas la fusion cellulaire, ce qui donne à penser que l'IN est la première cible de ces inhibiteurs..

Des oligonucléotides contenant des 6-oxocytosines inhibent également l'IN avec des valeurs d'IC<sub>50</sub> inférieures au micromolaire (Brodin et al., 2001). Le meilleur inhibiteur décrit possède une séquence identique à celle commençant à la septième base, à partir de l'extrémité 5' du brin non-clivé du LTR U5 de l'ADN viral mais dans laquelle deux cytosines sont substituées par des 6-oxocytosines (**tableau 13**). L'oligonucléotide non-modifié n'a pas d'activité inhibitrice sur l'IN (**tableau 13**) bien qu'il ait pour l'enzyme une affinité similaire à celle de l'oligonucléotide modifié (Brodin et al., 2002). L'inhibition est dépendante de la séquence, de la présence de 6-oxocytosine et de la position de la 6-oxocytosine à la fin de l'oligonucléotide (Brodin et al., 2002; Brodin et al., 2001). Elle résulte d'un mécanisme non-compétitif dans lequel l'oligonucléotide perturbe le complexe IN/ADN (Brodin et al., 2002).

#### **IV- Dicéto-acides et leurs dérivés**

Les dicéto-acides (DKA) et leurs dérivés, composés synthétiques découverts par la société Shionogi et par les laboratoires de recherche Merck, sont des inhibiteurs sélectifs du transfert de brins. Les premiers DKA de Merck ont été identifiés via un criblage aléatoire d'une bibliothèque de 250000 composés (Hazuda et al., 2000). Au cours de cette recherche, L-731,988 et L-708,906 (**figure 33**) ont été identifiées comme les deux molécules les plus actives dans les tests de transfert de brins. La concentration de L-731,988 nécessaire pour inhiber le 3'-processing est environ 70 fois plus élevée que celle requise pour inhiber le transfert de brins (6 µM contre 0,08 µM respectivement). L-731,988 et L-708,906 inhibent la réplication du VIH-1 avec des IC<sub>50</sub> dans la gamme du micromolaire. De plus, ces composés ciblent sélectivement l'IN dans les cellules infectées par le VIH. Le groupe Merck a ensuite découvert d'autres composés tels que des dérivés 8-hydroxy-(1,6)-naphthyridine-7-carboxamides. Ces composés peuvent être considérés comme des dérivés de type DKA vu qu'ils contiennent le motif structural β-hydroxyl-cétone et en raison de leur mode d'inhibition de l'IN ; ce sont également des inhibiteurs sélectifs du transfert de brins (Hazuda et al., 2004a; Hazuda et al., 2004b; Zhuang et al., 2003). Les études cliniques ont débuté pour une de ces molécules L-870,810 (**figure 33**) mais elles ont été arrêtées après qu'une toxicité inacceptable des cellules des reins et du foie ait été trouvée chez le chien. Par contre, un autre de ces

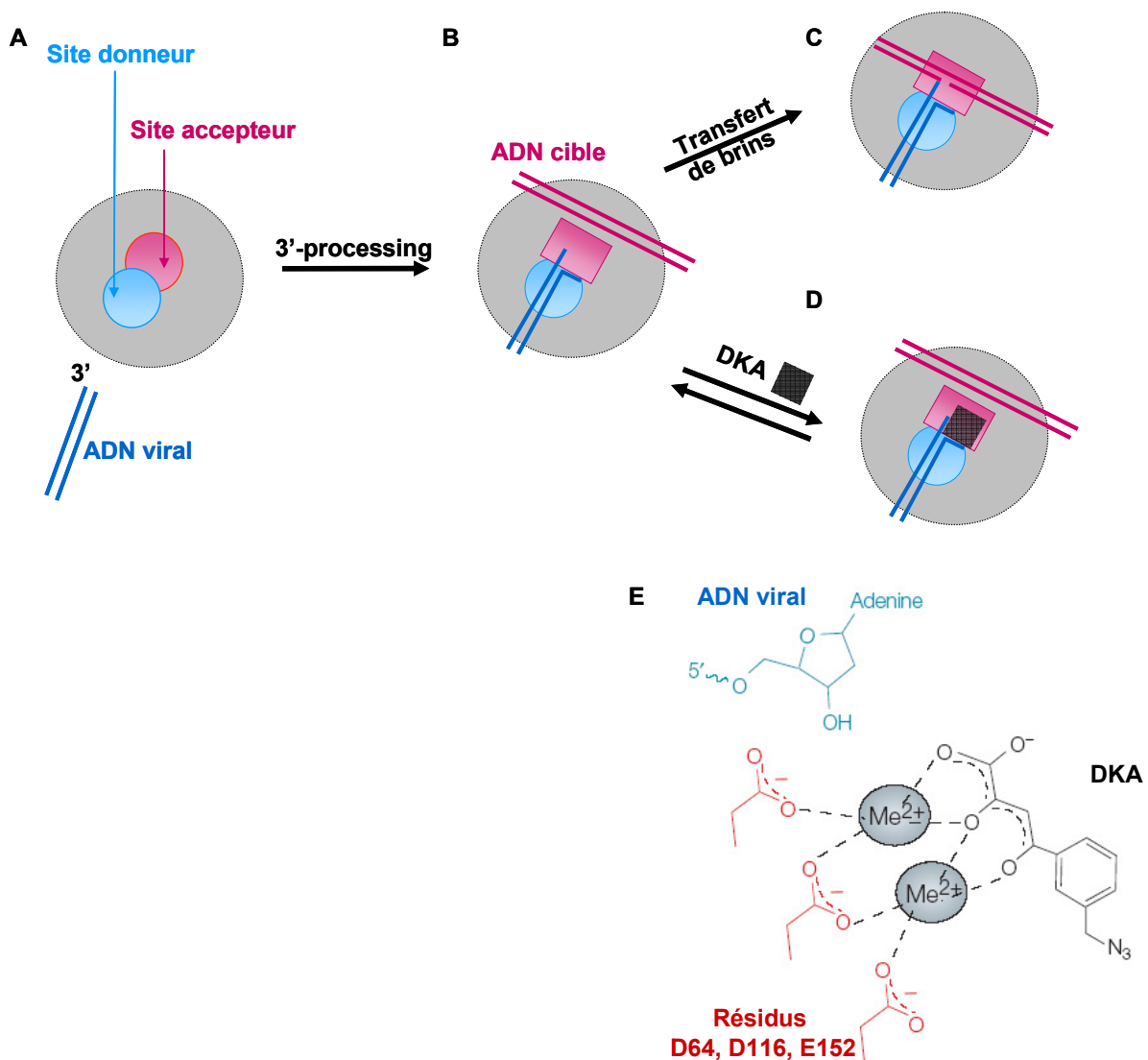




**Figure 34-** Structure cristallographique de la liaison du 5CITEP au domaine central de l'IN du VIH-1 (code PDB : 1QS4) (Goldgur et al., 1999). 5CITEP est coloré en noir. La partie du domaine central de l'IN représentée est en modèle ruban et colorée en gris clair. Les résidus de la triade catalytique DDE sont colorés en rouge et ceux impliqués dans des liaisons hydrogène avec le ligand DKA sont colorés en jaune. L'ion  $Mg^{2+}$  est représenté par une sphère verte.

composés L-870,812 (**figure 33**) est en test pré-clinique. L'activité antivirale de L-870,812 pour supprimer la virémie aussi bien que les infections chroniques des macaques, rhésus infectés par SIV, a été démontrée (Hazuda et al., 2004b).

Alors que Merck publiait ses premiers inhibiteurs DKA, la structure cristallographique d'un inhibiteur dicéto découvert par Shionogi, le 5CITEP (**figure 33**), complexé avec le domaine central de l'IN du VIH-1, a été résolue (code PDB : 1QS4) (Goldgur et al., 1999). Il s'agit de la seule structure cristallographique du domaine central de l'IN du VIH-1, liée à un inhibiteur localisé au site actif, actuellement disponible. Le 5CITEP appartient à la famille des DKA puisque l'on peut considérer le tétrazole comme un isostère du groupe acide carboxylique (Marchand et al., 2002). Il est environ 40 fois moins sélectif pour le transfert de brins que des DKA de Merck tel L-708,906. Son effet sur le transfert de brins est plus faible en présence de  $Mg^{2+}$  que de  $Mn^{2+}$  et il ne présente pas d'activité antivirale. Malgré cela, la résolution de la structure du complexe qu'il forme avec le domaine central de l'IN a apporté des informations importantes pour la conception de nouvelles drogues. Dans l'unité asymétrique cristallographique, trois monomères du domaine central ont été observés ; deux d'entre eux forment un dimère. Chaque molécule de 5CITEP se lie au centre du site actif de chaque monomère de l'IN. Cependant, la position de l'inhibiteur n'a pu être tracée sans ambiguïté que dans un seul cas. L'inhibiteur est placé entre les résidus du site actif D64, D116 et E152 ; il est lié à la surface par des liaisons hydrogènes entre le tétrazole et des acides aminés voisins : T66, N155, K156 et K159 (**figure 34**). Parmi ces acides aminés, K156 et K159 sont connus pour être impliqués dans la liaison de l'ADN viral (Jenkins et al., 1997). L'atome N1 de l'indole est impliqué dans une liaison hydrogène avec Q148 ; ce dernier a été identifié comme liant l'adénine 5' du brin non clivé de l'ADN viral (Esposito and Craigie, 1998). 5CITEP est également stabilisé dans le site actif par une liaison hydrogène avec un des résidus de la triade catalytique, E152. L'ion  $Mg^{2+}$  reste lié à D64 et D116, en présence de l'inhibiteur. Le site actif auquel l'inhibiteur tracé se lie, se trouve près de l'axe cristallographique d'ordre 2 et il interagit avec une autre molécule 5CITEP, à travers cet axe. Il est possible que cela puisse affecter le mode de liaison au site actif observé. Des études de docking soutiennent la liaison à deux sites reliés par un axe d'ordre 2 mais la liaison au domaine isolé prédite est différente de celle observée en cristallographie (Sotriffer et al., 2000). Dans une simulation de dynamique moléculaire, une molécule d'eau a été identifiée comme formant un pont potentiel entre le 5CITEP et les résidus D64 et N155 (Ni et al., 2001).



**Figure 35-** Le mécanisme d'action proposé pour les dicéto-acides et leurs dérivés (Pommier et al., 2005). Les DKA inhibent sélectivement le transfert de brins en se liant à l'interface de complexes IN/ADN. **A-** Les deux sites de liaisons proposés au niveau de l'IN : le site donneur pour l'ADN viral (cercle bleu) et le site accepteur pour l'ADN cible (cercle rose). **B-** Montre l'événement qui suit le 3'-processing : le complexe IN/ADN subit un changement structural qui rend le site accepteur (rectangle rose) capable de lier l'ADN cible. **C-** Dans des conditions normales, la liaison de l'ADN cible au site accepteur conduit au transfert de brins. **D-** L'inhibiteur DKA (rectangle noir) peut seulement se lier au site accepteur après le 3'-processing. **E-** Montre le détail de la liaison hypothétique des DKA (ici, MA-DKA) à l'interface du complexe IN/ADN/métal. L'extrémité 3' clivée de l'ADN viral (en bleue) est liée à l'IN (les trois résidus catalytiques sont montrés en rouge). Il a été proposé que les DKA chélatent deux ions métalliques divalents dans le site actif de l'IN et stabilisent le complexe IN/ADN après le 3'-processing.

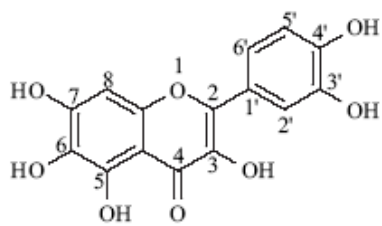
Un autre composé de type DKA, développé par la société Shionogi, S-1360 (**figure 33**), a atteint la phase II des tests cliniques mais son développement a été interrompu en 2003 pour des problèmes de biodisponibilité.

Les DKA sont les premiers inhibiteurs de l'IN à fournir la preuve que l'inhibition sélective du transfert de brins peut efficacement diminuer l'infection au VIH-1 (Hazuda et al., 2000; Hazuda et al., 2004b). Malgré leur sélectivité pour le transfert de brins, les DKA présentent également une activité inhibitrice du 3'-processing mais seulement à des concentrations beaucoup plus élevées (Hazuda et al., 2000; Marchand et al., 2003). Certains peuvent tout de même inhiber le 3'-processing à des concentrations plus faibles que d'autres comme le 5CITEP ou des DKA bifonctionnels de Merck tel DKA2 (**figure 33**). Cependant, l'inhibition du 3'-processing par les DKA n'a pas été observée dans les cellules traitées.

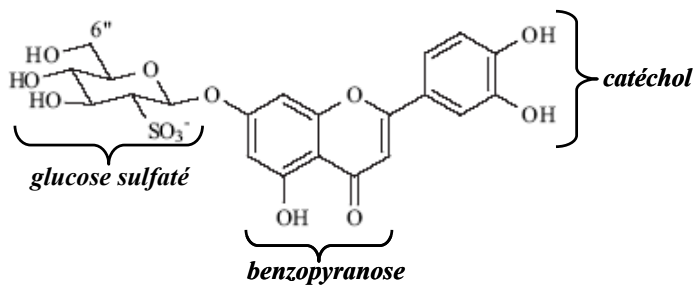
Un cofacteur métallique divalent est nécessaire pour la liaison des DKA à l'IN (Grobler et al., 2002) et l'inhibition de l'IN est métal-dépendant (Grobler et al., 2002; Marchand et al., 2003). Les fonctions dicéto ou  $\beta$ -hydroxyl-cétone sont connues pour être des chélateurs d'ions métalliques divalents. L'inhibition métal-dépendant des DKA a alors été interprétée comme indiquant une interaction directe de ces molécules avec l'ion métallique divalent dans le site actif de l'enzyme (Grobler et al., 2002; Marchand et al., 2003).

Un mode de liaison des DKA au domaine central de l'IN a été proposé (**figure 35**) (Marchand et al., 2002). Dans ce modèle, les deux sites catalytiques sont organisés autour de la triade catalytique DDE et le ou les ion(s) métallique(s) divalent(s). Le site donneur lie l'extrémité de l'ADN viral et catalyse le 3'-processing. Il en résulte un changement structural de l'IN qui permet la liaison de l'ADN cible dans le site accepteur puis le transfert de brins (Espeseth et al., 2000; Marchand et al., 2002). Il est suggéré que les DKA se lient sélectivement à une conformation unique du site accepteur, après la liaison de l'ADN viral et le 3' procesing. Le changement requis dans la conformation du site accepteur pour loger le ligand DKA devrait être produit par la liaison de l'ADN viral et le 3'-processing. Selon ce schéma, la coordination métallique serait cruciale pour la liaison des DKA au site accepteur. Bien que le nombre d'ions divalents requis pour l'activité de l'IN reste controversé, un modèle (**figure 35-E**) avec deux ions métalliques coordonnés aux résidus de la triade catalytique DDE, a été proposé (Grobler et al., 2002). Ce modèle est soutenu par des tests de proximité de scintillation (Espeseth et al., 2000; Grobler et al., 2002).

Les mutants d'IN sélectionnés par les DKA présentent des mutations dans le domaine central et une seule dans le domaine C-terminal (Billich, 2003; Fikkert et al., 2003; Hazuda et al., 2004a; Hazuda et al., 2000). Les acides aminés du domaine central qui confèrent la résistance,



**Quercétagétine**



**Thalassioline A**

**Figure 36-** Exemples d'inhibiteurs de type flavonoïdes de l'IN du VIH-1.

sont regroupés autour du site actif défini par la triade catalytique DDE. Ce groupe d'acides aminés chevauche le groupe de résidus identifiés pour être impliqués dans la liaison d'ADN. Le résidu Q148, commun aux deux groupes, semble être impliqué dans le positionnement correct de l'ADN viral, avant le transfert de brins (Johnson et al., 2006a). Le regroupement des résidus de l'IN impliqués dans la résistance aux DKA et dans les contacts avec l'ADN autour de la triade catalytique qui coordonne le ou les ion(s) métalliques, met en évidence la probabilité que les DKA se lient à l'interface du complexe ternaire ADN(viral)/IN/métal. La liaison des DKA aurait lieu spécifiquement à l'étape du processus catalytique de l'IN qui précède immédiatement le passage du 3'-processing au transfert de brins (**figure 35, étape B**). Une telle liaison stabiliserait l'intermédiaire du 3'-processing et empêcherait le transfert de brins.

Il est également proposé que les DKA bifonctionnels et le 5CITEP se lient aussi au site donneur (**figure 35**) comme ils sont moins sélectifs du transfert de brins (Marchand et al., 2002). Mais cette liaison se ferait avec une affinité plus faible que celle pour le site accepteur comme ce sont de meilleurs inhibiteurs du transfert de brins que du 3'-processing.

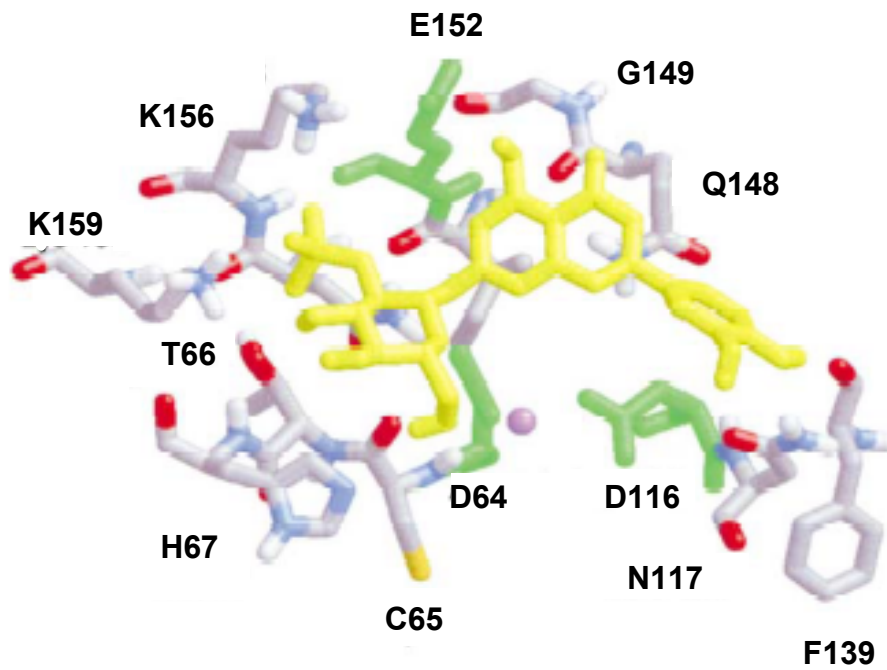
## **V- Polyphénols**

De nombreux inhibiteurs sont regroupés dans cette classe. Ils sont formés de groupes catéchols et/ou phénols et sont naturels ou synthétiques. Un grand nombre d'aromatiques polyhydroxylés sont fortement cytotoxiques mais certains d'entre eux sont habituellement consommés et exempts de toxicité. A partir de considérations structurales, on peut facilement déduire que les catéchols peuvent agir comme des chélateurs métalliques, être engagés dans des réactions redox ou s'intercaler dans l'ADN (quand le nombre de noyaux aromatiques est assez grand).

### **A- Des polyphénols naturels**

#### **A-1- Les Flavonoïdes**

Une des premières familles polyphénoliques naturelles à avoir été étudiée est celle des flavonoïdes (Fesen et al., 1994). La quercétagétine (**figure 36**) inhibe efficacement l'IN *in vitro* ( $IC_{50} = 0.8 \mu\text{M}$  pour 3'-processing,  $IC_{50} = 0.1 \mu\text{M}$  pour le transfert de brins et  $IC_{50} = 0.8 \mu\text{M}$  pour la désintégration). La quasi planéité, l'aromaticité et la polarité des flavonoïdes



**Figure 37-** Mode de liaison de la thalassioline A au domaine central de l'IN du VIH-1, déterminé par docking (Rowley et al., 2002). L'inhibiteur est montré en jaune. Les résidus catalytiques (D64, D116 et E152) sont colorés en vert alors que les autres acides aminés sont colorés par le code de couleur du type d'atome. L'ion  $Mg^{2+}$  est représenté par une sphère colorée en violet. A l'exception des deux résidus enfouis, I151 et N155, le nom de tous les autres acides aminés est mentionné.

peuvent permettre à ces composés de se lier à l'ADN par empilement avec une adénine ou une guanine. L'empilement avec des bases de l'ADN au niveau du site actif pourrait jouer un rôle dans le mécanisme inhibiteur de ces composés. Les flavonoïdes sont également connus pour être impliqués dans les réactions redox et dans la chélation d'ions métalliques divalents (Rice-Evans and Packer, 1998). Un rôle chélateur a été suggéré par le fait que l'inhibition de l'IN est plus efficace en présence de  $Mn^{2+}$  que de  $Mg^{2+}$ . Néanmoins, aucune corrélation claire n'a été établie entre la présence des groupes hydroxyles et l'inhibition de l'IN.

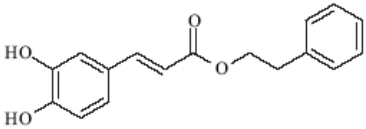
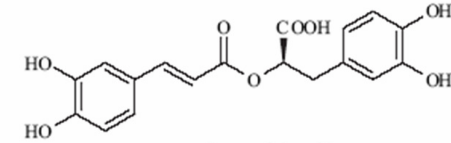
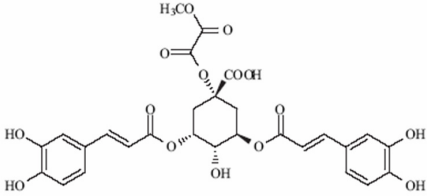
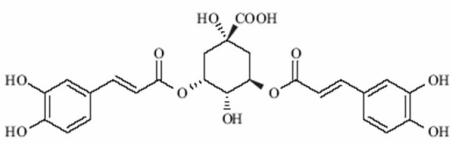
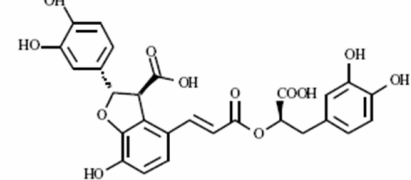
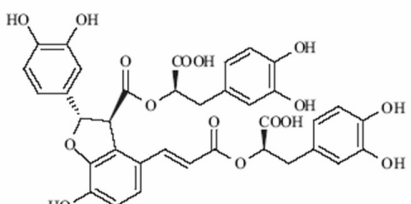
Plus récemment, il a été trouvé qu'une lutéonine substituée par un  $\beta$ -D-glucose sulfaté, à la position 7 et appelée la thalassioline A (**figure 36**), présentait une valeur d' $IC_{50}$  de 2.1  $\mu M$  pour le 3'-processing et de 0.4  $\mu M$  pour le transfert de brins (Rowley et al., 2002). Cette molécule a une activité antivirale avec une  $IC_{50}$  de 27  $\mu M$  et n'est relativement pas toxique. Des études de docking montrent une interaction favorable de la thalassioline A au domaine central de l'IN du VIH-1 (Rowley et al., 2002) (**figure 37**). L'oxygène 6'' est coordonné à l'ion  $Mg^{2+}$ . Le sulfate interagit avec K156 et K159. Le cycle benzopyranone se trouve près d'E152 et de Q148 tandis que le cycle catéchol s'étend dans l'espace situé entre Q148 et D116 et présente des interactions supplémentaires avec N117 et F139. Ces études de modélisation moléculaire ont été réalisées à partir de la structure cristallographique du domaine central de l'IN du VIH-1 complexé avec l'inhibiteur 5CITEP et contenant un ion  $Mg^{2+}$  (code PDB : 1QS4) (Goldgur et al., 1999). Le mode de liaison de la thalassioline A, montré par les études de docking, recouvre la localisation du 5CITEP, déterminée expérimentalement. Le glucose sulfaté de la thalassioline A occupe la même position de liaison que le cycle tétrazol du 5CITEP et la position du cycle benzopyranose montre des similarités avec le céto-énol et la partie chloroindole du 5CITEP.

#### A-2- Les esters d'acide caféique

Le premier dérivé d'acide caféique identifié comme inhibiteur de l'IN *in vitro* est l'ester phényléthylique d'acide caféique (CAPE), un produit naturel fabriqué par les abeilles (Fesen et al., 1993). CAPE est dix fois plus puissant en tant qu'inhibiteur du transfert de brins qu'en tant qu'inhibiteur du 3'-processing (**tableau 14**). La préincubation de CAPE avec l'IN (avant l'ajout des substrats) augmente énormément l'inhibition du 3'-processing mais faiblement l'inhibition du transfert de brins. L'inhibition de la désintégration par CAPE est métal-dépendante ( $IC_{50} = 65 \mu M$  ( $Mn^{2+}$ ) et  $155 \mu M$  ( $Mg^{2+}$ )) (Fesen et al., 1994).

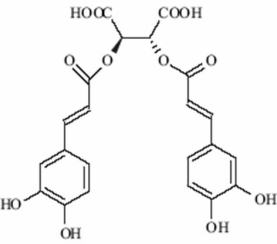


**Tableau 14-** Quelques esters d'acide caféique naturels et leurs activités d'inhibition contre l'IN.

Nom	Structure chimique	IC <sub>50</sub> (μM)	Réf.
CAPE (Ester Phényléthylrique d'Acide Caféique)		220 3'-P 19 TB	(Fesen et al., 1993)
Acide Rosmarinique		9 3'-P 4 TB	(Mazumder et al., 1997a)
1-MO-3,5-DCQA (l'acide 1-méthoxyoxalyl-3,5-dicaféoylquinique)		0,45 3'-P 0,43 TB	(Robinson, Jr. et al., 1996b)
3,5-DCQA (l'acide 3,5-dicaféoylquinique)		0,64 3'-P 0,66 TB	(Robinson, Jr. et al., 1996b)
Acide Lithospermique		0,84 3'-P 0,48 TB	(Abd-Elazem et al., 2002)
Acide Lithospermique B (dimère de l'acide rosmarinique)		0,48 3'-P 0,37 TB	(Abd-Elazem et al., 2002)

Les sigles 3'-P et TB signifient respectivement 3'-processing et transfert de brins. IC<sub>50</sub> déterminées par des tests in vitro, en présence de Mn<sup>2+</sup>.

**Tableau 15-** L'inhibition de l'IN et de la réplication virale par l'Acide L-chicorique (L-CA)

Structure chimique	IC <sub>50</sub> (μM)	ED <sub>50</sub> (μM)	CT <sub>5</sub> (μM)	Réf.
	0,15 3'-P 0,13 TB	4.2	264	(Robinson, Jr. et al., 1996b) (Reinke et al., 2002)

Les sigles 3'-P et TB signifient respectivement 3'-processing et transfert de brins.

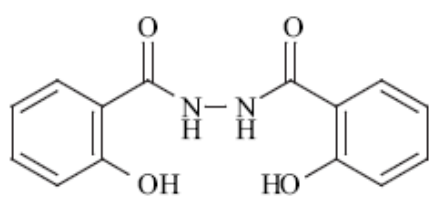
CT<sub>5</sub> (la dose cytotoxique de 5%) est la concentration de composé qui inhibe la croissance cellulaire de 5% ou moins sur une période de 72h. ED<sub>50</sub> (la dose effective de 50%) est la concentration de composé qui inhibe la réplication du VIH de 50%. IC<sub>50</sub> (la concentration inhibitrice de 50%) est la concentration en composé qui inhibe 50% de la réaction catalysée par l'IN

Plus tard, d'autres dérivés naturels de l'acide caféique, non sélectifs du transfert de brins et inhibant plus efficacement l'IN, ont été identifiés comme par exemple : l'acide rosmarinique, trouvé dans les espèces du *Boraginaceae* (Mazumder et al., 1997a), 1-MO-3,5-DCQA et 3,5-DCQA obtenues à partir de plantes médicinales (Robinson et al., 1996b) (**tableau 14**). Ces composés ont une activité antivirale à des concentrations non toxiques. Ils inhibent la réaction de désintégration, catalysée uniquement par le domaine central de l'IN. On en conclut que ces inhibiteurs doivent interagir avec ce domaine. De plus, des études de modélisation moléculaire montrent un mode de liaison de ces composés favorable au domaine central, au niveau d'une fissure à haut potentiel négatif, formée en partie par les résidus H67, E92, D116, N117, Q148, K159 et D64 (Robinson et al., 1996a). La différence d'énergie libre interne des complexes IN/ligand est corrélée avec la capacité de ces composés à inhiber l'IN *in vitro*. Cependant, il faut noter que les études de modélisation moléculaire ont été réalisées à partir de la structure du domaine central de l'IN cristallisée, en utilisant le cacodylate comme tampon et dont le site actif ne lie pas un ion métallique divalent (code PDB : 1ITG) (Dyda et al., 1994). L'acide lithospermique B (un dimère de l'acide rosmarinique) et l'acide lithospermique, isolés à partir d'extraits solubles dans l'eau d'une variété chinoise de sauge (*Salvia miltiorrhiza*), sont également actifs sur le 3'-processing et le transfert de brins avec des valeurs d'IC<sub>50</sub> inférieures au micromolaire (**tableau 14**). De plus, ces molécules sont actives sur l'infection au VIH-1 de cellules H9 (IC<sub>50</sub> = 6,9 µM et 2 µM) et elles ne présentent pas de cytotoxicité (Abd-Elazem et al., 2002).

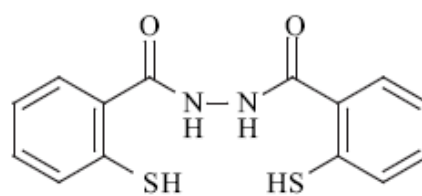
## **B- Des polyphénols synthétiques**

### **B-1-Les dérivés dicaféoyliques**

Parmi les dérivés dicaféoyliques synthétiques, l'acide L-chicorique (L-CA) (Robinson et al., 1996b) (**tableau 15**) est un composé très actif contre l'IN *in vitro*. Il inhibe également la croissance de cellules MT-2. Lorsque ce composé a été testé sur d'autres cibles virales potentielles, il a pu être démontré qu'il s'agit d'un inhibiteur sélectif de l'IN (McDougall et al., 1998). De plus, il a été prouvé qu'il agit sur le domaine central de l'IN (Robinson et al., 1996a). L'inhibition de l'IN a été aussi démontrée par la sélection d'un variant du VIH L-CA-résistant, suivi de l'isolement d'un virus mutant (G140S) (Reinke et al., 2002). Cependant, en culture cellulaire, L-CA doit son activité anti-VIH à une interaction avec gp120 (Pluymers et



**Salicylhydrazide**



**Mercaptosalicylhydrazide**

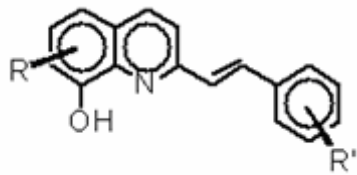
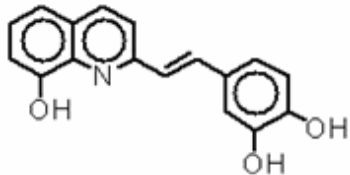
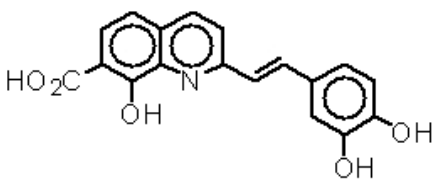
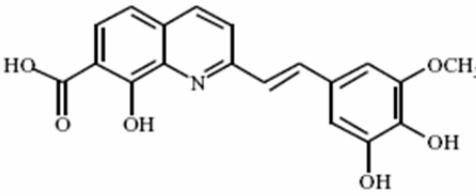
**Figure 38-** Des dérivés hydrazides

al., 2000). Par des passages répétés du virus en présence de L-CA, des mutations ont été trouvées dans les boucles V2, V3 et V4 de gp120 tandis qu'aucune mutation n'a été détectée dans l'IN. Cette dernière qui possède la mutation G140S semble être aussi sensible à l'effet de L-CA que l'IN sauvage. De plus, les expériences au cours desquelles l'inhibiteur est ajouté à différents temps après l'infection au VIH-1, corroborent une interaction de L-CA à l'étape d'absorption du virus. Par la suite, il a été rapporté que L-CA inhibe l'intégration à des concentrations de 500 nM à 10  $\mu$ M et l'entrée du virus à des concentrations voisines de 1  $\mu$ M. De plus, L-CA n'est pas un inhibiteur compétitif mais un inhibiteur réversible de l'IN *in vitro* et *in vivo* (Reinke et al., 2004). Ceci suggère que L-CA interagit probablement avec des acides aminés de l'IN autres que ceux qui lient le substrat ADN. Néanmoins, quelle que soit la controverse sur le mode d'action antiviral de L-CA, cela ne compromet pas nécessairement son développement en tant que médicament anti-VIH.

### B-2-Les salicylhydrazides et mercaptosalicylhydrazides

En utilisant une base de données tridimensionnelle de 152 composés dont l'activité inhibitrice sur l'IN était connue, Hong et al. ont proposé plusieurs modèles de pharmacophores (Hong et al., 1997) qui ont conduit à une nouvelle classe d'inhibiteurs, les hydrazides (Neamati et al., 1998; Neamati et al., 2002; Zhao et al., 1997). Ces composés présentent deux particularités : 1) Ce sont de vrais inhibiteurs compétitifs avec des  $IC_{50}$  dans la gamme des 10  $\mu$ M. Ils inhibent les deux étapes, qu'ils soient ajoutés au complexe ternaire IN/cation/ADN ou que l'ADN soit ajouté au complexe qu'ils forment avec l'IN et le cation. 2) Le salicylhydrazide inhibe l'IN en présence de  $Mn^{2+}$  mais pas en présence de  $Mg^{2+}$ . Par contre, les dérivés mercaptosalicylhydrazides inhibent l'IN aussi bien en présence de  $Mn^{2+}$  que de  $Mg^{2+}$ . Les calculs *ab initio* au niveau B3LYP montrent que  $Mg^{2+}$  forme avec le salicylhydrazide et son dérivé mercaptosalicylhydrazide des complexes de conformation différente. Lors d'une étude de docking du N,N-bis hydrazide (2-mercaptopbenzoyl) (**figure 38**) sur le domaine central de l'IN du VIH-1, aucune liaison hydrogène entre l'inhibiteur et l'enzyme n'a été observée. Cependant, durant une simulation de dynamique moléculaire, la distance la plus faible (4.86 Å) observée entre le soufre de C65 et un des atomes de soufre de l'inhibiteur suggère que la formation d'un pont disulfure est possible. Une fois la formation du pont disulfure avec C65 permise, l'inhibiteur s'est parfaitement positionné dans le site actif de l'IN en chélatant l'ion  $Mg^{2+}$ .

**Tableau 16-** Les dérivés des styrylquinoléines

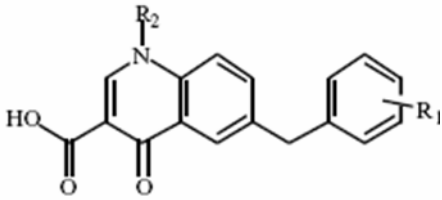
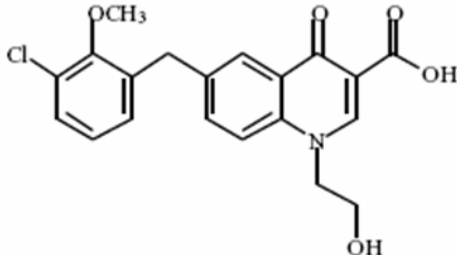
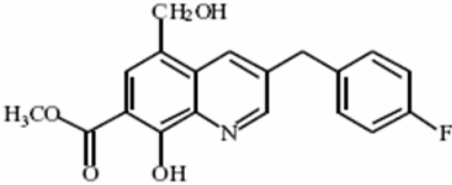
	
<p>Structure générale des SQ</p>	<p>Structure SQ « minimale » pour l'activité inhibitrice contre l'IN</p>
	
<p>KHD161 (CNRS)</p>	<p>FZ41 (Bioalliance)</p>

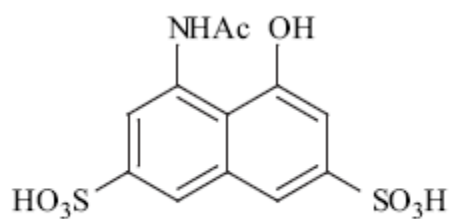
### B-3- Les dérivés des styrylquinoléines (SQ)

La présence d'ions métalliques divalents est essentielle à l'activité endonucléolytique de l'IN du VIH. Il y a quelques années, l'équipe de J. d'Angelo (CEP, Châtenay-Malabry) a conçu et synthétisé des dérivés des styrylquinoléines (SQ). Ces molécules ont été créées pour complexer  $Mg^{2+}$  dans le site actif de l'IN. Ces SQ ont été testées par l'équipe de J-F. Mouscadet (LBPA, ENS-Cachan). Plusieurs SQ sont des inhibiteurs de l'IN *in vitro* avec des valeurs d'IC<sub>50</sub> dans la gamme du micromolaire ou même inférieures au micromolaire pour les réactions de 3'-processing, de transfert de brins et de désintégration (Mekouar et al., 1998; Polanski et al., 2002; Zouhiri et al., 2000). Ces SQ bloquent également la réplication du VIH-1 *ex vivo* avec des IC<sub>50</sub> de l'ordre du micromolaire et sont dénuées de cytotoxicité.

Ces molécules sont des 8-hydroxyquinoléines reliées à un résidu aromatique hydroxylé par un espaceur oléfinique dont la structure générale est montrée dans le **tableau 16**. La SQ présente dans ce **tableau** (1<sup>ière</sup> ligne, colonne de droite) constitue la structure de base "minimale" nécessaire à l'activité biologique. La présence d'une fonction acide en C7 renforce de manière sensible l'activité de la molécule *in vitro* et en culture cellulaire ; elle abaisse considérablement la cytotoxicité. C'est ainsi que la SQ, KHD161 (**tableau 16**), possède une activité inhibitrice de l'IN *in vitro* (IC<sub>50</sub>) sensiblement micromolaire, une activité antivirale en culture cellulaire (IC<sub>50</sub>) micromolaire et une totale absence de cytotoxicité à 100  $\mu$ M. A la suite de cette découverte initiale, les relations structure-activité de cette famille de molécules ont été explorées (Zouhiri et al., 2000). L'influence des substituants de la partie quinoléine a été étudiée en préparant de très nombreux composés. Il a ainsi été mis en évidence l'importance du pharmacophore "acide salicylique" pour cette famille d'anti-IN. La fonction acide en C7 a été remplacée par un résidu aroyle ou hétéroaroyle (Normand-Bayle et al., 2005). Sur le plan biologique, ces composés se sont avérés peu actifs sur l'IN *in vitro* ; en revanche, certains conservent une activité cellulaire intéressante bien que souvent obérée par une certaine cytotoxicité. Le cycle catéchol a été remplacé par différents cycles hétéroaromatiques ou par un cycle phényl substitué par des groupes attracteurs ou donneurs d'électrons (Zouhiri et al., 2000). L'évaluation biologique *in vitro* de ces nouvelles molécules a montré que l'ensemble de ces composés inhibe l'IN à un bon niveau. Ceci semble indiquer que la partie droite des SQ interagit de manière non-spécifique avec l'IN, probablement par une interaction de type  $\pi$ . Par contre, en culture cellulaire, les composés dépourvus d'un groupe hydroxyle en C4' et d'un autre résidu oxygéné en C3' ou C5' sont peu ou pas actifs. Il a été synthétisé des molécules portant en C5 un groupement acide supplémentaire relié au

**Tableau 17-** Des dérivés quinoléiques

 <p>The image shows a general chemical structure of a quinoline derivative. It features a quinoline ring system with a nitrogen atom at position 1. A substituent R<sub>2</sub> is attached to the nitrogen. A carboxylic acid group (-COOH) is attached to the ring at position 2. A benzyl group (-CH<sub>2</sub>-C<sub>6</sub>H<sub>4</sub>-R<sub>1</sub>) is attached to the ring at position 6.</p>	 <p>The image shows the chemical structure of JTK-303. It consists of a quinoline ring system. At position 1, there is a 2-hydroxyethyl group (-CH<sub>2</sub>-CH<sub>2</sub>-OH). At position 2, there is a carboxylic acid group (-COOH). At position 6, there is a benzyl group (-CH<sub>2</sub>-C<sub>6</sub>H<sub>4</sub>-) where the benzene ring has a methoxy group (-OCH<sub>3</sub>) at the 3-position and a chlorine atom (-Cl) at the 4-position.</p>
<p>Structure générale des composés brevetés par la société japonaise Tobacco</p>	<p>JTK-303 (Gilead Sciences)</p>
 <p>The image shows a chemical structure of a quinoline derivative. It features a quinoline ring system with a nitrogen atom at position 1. At position 2, there is a hydroxyl group (-OH) and a methoxy group (-OCH<sub>3</sub>). At position 6, there is a benzyl group (-CH<sub>2</sub>-C<sub>6</sub>H<sub>4</sub>-F) where the benzene ring has a fluorine atom (-F) at the 4-position. At position 8, there is a hydroxymethyl group (-CH<sub>2</sub>OH).</p>	
<p>Composé breveté par la société Shionogi</p>	



**Y-3**

**Figure 39-** Structure chimique d'Y-3

noyau quinoléine soit directement, soit par l'intermédiaire d'un espaceur vinylique. Ces deux nouveaux composés se sont avérés actifs *in vitro* avec des  $IC_{50}$  inférieures au micromolaire mais sans gain appréciable de l'activité antivirale en culture cellulaire. Un travail de modulation du bras espaceur a également été effectué. En particulier, un espaceur amide maintenait l'activité biologique alors que celle-ci était perdue avec un bras urée ou hydrazone (Benard et al., 2004).

La SQ, FZ41 (société Bioalliance) est actuellement en test pré-clinique et devrait entrer en phase I d'étude clinique très prochainement. FZ41 présente une valeur d' $IC_{50}$  de 0,7  $\mu M$  pour le 3'-processing et de 1,7  $\mu M$  pour le transfert de brins. Ce composé possède une activité anti-VIH en culture cellulaire avec des valeurs d' $EC_{50}$  de 5 à 10  $\mu M$  dans les cellules CEM et de 1 à 4  $\mu M$  dans les cellules HeLa. De plus, la valeur de  $TC_{50}$  (la concentration qui réduit la viabilité cellulaire à 50%) est supérieure à 100  $\mu M$ .

#### B-4-D'autres dérivés quinoléiques

La société japonaise Shionogi a breveté des esters méthylés de l'acide 8-hydroxyquinoléine-7-carboxylique ayant en position 3 un groupe benzyl substitué. L'exemple montré dans le **tableau 17** est actif sur l'IN avec une  $IC_{50}$  de 0,2  $\mu M$ .

Des 6-benzyl-3-carboxy-4-quinoléones (société Japonaise Tobacco) présentent une activité contre l'IN avec des  $IC_{50}$  inférieures au micromolaire (**tableau 17**). Un de ces composés, JTK-303 (**tableau 17**) a déjà subi la phase I d'essai clinique au Japon et la phase II est en cours (Gilead Sciences). Il montre une  $IC_{50}$  de 1,7  $\mu M$  contre l'IN. Cependant, peu d'informations sur ce composé sont actuellement connues.

## **VI- Des dérivés de l'acide naphthalènesulfonique**

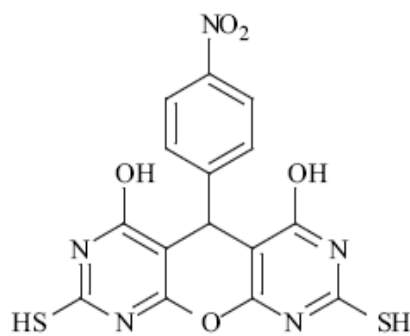
A partir de la connaissance d'inhibiteurs de l'IN du VIH-1 comme CAPE et les dérivés anthraquinones, il en a été déduit un pharmacophore à trois points pour la liaison d'inhibiteurs de l'IN (Nicklaus et al., 1997). Lors du criblage de bases de données tridimensionnelles avec ce pharmacophore, des dérivés de l'acide naphthalènesulfonique ont été identifiés et, en particulier, le disulfonate Y-3 (**figure 39**). Il s'agit d'un inhibiteur peu puissant ( $IC_{50} = 16.2 \mu M$  pour le 3'-processing et  $IC_{50} = 10.9 \mu M$  pour le transfert de brins) (Lubkowski et al., 1998b). Cependant, il est intéressant de le mentionner car il est également actif contre l'IN



**Tableau 18-** Divers produits naturels actifs contre l'IN du VIH-1 *in vitro*

Nom	Structure	IC <sub>50</sub> (μM)	Réf.
Intégracine A, R = COCH <sub>3</sub>		3,2 (3'-P+TB)	(Singh et al., 2002c)
Intégracine B, R = H		6,1 (3'-P+TB)	
Intégracine C		3,5 (3'-P+TB)	(Singh et al., 2002c)
Integramycine		4,0 (TB)	(Singh et al., 2002a)
Intégrastine A, R=CH <sub>2</sub> OH		1,1 (TB)	(Singh et al., 2002b)
Intégrastine B, R=CHO		2,5 (TB)	
Isocomplestatine		4,0 (TB) 0,2 (R)	(Singh et al., 2001)

3'-P+TB : mesure de l'inhibition couplée du 3'-processing et du transfert de brins. TB : mesure de l'inhibition du transfert de brins. R : mesure de l'inhibition de la répllication du VIH-1.



**V-165**

**Figure 40-** Structure chimique de V-165

d'ASV mais aussi parce que la structure cristallographique du complexe formé par le domaine central de l'IN d'ASV et Y-3 a été décrite (Lubkowski et al., 1998b). Cet inhibiteur se lie au domaine central de l'IN d'ASV, pas très loin du site actif mais de l'autre côté de la boucle catalytique. L'inhibiteur interagit avec un second monomère du domaine central de l'IN, à travers un axe d'ordre 2, soulevant la possibilité que l'effet cristallin joue un rôle dans la liaison. Y-3 a ensuite été docké sur le domaine central de l'IN du VIH-1 (Sotriffer et al., 2000). Le bicyclic aromatique est placé horizontalement dans le site actif au-dessus de D64. Un des deux sulfonates est placé entre K156 et K159 tandis que l'autre se trouve entre l'ion  $Mg^{2+}$  et l'amide de Q148. L'oxygène de la fonction amide d'Y-3 est coordonné avec l'ion  $Mg^{2+}$ . D'autres interactions ont été trouvées avec C65, T66 et N155.

## VII- Divers composés

### A- Les produits naturels

L'exploration des sources naturelles de molécules a mené à la découverte de divers inhibiteurs naturels de l'IN comme par exemple : les intégracines, les intégrastines, l'intégramycine et l'isocomplestatine (**tableau 18**). De nombreux inhibiteurs couvrant une grande diversité structurale, ont ainsi été identifiés.

### B- Les molécules synthétiques

De nombreuses molécules synthétiques couvrant une large diversité structurale ont été décrites. Parmi elles, on peut, en particulier, mentionner le composé V-165 (**figure 40**) qui a été identifié comme un candidat potentiel aux études cliniques. Il s'agit d'une pyranodipyrimidine qui inhibe l'activité de l'IN avec des valeurs d' $IC_{50}$  de 0,3  $\mu M$  pour le processus catalytique complet (3'-processing et transfert de brins), de 0,9  $\mu M$  pour le 3'-processing et de 16,1  $\mu M$  pour le transfert de brins. V-165 n'affecte pas directement la catalyse du 3'-processing mais la formation du complexe IN/ADN. Ce composé inhibe la réplication virale *ex vivo* (cellules MT-4) avec une valeur d' $IC_{50}$  de 8.9  $\mu M$  et une valeur de  $CC_{50}$  de 121  $\mu M$  (Pannecouque et al., 2002). Les expériences, au cours desquelles l'inhibiteur est ajouté à différents temps après l'infection des cellules par le VIH-1 et l'accumulation de cercles à 2 LTR, indiquent l'inhibition de l'IN par V-165 dans les cellules infectées. Des



mutations de l'IN menant à la résistance à V-165 ont été identifiées ; elles mettent également en évidence que l'IN est ciblée par ce composé, dans les cellules infectées. Cependant, V-165 inhibe la rétrotranscriptase dans la même gamme de concentration que son inhibition de l'IN, indiquant alors que cet inhibiteur pourrait avoir plus d'une cible.

Un grand nombre d'inhibiteurs de l'IN, ayant une grande diversité structurale, ont donc été décrits ; cela montre l'intérêt que représente l'IN en tant que cible pour la thérapie anti-VIH. Actuellement, quatre composés sont en étude clinique. Les progrès dans la recherche d'inhibiteurs de l'IN sont dus, en partie, à l'avancée des connaissances sur l'IN et au développement de tests plus robustes. Les inhibiteurs faux positifs sont fortement réduits avec les tests sur l'enzyme sauvage et des ions divalents  $Mg^{2+}$ . Plusieurs structures cristallographiques de domaines de l'IN sont disponibles pour des études de modélisation moléculaire. Les structures de deux inhibiteurs co-cristallisés avec le domaine central d'une IN (celle du VIH-1 ou celle d'ASV) ont également été résolues (Goldgur et al., 1998; Lubkowski et al., 1998b).

L'ensemble des inhibiteurs de l'IN peut être divisé en deux familles : ceux qui inhibent sélectivement le transfert de brins et les autres, c'est à dire la plupart des composés. Les informations concernant le mécanisme d'action des inhibiteurs sont relativement pauvres pour les molécules de la seconde famille. Parmi les inhibiteurs de celle-ci, les SQ sont très importantes en raison de leur activité antivirale dans la gamme du micromolaire et leur absence de cytotoxicité.



# ***RESULTATS ET DISCUSSION***

---

**PARTIE I**

**Mécanisme d'inhibition de l'IN du VIH-1 par  
des dérivés des Styrylquinoléines (SQ)**

*Ex vivo*, les dérivés des styrylquinoléines bloquent la réplication du VIH-1 dans la gamme du micromolaire. *In vitro*, les SQ inhibent efficacement l'activité de 3'-processing de l'IN avec des valeurs d'IC<sub>50</sub> comprises entre 0.5 et 5µM. Il est donc intéressant de comprendre comment ces molécules inhibent l'IN. Les groupes 7-COOH et 8-OH sont essentiels pour l'activité d'inhibition des SQ (Zouhiri et al., 2000). Des études de docking ont montré que les SQ se fixent près de l'ion Mg<sup>2+</sup> du domaine central d'ASV (Ouali et al., 2000). Les énergies d'interaction de ces complexes sont corrélées aux IC<sub>50</sub> des SQ contre l'IN d'ASV. Par contre, il n'y a pas de corrélation directe entre l'énergie de liaison des SQ à l'ion Mg<sup>2+</sup> et les IC<sub>50</sub> (Ouali et al., 2000). A la suite de ce travail, on pouvait légitimement penser que les SQ inhibent l'IN en entrant en compétition avec l'ADN substrat. Mais, à cause des propriétés d'agrégation de l'IN, il faut utiliser une concentration d'IN supérieure à la constante de dissociation K<sub>d</sub> du complexe de l'IN avec l'ADN substrat. Les méthodes habituelles ne peuvent pas être utilisées et Eric Deprez a du recourir à l'anisotropie de fluorescence pour mesurer la quantité d'ADN fixée à l'IN, ainsi qu'il est exposé dans l'article suivant publié dans Molecular pharmacology. Le travail montre notamment que les SQ préalablement incubées avec l'IN empêchent la fixation de l'ADN sur l'IN mais qu'elles ne peuvent pas déplacer l'ADN viral d'un complexe IN-ADN préalablement formé. Cette propriété est différente de celle des hydrazides salicyliques qui ont pourtant un IC<sub>50</sub> dans la gamme des 10<sup>-5</sup> M seulement. D'autre part, *in vitro*, les SQ inhibent la réaction de transfert de brins ce qui suggère qu'elles se lient à l'IN ou au complexe IN-ADN viral en un autre site. Pour suggérer une localisation de cet autre site, nous avons effectué des dockings des SQ sur le domaine central de l'IN et nous avons regroupé les 100 meilleurs complexes en 3 clusters. D'autre part, la stabilité des complexes a été prouvée en réalisant des dynamiques moléculaires.

Par la suite, il a été montré que les SQ n'interfèrent pas avec l'entrée du virus et que la cible cellulaire de ces drogues est bien l'IN (Bonnenfant et al., 2004). En effet, des mutations dans le gène de l'IN apparaissent lorsque les virus sont cultivés en présence de concentrations croissantes de SQ (doubles mutants V165I V249I ou simple mutant C280Y). Ces mutations amènent à des virus résistants aux SQ quand elles sont introduites dans la séquence sauvage



(Bonnenfant et al., 2004). Au contraire, les SQ sont entièrement actives contre les virus résistants aux inhibiteurs de la transcription inverse et des dicéto-acides.

Les SQ agissent donc sur l'IN *in vivo* alors qu'il est largement pensé que l'interaction IN-ADN dans le PIC empêche les drogues, agissant en tant qu'inhibiteurs compétitifs de l'ADN viral, d'être efficaces dans le contexte cellulaire. D'après nos données, deux hypothèses peuvent expliquer cette contradiction apparente. 1) L'activité *ex vivo* des SQ est reliée à leur capacité d'inhiber la réaction de transfert de brins. 2) Le site actif de l'IN est accessible, au moins transitoirement, pendant les premières étapes de la réplication virale ; la réaction de 3'-processing peut alors être affectée par les SQ.

Des études réalisées *ex vivo* indiquent que les SQ sont plus efficaces aux étapes précoces du cycle de réplication, avant le transport nucléaire et s'accordent donc avec la seconde hypothèse (Bonnenfant et al., 2004). Des analyses de PCR quantitatives montrent que le traitement de cellules infectées par le VIH-1 avec des SQ, réduit le taux d'ADN viral. Ceci suggère, pour la première fois, que les molécules ciblant l'IN peuvent affecter l'accumulation d'ADN durant la transcription inverse. Le motif KRK de l'IN (résidus 186 à 188) a été décrit pour jouer un rôle majeur dans l'efficacité de la transcription inverse (Tsurutani et al., 2000). Deux résidus de ce motif, K186 et K188, sont présents dans le cluster 2 de notre étude de docking.

Il a également été montré que les SQ inhibent spécifiquement et efficacement l'import nucléaire de l'IN *in vitro* sans affecter d'autres chemins d'importations (Mousnier et al., 2004). Cela suggère ainsi que les SQ pourraient également perturber l'interaction de l'IN avec des facteurs cellulaires requis pour son importation nucléaire. Le résidu V165 dont la mutation confère la résistance aux SQ est impliqué dans l'interaction avec le facteur LEDGF/p75 (Cherepanov et al., 2005b) décrit comme étant impliqué dans la localisation nucléaire de l'IN (Maertens et al., 2003). Ce résidu est proche du cluster 2 de notre étude de docking.

## Mechanism of HIV-1 Integrase Inhibition by Styrylquinoline Derivatives in Vitro

Eric Deprez, Sophie Barbe, Maciej Kolaski,<sup>1</sup> Hervé Leh, Fatima Zouhiri, Christian Auclair, Jean-Claude Brochon, Marc Le Bret, and Jean-François Mouscadet

Centre National de la Recherche Scientifique Unité Mixte Recherche 8113, Laboratoire de Biotechnologies et Pharmacologie génétique Appliquée, Ecole Normale Supérieure de Cachan, Cachan, France (E.D., S.B., M.K., C.A., J.-C.B., M.L.B., J.-F.M.); and Bioalliance Pharma, Paris, France (H.L., F.Z.)

Received July 3, 2003; accepted October 10, 2003

This article is available online at <http://molpharm.aspetjournals.org>

### ABSTRACT

Styrylquinoline derivatives (SQ) efficiently inhibit the 3'-processing activity of integrase (IN) with IC<sub>50</sub> values of between 0.5 and 5 μM. We studied the mechanism of action of these compounds in vitro. First, we used steady-state fluorescence anisotropy to assay the effects of the SQ derivatives on the formation of IN-viral DNA complexes independently of the catalytic process. The IC<sub>50</sub> values obtained in activity and DNA-binding tests were similar, suggesting that the inhibition of 3'-processing can be fully explained by the prevention of IN-DNA recognition. SQ compounds act in a competitive manner, with K<sub>i</sub> values of between 400 and 900 nM. In contrast, SQs did not inhibit 3'-processing when IN-DNA complexes were preassembled. Computational docking followed or not by molecular dynamics using the catalytic core of HIV-1 IN suggested a

competitive inhibition mechanism, which is consistent with our previous data obtained with the corresponding Rous sarcoma virus domain. Second, we used preassembled IN-preprocessed DNA complexes to assay the potency of SQs against the strand transfer reaction, independently of 3'-processing. Inhibition occurred even if the efficiency was decreased by about 5- to 10-fold. Our results suggest that two inhibitor-binding modes exist: the first one prevents the binding of the viral DNA and then the two subsequent reactions (i.e., 3'-processing and strand transfer), whereas the second one prevents the binding of target DNA, thus inhibiting strand transfer. SQ derivatives have a higher affinity for the first site, in contrast to that observed for the diketo acids, which preferentially bind to the second one.

Integration of the HIV-1 DNA into the host genome ensures stable maintenance of the viral genome and perpetuation of the virus in the host organism. Therefore, this reaction, catalyzed by integrase (IN), is a key process in the life cycle of the virus. Inhibitors of therapeutic interest for AIDS exist. Inhibitors that act early in the replication cycle target reverse transcriptase and those that act later target protease. The high mutation rate of HIV means that drug resistance is emerging, making it necessary to develop new drugs with different targets (combined therapies). Thus, it is essential to find drugs that target alternative steps of the HIV-1 replication cycle. IN, the third enzyme of the Pol precursor protein, is then an attractive target for novel drugs because of its central role in the HIV-1 life cycle.

IN is a 32-kDa protein (288 amino acids) that consists of three functional domains. The central domain (or core domain, residues 50–210) contains the catalytic triad (DDE) that is essential for enzymatic activity. This domain is flanked by the N-terminal domain (which has a role in the assembly of an active multimeric form of the enzyme) and the C-terminal domain (which has a role in the nonspecific DNA-binding activity). These three domains, protein oligomerization, the integrity of the catalytic triad, and a cationic cofactor such as magnesium are all necessary for IN activity (for review, see Craigie, 2001). The integration process requires two distinct catalytic steps. During the first step, termed *3'-processing*, IN specifically removes two nucleotides from each end of the linear viral DNA. The second step, called *strand transfer*, occurs after the translocation of the viral DNA into the nucleus. During this step, IN transfers both extremities of the viral DNA into the target DNA by a one-step transesterification reaction, resulting in full-site integration. Obviously, DNA-IN complex assembly is a prerequi-

This work was supported in part by the French national agency of research against AIDS (ANRS) and Ensemble Contre le Sida. J.-F.M., J.-C.B. and M.L.B. are funded by the Centre National de la Recherche Scientifique.

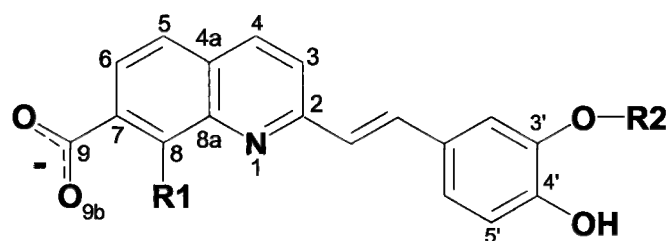
<sup>1</sup> Present address: Department of Theoretical Chemistry, Institute of Chemistry, Szkolna 9, 40-006 Katowice, Poland.

**ABBREVIATIONS:** IN, integrase; SQ, styrylquinoline; DTT, dithiothreitol; LTR, long-terminal repeat; DKA, diketo acids; MD, molecular dynamic(s); L-708,906, 4-(3,5-bis-benzyloxy-phenyl)-2,4-dioxo-butiric acid; 5-CITEP, 1-(5-chloroindo-3-yl)-3-hydroxy-3-(2H-tetrazol-5-yl)-propanone; EDC, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide.

site for the subsequent catalysis. In vitro, another reaction named disintegration can be observed when branched oligonucleotide substrates are used. In first approximation, this reaction represents the reversal of the half-site integration process (Chow et al., 1992). Interestingly, the central domain alone is capable of performing this reaction.

Many IN inhibitors have been characterized in vitro, but the biological activities of these drugs are generally poor, and little information concerning the mechanism of action is available in most cases (for review, see Pommier et al., 2000). It must be noted that most putative IN inhibitors were identified in vitro against  $Mn^{2+}$ - instead of  $Mg^{2+}$ -dependent IN activity. Because IN activity in the physiological context requires  $Mg^{2+}$  as the cationic cofactor, compounds that block the  $Mn^{2+}$ -dependent activity in vitro are less likely to possess antiviral activity. Active inhibitors can be divided into two groups: those that inhibit the 3'-processing reaction (most of the compounds) and those that preferentially inhibit the strand transfer reaction. The well known diketo acids (DKAs) and their more recent derivatives (Hazuda et al., 2000; Zhuang et al., 2003) belong to the second group and have good ex vivo activities against HIV replication, probably because of their ability to inhibit preassembled complexes. A recent study by Pommier and collaborators on a bifunctional DKA showed that the subdivision of IN inhibitors into two distinct groups is not so clear (Marchand et al., 2002). Our study on styrylquinoline (SQ) compounds leads to a similar conclusion.

SQ compounds (Fig. 1) described previously (Mekouar et al., 1998) act on the 3'-processing reaction in vitro, with  $IC_{50}$  values of between 500 nM and 5  $\mu$ M. Several of them are efficient ex vivo with  $IC_{50}$  values in the micromolar range. We demonstrated that the 7-COOH and the 8-OH groups are essential for the inhibition activity (Zouhiri et al., 2000). Nevertheless, the mechanism of action of SQs remains unknown. Do these drugs act at the catalytic level or do they prevent substrate DNA-binding? Such molecular insights about the mechanism of action have been slow to emerge at least in part because of the experimental difficulties to separate the DNA-binding step from the subsequent catalysis reactions. Using a fluorescence-based DNA-binding assay, we show here that SQs prevent recognition of DNA by IN.



Compound	R1	R2	$IC_{50}$ (3'-processing)
KHD161	OH	H	1.2 $\mu$ M
FZ55	OH	CH <sub>3</sub>	2.5 $\mu$ M
FZ117	H	CH <sub>3</sub>	>100 $\mu$ M

Fig. 1. Chemical structures and  $IC_{50}$  of styrylquinoline compounds.

The  $IC_{50}$  values are in good agreement with those obtained in classic inhibition studies of the 3'-processing activity. This competitive mechanism is supported by molecular docking of these drugs on the HIV-1 catalytic core. In contrast, no inhibition was observed when the IN-DNA substrate complex was formed before adding the drug. Nevertheless, SQs were found to be potent inhibitors of the strand transfer reaction performed by the preassembled, preprocessed DNA-IN complexes, suggesting that two SQ-binding modes exist. We used computational docking data to propose several possible SQ-binding sites to take into account experimental results.

## Materials and Methods

**SQ Compounds and Oligonucleotides.** KHD161, FZ55 and FZ117 (Fig. 1) were synthesized as described previously (Zouhiri et al., 2000; Polanski et al., 2002). Oligonucleotides U5B, 5'-GTGTG-GAAAATCTCTAGCAGT-3'; U5B-2, 5'-GTGTGGAAAATCTCTAGCA-3'; U5A, 5'-ACTGCTAGAGATTTTCCACAC-3'; U5B-F, 5'-F-GTGTG-GAAAATCTCTAGCAGT-3' (where F designates the fluorescein probe); LDS32Bp, 5'-p-ACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA-3'; LDS31A, 5'-ACTGCTAGAGATTTTCCACACTGACTAAAAG-3'; T1A, 5'-AGTGAATTAGCCCTTGGTCA-biotin; and T1B, 5'-TGAC-CAAGGGCTAATTCCT-biotin were purchased from Eurogentec (Seraing, Belgium) and further purified on an 18% denaturing acrylamide/urea gel. The phosphate residue present at the 5'-extremity of LDS32Bp was used to attach the target oligonucleotide covalently to the amino-Covalink surface (Merck Eurolab, Fontenay-sous-Bois, France). Terminal bases removed by IN during the 3'-processing reaction are underlined.

**Fluorescence Properties of SQ in the Presence of IN.** Excitation and emission spectra of FZ55 were performed on a Cary-Eclipse fluorescence spectrophotometer (Varian, Palo Alto, CA) thermostated at 25°C. Spectral study of FZ55 with varying concentrations of IN was performed in a Tris buffer (20 mM, pH 7.2) containing 100 mM NaCl. Quenching of the FZ55 steady-state fluorescence by  $Mg^{2+}$  was analyzed by a Stern-Volmer plot (i.e., by plotting  $F_0/F-1$  as a function of  $Mg^{2+}$  concentration, where  $F_0$  and  $F$  represent the fluorescence intensities in the absence and in the presence of quencher, respectively). Time-resolved fluorescence experiments for lifetime determination of FZ55 in the absence or in the presence of  $Mg^{2+}$  (60 mM) were performed by single-photon counting as described previously (Deprez et al., 2000, 2001) in the same conditions used as for the steady-state study (25°C, 30  $\mu$ M FZ55). Excitation and emission wavelengths were 320 and 538 nm ( $\Delta\lambda = 15$  nm), respectively, with the emission polarizer oriented at the magic angle (54.7°). No photobleaching was observed during the acquisition and photons were collected until the total count reached 12 to 15 million.

**Purification of IN and Standard Activity Assays.** IN was expressed and purified as described previously (Leh et al., 2000). For activity assays, 100 pmol of either U5B (3'-processing substrate) or U5B-2 (strand transfer substrate) were radiolabeled using T4 polynucleotide kinase and 50  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P]ATP (3000 Ci/mmol). The T4 kinase was then heat-inactivated, and unincorporated nucleotides were removed by filtration through a Sephadex G-25 column (Amersham Biosciences AB, Uppsala, Sweden). NaCl was added to a final concentration of 100 mM and the complementary unlabeled strand U5A was added. The mixture was incubated at 90°C for 3 min and allowed to anneal by slowly cooling to room temperature. The 3'-processing and strand transfer reactions were performed using 1.5 nM U5A/U5B and 10 nM U5A/U5B-2 double-stranded oligonucleotide substrates, respectively, in a buffer containing 20 mM HEPES, pH 6.8, 7.5 mM  $MgCl_2$ , and 2 mM DTT in the presence of various concentrations of IN. The reaction mixture was incubated for 1 h at 37°C and stopped by phenol-chloroform extraction. DNA products were precipitated with ethanol, dissolved in a loading buffer contain-

ing 7 M urea and 10 mM EDTA and subjected to electrophoresis on an 18% denaturing acrylamide/urea gel. Gels were dried and the reaction products were visualized using a STORM PhosphorImager (Amersham Biosciences).

**Immobilized IN-DNA Complex Activity Assay: Microtiter Plate Assay.** The strand transfer activity of the preformed complex was assayed in a 96-well microtiter plate coated with the U5-2 substrate as described previously (Hazuda et al., 1994). Briefly, the DNA substrate was coupled to the Covalink NH plates as follows: 75  $\mu$ l of 100 nM double-stranded oligonucleotides LSD32Bp/LSD31A were placed into each well of the Covalink NH plates. Twenty-five microliters of 0.2 M EDC dissolved in 10 mM 1-MeIm7 was then added, and plates were incubated at 50°C for 5 h. Free oligonucleotides were removed and unreacted amino groups were deactivated by adding 10 mM citraconic anhydride. After 30 min at 50°C, the wells were finally washed with 5 $\times$  standard saline citrate and 0.25% SDS. IN (100 nM) was then incubated with the substrate for 20 min at 25°C. The supernatant was discarded and 130 nM of the biotinylated target oligonucleotides were added in a final volume of 40  $\mu$ l (the target double-stranded DNA was obtained by annealing the two complementary strands T1A and T1B in a 20 mM Tris-HCl buffer, pH 7.2, containing 0.1 M NaCl). The reaction was allowed to proceed for 1 h at 37°C and stopped by the addition of 0.1 M NaCl and 50 mM EDTA. The plates were washed twice for 5 min in washing buffer (0.05% Tween 20 and 10  $\mu$ g/ml bovine serum albumin in phosphate-buffered saline). Extravidine phosphatase (Sigma; 100  $\mu$ l of a 1/10,000 dilution in washing buffer) was added and incubated for 15 min at room temperature. The plates were washed twice in a 0.1 M Tris buffer pH 9 containing Tween + bovine serum albumin, 0.1 M NaCl, 50 mM MgCl<sub>2</sub>. Finally, 100  $\mu$ l of AP substrate were added and the reaction was allowed to proceed for 1 h at 37°C. Luminescent signal was recorded on a microtiter plate photometer (Dynatech Labs, Chantilly, VA).

**DNA-Integrase Binding Assay.** The interaction between IN and the fluorescein-labeled oligonucleotide was detected by fluorescence anisotropy. The double-stranded oligonucleotide was obtained by mixing equimolar amounts of complementary strands U5B-F and U5A in a 20 mM Tris-HCl buffer, pH 7.2, containing 100 mM NaCl. The mixture was heated to 85°C for 5 min and allowed to anneal by slowly cooling to room temperature. The steady-state anisotropy measurements were recorded on a Beacon 2000 instrument (PanVera, Madison, USA). To determine the apparent  $K_d$  value, fluorescein-labeled DNA (1 nM) was incubated with increasing concentrations of IN in a 20 mM Tris buffer, pH 7.2, containing 1 mM DTT, 25 mM NaCl, and 5 mM MgCl<sub>2</sub> for 20 min at 25 or 37°C. Steady-state anisotropy was then recorded in a cell thermostated at 25 or 37°C. The fractional saturation  $Y$  was calculated as  $(A - A_f)/(A_b - A_f)$ , where  $A_b$  and  $A_f$  represent the bound and free DNA anisotropy, respectively, because no significant concomitant change in the fluorescence intensity was observed. For cooperative binding, the Hill number,  $n$ , was calculated from the plot LOG  $(Y/1 - Y) = F(\text{LOG}(\text{IN}))$ . The concentration of IN required to titrate the DNA to half saturation was  $\text{IN}_{50}$  or apparent  $K_d$  and was equal to  $K_d^{1/n}$ .

$\text{IC}_{50}$  values were determined by measuring  $Y$  in the presence of increasing concentrations of drugs. IN was preincubated with drugs for 10 min at 25 or 37°C in a 20 mM Tris buffer, pH 7.2, containing 25 mM NaCl, 5 mM MgCl<sub>2</sub>, and 1 mM DTT. Complexes with DNA were then allowed to form at 25 or 37°C after the addition of double-stranded fluorescein-labeled oligonucleotides. The midpoint or  $\text{IC}_{50}$  of the inhibition curve represents the total drug concentration giving 50% inhibition ( $Y = 0.5$ ) and was used to estimate the  $K_i$  value.  $K_i$  was determined by use of the following equations, which are based on a competitive inhibition model. At equilibrium,



with

$$K_d = \frac{[\text{IN}]^n \cdot [\text{DNA}]}{[\text{IN-DNA}]} \quad (1)$$

and

$$K_i = \frac{[\text{IN}] \cdot [i]}{[\text{IN-i}]} \quad (2)$$

where IN, DNA, and  $i$  are the concentrations of free integrase, DNA substrate, and inhibitor, respectively,  $\text{IN} \cdot \text{DNA}$  and  $\text{IN} \cdot i$  designate the complexes, and  $n$  is the average number of interacting sites.

The concentrations of total IN, DNA substrate, and inhibitor are calculated as follows,

$$[\text{IN}_T] = [\text{IN}] + n \cdot [\text{IN-DNA}] + [\text{IN-i}] \quad (3)$$

$$[i_T] = [i] + [\text{IN-i}] \quad (4)$$

$$[\text{DNA}_T] = [\text{DNA}] + [\text{IN-DNA}] \quad (5)$$

using eqs. 2 and 4,

$$\frac{1}{[\text{IN}]} = \frac{[i_T]}{K_i \cdot [\text{IN-i}]} - \frac{1}{K_i} \quad (6)$$

using eqs. 1 and 5,

$$\frac{1}{[\text{IN}]} = \left( \frac{1 - Y}{K_d \cdot Y} \right)^{\frac{1}{n}} \text{ with } Y = \frac{[\text{IN-DNA}]}{[\text{DNA}_T]} \quad (7)$$

Eqs. 6 and 7 yield:

$$\left( \frac{1}{K_d} \cdot \left( \frac{1}{Y} - 1 \right) \right)^{\frac{1}{n}} = \frac{1}{K_i} \cdot \left( \frac{[i_T]}{[\text{IN-i}]} - 1 \right) \quad (8)$$

Substitution of eq. 3, i.e.,

$$[\text{IN} - i] = [\text{IN}_T] - \left( \frac{K_d - Y}{1 - Y} \right)^{\frac{1}{n}} - n \cdot Y \cdot [\text{DNA}_T]$$

into eq. 8 yields:

$$\left( \frac{1}{K_d} \cdot \left( \frac{1}{Y} - 1 \right) \right)^{\frac{1}{n}} = \frac{1}{K_i} \cdot \left( \frac{[i_T]}{[\text{IN}_T] - \left( \frac{K_d \cdot Y}{1 - Y} \right)^{\frac{1}{n}} - n \cdot Y \cdot [\text{DNA}_T]} - 1 \right) \quad (9)$$

Eq. 9 is equivalent to:

$$[i_T] = \left( K_i \cdot \left( \frac{1}{K_d} \cdot \left( \frac{1}{Y} - 1 \right) \right)^{\frac{1}{n}} + 1 \right) \cdot \left( [\text{IN}_T] - \left( \frac{K_d \cdot Y}{1 - Y} \right)^{\frac{1}{n}} - n \cdot Y \cdot [\text{DNA}_T] \right) \quad (10)$$

Assuming saturation conditions in the absence of inhibitor (i.e.,  $Y = 1$ ), eq. (10) can be simplified at the  $\text{IC}_{50}$  value where  $Y = 0.5$ .

$$K_i = \left( \frac{\text{IC}_{50}}{[\text{IN}_T] - K_d^{\frac{1}{n}} - \frac{1}{2} \cdot n \cdot [\text{DNA}_T]} - 1 \right) \cdot K_d^{\frac{1}{n}} \quad (11)$$

The independent measurements of  $\text{IN}_{50}$  or  $K_d^{1/n}$  and  $\text{IC}_{50}$  allowed the calculation of  $K_i$  using eq. 11.

### Computer Studies and Protein Conformations

The computer studies used two models for the molecular target of the drug. The first one, hereafter called  $\alpha$ , was the catalytic core

(residues 50–212) of the HIV-1 IN containing the single mutation F185H with an  $Mg^{2+}$  cation in its active site. The conformation was the frame obtained after 3.4275 ns of production of a 4-ns molecular dynamics simulation (Laboulais et al., 2001). This frame was selected because the Glu152 carboxylic group pointed to the other two triad members Asp64 and Asp116 in a geometry that is compatible with the catalysis. The flexible loop (139–151) is not as stacked onto the catalytic core as found in the structure 1QS4 (Goldgur et al., 1999) but closely resembles the structures described by Maignan et al. (1998). The model was then stripped of the chloride anion. Two submodels were derived. In submodel  $a_0$ , no water molecules were retained, whereas in submodel  $a_4$ , the four water molecules coordinated to  $Mg^{2+}$  were explicitly retained. The other model, hereafter called  $b$ , was the 55–209 fragment of the HIV-1 IN catalytic core containing the mutations W131E, A133G, and F185K. The conformation was taken from the coordinates of monomer A in the crystal structure 1QS4 (Goldgur et al., 1999). Missing residues were forced to have the same backbone dihedrals as in the crystal structure 1BIS (Goldgur et al., 1998). The model was then stripped of all water molecules, giving the submodel  $b_0$ ; in submodel  $b_4$ , the four water molecules coordinated to  $Mg^{2+}$  were explicitly retained.

### Docking

The target ( $a_0$ ,  $a_4$ ,  $b_0$ , or  $b_4$ ) was digitized on an orthorhombic grid containing  $256 \times 256 \times 256$  meshes of size  $0.305 \times 0.305 \times 0.297 \text{ \AA}^3$ . The structure of the drug KHD161 was optimized as described previously (Ouali et al., 2000) and the tautomeric form with the 7-COO<sup>-</sup> and 8-OH groups was used. The drug was rotated by  $12^\circ$  increments and for each of these 12,660 orientations, the translations giving the best overlap were stored when the electrostatic drug-target interaction energy was favorable. As described previously (Ouali et al., 2000), the best 1000 overlapping complexes and the most favorable 200 complexes for the pure electrostatic interaction were minimized, with both drug and target being optimized. During the minimization, the aqueous medium was implicitly represented through a sigmoidal distance-dependent dielectric constant (Lavery et al., 1986). The 1200 complexes were sorted according to the drug-protein interaction energy, which is the sum of the van der Waals and Coulomb contributions.

### Clustering

The 100 most favorable complexes were stored and the  $100 \times 100$  root-mean-squared Euclidean distances of the drug atom coordinates were further transformed into a dendrogram representing the hierarchy of categories of clusters. The hierarchical and agglomerative processes work as follows. Starting from a set of  $n = 100$  clusters, one per drug position and orientation, a sequence of clusterings was grown until all  $n$  patterns were in a single cluster. At each step, the two closest clusters were reunited. The distance separating the remaining clusters from the newly created cluster was the weighted average of the distances to the members of the new cluster (Labou-

lais et al., 2002). The results of the clustering analysis are shown in Fig. 3 and Table 2.

### Molecular Dynamics

The catalytic core-drug complexes obtained using an implicit solvent representation at the end of the docking (conformation  $a_0$ ) had the net coulombic charge of one proton. A chloride anion was then added and the whole system was embedded in a  $52.3 \times 52.3 \times 52.3 \text{ \AA}^3$  box filled with 3638 transferable intermolecular potential (TIP3P) water molecules. The SANDER module of AMBER (P. Kollman, University of California, San Francisco) was used to simulate the dynamics. Long-range coulombic interactions were calculated using the particle-mesh technique for Ewald sums (Ewald, 1921) with a cut-off of 10  $\text{\AA}$  and a  $50 \times 50 \times 50$ -point grid. No correction was applied for the neglected long-range van der Waals interactions because they were expected to be small. Covalent bonds containing a proton were constrained to their equilibrium length by use of the SHAKE algorithm (Ryckaert et al., 1977). The elementary integration time step was 2 fs. Before dynamics could be reliably produced, the system was prepared by two heating steps. In the first step, only the water molecules were allowed to move and the temperature was increased from 25 to 300 K in 25-K steps during 14.6 ps. At the end of this first step, water molecules were cooled to 10 K. In a second step, the temperature of the whole system was increased from 10 to 300 K over 18 ps. The velocities were then equilibrated for 13 ps at 300 K, before starting dynamics production. Frames were then recorded at 0.1-ps intervals. The total duration of the molecular dynamics simulations was 2 ns.

## Results

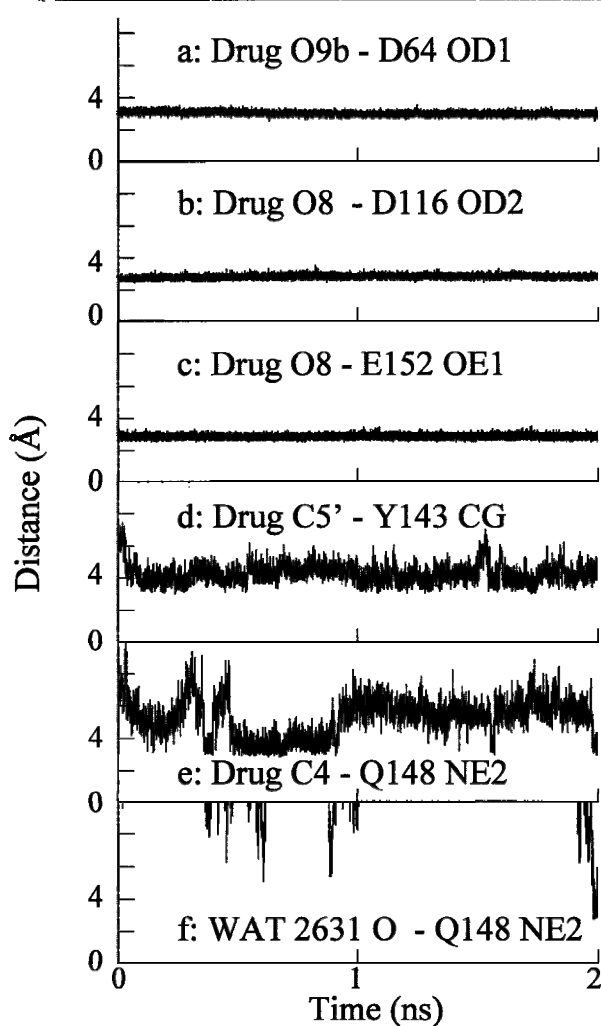
**Docking of KHD161 to the Catalytic Core Domain of IN.** To get some clue about the location of the binding site on the HIV-1 IN catalytic core, we carried out a number of docking simulations with the KHD161 compound (Fig. 1). The first step of the procedure was based on rigid docking and the second on an MD approach. After rigid docking and minimization using an implicit representation of the solvent (model  $a_0$ ), the best target-drug complexes were selected according to their interaction energies (coulombic and van der Waals). For all the best 17 complexes, KHD161 was found near the essential catalytic cofactor  $Mg^{2+}$ . In all cases, at least one oxygen atom of the drug was within 3  $\text{\AA}$  of the  $Mg^{2+}$  ion, and the drug-target interaction energies ranged from  $-115$  to  $-87$  kcal/mol. The best two complexes (d1 and d2) were used as starting conformations for MD. Some features of the dynamics d1 and d2 are summarized in Table 1.

In d1,  $Mg^{2+}$  bound to O8 and O9b of the quinolic ring. The catechol ring was stacked over the phenol group of Tyr143. Figure 2 shows how some distances changed over time in the

TABLE 1  
Parameters of HIV-1 catalytic core-KHD161 complexes according to computational docking and molecular dynamics studies

	Complex	
	d1	d2
Interaction energy (kcal/mol) (implicit solvent representation) before MD	-114.97	-97.4
Average energy in the presence of 3638 explicit water molecules (kcal/mol). Fluctuations are indicated in parentheses		
From 0 to 1 ns	-31,494.4 (353.8)	-31,479.8 (317.9)
From 1 to 2 ns	-31,513.5 (45.2)	-31,507.1 (46.8)
Contacting residues after docking and minimization and before MD	Thr66, Asp116, Phe139, Gly140, Ile141, Pro142, Tyr143	Asp64, Thr66, His67, Leu68, Asp69, Glu152, Lys159
Contacting residues after 2 ns of dynamic simulation	Asp64, Asp116, Tyr143, Gln148, Glu152	Asp64, Cys65, Thr66, Glu152, Asn155

subsequent MD. Time 0 is the beginning of the production phase (the preparation phase is not shown). Although some motion was observed, the drug remained in the same position as at the beginning of the dynamics. For instance, the average distances separating the oxygen atoms O9b (Fig. 2a) and O8 (not shown) from the Asp64 of the catalytic triad Asp64-Asp116-Glu152 were the same from the start of the docking step until the end of the dynamics simulation. Some changes occurred only during the preparation phase. Residues Asp116 and Glu152, which were 4 and 7 Å from the drug oxygen atoms O8 and O9b, moved nearer to these atoms during the preparation phase, but remained at the same distance throughout the production phase (Fig. 2, b and c). Other distances fluctuated much more, such as those separating the drug from Tyr143 and Gln148 (Fig. 2, d and e). During the simulation, water molecule 2631 came into close contact with Gln148 and the drug several times, but only transitorily. At the end of the simulation (2 ns), this water molecule joined Gln148 to the quinoline proton H4, Gln148 being close to the quinoline proton H5. The pattern shown in



**Fig. 2.** Changes in selected distances (Ångstroms) with time (nanoseconds) in the d1 dynamics. The time scale started at the beginning of the production phase. a, drug O9b-Asp64 OD1; b, drug O8-Asp116 OD2; c, drug O8-Glu152 OE1; d, drug C5'-Tyr143 CG; e, drug C4-Gln148 NE2; f, water 2631 O-Gln148 NE2.

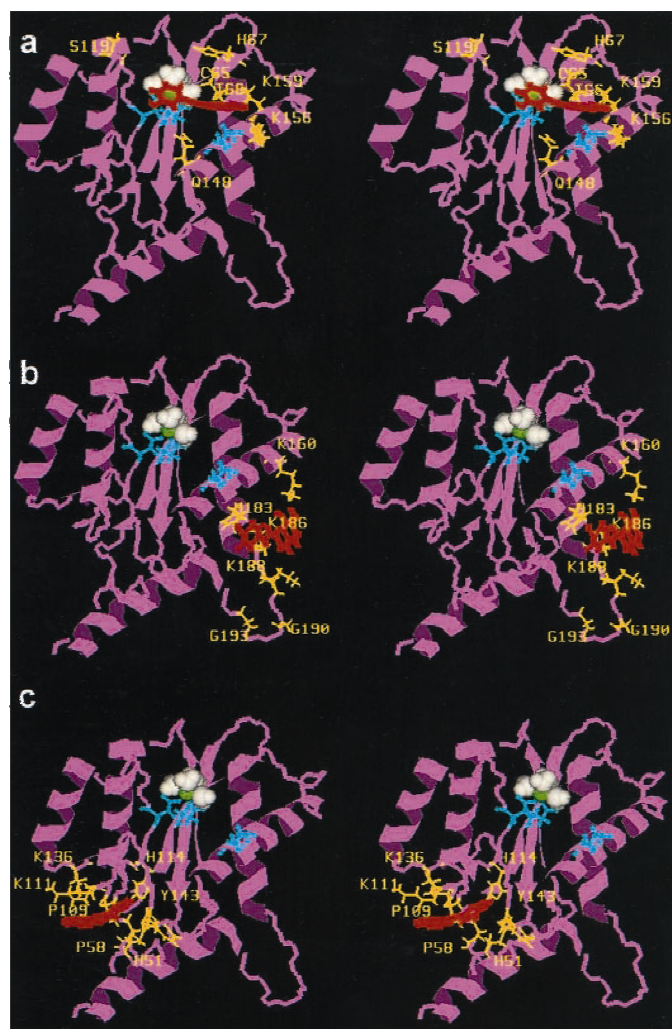
Fig. 2f suggests that the water molecule is not stable in the active site.

The conformation d2 was less favorable. In this case, the drug bound to  $Mg^{2+}$  through the hydroxyl groups of the catechol ring. At the end of the MD, the catechol ring remained near  $Mg^{2+}$ , but the quinoline ring protruded from the protein structure in the solvent (not shown). This result was not observed with d1. Our simulation results indicate that d1 is characterized by more favorable interaction energy after the docking and is also more stable in the active site during the dynamics. During both dynamics d1 and d2, no water molecule formed a steady interaction with both the drug and the protein. This is in contrast to the 1-(5-chloroindol-3-yl)-3-hydroxy-3-(2H-tetrazol-5-yl)-propanone complex (5-CITEP) (Ni et al., 2001). This explains why the interaction energies calculated using an implicit representation of the solvent and the average energies calculated over 1 ns were of the same order. Only the energy fluctuation was much lower during the second nanosecond than during the first (see Table 1). For both d1 and d2, some key residues found to form contacts with KHD161, for instance Tyr143, Gln148, and Lys159 (see Table 1), have previously been described to form contacts with DNA (Jenkins et al., 1997; Esposito and Craigie, 1998), suggesting that the location of SQ within the active site is incompatible with DNA-binding. The results were not significantly different when  $b_0$  was used as a target (structure 1QS4) in the docking and minimization processes instead of  $a_0$ .

In the simulations described above (model  $a_0$  or  $b_0$ ), the protein was completely stripped of all its water molecules, allowing the drug to form direct contacts with  $Mg^{2+}$ . This overestimates the electrostatic interaction. It is more likely that some water molecules are sandwiched between the drug and the cation. Additional docking simulations were then carried out, with four water molecules surrounding  $Mg^{2+}$ . The introduction of four explicit water molecules into the models (models  $a_4$  and  $b_4$ ) had the following general consequences compared with simulations using  $a_0$  and  $b_0$  (note that no major difference was observed between  $a_4$  and  $b_4$ , as was the case for  $a_0/b_0$ ): 1) Because of the increase in the shortest distance separating the drug from  $Mg^{2+}$  (caused by the presence of water), drug-protein interaction energies were less favorable. 2) The hierarchy of interaction energies for the different binding sites was not similar between the  $a_4/b_4$  models and the  $a_0/b_0$  models. For instance, a binding localization came up in the 3rd rank in model  $a_4$  compared with in the 18th rank in model  $a_0$ . 3) Although some differences were observed when  $a_4/b_4$  were compared with  $a_0/b_0$ , the drug was still in contact with the hydrated  $Mg^{2+}$  in the most favorable localization even if the presence of water molecules decreased the role of  $Mg^{2+}$  in the stabilization of the drug-protein complex. 4) Energy in the  $a_4$  or  $b_4$  model was not as widespread as in the  $a_0$  or  $b_0$  model.

Because no clear differences were observed between  $a_4$  and  $b_4$ , the best 100 complexes were pooled and classified by clustering analysis (see *Materials and Methods*) without any energy considerations to identify different sites. Three main sites were identified. Stereoviews of representative complexes for each cluster are shown in Fig. 3, and Table 2 summarizes the main features of the various clusters (e.g., population, energy statistics, and contacting residues). The contacting residues of cluster 1 (Fig. 3A) overlap the catalytic

triad. Note that d1 and d2 as obtained after rigid docking and MD belong to this cluster. Cluster 2 is actually made of a succession of overlapping sites and is spread over a large region of the protein surface (Fig. 3B) whereas a small number of conformations form the cluster 3 (Fig. 3C). The influence of  $Mg^{2+}$  is only significant in cluster 1 and explains why the average interaction energy was better for this cluster. As



**Fig. 3.** Stereoview of one SQ-IN complex conformation belonging to each cluster. The different clusters were obtained as described under *Materials and Methods*. Only a typical conformation is shown among various orientations which were clustered together. a, cluster 1; b, cluster 2; c, cluster 3. The detailed clustering analysis is shown in Table 2. Magenta, protein; green,  $Mg^{2+}$ ; white, water; red, drug. The catalytic triad (Asp64-Asp116-Glu152), is shown in cyan (sticks and balls), and contacting residues are in yellow.

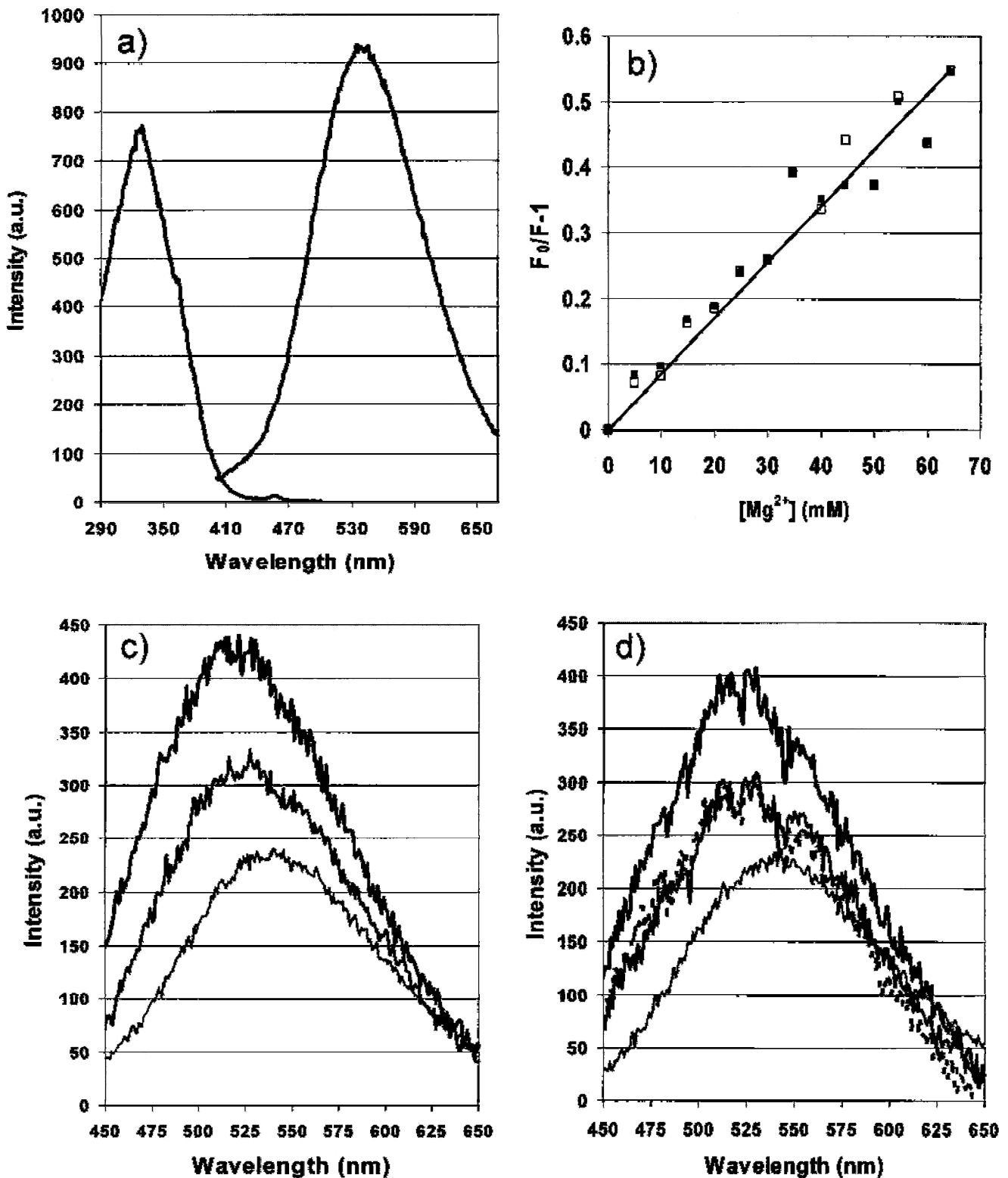
expected, the negative drug bound to lysine-rich regions. The lysine residues can be very roughly clustered into two groups: group 1 contains Lys156, Lys159, Lys160, Lys186, and Lys188, whereas group 2 contains Lys111 and Lys136. From this point of view, clusters 1 and 2 belong to group 1 whereas cluster 3 belongs to group 2. Residues Lys156 and Lys159 in cluster 1 and residue Lys136 in cluster 3 have been previously found to be involved in DNA binding (Mazumder et al., 1996; Jenkins et al., 1997). Besides the lysine residues, Gln148 and Tyr143, contained in clusters 1 and 3, respectively, are also key residues for DNA binding, as shown by photo-cross-linking experiments (Esposito and Craigie, 1998). Another residue in cluster 1, Ser119, is important for DNA binding because mutation of this residue affects target site selection (Harper et al., 2001). No residues from cluster 2 are known to be essential for DNA binding.

**Influence of IN on the Fluorescence Emission Spectrum of FZ55.** A fluorescence study of SQ in the absence or in the presence of IN was carried out to get experimental support of the existence of IN-SQ complexes. As found previously (Ouali et al., 2000; Zouhri et al., 2000) and confirmed in this study (see below), the *in vitro* inhibition properties of SQ derivatives originate in the simultaneous presence of both C7-COOH and C8-OH groups in the quinoline subunit and are less sensitive to modifications at the C3' or C4' positions. Accordingly, KHD161 and FZ55 are characterized by very similar inhibition properties. For stability reasons (Burdujan, 2002), FZ55 was preferred over KHD161 in the following spectral study. Excitation and emission spectra of FZ55 are shown in Fig. 4a. Excitation maximum is centered at 330 nm, whereas emission maximum is centered at 538 nm. During our study, we found that  $Mg^{2+}$  (the essential cofactor for IN activity) acts as a quencher of fluorescence for FZ55, suggesting that SQs could not be a competitive inhibitor but simply a chelation agent. In other words,  $Mg^{2+}$  could be displaced from the active site in the presence of SQs, and the inhibition process could then occur independently of the drug binding within the active site. We then performed Stern-Volmer analysis to quantify the quenching process (Fig. 4b). The Stern-Volmer plot was linear and similar slopes were obtained at two different excitation wavelengths, 320 and 345 nm. Note that quenching can be caused by two distinct phenomena at the molecular level, dynamic or static, each leading to the overall decrease of fluorescence intensity and linear Stern-Volmer plot. Only in the case of static quenching does a stable complex exist that is nonfluorescent. In contrast, dynamic quenching is caused by a collisional process that occurs during the lifetime of the excited state. This means that the lifetime is not influenced in the case of static quenching, whereas it decreases in the case of collisional quenching. Because the two phenomena can be distin-

TABLE 2

Clustering of the best 100 complexes obtained after docking of KHD161 to the  $\alpha_4$  and  $\beta_4$  IN catalytic core

Cluster	Population	Average Energy	Lowest Energy	Highest Energy	Contacting residues (<5 Å)	No of Lysine Groups
		<i>kcal/mol</i>				
1	19	-70.0	-92.8	-61.6	Asp64 Cys65 Thr66 His67 Asn92 Asp116 Asn117 Gly118 Ser119 Asn120 Gln148 Glu152 Asn155 Lys156 Lys159	1
2	42	-67.7	-77.1	-61.4	Glu157 Lys160 Ile161 Gln164 His183 Lys186 Lys188 Gly189 Gly190 Gly193 Tyr194 Ser195 Ala196 Arg199	1
3	7	-65.9	-70.9	-62.9	His51 Ser57 Pro58 Gly59 Ile60 Pro109 Lys111 Thr112 His114 Lys136 Pro142 Tyr143	2



**Fig. 4.** Fluorescence properties of free and IN-bound FZ55. a, excitation and emission spectra of FZ55 (30  $\mu\text{M}$ ). Excitation ( $\lambda_{\text{em}} = 538 \text{ nm}$ ) and emission ( $\lambda_{\text{exc}} = 345 \text{ nm}$ ) spectra were obtained at 25°C. Emission and excitation slit widths were 5 nm. b, Stern-Volmer representation for the quenching of FZ55 by  $\text{Mg}^{2+}$ .  $F_0$  and  $F$  are the intensities of emission fluorescence at 538 nm in the absence and in the presence of  $\text{Mg}^{2+}$ , respectively. FZ55 concentration, 30  $\mu\text{M}$ .  $\square$ , excitation at 320 nm.  $\blacksquare$ , excitation at 345 nm. c, influence of IN on the emission spectrum of FZ55 ( $\lambda_{\text{exc}} = 345 \text{ nm}$ ). Concentration of FZ55 was 3  $\mu\text{M}$  in a Tris buffer (20 mM, pH 7.2) containing 100 mM NaCl and 5 mM  $\text{Mg}^{2+}$ . Light gray, no IN; dark gray, IN 0.9  $\mu\text{M}$ ; black, IN 3  $\mu\text{M}$ . d, influence of  $\text{Mg}^{2+}$  on the emission spectrum of IN-bound FZ55 ( $\lambda_{\text{exc}} = 345 \text{ nm}$ ). Light gray, FZ55 3  $\mu\text{M}$  without IN in the presence of 5 mM  $\text{Mg}^{2+}$ ; dark gray, 3  $\mu\text{M}$  FZ55 with 3  $\mu\text{M}$  IN in the absence of 5 mM  $\text{Mg}^{2+}$ ; black, 3  $\mu\text{M}$  FZ55 with 3  $\mu\text{M}$  IN in the presence of 5 mM  $\text{Mg}^{2+}$ ; dashed line, 3  $\mu\text{M}$  FZ55 with 3  $\mu\text{M}$  IN in the presence of 1 mM EDTA. NaCl concentration was 100 mM.



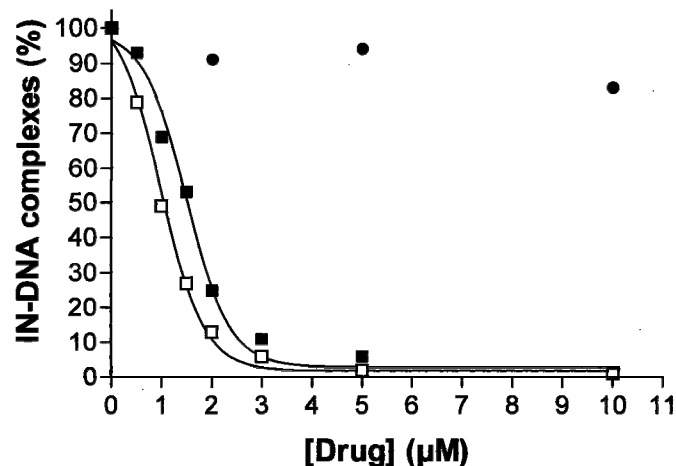
guished by lifetime measurements, we then performed time-resolved fluorescence experiments in the absence or in the presence of 60 mM  $Mg^{2+}$  ( $\lambda_{exc} = 320$  nm). The average lifetimes as well as lifetime distributions of FZ55 were unchanged in both conditions ( $\tau_1 = 18 \pm 2$  ps (33%)/ $\tau_2 = 70 \pm 11$  ps (5%)/ $\tau_3 = 242 \pm 4$  ps (54%)/ $\tau_4 = 490 \pm 6$  ps (8%)/ $\tau_{average} = 175$  ps) demonstrating that the  $Mg^{2+}$ -dependent quenching of FZ55 is static. The slope of the Stern-Volmer plot shown in Fig. 4b corresponds then to an association constant  $K_a = 8.6$   $M^{-1}$  (i.e.,  $K_d = 120$  mM). This value definitively ruled out the idea that IN inhibition by SQs, characterized by submicromolar or micromolar  $IC_{50}$  in the presence of 5 mM  $Mg^{2+}$ , could be simply caused by the chelation of  $Mg^{2+}$ . It is then unlikely that SQs inhibit IN by displacing the essential  $Mg^{2+}$  cofactor from the active site.

Next, we studied the influence of IN-binding on the fluorescent properties of FZ55. As shown in Fig. 4c, increasing concentrations of IN (entire protein) enhanced fluorescence intensity and blue-shifted the maximum of emission. This phenomenon is less pronounced in the absence of  $Mg^{2+}$  or in the presence of EDTA (Fig. 4d) indicating that first, FZ55 actually binds to IN and second, the stability of the complex is dependent on the presence of  $Mg^{2+}$ , in accordance with the docking results. Similar conclusions had been previously published in the case of the catalytic core (Burdujan, 2002). It is then very likely that the inhibition property of SQs is related to their capabilities to bind within the protein.

**Inhibitory Effect Is Caused by the Ability to Impair DNA Binding.** Even though we do not yet know how the IN active unit is organized, recent molecular docking of a dinucleotide onto the IN core domain has revealed two possible nucleic acid-binding sites on a single monomer (Perryman and McCammon, 2002). One of these sites encompasses a previously described nucleotide-binding site that involves three lysine residues of the catalytic core domain, Lys156, Lys159, and Lys160 (i.e., cluster 1 in our study) (Jenkins et al., 1997; Drake et al., 1998; Jing et al., 2000). These lysine residues are involved in the binding of viral DNA (i.e., donor DNA) (Jenkins et al., 1997). The second site is a putative binding site for the target or cell DNA (i.e., acceptor DNA). Our molecular docking study strongly suggested that the location of the SQ drug within the active site mainly disturbs the binding of viral DNA to the core domain of the IN. To address this hypothesis, we used steady-state fluorescence anisotropy to investigate the effects of SQs on the binding of IN to a short double-stranded substrate DNA (21-mer). One major difficulty encountered when studying of DNA binding properties of HIV-1 IN is the separation of the elementary DNA-binding process from the subsequent activity steps (i.e., 3'-processing and strand transfer). It is then a difficult task to separate the molecular and physicochemical effects on these different steps by only measuring the overall activity by classic enzymatic assays. Note that the enzymatic assays involving IN described in the literature do not correspond to Michaelis-Menten conditions, meaning that " $k_{cat}$ " and " $V_{max}$ " values should be interpreted with caution. Steady-state anisotropy measurements are based on the principle of depolarization of light by a fluorescent probe caused by the rotational diffusion that occurs between excitation and emission. The degree of depolarization is related to rotational motion and flexibility properties of the molecular edifice that contains an intrinsic or extrinsic fluorophore. Low anisotropy is

associated with a small edifice and high local flexibility, whereas high anisotropy is associated with a large edifice and low local flexibility. To study the DNA-IN interaction, fluorescein was attached to the extremity of a double-stranded oligonucleotide (21-mer) that mimicked the U5 viral LTR terminus. Fluorescein was attached through a 6-carbon linker to the opposite end from the processing site to prevent steric conflicts between IN and the fluorophore [similar 3'-processing and integration activities were detected when labeled or unlabeled DNA was used as a substrate (data not shown)].

The fluorescein-labeled oligonucleotides, free in solution, were characterized by a low anisotropy value, 0.048 and 0.060 at 37 and 25°C, respectively. When IN bound to the oligonucleotide, the steady-state fluorescence anisotropy increased. The upper limits of anisotropy values were 0.232 and 0.240 at 37 and 25°C, respectively, yielding a maximum amplitude of about 0.180. This value of the steady-state anisotropy may be explained by a combination of two phenomena, the slower global tumbling of the IN-DNA complex and the more restricted local motion of fluorescein. As shown in Fig. 5, this value decreased when the interaction was tested in the presence of KHD161. Data analysis of the KHD161 curves (■, □) yielded  $IC_{50}$  values of  $1.50 \pm 0.13$   $\mu M$  and  $0.98 \pm 0.09$   $\mu M$  at 37 and 25°C, respectively. The  $IC_{50}$  for the DNA-binding step at 37°C is compatible with the value found in activity tests (3'-processing reaction) at the same temperature in the presence of similar IN and substrate concentrations ( $IC_{50} = 1.2$   $\mu M$ ). In contrast, concentrations of up to 10  $\mu M$  of the structurally related compound FZ117 had no significant inhibitory effect [(DNA-IN interactions were weakened only above 50  $\mu M$  FZ117 (data not shown)]. Again, this result is in accordance with the activity tests, because FZ117 had no effect on IN activity. Because the experimental conditions used in the DNA-binding assays were similar to those used in enzymatic assays, the  $IC_{50}$  values obtained in both cases (1.2–1.5  $\mu M$ ) can be compared. Consequently,



**Fig. 5.** Percentage of integrase-oligonucleotide complexes as a function of drug concentration. □, KHD161 at 25°C; ■, KHD161 at 37°C; ●, FZ117 at 25°C. Steady-state fluorescence anisotropies were measured at 25°C or 37°C as described in the text. The values were normalized based on the anisotropy amplitude values (complex minus free oligonucleotide). One hundred percent corresponds to an anisotropy amplitude of 0.180 (see text). IN concentration, 120 nM. The fit of the KHD161 curves using a sigmoidal shape gave  $IC_{50}$  values of  $0.98 \pm 0.09$   $\mu M$  and  $1.50 \pm 0.13$   $\mu M$  at 25 and 37°C, respectively.

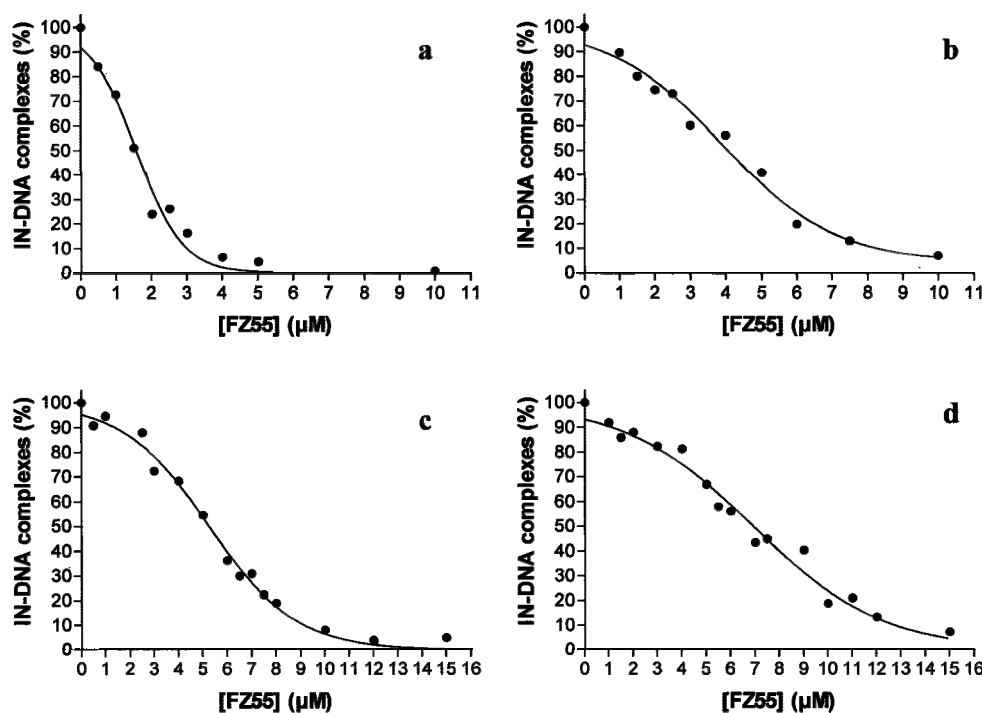
DNA-binding inhibition probably accounts for the whole inhibition of enzyme activity, thus suggesting a competitive mechanism of action. Unlike when IN and KHD161 were preincubated before adding the DNA, the IN-DNA complex was not disrupted when IN and DNA were preincubated before adding the drug (data not shown).

To test the hypothesis of a competitive mode of action, we determined  $IC_{50}$  values at different protein concentrations, because the theory suggests that  $IC_{50}$  is related to IN concentration. Moreover, if this is case,  $K_i$  can be estimated from eq. 11. Note that the well known Cheng and Prusoff model (1973), which estimates  $K_i$  from  $IC_{50}$  values, is only valuable for classic enzymatic conditions (i.e., low enzyme concentration and excess substrate), when free substrate and inhibitor concentrations can be assumed to be equal to total concentrations. However, in the case of IN, the target of the inhibitor, the enzyme, is in excess compared with the DNA substrate. Thus, the Cheng-Prusoff relationship cannot be used because DNA and  $i$  (which represent concentrations of free DNA and inhibitor, respectively) are not equal to  $DNA_T$  and  $i_T$ , respectively. In contrast to the Cheng-Prusoff relationship, the Linden equation (Linden, 1982) represents the exact solution of  $K_i$  because it is not based simply on the assumption that free and total concentrations are approximately equal (this last assumption is often true for Michaelis-Menten conditions). Unfortunately, as is shown below, the Hill slope of the binding of IN to DNA is not equal to 1. In this case, the Linden equation is not valid. In our study, the estimation of  $K_i$  was based on the knowledge of the  $IC_{50}$  (which corresponds to the total inhibitor concentration, not free), the IN concentration used in the assay (again, the total concentration, not free), and the apparent  $K_d$  value or  $IN_{50}$ , using eq. 11.

The experiments were carried out with the SQ derivative FZ55, which has an inhibitory potency similar to that of KHD161 (the  $IC_{50}$  value of FZ55 for the 3'-processing reac-

tion,  $2.5 \mu M$ , is slightly higher than those obtained in same experimental conditions for KHD161). The saturation fraction,  $Y$ , as a function of total inhibitor concentration ( $i_T$ ) is expected to be sigmoidal (see eq. 10). Accordingly, as shown in Fig. 6,  $Y = f(i_T)$  was sigmoidal at four different IN concentrations (Fig. 6, a–d).  $IC_{50}$  values were then obtained from a sigmoidal model. To calculate the  $K_i$  values at two temperatures, 25 and 37°C, we first determined the apparent  $K_d$  values using the isotherm binding curves shown in Fig. 7. The results are summarized in Table 3. The apparent  $K_d$  value was rather constant in the 25 to 37°C temperature range, indicating that the enthalpic contribution to the IN-DNA complex formation is minor. Furthermore, the binding of DNA to IN exhibited a positive cooperativity at both temperatures ( $n = 1.81$ – $1.88$ ). As expected from eq. 11, the  $IC_{50}$  values were strongly dependent upon the experimental conditions (i.e., the IN concentration) (Table 4). Taking into account the various IN concentrations, the  $K_i$  values were rather similar for a given temperature (see the fourth column of Table 4), and average  $K_i$  increased slightly (from 620 to 890 nM) when the temperature was increased from 25°C to 37°C. This indicates that in contrast to the IN-DNA interaction, the binding of FZ55 to IN is not entropy-driven. The estimated enthalpy for the IN-FZ55 complex was about  $-5.6$  kcal/mol, whereas the entropy was slightly negative. KHD161 is slightly better than FZ55. Taking into account the  $IC_{50}$  values found in Fig. 5, the calculated  $K_i$  for KHD161 at 25°C and 37°C are equal to about 400 and 680 nM, respectively. This agrees with the  $IC_{50}$  values found in activity tests. Note that  $K_i$  values can be calculated using eq. 11 only if fractional saturation is near or equal to 1 in the absence of inhibitor. This can be considered the case given the last column of Table 4.

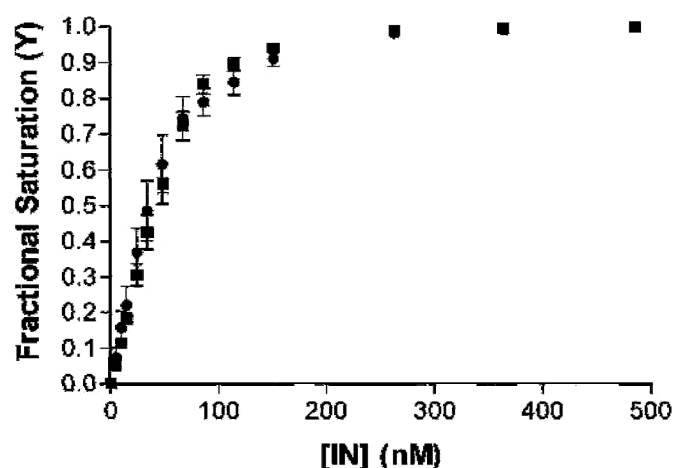
Only a model using  $n$  different sites was able to fit the binding isotherm data. The Hill number,  $n$ , was found to be equal to approximately 1.9 by the binding isotherms shown



**Fig. 6.** Percentage of integrase-oligonucleotide complexes as a function of FZ55 and integrase concentrations. Steady-state fluorescence anisotropies were measured at 25°C as described under *Materials and Methods*. a, [IN] = 120 nM; b, [IN] = 240 nM; c, [IN] = 360 nM; d, [IN] = 480 nM. The  $IC_{50}$  values together with those obtained at 37°C are reported in Table 4.

in Fig. 7. This finding was confirmed by simulations of inhibition curves using eq. 10 and according to the  $K_i$  and  $K_d$  values calculated in the previous section. Two examples are shown in Fig. 8, using two different IN concentrations [120 nM (Fig. 6a) and 240 nM (Fig. 6b)]. When  $n$  was equal to 1, the shape was hyperbolic and was not consistent with the experimental data. It was thus necessary to increase  $n$  to fit the experimental curves. A Hill coefficient of approximately 2.5 was found, which is compatible with the value of 1.9 found in direct binding isotherm experiments. The nonhyperbolic shape of the inhibition curves shown in Fig. 6 can therefore be explained by the positive cooperative binding of IN to DNA.

At higher DNA substrate concentrations (>10 nM), the 3'-processing activity decreased (data not shown). The most plausible explanation for this is that the active multimeric form of IN is not favorable at this high DNA substrate/IN ratio. According to the steady-state fluorescence anisotropy data, it seems that a simple 1:1 model of IN/viral DNA



**Fig. 7.** Binding isotherms for IN-oligonucleotide interactions. The concentration of fluorescein-labeled DNA was 1 nM. Various IN concentrations were preincubated with DNA at 25°C (■) or 37°C (○) for 20 min to reach equilibrium. Steady-state fluorescence anisotropies were then measured at the same temperatures.

TABLE 3

Parameters for the IN-DNA binding isotherm

	IN <sub>50</sub> or apparent $K_d$	Hill Coefficient	Anisotropy Amplitude
25°C	37.5 ± 3.5 nM	1.88 ± 0.04	0.183 ± 0.014
37°C	39.2 ± 8 nM	1.81 ± 0.08	0.180 ± 0.006

TABLE 4

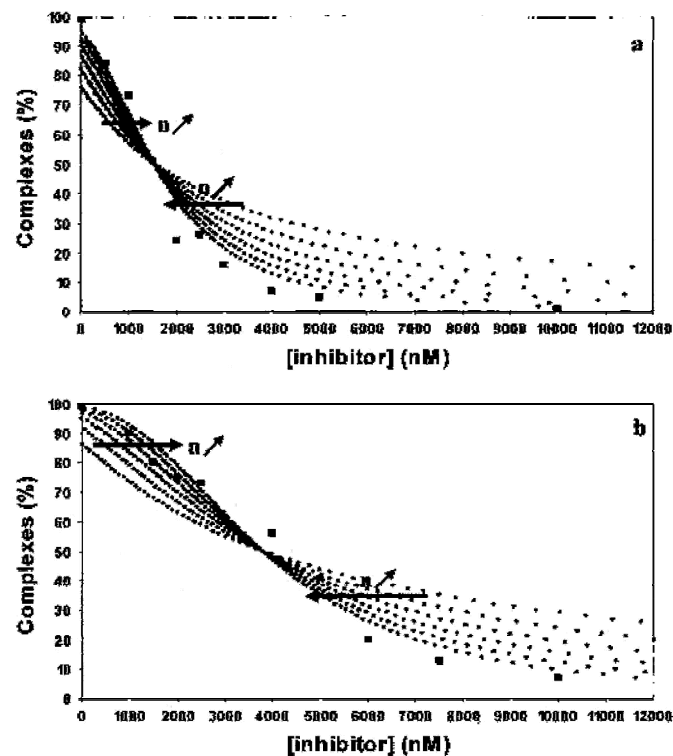
IC<sub>50</sub> and  $K_i$  for FZ55 at different IN concentrations

$K_i$  was calculated from eq. 11. The apparent  $K_d$  values at 25°C and 37°C in eq. 11 were obtained from the plot shown in Fig. 7.

	[Integrase]	IC <sub>50</sub>	$K_i$	$\frac{(IN)^n}{H[(IN)^n_H + (IN_{50})^n_H]}$
	nM	nM	nM	
25°C	120	1575 ± 197	670	0.899
	240	3930 ± 623	687	0.970
	360	5270 ± 227	573	0.986
	480	6930 ± 374	548	0.991
37°C	120	2340 ± 174	1090	0.883
	240	4540 ± 264	844	0.964
	360	6900 ± 411	802	0.982
	480	9810 ± 588	832	0.989

interaction is not sufficient to fit the experimental data of both binding isotherm and inhibition. A cooperative model is more appropriate and gives a Hill coefficient of between 1.9 and 2.5. In a previous study, we found a monomeric form of IN bound to DNA at 25°C and a monomer-dimer equilibrium at 37°C (Deprez et al., 2001). These results were obtained at a low IN/DNA ratio (from 5 to 10). Given the  $K_d$  value found here (apparent  $K_d$ , ~35 nM), these conditions were not favorable to observe multimeric forms of IN onto DNA. Here, the IN/DNA ratios were much higher at both 25°C and 37°C (from 120 to 480), which could explain the cooperativity of DNA-binding, which is clear from the direct binding of IN to DNA (Table 3) and the displacement of the equilibrium in the presence of SQs (Fig. 8).

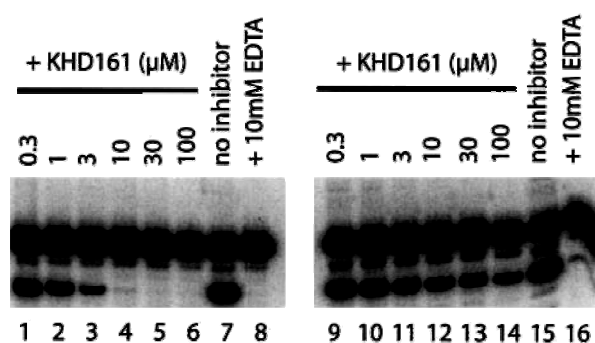
**Styrylquinolines Do Not Inhibit the 3'-Processing Activity of Preformed IN-DNA Complexes.** Our results show that the inhibition of 3'-processing is probably caused by a competition mechanism between the drug and the DNA substrate. Because IN-DNA complexes display half-lives of a few hours, without turnover of the enzymes during the reaction (Lee et al., 1995; Pemberton et al., 1996), SQs must act on free IN and can only efficiently block IN when preincubated with the enzyme. They should not inhibit 3'-processing performed by preformed IN-DNA complexes. This idea is reinforced by the results of the DNA-binding assays described above. To address this hypothesis, we compared the effect of SQs on the 3'-processing reaction when IN was preincubated with KHD161 before adding the DNA substrate and when IN was allowed to form a complex with its cognate DNA substrate before adding the inhibitor (Fig. 9). First, KHD161 was incubated with 50 nM IN before the addition of



**Fig. 8.** Simulation of the Hill coefficient effect on the inhibition curves. a, 120 nM IN; b, 240 nM IN. ■, experimental data. ○, simulations using equation 10. The Hill number ( $n$ ) increases from 1 to 2.5 (in increments of 0.3), as indicated by the black arrows.

the DNA substrate (Fig. 9, lanes 1–8). Submicromolar inhibition was observed, with an  $IC_{50}$  of  $0.4 \mu\text{M}$ . This value was lower than that determined above ( $1.2 \mu\text{M}$ ) because the concentration of IN was lower ( $50 \text{ nM}$  compared with  $120 \text{ nM}$ ). Second, IN/LTR complexes were performed for 20 min at  $25^\circ\text{C}$  in the presence of  $\text{Mg}^{2+}$ , because it has been shown previously that the amount of complexes is optimal at this time (Brodin et al., 2002). The drug was then added and the reaction was allowed to proceed at  $37^\circ\text{C}$  (Fig. 9, lanes 9–15). In a control experiment, a metal chelator, EDTA, was added immediately after the preincubation period to ensure that no 3'-processing could take place during this step. In the presence of EDTA, only a low level of residual activity was observed (lane 16), demonstrating that the 3'-processing activity observed in the absence of EDTA mainly occurred during the incubation at  $37^\circ\text{C}$ . In this condition, the inhibition by the SQ was dramatically decreased, with an  $IC_{50}$  of more than  $30 \mu\text{M}$ . Thus, this compound was much less efficient in inhibiting the 3'-processing activity of the preformed IN-DNA substrate complex because of its inability to disrupt the complex.

**Styrylquinoline Derivatives Inhibit Strand Transfer Activity of Immobilized IN-DNA Complex.** SQs cannot inhibit the 3'-processing of preformed complexes. However, because strand transfer requires the binding of a second DNA molecule to the processed DNA-enzyme complex, our data did not allow us to draw any conclusions about their abilities to act as strand transfer inhibitors. Unlike inhibitors such as DKA derivatives, which specifically impair the strand transfer process, SQ compounds efficiently inhibit the 3'-processing reaction. However, in overall standard *in vitro* activity assays using oligonucleotides that mimic one LTR extremity, the strand transfer products correspond to the transfer of a processed DNA into a homologous DNA. In the case of SQ, inhibition of the 3'-processing reaction obviously leads to a marked decrease in the amount of strand transfer products. This observation does not necessarily mean that SQs are potent inhibitors of the strand transfer process because the apparent inhibitory effect may simply result from

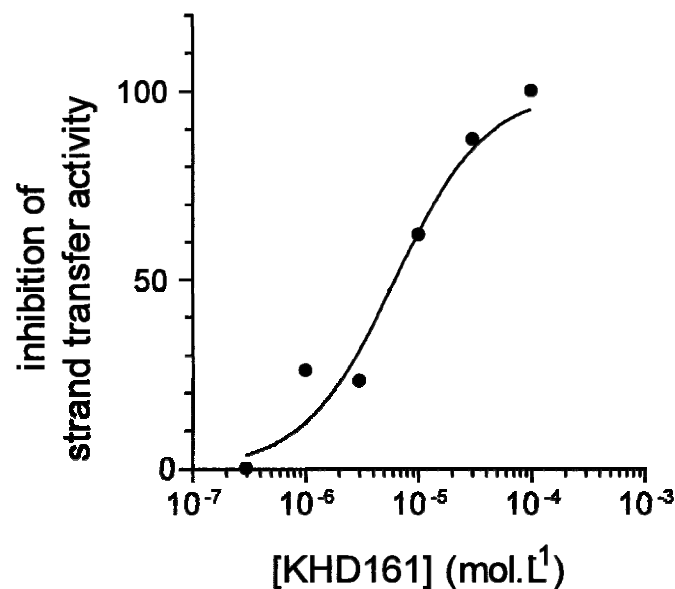


**Fig. 9.** Comparison of the effects of styrylquinoline on the free IN and preassembled DNA-IN complexes for the 3'-processing reaction. IN was incubated with the U5 blunt substrate for 20 min at room temperature to allow formation of the IN-DNA complex. The 3'-processing reaction was subsequently started by increasing the temperature to  $37^\circ\text{C}$  and allowed to proceed for 1 h. A control reaction was quenched by adding EDTA at  $t = 20 \text{ min}$  to determine the reaction yield during the preincubation period. Increasing concentrations of KHD161 were added before (lanes 1–8) or after (lanes 9–16) the formation of IN-DNA complexes. Drug concentrations: lanes 1 and 9,  $0.3 \mu\text{M}$ ; lanes 2 and 10,  $1 \mu\text{M}$ ; lanes 3 and 11,  $3 \mu\text{M}$ ; lanes 4 and 12,  $10 \mu\text{M}$ ; lanes 5 and 13,  $30 \mu\text{M}$ ; lanes 6 and 14,  $100 \mu\text{M}$ ; lanes 7 and 15, no inhibitor; lanes 8 and 16, no inhibitor and  $10 \text{ mM}$  EDTA.

the disappearance of the substrate of the strand transfer reaction. On the other hand, classic strand transfer assays using preprocessed DNAs may only reveal the propensity of drugs to inhibit the binding of IN to its DNA substrate. To test whether SQs can inhibit the strand transfer independently of the 3'-processing reaction, we assayed this activity on a preformed IN-DNA complex. The viral preprocessed DNA substrate (LDS31A/LDS32Bp) was immobilized on the NH<sub>2</sub>-activated surface of a microtiter plate and the IN-DNA complex was formed by allowing IN to bind to the substrate for 20 min. Free IN was removed, and biotin-labeled double-stranded DNA target was added in the presence of the SQ derivative. A dose-dependent inhibitory effect was observed when SQ was added after the formation of the IN/DNA complex (Fig. 10). The  $IC_{50}$  value in this condition was approximately  $6 \mu\text{M}$ . In a control experiment (not shown), KHD161 was added before IN to determine the inhibition of strand transfer caused by the impairment of IN-binding to its substrate DNA. The  $IC_{50}$  value was about  $1 \mu\text{M}$ . Although the  $IC_{50}$  value obtained for the preformed complex,  $6 \mu\text{M}$ , was higher than that observed when the drug was present before the formation of the complex, this result indicates that SQs also inhibit the strand transfer reaction.

## Discussion

Our experimental data indicate that SQ compounds are pure competitive inhibitors of the 3'-processing reaction. SQs inhibit the formation of the IN-viral DNA complexes with  $K_i$  values of  $0.4$  to  $0.9 \mu\text{M}$ . The fluorescence spectral study of one derivative strongly supports a direct binding of SQ to IN. Two of the three protein domains display a DNA-binding activity



**Fig. 10.** Inhibition of strand transfer reaction on immobilized IN-U5-2 complex by KHD161. HIV-1 IN ( $100 \text{ nM}$ ) was incubated with preprocessed U5 substrate covalently immobilized on a microtiter plate for 20 min at room temperature. Free IN was subsequently removed. The reaction was started by increasing the temperature to  $37^\circ\text{C}$  and allowed to proceed for 1 h. The reaction mixture included  $130 \text{ nM}$  of biotin-labeled double-stranded DNA target and increasing concentrations of KHD161 compound. The reaction was stopped by adding EDTA. Strand transfer products were quantified by measuring phosphatase activity as indicated under *Materials and Methods*. Inhibition is indicated as a percentage of the control activity.

and are therefore potential targets: the catalytic core, which contains the catalytic triad, and the C-terminal domain, which stabilizes the DNA-protein interaction. Previous inhibition experiments of the core domain-mediated disintegration reaction by SQs strongly suggested that this domain was the target (Zouhiri et al., 2000). The docking data with the HIV-1 catalytic core confirm this hypothesis. These data indicate the existence of stable interactions between SQ and the core domain near the DDE triad (cluster 1) and other residues known to make contacts with DNA. Moreover, these interactions remained stable over a 2-ns MD.

SQs were efficient when they were added before the formation of IN-DNA complexes, suggesting that once bound to the target, IN, they prevent DNA binding as suggested by computational docking. This mechanism was confirmed by fluorescence anisotropy assays. The addition of SQs prevented the formation of IN-DNA complexes in a manner that is compatible with the decrease in the 3'-processing activity. Therefore, SQs do not act in the subsequent catalysis step. Furthermore, SQs were unable to disrupt preformed complexes and did not inhibit the 3'-processing of the preformed IN-viral DNA complex. This is in agreement with other data showing that IN-DNA complexes are very stable ( $t_{1/2} = 1$  h) and characterized by a very slow turnover ( $k_{\text{cat}} = 0.04\text{--}0.06$  min<sup>-1</sup>) (Lee et al., 1995; Pemberton et al., 1996). Thus, the very slow dissociation rate of the IN-DNA complex does not allow competition with SQ to take place within the duration of the experiments described here.

The *in vitro* inhibition activity of styrylquinoline is mainly related to the simultaneous presence of a carboxyl group at C-7 (Zouhiri et al., 2000) and a hydroxyl group at C-8 in the quinoline subunit (this study; see FZ117). At experimental pH, close to neutrality, KHD161 and FZ55 probably have a net charge of -1 (reinforcing the hypothesis that DNA is not a ligand for these drugs) and are most likely stabilized in the active site through interactions between the *o*-hydroxy acid group (7-COO<sup>-</sup>/8-OH) and the Mg<sup>2+</sup> cofactor. The enol form is more favorable in aqueous medium, but the keto form has more favorable interaction energy in the active site context. Such interactions are compatible with an enthalpy-driven association of the drug with the active site as found using steady-state fluorescence anisotropy assay. In contrast, IN-DNA interactions seem to be driven primarily by entropy, probably because of a large solvent contribution (the release of water from large protein-DNA interfaces is entropically favorable).

Using assays that make it possible to test independently the inhibition effect on the strand transfer process performed by a preformed complex between IN and the preprocessed viral DNA substrate, we found that SQs may significantly inhibit the subsequent integration reaction into a heterologous target DNA. The corresponding IC<sub>50</sub> value was higher than that measured for 3'-processing (6 versus 1 μM, respectively). Two models can explain this result: 1) SQs may target and actively disrupt the preformed complex. This hypothesis is unlikely because we established that SQs do not efficiently inhibit the 3'-processing of the preformed IN-viral DNA complex. This means that once the IN-viral DNA or the IN-preprocessed viral DNA complex is established, it cannot be disrupted by competitive drugs. 2) SQs may bind to a second site within the active site in a manner that competitively inhibits the binding of the second substrate, the target DNA.

This secondary binding site that prevents the target DNA from binding could be close to the first one that prevents the viral DNA from binding, yet distinct in the active site. In this binding mode, SQs would act similarly to DKA compounds (Espeseth et al., 2000). However, one major difference between SQ and DKA compounds is that SQs have a lower affinity for the target DNA binding site; the IC<sub>50</sub> values for strand transfer are significantly higher than the values found for 3'-processing. The situation is very different for the DKAs because they do not bind efficiently to the active site of the DNA-free IN (Espeseth et al., 2000). Based on our results, we cannot determine whether SQs also bind to the second site in the context of the DNA-free IN or only when the first DNA substrate is already bound, as is observed with DKAs.

From computational simulations, it seems that cluster 1 is very likely to be the primary binding site for SQs. This cluster, which corresponds to one of the two proposed DNA-binding sites (Perryman and McCammon, 2002), contains the catalytic triad and several key residues for DNA binding. This cluster is characterized by a positive electrostatic potential created by residues Lys156, Lys159, and His67. The average interaction energy is better in cluster 1 than in other clusters. Moreover, unlike clusters 2 and 3, the interaction energy is strongly affected by the proximity of Mg<sup>2+</sup> in this cluster in accordance with experiments. In fact, IN has much less effect on the fluorescence properties of FZ55 in the absence of Mg<sup>2+</sup> or in the presence of EDTA. Nevertheless, clusters 2 and 3 may be candidates for the secondary binding site. It is not clear which one is actually the target, but cluster 3 contains two key residues for DNA-binding (Mazumder et al., 1996; Esposito and Craigie, 1998) and corresponds to one DNA binding site proposed by Perryman and McCammon (2002). Nevertheless, our docking data were obtained with the DNA-free protein, and we cannot rule out the idea that the secondary SQ-binding site is created by the binding of DNA in a manner similar to that described previously for DKAs. In that case, the role of a second Mg<sup>2+</sup> in the active site could be crucial for the stabilization of the SQ in the second binding site.

A recent study (Marchand et al., 2002) suggests that two different binding sites for DKAs may coexist in the active site and the donor and the acceptor sites. Accordingly, a DKA compound, such as L-708,906, that preferentially inhibits the strand transfer reaction, binds within the acceptor site. Bifunctional DKAs and a DKA-related compound, 5-CITEP, are less selective and seem to bind both donor and acceptor sites. SQs behaves similarly, and the existence of two binding modes may be extended to many compounds having both aromatic and carboxylate groups. Therefore, it seems that all these compounds share a common competitive mechanism of action, with a more pronounced effect against either donor or acceptor DNA-binding depending on their relative affinities for the corresponding binding sites. L-708,906 displays extreme behavior; it has a very low and a very high affinity for the donor and acceptor sites, respectively. The bifunctional DKA and 5-CITEP bind to both sites but have a higher affinity for the acceptor site because they are better able to inhibit the strand transfer than the 3'-processing. Finally, SQs can efficiently inhibit 3'-processing and exhibit significant proper activity on the strand transfer reaction starting from a preassembled IN-preprocessed viral DNA complex. According to the IC<sub>50</sub> values obtained for the two catalytic

steps, SQs probably have a greater affinity (5- to 10-fold) for the donor site than for the acceptor site, in contrast to that observed for Merck or Shionogi inhibitor.

It is widely believed that the stable IN-DNA interaction within the preintegration complex precludes inhibitors acting as competitor of donor DNA from being effective in the cellular context. If this is correct, compounds that inhibit 3'-processing in vitro in a competitive manner should be unable to block viral replication. However, although most 3'-processing inhibitors actually lack antiviral effect, this is not the case for SQs (Zouhri et al., 2000) and other 3'-processing inhibitors such as pyranodipyrimidine compounds (Pannecouque et al., 2002), which show a moderate activity ex vivo. Several hypotheses may account for this apparent discrepancy. First, in light of our data, the ex vivo activity of SQs could still be related to their ability to inhibit the strand transfer reaction. Because SQ derivatives are less efficient against strand transfer than 3'-processing in vitro, this hypothesis could explain the micromolar antiviral activity detected ex vivo. Recently, another series of IN inhibitors that is not specific for strand transfer in vitro was shown to effectively prevent HIV replication by inhibiting integration (Pannecouque et al., 2002). Second, we cannot rule out the possibility that the active site of IN is accessible, at least transiently, during the early steps of replication. If this is the case, the 3'-processing reaction could be affected by SQs. The early targeting of IN might also affect reverse transcriptase, because functional interactions between both proteins have been described. In fact, the KRK (186–188) motif of IN has been described to play a major role in the efficiency of reverse transcription (Tsurutani et al., 2000) and two residues of this motif, Lys186 and Lys188, were present in cluster 2 of our docking study. Finally, the IN-inhibitor interaction may alter the karyophilic properties of the enzyme, which are required for the nuclear translocation of the preintegration complex. It is noteworthy that a single structural protein domain may be simultaneously involved in both DNA-binding and nuclear localization (Nissen et al., 2000). Studies on the ex vivo mechanism(s) of inhibition of HIV replication by SQs are in progress.

#### Acknowledgments

We thank Françoise Simon for technical assistance. M.K. and M.L.B. are grateful for access to the O2000 multiprocessor of Pôle Parallélisme IdF Sud.

#### References

- Brodin P, Pinskaya M, Buckle M, Parsch U, Romanova E, Engels J, Gottikh M, and Mouscadet JF (2002) Disruption of HIV-1 integrase-DNA complexes by short 6-oxocytosine-containing oligonucleotides. *Biochemistry* **41**:1529–1538.
- Burdujan R (2002) *Dynamique réactionnelle d'inhibiteurs de l'intégrase du VIH-1: de la solution au milieu cellulaire*. Thesis, University Paris XI, Orsay, France.
- Cheng Y and Prusoff WH (1973) Relationship between the inhibition constant (K<sub>i</sub>) and the concentration of inhibitor which causes 50 per cent inhibition (I<sub>50</sub>) of an enzymatic reaction. *Biochem Pharmacol* **22**:3099–3108.
- Chow SA, Vincent KA, Ellison V, and Brown PO (1992) Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science (Wash DC)* **255**:723–726.
- Craigie R (2001) HIV integrase, a brief overview from chemistry to therapeutics. *J Biol Chem* **276**:23213–23216.
- Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, and Brochon JC (2000) Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. *Biochemistry* **39**:9275–9284.
- Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, Hawkins ME, and Brochon JC (2001) DNA binding induces dissociation of the multimeric form of HIV-1 integrase: a time-resolved fluorescence anisotropy study. *Proc Natl Acad Sci USA* **98**:10090–10095.
- Drake RR, Neamati N, Hong H, Pilon AA, Sunthakar P, Hume SD, Milne GW, and Pommier Y (1998) Identification of a nucleotide binding site in HIV-1 integrase. *Proc Natl Acad Sci USA* **95**:4170–4175.
- Espeseth AS, Felock P, Wolfe A, Witmer M, Grobler J, Anthony N, Egbertson M, Melamed JY, Young S, Hamill T, et al. (2000) HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc Natl Acad Sci USA* **97**:11244–11249.
- Esposito D and Craigie R (1998) Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO (Eur Mol Biol Organ) J* **17**:5832–5843.
- Ewald P (1921) Die Berechnung Optischer und Elektrostatischer Gitterpotentiale. *Ann Phys* **64**:253–287.
- Goldgur Y, Craigie R, Cohen GH, Fujiwara T, Yoshinaga T, Fujishita T, Sugimoto H, Endo T, Murai H, and Davies DR (1999) Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. *Proc Natl Acad Sci USA* **96**:13040–13043.
- Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, and Davies DR (1998) Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc Natl Acad Sci USA* **95**:9150–9154.
- Harper AL, Skinner LM, Sudol M, and Katzman M (2001) Use of patient-derived human immunodeficiency virus type 1 integrases to identify a protein residue that affects target site selection. *J Virol* **75**:7756–7762.
- Hazuda DJ, Felock P, Witmer M, Wolfe A, Stillmock K, Grobler JA, Espeseth A, Gabryelski L, Schleif W, Blau C, et al. (2000) Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science (Wash DC)* **287**:646–650.
- Hazuda DJ, Hastings JC, Wolfe AL, and Emini EA (1994) A novel assay for the DNA strand-transfer reaction of HIV-1 integrase. *Nucleic Acids Res* **22**:1121–1122.
- Jenkins TM, Esposito D, Engelman A, and Craigie R (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photocrosslinking. *EMBO (Eur Mol Biol Organ) J* **16**:6849–6859.
- Jing N, Marchand C, Liu J, Mitra R, Hogan ME, and Pommier Y (2000) Mechanism of inhibition of HIV-1 integrase by g-tetrad-forming oligonucleotides in vitro. *J Biol Chem* **275**:21460–21467.
- Laboulais C, Deprez E, Leh H, Mouscadet JF, Brochon JC, and Le Bret M (2001) HIV-1 integrase catalytic core: molecular dynamics and simulated fluorescence decays. *Biophys J* **81**:473–489.
- Laboulais C, Ouali M, Le Bret M, and Gabarro-Arpa J (2002) Hamming distance geometry of a protein conformational space: application to the clustering of a 4-nm molecular dynamics trajectory of the HIV-1 integrase catalytic core. *Proteins* **47**:169–179.
- Lavery R, Sklenar H, Zakrzewska K, and Pullman B (1986) The flexibility of the nucleic acids: (II). The calculation of internal energy and applications to mononucleotide repeat DNA. *J Biomol Struct Dyn* **3**:989–1014.
- Lee SP, Kim HG, Censullo ML, and Han MK (1995) Characterization of Mg<sup>2+</sup>-dependent 3'-processing activity for human immunodeficiency virus type 1 integrase in vitro: real-time kinetic studies using fluorescence resonance energy transfer. *Biochemistry* **34**:10205–10214.
- Leh H, Brodin P, Bischerour J, Deprez E, Tauc P, Brochon JC, LeCam E, Coulaud D, Auclair C, and Mouscadet JF (2000) Determinants of Mg<sup>2+</sup>-dependent activities of recombinant human immunodeficiency virus type 1 integrase. *Biochemistry* **39**:9285–9294.
- Linden J (1982) Calculating the dissociation constant of an unlabeled compound from the concentration required to displace radiolabel binding by 50%. *J Cyclic Nucleotide Res* **8**:163–172.
- Maignan S, Guilloleau JP, Zhou-Liu Q, Clement-Mella C, and Mikol V (1998) Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *J Mol Biol* **282**:359–368.
- Marchand C, Zhang X, Pais GC, Cowansage K, Neamati N, Burke TR Jr, and Pommier Y (2002) Structural determinants for HIV-1 integrase inhibition by β-diketone acids. *J Biol Chem* **277**:12596–12603.
- Mazumder A, Neamati N, Pilon AA, Sunder S, and Pommier Y (1996) Chemical trapping of ternary complexes of human immunodeficiency virus type 1 integrase, divalent metal and DNA substrates containing an abasic site. Implications for the role of lysine 136 in DNA binding. *J Biol Chem* **271**:27330–27338.
- Mekouar K, Mouscadet JF, Desmaele D, Subra F, Leh H, Savoure D, Auclair C, and d'Angelo J (1998) Styrylquinoline derivatives: a new class of potent HIV-1 integrase inhibitors that block HIV-1 replication in CEM cells. *J Med Chem* **41**:2846–2857.
- Ni H, Sotriffer CA, and McCammon JA (2001) Ordered water and ligand mobility in the HIV-1 integrase-5CITEP complex: a molecular dynamics study. *J Med Chem* **44**:3043–3047.
- Nissen P, Kjeldgaard M, and Nyborg J (2000) Macromolecular mimicry. *EMBO (Eur Mol Biol Organ) J* **19**:489–495.
- Ouali M, Laboulais C, Leh H, Gill D, Desmaele D, Mekouar K, Zouhri F, d'Angelo J, Auclair C, Mouscadet JF, et al. (2000) Modeling of the inhibition of retroviral integrases by styrylquinoline derivatives. *J Med Chem* **43**:1949–1957.
- Pannecouque C, Pluyers W, Van Maele B, Tetz V, Cherepanov P, De Clercq E, Witvrouw M, and Debyser Z (2002) New class of HIV integrase inhibitors that block viral replication in cell culture. *Curr Biol* **12**:1169–1177.
- Pemberton IK, Buckle M, and Buc H (1996) The metal ion-induced cooperative binding of HIV-1 integrase to DNA exhibits a marked preference for Mn(II) rather than Mg(II). *J Biol Chem* **271**:1498–1506.
- Perryman AL and McCammon JA (2002) AutoDocking dinucleotides to the HIV-1 integrase core domain: exploring possible binding sites for viral and genomic DNA. *J Med Chem* **45**:5624–5627.
- Polanski J, Zouhri F, Jeanson L, Desmaele D, d'Angelo J, Mouscadet JF, Gieleciak R, Gateiger J, and Le Bret M (2002) Use of the Kohonen neural network for rapid screening of ex vivo anti-HIV activity of styrylquinolines. *J Med Chem* **45**:4647–4654.

98 Deprez et al.

Pommier Y, Marchand C, and Neamati N (2000) Retroviral integrase inhibitors year 2000: update and perspectives. *Antiviral Res* **47**:139–148.

Ryckaert JP, Cicotti G, and Berendsen H (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comp Phys* **23**:327–341.

Tsurutani N, Kubo M, Maeda Y, Ohashi T, Yamamoto N, Kannagi M, and Masuda T (2000) Identification of critical amino acid residues in human immunodeficiency virus type 1 IN required for efficient proviral DNA formation at steps prior to integration in dividing and nondividing cells. *J Virol* **74**:4795–4806.

Zhuang L, Wai JS, Embrey MW, Fisher TE, Egbertson MS, Payne LS, Guare JP Jr, Vacca JP, Hazuda DJ, Felock PJ, et al. (2003) Design and synthesis of 8-hydroxy-

[1, 6]naphthyridines as novel inhibitors of HIV-1 integrase in vitro and in infected cells. *J Med Chem* **46**:453–456.

Zouhiri F, Mouscadet JF, Mekouar K, Desmaele D, Savoure D, Leh H, Subra F, Le Bret M, Auclair C, and d'Angelo J (2000) Structure-activity relationships and binding mode of styrylquinolines as potent inhibitors of HIV-1 integrase and replication of HIV-1 in cell culture. *J Med Chem* **43**:1533–1540.

---

**Address correspondence to:** Jean-François Mouscadet, UMR CNRS-8113, Laboratoire de Biotechnologies et Pharmacologie Génétique Appliquée (L.B.P.A.), Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson, 94235 Cachan Cedex, France. E-mail: mouscadet@lbpa.ens-cachan.fr

---

**PARTIE II**

**Conformation et flexibilité de nucléosides  
différemment substitués en 2' et effets  
d'analogues de l'ADN viral sur l'activité de  
3'-processing de l'IN**



L'interaction spécifique entre un ADN et une protéine est généralement un processus complexe de reconnaissance directe et indirecte. La reconnaissance directe se fait à travers des liaisons spécifiques entre les deux molécules. La reconnaissance indirecte implique la structure, la forme et la flexibilité des molécules mises en jeu dans l'interaction.

On a pu observer, sur de nombreux complexes, que la fixation d'une protéine induit des déformations de la molécule d'ADN. Dans d'autres cas, des distorsions préexistent dans la double hélice d'ADN mais sont accentuées par la liaison de la protéine.

Les analogues des nucléosides permettent de modifier la structure de l'ADN et donc les interactions d'un ADN avec sa protéine cible. Si la liaison de l'ADN cible induit ou accentue une déformation de l'ADN, le changement de flexibilité devrait influencer sur la vitesse de formation du complexe actif.

L'activité de 3'-processing de l'IN nécessite que l'IN reconnaisse de façon spécifique les extrémités LTR de l'ADN viral. Le processus de cette reconnaissance n'est actuellement pas compris. La détermination de ce mécanisme d'interaction, formant le complexe IN/ADN actif, permettrait d'orienter la conception de nouvelles drogues inhibitrices de l'IN.

Pour essayer de comprendre les déterminants de la spécificité de l'interaction IN/ADN viral, des analogues de nucléotides ont été insérés dans des oligonucléotides doubles brins mimant l'extrémité LTR U5 de l'ADN viral (équipe M. Gottikh) et les effets de ces modifications sur le 3'-processing (équipe J-F. Mouscadet) ont été mesurés. Pour comprendre ces effets, nous avons voulu déterminer qu'elles étaient les modifications structurales que pouvaient induire ces analogues de nucléotides. Nous avons alors réalisé une étude structurale, par mécanique quantique *ab initio*, de nucléosides différemment substitués en 2'. Nous avons voulu prédire la conformation adoptée par ces analogues de nucléosides dans une double hélice ainsi que leur flexibilité. On aurait pu penser que pour déterminer la structure adoptée par l'analogue de nucléoside, il suffisait de trouver la conformation du minimum absolu. Notre travail montre que la structure ne peut être prédite correctement que lorsque le squelette a une conformation compatible avec celui des acides nucléiques doubles brins. D'autre part, l'orientation des substituants et l'hydratation, jouent un rôle très important.

## Chapitre I

### Détermination *ab initio* de la flexibilité de 2'-aminoribonucléosides et de 2'-aminoarabinonucléosides

---

L'effet de l'introduction d'un 2'-NH<sub>2</sub> ribonucléoside dans une double hélice d'acide nucléique est mal compris. L'introduction d'un groupe amino en 2' dans un fragment d'ADN mais également dans un fragment d'ARN double brin fait décroître la température de fusion. Dans l'ARN, il a été montré par RMN que l'aminoribonucléoside est en Nord (Pham *et al.*, 2004). Si l'aminoribonucléoside adopte également une conformation Nord dans l'ADN, sa présence permet de comprendre la déstabilisation de l'ADN où tous les sucres sont en Sud. Par contre cette conformation n'explique pas pourquoi l'ARN, où tous les sucres sont en Nord, est également déstabilisée.

Pour essayer de comprendre l'effet d'un 2'-aminonucléoside dans les doubles hélices d'ADN et d'ARN, nous avons réalisé une étude structurale de ce nucléoside modifié par mécanique quantique *ab initio* (niveau théorique B3LYP/6-31G\*).

Nos calculs montrent que lorsque le groupe amino est orienté vers l'O3' (comme l'orientation préférentielle du 2'-OH des ribonucléotides (Auffinger and Westhof, 1997)), le 2'-NH<sub>2</sub> ribonucléoside est en Nord dans la conformation la plus favorable qui puisse être insérée dans une double hélice. Ceci explique que l'insertion d'un 2'-aminoribonucléoside déstabilise un ADN mais elle ne devrait pas affecter un ARN. Si une molécule d'eau est placée près du groupe 2'-NH<sub>2</sub> par analogie avec la molécule d'eau formant dans les ARN un pont hydrogène entre le groupe 2'-OH et la base (le N3 des purines ou l'O2 des pyrimidines) (Egli *et al.*, 1996), le 2'-aminoribonucléoside est plus flexible qu'un ribonucléoside; il est le plus souvent en Nord mais adopte une conformation Sud. Il aura tendance à être plus souvent en Sud dans

une double hélice normalement en Nord. Ces passages en Sud désordonnés déstabilisent la structure d'ARN. Un 2'-aminonucléoside déstabilise donc un ADN parce qu'il est en Nord et un ARN parce qu'il est plus flexible qu'un ribonucléoside (voir chapitre II pour les propriétés de flexibilité des ribonucléosides).

Aurup et al. ont montré que l'introduction d'un groupe amino protonné en 2' dans un fragment d'ADN double brin faisait également décroître la température de fusion (Aurup et al., 1994). Par nos calculs de mécanique quantique, nous avons pu montrer que le 2'-NH<sub>3</sub><sup>+</sup> ribonucléoside est en Sud dans la conformation la plus stable qui peut être insérée dans une double hélice. L'angle de torsion  $\epsilon$  dans cette conformation Sud est caractéristique de la forme B<sub>II</sub> d'une double hélice et non de B<sub>I</sub>. *In vacuo* et en solvant implicite, le nucléoside modifié est extrêmement rigide et ne peut pas être en Nord. De plus, sa structure présente un angle glycosidique  $\chi$  anormal qui est favorisé par une forte attraction d'un proton du groupe NH<sub>3</sub><sup>+</sup> par l'O2 de l'uracile ou le N3 de l'adénine. Seule, la présence d'une molécule d'eau, pontant le proton de l'amino et la base, supprime ce phénomène et maintient l'orientation attendue pour une base d'une double hélice. L'analogue de nucléoside devient également plus souple, en présence de la molécule d'eau. Une conformation Nord est alors plus accessible mais elle présente un  $\epsilon$  tout à fait anormal pour une structure en double hélice. La déstabilisation, observée par l'introduction du nucléoside modifié dans un ADN, ne peut pas s'expliquer par un effet du plissement du sucre mais pourrait plutôt être liée à la conformation du squelette sucre-phosphate.

Nous avons également déterminé la conformation et la flexibilité de 2'-aminoarabinonucléosides. Lorsque le groupe 2'-amino neutre est positionné en arabinose, le plissement Nord est favorisé. Le nucléoside modifié est rigide *in vacuo* et devient un peu plus souple lorsqu'il est en solvant implicite. L'introduction de cet analogue de nucléoside dans un ADN devrait déstabiliser la double hélice.

Le groupe amino protonné, orienté en arabinose, favorise le plissement Sud. Il est extrêmement rigide *in vacuo* et en solvant implicite.

Par ailleurs, nous avons calculé la carte d'énergie potentielle de la désoxyadénosine dans l'espace amplitude/pseudorotation afin de déterminer le chemin entre le plissement Sud et Nord des sucres. Dans les conformations qui peuvent être introduites dans les structures en double hélice, la vallée à l'Est est unique. Ainsi les transitions Sud $\leftrightarrow$ Nord des sucres peuvent uniquement se faire à travers la barrière Est.

L'ensemble de ce travail est présenté dans le manuscrit joint ci-après.

***Ab initio determination of the flexibility of 2'-aminoribonucleosides and 2'-aminoarabinonucleosides.***

Sophie Barbe\* and Marc Le Bret\*

\*UMR CNRS-8113, Laboratoire de Biotechnologies et Pharmacologie génétique Appliquée (L.B.P.A.) Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson, 94235 Cachan, France ;

Address reprint requests to Marc Le Bret, LBPA, ENS-Cachan, 61 avenue du Président Wilson, 94235 Cachan, France. Tél: 33-1-47-40-59-97; Fax: 33-1-47-40-76-71; E-mail: mlebret@lbpa.ens-cachan.fr.

Keywords: Sugar puckering, modified nucleosides, DNA stability, Solvation, hydration, Quantum mechanics.

Running title: 2'-aminonucleosides conformations

## **ABSTRACT**

The sugar puckering of adenosine and uridine nucleosides with a neutral or protonated amino group at 2' in the ribo or arabino orientations are determined using high level quantum mechanical calculations. The amino group has always been found to be pyramidal and its orientation governs the conformation of the sugar. However, north puckering is favored among the conformations where the amino group points to the O3' as the 2'-hydroxyl group does in RNA. An explicit water molecule set as in RNA decreases the energy gap between the north and the south conformations. The model then shows that the insertion of an amino ribonucleotide destabilizes double-stranded RNAs and also double-stranded DNAs. The protonated NH<sub>3</sub><sup>+</sup> group at 2' favors the S-type but distorts the backbone. In the arabino orientation, an NH<sub>2</sub> substituent at 2' favors north puckering and an NH<sub>3</sub><sup>+</sup> south puckering. The whole relaxed potential map, in the amplitude/pseudorotation space, shows that for natural deoxyadenosine there is only one valley in the east running from south to north puckering.

## INTRODUCTION

Double-stranded DNA is in a dynamic equilibrium between various conformations depending on base sequence and environmental conditions (1-3). In aqueous solution, DNA is mostly in the B-form, which is characterized by the C2' endo or South (S) conformation of the nucleotide sugar. The sugar may switch to the C3' endo or North (N) conformation in less hydrated medium or in protein-DNA complexes (4-6). Although DNA and RNA differ only by one hydroxyl group, this is enough to confine RNA double helices to a single structural family (the A form), whereas DNA is polymorphic and exists in a variety of structural families including the A, B and Z forms (7).

The substituent at the 2' position can have two different orientations. In ribonucleotides, it, and O<sub>3'</sub>, are in the *cis* position relative to the C<sub>2'</sub>-C<sub>3'</sub> bond of the furanose ring. In arabino nucleotides they are in the *trans* position.

Nucleoside and nucleotide analogues have for long been used as antiviral, antibacterial and anticancer agents (8, 9). In recent years, the substitution of the 2' position of ribonucleotides with an amino group has been used to enhance nuclease resistance (10): this modification increases the therapeutic efficiency of aptamers (11). Aminonucleosides are used as sites of post-synthetic chemical modifications, and play a central role in cross-linking strategies (12, 13). Amino groups in 2' accelerate the phosphodiester cleavage (14). 2'-amino substituted nucleotides are more acylated when they are incorporated into bulged sites of DNA duplexes (15). These properties are extensively used to determine the structure of ribozymes (16).

There have been many attempts to find a satisfactory rule for assessing the pucker of natural or modified nucleosides. Nucleosides modified at the 2' position have been synthesized to identify the factors that control the S to N equilibrium. The first rule concerned dihedrals containing electronegative atoms X and Y. According to the *gauche* effect, the X-C-C-Y dihedrals are more often in *gauche* than in *trans* conformations (17) and this rule was sufficient to simulate the conformation of 2'-fluoro-2'-deoxyribose correctly. Later, it was suggested that the N-S equilibrium results from two opposite trends. i) Because of steric hindrance the base should have an equatorial orientation relative to the furanose ring. This is favored in the furanose S-conformation. ii) The so-called anomeric effect is the interaction  $n(O4') \rightarrow \sigma^*(C1'-N)$  between the O4' lone pair and the base nitrogen. It is favored in the furanose N-conformation (18). It was found that the population of a nucleoside in the N conformation increases with the electronegativity of the substituent (19, 20). The rule applies to the 2'-OCH<sub>3</sub>, 2'-SCH<sub>3</sub> pair of nucleotides (21). Interestingly, the 2'-aminonucleoside is one of the nucleosides farthest away from the regression line (19). Although NH<sub>2</sub> is more electronegative than H, it does not increase greatly the population in the N conformation. A 2'-aminonucleosides by themselves in solution are mostly in the S conformation. However, Aurup et al. (22) observed decreases in both the thermal stabilities of RNA/RNA but also in those of RNA/DNA and DNA/DNA helices when 2'-aminocytidines are present. The 2'-aminonucleotides were thought to be in the N conformation in all duplexes (23) and this was later confirmed by NMR (13) and by X-ray crystallography (24) for the case of self complementary RNAs containing a unique modified 2'-aminonucleotide.

Gauche effects, anomeric effects and electronegativity can be tuned in a set of experimental structures and this would elucidate the conformation of a modified nucleoside. This *ab initio* approach is however the most reliable if only few or no NMR or no X-ray structures are available. Such quantum mechanical methods have proved their value for natural deoxynucleosides (25, 26) which are found in the S conformation. Moreover, compounds with a 2'OH furanose mostly adopt the N conformation (27) among conformations compatible with a double helical structure.

*Ab initio* methods can be used to determine conformations in the global and local minima and how these are populated from the energy gaps. Moreover the path connecting two conformations and the energy barrier can be estimated, giving some insight into the dynamics of the conformational change.

In this work, we determine the conformation of nucleosides carrying either a NH<sub>2</sub> or a NH<sub>3</sub><sup>+</sup> group at the 2' position either in ribo or arabino orientation using *ab initio* methods that take the electronic correlation into account. We found that solvation plays an important role, so we report calculations *in vacuo*, in a continuous medium having the dielectric constant of water and *in vacuo* in the presence of an explicit water molecule. When the 2'-aminonucleoside is *in vacuo* or surrounded by the continuous medium our prediction is that it will be found in the S conformation; it will be in the N conformation only when the 2'-amino group is oriented towards the O<sub>3'</sub> as the 2'-hydroxyl group does in RNA. The study of the conformations leads to an estimate of the populations and helps to understand why the insertion of a 2'-aminonucleoside destabilizes both double-stranded DNA and double-stranded RNA. We develop a model in which the N conformation is the most favored so that it agrees with NMR data. Moreover, the insertion of an N-type nucleoside into a DNA duplex would cause its destabilization. However, the energy gap between N and S should be significantly smaller than its value in a ribonucleoside to explain the destabilization of an RNA duplex.

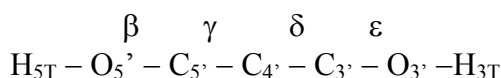
## METHODS

Quantum mechanical calculations were carried out with the Gaussian 98 and 03 programs (28, 29) using the 6-31 G\* basis set. Energy minimizations were performed to the default tolerances in the Gaussian programs. All degrees of freedom, other than those specified as being fixed, including all bond lengths and angles, were allowed to relax during the energy minimizations.

Two levels of theory were used for the minimizations: the Hartree-Fock (HF) level and the Density Functional Theory / Becke 3, Lee, Yang and Par (DFT/B3LYP) level. The B3LYP level considers the electron correlation effects explicitly. The effect of the solvent on the conformation was estimated in two different ways. i) The nucleoside was embedded in a continuous medium having the dielectric properties of water using the conductor-like polarizable continuum model (CPCM) method (30, 31). ii) A water molecule was added to the nucleoside and the complex was minimized *in vacuo*. In the case of NH<sub>3</sub>-RIB-U (Fig. 1), the water molecule was initially positioned with one of its protons, H<sub>2WT</sub>, in contact with the O<sub>2</sub> atom of the uracil and its oxygen, O<sub>WT</sub>, in contact with the H<sub>X2</sub> hydrogen of the 2'-group. In the case of NH<sub>2</sub>-RIB-U, the calculations were performed with two different initial orientations of the water molecule. In the first, WT<sup>1</sup>, the H<sub>2WT</sub> and O<sub>WT</sub> atoms of the water molecule form hydrogen bonds with the O<sub>2</sub> and H<sub>X2</sub> atoms of the 2'-amino nucleoside, respectively. In the second, WT<sup>2</sup>, the H<sub>2WT</sub> and H<sub>1WT</sub> protons of the water molecule form hydrogen bonds with the O<sub>2</sub> and N<sub>X</sub> atoms of the 2'-amino nucleoside, respectively..

-----  
- Insert Figure 1 about here -  
-----

The dihedral angles in the nucleosides (Fig. 1) are defined as follows:



The glycoside angle  $\chi$  is defined as:  $\text{O}_{4'} - \text{C}_{1'} - \text{N}_9 - \text{C}_4$  for purines and  $\text{O}_{4'} - \text{C}_{1'} - \text{N}_1 - \text{C}_2$  for pyrimidines. The five endocyclic dihedral angles were approximated by a Fourier analysis giving the phase or pseudorotation angle (P) and the amplitude ( $\tau$ ) according to (32) :

$$\tau_j = \tau \cos[P + 4\pi(j-2)/5] \quad (1)$$

From the measurements of the dihedrals  $\tau_j$  in a conformation, A and B are computed as:

$$A = 2/5 \sum_{j=0}^4 \tau_j \cos[4\pi(j-2)/5] \quad (2)$$

$$B = -2/5 \sum_{j=0}^4 \tau_j \sin[4\pi(j-2)/5] \quad (3)$$

which give the amplitude,  $\tau$ , and the pseudorotation, P, through:

$$\tau = (A^2 + B^2)^{1/2} \quad (4)$$

$$P = \tan^{-1} B/A \quad (5)$$

The amplitude is obviously a positive number. However, it can be used an algebraic number when  $\tau_2$  is zero, ( $\text{O}_{4'}$  endo and  $\text{O}_{4'}$  exo conformations): it is then positive when  $P=90^\circ$  and negative when  $P=-90^\circ$ . The pseudorotation space is divided into four equally sized quadrants centered around  $P=0.0^\circ$ ,  $P=90.0^\circ$ ,  $P=180.0^\circ$  and  $P=-90.0^\circ$ , referred to as the north (N), east (E), south(S), and west (W) quadrants, respectively.

When locating the north and south energy minima, the initial furanose conformations were set in canonical N and S puckering, respectively and the dihedral angles  $\beta$ ,  $\gamma$  and  $\chi$  were assigned modal values from the crystal DNA structure database (33) as follows:

$\beta = 175^\circ \pm 15$ ,  $\gamma = 57^\circ \pm 10$  and  $\chi = -161^\circ \pm 11$  for N-type conformations;

$\beta = 168^\circ \pm 25$ ,  $\gamma = 51^\circ \pm 16$  and  $\chi = -108^\circ \pm 11$  for S-type conformations.

The energy profile of a furanose cycle substituted in 3' by a  $\text{O}_{3'}\text{-P}$  group shows three minima when  $\varepsilon$  is varied (34). Therefore, in addition to the puckering and the dihedrals  $\beta$ ,  $\gamma$  and  $\chi$ , the dihedral  $\varepsilon$  was set in  $-ap$ ,  $-sc$  and  $+sc$  before optimization. In most cases, the dihedrals are unfrozen during the minimization. When the furanose of the minimized conformation is S-type, and  $\varepsilon$  in  $-ap$  or  $-sc$  the conformation is called  $\text{B}_I$  or  $\text{B}_{II}$  form. The minimum corresponding to  $\varepsilon$  in the  $+sc$  range we call  $\text{B}_{III}$ . Similarly, when the furanose is in the N conformation, the minima, with  $\varepsilon$  in  $-ap$ ,  $-sc$  and  $+sc$  are called A,  $\text{A}_{II}$  and  $\text{A}_{III}$  forms. For neutral amino nucleosides, certain conformations were minimized by freezing the  $\varepsilon$  dihedral angle to avoid the formation of the  $\text{H}_{3\text{T}} \dots \text{N}_X$  hydrogen bond which cannot exist in polynucleotide chains. The value chosen for  $\varepsilon$  depends on the conformation involved:  $-155^\circ$  for the A and  $\text{B}_I$  forms,  $-90^\circ$  for the  $\text{A}_{II}$  and  $\text{B}_{II}$  forms and  $+30^\circ$  or  $+60^\circ$  for the  $\text{A}_{III}$  and  $\text{B}_{III}$  forms. The conformations were prepared using our program Morcad (35) to satisfy all the constraints including the dihedrals to be frozen before they were submitted to Gaussian.

Potential energy differences were calculated to study the relationship between conformation and energy. The electronic energy of the  $\text{B}_I$  final conformation was used as the as reference ( $\Delta E=0$ ).

The east energy barrier separating the A and  $\text{B}_I$  conformations was identified and characterized as the  $\text{O}_{4'}$  endo conformation of the furanose (26) by freezing the endocyclic dihedral angle  $\tau_2$  ( $\text{C}_{1'} - \text{C}_{2'} - \text{C}_{3'} - \text{C}_{4'}$ ) to  $0.0^\circ$ .



To compute the relaxed potential map, the dihedral angles  $\beta$ ,  $\gamma$ ,  $\chi$  were set to their modal values and the sugar puckering ( $\tau$ , P) was set as required by freezing two of the endocyclic  $\tau$ s. When one of  $\beta$ ,  $\gamma$ ,  $\chi$  departed from the modal range during minimization, it was frozen and the minimization restarted. The conformation was then minimized with two endocyclic dihedrals frozen. Next, the constraint on one of the frozen intracyclic dihedrals was removed and an additional minimization was performed. The energy at a representative point in the ( $\tau$ , P) plane, was thus estimated with a different number of frozen dihedrals and a different number of optimizations according to its location. In a preliminary study, all the intermediate steps obtained during the minimization were retained to increase the number of energy evaluations in the ( $\tau$ , P) plane. When an endocyclic dihedral is frozen, the conformations obtained at each step of the minimization describe a straight descent line in the ( $\tau$ , P) plane (see Eq. 1) and yield the fully minimized conformation (when the gradient is perpendicular to the descent line). However, at the intersection of two descent lines we obtained inconsistent values and decided only to keep the fully minimized conformations. The points in the ( $\tau$ , P) plane representing fully minimized conformations were used to build Delaunay triangles to partition the ( $\tau$ , P) plane without overlap. A set of 3 points is a Delaunay triangle if its circumscribing circle does not contain any other point. Equipotential lines were drawn from a linear interpolation on the sides of the Delaunay triangles.

## RESULTS AND DISCUSSION

### H-A

Foloppe et al. (26) studied all the natural deoxynucleosides in the A and B<sub>I</sub> conformations using various *ab initio* methods including HF/6-31G\* and B3LYP/6-31G\*. Here, using HF/6-31G\*, B3LYP/6-31G\* and B3LYP/CPCM/6-31G\*, and starting from the six conformations described in the methods section, six unconstrained minima for the deoxyadenosine were found (see Table 1): three of them have N puckering and three of them S puckering. The results agree with literature (34) when the method and the starting conformation overlap. The values of  $\beta$ ,  $\gamma$ , and  $\delta$  dihedrals (not shown), and of  $\varepsilon$  and  $\chi$  (shown in Table 1) were in the modal ranges for each of the canonical A, B<sub>I</sub>, and B<sub>II</sub> DNA forms. The higher energy conformations, A<sub>II</sub>, A<sub>III</sub> and B<sub>III</sub>, can be associated with DNA forms observed by NMR (36) and in some rare protein - nucleic acid complexes deposited in the Protein Data Bank (33).

The *ab initio* computations show that deoxyadenosine has equilibrium conformations that can be inserted into a double stranded deoxynucleic acid with only small distortions.

The most favorable conformation at all theoretical levels is B<sub>I</sub>. When the deoxyadenosine is completely embedded in a continuous medium, as in the CPCM model, the A<sub>III</sub> and B<sub>III</sub> conformations are more favorable than the A form (Fig. 2), for which there is no experimental evidence. However, when deoxyadenosine is inserted in a double-stranded nucleic acid, it is not entirely in contact with the solvent. Therefore, the CPCM model exaggerates the effect of the solvent.

---

- Insert Table 1 about here -

---

-----  
 - Insert Figure 2 about here -  
 -----

Relaxed potential map.

Deoxyadenosine was minimized at the B3LYP level starting from many different conformations. Various endocyclic dihedrals were frozen during the minimizations to get an energy evaluation in the ( $\tau$ , P) plane (see methods section). Moreover, the backbone dihedrals must not depart substantially from the modal values. Fig. 3 shows the number of dihedrals that were frozen during the minimizations for each of the points. Most points in the middle of the map or on the eastern side were obtained with only one frozen dihedral. On the west, several points were obtained by freezing  $\beta$  at  $168^\circ$ . The isopotential lines show two minima (south and north, marked 0) and a unique saddle point on the east, most probably close to the  $\tau_2=0$  line as suggested by Foloppe et al. (26). The tangent plane at the inversion point (5 dihedrals frozen at  $0^\circ$ ) is not horizontal. There is no saddle point to the west of the inversion point. Starting from the point the farthest to the west, the eastern saddle point is obtained if  $\tau_2=0$  is the only constraint. Conformations along the  $\tau_2=0$  line were also studied with  $\beta$  close to  $80^\circ$  (Fig. 4): this gave a very shallow and energetically unfavorable western saddle point. The relaxed potential map is well established in the east, but is relatively arbitrary in the west because of the many constraints. Arora and Schlick (37) have simulated the dynamics of a DNA fragment using a MMMQ approach and did not detect any accumulation of west puckering, suggesting that there is no valley in the west.

-----  
 - Insert Figure 3 about here -  
 -----

-----  
 - Insert Figure 4 about here -  
 -----

## NH2-RIB-A and NH2-RIB-U

Molecular dynamic studies suggest that the 2'-OH bonds in S and N riboses of a tRNA<sup>Asp</sup> point preferentially toward the O<sub>3'</sub> oxygen of the same residue (38). When Foloppe and McKarrel (39) conducted MP2/6-31G\* minimization with a ribose attached to an imidazole group by initially fixing  $\varepsilon$  at  $180.0^\circ$  and the dihedral angle (C<sub>3'</sub>-C<sub>2'</sub>-O<sub>2'</sub>-H<sub>O'2</sub>) in g<sup>+</sup>, the N conformation was found to be more favorable than the S conformation. In the N and S minima, the 2'-OH remained oriented towards the O<sub>3'</sub> atom (39). There is a further difficulty if we want to transpose the results concerning ribonucleosides to 2'-aminonucleosides: the two hydrogen atoms H<sub>X1</sub> and H<sub>X2</sub> are not equivalent because the amino group at 2', in the 2'-aminonucleosides studied here, has always been found to be pyramidal. Therefore we studied two cases: models 1 and 2. The dihedral  $\lambda$  (H<sub>X2</sub>-N<sub>X</sub>-C<sub>2'</sub>-H<sub>2'2</sub>) is set so that H<sub>X2</sub> is as far as possible from O<sub>3'</sub> in model 1, and as close as possible to O<sub>3'</sub> in model 2. Once the starting conformation was obtained, a preliminary minimization was performed with  $\lambda$  frozen, and the resulting conformation was optimized during a second minimization with the dihedral  $\lambda$

unfrozen. The dihedral  $\lambda$  is thus a supplementary degree of freedom in addition to the three  $\varepsilon$  and the two puckers.

-----  
 - Insert Table 2 about here -  
 -----

Model 1:  $H_{X2}$  set far from  $O_{3'}$ .

When  $H_{X2}$  is set far from  $O_{3'}$ , minimizations performed *in vacuo*, at the B3LYP/6-31G\* level of theory, indicate that NH2-RIB-U has only two fully relaxed unconstrained minima, corresponding to  $B_{II}$  and  $A_{II}$  conformations. To avoid the  $H_{3T} \dots N_X$  hydrogen bond, the dihedral  $\varepsilon$  was frozen and three additional constrained minima were obtained corresponding to  $B_I$ ,  $B_{III}$  and  $A_{III}$  conformations (Table 2). These five minima appear in the order  $B_I < B_{II} < B_{III} < A_{III} < A_{II}$  and  $B_I$  is much favored over the four other conformations. S-type conformations are favored for a simple reason: the hydrogen bonds that stabilize the S and N minima of H-A also stabilize the S and N minima of NH2-RIB-U (Fig. 5). However, when  $H_{X2}$  diverges from  $O_{3'}$ , the S (but not N) conformations of NH2-RIB-U are stabilized by an additional hydrogen bond between the  $O_{3'}$  and the  $H_{X1}$  atoms, similar to the  $O_{2'}-H_{2'} \dots O_{3'}$  hydrogen bond observed in the A conformation of the ribonucleosides (Fig. 5). Analogous results were obtained with NH2-RIB-A (Table 2). Essentially, the S conformations of NH2-RIB-A and NH2-RIB-U are populated. The rate of interconversion between S and N is certainly very slow as the barrier is high for NH2-RIB-A. The conformation of the minimum obtained by freezing  $\tau_2=0^\circ$  is probably a good approximation of the east barrier. The east energy barrier (6.07 Kcal/mol) separating the  $B_I$  and A conformations is much greater than in the case of H-A. Moreover, the value of the puckering amplitude  $\tau$  at the east energy barrier ( $\tau=35.10$ ) is greater than that in the  $B_I$  and A conformations ( $\tau=32.60$  and  $\tau=28.00$ , respectively).

NMR experiments showed that the puckering of 2'-aminoribonucleosides inserted into RNA duplexes is N-type, so the B3LYP/6-31G\* vacuum model is questionable and we therefore investigated the effect of the solvent as a continuous medium.

When embedded in implicit solvent, at the B3LYP/CPCM/6-31G\* level, NH2-RIB-U has three unconstrained minima corresponding to  $B_{II}$ , A and  $A_{II}$  conformations, and three constrained minima corresponding to  $B_I$ ,  $B_{III}$  and  $A_{III}$  conformations (Table 2). These six minimized conformations appear in the order  $B_{II} < B_I < A < A_{III} < A_{II} < B_{III}$ . Therefore, in implicit solvent, NH2-RIB-U mainly adopts a  $B_{II}$  conformation and it is much less flexible than natural nucleosides such as H-A (compare to the corresponding results in table 1). As the most favorable conformation is S-type, the B3LYP/CPCM/6-31G\* model of NH2-RIB-U, where  $H_{X2}$  lies far from  $O_{3'}$ , should be rejected. Therefore, we minimized NH2-RIB-U, in N and S conformations with three initial values for  $\varepsilon$ , the hydrogen atom  $H_{X2}$  pointing towards  $O_{3'}$ .

-----  
 - Insert Figure 5 about here -  
 -----

Model 2:  $H_{X2}$  set close to  $O_{3'}$ .

In this case, two unconstrained minima (A and  $B_{II}$  conformations) were obtained (Table 2). The  $\epsilon$  dihedral angle had to be frozen in the  $B_I$ ,  $B_{III}$  and  $A_{III}$  conformations to avoid the formation of the  $O_{3'}-H_{3T}\dots N_X$  hydrogen bond (Table 2). These five minima appear in the order  $A < A_{III} < B_{II} < B_{III} < B_I$  (Table 2). Above, only the S puckering was favored by the  $N_X-H_{X1}\dots O_{3'}$  hydrogen bond, whereas both N and S conformations are stabilized by the  $N_X-H_{X2}\dots O_{3'}$  hydrogen bond. The S conformations are stabilized by  $C_{2'}-H_{2'}\dots O_{5'}$ ,  $C_6-H_6\dots O_{5'}$ ,  $C_6-H_6\dots O_{4'}$  and  $C_{1'}-H_{1'}\dots O_2$  hydrogen bonds. The N conformations are now stabilized by the  $C_{3'}-H_{3'}\dots O_{5'}$ ,  $C_6-H_6\dots O_{5'}$ ,  $C_6-H_6\dots O_{4'}$  and  $C_{1'}-H_{1'}\dots O_2$  hydrogen bonds and become the most favored conformations (Fig. 6).

According to this model, the insertion into a DNA duplex of a mainly N-type 2'-aminoribonucleoside should cause destabilization. In RNA, it is expected to be N-type as experimentally observed. However, the energy gap between the N-type and the S-type is very similar to our prediction for a 2'-ribonucleoside (Barbe & Le Bret, in preparation), and the substitution of nucleosides of same rigidity should not affect the stability of a double-stranded RNA. Therefore, we investigated other models.

-----  
 - Insert Figure 6 about here -  
 -----

Role of hydration.

Considerable effort has been devoted to determining the position and to the characterization of water molecules around nucleic acids. The high-resolution crystal structure of duplex RNA shows that water molecules form various clusters (Egli et al. (40)). Water molecules near the 2'-OH group can form an intra-residue bridge between the 2'-OH group and the pyrimidine  $O_2$ . Therefore, we studied the conformation of  $NH_2-RIB-U$  *in vacuo* but in the presence of an explicit water molecule forming a bridge between the  $O_2$  of the uracil and the 2'- $NH_2$  group. The calculations were carried out at the B3LYP/6-31G\* level with two different initial orientations of the water molecule  $WT^1$  in which  $H_{2WT}$  interacts with uracil  $O_2$  and  $O_{WT}$  with  $H_{X2}$ ; and  $WT^2$  in which  $H_{2WT}$  interacts with  $O_2$  and  $H_{1WT}$  with  $N_X$  (Fig. 7).

-----  
 - Insert Figure 7 about here -  
 -----

Using  $WT^1$  geometry, we obtained six minima: four constrained minima corresponding to the  $B_I$ ,  $B_{III}$ , A and  $A_{III}$  conformations and two unconstrained minima corresponding to the  $B_{II}$  and  $A_{II}$  conformations. These minimized conformations appear in the order:  $B_I < B_{II} < A < B_{III} < A_{III} < A_{II}$  (Table 2). The water molecule was free to move but the  $WT^1$  geometry remained when the structure was fully minimized. The  $B_I$  structure is very rigid because the energy gap separating  $B_I$  from the other minima is large. This does not agree with experimental data.

Using the  $WT^2$  geometry, we obtained five minima: four constrained minima corresponding to the  $B_I$ ,  $B_{III}$ , A and  $A_{III}$  conformations and one unconstrained minima corresponding to the  $B_{II}$

conformation. These minimized conformations appear in the order:  $A < B_{II} < B_{III} < B_I < A_{III}$  (Table 2). The energy gap separating the most favorable conformation from the next three conformations is small. Although the water molecule was free to move and distort, it remained in its initial orientation, in all the final minimized conformations. Moreover,  $O_{WT}$  was close to the  $H_{3'}$ . As in model 2, the S and N  $WT^2$  conformations are stabilized by the  $N_X H_{X2} \dots O_{3'}$  hydrogen bond (Fig. 7). The orientation of the 2'-amino group plays a crucial role in the puckering of NH2-RIB-U. The water molecule in the orientation  $WT^2$  gives flexibility to the system.

### NH3-RIB-A and NH3-RIB-U

*In vacuo*, NH3-RIB-A and NH3-RIB-U were been minimized from the N-type and S-type conformations with three values for  $\epsilon$ : starting from these six conformations, only one unconstrained minimum was obtained (Table 3). In this minimum, the dihedrals  $\beta$ ,  $\gamma$  and  $\delta$  (not shown here) are close to their B modal values but the dihedral angle  $\epsilon$  is typical of a  $B_{II}$  conformation. Moreover, the dihedral angles  $\chi$  are in  $+ap$  and  $-ap$  in NH3-RIB-A and NH3-RIB-U, respectively instead of being in  $-ac$  as in a  $B_{II}$  conformation. Therefore the unconstrained minimum is noted “ $B_{II}$ ”. The anomalous glycoside angle is favored because the  $H_{X2}$  proton is strongly attracted by the  $O_2$  atom of the uracil or the  $N_3$  atom of the adenine (Fig. 8). Because of the anomalous  $\chi$  the  $C_6-H_6 \dots O_5'$  and  $C_1'-H_1' \dots O_2$  hydrogen bonds vanished from NH3-RIB-U and the  $C_8-H_8 \dots O_5'$  and  $C_1'-H_1' \dots N_3$  hydrogen bonds from NH3-RIB-A. Moreover, the  $C_8-H_8 \dots O_4'$  hydrogen bond present only in the N conformations of NH2-RIB-A is observed in this S-type NH3-RIB-A.

We tried to decrease the strength of the  $N_X-H_{X2} \dots O_2$  hydrogen bond by embedding NH3-RIB-U into a continuous medium having the dielectric constant of water.

In implicit solvent, NH3-RIB-U has five unconstrained minima corresponding to “ $B_{II}$ ”,  $B_{II}$ ,  $B_{III}$ , A and  $A_{III}$  (Table 3). These conformations appear in the order “ $B_{II}$ ”  $< B_{II} < A < B_{III} < A_{III}$ . As the  $B_I$  conformation was not obtained, the  $\Delta E$  values were calculated relative to the “ $B_{II}$ ” conformation; all the other minima are energetically unfavorable.

To interfere with the  $N_X-H_{X2} \dots O_2$  hydrogen bonds, we introduced a water molecule to interact with the  $O_2$  atom of NH3-RIB-U. The result was two unconstrained minima for NH3-RIB-U (Table 3). The most favorable conformation is the normal  $B_{II}$  conformation with a glycoside angle in  $-ac$  (Fig. 8). The water molecule makes a bridge between the sugar and the uracil base that destroys the direct  $N_X-H_{X2} \dots O_2$  hydrogen bond:  $O_{WT}$  interacts with  $H_{X2}$  and  $H_{2WT}$  with  $O_2$ . Therefore, as expected in a  $B_{II}$  conformation, the glycoside angle is in  $-ac$ , and the  $C_6-H_6 \dots O_5'$  and  $C_1'-H_1' \dots O_2$  hydrogen bonds give stability to the structure. A second minimum with N puckering was obtained with an abnormal dihedral angle  $\epsilon$  in the  $-ac$  range (Table 3).

Thus, NH3-RIB-U mainly has S puckering. However, it has a  $B_{II}$  backbone and this might be the reason why its insertion into a DNA duplex causes destabilization.

-----  
- Insert Table 3 about here -  
-----

-----  
- Insert Figure 8 about here -  
-----

**NH2-ARA-A and NH2-ARA-U**

Both NH2-ARA-A and NH2-ARA-U have, at the B3LYP/6-31G\* level, three unconstrained minima corresponding to B<sub>I</sub>, B<sub>II</sub> and B<sub>III</sub> and three constrained minima corresponding to A, A<sub>II</sub> and A<sub>III</sub> where the dihedral angle  $\varepsilon$  had to be frozen to prevent the formation of the H<sub>3T</sub>...N<sub>X</sub> hydrogen bond (Table 4). For both compounds, the six minimized conformations appear in the same order: A < B<sub>I</sub> < A<sub>III</sub> < A<sub>II</sub> < B<sub>II</sub> < B<sub>III</sub>. The A conformation is the most energetically favorable conformation. The east energy barriers separating the B<sub>I</sub> and A conformations were estimated from the energy of the O4'-endo conformation obtained by freezing  $\tau_2=0^\circ$ : 3.44 kcal/mol for NH2-ARA-A and 4.10 kcal/mol for NH2-ARA-U. The puckering amplitudes at the east energy barrier ( $\tau=7.20$  for NH2-ARA-A and  $\tau=8.80$ , for NH2-ARA-U) are significantly smaller than the amplitudes in the B<sub>I</sub> and A conformations ( $\tau\approx 34$ ).

In implicit solvent, NH2-ARA-A has five unconstrained minima corresponding to B<sub>I</sub>, B<sub>II</sub>, B<sub>III</sub>, A<sub>II</sub> and A<sub>III</sub> and one constrained minimum corresponding to A (Table 4). Other than A<sub>III</sub>, the order is the same as *in vacuo*: A < A<sub>III</sub> < B<sub>I</sub> < A<sub>II</sub> < B<sub>II</sub> < B<sub>III</sub>.

-----  
 - Insert Table 4 about here -  
 -----

All N conformations of NH2-ARA-A are stabilized by the same hydrogen bonds *in vacuo* and in implicit solvent. In addition to the hydrogen bonds that stabilize H-A (Fig 5), we found N<sub>X</sub>-H<sub>X1</sub>...N<sub>3</sub> and N<sub>X</sub>-H<sub>X2</sub>...N<sub>9</sub> (Fig. 9). The hydrogen bonds stabilizing the S conformations are shown in Fig. 9.

-----  
 - Insert Figure 9 about here -  
 -----

**NH3-ARA-A and NH3-ARA-U**

*In vacuo*, NH3-ARA-A has four unconstrained minima (Table 5). In all these minima, the dihedrals  $\beta$  and  $\gamma$  (not shown here) and  $\varepsilon$  fall in their modal ranges. However, the S starting conformations do not have C2' endo puckering and display C1'exo puckering after minimization. Moreover, the glycoside angle  $\chi$  is *syn* instead of being *anti*. Therefore, these conformations are called "B<sub>I</sub>" and "B<sub>II</sub>". In the minima obtained from starting N conformations, the puckering is no longer C3' endo but C2' exo, and the dihedral angles  $\chi$  are not *-ap* but *+ac*. Therefore, we called these two minima "A<sub>II</sub>" and "A<sub>III</sub>". The four minima appear in the order "B<sub>I</sub>" < "B<sub>II</sub>" < "A<sub>II</sub>" < "A<sub>III</sub>". The anomalous values of P and  $\chi$  in these conformations are favored because the H<sub>X1</sub> atom is strongly attracted by N<sub>3</sub>, especially in the N conformations (Fig. 10). When the solvent is represented implicitly, the N<sub>X</sub>-H<sub>X1</sub>...N<sub>3</sub> hydrogen bond is weaker and variables including P and  $\chi$  become closer to their canonical values (see the results concerning NH3-ARA-A at the B3LYP/CPCM level show, Table 5).

*In vacuo*, NH<sub>3</sub>-ARA-U has five unconstrained minima (Table 5). The three S conformations present anomalies that are similar to those described above for NH<sub>3</sub>-ARA-A. Here the strong interaction of H<sub>X1</sub> with O<sub>2</sub> is probably involved.

-----  
- Insert Table 5 about here -  
-----

-----  
- Insert Figure 10 about here -  
-----

## CONCLUSION

In this work, we have confirmed that the conformations of deoxyadenosines (H-A) when embedded in double-stranded RNAs or DNAs are close to local minima of the nucleoside by itself *in vacuo*, when its backbone is forced to be close to the backbone in duplexes. The A and B DNA populations may be inferred from the local minima (39). Rare conformations correspond to the rather high energy B<sub>III</sub> local minimum that was found here. Therefore, *ab initio* calculations on nucleoside *in vacuo* reproduce the flexibility of duplexes. When the whole nucleoside is embedded in a continuous medium, the ordering of conformations is not as good. This may be because nucleosides in DNA are not fully in contact with water. Less is known about 2'-aminoribonucleosides than deoxynucleosides. It is clear that the simple search for the global minimum would not lead to the right solution. To reproduce the puckering of 2'-aminoribonucleosides *in silico*, we had to introduce an explicit water molecule into the model. The model was built by analogy with fine structural features of ribonucleosides. According to this model, the effect of the insertion of a modified nucleoside on the melting temperature and stability of a DNA or RNA duplex depends on its flexibility. The 2'-aminoribonucleoside prefers the N conformation, which destabilizes a DNA but it is less rigid than a natural ribose, which destabilizes a RNA. We also present the structure of 2'-aminoarabinonucleosides, and hopefully this structure will be confirmed experimentally. This work was motivated by attempts to understand the activity of integrase on viral DNA containing modified nucleosides (41). Indeed, integrase may be sensitive not only to the formation of new hydrogen bonds but also to a modification of the flexibility of DNA in some indirect recognition mechanism.

## REFERENCES

1. Franklin, S. E., and Gosling R.G. 1953. Molecular configuration in sodium thymonucleate. *Nature* 171:740-741.
2. Leslie, A. G., S. Arnott, R. Chandrasekaran, and R. L. Ratliff. 1980. Polymorphism of DNA double helices. *J. Mol. Biol.* 143:49-72.
3. Dickerson, R. E., H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini, and M. L. Kopka. 1982. The anatomy of A-, B- and Z- DNA. *Science* 216:475-485.
4. Lu, X. J., Z. Shakked, and W. K. Olson. 2000. A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* 300:819-840.
5. Tolstorukov, M. Y., V. I. Ivanov, G. G. Malenkov, R. L. Jernigan, and V. B. Zhurkin. 2001. Sequence-dependent B $\leftrightarrow$ A transition in DNA evaluated with dimeric and trimeric scales. *Biophys. J.* 81:3409-3421.
6. Tolstorukov, M. Y., R. L. Jernigan, and V. B. Zhurkin. 2004. Protein-DNA hydrophobic recognition in the minor groove is facilitated by sugar switching. *J. Mol. Biol.* 337:65-76.
7. Saenger W. 1984. *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.
8. Stein, D. S., and K. H. Moore. 2001. Phosphorylation of nucleoside analog antiretrovirals: a review for clinicians. *Pharmacotherapy* 21:11-34.
9. Gallant, J. E., P. Z. Gerondelis, M. A. Wainberg, N. S. Shulman, R. H. Haubrich, M. St Clair, E. R. Lanier, N. S. Hellmann, and D. D. Richman. 2003. Nucleoside and nucleotide analogue reverse transcriptase inhibitors: a clinical review of antiretroviral resistance. *Antivir. Ther.* 8:489-506.
10. Jellinek, D., L. S. Green, C. Bell, C. K. Lynott, N. Gill, C. Vargeese, G. Kirschenheuter, D. P. McGee, P. Abesinghe, W. A. Pieken, and . 1995. Potent 2'-amino-2'-deoxypyrimidine RNA inhibitors of basic fibroblast growth factor. *Biochemistry* 34:11363-11372.
11. Yan, X., X. Gao, and Z. Zhang. 2004. Isolation and characterization of 2'-amino-modified RNA aptamers for human TNF $\alpha$ . *Genomics Proteomics. Bioinformatics.* 2:32-42.
12. Sigurdsson, S. T., and F. Eckstein. 1996. Site specific labelling of sugar residues in oligoribonucleotides: Reactions of aliphatic isocyanates with 2' amino groups. *Nucleic Acids Research* 24:3129-3133.
13. Pham, J. W., I. Radhakrishnan, and E. J. Sontheimer. 2004. Thermodynamic and structural characterization of 2'-nitrogen-modified RNA duplexes. *Nucleic Acids Res.* 32:3446-3455.
14. Ora, M., H. Linjalahti, and H. Lonnberg. 2005. Phosphodiester cleavage of guanylyl-(3',3')-(2'-amino-2'-deoxyuridine): rate acceleration by the 2'-amino function. *J. Am. Chem. Soc* 127:1826-1832.
15. John, D. M., E. J. Merino, and K. M. Weeks. 2004. Mechanics of DNA flexibility visualized by selective 2'-amine acylation at nucleotide bulges. *J. Mol. Biol.* 337:611-619.
16. Borda, E. J., and S. T. Sigurdsson. 2005. Investigation of Mg $^{2+}$ - and temperature-dependent folding of the hairpin ribozyme by photo-crosslinking: effects of photo-crosslinker tether length and chemistry. *Nucleic Acids Res.* 33:1058-1068.
17. Olson W.K. 1982. How Flexible Is the Furanose Ring? 2. An Updated Potential Energy Estimate. *J. Am. Chem. Soc* 104:278-286.
18. Plavec, J., W. M. Tong, and J. Chattopadhyaya. 1993. How do the Gauche and Anomeric Effects Drive the Pseudorotational Equilibrium of the Pentofuranose Moiety of Nucleosides. *Journal of the American Chemical Society* 115:9734-9746.



19. Uesugi S., Miki H., Ikehara M., Iwahashi H., and Kyogoku Y. 1979. A linear relationship between electronegativity of 2'-substituents and conformation of adenine nucleosides. *Tetrahedron Letters* 42:4073-4076.
20. Guschlbauer, W., and K. Jankowski. 1980. Nucleoside conformation is determined by the electronegativity of the sugar substituent. *Nucleic Acids Res.* 8:1421-1433.
21. Venkateswarlu, D., K. E. Lind, V. Mohan, M. Manoharan, and D. M. Ferguson. 1999. Structural properties of DNA:RNA duplexes containing 2'-O-methyl and 2'-S-methyl substitutions: a molecular dynamics investigation. *Nucleic Acids Res.* 27:2189-2195.
22. Aurup, H., T. Tuschl, F. Benseler, J. Ludwig, and F. Eckstein. 1994. Oligonucleotide duplexes containing 2'-amino-2'-deoxycytidines: thermal stability and chemical reactivity. *Nucleic Acids Res.* 22:20-24.
23. Aurup, H., T. Tuschl, F. Benseler, J. Ludwig, and F. Eckstein. 1994. Oligonucleotide duplexes containing 2'-amino-2'-deoxycytidines: thermal stability and chemical reactivity. *Nucleic Acids Res.* 22:20-24.
24. Gherghe, C. M., J. M. Krahn, and K. M. Weeks. 2005. Crystal structures, reactivity and inferred acylation transition states for 2'-amine substituted RNA. *J. Am. Chem. Soc* 127:13622-13628.
25. Foloppe, N., and A. D. MacKerell, Jr. 1999. Intrinsic conformational properties of deoxyribonucleosides: implicated role for cytosine in the equilibrium among the A, B, and Z forms of DNA. *Biophys. J.* 76:3206-3218.
26. Foloppe, N., L. Nilsson, and A. D. MacKerell, Jr. 2001. Ab initio conformational analysis of nucleic acid components: intrinsic energetic contributions to nucleic acid structure and dynamics. *Biopolymers* 61:61-76.
27. Brameld, K. A., and W. A. Goddard. 1999. Ab initio quantum mechanical study of the structures and energies for the pseudorotation of 5'-dehydroxy analogues of 2'-deoxyribose and ribose sugars. *Journal of the American Chemical Society* 121:985-993.
28. Frisch M.J., Trucks G.W., Schlegel H.B., Scuseria G.E., Robb M.A., Cheeseman J.R., Zakrzewski V.G., Montgomery J.A., Stratmann Jr.R.E., Burant J.C., Dapprich S., Millam J.M., Daniels A.D., Kudin K.N., Strain M.C., Farkas O., Tomasi J., Barone V., Cossi M., Cammi R., Mennucci B., Pomelli C., Adamo C., Clifford S., Ochterski J., Petersson G.A., Ayala P.Y., Cui Q., Morokuma K., Malick D.K, Rabuck A.D., Raghavachari K., Foresman J.B., Cioslowski J., Ortiz J.V., Stefanov B.B., Liu G., Liashenko A., Piskorz P., Komaromi I., Gomperts R., Martin R.L., Fox D.J., Keith T., Al-Laham M.A., Peng C.Y., Nanayakkara A., Gonzalez C., Challacombe M., Gill P.M.W., Johnson B., Chen W., Wong M.W., Andres J.L, Gonzalez C., Head-Gordon M., Replogle E.S., and Pople J.A. 1998. Gaussian 98. Gaussian, Inc., Pittsburgh PA.
29. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, Jr. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C.

- Gonzalez, and J. A. Pople. 2003. Gaussian 03, Revision A.1. Gaussian, Inc., Pittsburgh PA, 2003.
30. Barone V., Cossi M., and Tomasi J. 1997. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem Phys* 107:3210-3221.
  31. Barone V., and Cossi M. 1998. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys Chem A*. 102:1995-2001.
  32. Rao, S., E. Westhof, and M. Sundralingam. 1981. Exact method for the calculation of pseudorotation parameters  $P$ ,  $[\tau]_m$  and their errors. A comparison of the Altona-Sundaralingam and Cremer-Pople treatment of puckering of five-membered rings. *Acta Crystallogr. A* 37:421-425.
  33. Berman, H. M., W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S. H. Hsieh, A. R. Srinivasan, and B. Schneider. 1992. The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* 63:751-759.
  34. Foloppe, N., L. Nilsson, and A. D. MacKerell, Jr. 2001. Ab initio conformational analysis of nucleic acid components: intrinsic energetic contributions to nucleic acid structure and dynamics. *Biopolymers* 61:61-76.
  35. Le Bret M., Gabarro-Arpa J., Gilbert J.C., and Lemarechal C. 1991. MORCAD, an object-oriented molecular modelling package running on IBM RS/6000 and SGI 4Dxxx workstations. *J. Chim. Phys.* 88:2489-2496.
  36. Isaacs, R. J., and H. P. Spielmann. 2001. NMR evidence for mechanical coupling of phosphate B(I)-B(II) transitions with deoxyribose conformational exchange in DNA. *J. Mol. Biol.* 311:149-160.
  37. Arora, K., and T. Schlick. 2003. Deoxyadenosine sugar puckering pathway simulated by the stochastic difference equation algorithm. *Chemical Physics Letters* 378:1-8.
  38. Auffinger, P., and E. Westhof. 1997. Rules governing the orientation of the 2'-hydroxyl group in RNA. *J. Mol. Biol.* 274:54-63.
  39. Foloppe, N., and A. D. MacKerell. 1998. Conformational Properties of the Deoxyribose and Ribose Moieties of Nucleic Acids: A Quantum Mechanical Study. *J. Phys. Chem. B*. 102:6669-6678.
  40. Egli, M., S. Portmann, and N. Usman. 1996. RNA hydration: a detailed look. *Biochemistry* 35:8489-8494.
  41. Agapkina, J., M. Smolov, S. Barbe, E. Zubin, T. Zatsopin, E. Deprez, M. Le Bret, J. F. Mouscadet, and M. Gottikh. 2006. Probing of HIV-1 integrase-DNA interactions using novel analogs of viral DNA. *J. Biol. Chem.*

**TABLE 1-** Conformational angles (degrees) and relative energy values (kcal/mol) of H-A conformations obtained by geometry optimization using several methods of calculation.

Methods of calculation	P	$\epsilon$	$\chi$	$\Delta E$	Forms
HF/6-31G*	16.29	-93.06	-158.19	0.95	A <sub>II</sub>
	13.94	69.12	-159.96	1.41	A <sub>III</sub>
	13.94	-169.00	-161.35	0.35	A
	165.17	-67.44	-126.73	0.73	B <sub>II</sub>
	169.00	59.93	-126.36	1.55	B <sub>III</sub>
	161.71	176.60	-131.79	0.00	B <sub>I</sub>
B3LYP/6-31G*	11.48	-88.25	-148.92	0.91	A <sub>II</sub>
	9.78	66.11	-150.68	0.94	A <sub>III</sub>
	13.92	-168.97	-161.42	0.48	A
	174.70	-63.96	-122.39	0.46	B <sub>II</sub>
	-176.23	56.69	-120.07	0.50	B <sub>III</sub>
	161.70	176.49	-131.68	0.00	B <sub>I</sub>
B3LYP/CPCM /6-31G*	7.00	-87.79	-151.02	1.25	A <sub>II</sub>
	8.16	59.68	-154.20	0.76	A <sub>III</sub>
	8.63	-165.20	-150.74	1.07	A
	174.67	-63.97	-122.43	0.04	B <sub>II</sub>
	-174.69	54.06	-120.43	0.85	B <sub>III</sub>
	178.95	175.96	-121.35	0.00	B <sub>I</sub>

**TABLE 2-** Conformational angles (degrees) and relative energy values (kcal/mol) of NH2-RIB-A and NH2-RIB-U conformations obtained by geometry optimization using several methods of calculation.

Nucleoside analogues	Methods of calculation	P	$\varepsilon$	$\chi$	$\Delta E$	Forms
NH2-RIB-A	B3LYP/6-31G*	8.75	-92.51	-155.51	5.80	A <sub>II</sub>
		3.23	30.00*	-161.29	9.70	A <sub>III</sub>
		16.42	-155.00*	-155.08	3.90	A
		170.70	-75.58	-130.14	1.48	B <sub>II</sub>
		180.00	41.36	-125.18	1.97	B <sub>III</sub>
		165.20	-155.00*	-134.06	0.00	B <sub>I</sub>
NH2-RIB-U	B3LYP/6-31G*	14.33	-96.54	-160.44	5.40	A <sub>II</sub>
		7.07	30.00*	-162.89	5.16	A <sub>III</sub>
		166.75	-77.55	-133.70	1.58	B <sub>II</sub>
		172.15	30.00*	-121.14	2.15	B <sub>III</sub>
		160.78	-155.00*	-138.96	0.00	B <sub>I</sub>
		B3LYP/6-31G* <sup>†</sup>	20.22	30.00*	-161.14	-1.50
	15.25		-154.29	-162.43	-1.61	A
	165.08		-65.77	-124.55	-0.39	B <sub>II</sub>
	171.97		30.00*	-121.19	-0.01	B <sub>III</sub>
	159.34		-155.00*	-126.42	0.00	B <sub>I</sub>
	B3LYP/CPCM /6-31G*		9.74	-88.82	-159.47	3.22
		9.10	30.00*	-159.81	3.19	A <sub>III</sub>
		11.62	171.42	-158.96	2.20	A
		171.80	-71.50	-131.59	-1.18	B <sub>II</sub>
		171.50	30.00*	-121.71	3.46	B <sub>III</sub>
		170.62	-155.00*	-126.53	0.00	B <sub>I</sub>
	B3LYP/6-31G* + WT <sup>‡</sup>	11.87	-89.48	-158.43	4.09	A <sub>II</sub>
		6.54	30.00*	-160.83	3.45	A <sub>III</sub>
		15.30	-155.00*	-157.53	2.19	A
		168.89	-79.65	-130.75	1.96	B <sub>II</sub>
		175.45	30.00*	-127.66	2.80	B <sub>III</sub>
		163.19	-155.00*	-134.28	0.00	B <sub>I</sub>
	B3LYP/6-31G* + WT <sup>¶</sup>	14.58	30.00*	-146.72	2.22	A <sub>III</sub>
		11.59	-155.00*	-152.47	-0.48	A
168.43		-66.87	-119.08	-0.41	B <sub>II</sub>	
173.79		30.00*	-121.83	-0.23	B <sub>III</sub>	
161.48		-155.00*	-126.26	0.00	B <sub>I</sub>	

\* Dihedral angle  $\varepsilon$  frozen during the minimization.<sup>†</sup> Model 2: H<sub>O2</sub> is distant from O<sub>3</sub>.<sup>‡</sup> First initial orientation of the water molecule: H<sub>2WT</sub> and O<sub>WT</sub> of the water molecule form hydrogen bonds with O<sub>2</sub> and H<sub>X2</sub> atoms of the 2'-aminonucleoside, respectively.<sup>¶</sup> Second initial orientation of the water molecule: H<sub>2WT</sub> and H<sub>1WT</sub> of the water molecule form hydrogen bonds with O<sub>2</sub> and N<sub>X</sub> of the 2'-aminonucleoside, respectively.

**TABLE 3-** Conformational angles (degrees) and relative energy values (kcal/mol) of NH3-RIB-A and NH3-RIB-U conformations obtained by geometry optimization using several methods of calculation.

Nucleoside analogues	Methods of calculation	P	$\varepsilon$	$\chi$	$\Delta E$	Forms
NH3-RIB-A	B3LYP/6-31G*	165.78	-73.91	177.21	0.00	“B <sub>II</sub> ”
NH3-RIB-U	B3LYP/6-31G*	-178.91	-74.87	-168.35	0.00	“B <sub>II</sub> ”
	B3LYP/CPCM /6-31G*	17.02	35.32	-164.66	5.61	A <sub>III</sub>
		15.62	-150.76	-161.04	3.16	A
		174.64	-90.20	-124.63	2.88	B <sub>II</sub>
		-170.21	-85.82	-166.52	0.00	“B <sub>II</sub> ”
		174.68	32.94	-123.99	3.95	B <sub>III</sub>
	B3LYP/6-31G* + WT	11.84	-116.10	-163.93	0.87	“A”
171.18		-77.14	-123.69	0.00	B <sub>II</sub>	

**TABLE 4-** Conformational angles (degrees) and relative energy values (kcal/mol) of NH2-ARA-A and NH2-ARA-U conformations obtained by geometry optimization using several methods of calculation.

Nucleoside analogues	Methods of calculation	P	$\varepsilon$	$\chi$	$\Delta E$	Forms		
NH2-ARA-A	B3LYP/6-31G*	9.86	-90.00*	-154.42	1.10	A <sub>II</sub>		
		7.00	60.00*	-151.10	0.86	A <sub>III</sub>		
		8.04	-155.00*	-155.87	-1.21	A		
		158.55	-64.79	-140.99	1.54	B <sub>II</sub>		
		160.80	60.41	-140.07	1.85	B <sub>III</sub>		
		151.08	178.33	-144.35	0.00	B <sub>I</sub>		
	B3LYP/CPCM /6-31G*	7.96	-88.92	-145.58	0.14	A <sub>II</sub>		
		7.05	63.75	-149.87	-0.32	A <sub>III</sub>		
		7.42	-155.00*	-148.25	-0.50	A		
		170.80	-64.69	-137.49	0.45	B <sub>II</sub>		
		171.87	56.23	-137.88	1.37	B <sub>III</sub>		
		174.13	170.70	-141.58	0.00	B <sub>I</sub>		
		NH2-ARA-U	B3LYP/6-31G*	15.16	-90.00*	-152.76	1.17	A <sub>II</sub>
				12.13	60.00*	-153.92	0.95	A <sub>III</sub>
13.82	-155.00*			-155.14	-1.18	A		
156.44	-65.84			-143.41	1.70	B <sub>II</sub>		
154.93	63.06			-143.40	1.92	B <sub>III</sub>		
150.34	177.67			-147.49	0.00	B <sub>I</sub>		

\* Dihedral angle  $\varepsilon$  frozen during the minimization

**TABLE 5-** Conformational angles (degrees) and relative energy values (kcal/mol) of NH3-ARA-A and NH3-ARA-U conformations obtained by geometry optimization using several methods of calculation.

Nucleoside analogues	Methods of calculation	P	$\varepsilon$	$\chi$	$\Delta E$	Forms	
NH 3-ARA-A	B3LYP/6-31G*	-23.56	-84.84	145.75	1.30	“A <sub>II</sub> ”	
		-26.74	49.95	140.94	2.06	“A <sub>III</sub> ”	
		138.44	-73.01	84.20	0.21	“B <sub>II</sub> ”	
		137.05	-164.0	82.87	0.00	“B <sub>I</sub> ”	
	B3LYP/CPCM /6-31G*	0.33	-94.82	-164.11	2.64	A <sub>II</sub>	
		0.43	55.88	-164.48	2.28	A <sub>III</sub>	
		-0.24	-141.10	-164.22	2.66	A	
		172.73	-68.09	-139.58	0.62	B <sub>II</sub>	
		171.90	48.91	-142.23	1.50	B <sub>III</sub>	
		166.77	-175.11	-144.53	0.00	B <sub>I</sub>	
	NH 3-ARA-U	B3LYP/6-31G*	5.86	-79.87	-160.62	5.87	A <sub>II</sub>
			3.43	42.63	-160.62	7.59	A <sub>III</sub>
			138.82	-75.23	107.63	0.32	“B <sub>II</sub> ”
			138.11	31.40	118.46	3.33	“B <sub>III</sub> ”
137.51			-163.41	107.30	0.00	“B <sub>I</sub> ”	

**FIGURE LEGENDS**

**FIGURE 1-** Atom numbering and chemical structure of the compounds studied.

**FIGURE 2-** B<sub>I</sub> and A conformations of H-A obtained by geometry optimization at the B3LYP/CPCM/6-31G\* level of theory. Hydrogen bonds are represented with dashed lines.

**FIGURE 3-** The relaxed potential map of H-A at the B3LYP/6-31G\* level of theory. The points show the conformations that have been minimized. The small numbers are the number of dihedrals frozen during the minimization. The inversion point is obtained when all five endocyclic dihedrals are zero and can be recognized by the mark 5. The circles are centered at the inversion point, when the ring is planar, and represent conformations of constant amplitudes  $\tau = 10^\circ, 20^\circ, 30^\circ$  and  $40^\circ$ . The lowest isopotential (0.5 kcal/mol) is represented by a dashed line circling the two unconstrained minima (mark 0). The continuous and dashed lines represent isopotentials separated by 0.5 kcal/mol. Integral energy values are shown on the periphery (big bold numbers). There is clearly only one valley on the east connecting the two unconstrained minima.

**FIGURE 4-** B3LYP/6-31G\* energy (kcal/mol) of H-A as a function of the pseudorotation amplitude  $\tau$ . Each point represents a minimization with two dihedral constraints.  $\tau_2=0$ .  $\tau_3$  is varied to monitor the pseudorotation. Starting conformation:  $\beta=168^\circ \chi = -80^\circ$ ,  $\Delta$ ;  $\beta=168^\circ \chi = 180^\circ$   $\square$ ;  $\beta=75^\circ \chi = -80^\circ$ ,  $\blacktriangle$ ;  $\beta=75^\circ \chi = 168^\circ$   $\blacksquare$ .

**FIGURE 5-** Conformations of NH<sub>2</sub>-RIB-U, H<sub>X2</sub> pointing away from O<sub>3</sub>' optimized at the B3LYP/6-31G\* level of theory. Left: B<sub>I</sub>; right: A<sub>III</sub>. Hydrogen bonds are represented by dashed lines

**FIGURE 6-** A conformation of NH<sub>2</sub>-RIB-U, H<sub>X2</sub> pointing toward O<sub>3</sub>, optimized at the B3LYP/6-31G\* level of theory. Hydrogen bonds are represented by dashed lines

**FIGURE 7-** Conformations of NH<sub>2</sub>-RIB-U with a water molecule. Left: the B<sub>I</sub> conformation favored by the water molecule in the orientation WT<sup>1</sup>. Right: the A conformation favored by the water molecule in the WT<sup>2</sup> orientation. Hydrogen bonds are represented by dashed lines.

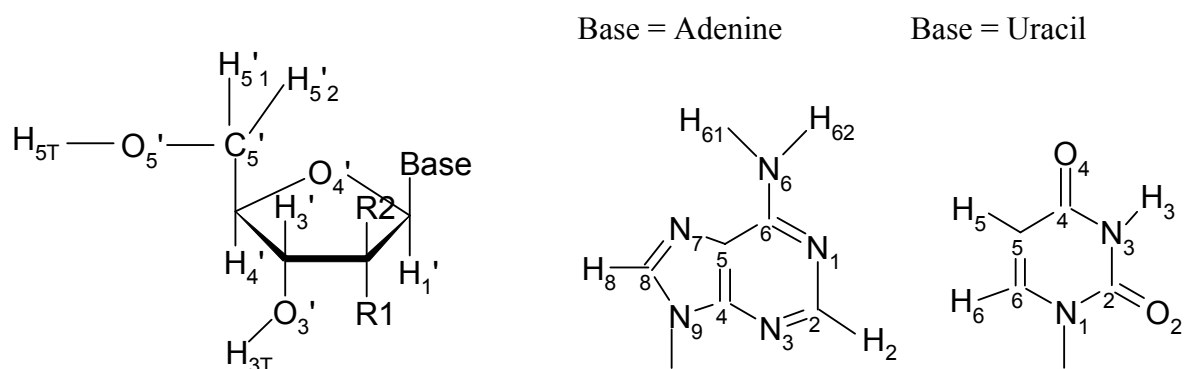
**FIGURE 8-** “B<sub>II</sub>” conformations of NH<sub>3</sub>-RIB-A and NH<sub>3</sub>-RIB-U obtained by geometry optimization at the B3LYP/6-31G\* level of theory (A). B<sub>II</sub> and A conformations of NH<sub>3</sub>-RIB-U obtained by geometry optimization from the B3LYP/6-31G\* + WT method of calculation (B). Hydrogen bonds are represented with dashed lines.

**FIGURE 9-** B<sub>I</sub> and A conformations of NH<sub>2</sub>-ARA-A obtained by geometry optimization at the B3LYP/6-31G\* level of theory. Hydrogen bonds are represented with dashed lines.

**FIGURE 10-** “B<sub>I</sub>” and “A<sub>II</sub>” conformations of NH<sub>3</sub>-ARA-A obtained by geometry optimization at the B3LYP/6-31G\* level of theory. Hydrogen bonds are represented with dashed lines.



FIGURE 1



Compounds	Bases	R1	R2
H-A	Adenine	$\begin{array}{c}   \\ \text{H}_{2,1}' \end{array}$	$\begin{array}{c}   \\ \text{H}_{2,2}' \end{array}$
NH2-RIB-A	Adenine	$\begin{array}{c}   \\ \text{H}_{x1} - \text{N}_x \\ \quad \quad \quad \backslash \\ \quad \quad \quad \text{H}_{x2} \end{array}$	$\begin{array}{c}   \\ \text{H}_{2,2}' \end{array}$
NH2-RIB-U	Uracil	$\begin{array}{c}   \\ \text{H}_{x1} - \text{N}_x \\ \quad \quad \quad \backslash \\ \quad \quad \quad \text{H}_{x2} \end{array}$	$\begin{array}{c}   \\ \text{H}_{2,2}' \end{array}$
NH3-RIB-A	Adenine	$\begin{array}{c}   \\ \text{H}_{x1} - \text{N}_x - \text{H}_{x3} \\ \quad \quad \quad   \\ \quad \quad \quad \text{H}_{x2} \end{array}$	$\begin{array}{c}   \\ \text{H}_{2,2}' \end{array}$
NH3-RIB-U	Uracil	$\begin{array}{c}   \\ \text{H}_{x1} - \text{N}_x - \text{H}_{x3} \\ \quad \quad \quad   \\ \quad \quad \quad \text{H}_{x2} \end{array}$	$\begin{array}{c}   \\ \text{H}_{2,2}' \end{array}$
NH2-ARA-A	Adenine	$\begin{array}{c}   \\ \text{H}_{2,1}' \end{array}$	$\begin{array}{c} \text{H}_{x2} \\ \quad \quad \quad \backslash \\ \quad \quad \quad \text{N}_x - \text{H}_{x1} \\ \quad \quad \quad   \end{array}$
NH2-ARA-U	Uracil	$\begin{array}{c}   \\ \text{H}_{2,1}' \end{array}$	$\begin{array}{c} \text{H}_{x2} \\ \quad \quad \quad \backslash \\ \quad \quad \quad \text{N}_x - \text{H}_{x1} \\ \quad \quad \quad   \end{array}$
NH3-ARA-A	Adenine	$\begin{array}{c}   \\ \text{H}_{2,1}' \end{array}$	$\begin{array}{c} \text{H}_{x2} \\ \quad \quad \quad \backslash \\ \text{H}_{x3} - \text{N}_x - \text{H}_{x1} \\ \quad \quad \quad   \end{array}$
NH3-ARA-U	Uracil	$\begin{array}{c}   \\ \text{H}_{2,1}' \end{array}$	$\begin{array}{c} \text{H}_{x2} \\ \quad \quad \quad \backslash \\ \text{H}_{x3} - \text{N}_x - \text{H}_{x1} \\ \quad \quad \quad   \end{array}$

**FIGURE 2**

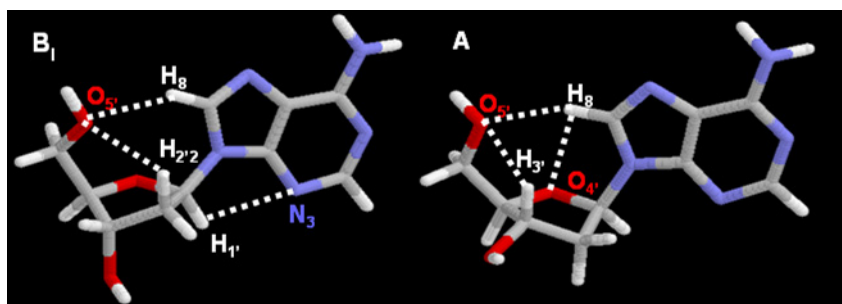


FIGURE 3

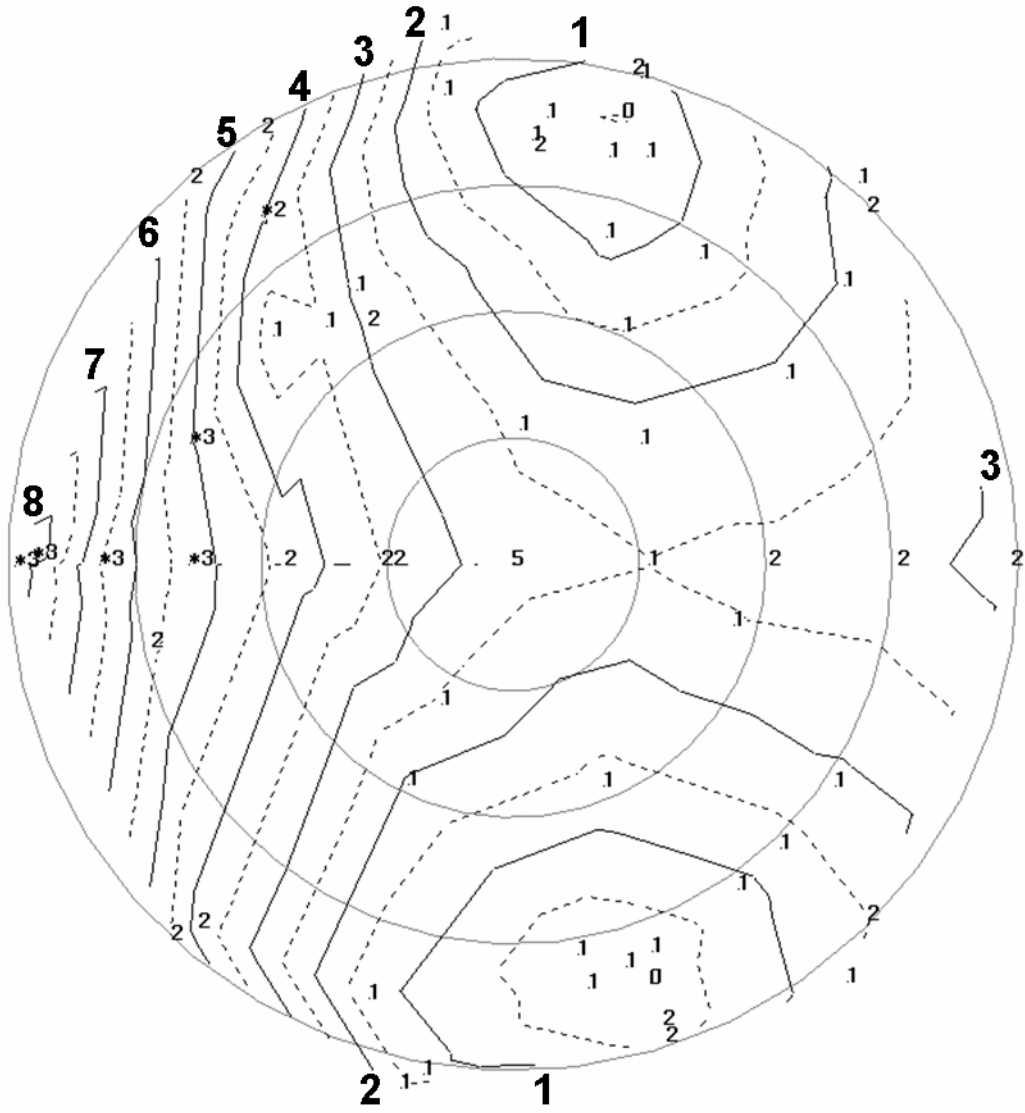
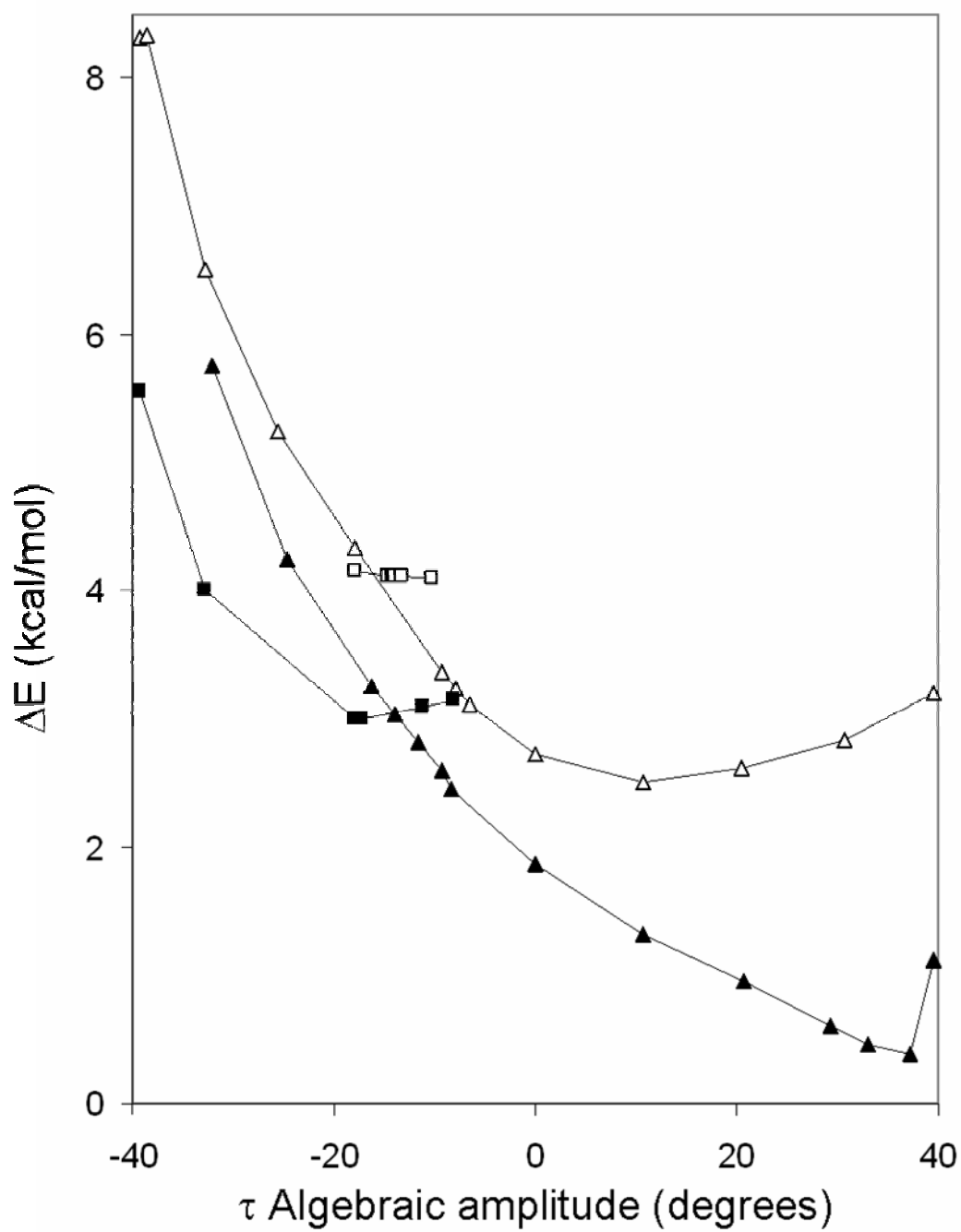
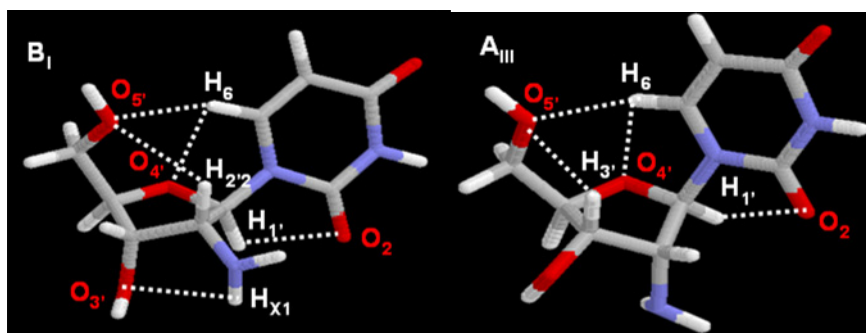


FIGURE 4



**FIGURE 5**



**FIGURE 6**

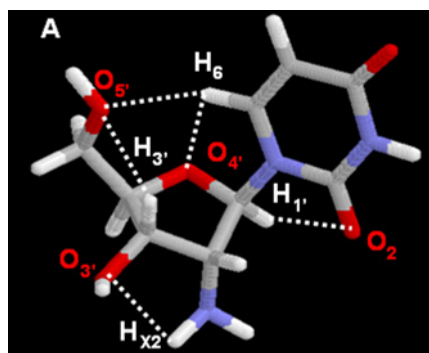


FIGURE 7

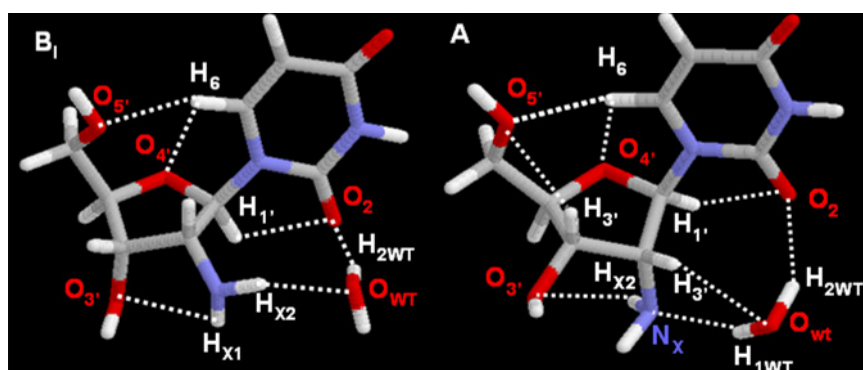
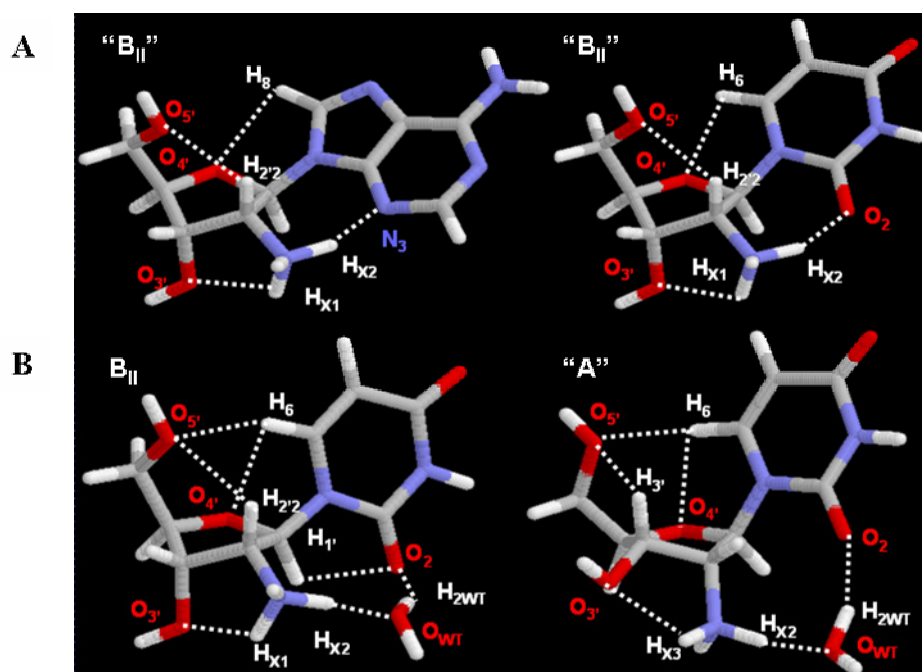


FIGURE 8





**FIGURE 9**

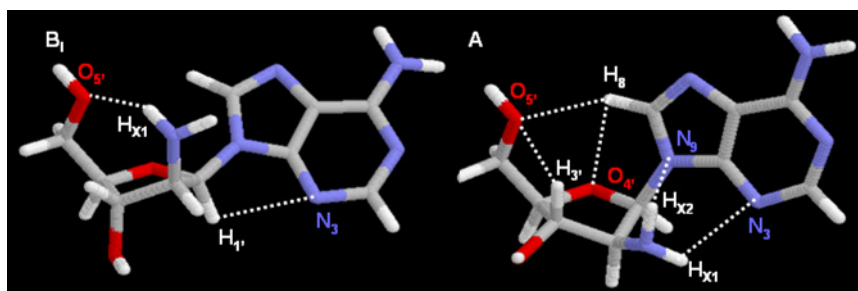
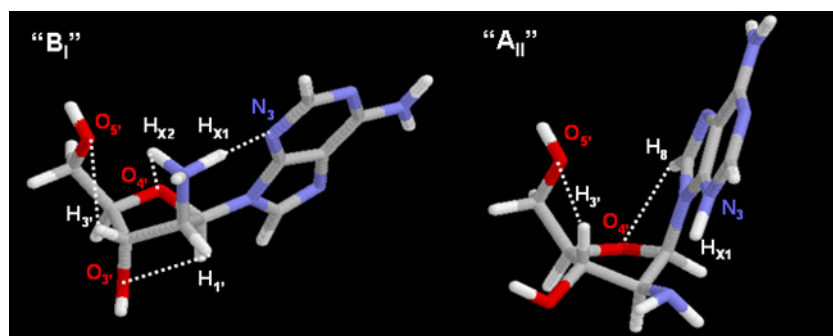


FIGURE 10



## Chapitre II

### Etude *ab initio* de l'hydratation et de la flexibilité de l'uridine, la 2'-désoxyuridine et la 2'-méthoxyuridine

---

Les doubles hélices d'ADN et d'ARN ont des propriétés de flexibilité différentes (Auffinger and Westhof, 2000; Auffinger and Westhof, 2001b; Cheatham and Kollman, 1997; Fujiwara and Shindo, 1985; Hagerman, 1988; Hagerman, 1997; Noy et al., 2004; Pan and MacKerell, 2003; Perez et al., 2004; Shindo et al., 1985). En particulier, les sucres de l'ARN apparaissent plus rigides que ceux de l'ADN (Auffinger and Westhof, 2000; Auffinger and Westhof, 2001b; Cheatham and Kollman, 1997; Pan and MacKerell, 2003). Dans l'ARN, les sucres ont un plissement Nord alors que dans l'ADN, ils sont en Sud mais ils peuvent transiter en Nord dans un milieu moins hydraté ou dans les complexes protéines/ADN (Lu et al., 2000; Tolstorukov et al., 2001; Tolstorukov et al., 2004). Le remplacement du 2'-H de l'ADN par un 2'-OH dans l'ARN modifie donc les propriétés conformationnelles et de flexibilité des sucres des doubles hélices.

Une double hélice d'ARN contenant des 2'-méthoxynucléotides apparaît structurellement plus stable que celle d'un ARN natif (Cummins et al., 1995; Egli et al., 1996; Lesnik and Freier, 1998). Le remplacement du 2'-OH hydrophile par un groupe 2'-méthoxy hydrophobe modifie donc les propriétés de flexibilité d'une double hélice.

Pour essayer de comprendre l'influence du groupe en 2' des sucres sur la conformation et la flexibilité des doubles hélices, nous avons réalisé une étude structurale par mécanique quantique *ab initio* (niveau théorique B3LYP/6-31G\*) de deux nucléosides naturels et d'un analogue de nucléoside qui ne diffèrent uniquement que par le substituant en position 2' : l'uridine, la 2'-désoxyuridine et la 2'-méthoxyuridine. L'hydratation jouant un rôle important

dans la conformation des acides nucléiques, les calculs ont été réalisés dans trois conditions : 1) *in vacuo* 2) en plongeant le système dans un milieu continu ayant les propriétés diélectriques de l'eau 3) *in vacuo* mais en présence d'une molécule d'eau explicite, décrite comme intervenant dans l'hydratation du petit sillon de l'ARN (molécule d'eau formant un pont hydrogène entre le groupe 2'-OH et l'O2 de l'uracile) et de l'ADN (molécule d'eau en contact avec l'O2 de l'uracile) (Auffinger and Westhof, 2000; Egli et al., 1996; Schneider et al., 1998).

Quelles que soient les conditions d'hydratation, le sucre de la 2'-désoxyuridine est en Sud dans la conformation la plus favorable qui peut être introduite dans une double hélice. *In vacuo*, le nucléoside est flexible et peut adopter une conformation Nord. En solvant implicite et en présence d'une molécule d'eau interagissant avec l'O2 de l'uracile, le nucléoside devient moins souple.

Dans des études antérieures de mécanique quantique, le plissement Nord de ribonucléosides ou d'un ribonucléotide n'était pas favorisé (Brameld and Goddard, 1999; Hocquet et al., 2000; Leulliot et al., 1999). Le groupe 2'-OH peut adopter différentes conformations. L'analyse de structures expérimentales combinée à des résultats de dynamique moléculaire indique que le 2'-OH s'oriente préférentiellement vers l'O3' dans les ribonucléotides Nord et Sud (Auffinger and Westhof, 1997). Nous avons pu montrer qu'avec une telle orientation du 2'-OH, le sucre de l'uridine est en Nord *in vacuo*, en présence et en absence d'une molécule d'eau pontant le 2'-OH et l'O2 de l'uracile. En l'absence de la molécule d'eau, l'uridine est flexible et peut adopter une conformation Sud. Par contre, en présence de la molécule d'eau explicite, l'uridine est rigide et ne peut être en Sud. La molécule d'eau est donc essentielle pour stabiliser un plissement Nord rigide du sucre.

Nous avons également réalisé l'étude conformationnelle de l'uridine avec le 2'-OH tourné du côté opposé à l'O3' : une orientation que peut adopter le 2'-OH des ribonucléotides Nord (Auffinger and Westhof, 1997). Dans ce cas, le plissement Nord du sucre est favorisé uniquement en présence de la molécule d'eau explicite. Le ribonucléoside est rigide et ne peut pas adopter une conformation Sud.

Ainsi, en présence d'un groupe OH en 2', une molécule d'eau formant un pont hydrogène entre la base et le sucre dans le petit sillon de l'ARN favorise un plissement Nord rigide du sucre des ribonucléosides.

L'analyse conformationnelle de la 2'-méthoxyuridine a été réalisée pour deux orientations du 2'-OCH<sub>3</sub>, semblables à celles du 2'-OH de l'uridine. Lorsque le 2'-OCH<sub>3</sub> est orienté vers

l'O3', le plissement Nord du sucre est favorisé *in vacuo*, en présence et absence de la molécule d'eau explicite. Cette conformation Nord est moins rigide que celle de l'uridine.

Des études cristallographiques (Adamiak et al., 1997; Lubini et al., 1994), de RMN (Popenda et al., 1997) et de dynamique moléculaire (Auffinger and Westhof, 2001a) indiquent que le 2'-OCH<sub>3</sub> s'oriente du côté opposé à l'O3' dans une double hélice. Avec une telle orientation du 2'-OCH<sub>3</sub>, le plissement Nord du sucre est également favorisé *in vacuo*, en présence et en absence de la molécule d'eau explicite. L'analogue de nucléoside est alors très flexible.

Le sucre de la 2'-méthoxyuridine a donc un plissement Nord beaucoup plus flexible que celui de l'uridine : ce qui confirme que la plus grande stabilité structurelle d'une double hélice contenant des 2'-méthoxynucléotides par rapport à celle d'un ARN natif (Cummins et al., 1995; Egli et al., 1996; Lesnik and Freier, 1998) n'est pas favorisée par une plus grande rigidité des sucres (Auffinger and Westhof, 2001a).

L'ensemble de ce travail est présenté dans le manuscrit joint ci-après.

## **Ab initio study of the hydration and the flexibility of uridine, 2'-deoxyuridine and 2'-methoxyuridine**

*Sophie Barbe\* and Marc Le Bret\**

\*UMR CNRS-8113, Laboratoire de Biotechnologies et Pharmacologie génétique Appliquée (L.B.P.A.) Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson, 94235 Cachan, France ;

Address reprint requests to Marc Le Bret, LBPA, ENS-Cachan, 61 avenue du Président Wilson, 94235 Cachan, France. Tél: 33-1-47-40-59-97; Fax: 33-1-47-40-24-79; E-mail: mlebret@lbpa.ens-cachan.fr.

Keywords: Quantum mechanics, 2' sugar group orientation, Nucleic acid stability, Solvation, Sugar puckering.

Running title: flexibility of natural and modified uridines,

## **ABSTRACT**

High-level DFT/B3LYP ab initio calculations are used to determine the pucker (north-type or south-type) of various compounds: uridine, the 2'-deoxyuridine and the 2'-O-methyl uridine. Even if the dihedrals of the backbone are set close to their experimental values in double stranded nucleic acids, calculations show that in vacuo or in a continuum mimicking the water dielectric properties, the north puckering conformations of uridine are not largely favored in contrast to the experimental data. We show here that the north puckering is favored when an explicit water molecule is introduced in the calculation. The influence of the orientations of the 2' group and of the water molecule in the prevalence of the north puckering is discussed. Several orientations of the water molecule binding uracil O2 and the 2' group are studied. The energy barriers in the path joining the north to south conformations are estimated. The north puckering is more favored in the 2'-OH than in the 2'-OCH3 compounds. When inserted in a double helix, uridine is expected to be more rigid than 2'-methoxy uridine or 2'-deoxy uridine.

## INTRODUCTION

Ribose and 2'-deoxyribose sugars are the basic subunits which differentiate RNA and DNA, respectively. The presence of the 2'-hydroxyl group on the RNA furanose sugar ring is at the origin of the profound structural and dynamical differences observed between RNA and DNA molecules. The only difference of an hydroxyl group is enough to confine RNA double helices to a single structural family (A form), whereas DNA is polymorphic and exists in a variety of structural families including the A, B and Z forms (1). Structural and biophysical studies have led to the conclusion that the additional hydroxyl group stabilizes the C3'-endo or North or N sugar pucker, typical of RNA, compared to the C2'-endo or South or S sugar pucker, prevalent in DNA systems (1-4). The 2'-hydroxyl group has major structural implications in that it is involved in recognition, processing, and catalytic properties of RNA, (5-7) such as the transesterification reactions involved in the Group I and Group II splicing reactions, (5-7) self-cleavage in lariat-RNA, (7) RNA catalysis in ribozyme and in ribonuclease action (6). The knowledge of the orientations and the dynamics of the 2'-hydroxyl bond is essential to understand its specific structural implications in the stabilization of helical regions, complex RNA tertiary folds and RNA/protein interactions. This knowledge is also important to study the mechanisms of catalytic RNA molecules. Crystallographic studies (8), RMN works (9) and dynamics simulations (10) suggest some 2'-OH orientations. Modified oligonucleotides are important for many applications, such as antisense therapeutics, diagnostics, profiling gene expressions with microarrays, identifying bands by Northern blots of gels and probing RNA structure. Oligonucleotides with 2'-O-alkyl modifications, as 2'-O methyl group can be particularly useful for these applications because they are easily synthesized, chemically stable and they form duplexes more stable than their unmodified counterparts, indicating that the replacement of a hydrophobic 2'-OH by a hydrophobic 2'-O-methyl group results in an additional stabilization of the helices (11-17). This effect is observed for the RNA, the DNA and hybrid duplexes. It is explained either by a stiffening of the N conformation of the ribose induced by the methyl group (18, 19) or a reduction of the hydration of the shallow groove (20) associated with decrease of the hydration enthalpy. It has also been proposed from molecular dynamic simulations of a fully modified 2'-O-methyl duplexes that, on the contrary, by creating optimal water binding pockets, the 2'-O-methylations lead to a significant stabilization of specific nucleotide-water interactions, which could favorably contribute to the enthalpy of hydration (21, 22).

Solvent interactions are a key factor in the conformational variability of nucleic acids (23-25) and binding of ligands to nucleic acids (26-32).

To predict the conformation of a modified nucleoside, gauche effects (33-38), anomeric effects (34, 35, 37, 39-42) and electronegativity (43, 44) could be tuned from a set of experimental structures. The ab initio approach is however more reliable when the flexibility of the nucleoside has to be determined. Natural deoxynucleosides (45, 46) are mainly found in the S conformation but their minor N population is correctly predicted. However, compounds with a 2'-OH furanose have been found prevalently in the N conformation (47, 48) when only the conformations compatible with the double helical conformations are considered. But the energy gap between the N and the S conformations is so small that in a double stranded RNA the A and B forms should be almost equally populated.

In this paper, we study the flexibility of three compounds which differ by the 2'-substituent, the uridine, the 2'-deoxyuridine and the 2'-O-methyl-2'-deoxyuridine. As the hydration plays an important role, we report calculations in vacuo, in a continuous medium mimicking the dielectric properties of water and in vacuo but in the presence of an explicit water molecule. The aim of this work is to determine the roles of the 2'-substituent and of the hydration on the



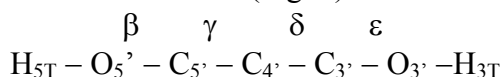
conformation of natural and 2'-modified nucleosides and to find a model such that the furanose puckering is well simulated.

## METHODS

Quantum mechanical calculations were carried out with the Gaussian 03 programs (49) using the 6-31 G\* basis set. Energy minimizations were performed to the default tolerances in the Gaussian programs. All degrees of freedom, other than those specified as being fixed, including all bond lengths and angles, were allowed to relax during the energy minimizations. One level of theory was used for the minimizations: the Density Functional Theory / Becke 3, Lee, Yang and Par (DFT/B3LYP) level which considers explicitly the correlation effects. The effect of the solvent has been estimated in two different ways. i) The nucleoside was embedded in a continuous medium having the dielectric properties of water using the conductor-like polarizable continuum model (CPCM) method (50, 51). ii) A water molecule was added to the nucleoside and the complex was minimized in vacuo. In H-U (Fig. 1), the water molecule was initially positioned with one of its proton, H<sub>2WT</sub>, in contact with the O<sub>2</sub> atom of the uracil. In the case of OH-RIB-U and OCH<sub>3</sub>-RIB-U (Fig. 1), the water molecule was initially positioned to form a bridge between the O<sub>2</sub> of the uracil and the 2'-group.

-----  
 - Insert Figure 1 about here -  
 -----

In the nucleosides (Fig. 1) the dihedral angles are defined as follows:



The glycosidic angle  $\chi$ , is defined as: O<sub>4'</sub> - C<sub>1'</sub> - N<sub>9</sub> - C<sub>4</sub> for purines and O<sub>4'</sub> - C<sub>1'</sub> - N<sub>1</sub> - C<sub>2</sub> for pyrimidines. The five endocyclic dihedral angles were approximated by a Fourier analysis giving the phase or pseudorotation angle (P) and the amplitude ( $\tau$ ) according to Rao et al. (52):

$$\tau_j = \tau \cos[P + 4\pi(j - 2)/5] \quad (1)$$

From the measurements of the dihedrals  $\tau_j$  in a conformation, A and B are computed as:

$$A = 2/5 \sum_{j=0}^4 \tau_j \cos[4\pi(j - 2)/5] \quad (2)$$

$$B = -2/5 \sum_{j=0}^4 \tau_j \sin[4\pi(j - 2)/5] \quad (3)$$

The amplitude,  $\tau$ , and the pseudorotation, P, are given by :

$$\tau = (A^2 + B^2)^{1/2} \quad (4)$$

$$P = \tan^{-1} B / A \quad (5)$$

The pseudorotation space is divided into four equally sized quadrants centered around P=0.0°, P=90.0°, P=180.0° and P=-90.0° that are referred to as the north (N), east (E), south(S), and west (W) quadrants, respectively.

When locating the north and south energy minima, the initial furanose conformations were set in canonical N and S puckering, respectively. In these calculations the dihedral angles  $\beta$ ,  $\gamma$  and

$\chi$  were assigned to modal values observed in the crystal nucleic acid structure database (53) as follows:

$\beta = 175^\circ \pm 15$ ,  $\gamma = 57^\circ \pm 10$  and  $\chi = -161^\circ \pm 11$  (anti<sup>(-ap)</sup>) for N-type conformations;

$\beta = 168^\circ \pm 25$ ,  $\gamma = 51^\circ \pm 16$  and  $\chi = -108^\circ \pm 11$  (anti<sup>(-ac)</sup>) for S-type conformations.

The  $\varepsilon$ 's energy profile of a furanose substituted in 3' by a O3'-P group shows three minima with both S and N puckering (46). Two minima, in *-ap* and *-sc* are much more favorable than the third. Therefore, in addition to the puckering and the dihedrals  $\beta$ ,  $\gamma$  and  $\chi$ , the dihedral  $\varepsilon$  was set in *-ap*, *-sc* before optimization. In most cases, the dihedrals are unfrozen during the minimization. When the furanose of the minimized conformation is S-type, and  $\varepsilon$ , in *-ap* or *-sc* the conformation is called B<sub>I</sub> or B<sub>II</sub> form. Similarly, when the furanose is in the N conformation, the minima, with  $\varepsilon$  in *-ap*, *-sc* are called A and A<sub>II</sub> forms. For OH-RIB-U and OCH3-RIB-U compounds, the minimizations of some A and B<sub>I</sub> conformations, were carried out by freezing the  $\varepsilon$  dihedral angle at  $-155.0^\circ$  to avoid the formation of the H<sub>3T</sub>...O<sub>2</sub>' hydrogen bond which can not exist in polynucleotide chains.

For OH-RIB-U and OCH3-RIB-U, calculations were performed starting from two different starting models depending on the dihedral  $\lambda$  : (H<sub>2</sub>'-C<sub>2</sub>'-O<sub>2</sub>'-H<sub>O</sub>'<sub>2</sub>) for OH-RIB-U and (H<sub>2</sub>'-C<sub>2</sub>'-O<sub>2</sub>'-C<sub>X2</sub>) for OCH3-RIB-U. In model 1, the 2'-group points toward the O<sub>3</sub>' atom. In some cases,  $\lambda$ , had to be frozen during a preliminary minimization. In model 2, the 2'-group was set away from the O<sub>3</sub>' atom.

Potential energy differences were calculated to understand the relation of conformation to energetic in the studied compounds. The energy of the B<sub>I</sub> conformation is taken as reference ( $\Delta E=0$ ).

The east energy barrier,  $B_e$ , separating the A and B<sub>I</sub> conformations, is approximated by the energy of the O<sub>4</sub>' endo conformation of the furanose (endocyclic dihedral angle  $\tau_2$  (C<sub>1</sub>'-C<sub>2</sub>'-C<sub>3</sub>'-C<sub>4</sub>') =  $0.0^\circ$ ) reported to either A or B<sub>I</sub> whichever is the more stable, like in similar works (46). The approximation was validated for desoxyadenosine by drawing the relaxed potential map in the (P,  $\tau$ ) space (Barbe and Le Bret in preparation).

## RESULTS

### H-U

With each method of calculations, four unconstrained minima were obtained: two with a N puckering and two others with a S puckering (Table 1). In each conformation, the values of dihedral angles  $\beta$ ,  $\gamma$ ,  $\delta$  (not shown here),  $\varepsilon$  and  $\chi$  (shown in Table 1) are close to their modal values. The four minimized conformations can easily be assimilated to the A, A<sub>II</sub>, B<sub>I</sub> and B<sub>II</sub> forms.

In vacuo, the four minimized conformations appear in the order B<sub>I</sub> < A < B<sub>II</sub> < A<sub>II</sub> (Table 1). As expected for a deoxyribonucleoside, the most favorable energetic conformation is B<sub>I</sub> and the most unfavorable conformation is A<sub>II</sub>. Analysis of the nucleic acid database (53) shows that the ratio B<sub>I</sub>/B<sub>II</sub> of the populations of B<sub>I</sub> and B<sub>II</sub> conformations is around 2.7 (54). This ratio agrees remarkably well with the energy gap, 0.6 kcal/mol, shown in Table 1. The energy gap between A and B<sub>I</sub> conformations is two times smaller than that determined at the MP2/6-31G\* level of theory (45).

In implicit solvent, the four minimized conformations appear in the order  $B_{II} < B_I < A < A_{II}$  (see Table 1). This model predicts a much larger  $B_{II}$  population than is actually observed in the data banks.

In DNA, a hydration site is observed in the vicinity of the pyrimidine O2 (4, 55). We then performed calculations in vacuo but in the presence of an explicit water molecule in contact with the O<sub>2</sub> atom of the uracil. In this case, the four minimized conformations appear in the order  $B_I < A < B_{II} < A_{II}$  (see Table 1). The energy classification of the four conformations is identical to that obtained in vacuo. However, the energy gap between the  $B_I$  and A conformations is approximately two times greater than that found in vacuo. This model predicts a  $B_{II}$  population which remains in agreement with that observed in the data banks. The  $A_{II}$  conformation is always the most unfavorable conformation, as expected.

-----  
 - Insert Table 1 about here -  
 -----

Because of the absence of strong hydrogen bonds in the phosphodiester chain, weak hydrogen bonds, such as C-H...O and C-H...N, contribute to the conformational stability of nucleic acids.

Figure 2-a shows the hydrogen bonds stabilizing the  $B_I$  and A conformations in vacuo. The S conformations, are stabilized by the C<sub>2</sub>'-H<sub>2</sub>'...O<sub>5</sub>' hydrogen bond. The N conformations, are stabilized by a C<sub>3</sub>'-H<sub>3</sub>'...O<sub>5</sub>' hydrogen bond. As experimentally observed (56, 57) and well simulated in silico (58), the formation of the C<sub>6</sub>-H<sub>6</sub>'...O<sub>5</sub>' hydrogen bond stabilizes the anti pyrimidine bases in nucleosides and nucleotides in the S and N conformations although the glycosidic angle  $\chi$  differs by approximately 30 degrees (59). Both S and N conformations are further stabilized by the bonds C<sub>1</sub>'-H<sub>1</sub>'...O<sub>2</sub> and C<sub>6</sub>-H<sub>6</sub>'...O<sub>4</sub>'.

The S and N conformations obtained in vacuo but in the presence of an explicit water molecule in contact with O<sub>2</sub> are stabilized by the same hydrogen bonds as those observed in vacuo and in implicit solvent. However, supplementary hydrogen bonds between the nucleoside and the water molecule were observed. The interactions H<sub>2</sub>WT ... O<sub>2</sub> and O<sub>WT</sub> ... H<sub>1</sub>' stabilize both  $B_I$  and A conformations (Fig. 2-b). In the S conformations, H<sub>1</sub>WT interacts with O<sub>3</sub>', but remains free in the N conformations. Therefore, the orientation of the water molecule differs in the N and S conformations.

-----  
 - Insert Figure 2 about here -  
 -----

## **OH-RIB-U**

The conformational analysis of the ribonucleoside OH-RIB-U is much more intricate than that of deoxyribouracil, because of the simultaneous presence of the hydroxyl groups at 2' and 3'. When a ribose alone (58) or a ribose attached to a pyrrole cycle (48) are minimized without constraints, the S conformation was found more favorable than the N conformation. The minimizations without constraints of the ribonucleosides rA, rU and rG give two minima, N and S, of almost the same energy (60). The N conformations are stabilized by a O<sub>2</sub>'-H<sub>O</sub>'...O<sub>3</sub>'

hydrogen bond and the S conformations by a  $O_{3'}-H_{3T}\dots O_{2'}$  hydrogen bond (60). Clearly, an intramolecular hydrogen bond, where the 3' hydroxyl is the hydrogen donor and the 2' hydroxyl the acceptor, does not pertain to RNA structures, except at the 3' termini. A tempting solution is to study 3'-methyl phosphate-uridine instead of uridine. However, at the B3LYP/6-31G<sup>†</sup> level of theory (58), the S conformation of the latter compound is found more stable than its N conformation by 0.44 kcal/mol (58). To eliminate the hydrogen bond donor properties of the  $O_{3'}-H_{3T}$  group, and thus disfavor the S conformation, the dihedral  $\varepsilon$  was frozen to a value observed in RNA: the pucker of a ribose attached to an imidazole ring (47) or a pyrrole cycle (48) was found in N, as wanted. Therefore, here, OH-RIB-U is first minimized freely. If the  $O_{3'}-H_{3T}\dots O_{2'}$  hydrogen bond is formed, the minimization was restarted freezing the dihedral  $\varepsilon$  to  $-155^\circ$ , a value that belongs to both the A and B<sub>I</sub> forms.

Egli et al. (8) have shown that water molecules extensively hydrate the minor groove. Their high-resolution crystal structure of an RNA duplex shows that water molecules form clusters, positioned around the  $O_{4'}$  and the  $O_{2'}$  atoms. Water molecules that are localized near the hydroxyl group at 2' are of particular significance because, in addition to their interactions with  $O_{2'}$ , they form secondary contacts that order the water structure in the minor groove. One of these water molecules forms an intra-residue bridge between  $O_{2'}$  and pyrimidine  $O_2$ . Simulation of the molecular dynamics of tRNA<sup>Asp</sup> shows that  $H_{O'2}$  in S-type and N-type riboses points preferentially toward the  $O_{3'}$  oxygen of the same residue (10). When the ribose is N-type, it can also point to pyrimidine  $O_2$ . When a ribose attached to an imidazole group (47) is minimized at the MP2/6-31G\* level, the A conformation is found more favorable than the B<sub>I</sub> conformation ( $\Delta E = -2.3$  kcal/mol) if  $H_{O'2}$  points to the  $O_{3'}$  atom in both conformations (47). It was then tempting to test on a natural nucleoside if the more stable conformation could be predicted from the orientation of  $H_{O'2}$  by itself. Two models are studied here. In model 1,  $H_{O'2}$  is set away from  $O_{3'}$ . In model 2,  $H_{O'2}$  points to  $O_{3'}$ .

#### Model 1 ( $H_{O'2}$ set away from $O_{3'}$ )

Here, the  $H_{O'2}$  proton of OH-RIB-U was set away from the  $O_{3'}$  atom in the starting conformations. But for the  $\varepsilon$  dihedral angle that was frozen in its modal values in the A and B<sub>I</sub> conformations, no other dihedral angle was forced during the minimizations. As for the study of H-U, the minimizations were performed in vacuo, B3LYP/6-31G\*, in implicit solvent, B3LYP/CPCM/6-31G\* and in vacuo but in presence of an explicit water molecule.

In vacuo, four minima were obtained: two of them are N-type and the other two are S-type (Table 2). In the N conformations, the values of dihedral angles  $\beta$ ,  $\gamma$ ,  $\delta$  (not shown here),  $\varepsilon$  and  $\chi$  (shown in Table 2) lie in the conformational ranges and can be called A and A<sub>II</sub> forms. For the same reasons, the S conformations could be called B<sub>I</sub> and B<sub>II</sub> conformations, if the dihedral angle  $\chi$  was  $-ac$ . As  $\chi$  is in the  $-ap$  range the S-type minima are called "B<sub>I</sub>" and "B<sub>II</sub>". The four minimized conformations appear in the order "B<sub>I</sub>" < A < "B<sub>II</sub>" < A<sub>II</sub> (see Table 2). The "B<sub>I</sub>" conformation is very favorable, which does not agree with experiential data. In the "B<sub>I</sub>" and "B<sub>II</sub>" conformations, the anomalous glycosidic angles are favored because the  $H_{O'2}$  hydrogen of the 2' hydroxyl is strongly attracted by the  $O_2$  atom of the uracil (Fig. 3-a). This interaction is so strong that it destroys the  $C_6-H_6\dots O_5$  and  $C_1-H_1\dots O_2$  hydrogen bonds, that are normally observed in the S conformations. The N conformations of OH-RIB-U are stabilized by the same four hydrogen bonds as those normally observed in the N conformations of H-U obtained in vacuo, and in implicit solvent. Minimizations in the presence of implicit water with the CPCM method were performed to dampen the effect of the  $O_{2'}-H_{O'2}\dots O_2$  hydrogen bond. We then obtained four minima corresponding to the A, A<sub>II</sub>, B<sub>I</sub> and B<sub>II</sub> conformations. These four minimized conformations appears in the order A < B<sub>I</sub> < B<sub>II</sub> < A<sub>II</sub> (see Table 2). In both S conformations, the  $O_{2'}-H_{O'2}\dots O_2$  hydrogen bond is absent

(Fig. 3-b). The small energy gap between the A and B<sub>I</sub> conformations does not agree with the prevalence of N-type in double stranded RNA.

We then performed calculations in vacuo, at the B3LYP/6-31G\* level of theory but in presence of an explicit water molecule forming a bridge between the O<sub>2</sub> of the uracil and the 2'-OH group. The H<sub>2WT</sub> and O<sub>WT</sub> atoms of the water molecule form hydrogen bonds with O<sub>2</sub> and H<sub>O'2</sub>, respectively (Fig. 3-c). We then obtained four energy minima corresponding to the A, A<sub>II</sub>, B<sub>I</sub> and B<sub>II</sub> conformations. The four minimized conformations appear in the order: A < A<sub>II</sub> < B<sub>I</sub> < B<sub>II</sub> (see Table 2). The energy gap between the B<sub>I</sub> and A conformations is large enough to favor the A conformation with regard to the B<sub>I</sub> conformation. The A<sub>II</sub> conformation is here more stable than either S-type conformations.

-----  
 - Insert Table 2 about here -  
 -----

-----  
 - Insert Figure 3 about here -  
 -----

#### Model 2 (H<sub>O'2</sub> points to O<sub>3'</sub>)

In vacuo, at the B3LYP/6-31G\* level of theory, we obtained three minima that agree very well with the A, B<sub>I</sub> and B<sub>II</sub> helix forms. For each conformation, the values of dihedral angles  $\beta$ ,  $\gamma$ ,  $\delta$  (not shown here),  $\varepsilon$  and  $\chi$  (shown in Table 3) lie in their conformational ranges. The A<sub>II</sub> conformation was not found. The three minima appear in the order A < B<sub>I</sub> < B<sub>II</sub> (see Table 3). The most favorable conformation is A, as wanted. In N and S conformations, H<sub>O'2</sub> interacts with O<sub>3'</sub> atom (Fig. 4-a). The O<sub>2'</sub>-H<sub>O'2</sub>...O<sub>3'</sub> hydrogen bond is stronger in the A conformation in agreement with the work of Foloppe and Mackerell (47). However, the absolute value of the energy gap between the A and B<sub>I</sub> conformations estimated here for an uracil at the B3LYP/6-31G\* level of theory is much less than the corresponding value (2.3 kcal/mol) found for a ribose attached to an imidazole group at the MP2/6-31G\* level (47).

In implicit solvent, at the B3LYP/CPCM/6-31G\* level of theory, we obtained three minima that agree with the A, B<sub>I</sub> and B<sub>II</sub> helix forms (Table 3). For each conformation, the values of dihedral angles  $\beta$ ,  $\gamma$ ,  $\delta$  (not shown here),  $\varepsilon$  and  $\chi$  (shown in Table 3) lie in their conformational ranges. As in vacuo, the minimization of a starting conformation similar to A<sub>II</sub> gave the A conformation. The three minima appear in the order B<sub>I</sub> < A < B<sub>II</sub> (see Table 3). Because B<sub>I</sub> is more stable than A, these calculations do not agree with the experimental data concerning double stranded RNAs.

An explicit water molecule was then considered, its protons H<sub>1WT</sub> and H<sub>2WT</sub> interacting with O<sub>2'</sub> and O<sub>2</sub>, respectively. Three energy minima corresponding to the A, B<sub>I</sub> and B<sub>II</sub> conformations were obtained. As in vacuo and in implicit solvent, the A<sub>II</sub> conformation was not found. The three minimized conformations appear in the order: A < B<sub>II</sub> < B<sub>I</sub> (see Table 3) and the A conformation is by far, the most favorable conformation. In the three minimized conformations, H<sub>O'2</sub> and the water molecule remained after minimization in similar orientations that those initially chosen (Fig. 4-b). After minimizations, the H<sub>O'2</sub> still interacts with the O<sub>3'</sub> atom, in the N and S conformations. In the A, B<sub>II</sub> and B<sub>I</sub> conformations, the distances separating O<sub>3'</sub> from H<sub>O'2</sub> are equal to 2.05 Å, 2.06 Å and 2.22 Å, respectively. In the B<sub>I</sub> conformation, the O<sub>2'</sub>-H<sub>O'2</sub>...O<sub>3'</sub> hydrogen bond is not as strong as in the A and B<sub>II</sub>

conformations. After minimization, in the N and S conformations, the H<sub>1WT</sub> and H<sub>2WT</sub> hydrogen atoms of the water molecule always form hydrogen bonds with the O<sub>2'</sub> and O<sub>2</sub> atoms of OH-RIB-U, respectively. In the A, B<sub>II</sub> and B<sub>I</sub> conformations, the O<sub>2'</sub> to H<sub>1WT</sub> distance is equal to 2.10 Å, 2.02 Å and 2.03 Å; and the O<sub>2</sub> to H<sub>2WT</sub> distance is equal to 2.06 Å, 2.14 Å and 2.14 Å, respectively. In the A conformation, there is a supplementary hydrogen bond between the water molecule and the ribonucleoside. Indeed, the O<sub>WT</sub> oxygen of the water molecule forms a hydrogen bond with the H<sub>1'</sub> hydrogen of the sugar (Fig. 4-b). The other hydrogen bonds stabilizing the N and S conformations are similar to those normally stabilizing the N and S conformations, as those observed in the N and S conformations obtained in vacuo and in implicit solvent.

-----  
 - Insert Table 3 about here -  
 -----

-----  
 - Insert Figure 4 about here -  
 -----

To easily compare model 1 to model 2, the energy of the minimized conformations of model 2 are reported to the energy of the conformation B<sub>I</sub> of model 1. (Table 3, column  $\Delta E_1$ ). They can easily be compared to the values,  $\Delta E_1$ , of model 1 reported in Table 2. In vacuo, the most favorable conformation is the model 1 “B<sub>I</sub>” conformation and the next N-type conformation is the model 2 A conformation. The energy gap is only 0.16 kcal/mol, but in favor of the S-type form. In the presence of a continuous medium, the most favorable conformation is model 1 B<sub>I</sub> conformation, by a very narrow edge since the next N-type conformation is model 1 A conformation. The N-type form is clearly favored over S-type forms when a water molecule is set between O<sub>2'</sub> and O<sub>2</sub>. We have to conclude that water plays a crucial role. Once water is introduced explicitly, the N-type form is favored in both models 1 and 2.

### OCH3-RIB-U

When OCH3-RIB-U is alone in free solution, the terminal H<sub>3T</sub> hydrogen can be involved in a hydrogen bond with the adjacent 2' group. This hydrogen bond cannot exist when the nucleoside analogue is inserted in a nucleic acid chains because of the replacement of the 3'-OH group by a phosphate moiety. In all B<sub>I</sub> conformations and some A conformations the O<sub>3'</sub>-H<sub>3T</sub>... O<sub>2'</sub> bond was detected, and the minimization was restarted with the dihedral angle  $\epsilon$  frozen to -155.0°. As for OH-RIB-U, we must take care of the orientation of the methyl group either far from O<sub>3'</sub> (model 1) or pointing to O<sub>3'</sub> (model 2) in the starting conformations. As in the case of OH-RIB-U, a preliminary minimization had sometimes in model 1 to be performed freezing the dihedral angle  $\lambda$ . Then,  $\lambda$  was unfrozen and a second minimization followed. The orientation of the methyl group remained as wanted in the final minimized conformation.

Model 1: Methyl group away from O<sub>3'</sub>.

Molecular dynamics simulations (20, 21) as well as crystallographic (61, 62) and RMN (63) data, show that the 2'-OCH3 group is preferentially oriented away from O<sub>3'</sub> in 2'-OCH3

(CpG) duplexes. Moreover, molecular dynamics simulation of DNA/RNA hybrid sequence d(CCAACGTTGG).r(CCAACGUUGG) with modifications at the C2' positions of the DNA strand by 2'-OCH<sub>3</sub> shows that the 2'-OCH<sub>3</sub> is also oriented on the opposite site of the O<sub>3'</sub> atom (17). The same orientation was also found in the crystal structure of a 2'-OCH<sub>3</sub> adenosine incorporated in a DNA duplex (64). In all these oligonucleotides, the 2'-OCH<sub>3</sub> nucleotides are in a N conformation.

The results concerning model 1 are shown in Table 4. Whatever the condition of hydration, four minima were obtained: two with a N puckering and two others with a S puckering. As the values of dihedral angles  $\beta$ ,  $\gamma$ ,  $\delta$  (not shown here),  $\epsilon$  and  $\chi$  (shown in table 6) lie in their modal ranges, the minimized conformations agree very well with the forms: A, A<sub>II</sub>, B<sub>I</sub> and B<sub>II</sub>.

In vacuo, the four minimized conformations appear in the order A < B<sub>I</sub> < A<sub>II</sub> < B<sub>II</sub> (Table 4). The A conformation is the most energetically favorable conformation. The A conformation is more stable than next B<sub>I</sub> conformation by 0.17 kcal/mol only. The A<sub>II</sub> conformation is less favorable than the B<sub>I</sub> conformation and the B<sub>II</sub> conformation is much discriminated. In the final optimized conformations the 2'-OCH<sub>3</sub> remained away from the O<sub>3'</sub> atom (Fig. 5-a). Figure 5-a shows that the hydrogen bonds stabilizing the B<sub>I</sub> and the A conformations of OCH<sub>3</sub>-RIB-U are the four canonical hydrogen bonds that respectively stabilize a S and a N conformation.

In implicit solvent, the minimized conformations appear in the order B<sub>I</sub> < A < A<sub>II</sub> < B<sub>II</sub> (Table 4). The B<sub>I</sub> conformation is more stable than the A conformation by a small difference. Moreover, the energy differences with the other conformations are also relatively weak. Although, the B<sub>II</sub> conformation is the most unfavorable conformation, it is much less discriminated than in vacuo. Therefore OCH<sub>3</sub>-RIB-U is very flexible, in implicit solvent. The hydrogen bonds stabilizing the S and N conformations of OCH<sub>3</sub>-RIB-U in implicit solvent are similar to those stabilizing the S and N conformations of OCH<sub>3</sub>-RIB-U in vacuo, respectively.

In vacuo but in presence of an explicit water molecule forming a bridge between the O<sub>2</sub> of the uracil and the 2'-group, the minimized conformations appear in the order A < B<sub>I</sub> < A<sub>II</sub> < B<sub>II</sub> (Table 4) which is identical to that the ranking obtained in vacuo without the explicit water molecule. In presence of the water molecule, the A conformation is more stable than the B<sub>I</sub> conformation by 0.44 kcal/mol while in absence of the water molecule, the A conformation is more stable than the B<sub>I</sub> conformation by only 0.17 kcal/mol. However, the energy gap between the A and B<sub>I</sub> conformations is much smaller with OCH<sub>3</sub>-RIB-U than with OH-RIB-U (Table 2). Therefore, in vacuo but in presence of an explicit water molecule, the prevalence of the A conformation is more significant for OH-RIB-U than for OCH<sub>3</sub>-RIB-U. Like in absence of the water molecule, the B<sub>II</sub> conformation of OCH<sub>3</sub>-RIB-U is much discriminated. Therefore, in vacuo, in the presence or in the absence of the water molecule, OCH<sub>3</sub>-RIB-U is preferentially in A conformation but the populations in B<sub>I</sub> and A<sub>II</sub> should be detectable.

In the S and N conformations, the water molecule forms a bridge between O<sub>2'</sub> and O<sub>2</sub> (Fig 5-b). The atoms of the water molecule, H<sub>2WT</sub> and O<sub>WT</sub> are involved in hydrogen bonds with O<sub>2</sub> of the uracil and H<sub>X<sub>2</sub>B</sub> of the 2'-OCH<sub>3</sub> group, respectively. In the A, B<sub>I</sub>, A<sub>II</sub> and B<sub>II</sub> conformations, the O<sub>WT</sub> to H<sub>X<sub>2</sub>B</sub> distance is equal to 2.48 Å, 2.44 Å, 2.46 Å and 2.46 Å, respectively, and the O<sub>2</sub> to H<sub>2WT</sub> distance is equal to 1.94 Å, in the four conformations. In the N and S conformations, the orientation of the methyl group is similar to that observed in the N and S conformations obtained in implicit solvent and in vacuo without the explicit water molecule. The others hydrogen bonds stabilizing the N and S conformations are similar as those normally stabilizing the N and S conformations, as those observed in the N and S conformations obtained in implicit solvent and in vacuo without the explicit water molecule.

To summarize, in the model 1 orientation, the A conformation is the most stable conformation, in vacuo, in presence and in absence of a water molecule forming a bridge between the 2'-OCH3 and the O<sub>2</sub> of the uracil. The energy difference between the N and S is small.

-----  
 - Insert Table 4 about here -  
 -----

-----  
 - Insert Figure 5 about here -  
 -----

Model 2 (Methyl group points to O<sub>3'</sub>).

When the methyl group of the substituent at 2' is oriented towards O<sub>3'</sub> in the starting conformation, the minimizations gave the results show in Table 5. With each condition of hydration, four minima were obtained: two with a N puckering and two others with a S puckering. These minimized conformations agree very well with the different forms: A, A<sub>II</sub>, B<sub>I</sub> and B<sub>II</sub>. Indeed, for each conformation, the values of dihedral angles  $\beta$ ,  $\gamma$ ,  $\delta$  (not shown here),  $\varepsilon$  and  $\chi$  (shown in Table 5) are positioned in their conformational ranges observed in each corresponding form.

In vacuo, the four minimized conformations appear in the order  $A < A_{II} < B_I < B_{II}$  (Table 5). The A conformation is the most energetic favorable conformation. The two N conformations, A et A<sub>II</sub> are more stable than the two S conformations, B<sub>I</sub> and B<sub>II</sub>. Therefore, in vacuo, OCH3-RIB-U is N-type, preferentially in A and to a lesser degree in A<sub>II</sub>. The results agree with calculations concerning the 2'-O-methyl-2'-deoxyguanosine at the HF/6-31G\* and MP2/6-31G\*//HF/6-31G\* levels of theory that also indicate the preference for N puckering (17). The energy difference between the A and B<sub>I</sub> conformations is greater with OCH3-RIB-U than with OH-RIB-U (Table 3). In vacuo, the two S conformations are stabilized by the four same hydrogen bonds as those observed in the S conformations of H-U, obtained in vacuo and in implicit solvent. Fig. 6-a shows these four hydrogen bonds in the B<sub>I</sub> conformation of OCH3-RIB-U. In the S conformations, the 2'-OCH3 group remained oriented toward O<sub>3'</sub> but without forming a hydrogen bond with O<sub>3'</sub> atom. In contrast, in the two N conformations of OCH3-RIB-U, a weak C<sub>X2</sub>-H<sub>X2B</sub>...O<sub>3'</sub> hydrogen bond was maintained (Fig. 6-a). In the A and A<sub>II</sub> conformations, the O<sub>3'</sub> to H<sub>X2B</sub> distance is equal to 2.54 Å and 2.48 Å, respectively. The C<sub>X2</sub>-H<sub>X2B</sub>...O<sub>3'</sub> hydrogen bonds are much weaker than the O<sub>2'</sub>-H<sub>O'2</sub>...O<sub>3'</sub> hydrogen bonds stabilizing the N conformations of OH-RIB-U. The four other hydrogen bonds stabilizing the two N conformations are identical to those normally observed in the N conformations, as in the N conformations of H-U and OH-RIB-U obtained in vacuo and in implicit solvent.

In implicit solvent, the obtained minimized conformations appear in the order  $B_{II} < A < B_I < A_{II}$  (Table 5). The B<sub>II</sub> conformation is the most energetic favorable and the A<sub>II</sub> conformation is the less stable. The energy differences between the different conformations are relatively small. In implicit solvent, OCH3-RIB-U is very flexible since the energy difference between the A and B<sub>I</sub> conformations is only of 0.16 kcal/mol. The hydrogen bonds stabilizing the S and N conformations of OCH3-RIB-U in implicit solvent and in vacuo are similar. In the N conformations, the C<sub>X2</sub>-H<sub>X2B</sub>...O<sub>3'</sub> hydrogen bonds are always present and the O<sub>3'</sub> to H<sub>X2B</sub>



distances are equal to 2.49 Å and 2.45 Å in the A and A<sub>II</sub> conformations, respectively. These distances are a little shorter than those observed in the N conformations obtained in vacuo.

In vacuo but in presence of an explicit water molecule forming a bridge between the O<sub>2</sub> of the uracil and the 2'-group, the minimized conformations appear in the order A < A<sub>II</sub> < B<sub>II</sub> < B<sub>I</sub> (Table 5) which is similar to the ranking obtained in vacuo without the explicit water molecule. The two N conformations are more favorable than the two S conformations. Therefore, in vacuo, in the presence or in the absence of the water molecule, OCH<sub>3</sub>-RIB-U is in N conformation, preferentially in A but can also be in A<sub>II</sub> without that there is a lot of difference of energy. The energy gap between the A and B<sub>II</sub> conformations is weaker with OCH<sub>3</sub>-RIB-U than with OH-RIB-U (Table 3). Therefore, in vacuo but in presence of an explicit water molecule, the preference for the A conformation is more significant for OH-RIB-U than for OCH<sub>3</sub>-RIB-U, contrary to the results obtained in vacuo but in absence of a water molecule. In the S and N conformations, the water molecule forms a bridge between O<sub>2</sub>' and O<sub>2</sub> (Fig. 6-b). The hydrogen atoms of the water molecule, H<sub>1WT</sub> and H<sub>2WT</sub> are involved in hydrogen bonds with the O<sub>2</sub>' of the 2'-OCH<sub>3</sub> group and the O<sub>2</sub> of the uracil, respectively. In the A, A<sub>II</sub>, B<sub>II</sub> and B<sub>I</sub> conformations, the O<sub>2</sub>' to H<sub>1WT</sub> distance is equal to 2.09 Å, 2.09 Å, 2.20 Å and 2.14 Å, respectively, and the O<sub>2</sub> to H<sub>2WT</sub> distance is equal to 2.07 Å, 2.07 Å, 2.05 Å and 2.03 Å, respectively. In the A conformation, there is a supplementary hydrogen bond between O<sub>WT</sub> and H<sub>1</sub>' (Fig. 6-b), as in the N conformations of OH-RIB-U (Fig 4-b). In the N and S conformations, the orientation of 2'-OCH<sub>3</sub> is similar to that observed in the N and S conformations obtained in vacuo and in implicit solvent. In the N conformations, the C<sub>X2</sub>-H<sub>X2B</sub>...O<sub>3</sub>' hydrogen bonds are always present and the O<sub>3</sub>' to H<sub>X2B</sub> distances are equal to 2.44 Å and 2.39 Å in the A and A<sub>II</sub> conformations, respectively. The other hydrogen bonds stabilizing the N and S conformations are similar as those normally stabilizing the N and S conformations, as those observed in the N and S conformations obtained in implicit solvent and in vacuo without the explicit water molecule.

To summarize, in the model 2 orientation the A conformation is the most stable conformation, in vacuo, in presence and in absence of a water molecule forming a bridge between the 2'-OCH<sub>3</sub> and the O<sub>2</sub> of the uracil. In implicit solvent, the B<sub>II</sub> conformation is the most stable conformation but the next conformation, A, is very close.

-----  
 - Insert Table 5 about here -  
 -----

-----  
 - Insert Figure 6 about here -  
 -----

To easily compare the two models, the energies of the various model 2 minima are reported to the energy of the model 1 B<sub>I</sub> conformation. (Table 5, col ΔE<sub>1</sub>). The most favorable N conformation is the A conformation whatever the treatment of hydration. In contrast, the most favorable S conformation is the B<sub>I</sub> or B<sub>II</sub> conformation according to the hydration conditions. In vacuo, the most stable N and S conformations are the model 1 A and B<sub>I</sub> conformations. The A conformation is more stable than the B<sub>I</sub> conformation by 0.17 kcal/mol (Table 4). In implicit solvent, the most stable N and S conformations are the model 2 A and B<sub>II</sub> conformations. Here, again the energy difference is very small, B<sub>II</sub> being favored by 0.36 kcal/mol. In vacuo but in presence of an explicit water molecule bridging the O<sub>2</sub>' and O<sub>2</sub> of

the uracil, the most stable N and S conformations are the model 2 A and B<sub>II</sub> conformations. The A conformation is more stable than the B<sub>II</sub> conformation by 1.25 kcal/mol. In the three treatments of hydration, the energy gaps are relatively small so that OCH<sub>3</sub>-RIB-U is always found more flexible than OH-RIB-U. The water molecule is not required to find A as the most stable conformation.

### East energy barriers

We calculated the east energy barriers, Be, in H-U, OH-RIB-U (models 1 and 2) and OCH<sub>3</sub>-RIB-U (models 1 and 2) (see Table 6). For OH-RIB-U and OCH<sub>3</sub>-RIB-U, in the O4' endo conformations, the  $\epsilon$  dihedral angle was frozen at -155.0° to avoid the formation of the O<sub>3</sub>'-H<sub>3</sub>T...O<sub>2</sub>' hydrogen bond. The 2' group was oriented as in the models 1 or 2 by freezing  $\lambda$  in a preliminary minimization. The resulting conformation was optimized during a second minimization with the dihedral  $\lambda$  unfrozen.

The Be values are different according to the hydration conditions (Table 6).

#### H-U

The Be determined at the B3LYP/6-31G\* level of theory (Table 6) is approximately twice weaker than calculated at the MP2/6-31G\* level of theory (45). To contrast, this Be value is concordance with those in others deoxynucleosides estimated at the same level of theory (46, 65). Therefore, the Be are significantly sensitive to the level of theory used.

The O4' endo conformation of H-U is associated with a significant flattening of the furanose ring, as shown in others works (45) (Barbe and Le Bret, in preparation).

#### OH-RIB-U

The Be are higher in the model 1 than in the model 2 (Table 6).

##### Model 1.

In vacuo, in implicit solvent and in presence of the explicit water molecule, the Be in OH-RIB-U are higher than those in H-U of 0.42 kcal/mol, 0.86 kcal/mol and 1.60 kcal/mol, respectively. The empirical potential used by Olson (33) predicted that the east energy barrier would be approximately 2.0 kcal/mol higher in ribose than in deoxyribose. The value in presence of the explicit water molecule (1.60 kcal/mol) is the closest to Olson's prediction (33). Only in presence of the explicit water molecule, the amplitude  $\tau$  at the east barrier is significantly lower than in the N and S energy minima, as observed in H-U.

##### Model 2.

The Be determined at the B3LYP/6-31G\* level of theory (Table 6) is twice weaker than the Be calculated at the MP2/6-31G\* level for a ribose bound to an imidazole group (47). This difference may result of the replacement of the base by an imidazole group or of the level of theory used, as shown for H-U. In vacuo, in implicit solvent and in presence of the explicit water molecule, the Be in OH-RIB-U are higher than those in H-U of 0.12 kcal/mol, 0.14 kcal/mol and 0.56 kcal/mol, respectively. These values are weaker than the difference between the Be in ribose and deoxyribose predicted by Olson (33). To contrast, the value determined in presence of the explicit water molecule (0.56 kcal/mol) is similar to the difference between the Be in a ribose and a deoxyribose bounded to an imidazole group (47). The furanose amplitude  $\tau$  at the east barriers is lower than in the N and S energy minima. However, this flattening of the furanose ring is less significant than in H-U.

### OCH3-RIB-U

The Be are higher in the model 2 than in the model 1 (Table 6).

Model 1.

In vacuo, in absence and in presence of the explicit water molecule, the Be are higher than those in H-U by 0.24 kcal/mol and 1.66 kcal/mol, respectively. The Be are weaker than those in the model 1 of OH-RIB-U except in presence of the water molecule. To contrast, nearly all Be are higher than those in model 2 of OH-RIB-U, especially in presence of the water molecule. The furanose amplitude  $\tau$  at the east barriers is lower than in the N and S energy minima.

Model 2 :

In vacuo, in implicit solvent and in presence of the explicit water molecule, the Be are higher than those in H-U (by 1.56 kcal/mol, 0.35 kcal/mol and 1.87 kcal/mol) and in the model 2 of OH-RIB-U (by 1.44 kcal/mol, 0.21 kcal/mol and 1.31 kcal/mol). Nearly all Be are also higher than those in the model 1 of OH-RIB-U. Globally, the Be are significantly higher than in ribonucleoside and deoxynucleoside. The furanose ring, in the O4' endo conformations is not or slightly flattened.

## CONCLUSION

Ab initio calculations are useful to pinpoint which interaction is responsible for the conformations nucleosides have when inserted into a duplex. The conformation is not the absolute but a secondary minimum that depends on the treatment of hydration. The treatments in vacuo, with or without an explicit water molecule give as a rule more similar results than when the solvent is treated implicitly, where B<sub>II</sub> forms seemed to be favored. In vacuo, the conformations of deoxyuridine, H-U, found here are very similar to the conformations of deoxyadenosine (H-A) that we described elsewhere. The nucleoside is found very flexible. The hydroxyl group at 2' by itself does not make it rigid as it is found experimentally. Uridine is found rigid only in the presence of an explicit water molecule. As a rule, the nucleoside with a methoxyl group at 2' keeps its flexibility. It has many conformations of almost the same energy. Hopefully, this work is an incentive for more experimental determinations.

## REFERENCES

1. Saenger W. 1984. Principles of Nucleic Acid Structure. Springer-Verlag, New York.
2. Shindo, H., T. Fujiwara, H. Akutsu, U. Matsumoto, and Y. Kyogoku. 1985. Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 1. Static geometry of DNA double helices. *Biochemistry* 24:887-895.
3. Cheatham, T. E., and Kollman P.A. 1997. Molecular dynamics simulations highlight the structural differences among DNA:RNA, RNA:RNA, and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc* 119:4805-4825.
4. Auffinger, P., and E. Westhof. 2000. Water and ion binding around RNA and DNA (C,G) oligomers. *J. Mol. Biol.* 300:1113-1131.
5. Cech, T. R. 1990. Self-splicing of group I introns. *Annu. Rev. Biochem.* 59:543-568.
6. Fersht, A. R. 2005. Enzyme Structure and Mechanism. W. H. Freeman & Co, San Francisco.
7. Rouse, B., N. Puri, G. Viswanadham, P. Agback, C. Glemarec, A. Sandstrom, C. Sund, and J. Chattopadhyaya. 1994. Solution Conformation of Hexameric and Heptameric Lariat-Rnas and Their Self-Cleavage Reactions Which Give Products Mimicking Those from Some Catalytic Rnas (Ribozymes). *Tetrahedron* 50:1777-1810.
8. Egli, M., S. Portmann, and N. Usman. 1996. RNA hydration: a detailed look. *Biochemistry* 35:8489-8494.
9. Acharya, P., and J. Chattopadhyaya. 2002. The hydrogen bonding and hydration of 2' OH in adenosine and adenosine 3'-ethyl phosphate. *Journal of Organic Chemistry* 67:1852-1865.
10. Auffinger, P., and E. Westhof. 1997. Rules governing the orientation of the 2'-hydroxyl group in RNA. *J. Mol. Biol.* 274:54-63.
11. Cramer, H., and W. Pfeleiderer. 2000. Nucleotides LXIV[1]: synthesis hybridization and enzymatic degradation studies of 2'-O-methyloligoribonucleotides and 2'-O-methyl/deoxy gapmers. *Nucleosides Nucleotides Nucleic Acids* 19:1765-1777.
12. Freier, S. M., and K. H. Altmann. 1997. The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res.* 25:4429-4443.
13. Lesnik, E. A., and S. M. Freier. 1998. What affects the effect of 2'-alkoxy modifications? 1. Stabilization effect of 2'-methoxy substitutions in uniformly modified DNA oligonucleotides. *Biochemistry* 37:6991-6997.
14. Majlessi, M., N. C. Nelson, and M. M. Becker. 1998. Advantages of 2'-O-methyl oligoribonucleotide probes for detecting RNA targets. *Nucleic Acids Res.* 26:2224-2229.
15. Sproat, B. S., A. I. Lamond, B. Beijer, P. Neuner, and U. Ryder. 1989. Highly efficient chemical synthesis of 2'-O-methyloligoribonucleotides and tetrabiotinylated derivatives; novel probes that are resistant to degradation by RNA or DNA specific nucleases. *Nucleic Acids Res.* 17:3373-3386.
16. Cummins, L. L., S. R. Owens, L. M. Risen, E. A. Lesnik, S. M. Freier, D. McGee, C. J. Guinasso, and P. D. Cook. 1995. Characterization of fully 2'-modified oligoribonucleotide hetero- and homoduplex hybridization and nuclease sensitivity. *Nucleic Acids Res.* 23:2019-2024.

17. Venkateswarlu, D., K. E. Lind, V. Mohan, M. Manoharan, and D. M. Ferguson. 1999. Structural properties of DNA:RNA duplexes containing 2'-O-methyl and 2'-S-methyl substitutions: a molecular dynamics investigation. *Nucleic Acids Res.* 27:2189-2195.
18. Yathindra, N., and M. Sundaralingam. 1979. Effect of O(2')-methylation on the stability of polynucleotide helices. *Biopolymers* 18:2721-2731.
19. Kawai, G., Y. Yamamoto, T. Kamimura, T. Masegi, M. Sekine, T. Hata, T. Iimori, T. Watanabe, T. Miyazawa, and S. Yokoyama. 1992. Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* 31:1040-1046.
20. Kulinska, K., T. Kulinski, A. Lyubartsev, A. Laaksonen, and R. W. Adamiak. 2000. Spatial distribution functions as a tool in the analysis of ribonucleic acids hydration - molecular dynamics studies. *Computers & Chemistry* 24:451-457.
21. Auffinger, P., and E. Westhof. 2001. Hydrophobic groups stabilize the hydration shell of 2'-O-methylated RNA duplexes. *Angewandte Chemie-International Edition* 40:4648-+.
22. Rozners, E., and J. Moulder. 2004. Hydration of short DNA, RNA and 2'-OMe oligonucleotides determined by osmotic stressing. *Nucleic Acids Res.* 32:248-254.
23. Westhof, E. 1988. Water - An Integral-Part of Nucleic-Acid Structure. *Annual Review of Biophysics and Biophysical Chemistry* 17:125-144.
24. Saenger, W. 1987. Structure and dynamics of water surrounding biomolecules. *Annu. Rev. Biophys. Biophys. Chem.* 16:93-114.
25. Buckin, V. A. 1987. Experimental studies of DNA-water interaction. *Mol. Biol.* 21:512-525.
26. Otwinowski, Z., R. W. Schevitz, R. G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, and P. B. Sigler. 1988. Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* 335:321-329.
27. Shakked, Z., G. Guzikovich-Guerstein, F. Frolow, D. Rabinovich, A. Joachimiak, and P. B. Sigler. 1994. Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature* 368:469-473.
28. Qian, Y. Q., G. Otting, and K. Wuthrich. 1993. Nmr1 Detection of Hydration Water in the Intermolecular Interface of A Protein-Dna Complex. *Journal of the American Chemical Society* 115:1189-1190.
29. Robinson, C. R., and S. G. Sligar. 1993. Molecular Recognition Mediated by Bound Water - A Mechanism for Star Activity of the Restriction-Endonuclease EcorI. *Journal of Molecular Biology* 234:302-306.
30. Chalikian, T. V., G. E. Plum, A. P. Sarvazyan, and K. J. Breslauer. 1994. Influence of Drug-Binding on Dna Hydration - Acoustic and Densimetric Characterizations of Netropsin Binding to the Poly(Dadt)Center-Dot-Poly(Dadt) and Poly(Da)Center-Dot-Poly(Dt) Duplexes and the Poly(Dt)Center-Dot-Poly(Da)Center-Dot-Poly(Dt) Triplex at 25-Degrees-C. *Biochemistry* 33:8629-8640.
31. Olmsted, M. C. 1996. The effect of nucleic acid geometry on counterion association. *Journal of Biomolecular Structure & Dynamics* 13:885-902.
32. Misra, V. K., K. A. Sharp, R. A. Friedman, and B. Honig. 1994. Salt Effects on Ligand-Dna Binding - Minor-Groove Binding Antibiotics. *Journal of Molecular Biology* 238:245-263.
33. Olson W.K. 1982. How Flexible Is the Furanose Ring? 2. An Updated Potential Energy Estimate. *J. Am. Chem. Soc* 104:278-286.

34. Plavec, J., W. M. Tong, and J. Chattopadhyaya. 1993. How do the Gauche and Anomeric Effects Drive the Pseudorotational Equilibrium of the Pentofuranose Moiety of Nucleosides. *Journal of the American Chemical Society* 115:9734-9746.
35. Thibaudeau, C., J. Plavec, and J. Chattopadhyaya. 1994. Quantitation of the Anomeric Effect in Adenosine and Guanosine by Comparison of the Thermodynamics of the Pseudorotational Equilibrium of the Pentofuranose Moiety, in N-Nucleosides and C-Nucleosides. *Journal of the American Chemical Society* 116:8033-8037.
36. Phillips, L., and V. Wray. 1973. Conformational and Chemical Consequences of Vicinal Polar Bonds. *Journal of the Chemical Society-Chemical Communications* 90-91.
37. Murcko, M. A., and R. A. Dipaola. 1992. Abinitio Molecular-Orbital Conformational-Analysis of Prototypical Organic-Systems .1. Ethylene-Glycol and 1,2-Dimethoxyethane. *Journal of the American Chemical Society* 114:10010-10018.
38. Wiberg, K. B., M. A. Murcko, K. E. Laidig, and P. J. Macdougall. 1990. Origin of the Gauche Effect in Substituted Ethanes and Ethenes. *Journal of Physical Chemistry* 94:6956-6959.
39. Oleary, D. J., and Y. Kishi. 1994. Preferred Conformation of C-Glycosides .13. A Comparison of the Conformational Behavior of Several C-Furanosides, N-Furanosides, and O-Furanosides. *Journal of Organic Chemistry* 59:6629-6636.
40. Plavec, J., C. Thibaudeau, and J. Chattopadhyaya. 1996. How do the energetics of the stereoelectronic gauche and anomeric effects modulate the conformation of nucleos(t)ides?. *Pure and Applied Chemistry* 68:2137-2144.
41. Kneisler, J. R., and N. L. Allinger. 1996. Ab initio and density functional theory study of structures and energies for dimethoxymethane as a model for the anomeric effect. *Journal of Computational Chemistry* 17:757-766.
42. Jaffe, R. L., G. D. Smith, and D. Y. Yoon. 1993. Conformations of 1,2-Dimethoxyethane from Ab-Initio Electronic-Structure Calculations. *Journal of Physical Chemistry* 97:12745-12751.
43. Uesugi S., Miki H., Ikehara M., Iwahashi H., and Kyogoku Y. 1979. A linear relationship between electronegativity of 2'-substituents and conformation of adenine nucleosides. *Tetrahedron Letters* 42:4073-4076.
44. Guschlbauer, W., and K. Jankowski. 1980. Nucleoside conformation is determined by the electronegativity of the sugar substituent. *Nucleic Acids Res.* 8:1421-1433.
45. Foloppe, N., and A. D. MacKerell, Jr. 1999. Intrinsic conformational properties of deoxyribonucleosides: implicated role for cytosine in the equilibrium among the A, B, and Z forms of DNA. *Biophys. J.* 76:3206-3218.
46. Foloppe, N., L. Nilsson, and A. D. MacKerell, Jr. 2001. Ab initio conformational analysis of nucleic acid components: intrinsic energetic contributions to nucleic acid structure and dynamics. *Biopolymers* 61:61-76.
47. Foloppe, N., and A. D. MacKerell. 1998. Conformational Properties of the Deoxyribose and Ribose Moieties of Nucleic Acids: A Quantum Mechanical Study. *J. Phys. Chem. B.* 102:6669-6678.
48. Brameld, K. A., and W. A. Goddard. 1999. Ab initio quantum mechanical study of the structures and energies for the pseudorotation of 5'-dehydroxy analogues of 2'-deoxyribose and ribose sugars. *Journal of the American Chemical Society* 121:985-993.
49. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, Jr. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J.

- Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. 2003. Gaussian 03, Revision A.1. Gaussian, Inc., Pittsburgh PA, 2003.
50. Barone V., Cossi M., and Tomasi J. 1997. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem Phys* 107:3210-3221.
  51. Barone V., and Cossi M. 1998. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys Chem A*. 102:1995-2001.
  52. Rao, S., E. Westhof, and M. Sundralingam. 1981. Exact method for the calculation of pseudorotation parameters  $P$ ,  $m$  and their errors. A comparison of the Altona-Sundralingam and Cremer-Pople treatment of puckering of five-membered rings. *Acta Crystallogr. A* 37:421-425.
  53. Berman, H. M., W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S. H. Hsieh, A. R. Srinivasan, and B. Schneider. 1992. The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* 63:751-759.
  54. Winger, R. H., K. R. Liedl, S. Rudisser, A. Pichler, A. Hallbrucker, and E. Mayer. 1998. B-DNA's B-I  $\rightarrow$  B-II conformer substate dynamics is coupled with water migration. *Journal of Physical Chemistry B* 102:8934-8940.
  55. Schneider, B., K. Patel, and H. M. Berman. 1998. Hydration of the phosphate group in double-helical DNA. *Biophys. J.* 75:2422-2434.
  56. Desiraju, G. R. 1996. The C-H center dot center dot center dot O hydrogen bond: Structural implications and supramolecular design. *Accounts of Chemical Research* 29:441-449.
  57. Wahl, M. C., and M. Sundralingam. 1997. C-H center dot center dot center dot O hydrogen bonding in biology. *Trends in Biochemical Sciences* 22:97-102.
  58. Leulliot, N., M. Ghomi, G. Scalmani, and G. Berthier. 1999. Ground state properties of the nucleic acid constituents studied by density functional calculations. I. Conformational features of ribose, dimethyl phosphate, uridine, cytidine, 5'-methyl phosphate-uridine, and 3'-methyl phosphate-uridine. *Journal of Physical Chemistry A* 103:8716-8724.
  59. Lu, X. J., Z. Shakked, and W. K. Olson. 2000. A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* 300:819-840.
  60. Hocquet, A., N. Leulliot, and M. Ghomi. 2000. Ground-state properties of nucleic acid constituents studied by density functional calculations. 3. Role of sugar puckering and base orientation on the energetics and geometry of 2'-deoxyribonucleosides and ribonucleosides. *Journal of Physical Chemistry B* 104:4560-4568.
  61. Adamiak, D. A., J. Milecki, M. Popenda, R. W. Adamiak, Z. Dauter, and W. R. Rypniewski. 1997. Crystal structure of 2'-O-Me(CGCGCG)<sub>2</sub>, an RNA duplex at 1.30 Å resolution. Hydration pattern of 2'-O-methylated RNA. *Nucleic Acids Res.* 25:4599-4607.

62. Adamiak, D. A., W. R. Rypniewski, J. Milecki, and R. W. Adamiak. 2001. The 1.19 Å X-ray structure of 2'-O-Me(CGCGCG)<sub>2</sub> duplex shows dehydrated RNA with 2-methyl-2,4-pentanediol in the minor groove. *Nucleic Acids Res.* 29:4144-4153.
63. Popenda, M., E. Biala, J. Milecki, and R. W. Adamiak. 1997. Solution structure of RNA duplexes containing alternating CG base pairs: NMR study of r(CGCGCG)<sub>2</sub> and 2'-O-Me(CGCGCG)<sub>2</sub> under low salt conditions. *Nucleic Acids Res.* 25:4589-4598.
64. Lubini, P., W. Zurcher, and M. Egli. 1994. Stabilizing effects of the RNA 2'-substituent: crystal structure of an oligodeoxynucleotide duplex containing 2'-O-methylated adenosines. *Chem. Biol.* 1:39-45.
65. Arora, K., and T. Schlick. 2003. Deoxyadenosine sugar puckering pathway simulated by the stochastic difference equation algorithm. *Chemical Physics Letters* 378:1-8.



**TABLE 1-** Conformational angles (degrees) and relative energies values,  $\Delta E$  (kcal/mol), of H-U optimized using the methods of calculation shown in column 1.

Methods of calculation	P	$\tau$	$\varepsilon$	$\chi$	$\Delta E$	Forms
B3LYP/6-31G*	14.30	35.8	-89.32	-158.61	0.98	A <sub>II</sub>
	12.96	34.7	-169.54	-159.72	0.43	A
	167.75	34.8	-65.04	-126.33	0.60	B <sub>II</sub>
	165.08	34.8	174.31	-130.14	0.00	B <sub>I</sub>
B3LYP/CPCM/6-31G*	11.61	34.4	-89.04	-157.82	1.13	A <sub>II</sub>
	11.80	34.9	-167.99	-156.50	1.12	A
	173.24	35.0	-66.06	-126.38	-0.03	B <sub>II</sub>
	172.90	35.1	177.42	-127.47	0.00	B <sub>I</sub>
B3LYP/6-31G* + WT	17.17	35.4	-86.82	-158.90	1.46	A <sub>II</sub>
	16.17	33.9	-170.35	-159.62	1.00	A
	158.59	38.5	-58.54	-146.11	1.06	B <sub>II</sub>
	158.87	38.4	-166.19	-139.08	0.00	B <sub>I</sub>

**TABLE 2-** Conformational angles (degrees) and relative energies values (kcal/mol) of OH-RIB-U in the model 1 (H<sub>2</sub>O set away from O<sub>3</sub>) optimized using the methods of calculation shown in column 1.  $\Delta E_1$  is reported to the energy of the B<sub>I</sub> conformation of model 1.

Methods of calculation	P	$\tau$	$\varepsilon$	$\chi$	$\Delta E_1$	Forms
B3LYP/6-31G*	16.17	38.9	-92.96	-162.78	3.60	A <sub>II</sub>
	17.07	35.5	-155.00*	-163.34	2.67	A
	-170.69	34.0	-75.33	-172.25	3.17	“B <sub>II</sub> ”
	-177.51	33.1	-155.00*	-171.71	0.00	“B <sub>I</sub> ”
B3LYP/CPCM/6-31G*	9.96	38.8	-92.77	-159.36	0.42	A <sub>II</sub>
	10.40	37.7	-155.00*	-160.08	-0.02	A
	173.91	37.9	-67.33	-124.52	0.38	B <sub>II</sub>
	172.41	36.7	-155.00*	-123.68	0.00	B <sub>I</sub>
B3LYP/6-31G* + WT	13.30	41.2	-90.64	-159.01	-1.33	A <sub>II</sub>
	12.41	39.2	-155.00*	-159.46	-2.06	A
	171.28	35.3	-74.45	-128.21	3.58	B <sub>II</sub>
	167.81	33.6	-155.00*	-129.29	0.00	B <sub>I</sub>

\* Dihedral angle  $\varepsilon$  frozen during the minimization

**TABLE 3-** Conformational angles (degrees) and energies values (kcal/mol) of OH-RIB-U in the model 2 (H<sub>2</sub>O points to O<sub>3</sub>) optimized using the methods of calculation shown in column 1.  $\Delta E_2$  is reported to the B<sub>I</sub> conformation of model 2.  $\Delta E_1$  is reported to the energy of the B<sub>I</sub> conformation of model 1.

Methods of calculation	P	$\tau$	$\varepsilon$	$\chi$	$\Delta E_2$	$\Delta E_1$	Forms
B3LYP/6-31G*	13.78	36.8	-157.50	-158.53	-0.48	0.17	A
	164.26	36.2	-81.56	-124.53	0.15	0.80	B <sub>II</sub>
	158.44	33.7	-155.00*	-127.36	0.00	0.65	B <sub>I</sub>
B3LYP/CPCM/6-31G*	12.09	35.5	-157.50	-157.22	0.09	-1.79	A
	172.32	36.8	-81.72	-125.46	0.22	-1.66	B <sub>II</sub>
	171.43	32.8	-166.81	-125.28	0.00	-1.88	B <sub>I</sub>
B3LYP/6-31G* + WT	16.67	35.4	-155.84	-163.47	-2.34	0.51	A
	165.10	35.1	-85.65	-126.98	-0.15	2.70	B <sub>II</sub>
	159.68	30.8	-155.72	-129.59	0.00	2.85	B <sub>I</sub>

\* Dihedral angle  $\varepsilon$  was frozen during the minimization.

**TABLE 4-** Conformational angles (degrees) and relative energies values (kcal/mol) of OCH3-RIB-U in the model 1 (methyl group away from O<sub>3'</sub>) optimized using the methods of calculation shown in column 1.  $\Delta E_1$  is reported to the energy of the B<sub>I</sub> conformation of model 1.

Methods of calculation	P	$\tau$	$\varepsilon$	$\chi$	$\Delta E_1$	Forms
B3LYP/6-31G*	14.69	40.2	-92.23	-158.26	0.43	A <sub>II</sub>
	13.66	38.1	-155.00*	-158.77	-0.17	A
	170.62	35.9	-71.55	-126.10	2.83	B <sub>II</sub>
	166.64	34.3	-155.00*	-127.95	0.00	B <sub>I</sub>
B3LYP/CPCM/6-31G*	11.80	38.1	-92.33	-156.32	0.44	A <sub>II</sub>
	11.38	37.0	-155.00*	-155.91	0.04	A
	177.34	36.6	-68.31	-125.93	0.63	B <sub>II</sub>
	175.59	34.9	-155.00*	-125.34	0.00	B <sub>I</sub>
B3LYP/6-31G* + WT	12.76	41.0	-93.50	-157.65	0.09	A <sub>II</sub>
	12.15	38.9	-155.00*	-157.38	-0.44	A
	170.51	37.0	-72.53	-120.26	2.98	B <sub>II</sub>
	166.99	35.4	-155.00*	-121.25	0.00	B <sub>I</sub>

\* Dihedral angle  $\varepsilon$  frozen during the minimization

**TABLE 5-** Conformational angles (degrees) and relative energies values (kcal/mol) of OCH<sub>3</sub>-RIB-U in model 2 (methyl group points to O<sub>3'</sub>) optimized using the methods of calculation shown in column 1.  $\Delta E_2$  is reported to the B<sub>I</sub> conformation of model 2.  $\Delta E_1$  is reported to the energy of the B<sub>I</sub> conformation of model 1.

Methods of calculation	P	$\tau$	$\varepsilon$	$\chi$	$\Delta E_2$	$\Delta E_1$	Forms
B3LYP/6-31G*	13.22	40.5	-91.88	-161.47	-0.85	0.48	A <sub>II</sub>
	11.16	39.6	-157.20	-162.96	-1.17	0.16	A
	166.08	35.5	-67.09	-122.50	0.50	1.83	B <sub>II</sub>
	163.69	36.7	-155.00*	-124.66	0.00	1.33	B <sub>I</sub>
B3LYP/CPCM/6-31G*	11.03	37.7	-96.43	-157.53	0.47	0.55	A <sub>II</sub>
	10.42	37.7	-156.91	-157.49	-0.16	-0.08	A
	172.66	35.7	-70.57	-124.37	-0.53	-0.45	B <sub>II</sub>
	172.98	36.7	-155.00*	-123.82	0.00	0.08	B <sub>I</sub>
B3LYP/6-31G* + WT	16.44	39.5	-90.28	-163.61	-1.05	-2.81	A <sub>II</sub>
	13.96	38.6	-153.35	-164.54	-1.30	-3.06	A
	161.75	35.2	-62.68	-132.70	-0.05	-1.81	B <sub>II</sub>
	160.83	35.7	-155.00*	-132.67	0.00	-1.76	B <sub>I</sub>

\* Dihedral angle  $\varepsilon$  frozen during the minimization

**TABLE 6-** Estimates of the east barriers for the compounds studied here. The energy Be (kcal/mol), and the puckering amplitude  $\tau$  are calculated for the O4' endo ( $\tau_2=0^\circ$ ) conformations.

Compounds	Models	Methods of calculation	Be	$\tau$
H-U		B3LYP/6-31G*	2.34	17.7
		B3LYP/CPCM/6-31G*	2.82	12.0
		B3LYP/6-31G* + WT	1.68	15.4
OH-RIB-U	1	B3LYP/6-31G*	2.76	29.4
		B3LYP/CPCM/6-31G*	3.68	36.5
		B3LYP/6-31G* + WT	3.28	14.7
OH-RIB-U	2	B3LYP/6-31G*	2.46	22.9
		B3LYP/CPCM/6-31G*	2.96	20.0
		B3LYP/6-31G* + WT	2.44	24.0
OCH3-RIB-U	1	B3LYP/6-31G*	2.58	19.5
		B3LYP/CPCM/6-31G*	2.71	15.3
		B3LYP/6-31G* + WT	3.34	17.4
OCH3-RIB-U	2	B3LYP/6-31G*	3.90	26.1
		B3LYP/CPCM/6-31G*	3.17	38.9
		B3LYP/6-31G* + WT	3.55	32.2

## FIGURE LEGENDS

**FIGURE 1-** Atom numbering and chemical structure of the compounds studied.

**FIGURE 2-** B<sub>I</sub> and A conformations of H-U obtained by geometry optimization at the B3LYP/6-31G\* level of theory (a). B<sub>II</sub> and A conformations of H-U obtained by geometry optimization from the B3LYP/6-31G\* + WT method of calculation (b). Hydrogen bonds are represented with dashed lines.

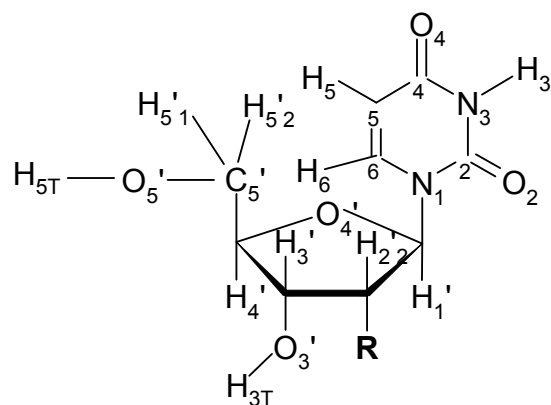
**FIGURE 3-** “B<sub>I</sub>” and A conformations of OH-RIB-U (model 1: H<sub>O</sub>·<sub>2</sub> away from O<sub>3</sub>·) obtained by geometry optimization at the B3LYP/6-31G\* level of theory (a). B<sub>I</sub> conformation of OH-RIB-U (model 1) obtained by geometry optimization from the B3LYP/CPCM/6-31G\* method of calculation (b). B<sub>I</sub> and A conformations of OH-RIB-U (model 1) obtained by geometry optimization from the B3LYP/6-31G\* + WT method of calculation (c). Hydrogen bonds are represented with dashed lines.

**FIGURE 4-** B<sub>I</sub> and A conformations of OH-RIB-U (model 2: H<sub>O</sub>·<sub>2</sub> points to O<sub>3</sub>·) obtained by geometry optimization at the B3LYP/6-31G\* level of theory (a). B<sub>I</sub> and A conformations of OH-RIB-U (model 2) obtained by geometry optimization from the B3LYP/6-31G\* + WT method of calculation (b). Hydrogen bonds are represented with dashed lines..

**FIGURE 5-** - B<sub>I</sub> and A conformations of OCH<sub>3</sub>-RIB-U (model 1: methyl group away from O<sub>3</sub>·) obtained by geometry optimization at the B3LYP/6-31G\* level of theory (a). B<sub>I</sub> and A conformations of OCH<sub>3</sub>-RIB-U (model 1) obtained by geometry optimization from the B3LYP/6-31G\* + WT method of calculation (b). Hydrogen bonds are represented with dashed lines.

**FIGURE 6-** B<sub>I</sub> and A conformations of OCH<sub>3</sub>-RIB-U (model 2: methyl group points to O<sub>3</sub>·) obtained by geometry optimization at the B3LYP/6-31G\* level of theory (a). B<sub>I</sub> and A conformations of OCH<sub>3</sub>-RIB-U (model 2) obtained by geometry optimization from the B3LYP/6-31G\* + WT method of calculation (b). Hydrogen bonds are represented with dashed lines.

FIGURE 1



Compounds	R
H-U	$\begin{array}{c}   \\ \text{H}_{2'1} \end{array}$
OH-RIB-U	$\begin{array}{c}   \\ \text{O}_{2'} - \text{H}_{\text{O}2} \end{array}$
OCH3-RIB-U	$\begin{array}{c}   \\ \text{O}_{2'} - \text{C}_{\text{X}2} \begin{array}{l} \text{---} \text{H}_{\text{X}2\text{A}} \\ \text{---} \text{H}_{\text{X}2\text{B}} \\ \text{---} \text{H}_{\text{X}2\text{C}} \end{array} \end{array}$



FIGURE 2

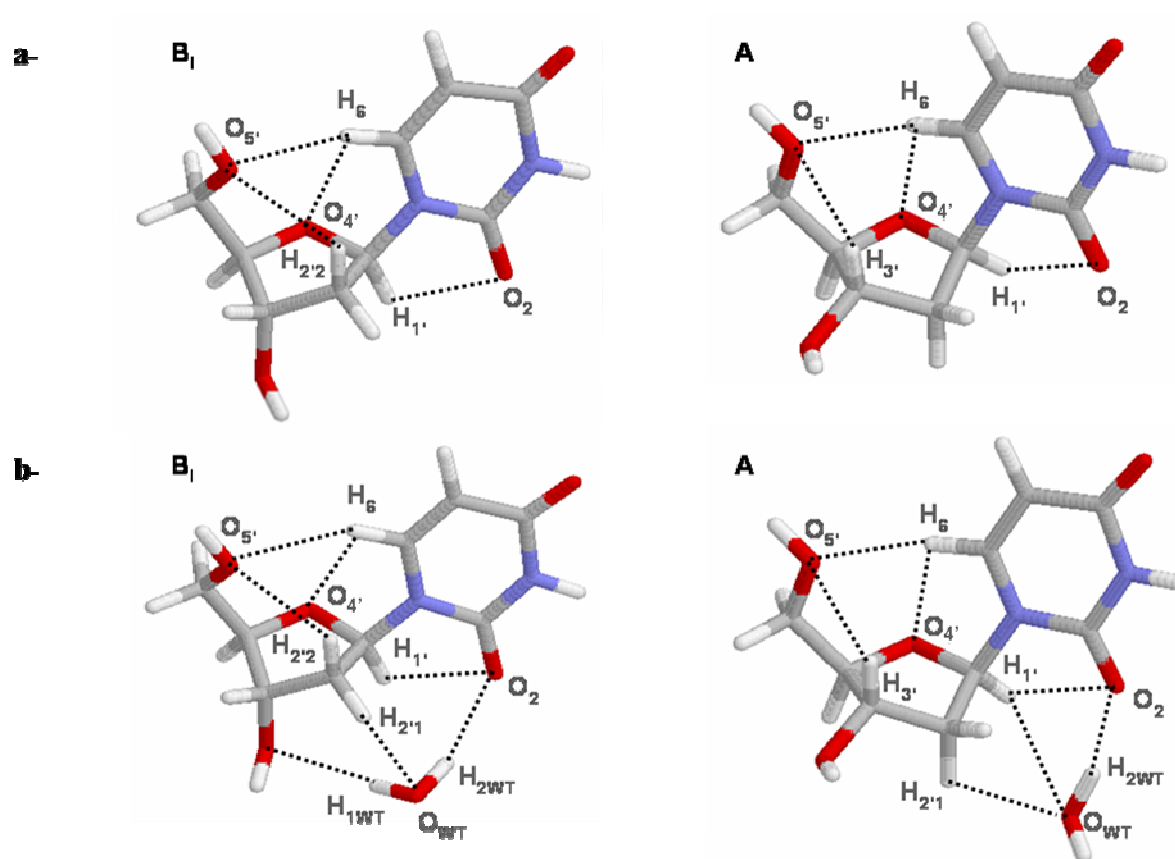
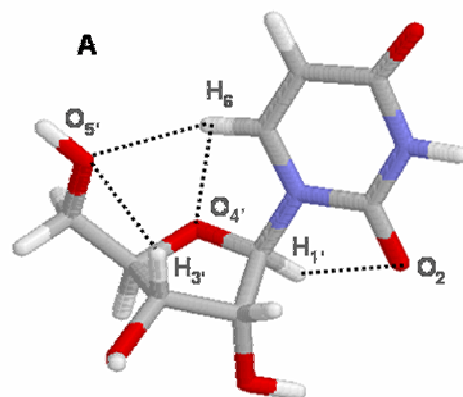
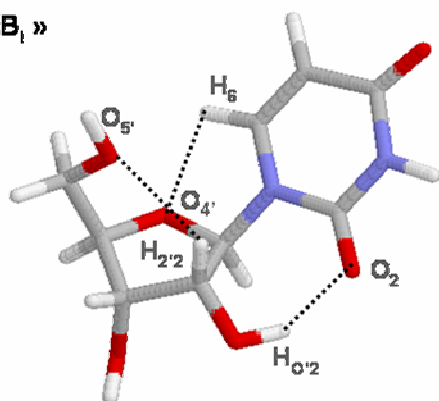
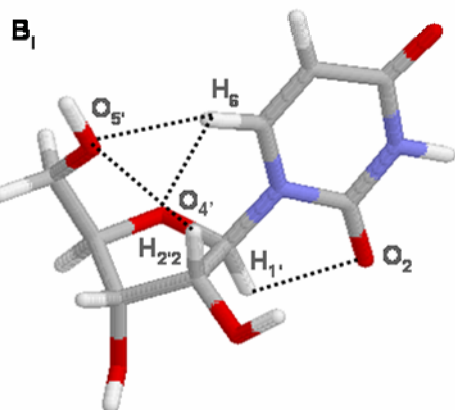


FIGURE 3

a- « B<sub>1</sub> »

b-



c-

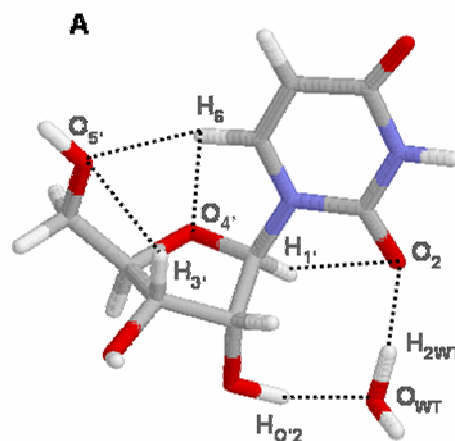
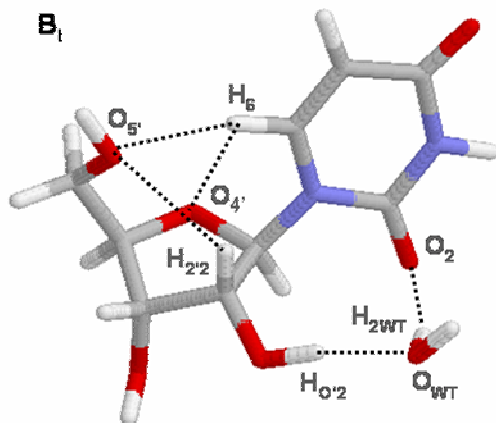


FIGURE 4

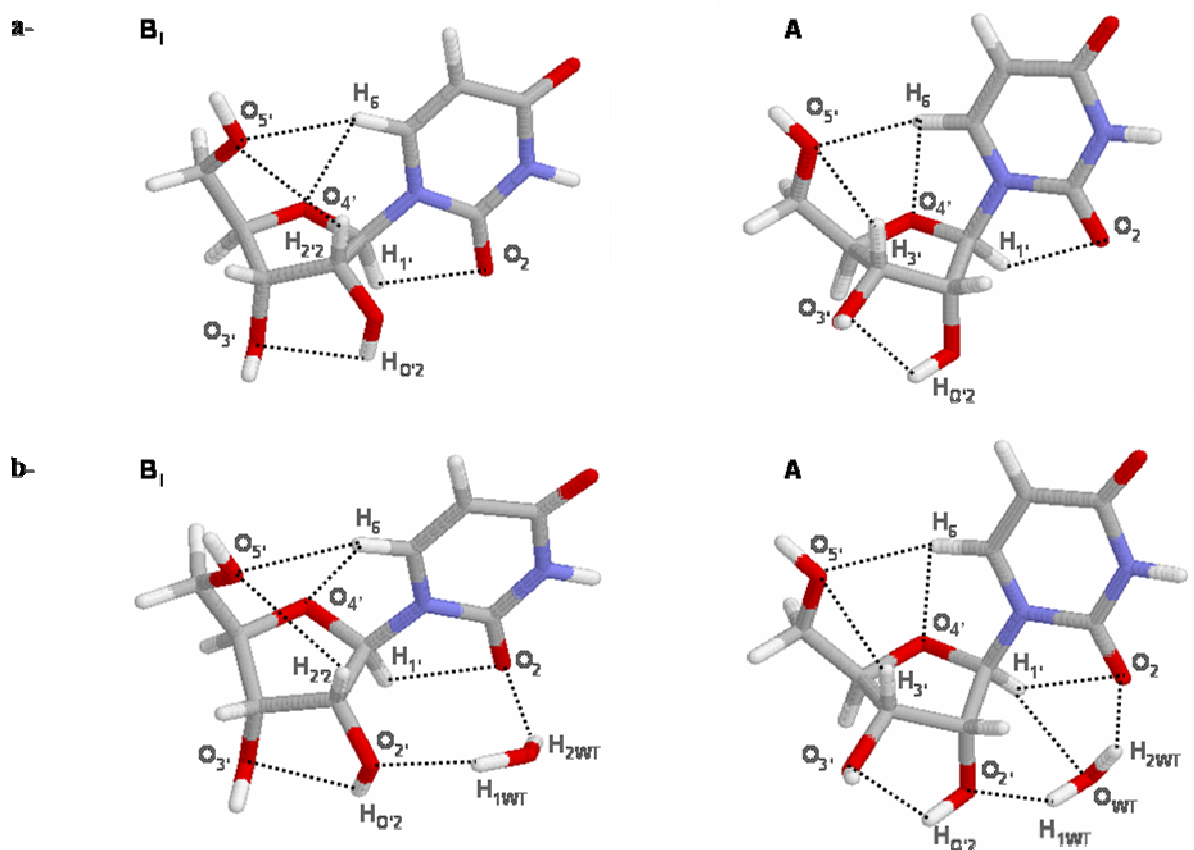


FIGURE 5

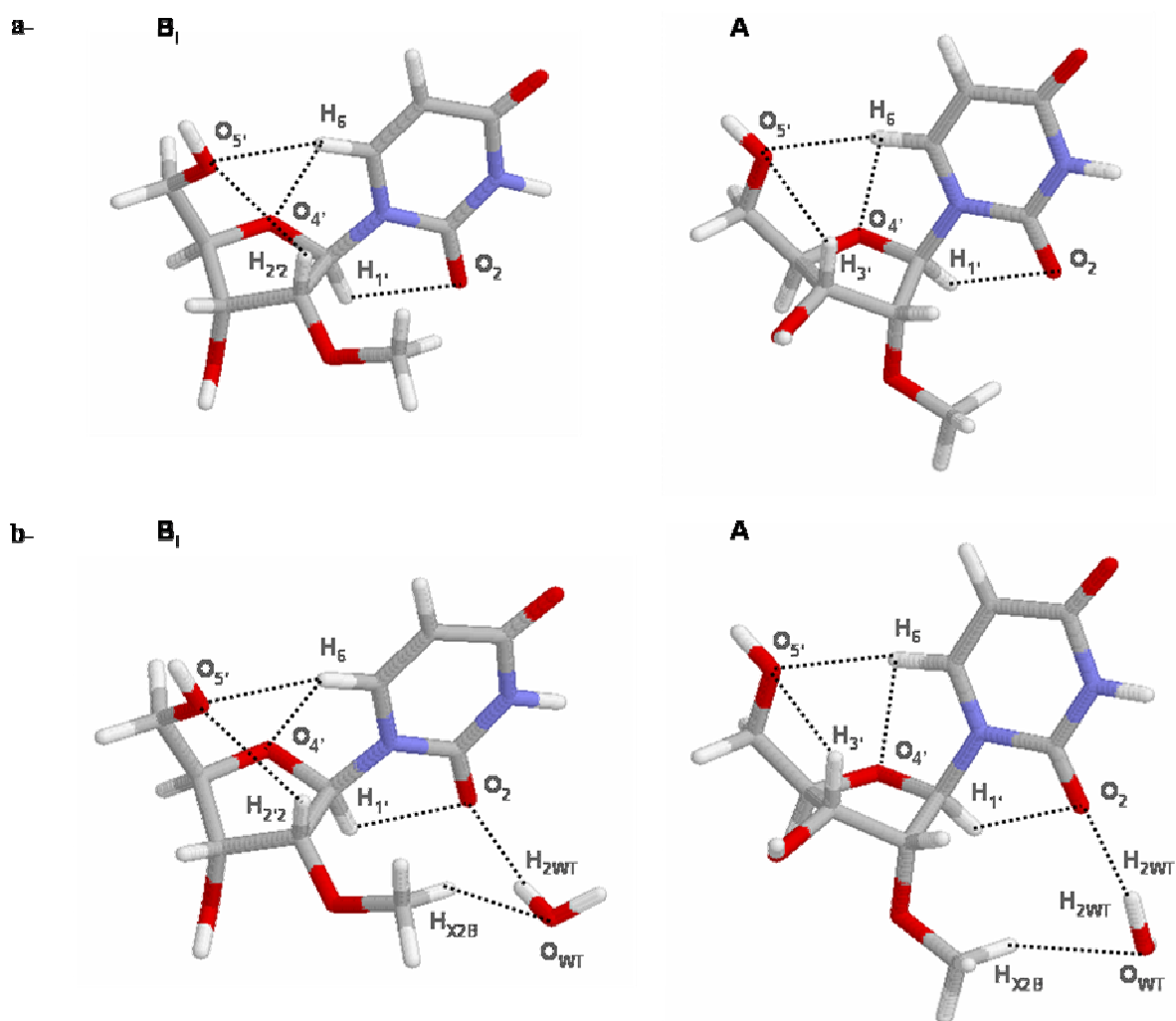
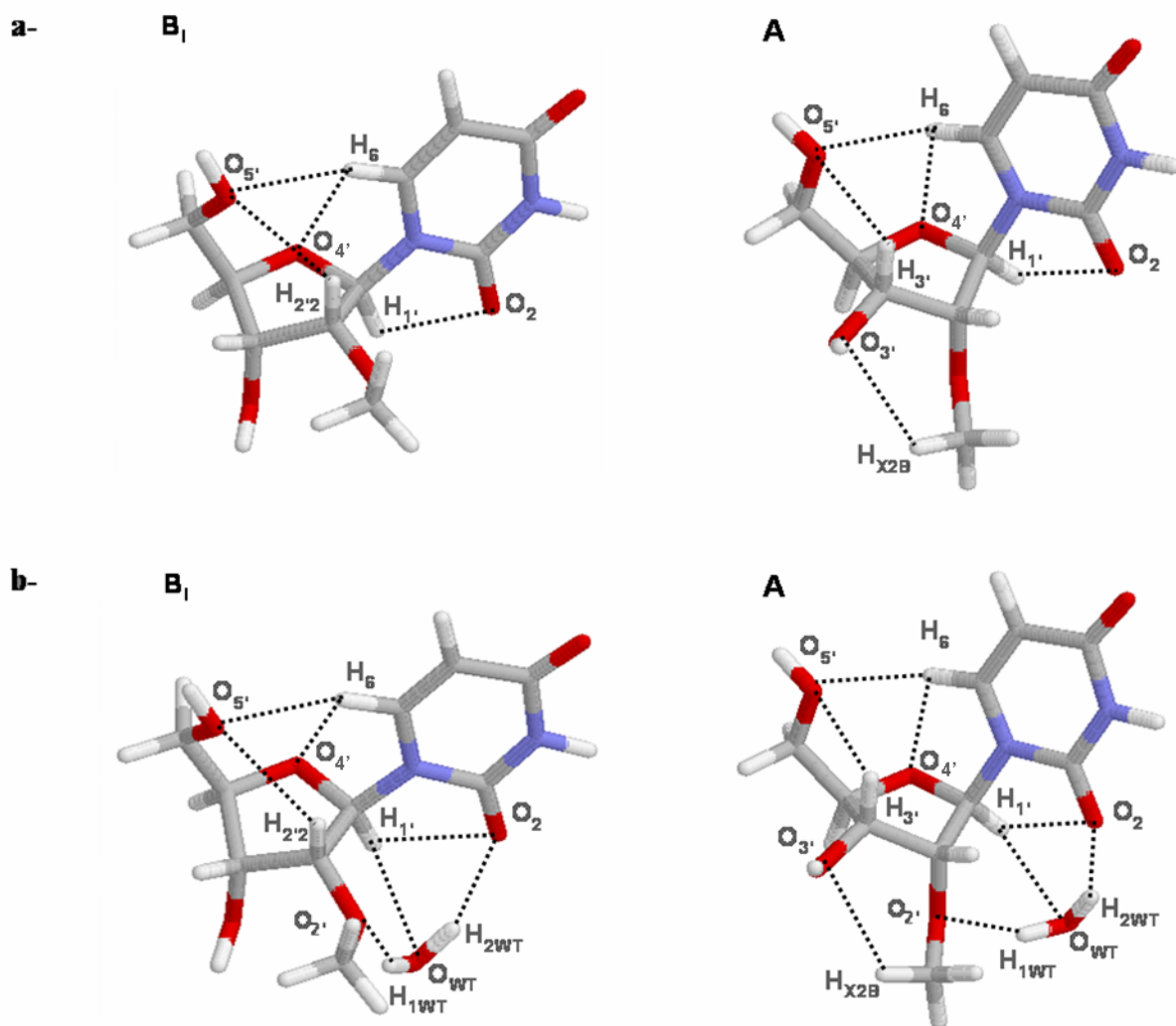


FIGURE 6



## Chapitre III

### Etude de l'interaction IN/ADN en utilisant des analogues de l'ADN viral

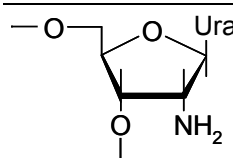
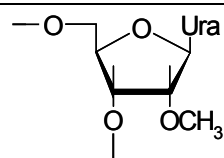
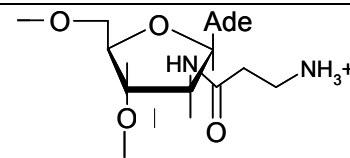
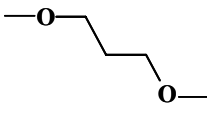
Pour catalyser la réaction de 3'-processing, l'IN se lie spécifiquement à certaines séquences localisées aux extrémités LTR de l'ADN viral. Cependant, les interactions à travers lesquelles l'IN forme un complexe stable et actif avec l'ADN viral ne sont pas encore comprises.

Dans le but d'identifier les caractéristiques structurales de l'ADN viral déterminant sa reconnaissance et son clivage par l'IN, l'effet d'analogues de l'ADN viral sur le 3'-processing a été mesuré. Ces analogues sont des dodécanucléotides doubles brins correspondant à l'extrémité LTR U5 de l'ADN viral (**tableau 1**) mais contenant des nucléotides modifiés en 2' ou le 1,3-propanediol (**figure 1**). Ces modifications ont été introduites à différentes positions sur l'un des deux brins ou sur les deux brins.

**Tableau 1-** Dodécanucléotide double brin correspondant à l'extrémité LTR U5 de l'ADN viral

brin U5B (scissible)	5'	A12	T11	C10	T9	C8	T7	A6	G5	C4	A3	G2	T1	3'
brin U5A	3'	T12	A11	G10	A9	G8	A7	T6	C5	G4	T3	C2	A1	5'

			
NH2-RIB-U (UN)	OCH3-RIB-U (UM)	NX-ARA-A (aAN)	1,3- propanediol (P)

**Figure 1-** Les nucléotides modifiés en 2' et le 1,3-propanediol

Les mesures des taux initiaux de 3'-processing, en présence des cofacteurs  $Mn^{2+}$  ou  $Mg^{2+}$  montrent que certaines modifications affectent considérablement le 3'-processing comme exposé dans l'article joint ci-après et publié dans *The Journal of Biological Chemistry*. Par contre, aucun effet significatif sur la liaison de l'IN n'a été observé. Il est vraisemblable que l'ADN se lie à l'IN puis glisse sur l'IN ou bien qu'il y ait une adaptation conformationnelle du complexe initial IN/ADN avant que le complexe chimiquement actif ne se forme.

Les effets des modifications aux positions 5, 6 et 3 dépendent fortement du cofacteur utilisé. Ces effets sont beaucoup plus prononcés en présence de  $Mg^{2+}$  que de  $Mn^{2+}$ , certainement parce que l'IN est plus rigide en présence de  $Mg^{2+}$  (Esposito and Craigie, 1998). Les effets des modifications des paires de bases 7 à 9 sont identiques en présence des deux ions. D'après ces résultats, les positions 3 à 6 de l'ADN viral interagissent avec le domaine central de l'IN qui lie le cofacteur métallique. Le domaine C-terminal connu pour interagir avec l'ADN viral et qui n'est pas directement impliqué dans la liaison de l'ion métallique, pourrait donc interagir avec les positions 7 à 9 de l'ADN viral. Ces résultats sont en accord avec les données de pontage (Jenkins *et al.*, 1997; Gerton *et al.*, 1998; Esposito and Craigie, 1998; Gao *et al.*, 2001).

# Probing of HIV-1 Integrase/DNA Interactions Using Novel Analogs of Viral DNA<sup>\*[5]</sup>

Received for publication, November 15, 2005, and in revised form, February 21, 2006. Published, JBC Papers in Press, February 24, 2006, DOI 10.1074/jbc.M512271200

Julia Agapkina<sup>‡</sup>, Maksim Smolov<sup>‡</sup>, Sophie Barbe<sup>§</sup>, Evgenii Zubin<sup>‡</sup>, Timofei Zatsepin<sup>‡</sup>, Eric Deprez<sup>§</sup>, Marc Le Bret<sup>§</sup>, Jean-François Mouscadet<sup>§1</sup>, and Marina Gottikh<sup>‡</sup>

From the <sup>‡</sup>Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119992 Moscow, Russia and <sup>§</sup>LBPA, UMR 8113 CNRS, Ecole Normale Supérieure de Cachan, 61 Avenue du Président Wilson, 94235 Cachan Cedex, France

The specific activity of the human immunodeficiency virus, type 1 (HIV-1), integrase on the viral long terminal repeat requires the binding of the enzyme to certain sequences located in the U3 and U5 regions at the ends of viral DNA, but the determinants of this specific DNA-protein recognition are not yet completely understood. We synthesized DNA duplexes mimicking the U5 region and containing either 2'-modified nucleosides or 1,3-propanediol insertions and studied their interactions with HIV-1 integrase, using Mn<sup>2+</sup> or Mg<sup>2+</sup> ions as integrase cofactors. These DNA modifications had no strong effect on integrase binding to the substrate analogs but significantly affected 3'-end processing rate. The effects of nucleoside modifications at positions 5, 6, and especially 3 strongly depended on the cationic cofactor used. These effects were much more pronounced in the presence of Mg<sup>2+</sup> than in the presence of Mn<sup>2+</sup>. Modifications of base pairs 7–9 affected 3'-end processing equally in the presence of both ions. Adenine from the 3rd bp is thought to form at least two hydrogen bonds with integrase that are crucial for specific DNA recognition. The complementary base, thymine, is not important for integrase activity. For other positions, our results suggest that integrase recognizes a fine structure of the sugar-phosphate backbone rather than heterocyclic bases. Integrase interactions with the unprocessed strand at positions 5–8 are more important than interactions with the processed strand for specific substrate recognition. Based on our results, we suggest a model for integrase interaction with the U5 substrate.

Following reverse transcription, a DNA copy of the human immunodeficiency virus, type 1 (HIV-1),<sup>2</sup> RNA is integrated into the genome of infected cells. Integration is a prerequisite for viral replication and is catalyzed by the viral enzyme integrase (IN). IN binds to sequences located at the end of U3 and U5 parts of long terminal repeats (LTRs) of viral DNA and catalyzes the trimming, or 3'-end processing, of the terminal dinucleotide from the 3'-ends of both strands of the DNA. IN then mediates a strand transfer reaction that inserts the viral DNA into the host DNA. During this reaction, IN must bind simultaneously to viral and target DNA. However, IN interacts with these two DNA mol-

ecules in different ways as follows: binding to host DNA does not depend directly on host DNA sequence, whereas interaction with the viral DNA is a sequence-specific process. Nevertheless, the U5 and U3 sequences recognized by IN are not exactly identical.

Strand transfer and 3'-end processing reactions may be carried out *in vitro*, using recombinant HIV IN, DNA duplexes mimicking U3 or U5 sequences of LTRs, and divalent metal ions, such as Mg<sup>2+</sup> or Mn<sup>2+</sup>. However, the Mn<sup>2+</sup>- and Mg<sup>2+</sup>-dependent activities of IN are not equivalent, with lower specificity reported for Mn<sup>2+</sup>-dependent IN (1, 2). Moreover, the inhibition of HIV-1 IN by compounds such as  $\beta$ -diketo acids, which interact with the active site of HIV-1 IN, is also metal-dependent (3). The metal ion is coordinated within the IN catalytic core by two Asp residues (Asp-64 and Asp-116) and water molecules (4–6). These Asp residues are part of the D,D(35)E motif, a triad of invariant acidic amino acids found in all retroviral integrases (7) and ultimately required for catalytic activity (2, 8–10). Binding of the metal cofactor by the amino acids of the D,D(35)E motif causes both local and long range conformational changes and the activation of IN (5, 11). The IN catalytic core binds both the metal cofactor and the substrate. These conformational changes therefore influence the specificity of IN interactions with substrate DNA. Differences in the coordination properties of Mg<sup>2+</sup> and Mn<sup>2+</sup> ions may result in different conformational changes in the IN catalytic core. This may account for differences in the specificity of HIV-1 IN interactions with substrate DNA (1, 2), inhibitors (3, 12), and the nucleophiles used to carry out 3'-end processing (13). As IN is more sensitive to the structure of molecules interacting with its catalytic core in the presence of magnesium ions, it has been suggested that the catalytic core of this molecule is more rigid in the presence of magnesium ions, as reported for other magnesium-dependent enzymes (1).

IN has three domains: the N-terminal, catalytic, and C-terminal domains. The structure of each domain, and of two-domain fragments, has been determined by x-ray crystallography or NMR spectroscopy (4, 6, 14–16). However, neither the structure of the entire enzyme nor that of the IN-viral DNA complex has yet been determined. Cross-linking experiments have been used to study contacts between IN and viral DNA, and it has been found that the catalytic core (residues 50–212) and C-terminal domain (residues 213–288) are the principal areas involved in DNA binding. It has been shown that the Lys-159 residue of HIV-1 integrase may interact with the N-7 position of the third adenosine of the processed viral DNA strand (17) and that Glu-152 may also interact with this adenosine (18). Gln-148 and Tyr-143 have been shown to bind to the terminal nucleotides of the processed strands of viral DNA, and the thymidine at position 7 interacts with amino acids from the C-terminal domain (1). Glu-246 is also thought to be located near the C-2' atom of the 2'-deoxyadenosine at position 7 in the unprocessed strand (19). Nevertheless, the protein/DNA interactions through which IN forms a stable and catalytically active complex with viral DNA are not yet completely understood.

\* This work was supported by the TRIOH European Project FP6 Grant 503480, the Russian Foundation for Basic Research Grants 04-04-22000 and 05-04-48743, the CNRS (PIC527), and the French National Agency for Research on AIDS (ANRS). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[5] The on-line version of this article (available at <http://www.jbc.org>) contains Tables S1 and S2 and Fig. S1.

<sup>1</sup> To whom correspondence should be addressed: LBPA-CNRS UMR8113, Ecole Normale Supérieure de Cachan, 61 Avenue du Président Wilson, 94235 Cachan Cedex, France. Tel.: 33-1-47-40-76-75; Fax: 33-1-47-40-76-84; E-mail: mouscadet@lbpa.ens-cachan.fr.

<sup>2</sup> The abbreviations used are: HIV-1, human immunodeficiency virus type 1; LTR, long terminal repeat; IN, integrase; DTT, dithiothreitol; DAP, 2,6-diaminopurine.



## New HIV DNA Analogs for Studying Interactions with Integrase

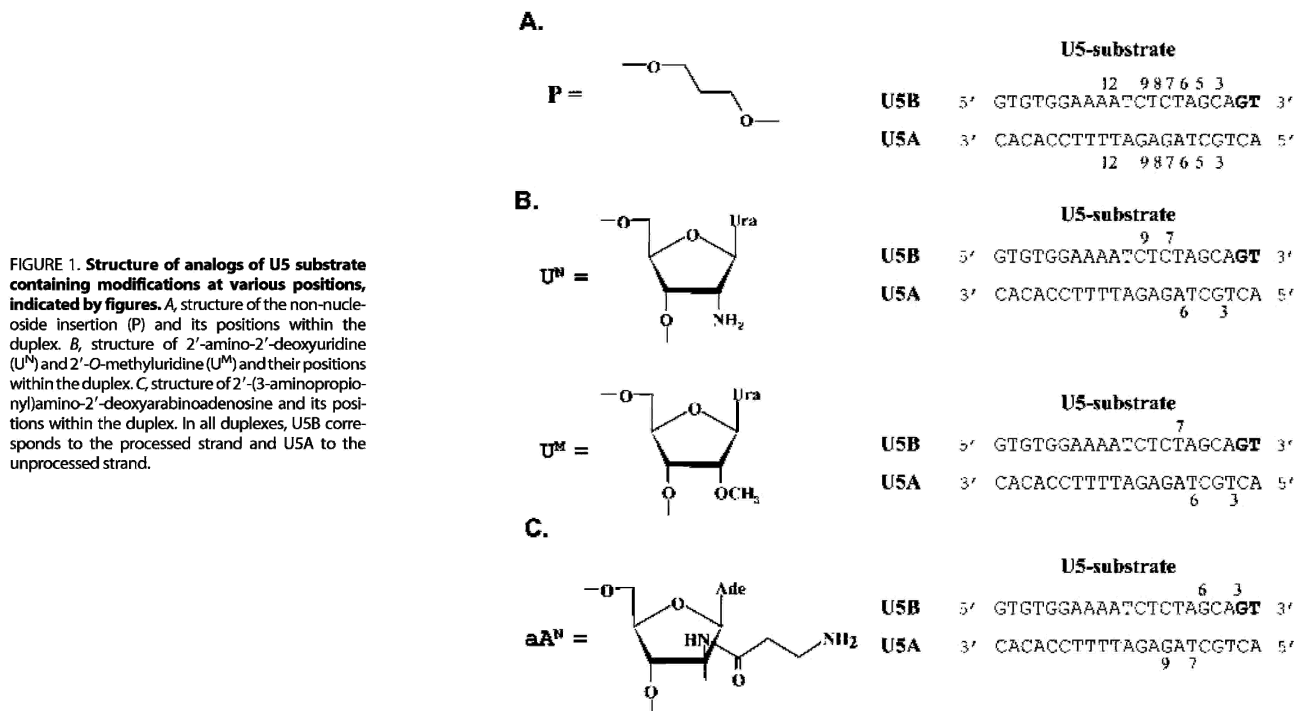


FIGURE 1. Structure of analogs of U5 substrate containing modifications at various positions, indicated by figures. A, structure of the non-nucleoside insertion (P) and its positions within the duplex. B, structure of 2'-amino-2'-deoxyuridine (U<sup>N</sup>) and 2'-O-methyluridine (U<sup>M</sup>) and their positions within the duplex. C, structure of 2'-(3-aminopropionyl)amino-2'-deoxyarabinoadenosine and its positions within the duplex. In all duplexes, USB corresponds to the processed strand and USA to the unprocessed strand.

The third and fourth nucleotides of the processed strand of the viral DNA (the so-called conserved 5'-CA-3') are important for 3'-end processing and strand transfer activities of IN (1, 20, 21). It has been shown that the region from nucleotides 4 to 9 of the viral DNA is important for efficient 3'-end processing (22–25). Modification of the nucleosides at positions 5 and 6 in the processed strand significantly decreases 3'-end processing efficiency (1, 20, 25), and it has been suggested that IN comes into contact with the heterocyclic bases of these nucleosides (26). However, the mechanism of IN interactions with the processed and unprocessed strands of the HIV DNA remains to be elucidated.

We therefore investigated the effect of the double-helix structure of viral DNA on HIV-1 integrase activity in the 3'-end processing reaction, as a means of identifying the structural features of the viral DNA determining the sequence specificity of its recognition and processing by IN. We synthesized a set of modified DNA duplexes, with sequences mimicking the U5 sequence of LTR of the viral DNA (U5 substrate). Nucleosides at various positions in the processed strand, the unprocessed strand, or both were replaced by a non-nucleoside insertion, 1,3-propanediol, or nucleosides containing modified sugar residues, 2'-amino-nucleosides and 2'-O-methylnucleosides (Fig. 1). Because the conserved CA sequence adenosine is considered crucial for 3'-end processing and strand transfer activities, we also prepared a number of U5 substrate analogs in which the A/T pair was replaced by unpaired nucleotides in the 3rd position and a 2,6-diaminopurine replaced the conserved adenine.

We have shown that IN probably recognizes the viral DNA by forming at least two hydrogen bonds with the adenine at position 3. The complementary base, thymine, does not interact with IN. We suggest that IN recognizes the fine structure of the sugar-phosphate backbone rather than heterocyclic bases at other positions. We also found that interactions with the unprocessed strand were more important for specific substrate recognition by IN than interactions with the processed strand. We showed that local destabilization of the A/T pair at position 3 was required for 3'-end processing, allowing such processing to occur even at internal positions of the U5 substrate within DNA. Finally, we

studied IN/DNA interactions in the presence of Mn<sup>2+</sup> and Mg<sup>2+</sup> ions, investigating the differences in substrate recognition specificity in the presence of these two cofactors. We found that modification of the nucleosides in the 6th, 5th, and especially the 3rd positions influenced 3'-end processing more strongly in the presence of Mg<sup>2+</sup> than in the presence of Mn<sup>2+</sup> ions, whereas modification of base pairs 7–9 affected 3'-end processing equally in the presence of both ions. We suggest that positions 3–6 within the substrate DNA interact with the IN catalytic core that binds the metal cofactor.

## EXPERIMENTAL PROCEDURES

**Oligonucleotide Synthesis**—Oligonucleotides were synthesized using an Applied Biosystems 380B DNA synthesizer and the standard  $\beta$ -cyanoethyl phosphoramidite procedure. The resulting oligonucleotides were purified by electrophoresis in a 20% denaturing polyacrylamide gel. Oligonucleotides containing 2'-(3-aminopropionyl)amino-2'-deoxyarabinoadenosine (Fig. 1C) and 2'-amino-2'-deoxyuridine (Fig. 1B) were synthesized as described previously (27, 28). Oligonucleotides containing 1,3-propanediol insertions (Fig. 1A) were synthesized using the commercially available Spacer C3 (Glen Research). An oligonucleotide containing 2,6-diaminopurine was synthesized using commercially available 2-amino-dA-CE phosphoramidite (Glen Research). Structures of all modified oligonucleotides were confirmed by matrix-assisted laser desorption/ionization time of flight. Hybridization of oligonucleotides U5B and U5A (U5B being the processed strand) gave the wild-type U5 integrase substrate.

The fluorescein-labeled oligonucleotides U5B-fl (5'-fl-GTGTGG-AAAATCTCTAGCAGT-3') and U5A-fl (3'-fl-CACACCTTTTAGAGATCGTCA-5') used for IN/DNA binding studies were synthesized as follows: oligonucleotide U5B, with an amino group at its 5'-end, and oligonucleotide U5A, with an amino group at its 3'-end, were synthesized with 5'-amino-modifier C6 reagent and 3'-amino-modifier C3 controlled pore glass (Glen Research), respectively. The resulting oligonucleotides were dissolved in a mixture of 60  $\mu$ l of H<sub>2</sub>O, 20  $\mu$ l of 1 M

## New HIV DNA Analogs for Studying Interactions with Integrase

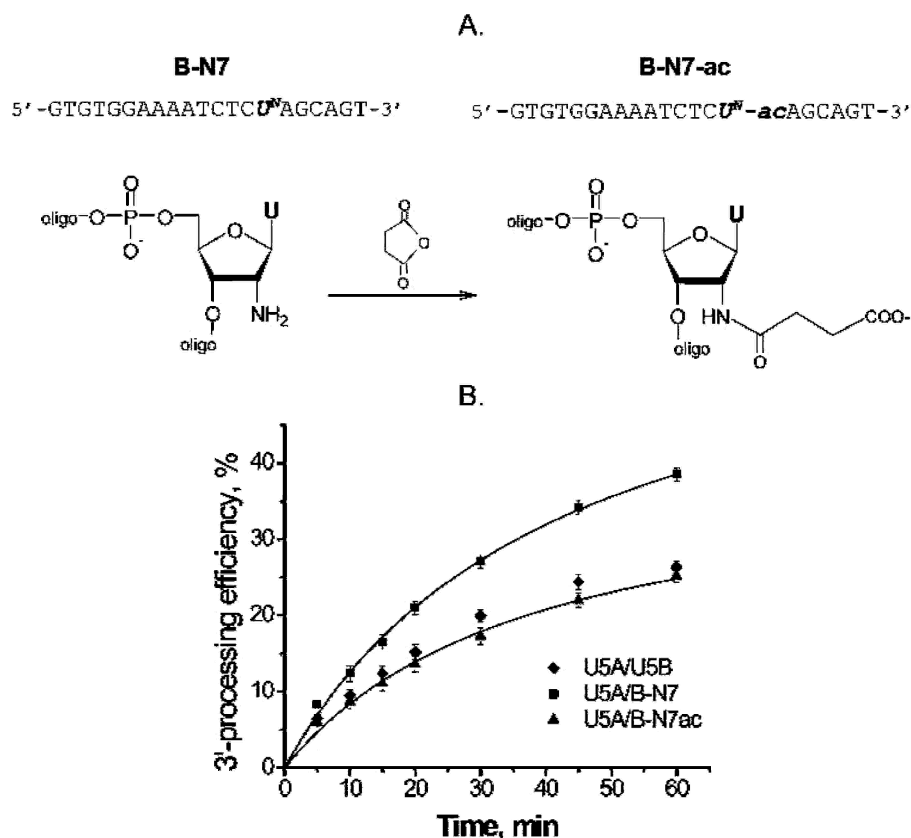


FIGURE 2. Stimulation of 3'-end processing because of the presence of the  $\text{NH}_2$  group in the sugar 2'-position of the 2'-aminouridine at position 7. A, acylation scheme for the amino group of oligonucleotide B-N7. B, kinetic curves for the 3'-end processing of the U5 substrate, U5A/B-N7, and U5A-B-N7-ac duplexes in the presence of  $\text{Mg}^{2+}$  ions.

carbonate buffer (pH 11), and 80  $\mu\text{l}$  of dimethylformamide. We then added 1 mg of fluorescein isothiocyanate to each solution. The mixtures were incubated in the dark at room temperature for 4–12 h, and 1 ml of 2%  $\text{LiClO}_4$  in acetone was added to precipitate the oligonucleotides. The reaction mixtures were analyzed by reversed-phase high performance liquid chromatography in ion pair mode, and the labeling efficiency was found to be greater than 90%. The conjugates were purified by electrophoresis in a 20% polyacrylamide denaturing gel, and the presence of fluorescein was confirmed by absorption spectroscopy.

**Synthesis of the Oligonucleotide Containing Acylated 2'-Amino Group**—A solution of 10 nmol of 2'-aminouridine containing oligonucleotide (B-N7) in 100  $\mu\text{l}$  of 1 M sodium borate buffer (pH 8.3) was supplemented with 100  $\mu\text{l}$  of a 0.1 M solution of succinic anhydride in acetonitrile, and the reaction mixture was incubated at 37  $^\circ\text{C}$  for 3 h (Fig. 2A). Acetonitrile was then evaporated, and the acylated oligonucleotide (B-N7-ac) was purified by gel filtration using NAP-10 columns (Amersham Biosciences). The efficiency of the amino group acylation determined by reversed-phase high performance liquid chromatography analysis in ion pair mode was about 95%. The acylated oligonucleotide (B-N7-ac) was further used for 3'-processing experiments without additional purification.

**Integrase**—Detergent-free recombinant integrase protein was produced in *Escherichia coli* and purified as described previously (29).

**Thermal Denaturation Studies**—Oligonucleotide duplexes (1  $\mu\text{M}$  in 20 mM Tris-HCl buffer (pH 7.0), 7.5 mM  $\text{MgCl}_2$ ) were heated at 90  $^\circ\text{C}$  for 3 min and slowly cooled to 10  $^\circ\text{C}$ . Melting studies were carried out in Teflon-stoppered 1-cm path length quartz cells, using a Hitachi 100–20 spectrophotometer equipped with a thermal programmer. We monitored absorbance at 260 nm as the temperature was increased at a rate of 0.5  $^\circ\text{C}/\text{min}$ . Melting temperatures ( $T_m$ ) were determined by fitting the

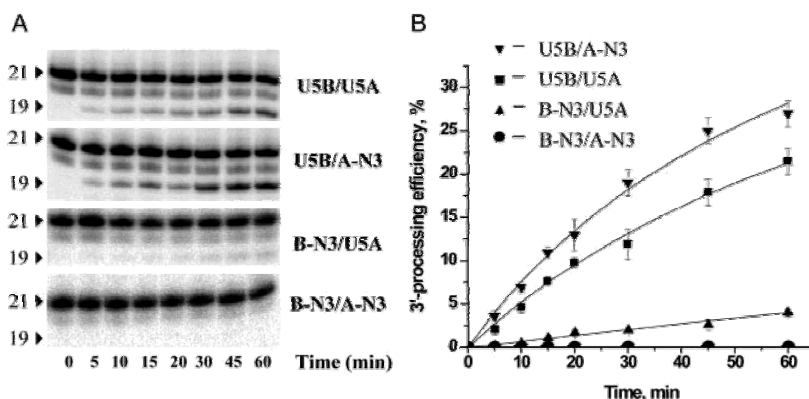
first derivative of absorbance with respect to  $1/T$ . Uncertainty in  $T_m$  was estimated to a precision of  $\pm 0.5$   $^\circ\text{C}$  by repeating experiments.

**Preparation of 5'-Radiolabeled DNA Duplexes**—For binding and processing assays, 10 pmol of the oligonucleotide U5B, or one of its modified analogs, was labeled using T4 polynucleotide kinase and 50  $\mu\text{Ci}$  of  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  (3000 Ci/mmol). T4 polynucleotide kinase was then inactivated by adding EDTA and heating at 65  $^\circ\text{C}$  for 5 min. The corresponding complementary unlabeled oligonucleotide, U5A, or its modified analog (10 pmol) was then added to each of the labeled oligonucleotides. The mixtures were heated at 90  $^\circ\text{C}$  for 3 min, and the DNA duplexes were allowed to anneal by slowly cooling to room temperature. DNA duplexes were purified using Micro Bio-Spin columns P-6 (Bio-Rad).

**3'-End Processing Assay**—The 3'-end processing reaction was carried out in a buffer containing 20 mM HEPES (pH 7.2), 1 mM DTT, and 7.5 mM  $\text{MnCl}_2$  or  $\text{MgCl}_2$  using the IN substrate U5B/U5A or an analog containing modifications in one or both strands.  $^{32}\text{P}$ -Labeled DNA substrates (1.25 nM) were incubated in 20  $\mu\text{l}$  of the buffer with 100 nM integrase at 37  $^\circ\text{C}$ . Aliquots were taken after 5, 10, 20, 30, 60, and 120 min. The reaction was stopped by adding 80  $\mu\text{l}$  of a stop solution (7 mM EDTA, 0.3 M sodium acetate, 10 mM Tris-HCl (pH 8)). IN was extracted in phenol/chloroform, and DNA fragments were precipitated with ethanol. The products were suspended in loading buffer (80% formamide, 0.05% bromphenol blue, 0.05% xylene cyanol) and separated on 20% polyacrylamide denaturing gels. Gels were analyzed on a Storm 840<sup>TM</sup> PhosphorImager (Amersham Biosciences) and quantified with ImageQuant<sup>TM</sup> 4.1 software (Fig. 3). We fitted a nonlinear regression curve to the experimental data, and initial rates were calculated as the tangent to the kinetic curves for  $t = 0$ . Rate data for modified substrates were

## New HIV DNA Analogs for Studying Interactions with Integrase

**FIGURE 3.** Effect on 3'-end processing of replacing nucleosides at position 3 by 2'-aminonucleosides in the presence of  $Mg^{2+}$  ions. **A**, electrophoresis analysis of 3'-end processing products in the case of the wild-type substrate (USB/USA) and its analogs containing the USB strand and  $U^N$  in the USA strand (USB/A-N3),  $aA^N$  in the USB strand and the USA strand (B-N3/USA), and 2'-aminonucleosides in both strands (B-N3/A-N3). **B**, kinetics of 3'-end processing product accumulation for wild-type U5 substrate and its analogs containing 2'-aminonucleosides in one or both strands.



normalized to the initial rate of 3'-end processing of the natural U5 substrate. All 3'-end processing experiments were carried out at least three times; the standard deviation was between 5 and 10%.

**DNA Binding Assay**— $^{32}P$ -Labeled duplexes (5 nM) were incubated with IN (10, 25, 50, 100, and 300 nM) in 10  $\mu$ l of a buffer containing 20 mM HEPES (pH 7.2), 1 mM DTT, 7.5 mM  $MgCl_2$ , and 5% (v/v) glycerol for 25 min at room temperature. A constant level of 20 mM NaCl was maintained in all reaction mixtures. The samples were then loaded on a 4% nondenaturing acrylamide gel containing 0.2% glycerol. Electrophoresis was carried out in a running buffer containing 50 mM Tris borate (pH 7.5) and 1 mM EDTA for 2 h at 4  $^{\circ}C$  (5.6 V/cm). A Storm 840 PhosphorImager (Amersham Biosciences) was used for analysis and quantification.

**Steady-state Fluorescence Anisotropy Measurements**—Steady-state fluorescence anisotropy values were recorded on a Beacon 2000 apparatus (Panvera, Madison, WI) in a thermostated cell, at 37  $^{\circ}C$  for kinetics studies or 20  $^{\circ}C$  for competition experiments. For kinetics studies, fluorescein-labeled DNA duplexes (4 nM) (wild-type substrate or modified oligonucleotides) were added to a 200- $\mu$ l buffered solution containing 20 mM HEPES (pH 7.2), 1 mM DTT, and 7.5 mM  $MgCl_2$ . IN-DNA complex formation was then triggered by adding IN (200 nM final concentration). A constant level of 20 mM NaCl was maintained in all reaction mixtures. Fluorescence anisotropy values were then recorded, and the binding curves were used to determine the pseudo first-order binding rate constant,  $k_{on} = k_1 \times [IN]$  (where  $k_1$  is the second-order binding rate constant). A single-exponential model was fitted to the data for DNA-IN complex formation.

For competition experiments, fluorescein-labeled DNA duplexes (2 nM) were first incubated with various concentrations of unlabeled DNA duplexes (either unmodified or modified oligonucleotides) in a buffered solution containing 20 mM HEPES (pH 7.2), 1 mM DTT, 20 mM NaCl, and 5 mM  $MgCl_2$ . IN (120 nM final concentration) was then added, and steady-state anisotropy values ( $r$ ) were recorded after equilibrium was reached. The  $r$  value corresponding to free fluorescein-labeled DNA was subtracted from the measured  $r$  value to give  $\Delta r$ . Finally,  $\Delta r$  values were plotted against competitor concentration to determine apparent  $K_d$  (concentration of unlabeled DNA decreasing the initial  $\Delta r$  value by 50%).

## RESULTS

This study aimed to identify the elements of viral DNA structure determining the sequence specificity of interactions with HIV-1 IN. We synthesized 21-mer DNA duplexes containing modifications inducing local distortions in the DNA double helix (27, 28, 30–32) and, therefore, DNA destabilization (supplemental Table S1). Oligonucleotides mim-

icking the HIV-1 U5 viral DNA end were synthesized with single modifications at a particular residue (Fig. 1). Modified substrate analogs contained only one modification in the processed (U5B) strand, the unprocessed (U5A) strand, or both. The native nucleosides in the U5 substrate were replaced by non-nucleoside insertions (1,3-propanediol residues), 2'-aminonucleosides, or 2'-*O*-methyluridine. It has been reported that the most important nucleosides for integration are those at positions 3–9 of the processed strand (17, 20, 21, 24, 26). However, nucleoside mutations distal to the LTR end (positions 10–13) have also been found to affect 3'-end processing, but only in the presence of  $Mg^{2+}$  ions (1). We assessed the importance of various nucleosides within the LTR sequence, by making modifications at positions 3, 5–9, and 12 of the U5 substrate (Fig. 1). Position 4 was not modified because the modification of this base pair has been shown to inhibit 3'-end processing (20, 21, 24, 26).

**Modifications at Position 12**—We investigated the influence of U5 substrate modifications on the ability of IN to mediate 3'-end processing. We found that replacement of the nucleosides at position 12 by the non-nucleoside insertion P did not affect IN activity in presence of  $Mg^{2+}$  or  $Mn^{2+}$  ions (supplemental Table S2). These nucleosides therefore do not form contacts with IN that are important for catalytic activity. By contrast, modifications at other sites on the U5 substrate significantly influenced the processing rate.

**Modifications at Positions 7–9**—Both non-nucleoside insertions and nucleoside modifications at position 9 decreased the 3'-end processing rate (Table 1). Similar effects were observed for modifications of the processed and unprocessed strands. Non-nucleoside insertions decreased the reaction rate more strongly than the insertion of a modified nucleoside. The 3'-end processing rate was lower if both substrate strands were modified. Therefore, nucleosides at position 9 in both strands of the DNA are involved in interactions with IN.

For positions 7 and 8, we found that all nucleoside modifications on the unprocessed strand, U5A, decreased the reaction rate (Table 1). By contrast, non-nucleoside insertions in the processed strand, U5B, had no effect on 3'-end processing rate (Table 1). Similar results were obtained if the thymidine at position 7 was replaced by 2'-*O*-methyluridine. Therefore, IN interacts principally with the nonprocessed strand of the U5 substrate at positions 7 and 8.

It should be especially noted that replacing position 7 of the U5B strand with 2'-aminouridine increased 3'-end processing rate (Table 1). This can be explained by the formation of an additional contact between the amino group of 2'-aminouridine and an amino acid inside IN. It may be an extra hydrogen bond or an electrostatic interaction that occurs if this amino group is at least partially protonated ( $pK_a = 6.2$  (33, 34)). We verified this by treating an oligonucleotide containing 2'-amino-

## New HIV DNA Analogs for Studying Interactions with Integrase

TABLE 1

The relative initial rates of 3'-end processing of the duplexes containing modifications at positions 7–9

The U5 substrate initial rate is considered to be equal to 1.

Position	Cofactor	Initial processing rate						
		U5B strand			U5A strand		Both strands	
		P	U <sup>N</sup>	U <sup>M</sup>	P	aA <sup>N</sup>	P	U <sup>N</sup> /aA <sup>N</sup>
7	Mg <sup>2+</sup>	1.10 ± 0.09	1.81 ± 0.21	0.91 ± 0.11	0.43 ± 0.06	0.73 ± 0.09	0.69 ± 0.08	1.58 ± 0.12
	Mn <sup>2+</sup>	0.83 ± 0.07	1.42 ± 0.12	0.82 ± 0.10	0.41 ± 0.05	0.57 ± 0.08	0.44 ± 0.06	1.24 ± 0.10
8	Mg <sup>2+</sup>	0.92 ± 0.09	— <sup>a</sup>	—	0.44 ± 0.06	—	0.79 ± 0.09	—
	Mn <sup>2+</sup>	0.93 ± 0.09	—	—	0.68 ± 0.08	—	0.81 ± 0.11	—
9	Mg <sup>2+</sup>	0.62 ± 0.08	0.94 ± 0.08	—	0.61 ± 0.07	0.79 ± 0.08	0.53 ± 0.07	0.67 ± 0.08
	Mn <sup>2+</sup>	0.59 ± 0.07	0.92 ± 0.10	—	0.62 ± 0.08	0.83 ± 0.10	0.48 ± 0.07	0.54 ± 0.07

<sup>a</sup> — indicates 3'-end processing assay was not carried out.

TABLE 2

The relative initial rates of 3'-end processing of the duplexes containing modifications at positions 5–6

The U5 substrate initial rate is considered to be equal to 1.

Position	Cofactor	Initial processing rate						
		U5B strand		U5A strand			Both strands	
		P	aA <sup>N</sup>	P	U <sup>N</sup>	U <sup>M</sup>	P	aA <sup>N</sup> /U <sup>N</sup>
5	Mg <sup>2+</sup>	0.10 ± 0.02	— <sup>a</sup>	0	—	—	0	—
	Mn <sup>2+</sup>	0.31 ± 0.04	—	0.11 ± 0.03	—	—	0.09 ± 0.01	—
6	Mg <sup>2+</sup>	0.21 ± 0.03	0.29 ± 0.04	0	0	0.10 ± 0.02	0	0
	Mn <sup>2+</sup>	0.34 ± 0.05	0.46 ± 0.07	0.18 ± 0.03	0.27 ± 0.04	0.18 ± 0.02	0.21 ± 0.04	0.19 ± 0.02

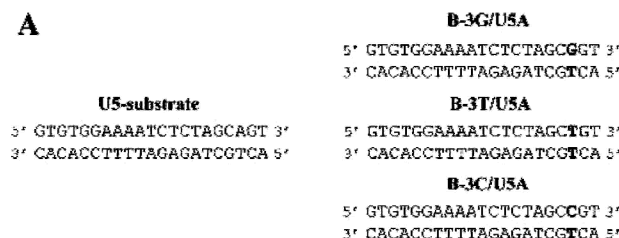
<sup>a</sup> — indicates 3'-end processing assay was not carried out.

uridine at position 7 (B-N7) with succinic anhydride, to give an oligonucleotide containing a carboxyl group (B-N7-ac) (Fig. 2A). We found that the 3'-processing rates of the duplex with this carboxyl group (B-N7-ac/U5A) and nonmodified U5 substrate were the same (Fig. 2B). Therefore, the increase in the 3'-processing rate of the substrate containing 2'-aminonucleoside at the processed strand position 7 is because of the presence of the 2'-amino group. This processing rate increase was also observed when the seventh nucleosides in both strands were modified. However, the effect was less, probably because of the opposition of two different effects: an increase in processing rate caused by the modification of the processed strand, and a decrease caused by the modification of the non-processed strand. Finally, for all modifications at positions 7–9, an equal effect on processing rate was observed in the presence of either Mn<sup>2+</sup> or Mg<sup>2+</sup> ions (Table 1).

**Modifications of Positions 5 and 6**—By contrast with the positions 7 and 8, modifications of the nucleosides in either strand at positions 5 and 6 affected the 3'-processing rate (Table 2). However, the rate decrease was much greater when the nonprocessed strand was modified. Replacement of native nucleosides by both the non-nucleoside insertion (P) and the modified nucleosides, 2'-aminonucleosides or 2'-O-methylnucleosides, resulted in a similar effect (Table 2). Therefore, IN was sensitive to any distortion of the double helix in this DNA region and not especially affected by heterocyclic base removing. The inhibition effect of the nucleoside modification was much greater in the presence of Mg<sup>2+</sup> than in the presence Mn<sup>2+</sup> ions (Table 2).

**Processing of Substrates Containing Modifications at Position 3**—Many studies have shown that adenine at position 3 is important for the efficient processing of HIV-1 DNA (1, 17, 20, 21, 24, 26). However, some modifications of the third nucleoside do not inhibit processing (25, 35). We investigated the interaction between IN and the adenine at position 3 by studying the 3'-end processing of U5 substrate analogs in which this adenine was replaced by guanine, thymine, or cytosine (Fig. 4A). All these duplexes had noncomplementary base pairs instead of A/T at position 3, because the wild-type oligonucleotide U5A was used as the unprocessed strand in all cases. In the presence of Mg<sup>2+</sup> ions, replacement of the conserved adenine by any other base decreased the proc-

A



B

duplex	U5-substrate	B-3G/U5A	B-3C/U5A	B-3T/U5A
<i>V<sub>o</sub> rel. (Mg<sup>2+</sup>)</i>	1.00	0.21 ± 0.05	0.19 ± 0.07	0.29 ± 0.06
<i>V<sub>o</sub> rel. (Mn<sup>2+</sup>)</i>	1.00	0.62 ± 0.07	0.51 ± 0.08	0.88 ± 0.07

FIGURE 4. Replacement of the adenine at position 3 by other heterocyclic bases. A, structures of U5 substrate analogs in which the adenine at position 3 was replaced by guanine, thymine, and cytosine. B, initial 3'-processing rates. The initial rate of U5 substrate 3'-end processing was set at 1.

essing rate by a factor of 3–4 (Fig. 4B). In the presence of Mn<sup>2+</sup> ions IN cleaved the mismatched duplexes (B-3G/U5A, B-3T/U5A, and B-3C/U5A) more efficiently than in presence of Mg<sup>2+</sup>, and 3'-end processing rates depended on the base replacing the adenine (Fig. 4B). The processing rate of the T-containing analog was similar to that of the wild-type substrate, whereas replacement of the A by a C or a G approximately halved the processing rate.

Replacing 2'-deoxyadenosine by 2'-aminonucleoside (aA<sup>N</sup>) at position 3 had no significant effect on processing rate in the presence of Mn<sup>2+</sup> ions (Table 3). In contrast, processing was almost completely inhibited in the presence of Mg<sup>2+</sup> ions (Fig. 3 and Table 3). Although 2'-aminonucleoside (aA<sup>N</sup>) contains adenine, it is not in the same position as within the wild-type substrate. IN may therefore recognize the adenine in the modified substrate only if it has a flexible structure. The active site of IN is thought to be more rigid in the presence of Mg<sup>2+</sup> than in the presence of Mn<sup>2+</sup> (1). Therefore, changing the position of adenine inhibits 3'-end processing only in the presence of Mg<sup>2+</sup> ions, because IN may adopt a new conformation in the presence of Mn<sup>2+</sup> ions, providing greater access to the modified substrate structure.

## New HIV DNA Analogs for Studying Interactions with Integrase

TABLE 3

The relative initial rates of 3'-end processing of the duplexes containing modifications at the 3rd position

The U5 substrate initial rate is considered to be equal to 1.

Position	Cofactor	Initial processing rate						
		U5B strand		U5A strand			Both strands	
		P	aA <sup>N</sup>	P	U <sup>N</sup>	U <sup>M</sup>	P	aA <sup>N</sup> /U <sup>N</sup>
3	Mg <sup>2+</sup>	0	0.10 ± 0.02	1.31 ± 0.14	1.25 ± 0.12	1.80 ± 0.19	0	0
	Mn <sup>2+</sup>	0	0.86 ± 0.08	1.28 ± 0.07	1.27 ± 0.09	1.27 ± 0.12	0	2.14 ± 0.11

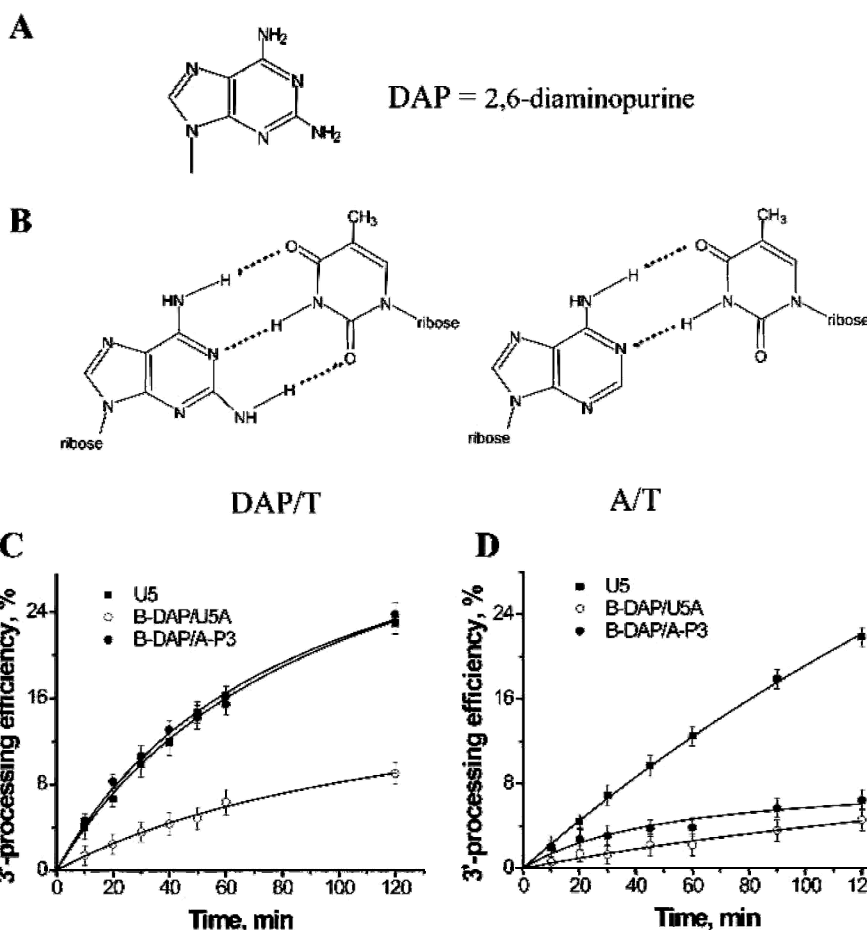


FIGURE 5. Effect on 3'-end processing of replacing the adenine at position 3 by DAP. *A*, structure of DAP. *B*, formation of Watson-Crick base pairs between thymine and adenine (A/T) and between thymine and 2,6-diaminopurine (DAP/T). Kinetics of 3'-end processing product accumulation in the case of U5 substrate and its analogs containing 2,6-diaminopurine in the U5B strand and the wild-type U5A strand (B-DAP/U5A) or the U5A strand with a non-nucleotide insertion at position 3 (B-DAP/A-P3), in the presence of Mn<sup>2+</sup> ions (*C*) and Mg<sup>2+</sup> ions (*D*) is shown.

Replacement of the 2'-deoxyadenosine at position 3 by 1,3-propanediol (P), resulting in the elimination of all functional groups of the adenine, completely inhibited 3'-end processing, regardless of the cofactor used (Table 3).

Replacement of the thymidine at position 3 in the unprocessed strand by 1,3-propanediol (P) or any modified nucleoside (U<sup>N</sup> or U<sup>M</sup>) increased the 3'-end processing rate (Table 3 and Fig. 3), regardless of the cofactor used. All these modifications at position 3 destabilized the DNA double helix (supplemental Table S1). This destabilization may stimulate 3'-end processing, consistent with previous data demonstrating an increase in 3'-processing activity because of destabilization of the 3rd bp (26, 35, 36).

The duplex with non-nucleoside insertions in both strands was not processed, regardless of the cofactor used (Table 3). In the presence of Mg<sup>2+</sup> ions, no 3'-processing was also observed for the duplex containing 2'-aminonucleosides in both strands, because of the inhibitory effect of the processed strand modification (Fig. 3 and Table 3). As replacing

the native A by aA<sup>N</sup> did not affect 3'-end processing in the presence of Mn<sup>2+</sup> ions, we observed an increase in reaction rate for the substrate analog containing two 2'-aminonucleosides in presence of this ion (Table 3). This effect may be due to the modification of both strands, further destabilizing the duplex.

As destabilization of the A/T base pair at position 3 stimulated 3'-end processing, we hypothesized that the stabilization of this base pair would decrease the processing rate. We tested this hypothesis by studying 3'-end processing of U5 substrate analogs containing 2,6-diaminopurine (DAP) instead of the adenine at position 3 of the U5B strand (Fig. 5A). The DAP base stabilizes the DNA structure by forming three hydrogen bonds with the complementary thymine, instead of two bonds for the normal Watson-Crick pair A/T (Fig. 5B) (37). Processing was strongly inhibited in the presence of either metallic cofactor following the incorporation of DAP into the U5B strand (duplex B-DAP/U5A) (Fig. 5, C and D). However, the 3'-end processing rate was increased by elimination of the hydrogen bonding between DAP and T because of

## New HIV DNA Analogs for Studying Interactions with Integrase

replacement of the T residue by 1,3-propanediol (duplex B-DAP/A-P3). Processing was as rapid as for the wild-type substrate in the presence of  $Mn^{2+}$  ions (Fig. 5C). A smaller increase in processing rate was observed in presence of  $Mg^{2+}$  ions (Fig. 5D).

We also studied interactions between IN and two 26-mer duplexes, because we assumed that A/T pair destabilization would enable IN to mediate cleavage of the U5 sequence in a more extended DNA duplex (Fig. 6). One duplex contained the sequence of the U5 substrate with 5 additional base pairs (26B/A). In the second duplex (26B/Amut), we replaced the thymine preceding the conserved adenine by an adenine, resulting in the noncomplementary pair A/A (Fig. 6A). The 3'-end proc-

essing efficiency of the 26B/A duplex was very low in the presence of  $Mn^{2+}$  ions (Fig. 6B), and no processing was observed in the presence of  $Mg^{2+}$  ions (Fig. 6C). In the case of the duplex containing the noncomplementary pair (26B/Amut), the 3'-processing efficiency increased in the presence of  $Mn^{2+}$  ions (Fig. 6B). Moreover, IN was able to cleave the mutated substrate even in presence of  $Mg^{2+}$  ions (Fig. 6C). Thus, A/T pair destabilization is essential for DNA cleavage by IN.

**Integrase Binding to Modified Substrate Analogs**—DNA substrate modification may influence the overall enzymatic reaction, either by affecting the substrate binding step or by influencing the catalytic activity of the enzyme. We therefore carried out electrophoretic mobility shift assays and steady-state fluorescence anisotropy experiments to test the effect of U5 substrate modifications on DNA binding by IN. We showed previously that the anisotropy parameter ( $r$ ) is particularly useful for studies of IN/DNA interactions using fluorescein-labeled DNA, with monitoring of the difference in  $r$  value between the IN-free and the IN-bound DNA (38). We studied IN binding to substrate analogs containing the modifications with the most pronounced effect (inhibition or stimulation) on 3'-end processing rate (Table 4). All experiments were carried out in the presence of  $Mg^{2+}$  ions.

We first studied the binding isotherms for DNA/IN interactions, for the modified U5 substrate analogs, by means of electrophoretic mobility shift assays (supplemental Fig. S1). IN bound equally well to all the modified duplexes tested. The binding curves for the modified substrates were similar to that for the control wild-type U5 substrate (Fig. 7). Thus, substrate modifications with major effects on 3'-end processing seemed to have no effect on substrate-IN affinity.

We then investigated IN/DNA interactions by steady-state fluorescence anisotropy (Table 4). The fluorescein-labeled wild-type DNA substrate was incubated in the presence of various concentrations of unlabeled modified DNA duplexes (Table 4) before the addition of IN. We calculated apparent  $K_d$  values from the competition curves for the modified DNA duplexes (Table 4). Consistent with the results of electrophoretic mobility shift assays, the apparent affinities of IN for the various modified duplexes were comparable with the one for wild-type substrate. A slight affinity increase was obtained for substrates containing the abasic site, which display no activity under magnesium conditions (U5B/A-P5 and B-P3/U5A  $K_d \approx 8.5$  nM, as compared with U5B/U5A  $K_d \approx 21$  nM). Moreover, other DNA substrates (U5B/A-N6 and B-N3/U5A) that display no 3'-processing activity have an apparent affinity for IN comparable with the wild-type substrate. Finally, processing of the modified duplex B-N7/U5A, which is characterized by the weakest affinity ( $K_d \approx 36$  nM), was stimulated (Table 1). Therefore, no

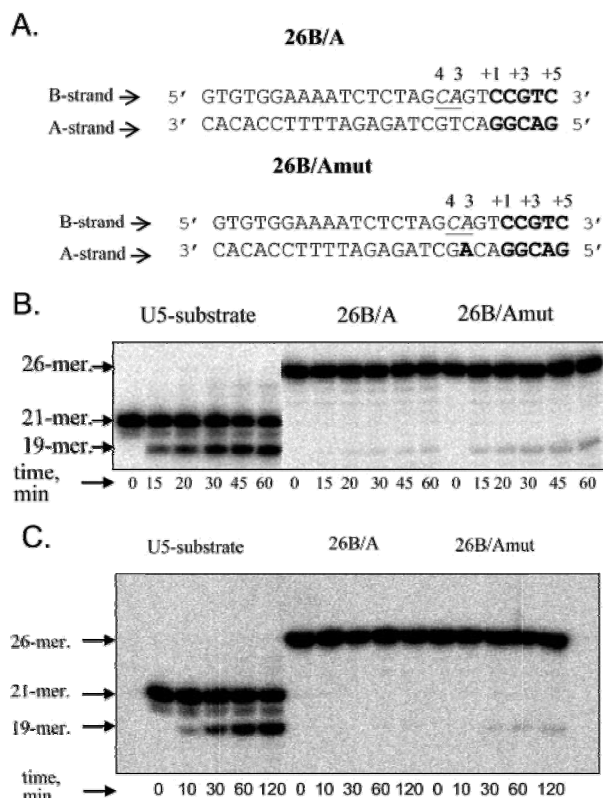


FIGURE 6. 3'-End processing of substrate analogs containing additional base pairs. A, structure of substrate analogs containing additional base pairs. Analysis of the processing products in the presence of  $Mn^{2+}$  ions (B) and  $Mg^{2+}$  ions (C) is shown.

TABLE 4

Study of the IN binding with U5 substrate and its modified analogs

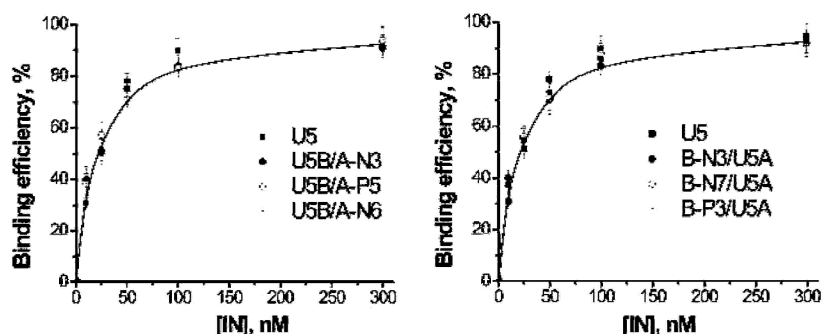
The type of the strand (A or B) modification is noted by N for 2'-aminonucleosides and P for 1,3-propanediol insertion; position of the modified nucleoside in the strand is noted by a corresponding figure.

DNA duplex	Sequence	$K_d'$ nM	$k_{on}^a$ $min^{-1}$
U5 substrate	5'-GTGTGGAAAATCTCTAGCAGT-3' U5B strand	21.0 ± 3.5	0.74 ± 0.04
U5B/U5A	3'-CACACCTTTTAGAGATCGTCA-5' U5A strand		
U5B/A-N3	GTGTGGAAAATCTCTAGCAGT CACACCTTTTAGAGATCGTCA	26 ± 3	0.89 ± 0.04
U5B/A-P5	GTGTGGAAAATCTCTAGCAGT CACACCTTTTAGAGATCGTCA	8.6 ± 0.4	1.07 ± 0.05
U5B/A-N6	GTGTGGAAAATCTCTAGCAGT CACACCTTTTAGAGATCGTCA	20 ± 2	1.07 ± 0.05
B-P3/U5A	GTGTGGAAAATCTCTAGCAGT CACACCTTTTAGAGATCGTCA	8.2 ± 0.8	1.00 ± 0.05
B-N3/U5A	GTGTGGAAAATCTCTAGCAGT CACACCTTTTAGAGATCGTCA	20 ± 3	0.71 ± 0.04
B-N7/U5A	GTGTGGAAAATCTCTAGCAGT CACACCTTTTAGAGATCGTCA	36 ± 5	1.13 ± 0.06

<sup>a</sup>  $k_{on}$  and  $K_d'$  (apparent  $K_d$ ) values were obtained from steady-state anisotropy measurements as indicated under "Experimental Procedures."

## New HIV DNA Analogs for Studying Interactions with Integrase

FIGURE 7. Binding of modified DNA substrates to IN (5 nM duplex, 10–300 nM integrase) in the presence of  $Mg^{2+}$  ions. Curves corresponding to the accumulation of IN-bound substrates were calculated using electrophoretic mobility shift assays data (supplemental Fig. S1). Structures of the modified substrates containing 2'-amino-nucleotides (N) or 1,3-propanediol insertion (P) are shown in the Table 4.



correlation between the apparent affinity and the specific activity was detected.

As we showed previously that IN-DNA complexes form fairly slowly (38, 39), this step might be differentially affected by DNA modifications, with effects on subsequent IN activity. We therefore investigated the effects of substrate modifications on the kinetics of DNA binding to IN (Table 4). Fluorescein was attached to the 5'-end of the U5B strand of the duplexes when the U5A strand was modified and was attached to the 3'-end of the U5A strand when the U5B strand was modified. The pseudo first-order rate constants ( $k_{on}$ ) were calculated for the various modified DNA duplexes (Table 4). Again, no significant difference was observed between modified and unmodified substrates, and no correlation between the effect of a given modification on DNA-binding rate and 3'-end processing rate was observed. Thus, the catalytic step is more sensitive to DNA modifications than the binding step. IN is known to have the intrinsic ability to bind target DNA in a nonspecific manner. This nonspecific binding may lead to underestimation, at least *in vitro*, of specific binding, minimizing the structural effects of DNA modifications on IN/DNA recognition. In the 3'-end processing assay, a smaller population of IN-DNA complexes is observed, the subpopulation corresponding to the correct and specific positioning of IN on its DNA substrate. Structural effects of DNA on this subpopulation may be more pronounced and easier to detect in activity assays than in DNA binding assays.

## DISCUSSION

In the absence of crystallographic or NMR structures for IN-DNA complexes, most of what is known about interactions between IN and its substrates is based on functional studies using modified DNA analogs. We tried to identify the elements of viral DNA structure important for IN binding and 3'-end processing. We first used a novel type of U5 substrate analog containing sugar-phosphate backbone modifications at various positions in one or both strands (Fig. 1). We investigated two different types of DNA modification: removal of the nucleoside by replacement with a 1,3-propanediol residue (P) or modification of the sugar conformation, changing the furanose 2'-position. It has been shown that the replacement of thymidines in the U5 substrate by 2'-deoxyuridines has no major influence on 3'-end processing (25), indicating that the methyl groups of the thymine residues in the U5 substrate are not involved in interactions with IN. The effect on the 3'-end processing rate of replacing T residues by 2'-modified uridines as demonstrated here therefore results from the unusual furanose structure rather than the elimination of methyl groups from heterocyclic bases.

It is important that all our modifications introduced a local distortion in the DNA double helix structure, decreasing  $T_m$  (supplemental Table S1). The instability caused by 2'-amino modification probably results from heterocyclic base destacking (27), causing base displacement in the double helix structure. In the case of  $U^M$ , the

destabilization may result from its C-3'-endo sugar puckering typical of the A-form geometry of RNA (32, 40). However, 2'-modified nucleosides retain their heterocyclic bases and may maintain at least some contacts with IN, whereas the incorporation of 1,3-propanediol results in the complete elimination of all functional groups that might interact with IN.

Most of the substrate positions involved in the specific activity of IN are located upstream from the 9th bp (20, 22, 25). However, nucleotides at positions 10–13 have also been shown to be important (1). We found that modifications at position 12 did not alter IN activity. We therefore focused on positions 3–9. We found that their modifications altered the 3'-end processing rate. As the conflicting results reported in previous publications may arise from the differential use of  $Mn^{2+}$  or  $Mg^{2+}$  ions, we also compared, for each modification, the 3'-end processing rate in the presence of these two cations.

We found that the catalytic step was sensitive to modifications to DNA-substrate structure but that the modifications had no effect on substrate binding to IN. In our hands, it was not possible to distinguish between specific and nonspecific binding of IN to DNA. The  $K_d$  value derived from binding to the U5 sequence was very close to the one obtained with nonspecific double-stranded DNA (25 nM; data not shown). Under our enzymatic conditions ( $E > S$ ), IN activity was correlated to the amount of IN-DNA complexes. It was therefore crucial to verify whether the substrate modifications did not alter the quantity of complexes formed *in vitro*. No modification of IN binding was observed, thus indicating that the alteration of the catalysis was related to a post-binding event, most likely a modification of the DNA conformation, which prevented the cleavage to occur.

*Integrase Interacts with the DNA Sugar-Phosphate Backbone at Positions 5–9*—We identified three different sites, between nucleosides 5 and 9 inclusive, where modification affected 3'-end processing rate in different ways.

All modifications at position 9 affected processing rate, but this effect was moderate in all cases (Table 1). The strongest effect was obtained by replacing the base pair by 1,3-propanediol (P). However, despite the P insertion causing the largest DNA distortion (the  $T_m$  shift was 8.3 °C) because of the disruption of both hydrogen bonding and heterocyclic base stacking, the processing rate was no more than halved. This suggests that IN interacts with the sugar-phosphate backbone, without making specific contacts with the bases. Moreover, single modifications on either strand affected IN activity equally, suggesting that this interaction likely involves both strands.

At positions 7 and 8, the total removal of heterocyclic bases from the processed strand had no significant effect on 3'-end processing rate (Table 1, modification P), ruling out the possibility of specific contact with heterocyclic bases. By contrast, similar modifications of the unprocessed strand decreased IN activity (Table 1, modification P) thus allowing specific interactions with IN for heterocyclic bases from this strand. It has been

## New HIV DNA Analogs for Studying Interactions with Integrase

shown, however, that substitutions at positions 7 and 8 with other complementary base pairs, which do not disturb the DNA double helix structure but change the pattern of putative bases contacts with IN, do not affect 3'-end processing (1). Thus, the inhibitory effect of P insertion results from the loss of IN contacts with the sugar-phosphate backbone because of its local distortion rather than a loss of contacts with heterocyclic bases.

Any modification of the fifth and sixth native nucleosides, either by P insertion or by replacement with 2'-modified nucleosides, led to a large decrease in the 3'-end processing rate (Table 2). Similar levels of inhibition were observed with all substrate modifications at position 6, suggesting that IN recognizes the conformation of the DNA backbone rather than nucleoside bases. Previous studies have shown that replacing the guanine at position 5 with  $O^6$ -methylguanine decreases substrate cleavage (26). It has been suggested that IN interacts with this heterocyclic base via the major groove of the DNA. However, the 5th and 6th bp are inverted in the U5 and U3 extremities (5'-A(G/C)T-3' in U5 and 5'-T(C/G)A-3' in U3), although both sequences are substrates of the enzyme. This observation is not consistent with the direct recognition of the bases at these positions by IN. Moreover, the effect of  $O^6$ -methylguanine may also result from the DNA structure distortion, because guanine  $O^6$ -methylation modifies the conformation of the phosphodiester backbone at the modification site (41). It therefore seems more likely that IN recognizes a DNA fine structure generated by the G/C and A/T pairs at positions 5 and 6, respectively. Deformation of the double helix by either base replacement or nucleoside modification prevents DNA recognition, thereby strongly inhibiting 3'-end processing (Table 2).

We conclude that IN recognition involves only interactions with the DNA sugar-phosphate backbone at positions 5–9. Both nucleosides of the 9th bp participate in these interactions. For positions 5–8, modifications of the unprocessed strand decrease 3'-end processing rate more strongly than modifications of the processed strand (Tables 1 and 2), indicating that active IN-DNA complex formation primarily requires interactions with the unprocessed strand. Nevertheless, at positions 5 and 6, contact between IN and both strands is important for specific DNA recognition (Table 2).

**Destabilization of the A/T Pair at Position 3 Is Necessary for 3'-End Processing**—Modification of the thymidine at position 3 in the unprocessed strand, by P insertion or 2'-modified nucleosides ( $U^N$  and  $U^M$ ), increased 3'-end processing rate (Table 3) probably because of local destabilization of the A/T base pair, as suggested previously (36). The processing stimulation following the thymidine replacement likely points to an absence of specific contact between IN and this base. Alternatively, one could consider that mispairing may result in a loss of a specific contact otherwise required for the unpairing step. However it must be noted that according to Ref. 42, binding of the avian homolog of HIV integrase to its DNA substrate enhances the accessibility and the subsequent modification of the third thymidine by  $KMnO_4$ . Thus, although integrase disrupts the A/T base pair, it does not protect the heterocyclic base from the modification, thereby ruling out a close contact between the base and the enzyme. The stimulation caused by A/T base pair destabilization is also consistent with the inhibition elicited by replacing the adenine at position 3 with DAP, a base known to stabilize the double helix as a result of the formation of three hydrogen bonds (Fig. 5B) (37). Accordingly, the disruption of these hydrogen bonds by replacing thymidine in the DAP/T base pair by 1,3-propanediol (duplex B-DAP/A-P3) led to an increase in 3'-end processing rate (Fig. 5, C and D).

The negative effect of DAP substitution on 3'-processing has been reported before, and it has been suggested that DAP disrupts a specific contact between IN and the C-2 position of the adenine at position 3

(26). However, such a contact is not consistent with the efficient cleavage of the mismatched substrate B-3T/U5A (Fig. 4) or the stimulation observed if the thymidine complementary to DAP is replaced by 1,3-propanediol (Fig. 5C). We therefore suggest that DAP prevents the local A/T pair destabilization that is essential for efficient processing. The necessity of this destabilization also found for avian sarcoma virus integrase (42) may account for the strict conservation of the CA/TG sequence in integrase substrates, as intrinsic conformational mobility is known to favor DNA flexibility (43–47). Such a local destabilization allows IN activity, even at an internal position, as seen in the internal processing of a 26-mer duplex (26B/A). As predicted, processing efficiency increased considerably when the A/T base pair was replaced by a noncomplementary A/A pair (duplex 26B/Amut) (Fig. 6).

**Specific and Metal-dependent Interactions with the Adenosine at Position 3 in the Processed Strand**—It has been reported that replacement of the A/T base pair at position 3 by other complementary pairs (1, 35) or replacement of the adenine in this position with  $N^7$ -deaza-adenine (26) decreases 3'-end processing efficiency, suggesting that specific contacts with this adenine are required. However, other studies have shown that replacement of the A/T pair by either noncomplementary base pairs (35) or an abasic site (AP-site) (25) does not affect 3'-end processing efficiency. Our study reconciles these contradictory data by demonstrating the important role played by the metal cofactor. Indeed, we observed that the replacement of 2'-deoxyadenosine by P insertion completely inhibited 3'-processing in the presence of both  $Mg^{2+}$  and  $Mn^{2+}$  (Table 3). By contrast, the modification of this residue by 2'-aminonucleoside,  $aA^N$ , yielded a metal-dependent effect; 3'-end processing was completely inhibited in the presence of  $Mg^{2+}$  ions, whereas no inhibition was observed in the presence of  $Mn^{2+}$  (Table 3). It has been suggested that  $Mg^{2+}$  coordination makes the active site of IN more rigid, whereas  $Mn^{2+}$  coordination does not (1, 4). The loss of flexibility caused by  $Mg^{2+}$  coordination may prevent IN from establishing specific contacts with the adenine within  $aA^N$ , as the sugar modification changes the location of the adenine. As a result, 3'-processing of the  $aA^N$ -containing substrate is inhibited in the presence of  $Mg^{2+}$  but unaffected in the presence of  $Mn^{2+}$ , which maintains structural flexibility of IN.

We investigated possible specific contacts between IN and the adenine at position 3 by studying the processing of substrates in which this adenosine was replaced by other nucleosides (Fig. 4). Adenine, which is simultaneously involved in Watson-Crick base pairing, can form hydrogen bonds with IN via its N-7 atom (proton acceptor) and the *exo*-cyclic amino group (proton donor) in the major groove and via the N-3 atom (proton acceptor) in the minor groove. A specific contact between IN and the N-7 atom has been suggested, based on the decrease in processing efficiency by a factor of 3 caused by replacing the adenine at position 3 with  $N^7$ -deaza-adenine, which cannot maintain this contact (26). The low processing rate observed for the C-containing duplex (B-3C/U5A) (Fig. 4B), for which a contact similar to that formed by N-7 is impossible, confirmed the necessity of this contact. However, the strong decrease in processing rate for the B-3G/U5A duplex (Fig. 4B), despite the presence of the same proton acceptors in guanine and adenine (N-3 and N-7 atoms), indicates that the  $NH_2$  group of adenine is also involved in an interaction with IN. The suggestion that two contacts are formed between the adenine and IN is supported by the finding that the T-containing duplex (B-3T/U5A) was processed almost as efficiently as the wild-type substrate (Fig. 4B). Indeed, thymine, which is not involved in Watson-Crick base pairing in this duplex, may form two hydrogen bonds with IN, involving O-4 as the proton acceptor (similar to N-7 of adenine) and the  $NH$  group as the proton donor. As the geometry of the bonds formed by thymine is not identical to that of the bonds



## New HIV DNA Analogs for Studying Interactions with Integrase

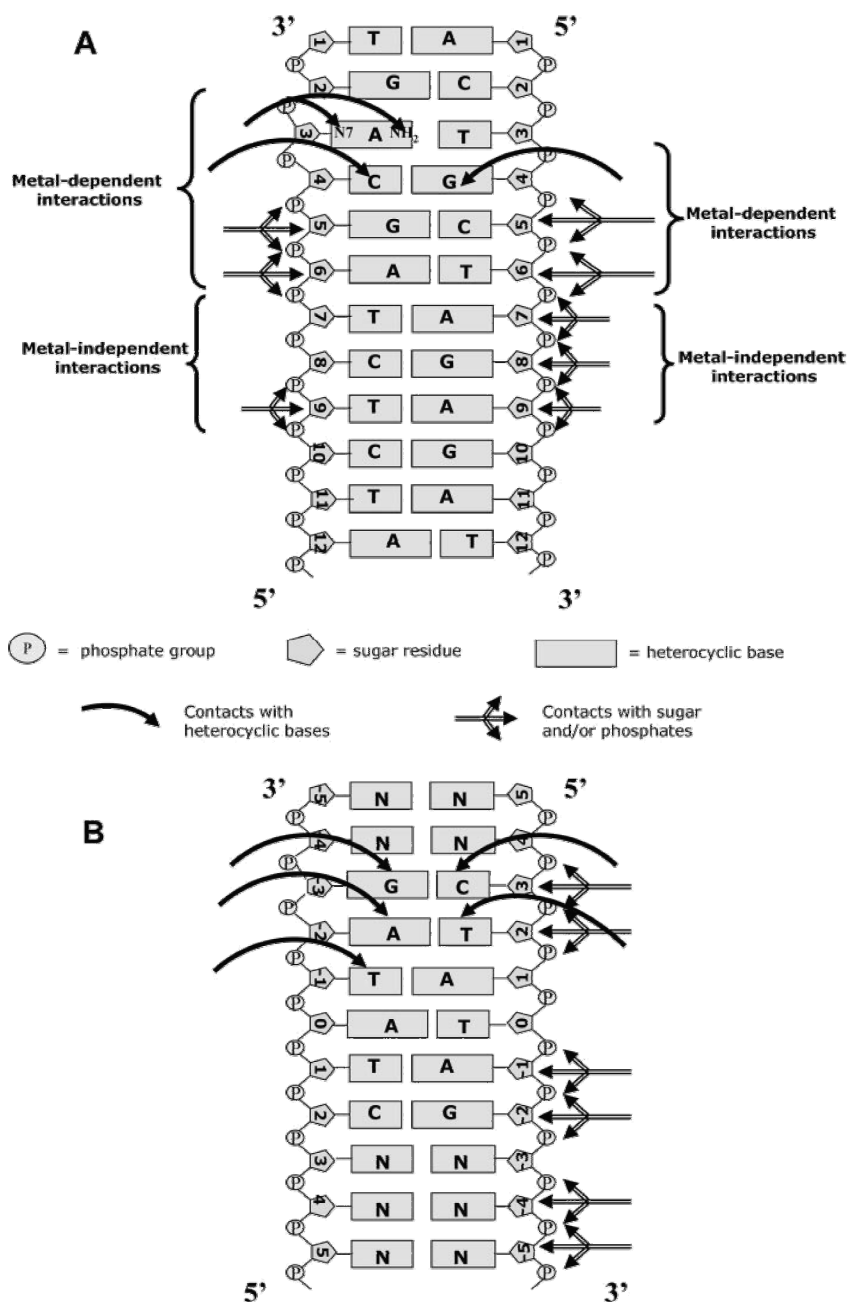


FIGURE 8. Scheme of contacts formed between HIV-1 IN and the U5 substrate proposed in this study (A) and EcoRV restriction endonuclease and its DNA substrate described in Ref. 51 (B). Contacts of IN with heterocyclic bases at position 4 of the U5 substrate have been described elsewhere (1, 20, 21). The size of the arrows corresponds to the importance of contacts for 3'-end processing rate.

formed by adenine, these bonds can be generated only if the structure of IN structure is flexible enough to accept T-containing duplexes as a substrate, *i.e.* in presence of  $Mn^{2+}$  ions. Clearly, these contacts cannot be formed if the thymine participates in base pairing, and as result, the 3'-end processing of a substrate analog containing the T/A pair instead of A/T is inefficient (35).

Metal-dependent sensitivity of IN to substrate structure was also observed when positions 5 and 6 were modified (Table 2), although this sensitivity was less marked than for modifications at position 3. As the cation present primarily affects the conformation of the core domain (4–6), metal-dependent sensitivity indicates possible interactions with this domain. We therefore suggest that nucleosides at positions 3–6 within the DNA substrate interact with the catalytic core. By

contrast, the C-terminal domain is not directly involved in metal ion binding, and the structure of this domain therefore depends less on the metal present than does that of the catalytic core. Consequently, modifications of nucleosides interacting with the C-terminal domain should affect 3'-end processing regardless of the cofactor used. This was indeed found to be the case for nucleosides at positions 7–9 (Tables 1), consistent with known interaction of the C-terminal domain with the thymidine at position 7 (1).

**Summary of IN/DNA Interactions**—The interactions between IN and DNA substrate identified in this study are summarized in Fig. 8A. This scheme provides new insight into mechanistic similarities between IN and restriction endonucleases. Retroviral integrases and type II restriction enzymes have been reported to share conserved structural features

## New HIV DNA Analogs for Studying Interactions with Integrase

(48). Thus, in the archetypal EcoRV/DNA interaction model, the protein contacts the base just upstream from the scissile bond but does not contact the complementary base in the opposite strand (Fig. 8B). Similarly, IN interacts with the adenine at position 3, but not its complementary thymine (Fig. 8A). Position -2 in the EcoRV recognition sequence, which is similar position 4 in the HIV-1 DNA, displays the highest level of contact within the EcoRV-DNA complex, with contacts established on both strands. It has been suggested that heterocyclic bases from the 4th bp interact with IN (1, 20, 21). Moreover, in addition to base-specific contacts, numerous contacts of EcoRV with the DNA backbone allow recognition of the cognate site through a sequence-dependent backbone conformation (49, 50). Regarding HIV-1 IN, it was found that specific protein mutations altering ionic interactions or hydrogen bonding result in a change of DNA recognition specificity when the mutated amino acids are located close to DNA positions 5–11 (51). These data support our conclusion that IN recognizes the fine conformation of DNA backbone at positions 5–9 rather than nucleoside bases.

Finally, EcoRV endonuclease initially binds DNA nonspecifically, subsequently forming a specific complex as a result of conformational changes in both the DNA and the enzyme (49). The DNA is locally distorted from a regular B form to a highly strained kinked conformation, leading to unwinding of the DNA, unstacking of the two central base pairs of the recognition site, and bending of the DNA (50). For IN, this bending and unstacking may be favored by the conformational mobility of the CA/TG fragment (43–44) and a loss of regular base stacking (45).

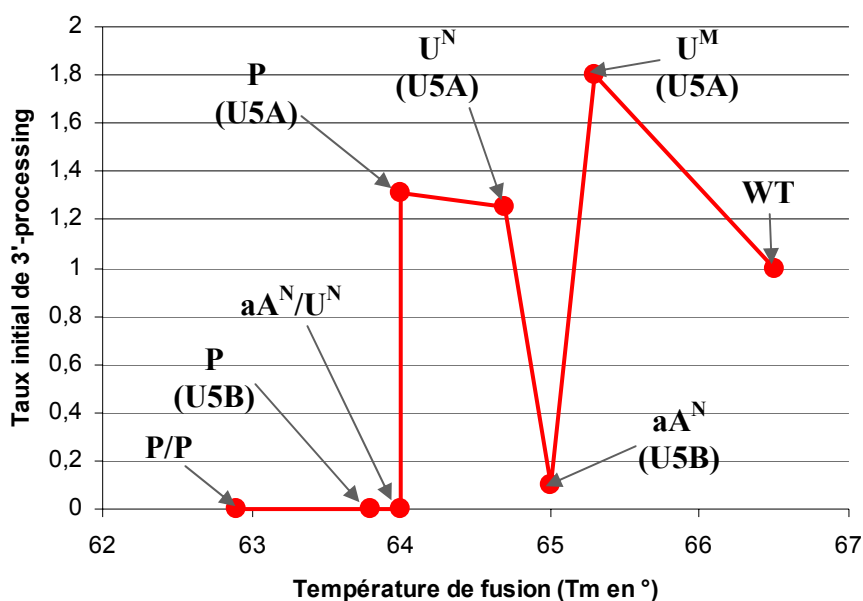
*Acknowledgment*—We thank Françoise Simon for technical assistance.

## REFERENCES

- Esposito, D., and Craigie, R. (1998) *EMBO J.* **17**, 5832–5843
- Engelman, A., and Craigie, R. (1995) *J. Virol.* **69**, 5908–5911
- Marchand, C., Johnson, A. A., Karki, R. G., Pais, G. C., Zhang, X., Cowansage, K., Patel, T. A., Nicklaus, M. C., Burke, T. R., Jr., and Pommier, Y. (2003) *Mol. Pharmacol.* **64**, 600–609
- Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R., and Davies, D. R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9150–9154
- Asante-Appiah, E., Seeholzer, S. H., and Skalka, A. M. (1998) *J. Biol. Chem.* **273**, 35078–35087
- Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clement-Mella, C., and Mikol, V. (1998) *J. Mol. Biol.* **282**, 359–368
- Polard, P., and Chandler, M. (1995) *Mol. Microbiol.* **15**, 13–23
- Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P., and Skalka, A. M. (1992) *Mol. Cell. Biol.* **12**, 2331–2338
- Drelich, M., Wilhelm, R., and Mous, J. (1992) *Virology* **188**, 459–468
- Gao, K., Wong, S., and Bushman, F. (2004) *J. Virol.* **78**, 6715–6722
- Asante-Appiah, E., and Skalka, A. M. (1997) *J. Biol. Chem.* **272**, 16196–16205
- Neamati, N., Lin, Z., Karki, R. G., Orr, A., Cowansage, K., Strumberg, D., Pais, G. C., Voigt, J. H., Nicklaus, M. C., Winslow, H. E., Zhao, H., Turpin, J. A., Yi, J., Skalka, A. M., Burke, T. R., Jr., and Pommier, Y. (2002) *J. Med. Chem.* **45**, 5661–5670
- Katzman, M., and Sudol, M. (1996) *J. Virol.* **70**, 2598–2604
- Chen, J. C., Krucinski, J., Miercke, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8233–8238
- Esposito, D., and Craigie, R. (1999) *Adv. Virus Res.* **52**, 319–333
- Wang, J. Y., Ling, H., Yang, W., and Craigie, R. (2001) *EMBO J.* **20**, 7333–7343
- Jenkins, T. M., Esposito, D., Engelman, A., and Craigie, R. (1997) *EMBO J.* **16**, 6849–6859
- Gerton, J. L., Ohgi, S., Olsen, M., DeRisi, J., and Brown, P. O. (1998) *J. Virol.* **72**, 5046–5055
- Gao, K., Butler, S. L., and Bushman, F. (2001) *EMBO J.* **20**, 3565–3576
- LaFemina, R. L., Callahan, P. L., and Cordingley, M. G. (1991) *J. Virol.* **65**, 5624–5630
- Sherman, P. A., Dickson, M. L., and Fyfe, J. A. (1992) *J. Virol.* **66**, 3593–3601
- Katzman, M., and Sudol, M. (1996) *J. Virol.* **70**, 9069–9073
- Mazumder, A., Neamati, N., Pilon, A. A., Sunder, S., and Pommier, Y. (1996) *J. Biol. Chem.* **271**, 27330–27338
- van den Ent, F. M. I., Vink, C., and Plasterk, R. H. A. (1994) *J. Virol.* **68**, 7825–7832
- Mazumder, A., and Pommier, Y. (1995) *Nucleic Acids Res.* **23**, 2865–2871
- Wang, T., Balakrishnan, M., and Jonsson, C. B. (1999) *Biochemistry* **38**, 3624–3632
- Zubin, E. M., Antsyovich, S. I., Oretskaya, T. S., Romanova, E. A., Volkov, E. M., Tashlitsky, V. N., Dolinnaya, N. G., and Shabarova, Z. A. (1998) *Nucleosides Nucleotides* **17**, 425–440
- Kuznetsova, L. G., Volkov, E. M., Romanova, E. A., Tashlitsky, V. N., Oretskaya, T. S., and Shabarova, Z. A. (1991) *Bioorg. Khim.* **17**, 1289–1291
- Leh, H., Brodin, P., Bischerour, J., Deprez, E., Tauc, P., Brochon, J. C., LeCam, E., Coulaud, D., Auclair, C., and Mouscadet, J. F. (2000) *Biochemistry* **39**, 9285–9294
- Gelfand, C. A., Plum, G. E., Grollman, A. P., Johnson, F., and Breslauer, K. J. (1998) *Biochemistry* **37**, 7321–7327
- Hendrix, C., Devreese, B., Rozenski, J., Van Aerschoot, A., De Bruyn, A., Van Seeumen, J., and Herdewijn, P. (1995) *Nucleic Acids Res.* **23**, 51–57
- Lubini, P., Zurcher, W., and Egli, M. (1994) *Chem. Biol.* **1**, 39–45
- Aurup, H., Tuschl, T., Bueseler, F., Ludwig, J., and Eckstein, F. (1994) *Nucleic Acids Res.* **22**, 20–24
- Kan, L.-S., Bauman, R., Bhan, P., and Miller, P. S. (1993) *J. Chin. Chem. Soc.* **40**, 469–473
- Scottoline, B. P., Chow, S., Ellison, V., and Brown, P. O. (1997) *Genes Dev.* **11**, 371–382
- Agapkina, J., Smolov, M., Zubin, E., Mouscadet, J.-F., and Gottikh, M. (2004) *Eur. J. Biochem.* **271**, 205–211
- Bailey, C., and Waring, M. J. (1998) *Nucleic Acids Res.* **26**, 4309–4314
- Deprez, E., Barbe, S., Kolaski, M., Leh, H., Zouhiri, F., Auclair, C., Brochon, J.-C., Le Bret, M., and Mouscadet, J.-F. (2004) *Mol. Pharmacol.* **65**, 85–98
- Brodin, P., Pinskaya, M., Buckle, M., Parsch, U., Romanova, E., Engels, J., Gottikh, M., and Mouscadet, J.-F. (2002) *Biochemistry* **41**, 1529–1538
- Venkateswarlu, D., Lind, K. E., Mohan, V., Manoharan, M., and Ferguson, D. M. (1999) *Nucleic Acids Res.* **27**, 2189–2195
- Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., and Jones, R. A. (1986) *Biochemistry* **25**, 1027–1036
- Katz, R. A., DiCandeloro, P., Kukolj, G., and Skalka, A. M. (2001) *J. Biol. Chem.* **276**, 34213–34220
- Dickerson, R. E. (1998) *Nucleic Acids Res.* **26**, 1906–1926
- Packer, M. J., Dauncey, M. P., and Hunter, C. A. (2000) *J. Mol. Biol.* **295**, 71–83
- Renisio, J. G., Cosquer, S., Cherrak, I., Antri, S., Mauffret, O., and Fermandjian, S. (2005) *Nucleic Acids Res.* **33**, 1970–1981
- Dickerson, R. E., and Chiu, T. K. (1997) *Biopolymers* **44**, 361–403
- Jones, S., van Heyningen, P., Berman, H. M., and Thornton, J. M. (1999) *J. Mol. Biol.* **287**, 877–896
- Venclovas, C., and Siksnys, V. (1995) *Nat. Struct. Biol.* **2**, 838–841
- Pingoud, A., and Jeltsch, A. (2001) *Nucleic Acids Res.* **29**, 3705–3727
- Pingoud, A., and Jeltsch, A. (1997) *Eur. J. Biochem.* **246**, 1–22
- Chen, A., Weber, I. T., Harrison, R., and Leis, J. (2006) *J. Biol. Chem.* **281**, 4173–4182

## Modifications de l'ADN viral en position 3 : conséquences sur sa conformation locale

Les modifications étudiées font décroître la température de fusion de l'oligonucléotide : elles déstabilisent donc la double hélice d'ADN. Mais il n'y a pas de relation entre le  $T_m$  et le taux initial de la réaction de 3'-processing (**figure 2**). Il est donc nécessaire de faire une étude plus fine.

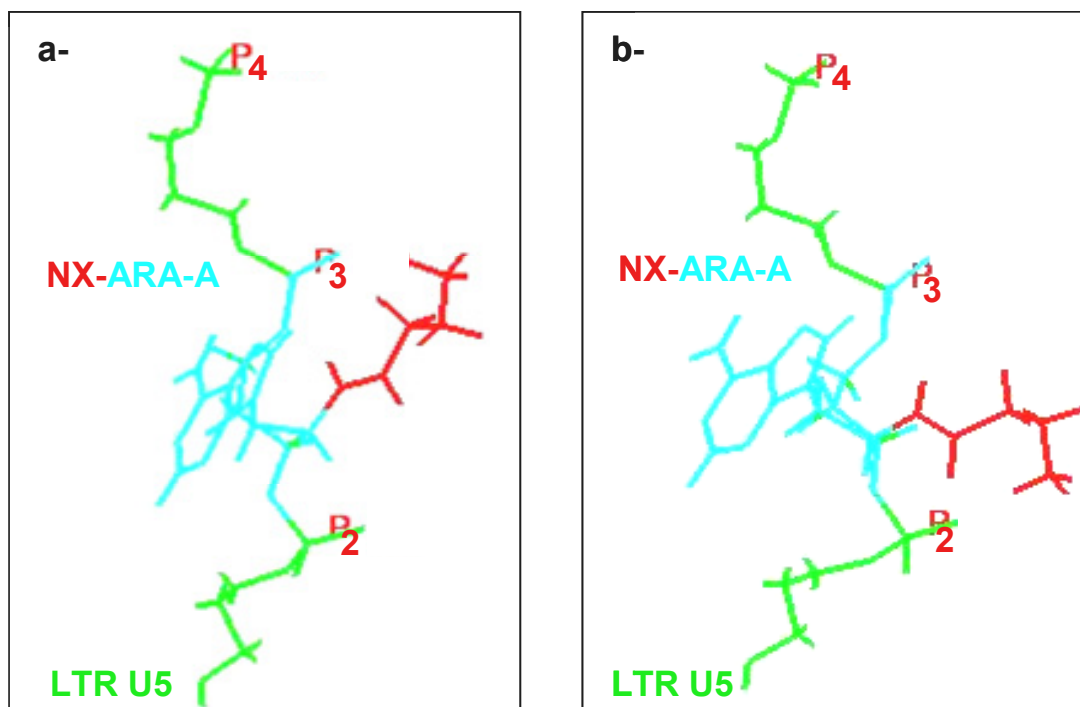


**Figure 2-** Le taux initial de 3'-processing ( $Mg^{2+}$ ) d'analogues de l'ADN viral (modifiés en position 3) en fonction de la température de fusion de ces analogues. **WT** : oligonucléotide sauvage ; **P/P** : A3 et T3 remplacés par P ; **P (U5B)** : A3 remplacé par P ; **P (U5A)** : T3 remplacé par P ; **aA<sup>N</sup> (U5B)** : A3 remplacé par aA<sup>N</sup> ; **U<sup>N</sup> (U5A)** : T3 remplacé par P ; **U<sup>M</sup> (U5A)** : T3 remplacé par U<sup>M</sup> ; **aA<sup>N</sup>/U<sup>N</sup>** : A3 et T3 remplacés par aA<sup>N</sup> et U<sup>N</sup>.

### Conformation d'aA<sup>N</sup>

Nous avons réalisé une étude structurale d'aA<sup>N</sup> (NX-ARA-A). Nous avons tout d'abord déterminé la conformation de la chaîne NX (le 2'-(3-amino-propionyl)amino) lorsque l'analogue de nucléotide est introduit à la place de la désoxyadenosine en position 3, dans un oligonucléotide de 12 paires de bases mimant l'extrémité LTR U5 de l'ADN viral et selon que son plissement soit Nord ou Sud. On a attribué 6 valeurs également espacées de 0° à 360° à chacun des trois angles dièdres libres de la chaîne NX. Par contre, la liaison peptidique a été conservée rigide en *trans*. Les 216 structures ont alors été optimisées avec un minimiseur quasi-newtonien et en utilisant le champ de force AMBER. Dans la meilleure conformation Nord obtenue (**figure 3-b**), la chaîne NX se trouve au-dessus de la liaison phosphodiester O3'-P et forme une liaison hydrogène avec le phosphate en 2. Une liaison hydrogène intra-

chaîne entre le CO et le NH<sub>3</sub><sup>+</sup> stabilise la conformation de la chaîne NX. Dans la conformation Sud la plus favorable (**figure 3-a**), la chaîne recouvre la liaison P-O5' et forme une liaison hydrogène avec le phosphate en 3.



**Figure 3-** Conformations Sud (**a**) et Nord (**b**) les plus favorables d'aA<sup>N</sup> (NX-ARA-A) inséré à la place d'A3 dans un dodécanuléotide mimant l'extrémité LTR U5 de l'ADN viral. Conformations obtenues par optimisation géométrique avec le minimiseur AMBER.

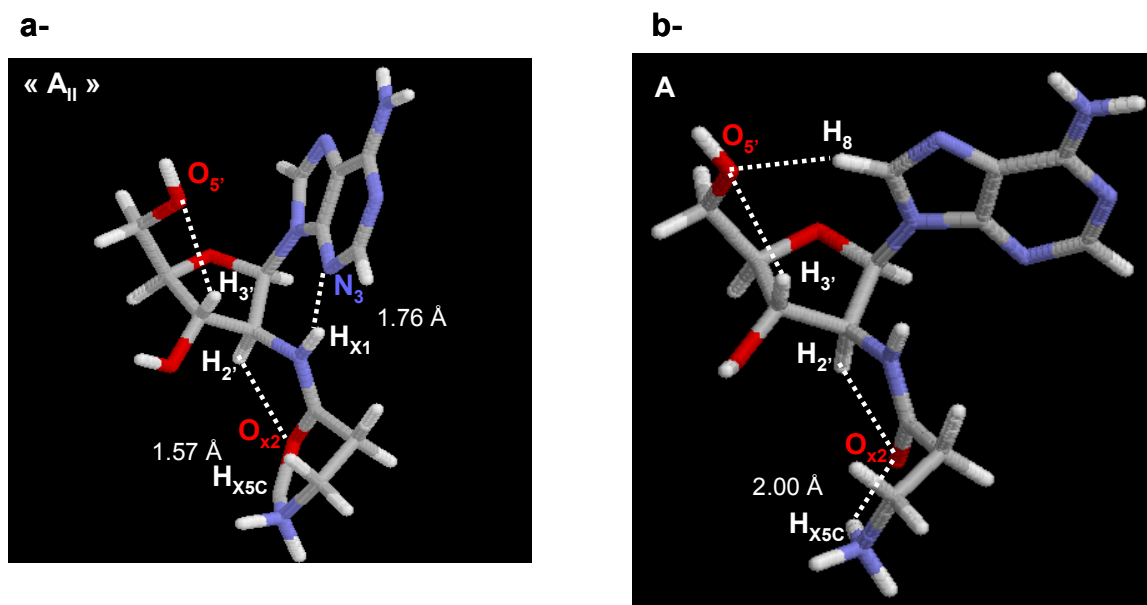
Nous avons ensuite effectué des calculs de mécanique quantique comme décrit dans le chapitre I afin de déterminer le plissement du sucre le plus favorable. Dans les plissements Sud et Nord, la chaîne NX a été positionnée et maintenue comme dans les conformations Sud et Nord obtenues lors des minimisations avec AMBER. Dans le vide, une conformation ayant un plissement Nord en C2' exo (au lieu de C3' endo) et un angle glycosidique  $\chi$  en *highsyn* (au lieu d'*anti*) est très favorisée (**figure 4-a**). Ces déformations sont dues à la très forte attraction du proton du NH de la chaîne par le N3 de l'adénine. De plus, cette conformation est stabilisée par une liaison intra-chaîne (entre le CO et le NH<sub>3</sub><sup>+</sup>) absente dans les conformations Sud. Nous avons donc refait les calculs en solvant implicite pour atténuer l'effet de ces liaisons. Dans ce cas, une conformation Nord normale (forme A) est favorisée (**figure 4-b**) (**tableau 2**). Le NH de la chaîne n'interagit plus avec le N3 de la base. La liaison hydrogène intra-chaîne est toujours présente dans cette conformation mais elle est plus faible. Pour vérifier si la conformation Sud n'est pas seulement défavorisée en raison de l'absence de la liaison intra-chaîne, nous avons refait des calculs avec des plissements Sud initiaux dans

lesquels la conformation de la chaîne NX est stabilisée par la liaison hydrogène intra-chaîne. Malgré la présence de cette liaison, les plissements Sud sont toujours défavorisés *in vacuo* et en solvant implicite (**tableau 3**). L'analogue de nucléoside,  $aA^N$ , est en Nord et il est rigide ; il ne peut pas adopter une conformation Sud.

**Tableau 2-** Angles ( $P$ ,  $\varepsilon$ , et  $\chi$  en degré) et écarts d'énergie potentielle (en kcal/mol) des conformations d' $aA^N$  obtenues par minimisations d'énergie avec Gaussian 03, en solvant implicite (niveau théorique B3LYP/CPCM /6-31G(d)).

P	$\varepsilon$	$\chi$	$\Delta E$	Formes
9.10	-91.46	-142.05	-1.95	A <sub>II</sub>
8.20	57.88	-142.81	-2.08	A <sub>III</sub>
7.46	-148.50	-144.37	-2.60	A
163.41	-60.06	-129.45	0.33	B <sub>II</sub>
155.83	48.61	-129.74	1.78	B <sub>III</sub>
155.36	-178.51	-130.81	0.00	B <sub>I</sub>

Dans les plissements Nord et Sud initiaux, la chaîne NX a été mise comme dans les conformations Sud et Nord obtenues lors des minimisations avec AMBER. La conformation de la chaîne NX a été maintenue fixe durant ces optimisations géométriques avec Gaussian 03.



**Figure 4-** **a-** Conformation « A<sub>II</sub> » d' $aA^N$  obtenue par minimisation d'énergie avec Gaussian 03 dans le vide. **b-** Conformation « A » d' $aA^N$  obtenue par minimisation d'énergie avec Gaussian 03 en solvant implicite.

**Tableau 3-** Angles ( $P$ ,  $\varepsilon$ , et  $\chi$  en degré) et écarts d'énergie potentielle (en kcal/mol) des conformations d'aA<sup>N</sup> obtenues par minimisations d'énergie avec Gaussian 03 *in vacuo* et en solvant implicite.

Méthode	P	$\varepsilon$	$\chi$	$\Delta E$	Formes
B3LYP/6-31G(d)	-26.32	-75.08	117.93	-5.48	“A <sub>II</sub> ”
	0.99	53.78	-158.46	-0.65	A <sub>III</sub>
	157.32	-58.84	-164.44	0.67	B <sub>II</sub>
	155.87	50.42	-161.57	1.17	B <sub>III</sub>
	142.71	179.46	-171.71	0.00	B <sub>I</sub>
B3LYP/CPCM /6-31G(d)	9.10	-91.46	-142.05	-0.79	A <sub>II</sub>
	8.20	57.88	-142.81	-1.07	A <sub>III</sub>
	7.46	-148.50	-144.37	-1.30	A
	174.40	-64.83	-121.00	0.58	B <sub>II</sub>
	175.72	44.03	-121.06	1.80	B <sub>III</sub>
	172.49	-179.87	-123.09	0.00	B <sub>I</sub>

Dans les plissements Nord et Sud initiaux, la chaîne NX a été mise comme dans la conformation Nord obtenue lors des minimisations avec AMBER. La conformation de la chaîne NX a été maintenue fixe durant ces optimisations géométriques avec Gaussian 03.

### Conformation d'U<sup>N</sup> et U<sup>M</sup>

Nous avons déterminé, lors des études structurales *ab initio*, présentées dans les chapitres I et II que les analogues de nucléosides U<sup>N</sup> et U<sup>M</sup> ont un plissement Nord et sont flexibles.

### Conséquences de l'insertion des analogues de nucléotides en position 3

Les remplacements de la thymidine, en position 3 (T3) par U<sup>N</sup> et U<sup>M</sup> entraînent une augmentation du taux initial de 3'-processing quel que soit le cofacteur métallique présent. Ainsi, si U<sup>N</sup> et U<sup>M</sup> ont un plissement Nord lorsqu'ils remplacent T3 de l'ADN viral, cette conformation est favorable à la réaction de 3'-processing.

Lorsqu'A3 est remplacé par aA<sup>N</sup> en Nord, la chaîne NX est au dessus de la liaison scissible par l'IN et le 3'-processing est inhibé en présence de Mg<sup>2+</sup>. Par contre cette modification est sans effet significatif sur le taux de 3'-processing, en présence de Mn<sup>2+</sup>. En présence de Mn<sup>2+</sup>, l'IN est flexible et un de ses groupes négatifs peut attirer la chaîne, rendant la liaison scissible accessible. La conformation Nord d'aA<sup>N</sup> n'affecte pas le taux de 3'-processing en présence de Mn<sup>2+</sup>.

De plus, la modification en position 3 des nucléotides des deux brins par aA<sup>N</sup> et U<sup>N</sup>, stimule considérablement le 3' processing en présence de Mn<sup>2+</sup>. La chaîne NX d'aA<sup>N</sup> en Nord est certainement attirée par un groupe négatif de l'IN flexible en présence de Mn<sup>2+</sup>. La liaison scissible est accessible et l'IN peut avoir toute son activité. Si U<sup>M</sup> a un plissement Nord lorsqu'il remplace T3, des conformations Nord au niveau de T3 et d'A3 sont alors très

favorables à la réaction de 3'-processing, en présence de  $Mn^{2+}$ . Ces mêmes modifications inhibent le 3'-processing en présence de  $Mg^{2+}$  comme la chaîne NX d'aA<sup>N</sup> en Nord empêche l'accès à la liaison scissible par l'IN.

Ainsi, si les analogues de nucléotides ont un plissement Nord lorsqu'ils remplacent A3 et T3 dans l'ADN viral, ces conformations n'affectent pas ou même stimulent le 3'-processing. Une transformation conformationnelle B → A au niveau de la 3<sup>ème</sup> paire de bases pourrait être nécessaire pour la formation du complexe IN/ADN actif. Une telle déformation locale est très observée dans les complexes protéines/ADN (Berman et al., 1992), en particulier dans ceux formés entre l'ADN et des enzymes catalysant des réactions de formation ou de clivage des liaisons phosphodiester O3'-P telles que des polymérases, des endonucléases et des transposases (Lu et al., 2000). La transformation B → A expose sélectivement les atomes du squelette sucre-phosphate qui sont habituellement enfouis à l'intérieur de la double hélice. Les interactions de la protéine avec le petit sillon sont favorisées. Or, les modifications au niveau du petit sillon de l'ADN viral affectent davantage la réaction de 3'-processing que celles ayant lieu dans le grand sillon (Wang et al., 1999).

Par ailleurs, le coût énergétique de ce type de transition est séquence dépendante (Minchenkova et al., 1986; Tolstorukov et al., 2001) et pourrait donc augmenter la sélectivité de la liaison IN/ADN.

Ce genre de déformation peut être induite par la liaison de la protéine ou bien, être préexistante dans la structure de l'ADN libre et accentuée par la liaison de la protéine.

Afin d'évaluer l'hypothèse qu'une transformation B → A soit impliquée dans la formation du complexe IN/ADN actif pour le 3'-processing, de nouvelles modifications des oligonucléotides vont être testées expérimentalement : le remplacement d'A3 et/ou de T3 par des ribonucléotides et arabinonucléotides qui sont normalement stabilisés en Nord. Les répercussions de ces modifications sur l'extrémité LTR U5 de l'ADN viral ont été étudiées par des simulations de dynamique moléculaire et sont présentées dans la partie III.

***PARTIE III***

**Structure fine d’analogues de l’extrémité LTR  
U5 de l’ADN viral**





Nous avons mis en évidence, dans la partie précédente, l'importance de la structure fine de l'extrémité LTR U5 de l'ADN viral dans son interaction avec l'IN afin que le complexe actif puisse se former et que la réaction de 3'-processing se produise. Nos données, en particulier, nous ont permis d'émettre l'hypothèse qu'une conformation de type A, au niveau de la troisième paire de bases de l'extrémité LTR U5 de l'ADN viral pourrait être impliquée dans la formation du complexe IN/ADN actif. Une transformation conformationnelle B → A pourrait être induite ou accentuée par la liaison de la protéine et serait déterminante dans l'interaction séquence-spécifique.

Cette hypothèse nous a amenés à étudier la structure d'analogues de l'extrémité LTR U5 de l'ADN viral modifiés à la position 3 par des ribonucléotides ou un arabinonucléotide. Nous avons voulu déterminer l'effet de l'introduction, en position 3, de composés nucléotidiques ayant un plissement Nord sur la structure fine de l'ADN viral au niveau et au voisinage du site de clivage de l'IN. Pour réaliser cette étude, nous avons simulé des dynamiques moléculaires de 2,5 ns d'oligonucléotides dont la séquence correspond à celle des 21 dernières paires de bases de l'extrémité LTR U5 de l'ADN viral mais dont les nucléotides A3 et, ou T3 ont été remplacés par un ribonucléotide ou un arabinonucléotide (**figure 1**). Nous avons également simulé la trajectoire de l'oligonucléotide sauvage WT et de deux oligonucléotides dont les bases ont été modifiées en position 3 (**figure 1**) :

- [A/U] avec lequel le taux initial de 3'-processing est similaire à celui de WT (Mazumder and Pommier, 1995)
- [T/A] avec lequel le taux initial de 3'-processing est considérablement diminué (Scottoline *et al.*, 1997)

A partir de ces trajectoires de simulation, ont été déterminés :

- les paramètres moyens de bases et de pas et leur fluctuation,
- les pourcentages de plissements Nord (N), Sud (S) et Est (E) adoptés par chaque nucléotide,

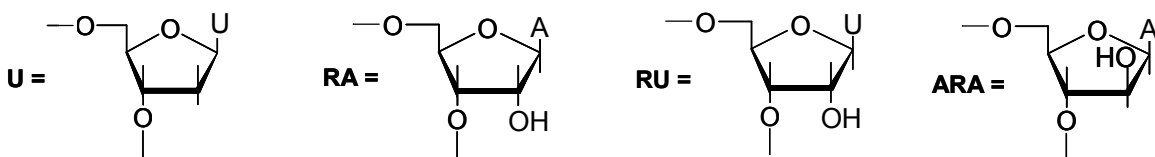


- les valeurs moyennes des angles dièdres du squelette sucre-phosphate et des angles glycosidiques ainsi que leur fluctuation.
- les distances moyennes des liaisons hydrogènes impliquées dans les paires de bases et leur fluctuation.
- la distribution de l'orientation du HO'2 des ribonucléotides (l'hydrogène du groupe hydroxyl en 2' des ribonucléotides) qui joue un rôle important pour leur conformation et leur interaction avec l'environnement (Auffinger and Westhof, 1997), ainsi que celle du HO'2 de l'arabinonucléotide.

La structure de chacun des 7 oligonucléotides étudiés a ainsi été caractérisée.

<b>WT</b>	5' G <sub>21</sub> T <sub>20</sub> G <sub>19</sub> T <sub>18</sub> G <sub>17</sub> G <sub>16</sub> A <sub>15</sub> A <sub>14</sub> A <sub>13</sub> A <sub>12</sub> T <sub>11</sub> C <sub>10</sub> T <sub>9</sub> C <sub>8</sub> T <sub>7</sub> A <sub>6</sub> G <sub>5</sub> C <sub>4</sub> <b>A</b> <sub>3</sub> G <sub>2</sub> T <sub>1</sub> 3'
	3' C <sub>21</sub> A <sub>20</sub> C <sub>19</sub> A <sub>18</sub> C <sub>17</sub> C <sub>16</sub> T <sub>15</sub> T <sub>14</sub> T <sub>13</sub> T <sub>12</sub> A <sub>11</sub> G <sub>10</sub> A <sub>9</sub> G <sub>8</sub> A <sub>7</sub> T <sub>6</sub> C <sub>5</sub> G <sub>4</sub> <b>T</b> <sub>3</sub> C <sub>2</sub> A <sub>1</sub> 5'
<b>[A/U]</b>	5' G <sub>21</sub> T <sub>20</sub> G <sub>19</sub> T <sub>18</sub> G <sub>17</sub> G <sub>16</sub> A <sub>15</sub> A <sub>14</sub> A <sub>13</sub> A <sub>12</sub> T <sub>11</sub> C <sub>10</sub> T <sub>9</sub> C <sub>8</sub> T <sub>7</sub> A <sub>6</sub> G <sub>5</sub> C <sub>4</sub> <b>A</b> <sub>3</sub> G <sub>2</sub> T <sub>1</sub> 3'
	3' C <sub>21</sub> A <sub>20</sub> C <sub>19</sub> A <sub>18</sub> C <sub>17</sub> C <sub>16</sub> T <sub>15</sub> T <sub>14</sub> T <sub>13</sub> T <sub>12</sub> A <sub>11</sub> G <sub>10</sub> A <sub>9</sub> G <sub>8</sub> A <sub>7</sub> T <sub>6</sub> C <sub>5</sub> G <sub>4</sub> <b>U</b> <sub>3</sub> C <sub>2</sub> A <sub>1</sub> 5'
<b>[T/A]</b>	5' G <sub>21</sub> T <sub>20</sub> G <sub>19</sub> T <sub>18</sub> G <sub>17</sub> G <sub>16</sub> A <sub>15</sub> A <sub>14</sub> A <sub>13</sub> A <sub>12</sub> T <sub>11</sub> C <sub>10</sub> T <sub>9</sub> C <sub>8</sub> T <sub>7</sub> A <sub>6</sub> G <sub>5</sub> C <sub>4</sub> <b>T</b> <sub>3</sub> G <sub>2</sub> T <sub>1</sub> 3'
	3' C <sub>21</sub> A <sub>20</sub> C <sub>19</sub> A <sub>18</sub> C <sub>17</sub> C <sub>16</sub> T <sub>15</sub> T <sub>14</sub> T <sub>13</sub> T <sub>12</sub> A <sub>11</sub> G <sub>10</sub> A <sub>9</sub> G <sub>8</sub> A <sub>7</sub> T <sub>6</sub> C <sub>5</sub> G <sub>4</sub> <b>A</b> <sub>3</sub> C <sub>2</sub> A <sub>1</sub> 5'
<b>[RA/T]</b>	5' G <sub>21</sub> T <sub>20</sub> G <sub>19</sub> T <sub>18</sub> G <sub>17</sub> G <sub>16</sub> A <sub>15</sub> A <sub>14</sub> A <sub>13</sub> A <sub>12</sub> T <sub>11</sub> C <sub>10</sub> T <sub>9</sub> C <sub>8</sub> T <sub>7</sub> A <sub>6</sub> G <sub>5</sub> C <sub>4</sub> <b>RA</b> <sub>3</sub> G <sub>2</sub> T <sub>1</sub> 3'
	3' C <sub>21</sub> A <sub>20</sub> C <sub>19</sub> A <sub>18</sub> C <sub>17</sub> C <sub>16</sub> T <sub>15</sub> T <sub>14</sub> T <sub>13</sub> T <sub>12</sub> A <sub>11</sub> G <sub>10</sub> A <sub>9</sub> G <sub>8</sub> A <sub>7</sub> T <sub>6</sub> C <sub>5</sub> G <sub>4</sub> <b>T</b> <sub>3</sub> C <sub>2</sub> A <sub>1</sub> 5'
<b>[A/RU]</b>	5' G <sub>21</sub> T <sub>20</sub> G <sub>19</sub> T <sub>18</sub> G <sub>17</sub> G <sub>16</sub> A <sub>15</sub> A <sub>14</sub> A <sub>13</sub> A <sub>12</sub> T <sub>11</sub> C <sub>10</sub> T <sub>9</sub> C <sub>8</sub> T <sub>7</sub> A <sub>6</sub> G <sub>5</sub> C <sub>4</sub> <b>A</b> <sub>3</sub> G <sub>2</sub> T <sub>1</sub> 3'
	3' C <sub>21</sub> A <sub>20</sub> C <sub>19</sub> A <sub>18</sub> C <sub>17</sub> C <sub>16</sub> T <sub>15</sub> T <sub>14</sub> T <sub>13</sub> T <sub>12</sub> A <sub>11</sub> G <sub>10</sub> A <sub>9</sub> G <sub>8</sub> A <sub>7</sub> T <sub>6</sub> C <sub>5</sub> G <sub>4</sub> <b>RU</b> <sub>3</sub> C <sub>2</sub> A <sub>1</sub> 5'
<b>[RA/RU]</b>	5' G <sub>21</sub> T <sub>20</sub> G <sub>19</sub> T <sub>18</sub> G <sub>17</sub> G <sub>16</sub> A <sub>15</sub> A <sub>14</sub> A <sub>13</sub> A <sub>12</sub> T <sub>11</sub> C <sub>10</sub> T <sub>9</sub> C <sub>8</sub> T <sub>7</sub> A <sub>6</sub> G <sub>5</sub> C <sub>4</sub> <b>RA</b> <sub>3</sub> G <sub>2</sub> T <sub>1</sub> 3'
	3' C <sub>21</sub> A <sub>20</sub> C <sub>19</sub> A <sub>18</sub> C <sub>17</sub> C <sub>16</sub> T <sub>15</sub> T <sub>14</sub> T <sub>13</sub> T <sub>12</sub> A <sub>11</sub> G <sub>10</sub> A <sub>9</sub> G <sub>8</sub> A <sub>7</sub> T <sub>6</sub> C <sub>5</sub> G <sub>4</sub> <b>RU</b> <sub>3</sub> C <sub>2</sub> A <sub>1</sub> 5'
<b>[ARA/T]</b>	5' G <sub>21</sub> T <sub>20</sub> G <sub>19</sub> T <sub>18</sub> G <sub>17</sub> G <sub>16</sub> A <sub>15</sub> A <sub>14</sub> A <sub>13</sub> A <sub>12</sub> T <sub>11</sub> C <sub>10</sub> T <sub>9</sub> C <sub>8</sub> T <sub>7</sub> A <sub>6</sub> G <sub>5</sub> C <sub>4</sub> <b>ARA</b> <sub>3</sub> G <sub>2</sub> T <sub>1</sub> 3'
	3' C <sub>21</sub> A <sub>20</sub> C <sub>19</sub> A <sub>18</sub> C <sub>17</sub> C <sub>16</sub> T <sub>15</sub> T <sub>14</sub> T <sub>13</sub> T <sub>12</sub> A <sub>11</sub> G <sub>10</sub> A <sub>9</sub> G <sub>8</sub> A <sub>7</sub> T <sub>6</sub> C <sub>5</sub> G <sub>4</sub> <b>T</b> <sub>3</sub> C <sub>2</sub> A <sub>1</sub> 5'

## B.



**Figure 1- A.** Séquence des oligonucléotides doubles brins étudiés (1<sup>ier</sup> brin : brin b, clivé par l'IN lors du 3'-processing, 2<sup>nd</sup> brin : brin a). **WT** : oligonucléotide sauvage correspondant aux 21 dernières paires de bases de l'extrémité LTR U5 de l'ADN viral ; **A/U** : remplacement du nucléotide T3 par U ; **T/A** : remplacement de la paire de bases A3/T3 par T/A ; **RA/T** : remplacement du nucléotide A3 par RA ; **A/RU** : remplacement du nucléotide T3 par RU. **ARA/T** : remplacement du nucléotide A3 par ARA.

**B.** Structure des nucléotides insérés en position 3.



## Méthode

### Le plissement du sucre des nucléotides

Deux paramètres, l'angle de pseudorotation  $P$  et l'amplitude maximale du plissement  $\tau$ , suffisent pour définir les cinq angles dièdres endocycliques du sucre  $\tau_j$  et donc pour caractériser le plissement (Rao et al., 1981) :

$$\tau_j = \tau \cos[P + 4\pi(j - 2)/5] \quad (1)$$

A partir de la mesure des angles  $\tau_j$ , dans une conformation,  $A$  et  $B$  sont calculés ainsi :

$$A = 2/5 \sum_{j=0}^4 \tau_j \cos[4\pi(j - 2)/5] \quad (2)$$

$$B = -2/5 \sum_{j=0}^4 \tau_j \sin[4\pi(j - 2)/5] \quad (3)$$

L'amplitude maximale  $\tau$  et l'angle de pseudorotation  $P$  sont alors donnés par :

$$\tau = (A^2 + B^2)^{1/2} \quad (4)$$

$$P = \tan^{-1} B/A \quad (5)$$

L'espace de pseudorotation est divisé en quatre cadrans de tailles égales, centrés autour de  $P=0,0^\circ$ ,  $P=90,0^\circ$ ,  $P=180,0^\circ$  et  $P=-90,0^\circ$  qui correspondent aux plissements Nord (N), Sud (S), Est (E) et Ouest (W) des sucres. Dans Morcad (Le Bret et al., 1991), les plissements sont séparés en Nord de  $-20^\circ$  à  $+54^\circ$ , en Est de  $+54^\circ$  à  $+126^\circ$  et en Sud de  $+126^\circ$  à  $+160^\circ$ .

### Les dynamiques moléculaires

Les dynamiques moléculaires des 7 oligonucléotides de 21 paires de bases (**figure 1**) ont été simulées avec le module SANDER d'AMBER 8 (Case et al., 2004) et le champ de force parm99 ; 40 ions sodium ont été ajoutés pour neutraliser le système. Le soluté (l'ADN et ses contre-ions) a été mis dans une boîte octaédrique tronquée ayant une dimension face à face d'environ 100 Å. Cette boîte permet à la couche d'hydratation de s'étendre d'au moins 10 Å autour de l'ADN. Le soluté a été hydraté avec des molécules d'eau TIP3P (Jorgensen, 1981). Les simulations ont été réalisées, dans des conditions périodiques, aux limites avec la cellule centrale contenant environ 14200 molécules d'eau.

Dans la conformation de départ, les oligonucléotides ont été mis dans une forme B canonique d'Arnott. Un plissement Nord canonique a été donné aux ribonucléotides RA et RU et à l'arabinonucléotide ARA. Les angles dièdres  $\varepsilon$  ( $C4'-C3'-O3'-P$ ) et  $\chi$  ( $O4'-C1'-N1-C2$  des pyrimidines et  $O4'-C1'-N9-C4$  des purines) des nucléotides RA, RU et ARA ont été mis à



leur valeur modale observée dans les doubles hélices de forme A :  $\varepsilon = -155^\circ$  et  $\chi = 161^\circ$  (Berman et al., 1992).

#### *Phase de préparation des oligonucléotides WT, A/U et T/A*

La préparation a débuté par une minimisation de 1000 pas avec des contraintes harmoniques de  $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  sur les atomes du soluté. Un chauffage lent jusqu'à 310 K a ensuite été réalisé à pression constante (1 bar) et avec des contraintes harmoniques de  $20 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . Ces contraintes ont été relâchées de 10 à  $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  durant une série de 4 équilibrations de 50 ps, réalisées dans des conditions de température (310 K) et de pression (1 bar) constantes, en utilisant l'algorithme de Berendsen (avec une constante de 0,2 ps pour les deux paramètres) (Berendsen et al., 1984). Une dernière équilibration de 50 ps a été effectuée sans contrainte.

#### *Phase de préparation des oligonucléotides RA/T, A/RU, RA/RU et ARA/T*

La préparation a débuté par une minimisation de 1000 pas avec des contraintes harmoniques de  $0.5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  sur les atomes du soluté et des contraintes de  $80 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  sur les angles dièdres :

- $\tau_0$  à  $\tau_5$  du sucre de tous les nucléotides afin que RA, RU et ARA soient en Nord et que les autres nucléotides soient en Sud
- $\varepsilon$  et  $\chi$  des nucléotides RA, RU et ARA afin qu'ils soient à leur valeur modale observée dans les doubles hélices de forme A
- $\lambda$  (HO'2-O2'-C2'-H2') des nucléotides RA et RU afin que l'hydroxyl en 2' soit orienté vers l'O3' (Auffinger et al., 1997).

Une deuxième minimisation de 1000 pas, suivie d'un chauffage jusqu'à 310 K à pression constante (1 bar) ont été réalisés avec uniquement les contraintes de  $80 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  sur les dièdres. Ensuite, ces contraintes ont été progressivement relâchées :

- de  $80$  à  $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  durant une série de 15 minimisations de 1000 pas
- de  $10$  à  $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  durant une série de 6 équilibrations de 50 ps réalisées dans des conditions de température (310 K) et de pression (1 bar) constantes, via l'algorithme de Berendsen (avec une constante de 0,2 ps pour les deux paramètres) (Berendsen H.J. et al., 1984).

Une dernière équilibration de 50 ps a été effectuée sans contrainte.





### *Phase de production des 7 oligonucléotides*

La phase de production a été simulée durant 2,5 ns avec un pas d'intégration de 2 fs et dans des conditions de température (310 K) et de pression (1 bar) constantes, via l'algorithme de Berendsen (avec une constante de 5 ps pour les deux paramètres) (Berendsen H.J. *et al.*, 1984). Les interactions électrostatiques ont été traitées en utilisant l'algorithme Particle Mesh Ewald (PME) (Essman *et al.*, 1995) avec un cutoff de 10 Å et un espacement, entre les points de la grille, d'environ 1 Å. Les longueurs des liaisons impliquant des atomes d'hydrogène sont maintenues fixes (Ryckaert *et al.*, 1977). Le centre de masse a été remis au centre de la boîte tous les 5000 pas, pour éviter les problèmes méthodologiques décrits antérieurement (Harvey *et al.*, 1998). Les conformations du système ont été sauvegardées toutes les 0,2 ps pour réaliser les analyses.

### **Les paramètres de bases et de pas**

Le programme OCL (Gabarro-Arpa *et al.*, 1992) est conçu pour donner les outils de base permettant de générer des trièdres de référence et de calculer l'orientation de trièdres les uns par rapport aux autres. Le trièdre de référence est calculé à partir des axes d'inertie des objets qui, par définition, n'ont pas d'orientation. Pour tout de même orienter le trièdre, on a recours au produit scalaire de l'axe avec des vecteurs liés aux atomes de l'objet. C'est à ce niveau qu'entre une condition qui rend l'algorithme non différentiable : si le produit scalaire est négatif, le sens de l'axe doit être inversé. Il est plus simple de calculer le trièdre à partir de la rotation, permettant d'ajuster au mieux une paire de bases idéale (Olson *et al.*, 2001) à la paire de bases réelle de l'ADN étudié. Curieusement, l'algorithme de Lu et Olson : 3DNA comporte également une condition sur l'orientation du trièdre calculée à partir d'un produit scalaire. En fait, il est possible d'adapter la procédure de McLachlan (McLachlan, 1979) pour que l'ajustement se fasse automatiquement, toujours de la même façon, sans condition. Pour faire marcher cet algorithme dans Morcad (Le Bret M. *et al.*, 1991), il faut indiquer au programme les résidus qui ont des appariements Watson-Crick par une instruction du type : wc 1-12 : 17-28.



## Résultats

### L'oligonucléotide sauvage WT

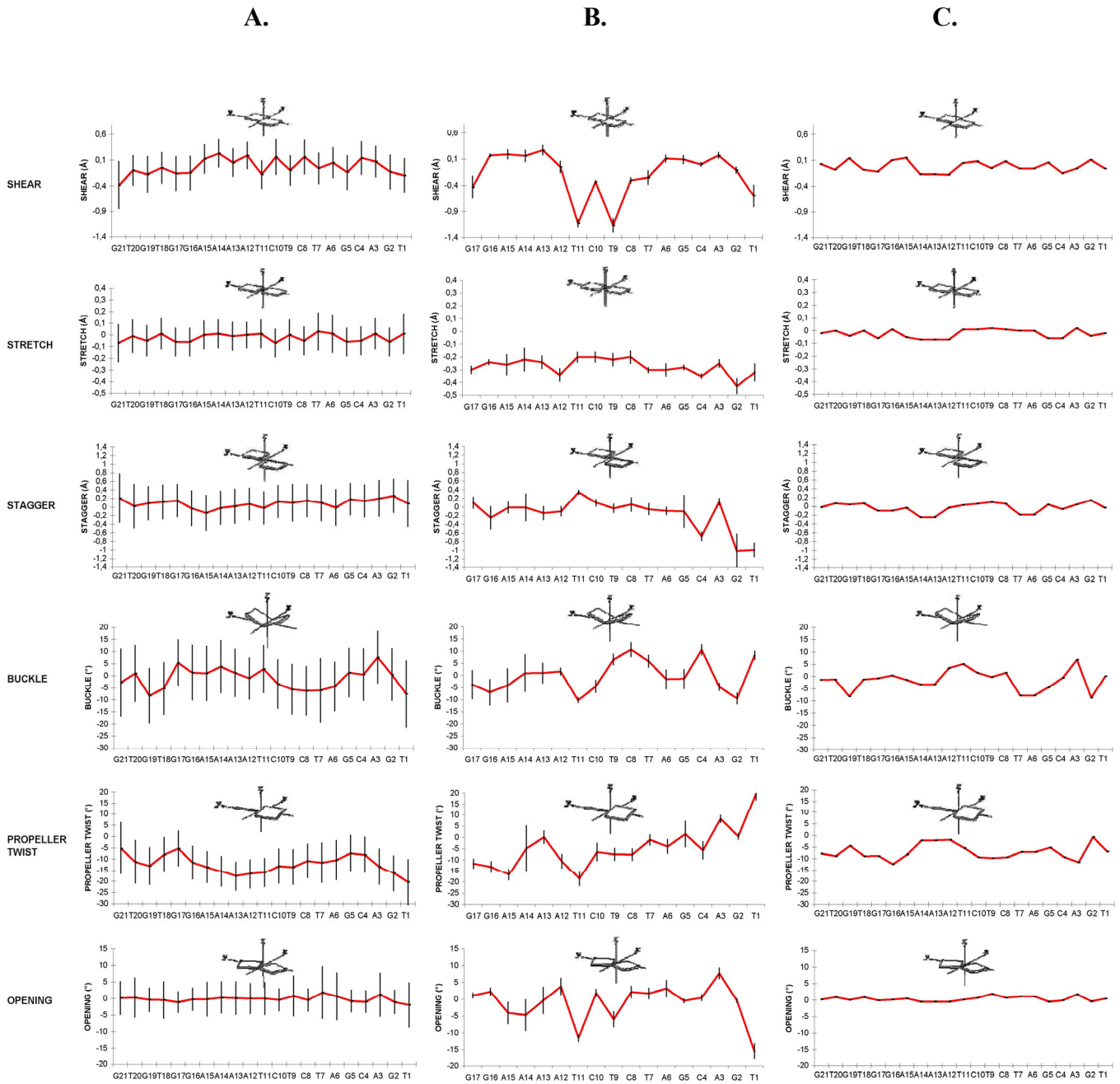
Après 2,5 ns de dynamique moléculaire, l'oligonucléotide WT présente une structure en double hélice ayant toujours une forme générale d'un ADN B<sub>I</sub>. La plupart des paramètres de bases (**figure 2-A.**) et de pas (**figure 3-A.**) moyens calculés sur les 1,5 dernières ns restent globalement dans leur gamme de valeurs modales observées dans l'ADN de forme B (Berman *et al.*, 1992). Cependant, certains de ces paramètres ont des valeurs particulières, notamment au niveau et au voisinage du site de clivage du 3'-processing par l'IN :

- un buckle très positif ( $\approx +8^\circ$ ) au niveau de la paire de bases A3/T3 et négatif au niveau de la première paire de bases ( $\approx -8^\circ$ ) et des paires 7 à 9 ( $\approx -6^\circ$ ) (**figure 2-A**).
- un propeller twist très négatif au niveau des paires de bases 1 ( $\approx -20^\circ$ ) et 2 ( $\approx -17^\circ$ ) (**figure 2-A**).
- un roll important au niveau des pas C4-A3/G4-T3 ( $\approx +9^\circ$ ) et T7-A6/A7-T6 ( $\approx +13^\circ$ ) (**figure 3-A**).
- un twist élevé ( $\approx +38^\circ$ ) au niveau du pas A3-G2/T3-C2 et faible au niveau des pas C4-A3/G4-T3 ( $\approx +29^\circ$ ) et T7-A6/A7-T6 ( $\approx +26^\circ$ ) (**figure 3-A**).

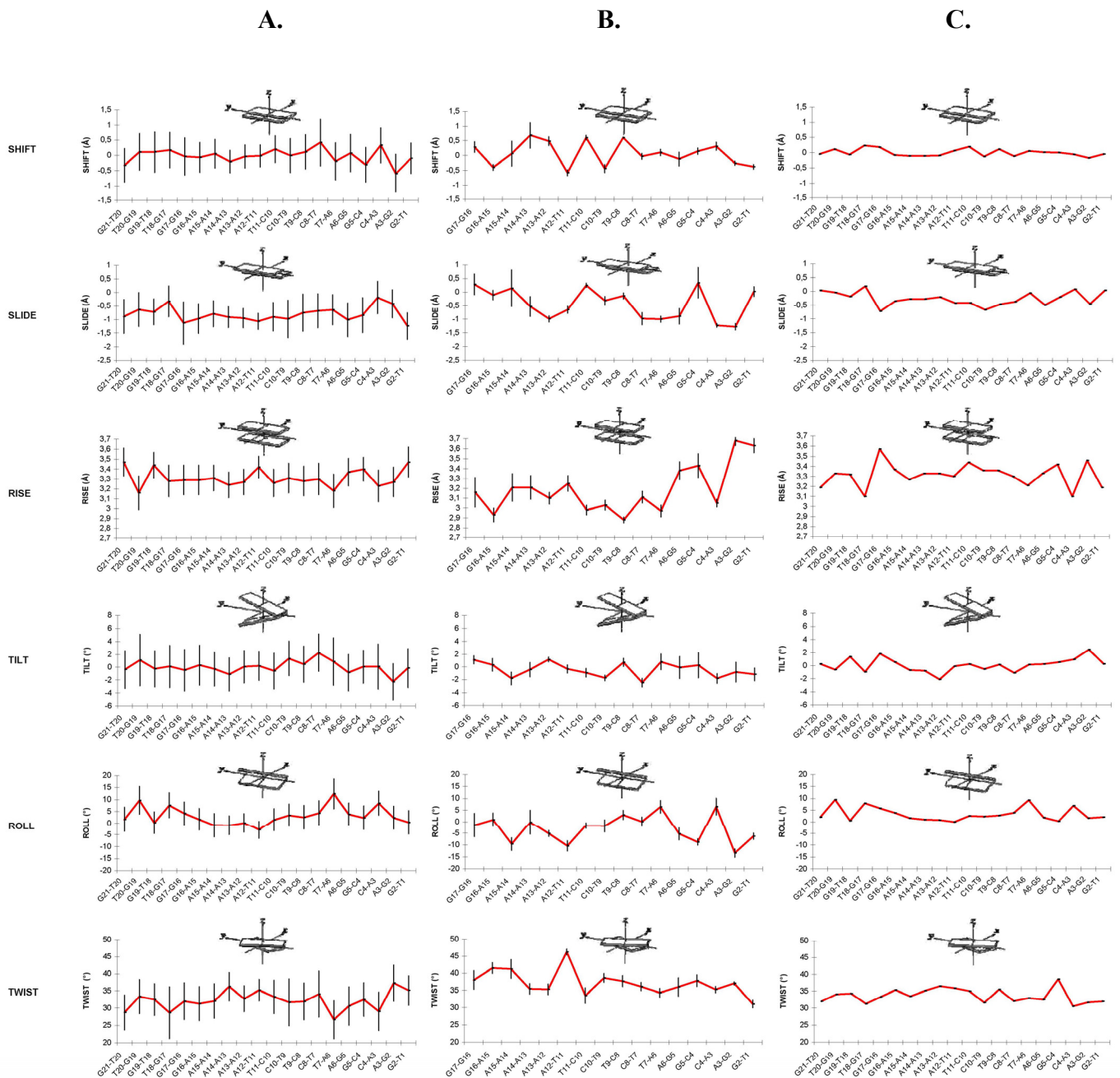
Les valeurs de roll et de twist, au niveau des pas C4-A3/G4-T3 et T7-A6/A7-T6, appartiennent aux gammes de valeurs modales de ces paramètres des doubles hélices d'ADN de forme A.

Les 21 paires de bases présentent des appariements stables de Watson-Crick comme le montrent les valeurs et les fluctuations du Stretch et du Stagger (**figure 2-A**) ainsi que les mesures des longueurs des liaisons hydrogènes impliquées dans ces appariements (données non montrées ici).

La majorité des angles dièdres du squelette sucre-phosphate ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\varepsilon$  et  $\zeta$ ) et des angles glycosidiques  $\chi$  restent durant toute la dynamique dans leur gamme de valeurs modales observées dans les doubles hélices de forme B<sub>I</sub> (Berman *et al.*, 1992).



**Figure 2-** Les paramètres de bases. **A.** Les paramètres de bases moyens (calculés avec Morcad) de l'oligonucléotide WT durant les 1,5 dernières ns de dynamique moléculaire. Les barres verticales indiquent la fluctuation de ces paramètres. **B.** Les paramètres de bases moyens (calculés avec Morcad) des 9 modèles RMN d'un oligonucléotide correspondant aux 17 dernières paires de bases du LTR U5 de l'ADN viral (code PDB : 1TQR) (Renisio et al., 2005). Les barres verticales indiquent l'écart de ces paramètres dans les 9 modèles. **C.** Les paramètres de bases calculés avec le serveur MDDNA (Dixit and Beveridge, 2006) à partir de la séquence nucléotidique de WT. Les schémas des paramètres de bases représentent des valeurs positives.



**Figure 3-** Les paramètres de pas. **A.** Les paramètres de pas moyens (calculés avec Morcad) de l'oligonucléotide WT durant les 1,5 dernières ns de dynamique moléculaire. Les barres verticales indiquent la fluctuation de ces paramètres. **B.** Les paramètres de pas moyens (calculés avec Morcad) des 9 modèles RMN d'un oligonucléotide correspondant aux 17 dernières paires de bases du LTR U5 de l'ADN viral (code PDB : 1TQR) (Renisio *et al.*, 2005). Les barres verticales indiquent l'écart de ces paramètres dans les 9 modèles. **C.** Les paramètres de pas calculés avec le serveur MDDNA (Dixit *et al.*, 2006) à partir de la séquence nucléotidique de WT. Les schémas des paramètres de bases représentent des valeurs positives.



La majorité des nucléotides de WT sont stabilisés dans un plissement Sud durant les 1,5 dernières ns de la dynamique moléculaire (**tableau 1**). Par contre, quelques nucléotides adoptent plus souvent un plissement Nord. Il s'agit en particulier, des désoxythymidines T3 et T6 du brin a et T7 du brin b. Ces plissements Nord plus fréquents pourraient être liés au fait que les valeurs de roll et de twist des pas C4-A3/G4-T3 et T7-A6/A7-T6 sont caractéristiques de l'ADN de forme A.

**Tableau 1-** Les pourcentages des plissements Nord (N), Sud (S) et Est (E) adoptés par les 12 premiers nucléotides des brins a et b de WT au cours des 1,5 dernières ns de la dynamique moléculaire

Brin b	% en N	% en E	% en S	Brin a	% en N	% en E	% en S
<b>5' A12</b>	0,1	8,9	91,0	<b>3' T12</b>	1,1	10,3	88,6
<b>T11</b>	0,6	8,2	91,2	<b>A11</b>	0,3	6,4	93,3
<b>C10</b>	1,1	16,0	82,9	<b>G10</b>	0,1	5,6	94,3
<b>T9</b>	1,4	16,7	81,9	<b>A9</b>	0,5	9,5	90
<b>C8</b>	4,0	22,1	73,9	<b>G8</b>	0,7	6,8	92,5
<b>T7</b>	<b>9,4</b>	33,5	57,1	<b>A7</b>	2	8,8	89,2
<b>A6</b>	1,0	16,0	83,0	<b>T6</b>	<b>8,6</b>	21,7	69,7
<b>G5</b>	1,0	10,8	88,3	<b>C5</b>	2,5	22,7	74,8
<b>C4</b>	1,0	14,4	84,6	<b>G4</b>	1,1	3,9	95,1
<b>A3</b>	0,4	3,0	<b>96,5</b>	<b>T3</b>	<b>10,3</b>	41,7	47,9
<b>G2</b>	2,0	3,8	94,2	<b>C2</b>	0,5	12,4	87,2
<b>3' T1</b>	2,2	15,0	82,8	<b>5' A1</b>	2,5	4,3	93,2

L'oligonucléotide WT présente donc des déformations structurales au niveau et au voisinage du site de clivage et en particulier, une tendance du pas C4-A3/G4-T3 à adopter une conformation de type A qui pourrait être significative pour la formation du complexe actif IN/ADN.

Nous avons comparé la structure fine de WT avec les paramètres de bases et de pas fournis par le serveur MDDNA, à partir de la séquence des 21 paires de bases de WT (Dixit *et al.*, 2006) (**figures 2-C et 3-C**). Cet outil construit un modèle atomique de l'oligonucléotide à partir de sa séquence sur la base d'informations structurales stockées dans une base de données relationnelles. Cette base contient les détails conformationnels d'ADN obtenus à partir de 39 trajectoires de dynamique moléculaire différentes, comprenant tous les 136 pas des tétranucléotides uniques.

Les profils (courbes en rouge des **figures 2-A, 2-C, 3-A et 3-C**) de plusieurs paramètres : le stretch, le stagger, l'opening et le roll sont très similaires dans les deux modèles. Certains paramètres ont des profils qui diffèrent davantage mais les valeurs des paramètres du modèle





prédit à partir de MDDNA restent dans la barre de fluctuations de ceux de notre structure ; ceci est le cas du shear, du buckle, du shift et du tilt. D'autres paramètres de bases ou de pas du modèle fourni par MDDNA ont des valeurs qui n'appartiennent pas à la gamme de fluctuations des paramètres de notre structure, au niveau de certaines paires de bases ou pas : ce sont les cas du propeller twist, du rise, du slide et du twist.

Des caractéristiques structurales observées dans notre structure au niveau ou au voisinage du site de clivage se retrouvent dans ce modèle :

- le buckle très positif au niveau d'A3/T3 ( $\approx +8^\circ$ ) et négatif au niveau de T7/A7 ( $\approx +9^\circ$ ) (**figure 2-C**)
- le roll important au niveau des pas C4-A3/G4-T3 ( $\approx +7^\circ$ ) et T7-A6/A7-T6 ( $\approx +11^\circ$ ) (**figure 3-C**).

Globalement, les deux modèles sont relativement proches.

Nous avons calculé avec Morcad les paramètres moyens de bases et de pas des 9 modèles, déterminés par RMN (code PDB : 1TQR) (Renisio *et al.*, 2005), d'un oligonucléotide de 17 pb correspondant à l'extrémité LTR U5 de l'ADN viral. Dans certaines régions de cette structure, les paramètres de bases et de pas (**figures 2-B et 3-B**) sont très éloignés de leur gamme de valeurs modales observées dans des doubles hélices d'ADN (Berman *et al.*, 1992). Les écarts des paramètres entre les 9 modèles sont faibles. Les profils des paramètres (**figures 2-B et 3-B**) diffèrent de ceux de notre structure (**figures 2-A et 3-A**). La seule caractéristique structurale observée dans notre modèle au niveau et au voisinage du site de clivage et présente dans cette structure, consiste en un roll positif élevé au niveau des pas C4-A3/G4-T3 et T7-A6/A7-T6.

La structure RMN de l'extrémité LTR U5 de l'ADN viral est donc très particulière ; elle n'est pas retrouvée par dynamique moléculaire.

## Les oligonucléotides [A/U] et [T/A]

### *L'oligonucléotide [A/U]*

L'oligonucléotide [A/U] dans lequel le T3 du brin a, a été remplacé par U3, se présente, comme l'oligonucléotide WT, sous la forme générale d'un ADN B<sub>1</sub> après 2,5 ns de dynamique moléculaire. Les profils des paramètres de bases et de pas de [A/U] (courbes en rouge, **figures 4-B et 5-B**) sont similaires à ceux des paramètres de WT (**figures 4-A et 5-A**).



Dans tous les cas, les valeurs moyennes des paramètres de [A/U] appartiennent aux gammes de fluctuations de ceux de WT.

La plupart des particularités structurales observées dans WT au niveau et au voisinage du site de clivage du 3'-processing, se retrouvent dans [A/U] :

- un buckle négatif au niveau de la première paire de bases ( $\approx -12^\circ$ ) (**figure 4-B**).
- un propeller twist très négatif au niveau des paires de bases 1 ( $\approx -17^\circ$ ) et 2 ( $\approx -18^\circ$ ) (**figure 4-B**).
- un roll important au niveau des pas C4-A3/G4-U3 ( $\approx +8^\circ$ ) et T7-A6/A7-T6 ( $\approx +9^\circ$ ) (**figure 5-B**).
- un twist faible au niveau des pas C4-A3/G4-U3 ( $\approx +30^\circ$ ) et T7-A6/A7-T6 ( $\approx +27^\circ$ ) (**figure 5-B**).

Les valeurs de roll et de twist, au niveau des pas C4-A3/G4-U3 et T7-A6/A7-T6 caractéristiques de l'ADN de forme A, sont donc observées dans [A/U]. Les nucléotides U3, T6 et T7 sont plus fréquemment en Nord que la plupart des autres nucléotides de la double hélice (**tableau 2**).

**Tableau 2-** Les pourcentages des plissements Nord (N), Sud (S) et Est (E) adoptés par les 12 premiers nucléotides des brins a et b de [A/U] au cours des 1,5 dernières ns de la dynamique moléculaire

Brin b	% en N	% en E	% en S	Brin a	% en N	% en E	% en S
<b>5' A12</b>	0.4	9.4	90.2	<b>3' T12</b>	3.4	6.1	90.5
<b>T11</b>	0.3	11.5	88.3	<b>A11</b>	0.0	0.8	99.2
<b>C10</b>	0.1	1.6	98.3	<b>G10</b>	0.6	13.2	86.1
<b>T9</b>	2.3	5.2	92.5	<b>A9</b>	0.4	6.6	92.9
<b>C8</b>	0.6	16.5	82.9	<b>G8</b>	0.1	4.7	95.2
<b>T7</b>	<b>6.8</b>	28.8	64.4	<b>A7</b>	3.6	11.3	85.1
<b>A6</b>	0.4	8.3	91.3	<b>T6</b>	<b>5.8</b>	25.2	69.0
<b>G5</b>	0.4	7.1	92.5	<b>C5</b>	3.3	17.8	78.9
<b>C4</b>	3.0	14.4	82.6	<b>G4</b>	3.1	5.8	91.1
<b>A3</b>	0.3	5.2	94.5	<b>U3</b>	<b>10.2</b>	23.6	66.2
<b>G2</b>	0.1	6.5	93.4	<b>C2</b>	1.9	18.7	79.4
<b>3' T1</b>	2.4	20.2	77.5	<b>5' A1</b>	1.4	6.2	92.4

Les 21 paires de bases présentent des appariements stables de Watson-Crick pendant toute la durée de la simulation. La majorité des angles dièdres du squelette sucre-phosphate ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\epsilon$  et  $\zeta$ ) et de l'angle glycosidique  $\chi$  des nucléotides sont dans leur gamme de valeurs modales observées dans les doubles hélices de forme B<sub>1</sub> (Berman *et al.*, 1992), durant toute la simulation.



La structure d'ensemble de l'oligonucléotide [A/U] est similaire à celle de WT. Le remplacement de la thymine par un uracile en position 3 n'a pas d'effet significatif, même localement, sur la structure de l'extrémité LTR de l'ADN viral.

#### *L'oligonucléotide [T/A]*

L'oligonucléotide [T/A] dans lequel A3 du brin b et T3 du brin a ont été respectivement remplacés par T3 et A3, présente après 2,5 ns de dynamique moléculaire une structure en double hélice uniquement sur 20 paires de bases. La paire de bases T1/A1 est fondue (**tableau 3**) alors que les 20 autres sont stabilisées dans des appariements de Watson-Crick et elles sont organisées en une double hélice ayant une forme générale d'un ADN B<sub>I</sub>.

**Tableau 3-** Les distances moyennes entre les hétéroatomes normalement impliqués dans les deux liaisons hydrogènes de la paire T1/A1, mesurées au cours des 1,5 dernières ns de la dynamique moléculaire de [T/A].

Distances mesurées	Distances moyennes (Å)	Fluctuations (Å)	Distances attendues (Å)
O4 T1 – N6 A1	4.63	1.98	2.95
N3 T1 – N1 A1	4.54	1.94	2.82

La majorité des angles dièdres du squelette sucre-phosphate et des angles glycosidiques sont dans leur gamme de valeurs modales observées dans les doubles hélices de forme B<sub>I</sub> (Berman *et al.*, 1992) durant toute la simulation.

Les profils des paramètres de bases, shear, stretch, stagger, et opening des nucléotides en double hélice (2 à 21) de [T/A] (**figures 4-C**) sont semblables à ceux de WT (**figure 4-A**). Par contre, le buckle et le propeller twist présentent des différences, en particulier au niveau ou proche de la paire de bases qui a été inversée :

- le buckle est négatif ( $\approx -9^\circ$ ) au niveau de T3/A3 alors qu'il est positif ( $\approx +8^\circ$ ) au niveau de la paire A3/T3 de WT
- le propeller twist est dans sa gamme de valeurs modales d'un ADN B (Berman *et al.*, 1992) ; la valeur particulièrement négative observée pour la paire de bases 2 dans WT n'est pas retrouvée dans [T/A].

Les paramètres de pas des nucléotides en double hélice de [T/A] (**figures 5-C**) présentent aussi des différences avec ceux de WT au niveau de la paire de bases inversée et donc du site de clivage du 3'-processing. En particulier, les valeurs de roll et de twist caractéristiques de



l'ADN A au niveau du pas C4-A3/G4-T3 dans WT ne sont pas retrouvées pour le pas C4-T3/G4-A3 de [T/A]. Ces deux paramètres sont dans leur gamme de valeurs modales d'un ADN B (Berman *et al.*, 1992) pour le pas C4-T3/G4-A3 de [T/A]. La tendance de T3 à adopter plus fréquemment un plissement Nord dans WT n'est également pas observée dans [T/A] (**tableau 4**).

Les seules particularités structurales observées dans WT qui se retrouvent dans [T/A] sont :

- un buckle très négatif au niveau des paires de bases 8 et 9 ( $\approx 7^\circ$ ) (**figures 4-C**)
- un roll important au niveau du pas T7-A6/A7-T6 ( $\approx +8^\circ$ ) (**figure 5-C**).
- un twist faible au niveau du pas T7-A6/A7-T6 ( $\approx +28^\circ$ ) (**figure 5-C**).

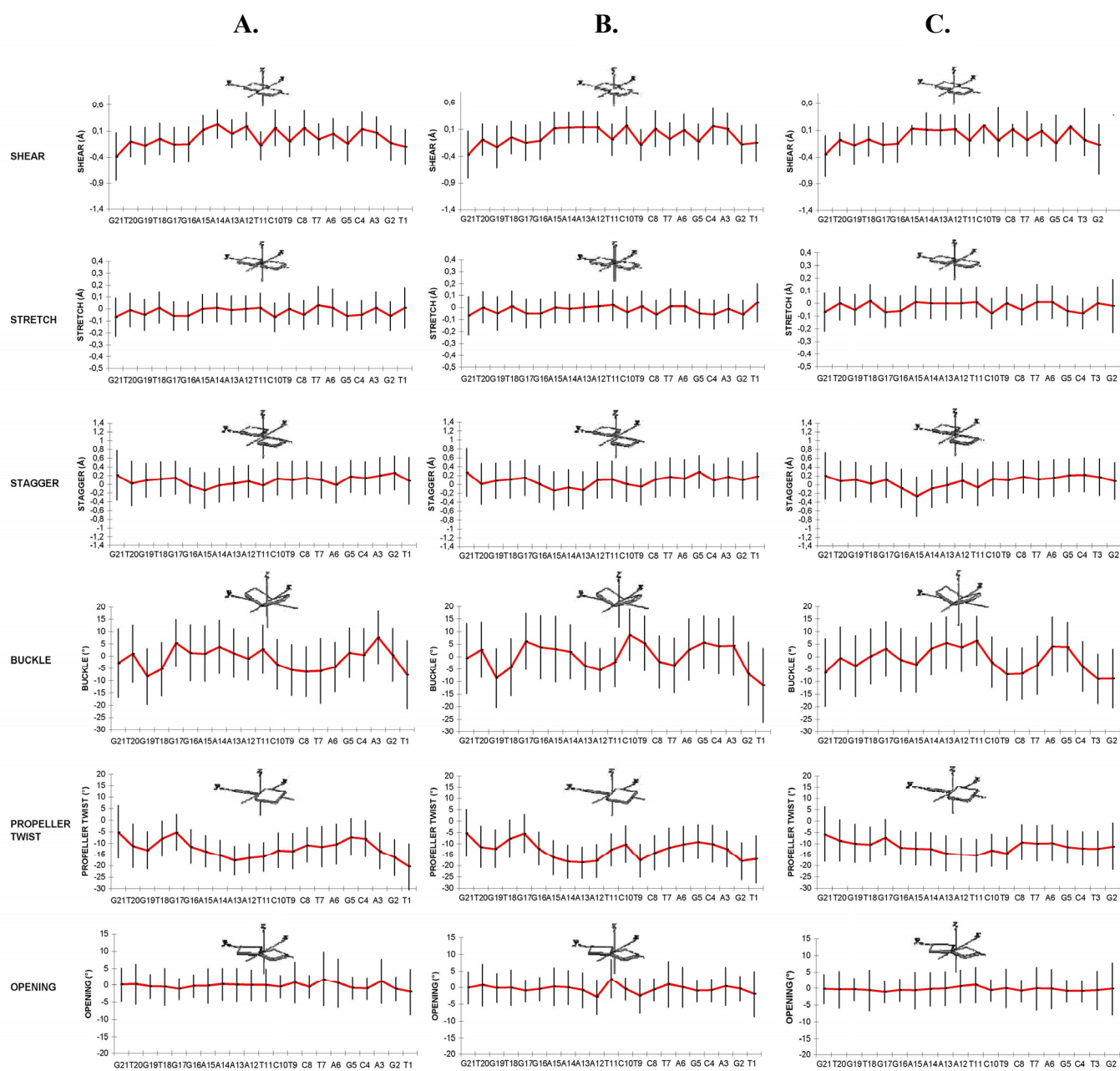
Le roll et le twist caractéristiques d'un ADN A, au niveau du pas T7-A6/A7-T6 sont certainement reliés à la plus grande tendance des désoxythymidines T6 et T7 à adopter un plissement Nord par rapport à la majorité des nucléotides qui sont stabilisés dans un plissement Sud (**tableau 4**).

Des différences dans la structure fine de la double hélice, au niveau et au voisinage du site de clivage, sont donc observées entre les deux oligonucléotides, WT et [T/A].

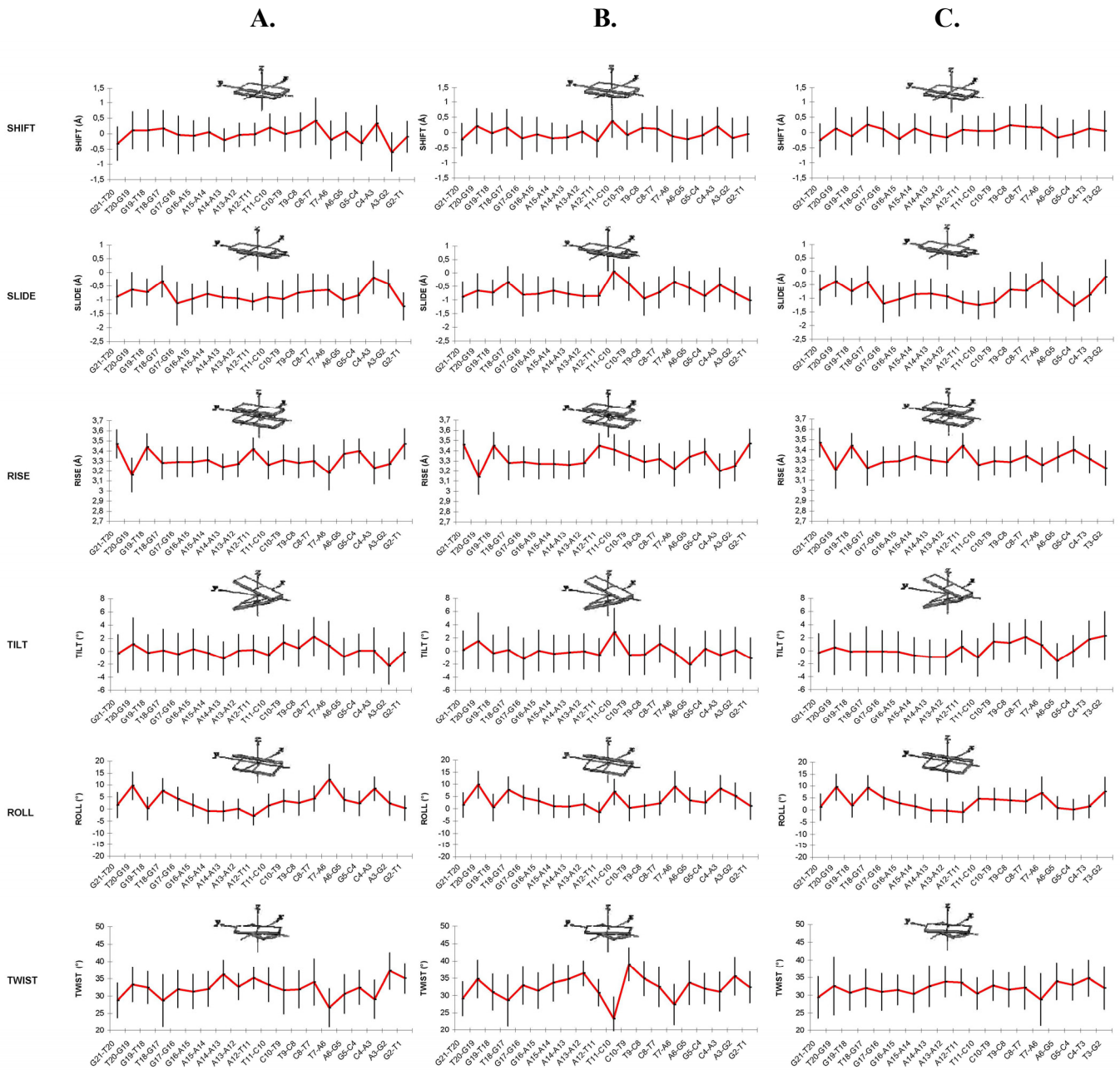
**Tableau 4-** Les pourcentages des plissements Nord (N), Sud (S) et Est (E) adoptés par les nucléotides 2 à 12 des brins a et b de [T/A] au cours des 1,5 dernières ns de la dynamique moléculaire

Brin b	% en N	% en E	% en S	Brin a	% en N	% en E	% en S
<b>5' A12</b>	1.0	7.7	91.2	<b>3' T12</b>	1.6	14.0	84.3
<b>T11</b>	1.5	17.0	81.5	<b>A11</b>	2.3	14.6	83.1
<b>C10</b>	0.4	20.3	79.3	<b>G10</b>	0.2	9.2	90.6
<b>T9</b>	2.0	25.0	73.0	<b>A9</b>	1.6	9.1	89.4
<b>C8</b>	1.9	24.6	73.5	<b>G8</b>	1.5	11.4	87.1
<b>T7</b>	<b>9.3</b>	29.4	61.2	<b>A7</b>	2.9	8.8	88.3
<b>A6</b>	0.7	7.2	92.0	<b>T6</b>	<b>6.6</b>	26.0	67.4
<b>G5</b>	0.0	5.0	95.0	<b>C5</b>	2.2	13.1	84.6
<b>C4</b>	0.4	9.4	90.3	<b>G4</b>	0.1	4.8	95.2
<b>T3</b>	1.5	13.7	84.8	<b>A3</b>	0.5	4.3	95.2
<b>3' G2</b>	0.2	4.1	95.8	<b>5'C2</b>	4.4	20.6	74.9





**Figure 4-** Les paramètres de bases moyens (calculés avec Morcad) des 1,5 dernières ns de dynamique moléculaire des oligonucléotides WT (A.), [A/U] (B.) et [T/A] (C.). Pour l'oligonucléotide [T/A], les paramètres impliquant la paire de bases T1/A1 et le pas G2-T1/C2-A1 ne sont pas représentés dans les graphes. Les barres verticales indiquent la fluctuation de ces paramètres.



**Figure 5-** Les paramètres de pas moyens (calculés avec Morcad) des 1,5 dernières ns de dynamique moléculaire des oligonucléotides WT (A.), [A/U] (B.) et [T/A] (C.). Pour l'oligonucléotide [T/A], les paramètres impliquant la paire de bases T1/A1 et le pas G2-T1/C2-A1 ne sont pas représentés dans les graphes. Les barres verticales indiquent la fluctuation de ces paramètres.



## Les oligonucléotides [RA/T], [A/RU] et [RA/RU]

### L'oligonucléotide [RA/T]

Dans l'oligonucléotide [RA/T], le ribonucléotide RA3 est stabilisé en Nord durant la dynamique moléculaire (**tableau 5**). Il favorise un plissement Nord plus fréquent de C4. Comme dans WT, les désoxythymidines T3, T6 et T7 sont plus souvent en Nord que la plupart des autres nucléotides qui sont stabilisés en Sud.

**Tableau 5-** Les pourcentages des plissements Nord (N), Sud (S) et Est (E) adoptés par les 12 premiers nucléotides des brins a et b de [RA/T] au cours des 1,5 dernières ns de la dynamique moléculaire

Brin b	% en N	% en E	% en S	Brin a	% en N	% en E	% en S
<b>5' A12</b>	2.5	8.0	82.5	<b>3' T12</b>	0.2	8.8	91.0
<b>T11</b>	0.3	8.2	91.5	<b>A11</b>	0.1	6.0	93.8
<b>C10</b>	1.1	15.7	83.2	<b>G10</b>	0.3	8.7	91.0
<b>T9</b>	0.2	16.4	83.4	<b>A9</b>	0.4	9.9	89.7
<b>C8</b>	1.4	18.5	80.1	<b>G8</b>	0.2	7.8	92.0
<b>T7</b>	<b>7.8</b>	24.4	67.8	<b>A7</b>	1.5	12.5	85.9
<b>A6</b>	1.5	12.3	86.2	<b>T6</b>	<b>7.7</b>	19.2	73.1
<b>G5</b>	0.4	11.8	87.8	<b>C5</b>	2.0	16.3	81.7
<b>C4</b>	<b>14.0</b>	27.1	58.9	<b>G4</b>	1.3	9.7	89.0
<b>RA3</b>	<b>93.2</b>	6.8	0.0	<b>T3</b>	<b>8.3</b>	26.0	65.7
<b>G2</b>	4.9	1.9	93.2	<b>C2</b>	1.9	23.6	74.5
<b>3' T1</b>	4.4	14.7	80.8	<b>5' A1</b>	2.3	3.9	93.8

Les 21 paires de bases de [RA/T] présentent des appariements stables de Watson-Crick pendant toute la durée de la simulation. Les valeurs moyennes des angles dièdres du squelette sucre-phosphate et de l'angle glycosidique de RA3 sont dans leur gamme de valeurs modales de l'ADN A (Berman *et al.*, 1992). Pour les autres nucléotides, la majorité des angles dièdres du squelette sucre-phosphate et les angles glycosidiques ont des valeurs moyennes caractéristiques de l'ADN B<sub>I</sub> (Berman *et al.*, 1992).

Les profils des paramètres de bases et de pas (**figures 8-A et 9-A**) de [RA/T] sont globalement proches de ceux de WT (**figures 2-A et 3-A**). Les différences majeures observées se situent au niveau de la 3<sup>ième</sup> paire de bases. Dans [RA/T], au niveau de cette paire de bases (**figure 8-A**) :

- le buckle est négatif ( $\approx -3^\circ$ )
- l'opening est important ( $\approx +7^\circ$ )
- le propeller twist est très négatif ( $\approx -17^\circ$ )



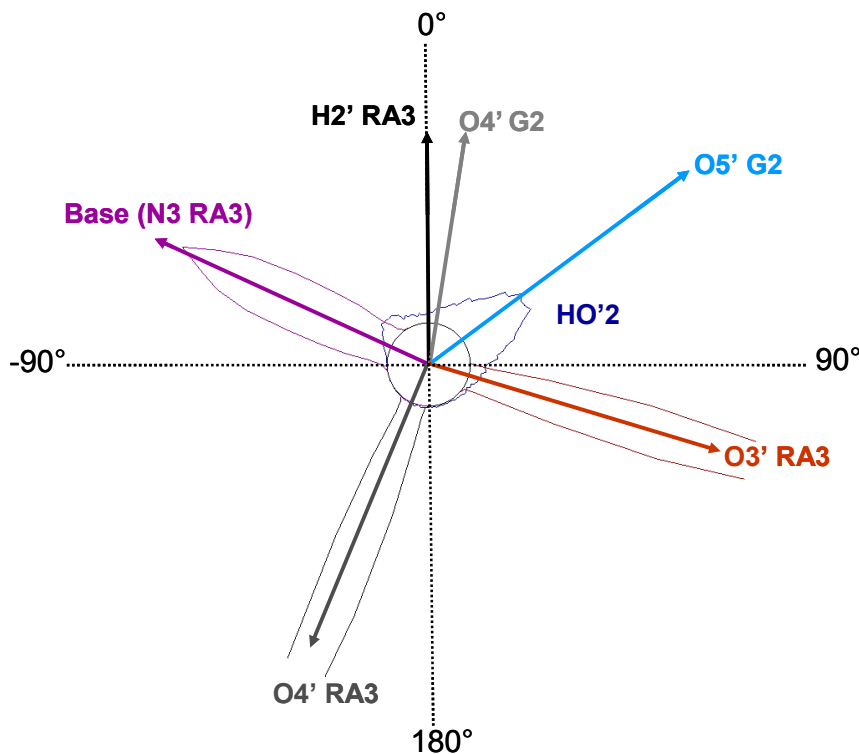
Cependant, des particularités structurales observées dans WT au niveau et au voisinage du site de clivage, se retrouvent dans [RA/T] :

- un propeller twist très négatif au niveau de la paire de bases 1 ( $\approx -16^\circ$ ) (**figure 8-A**).
- un roll important au niveau des pas C4-RA3/G4-T3 ( $\approx +10^\circ$ ) et T7-A6/A7-T6 ( $\approx +8^\circ$ ) (**figure 9-A**).
- un twist faible au niveau des pas C4-RA3/G4-T3 ( $\approx +30^\circ$ ) et T7-A6/A7-T6 ( $\approx +28^\circ$ ) (**figure 9-A**).

Les oligonucléotides WT et [RA/T] ont donc les mêmes caractéristiques structurales de type A du roll et du twist au niveau des pas C4-RA3/G4-T3 et T7-A6/A7-T6.

La présence d'un ribonucléotide en Nord, à la place d'A3 du WT, ne modifie pas significativement la structure d'ensemble de l'oligonucléotide mais seulement localement. En particulier, la présence de RA3, accentue la transformation conformationnelle B  $\rightarrow$  A au niveau du pas C4-RA3/G4-T3.

A partir de la trajectoire simulée de [RA/T], nous avons calculé la distribution de l'angle dièdre  $\lambda$  (HO'2-O2'-C2'-H2') de RA3. L'HO'2 de RA3 est orienté vers l'O5' de G2 ou vers le N3 de la base de RA3 (**figure 6**).



**Figure 6-** Orientations du HO'2 de RA3 qui est en Nord dans [RA/T].



*L'oligonucléotide [A/RU]*

L'oligonucléotide [A/RU] dans lequel T3 du brin a de WT a été remplacé par RU3, présente après 2,5 ns de dynamique moléculaire une structure en double hélice uniquement sur 20 paires de bases. La paire de bases T1/A1 est fondue (**tableau 6**) alors que les 20 autres sont stabilisées dans des appariements de Watson-Crick et elles forment une double hélice ayant une forme générale d'un ADN B<sub>I</sub>. Le ribonucléotide RU3 est stabilisé en Sud durant les 1,5 dernières ns de dynamique moléculaire (**tableau 7**). Comme dans WT, T6 et T7 ont tendance à être plus fréquemment en Nord que la plupart des autres nucléotides.

La majorité des angles dièdres du squelette sucre-phosphate et des angles glycosidiques sont dans leur gamme de valeurs modales observées dans les doubles hélices de forme B<sub>I</sub> (Berman *et al.*, 1992) durant toute la simulation.

**Tableau 6-** Les distances moyennes entre les hétéroatomes normalement impliqués dans les deux liaisons hydrogènes de la paire T1/A1, mesurées au cours des 1,5 dernières ns de la dynamique moléculaire de [A/RU].

Distances mesurées	Distances moyennes (Å)	Fluctuations (Å)	Distances attendues (Å)
O4 T1 – N6 A1	5.15	2.33	2.95
N3 T1 – N1 A1	4.95	1.70	2.82

**Tableau 7-** Les pourcentages des plissements Nord (N), Sud (S) et Est (E) adoptés par chaque nucléotide de [A/RU] au cours des 1,5 dernières ns de la dynamique moléculaire.

Brin b	% en N	% en E	% en S	Brin a	% en N	% en E	% en S
<b>5' A12</b>	1.4	9.5	89.1	<b>3' T12</b>	0.2	11.0	88.8
<b>T11</b>	1.8	9.5	88.7	<b>A11</b>	0.1	8.7	91.2
<b>C10</b>	2.9	14.5	82.6	<b>G10</b>	0.2	6.3	93.4
<b>T9</b>	0.8	17.7	81.5	<b>A9</b>	5.9	10.4	83.8
<b>C8</b>	1.5	19.2	79.3	<b>G8</b>	0.9	6.7	92.4
<b>T7</b>	<b>8.5</b>	33.5	58.0	<b>A7</b>	0.9	6.7	92.4
<b>A6</b>	1.8	7.0	91.2	<b>T6</b>	<b>6.9</b>	22.0	71.1
<b>G5</b>	0.0	3.2	96.7	<b>C5</b>	1.1	13.1	85.9
<b>C4</b>	0.7	17.0	82.3	<b>G4</b>	0.0	6.1	93.8
<b>A3</b>	0.4	13.5	86.0	<b>RU3</b>	3.2	0.0	<b>96.7</b>
<b>3' G2</b>	0.0	2.2	97.7	<b>5' C2</b>	0.0	0.5	99.2

Les profils des paramètres de bases et de pas (**figures 8-B et 9-B**) des nucléotides en double hélice (2 à 21) de [A/RU] sont globalement proches de ceux de WT (**figures 2-A et 3-A**). Plusieurs particularités structurales observées dans WT au niveau et au voisinage du site de clivage se retrouvent dans [A/RU] :



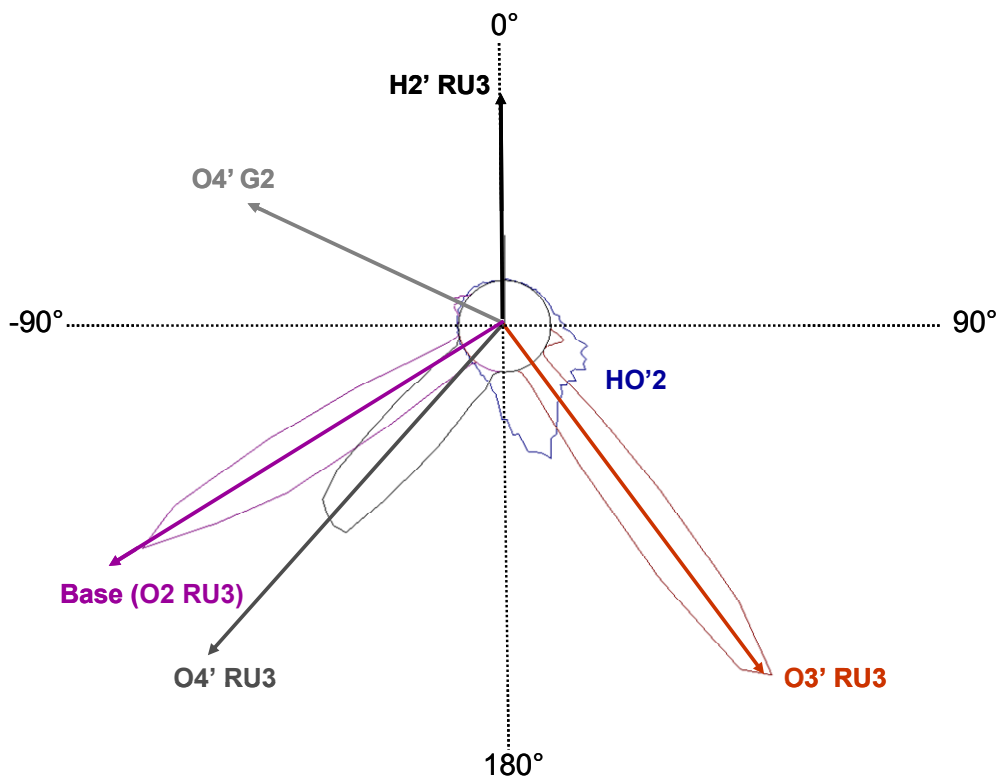


- un buckle négatif au niveau des paires de bases 7 à 9 ( $\approx -8^\circ$ ) (**figure 8-B**).
- un roll important au niveau des pas C4-A3/G4-RU3 ( $\approx +9^\circ$ ) et T7-A6/A7-T6 ( $\approx +9^\circ$ ) (**figure 9-B**).
- un twist faible au niveau du pas T7-A6/A7-T6 ( $\approx +26^\circ$ ) (**figure 9-B**).

Le pas T7-A6/A7-T6 a donc un roll et un twist caractéristiques d'une conformation de type A comme dans WT. Par contre, au niveau du pas C4-A3/G4-RU3, seul le roll est caractéristique de l'ADN A. Deux autres différences entre [A/RU] et WT peuvent être notées:

- le buckle qui est légèrement négatif ( $\approx -3^\circ$ ) au niveau d'A3/RU3 alors qu'il est positif ( $\approx +8^\circ$ ) au niveau de la paire A3/T3 de WT.
- le propeller twist de la paire de bases 2 qui est dans sa gamme de valeurs modales d'un ADN B (Berman *et al.*, 1992). alors qu'il est très négatif ( $\approx -17^\circ$ ) dans WT

A partir de la trajectoire simulée de [A/RU], nous avons calculé la distribution de l'angle dièdre  $\lambda$  (HO'2-O2'-C2'-H2') de RU3. L'HO'2 de RU3 est orienté vers l'O3' (**figure 7**).



**Figure 7-** Orientation du HO'2 de RU qui est en Sud dans [A/RU].



*L'oligonucléotide [RA/RU]*

L'oligonucléotide [RA/RU] dans lequel A3 et T3 de WT ont été remplacés par les ribonucléotides RA3 et RU3, présente après 2,5 ns de dynamique moléculaire une structure en double hélice uniquement sur 20 paires de bases. La paire de bases RA3/RU3 est fondue (**tableau 8**) alors que les 20 autres sont stabilisées dans des appariements de Watson-Crick et elles sont organisées en une double hélice ayant une forme générale d'un ADN B<sub>I</sub>. Les ribonucléotides RA3 et RU3 sont stabilisés en Nord durant la dynamique moléculaire (**tableau 9**). Comme dans WT, T6 et T7 ont tendance à être plus fréquemment en Nord que la plupart des autres nucléotides.

**Tableau 8-** Les distances moyennes entre les hétéroatomes normalement impliqués dans les deux liaisons hydrogènes de la troisième paire de bases, mesurées au cours des 1,5 dernières ns de la dynamique moléculaire de [RA/RU].

Distances mesurées	Distances moyennes (Å)	Fluctuations (Å)	Distances attendues (Å)
N6 RA3 – O4 RU3	6.50	1.45	2.95
N1 RA3 – N3 RU3	4.29	0.77	2.82

**Tableau 9-** Les pourcentages des plissements Nord (N), Sud (S) et Est (E) adoptés par les 12 premiers nucléotides des brins a et b de [RA/RU] au cours des 1,5 dernières ns de la dynamique moléculaire.

Brin b	% en N	% en E	% en S	Brin a	% en N	% en E	% en S
<b>5' A12</b>	1.3	7.5	91.2	<b>3' T12</b>	0.1	10.7	89.2
<b>T11</b>	1.3	11.9	86.8	<b>A11</b>	0.2	9.5	90.2
<b>C10</b>	0.3	13.9	85.8	<b>G10</b>	0.1	7.0	92.9
<b>T9</b>	1.0	19.2	79.8	<b>A9</b>	0.8	8.7	90.5
<b>C8</b>	6.8	33.5	59.8	<b>G8</b>	0.0	2.7	97.3
<b>T7</b>	<b>9.7</b>	30.0	61.3	<b>A7</b>	0.9	7.4	91.7
<b>A6</b>	3.6	8.9	87.5	<b>T6</b>	<b>6.0</b>	26.4	67.6
<b>G5</b>	0.4	7.7	91.9	<b>C5</b>	1.5	13.7	84.9
<b>C4</b>	1.6	12.9	85.5	<b>G4</b>	1.5	2.2	96.2
<b>RA3</b>	<b>94.2</b>	5.7	0.0	<b>RU3</b>	<b>90.6</b>	9.4	0.0
<b>G2</b>	0.0	1.6	98.3	<b>C2</b>	0.8	9.6	89.7
<b>3' T1</b>	4.7	21.6	73.7	<b>5' A1</b>	4.0	4.4	91.5

La majorité des angles dièdres de squelette sucre-phosphate et les angles glycosidiques des nucléotides en double hélice sont dans leur gamme de valeurs modales observées dans les doubles hélices de forme B<sub>I</sub> (Berman *et al.*, 1992) durant toute la simulation.

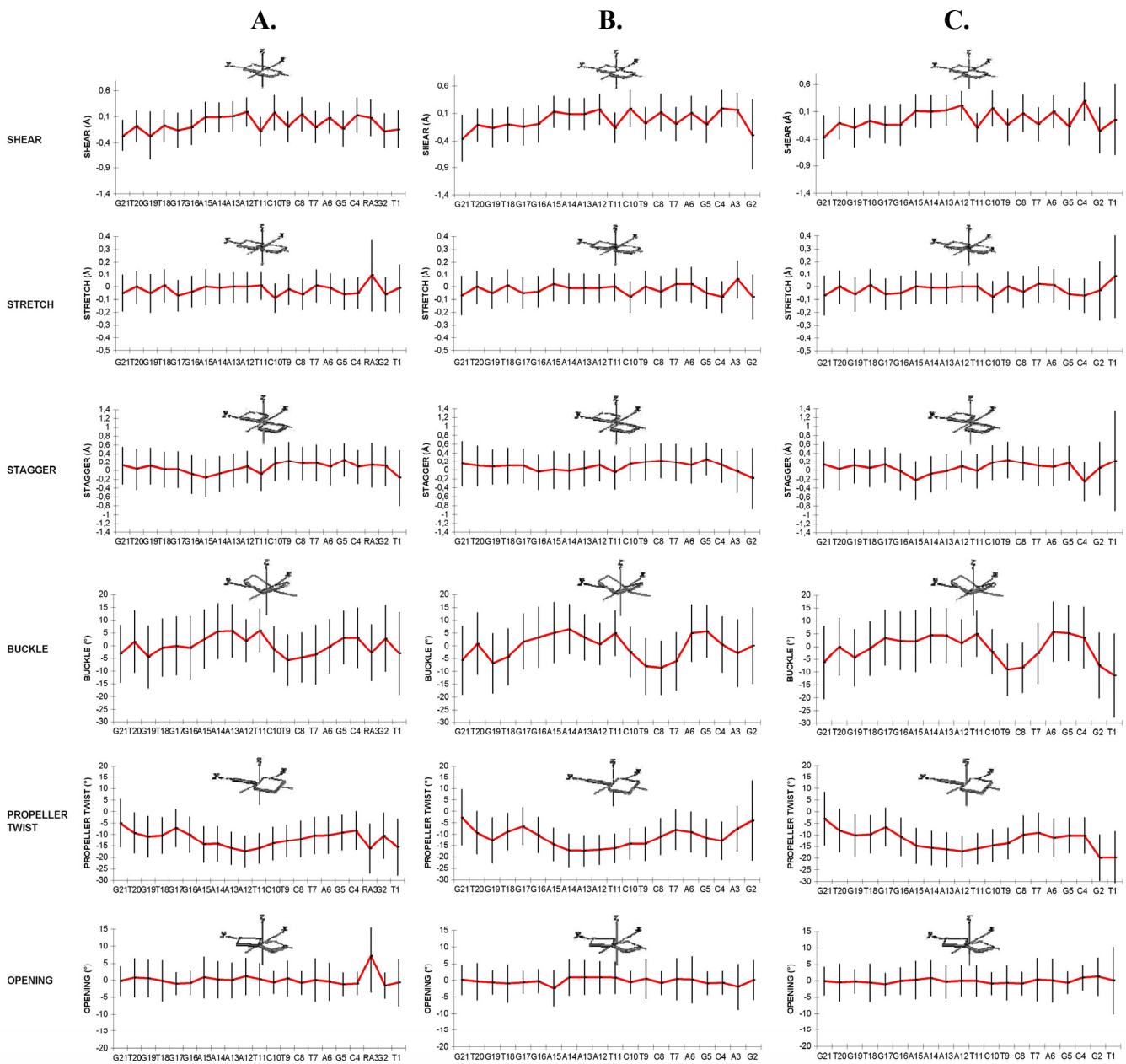


Les paramètres de bases et de pas des nucléotides en double hélice (nucléotides 1 à 2 et 4 à 21) ont des profils globalement proches de ceux de WT (**figures 8-C et 9-C**). Des particularités structurales observées au voisinage du site de clivage dans WT sont retrouvées dans [RA/RU] :

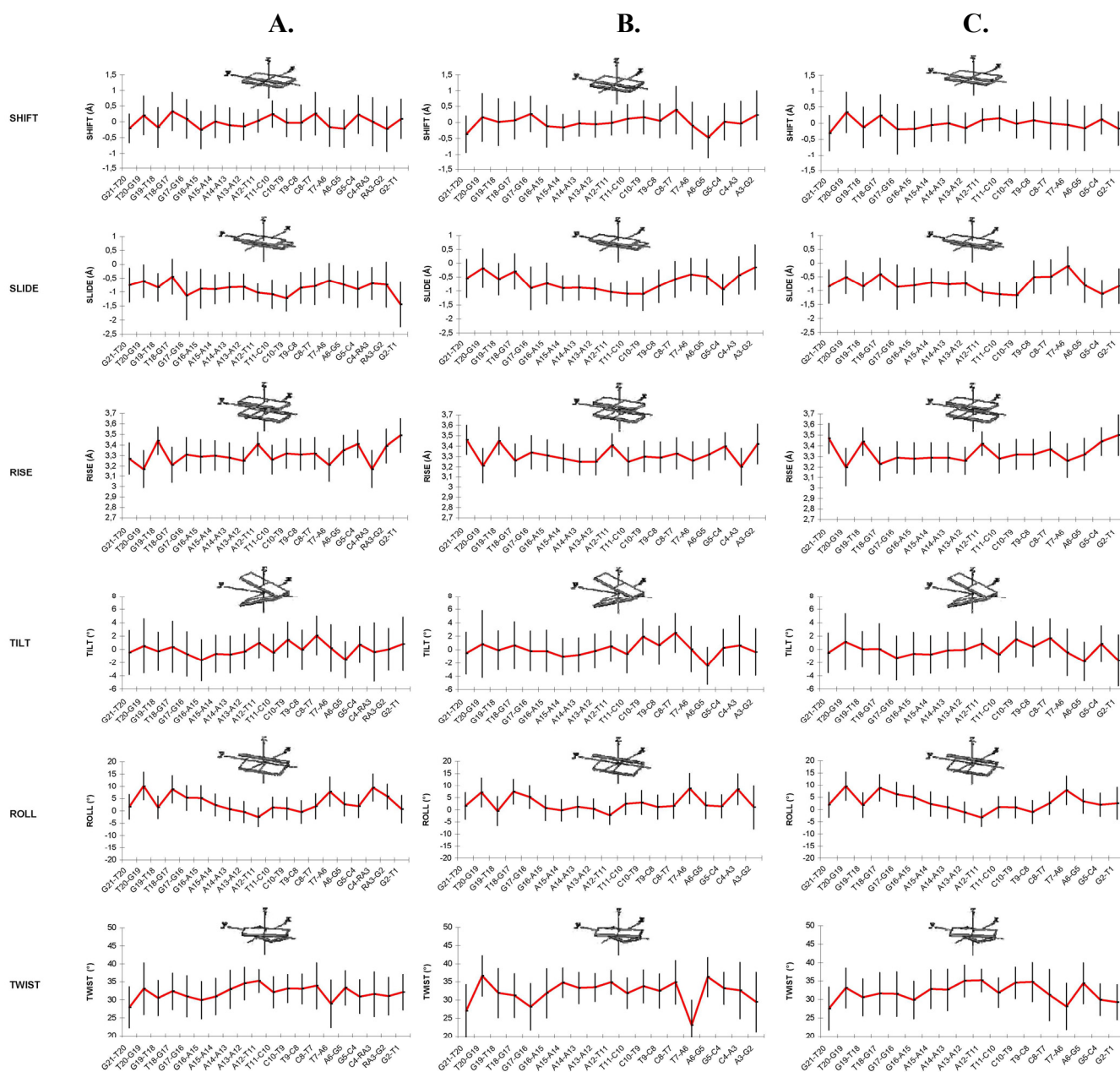
- un buckle très négatif au niveau des paires de bases 1 ( $\approx -11^\circ$ ), 8 ( $\approx -8^\circ$ ) et 9 ( $\approx -9^\circ$ ) (**figure 8-C**).
- un propeller twist très négatif au niveau des paires de bases 1 et 2 ( $\approx -20^\circ$ ) (**figure 8-C**).
- un roll important au niveau du pas T7-A6/A7-T6 ( $\approx +8^\circ$ ) (**figure 9-C**).
- un twist faible au niveau du pas T7-A6/A7-T6 ( $\approx +28^\circ$ ) (**figure 9-C**).

Le roll et le twist au niveau du pas T7-A6/A7-T6 sont donc caractéristiques de l'ADN de forme A comme dans WT et sont associés à un plissement Nord plus fréquent de T6 et T7 (**tableau 9**).

A partir de la trajectoire simulée de [RA/RU], nous avons calculé la distribution de l'angle dièdre  $\lambda$  de RA3 et de RU3 qui sont en Nord. L'HO'2 de RA3 est orienté vers l'O5' de G2 ou vers le N3 de RA3. L'HO'2 de RU est orienté vers l'O5' de G4 ou vers l'O2 de RU3.



**Figure 8-** Les paramètres de bases moyens (calculés avec Morcad) des 1,5 dernières ns de dynamique moléculaire des oligonucléotides [RA/T] (A.), [A/RU] (B.) et [RA/RU] (C.). Pour l'oligonucléotide [A/RU], la paire de bases T1/A1 étant fondue, les paramètres impliquant T1 et A1 ne sont pas représentés. Pour l'oligonucléotide [RA/RU], la paire de bases RA3/RU3 étant fondue, les paramètres impliquant RA3 et RU3 ne sont pas représentés. Les barres verticales indiquent la fluctuation de ces paramètres.



**Figure 9-** Les paramètres de pas moyens (calculés avec Morcad) des 1,5 dernières ns de dynamique moléculaire des oligonucléotides [RA/T] (A.), [A/RU] (B.) et [RA/RU] (C.). Pour l'oligonucléotide [A/RU], la paire de bases T1/A1 étant fondue, les paramètres impliquant T1 et A1 ne sont pas représentés. Pour l'oligonucléotide [RA/RU], la paire de bases RA3/RU3 étant fondue, les paramètres impliquant RA3 et RU3 ne sont pas représentés. Les barres verticales indiquent la fluctuation de ces paramètres.





### L'oligonucléotide [ARA/T]

L'oligonucléotide [ARA/T] dans lequel A3 du brin a de WT a été remplacé par ARA3, présente après 2,5 ns de dynamique moléculaire une structure en double hélice, uniquement sur 19 paires de bases. Les paires de bases T1/A1 et G2/C2 sont fondues (**tableau 10**) alors que les 19 autres sont stabilisées dans des appariements de Watson-Crick ; elles forment une double hélice ayant une forme générale d'un ADN B<sub>I</sub>. L'arabinonucléotide ARA3 est stabilisé en Sud durant les 1,5 dernières ns de dynamique moléculaire (**tableau 11**). Comme dans WT, T6 et T7 ont tendance à être plus fréquemment en Nord que la plupart des autres nucléotides. Par contre, T3 est stabilisé en Sud.

Les angles dièdres du squelette sucre-phosphate et l'angle glycosidique  $\chi$  sont dans leur gamme de valeurs modales observées dans les doubles hélices de forme B<sub>I</sub> (Berman *et al.*, 1992) durant toute la simulation.

**Tableau 10-** Les distances moyennes, entre les hétéroatomes normalement impliqués dans les liaisons hydrogènes des paires de bases 1 et 2, mesurées au cours des 1,5 dernières ns de la dynamique moléculaire de [ARA/T].

Paires de bases	Distances mesurées	Distances moyennes (Å)	Fluctuations (Å)	Distances attendues (Å)
T1/A1	O4 T1 – N6 A1	9.16	2.58	2.95
	N3 T1 – N1 A1	6.76	1.91	2.82
G2/C2	O6 G2 – N4 C2	8.03	0.98	2.91
	N1 G2 – N3 C2	7.90	0.97	2.95
	N2 G2 – O2 C2	7.90	1.07	2.86

**Tableau 11-** Les pourcentages des plissements Nord (N), Sud (S) et Est (E) adoptés par les nucléotides 3 à 12 des brins a et b de [ARA/T] au cours des 1,5 dernières ns de la dynamique moléculaire.

Brin b	% en N	% en E	% en S	Brin a	% en N	% en E	% en S
<b>5' A12</b>	1.8	11.6	86.7	<b>3' T12</b>	2.8	14.2	83.0
<b>T11</b>	1.5	16.3	82.2	<b>A11</b>	0.7	8.9	90.5
<b>C10</b>	1.8	15.5	82.7	<b>G10</b>	0.2	8.2	91.6
<b>T9</b>	0.7	14.9	84.4	<b>A9</b>	0.1	3.9	96.0
<b>C8</b>	0.7	11.6	87.7	<b>G8</b>	0.3	4.4	95.1
<b>T7</b>	<b>9.7</b>	16.7	73.6	<b>A7</b>	2.5	17.0	80.5
<b>A6</b>	0.0	0.2	99.8	<b>T6</b>	<b>11.1</b>	34.8	54.1
<b>G5</b>	5.4	5.4	89.0	<b>C5</b>	0.0	1.3	98.7
<b>C4</b>	0.2	4.2	95.6	<b>G4</b>	0.7	10.4	88.9
<b>3' ARA3</b>	0.0	2.2	97.8	<b>5' T3</b>	0.0	0.2	99.8

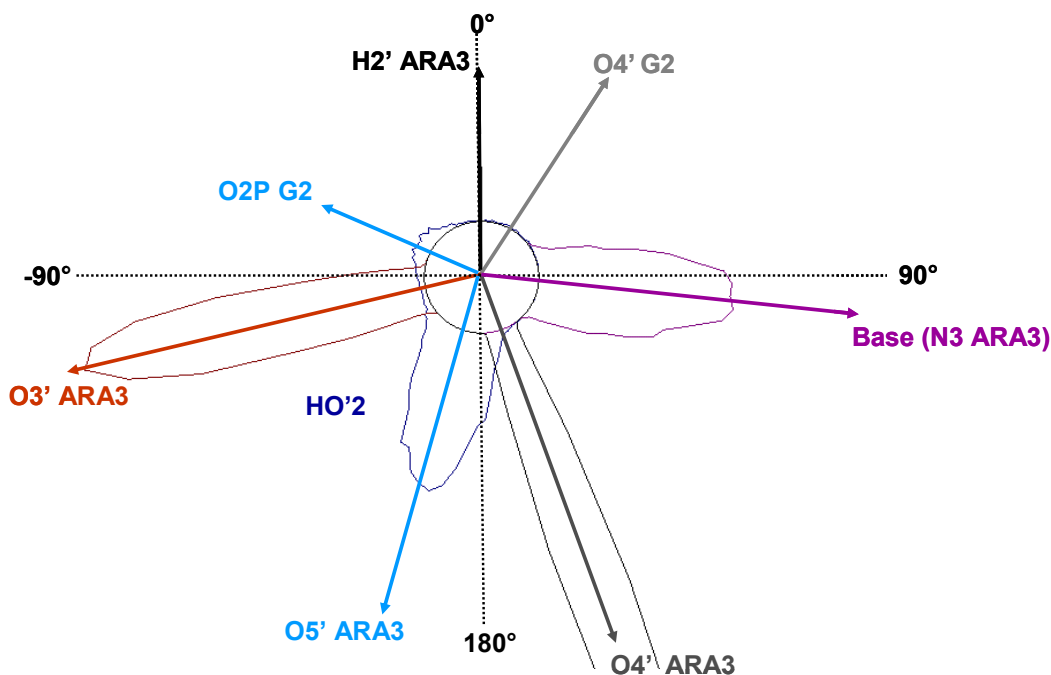


Les profils des paramètres de bases et de pas (**figure 11**) des nucléotides en double hélice (3 à 22) de [ARA/T] sont globalement proches de ceux de WT (**figures 2-A et 3-A**). Certaines particularités structurales observées dans WT au niveau et au voisinage du site de clivage se retrouvent dans [ARA/T] :

- un buckle négatif au niveau des paires de bases 7 à 8 ( $\approx -8^\circ$ )
- un roll important au niveau des pas C4-ARA3/G4-T3 ( $\approx +7^\circ$ ) et T7-A6/A7-T6 ( $\approx +8^\circ$ )
- un twist faible au niveau du pas T7-A6/A7-T6 ( $\approx +23^\circ$ ).

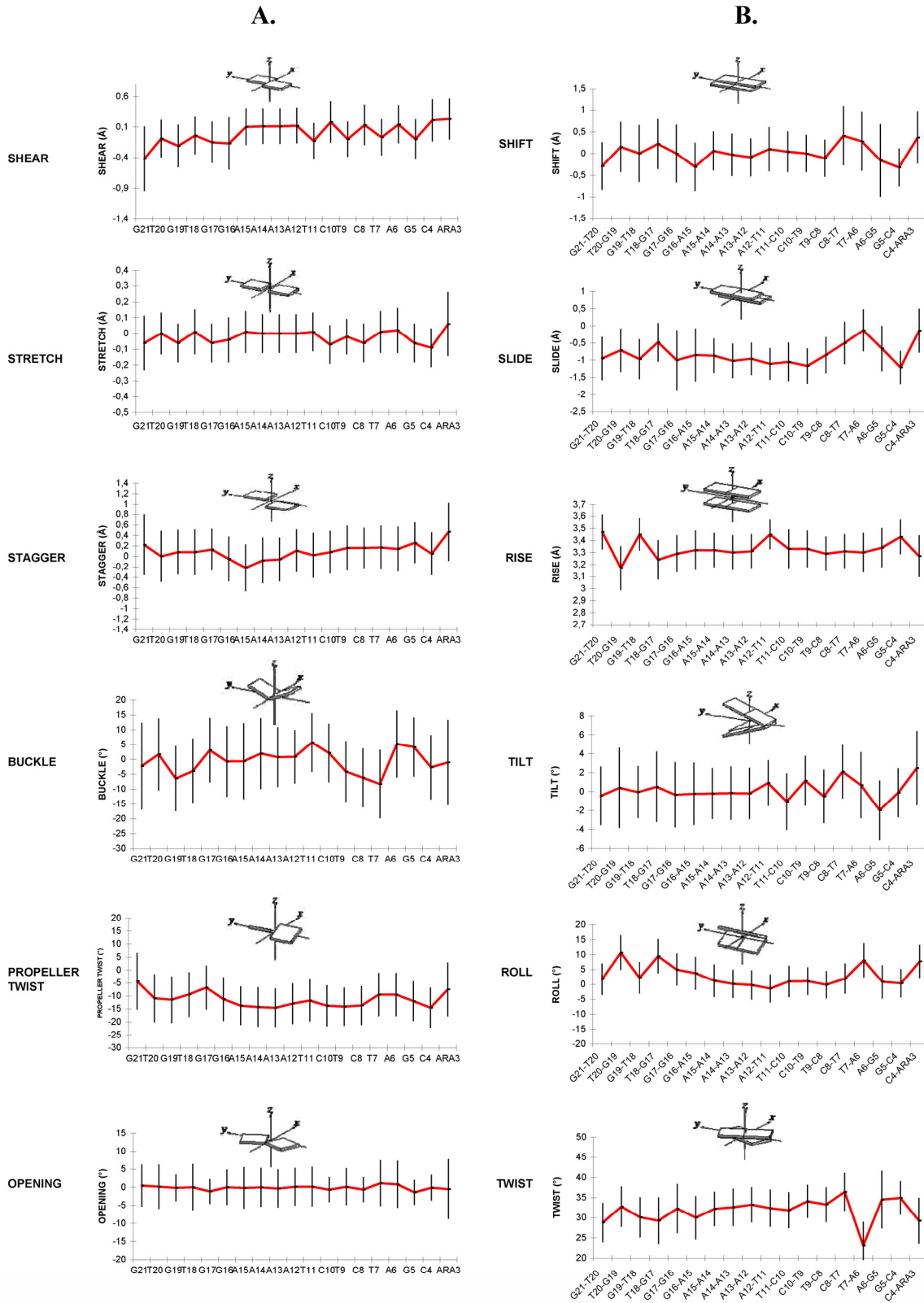
Le pas T7-A6/A7-T6 a donc un roll et un twist, caractéristiques d'une conformation de type A, comme dans WT. Par contre, au niveau du pas C4-ARA3/G4-T3, seul le roll est caractéristique de l'ADN A.

A partir de la trajectoire simulée de [ARA/T], nous avons calculé la distribution de l'angle dièdre  $\lambda$  (HO'2-O2'-C2'-H2') d'ARA3. L'HO'2 d'ARA3 est orienté vers l'O5' d'ARA3 ou vers l'O2P de G2 (**figure 10**).



**Figure 10-** Orientations du HO'2 d'ARA3 qui est en Sud dans [ARA/T].





**Figure 11-** Les paramètres de bases (A.) et de pas (B.) moyens (calculés avec Morcad) des 1,5 dernières ns de dynamique moléculaire de l'oligonucléotide [ARA/T]. Les paires de bases T1/A1 et G2/C2 étant fondues, les paramètres impliquant T1, A1, G2 et C2 ne sont pas représentés. Les barres verticales indiquent la fluctuation des paramètres.



## Discussion

L'oligonucléotide WT a une forme générale d'un ADN B<sub>I</sub> mais avec des déformations structurales au niveau et au voisinage du site de clivage de l'IN. En particulier, les pas C4-A3/G4-T3 et T7-A6/A7-T6 ont des roll et twist caractéristiques d'un ADN A.

Notre structure de WT est relativement proche de celle prédite à partir de MDDNA, même au niveau des bases terminales alors que l'on se serait attendu à avoir un effet de bord. La structure RMN est plus éloignée de notre modèle et elle présente davantage de particularités structurales. Il serait intéressant de faire une dynamique moléculaire à partir de cette structure afin d'évaluer sa stabilité. La tendance du pas C4-A3/G4-T3 à adopter une conformation de type A est retrouvée dans la structure RMN, comme dans notre modèle.

Le remplacement de T3 de WT par U3 dans [A/U] n'entraîne pas de modifications structurales significatives entre les deux oligonucléotides. Ceci explique pourquoi le taux initial de 3'-processing est similaire avec les deux substrats, WT et [A/U] (Mazumder *et al.*, 1995). Les particularités structurales mises en évidence dans WT et qui ne sont pas retrouvées dans [A/U] ne sont donc pas importantes pour la formation du complexe actif. Il s'agit :

- du buckle très positif au niveau de la paire de bases 3 et négatif au niveau des paires de bases 7 à 9
- du twist élevé au niveau du pas A3-G2/T3-C2

Des différences dans la structure fine de la double hélice au niveau et au voisinage du site de clivage sont observées entre les deux oligonucléotides, WT et [T/A].

Le taux initial de 3'-processing diminue énormément avec le substrat [T/A], en présence de Mn<sup>2+</sup> (Scottoline *et al.*, 1997). Cette inhibition a été initialement attribuée à l'absence de l'adénine en position 3 et donc des contacts spécifiques impliquant N7 et 6-NH2 qu'elle formerait avec l'IN. Cependant, le remplacement d'A3/T3 par T3/T3 n'a pas d'effet significatif sur le 3'-processing en présence de Mn<sup>2+</sup> (Agapkina *et al.*, 2006). Nous avons réalisé une dynamique moléculaire avec la paire de bases T/T en position 3 (données non présentées ici). Nous avons ainsi pu montrer que cette paire de bases est stabilisée par deux liaisons hydrogènes durant toute la simulation. Ces liaisons se forment entre :

- l'O4 de T3 du brin b et l'H3 de T3 du brin a
- l'H3 de T3 du brin b et l'O2 de T3 du brin a.

La paire de bases T/T a donc une structure en Wobble, W↓ (Gervais *et al.*, 1995; Trotta and Paci, 1998).





Un contact similaire à celui du 6-NH2 d'A3 et l'IN ne peut donc pas avoir lieu avec le T3 du brin b. La seule liaison hydrogène que l'IN flexible en présence de  $Mn^{2+}$  pourrait former avec le T3 se situe au niveau du O4. La géométrie de la liaison formée serait tout de même très différente de celle entre le N7 d'A3 et l'IN.

Il apparaît donc que le 3'-processing peut se produire en l'absence d'un contact impliquant le 6-NH2 d'A3. Ceci a été confirmé par de récents travaux réalisés en présence de  $Mn^{2+}$  (Johnson et al., 2006b). Dans ce cas, l'inhibition du 3'-processing mesurée avec [T/A] ne peut pas être expliquée par l'absence de l'adénine en 3 du brin b ; elle pourrait plutôt être engendrée par des différences conformationnelles. Les particularités structurales absentes dans [T/A] et présentes dans WT et [A/U] sont donc à noter. Il s'agit :

- de la tendance du pas C4-A3/G4-T3 à adopter une conformation de type A (roll et twist caractéristiques de l'ADN A et plissement Nord plus fréquent de T3)
- du propeller twist très négatif au niveau des paires de bases 1 et 2.

Ces déformations de l'extrémité LTR U5 de l'ADN viral semblent donc critiques pour la formation d'un complexe IN/ADN actif. Les caractéristiques structurales de type A du pas C4-A3/G4-T3 paraissent donc jouer un rôle important dans l'interaction, contrairement à celles du pas T7-A6/A7-T6.

Il a été observé que, dans de nombreux cas de liaison d'ADN séquence-spécifique, les perturbations de la double hélice préexistent dans l'ADN non lié mais sont accentuées par la liaison de la protéine. Un des meilleurs exemples de structures pré-organisées dans les acides nucléiques qui aident à la formation des interactions séquence-spécifique, est fourni par le site de liaison ADN de CAP (catabolite activator protein) (Dixit and Beveridge, 2005). Des études de dynamique moléculaire révèlent que 40% de la courbure de l'ADN lié à la protéine sont préexistants dans l'ADN libre et que 60% sont induits par la liaison de la protéine.

Les caractéristiques structurales de type A, au niveau du pas C4-A3/G4-T3, observées dans l'ADN viral non lié pourraient être accentuées par la liaison de la protéine : la liaison phosphodiester O3'-P serait davantage exposée à l'attaque par l'IN dans le petit sillon.

Le remplacement d'A3 par RA3 accentue la transformation B  $\rightarrow$  A du pas C4-A3/G4-T3. Des tests expérimentaux de 3'-processing avec le substrat [RA3/T] devraient permettre d'évaluer l'importance de la conformation du pas C4-A3/G4-T3 pour la formation du complexe IN/ADN actif



Les modifications d'A3 et T3 par RA3 et RU3 entraînent un désappariement de la paire de bases. Des études ont suggéré qu'une déstabilisation des appariements de la 3<sup>ième</sup> paire de bases pourrait être importante pour le 3'-processing (Scottoline *et al.*, 1997; Agapkina *et al.*, 2006). Pour évaluer cette hypothèse, la mesure du taux initial de 3'-processing avec l'oligonucléotide [RA3/RU3] sera donc intéressante.

Le résultat de la dynamique de [A3/RU3] nécessite un examen plus approfondi en raison de la conformation Sud de RU3.

Dans l'oligonucléotide [ARA3/T], ARA3 est stabilisé en Sud et l'HO'2 est préférentiellement orienté vers l'O5'. Un plissement Nord d'un arabinonucléotide, l'araC introduit dans un ADN est stabilisé par une liaison hydrogène entre l'HO'2 et l'O4' du nucléotide en 3' (Chrencik *et al.*, 2003). Il faudrait donc relancer une simulation de dynamique moléculaire en imposant une liaison entre l'HO'2 d'ARA3 et l'O4' de G2 durant les premières phases de préparation puis en relâchant progressivement la contrainte.



# ***CONCLUSION***

---



Au cours de cette thèse, nous avons utilisé des méthodes variées et complémentaires de modélisation moléculaire et de chimie computationnelle afin de prédire des modèles structuraux permettant de corroborer ou d'interpréter des phénomènes expérimentaux et de suggérer de nouvelles expériences. Nous ne disposons pas de données structurales expérimentales se rapportant à la plupart des systèmes étudiés. Afin de réaliser des modèles cohérents et fiables, nous avons utilisé des informations structurales expérimentales et/ou théoriques se rapportant à des systèmes analogues. Tout au long de ce travail, notre approche théorique est restée étroitement liée aux études expérimentales.

Nos études de docking et de dynamique moléculaire ont corroboré le mode d'inhibition des SQ montré expérimentalement. Nous avons déterminé trois clusters de localisation favorable des SQ sur le domaine central de l'IN. Le cluster 1 est le premier site de liaison des SQ qui empêche la liaison de l'ADN viral et inhibe le 3'-processing. Nous avons montré que ses interactions sont stables sur 2 ns de simulation moléculaire. Les deux autres clusters sont des candidats pour le deuxième site de liaison des SQ qui empêcherait la liaison de l'ADN cible et inhiberait ainsi le transfert de brins. Cependant, il n'est pas exclu que le deuxième site des SQ soit créé par la liaison de l'ADN viral, de manière semblable à ce qui a été décrit pour les DKA. La détermination du deuxième site de liaison nécessiterait dans ce cas la connaissance de la structure du complexe IN/ADN viral afin de pouvoir réaliser des docking des SQ sur ce complexe. Les SQ inhibant moins efficacement le transfert de brins que le 3'-processing, on s'attendrait alors à ce qu'ils se lient moins bien au complexe IN/ADN viral qu'à l'IN nue.

La prédiction de la conformation adoptée par des analogues de nucléosides dans une double hélice a nécessité beaucoup de précision dans la préparation des conformations initiales. En utilisant une approche de mécanique *ab initio*, on aurait pu s'attendre à trouver la structure adoptée par le composé en déterminant la conformation du minimum absolu. Mais, en réalité, il existe de nombreux états d'énergie très semblables. Les unités nucléosidiques des doubles





hélices d'acides nucléiques ne sont ni dans leur conformation d'énergie minimale ni contraintes ; elles se trouvent au niveau de minima secondaires. Le choix des conformations secondaires a dû être guidé par des données expérimentales ou théoriques se rapportant à des systèmes voisins. Tout d'abord, les angles dièdres du squelette ont dû être orientés dans une conformation compatible avec celle observée dans les désoxynucléotides ou ribonucléotides des acides nucléiques doubles brins. L'hydratation et l'orientation du groupe en 2' se sont également avérées critiques. La prédiction de la structure des analogues de nucléosides substitués en 2' par un groupe positionné en ribose a été réalisée par analogie avec l'orientation et l'hydratation du groupe 2'-hydroxyl des ribonucléotides. Les modèles prédits ont été confrontés aux données de  $T_m$  de fragments d'ADN et/ou d'ARN contenant ces analogues de nucléosides.

L'étude structurale réalisée sur les analogues de nucléotides nous a permis d'interpréter les résultats expérimentaux des taux initiaux de 3'-processing de manière plus critique. En confrontant les données structurales et expérimentales, nous avons pu émettre l'hypothèse qu'une conformation de type A au niveau de la 3<sup>ème</sup> paire de bases pouvait être nécessaire à la formation du complexe IN/ADN actif. Ceci nous a amenés à proposer de nouveaux analogues de l'ADN viral, à tester expérimentalement.

Les dynamiques moléculaires d'analogues oligonucléotidiques de l'extrémité LTR U5 de l'ADN viral ont nécessité, comme pour les études *ab initio*, l'apport d'informations expérimentales ou théoriques dans la préparation des conformations initiales. Pour les ribonucléotides, nous avons tenu compte de l'orientation préférentielle du 2'-hydroxyle. Concernant l'arabinonucléotide, nous avons relancé une simulation, en orientant le groupe 2'-hydroxyle comme observé pour l'araC du substrat ADN de la topoisomérase I.

En confrontant les analyses structurales des oligonucléotides WT, [A3/U3] et [T3/A3] et les données des taux initiaux de 3'-processing associés à ces substrats, nous avons pu mettre en évidence des déformations structurales intrinsèques à l'extrémité LTR U5 de l'ADN viral qui paraissent déterminantes pour que le complexe IN/ADN actif se forme et que la réaction de 3'-processing se produise. Ainsi, des caractéristiques structurales de type A du pas C4-A3/G4-T3 sont apparues importantes pour la spécificité de l'interaction IN/ADN viral. Ces résultats ont conforté notre hypothèse initialement émise à partir du regroupement des



données structurales des analogues de nucléosides et des taux initiaux de 3'-processing mesurés avec les substrats contenant ces analogues.

Ainsi, nous avons pu proposer qu'une reconnaissance indirecte soit impliquée dans l'interaction IN/ADN viral. Pour consolider ce modèle, il faudrait réaliser des études supplémentaires.

Des tests de 3'-processing avec les analogues contenant les ribonucléotides, [RA3/T3] et [A3/RU3] devraient permettre d'évaluer les effets d'une accentuation et d'une atténuation des caractéristiques structurales de type A du pas C4-A3/G4-T3.

Il serait également intéressant de réaliser des simulations moléculaires d'autres analogues de l'ADN viral dans lesquels la paire de bases A3/T3 serait remplacée par :

- C/T ou G/T, ce qui entraîne une inhibition de la réaction de 3'-processing en présence de  $Mn^{2+}$  et de  $Mg^{2+}$
- C/A ou T/G, ce qui n'affecte pas la réaction de 3'-processing en présence de  $Mn^{2+}$
- A/U<sup>N</sup> ou A/U<sup>M</sup>, ce qui stimule le 3'-processing en présence des deux cofacteurs.

Ces simulations permettraient également d'évaluer si des contacts spécifiques avec l'adénine en 3, sont réellement nécessaires pour que la réaction se produise.

La modélisation moléculaire et la chimie computationnelle ne prétendent pas apporter des solutions à tous les problèmes structuraux, surtout s'ils sont noyés dans des systèmes de plus grande importance. Mais, ce sont des outils puissants capables de mettre en évidence des phénomènes de taille atomique et de suggérer des explications aux problèmes posés. La biologie moléculaire n'est pas à même d'expliquer les phénomènes complexes survenant à l'échelle cellulaire mais elle apporte des réponses sur le matériel isolé dont elle se sert. Il en est de même des études *in silico* par rapport à la biologie moléculaire. Le rassemblement des briques élémentaires d'informations participe à l'élucidation des phénomènes survenant à l'échelle supérieure.



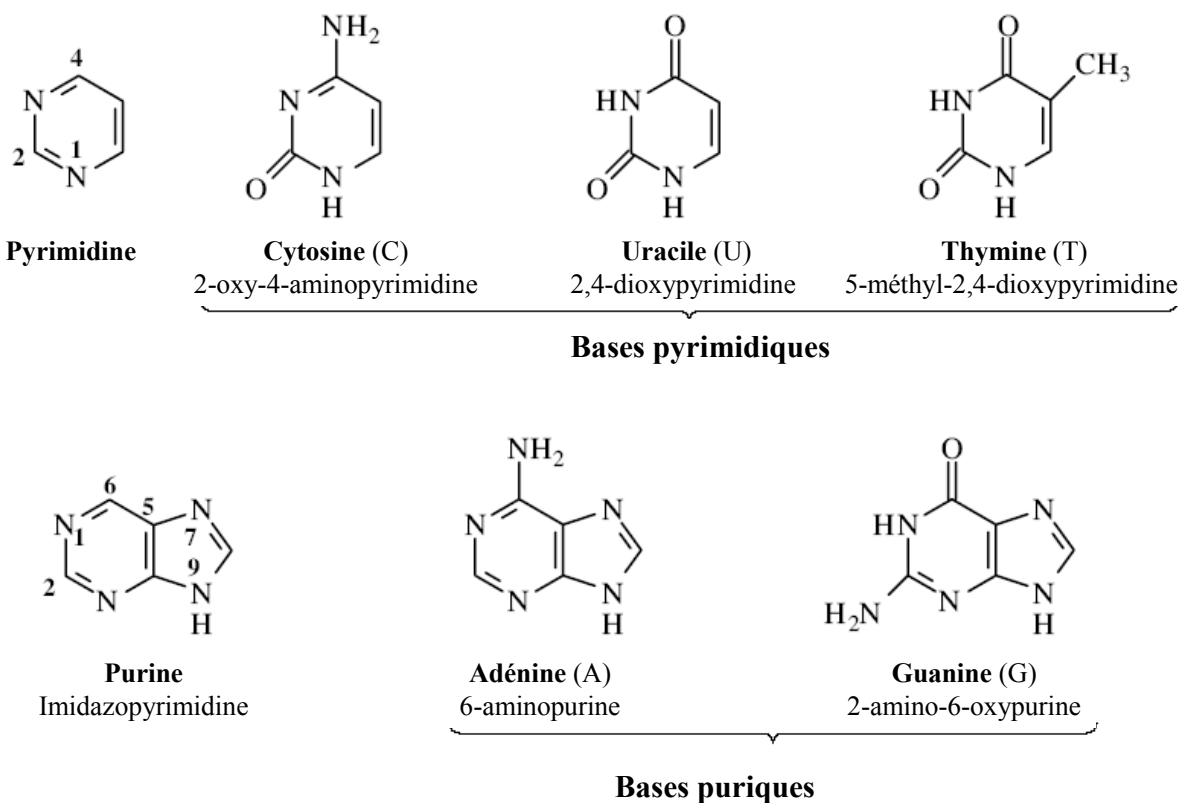
***ANNEXE***

---









**Figure 1-** Les cinq bases usuelles des acides nucléiques partagées en deux groupes : les bases pyrimidiques et les bases puriques. Les noms courants des différentes bases n'ont aucun lien avec la nomenclature classique de chimie organique, certains font référence à leurs conditions de découverte.

## STRUCTURE ET CONFORMATIONS DES DOUBLES HELICES D'ACIDES NUCLEIQUES

Depuis le modèle structural de l'ADN proposé par J. D. Watson et F. Crick (Watson and Crick, 1953), l'importance accordée aux acides nucléiques dans le monde du vivant s'est accrue et parallèlement, notre connaissance à leur sujet s'est considérablement enrichie. Il est désormais clair que la structure, la conformation et la flexibilité des acides nucléiques jouent un rôle clé dans la vie cellulaire.

L'ADN a une structure en double hélice formée par l'enroulement autour d'un même axe de deux chaînes polynucléotides antiparallèles et complémentaires. Chaque chaîne polynucléotidique est un enchaînement régulier de nucléotides. Un nucléotide résulte de :

- la condensation d'un sucre (furanose) avec une base qui produit un nucléoside
- l'estérification du sucre d'un nucléoside par un phosphate.

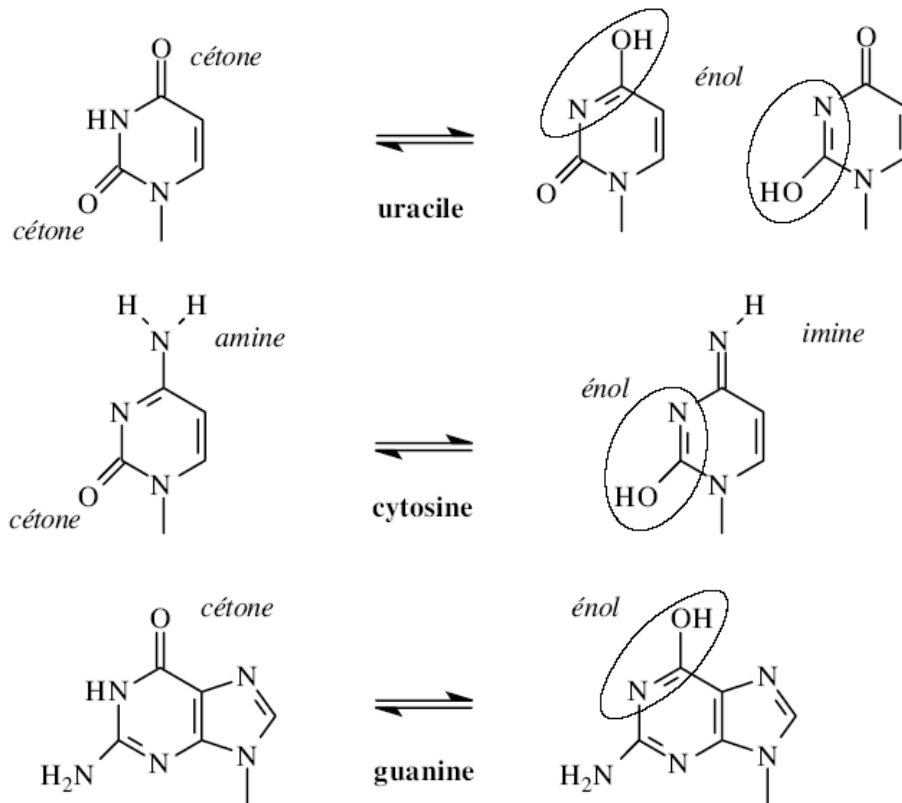
L'ARN se distingue de l'ADN par sa composition en bases et par son sucre. Sa structure secondaire se compose de « tiges » ou zones appariées formant une structure en double hélice et de « boucles » ou zones non appariées entre deux « tiges ».

L'ADN est polymorphe et existe dans diverses familles structurales, incluant les formes A, B et Z alors que les doubles hélices d'ARN se trouvent uniquement en forme A (Saenger, 1984).

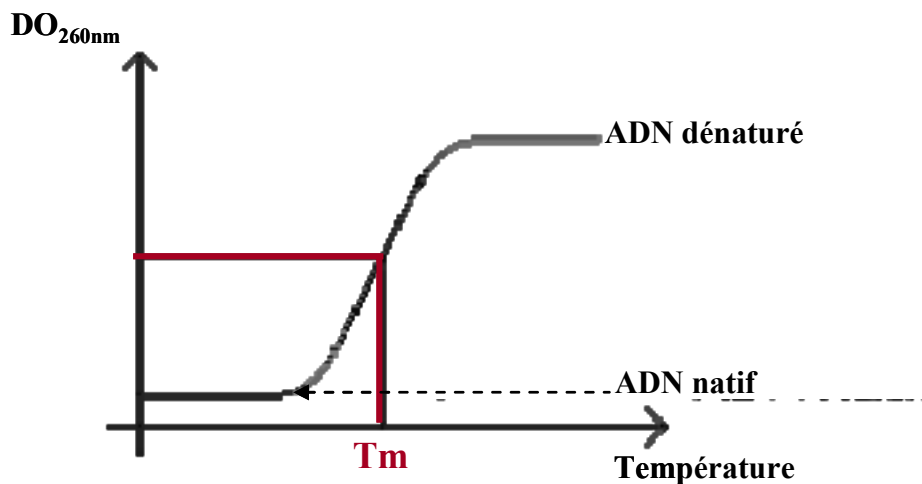
### **I- Les bases**

Trois bases pyrimidiques majeures dérivant de la pyrimidine (un cycle aromatique à 6 atomes) sont trouvées dans les acides nucléiques : la Cytosine (C), la Thymine (T) et l'Uracile (U) (**figure 1**). Deux bases puriques majeures dérivant du noyau purine (un premier cycle à six atomes et un second cycle à cinq atomes) que l'on peut considérer lui-même comme résultant de l'association d'un noyau pyrimidine avec un noyau imidazole, sont trouvées dans les acides nucléiques : l'Adénine (A) et la Guanine (G) (**figure 1**). L'ADN comme l'ARN ne comportent que quatre de ces bases :

- deux bases puriques communes aux deux types d'acides nucléiques, l'adénine et la guanine
- une base pyrimidique commune, la cytosine
- une base pyrimidique spécifique, l'uracile pour l'ARN et son dérivé méthylé, la thymine pour l'ADN.



**Figure 2-** Les formes tautomériques de différentes bases. Les formes lactame (cétone) et amine sont à gauche. Les formes lactime (énoles) et imine sont à droite.



**Figure 3-** La température de fusion ( $T_m$ ) d'un ADN (ou ARN en double hélice). Les bases puriques et pyrimidiques absorbent à 260 nm. Leurs moments de transition sont dans le plan des bases. Lorsque les bases sont parallèles, le moment de transition d'une base induit un dipôle qui diminue le moment de transition des autres bases : c'est l'effet hypochromique. Lorsque la température augmente, les bases ne sont plus parallèles, l'hypochromicité diminue et l'absorption augmente. La température de fusion ou  $T_m$  est la température où la moitié des paires de bases est dissociée.

A côté de ces bases principales, on rencontre dans certains acides nucléiques, en faible quantité, des dérivés de ces bases (en particulier des dérivés méthylés).

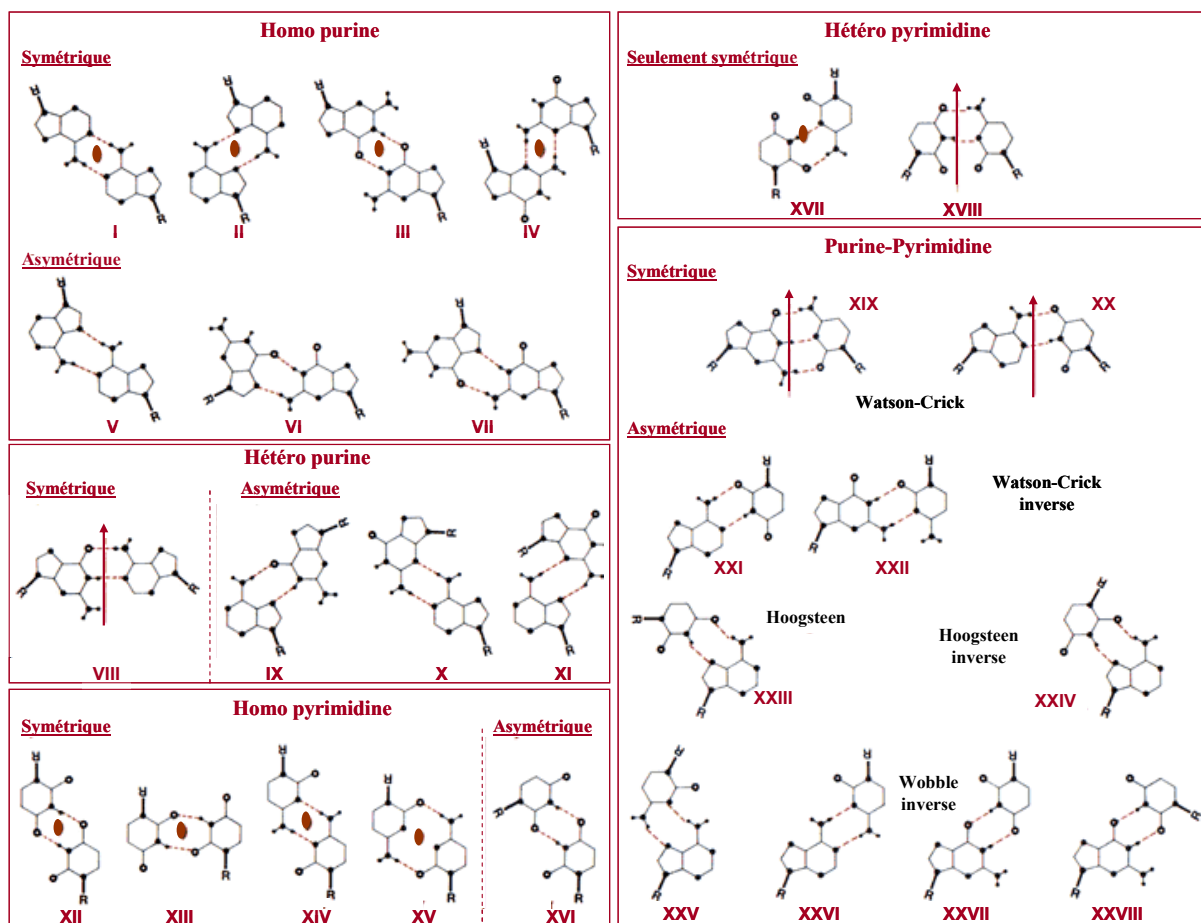
Quelle que soit l'origine de l'ADN, les rapports  $[A]/[T]$  et  $[G]/[C]$  sont approximativement égaux à 1 (Chargaff, 1951). Le nombre de purines est donc approximativement égal au nombre de pyrimidines : ( $[Pur] \approx [Pyr]$  ou encore  $[A]+[G] \approx [T]+[C]$ ). Les bases A et T (ou U) sont dites complémentaires, il en est de même pour G et C. Les proportions ( $[A]+[T]$ ) et ( $[G]+[C]$ ) sont différentes et varient dans la plupart des organismes.

En raison de leur caractère aromatique, les bases :

- sont des molécules planes.
- existent sous différentes formes tautomères (**figure 2**). Les formes lactame (cétone) et amine sont prépondérantes au pH physiologique. La population des formes lactime (énol) et imine est de  $10^{-4}$ .
- présentent des spectres d'absorption caractéristiques dans l'ultraviolet. Ceci est utilisé par exemple pour déterminer la température de fusion,  $T_m$ , d'un acide nucléique en double hélice afin de juger de sa stabilité (**figure 3**).

Les bases sont empilées les unes sur les autres le long de chacune des deux chaînes de la double hélice. Les interactions entre les systèmes conjugués des cycles aromatiques des bases d'une même chaîne stabilisent la structure hélicoïdale des acides nucléiques. L'énergie d'empilement provient essentiellement des interactions de type Van der Waals. Dans la double hélice, l'enthalpie d'empilement entre deux paires adjacentes est estimée entre -5 et -12 kcal/mol (Marky and Breslauer, 1982).

Les bases présentent des groupes accepteurs et donneurs de liaisons hydrogènes. Ces groupes permettent aux bases d'interagir entre elles par des liaisons hydrogène pour former des paires de bases. Ces interactions contribuent également à stabiliser les doubles hélices d'acides nucléiques. Des interactions entre des bases identiques (homo) et différentes (hétéro) ont été observées dans les analyses de structure cristallographique de bases, de nucléosides, de nucléotides et dans des complexes formés par plusieurs composés de ce type (Voet and Rich, 1970). Vingt huit arrangements possibles de paires de bases impliquant aux moins deux

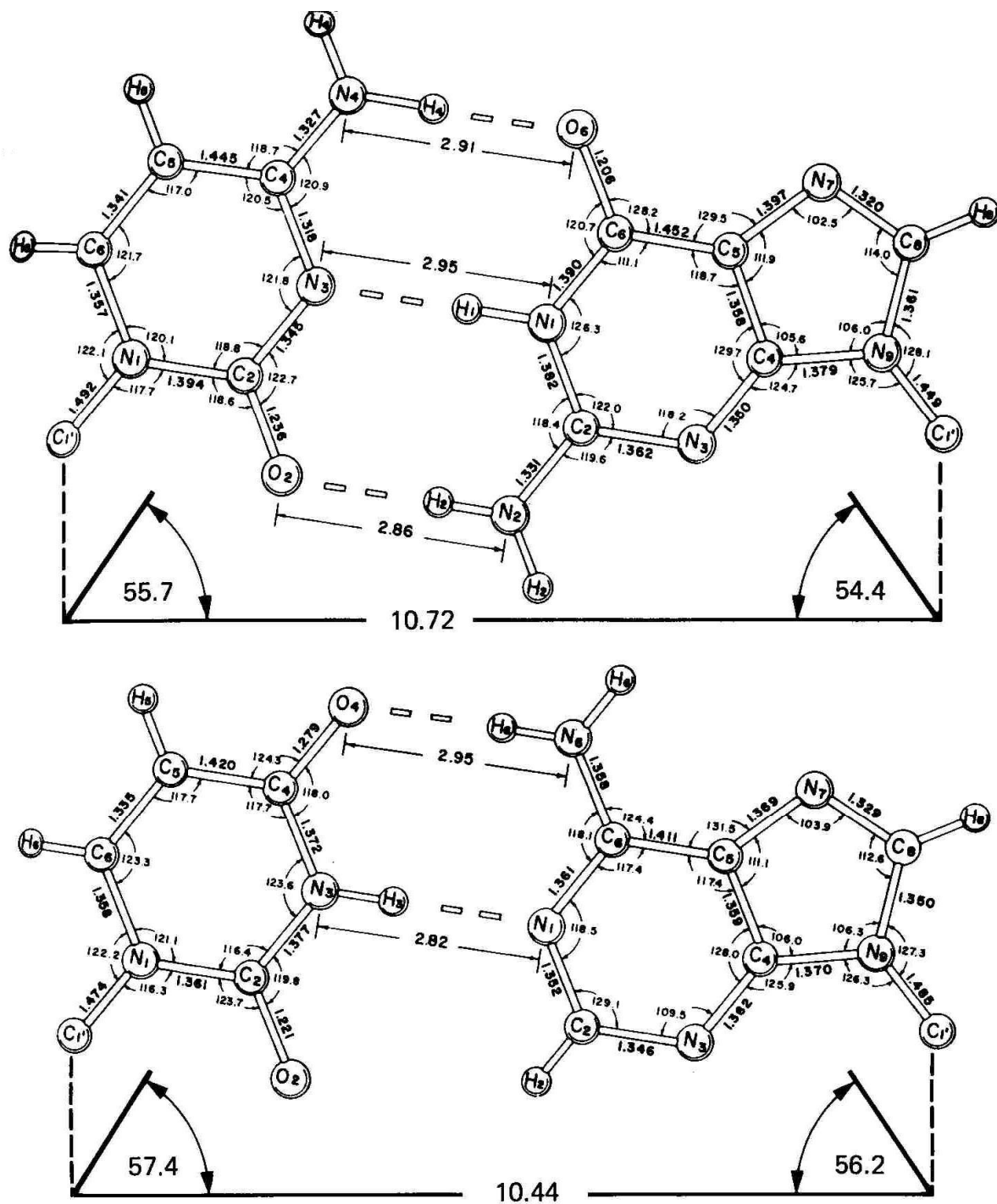


**Figure 4-** Les 28 paires de bases possibles pour A, G, U(T) et C impliquant au moins deux liaisons hydrogènes (Saenger W., 1984). Les atomes d'hydrogène et d'azote sont montrés par de petits et grands cercles remplis. Les atomes d'oxygène sont représentés par des cercles vides. La liaison glycosidique entre la base (N1 des pyrimidines et N9 des purines et le sucre R(C1') est représentée par un trait épais. Les paires de bases sont rassemblées selon leur composition: homo purine (deux purines identiques), hétéro purine (deux purines différentes), homo pyrimidine (deux pyrimidines identiques), hétéro pyrimidine (deux pyrimidines différentes) et purine-pyrimidine (une purine et une pyrimidine). Les paires de bases sont également regroupées selon qu'elles soient symétriques ou asymétriques. Les éléments de symétrie  $\uparrow$  et  $\bullet$  sont des axes de rotation d'ordre 2 verticaux et à l'intérieur du plan des paires de bases. Une pseudosymétrie reliant seulement les liens glycosidiques mais pas les atomes des bases individuelles est observée dans les paires de bases de Watson-Crick (XIX et XX) et dans les paires de bases VIII, XVII et XVIII.

liaisons hydrogènes sont montrés en **figure 4**. Ces paires de bases sont nommées selon la classification de Saenger (Saenger, 1984). Les doubles hélices d'ADN et d'ARN sont essentiellement formées par des paires de bases de Watson-Crick (XIX et XX selon Saenger). Ce sont les seules, parmi les paires hétéro, à pouvoir former des structures en double hélice régulière car elles sont isomorphes (leur forme globale et leurs dimensions sont comparables) (**figure 5**). Ces paires de bases, A/T (ou A/U dans les doubles hélices d'ARN) et G/C sont pseudosymétriques : un axe de rotation d'ordre 2 relie les liens glycosidiques (lien entre la base (N1 des pyrimidines et N9 des purines) et le C1' du sucre (R dans la figure 4)) mais pas les atomes des bases individuelles. Les bases, dans ces paires de Watson-Crick, ne sont pas coplanaires. Elles sont reliées par des liaisons hydrogène entre les groupes donneurs (N6 de l'adénine, N3 de la thymine (ou de l'uracile), N1 et N2 de la guanine et N4 de la cytosine) et les groupes accepteurs (N1 de l'adénine, O4 de la thymine (ou de l'uracile), O6 de la guanine et N3 et O2 de la cytosine). La paire de bases A/T (ou A/U) implique deux liaisons hydrogènes alors que trois liaisons sont présentes dans la paire de bases G/C qui est donc plus stable. Ainsi, plus une double hélice est riche en G/C, plus elle est stable et plus son  $T_m$  est élevé.

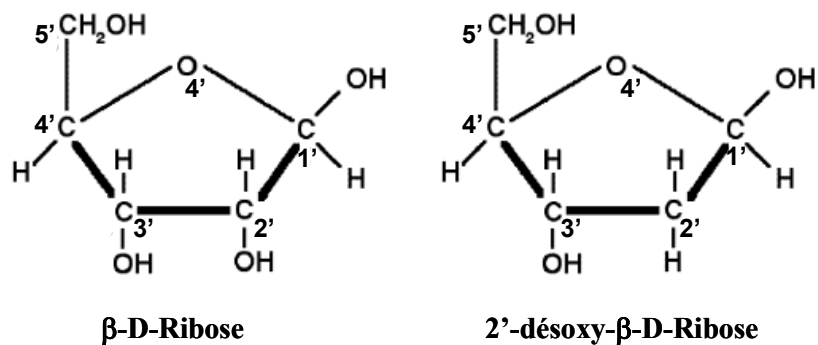
Il faut noter que dans les solvants polaires, les molécules d'eau rentrent en compétition avec les bases pour la formation des liaisons hydrogènes avec les bases complémentaires, réduisant ainsi l'énergie libre d'appariement entre les paires de bases. Les expériences thermodynamiques effectuées dans le chloroforme et dans le diméthyl-sulfoxyde estiment aux alentours de -2 à -5 kcal/mol l'enthalpie d'appariement entre deux bases (Newmark and Cantor, 1968; Uesugi et al., 1979). De récents calculs d'énergie libre estiment l'énergie libre d'appariement des paires A/T vers -4.3 kcal/mol et vers -5.8 kcal/mol pour les paires G/C (Arora and Schlick, 2003; Stofer et al., 1999). Ces calculs montrent que les interactions entre les bases complémentaires sont certainement plus faibles que celles de l'empilement.



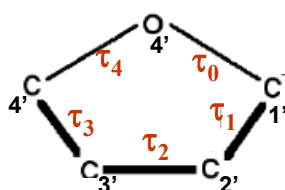


**Figure 5-** Les paires de bases de Watson-Crick observées dans les structures cristallographiques de GpC (A) et d'ApU (B) (Saenger, 1984). Les atomes d'hydrogènes n'ont pas été localisés expérimentalement mais ont été calculés à partir des positions des autres atomes.



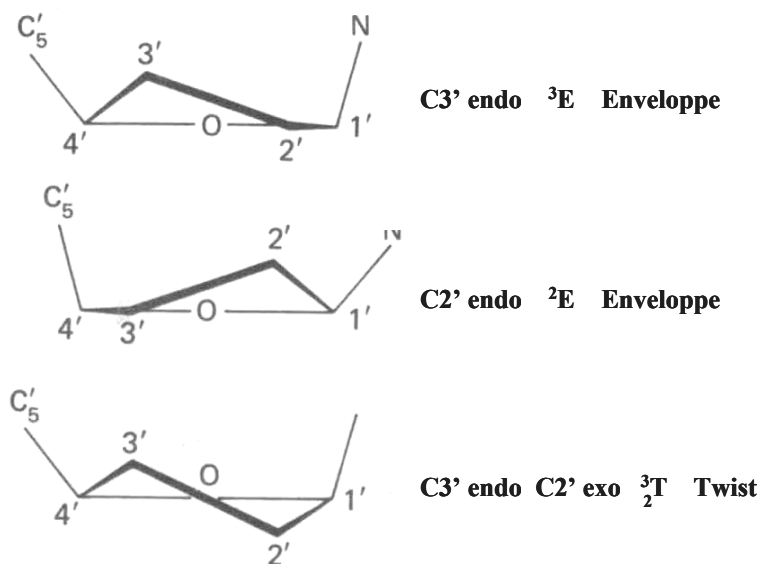


**Figure 6-** Représentation schématique des sucres : ribose (à gauche) et désoxyribose (à droite).



- $\tau_0$  (C4'-O4'-C1'-C2')
- $\tau_1$  (O4'-C1'-C2'-C3')
- $\tau_2$  (C1'-C2'-C3'-C4')
- $\tau_3$  (C2'-C3'-C4'-O4')
- $\tau_4$  (C3'-C4'-O4'-C1')

**Figure 7-** Définition des cinq angles dièdres endocycliques du sucre.



**Figure 8-** Exemples de plissements du sucre.

## II- Les sucres

Les sucres sont des furanoses, un  $\beta$ -D-ribose pour l'ARN et un 2'-désoxy- $\beta$ -D-ribose dans le cas de l'ADN (**figure 6**). Ils présentent donc les caractéristiques suivantes :

- l'énantiomère est de la série D
- le sucre est sous forme hémiacétalique (furanose)
- l'anomère est en conformation  $\beta$

Les cinq angles dièdres endocycliques du sucre sont notés  $\tau_0$  à  $\tau_4$  (**figure 7**).

Le cycle furanose n'est pas plan. Dans une première approximation, il peut être plissé sous la forme:

- d'une enveloppe (ou forme E) avec un atome hors du plan
- d'un twist (ou forme T) avec deux atomes adjacents hors du plan, l'un au-dessus et l'autre au-dessous du plan.

Les conformations sont généralement définies selon la direction et le numéro du ou des atomes déplacés hors du plan. Le terme endo (respectivement exo) est employé pour désigner la direction d'un atome déplacé vers le groupe exocyclique 5' (respectivement à l'opposé du groupe exocyclique 5') (**figure 8**). Comme le cycle furanose est composé de cinq atomes, on obtient un ensemble de 20 conformations : 10 formes E et 10 formes T.

Deux paramètres, l'angle de pseudorotation (P) et l'amplitude maximale du plissement  $\tau$ , définissent les cinq angles dièdres endocycliques et donc caractérisent le plissement. Selon Rao et al. (Rao et al., 1981):

$$\tau_j = \tau \cos[P + 4\pi(j-2)/5] \quad (1)$$

A partir de la mesure des angles  $\tau_j$  dans une conformation, A et B sont calculés ainsi :

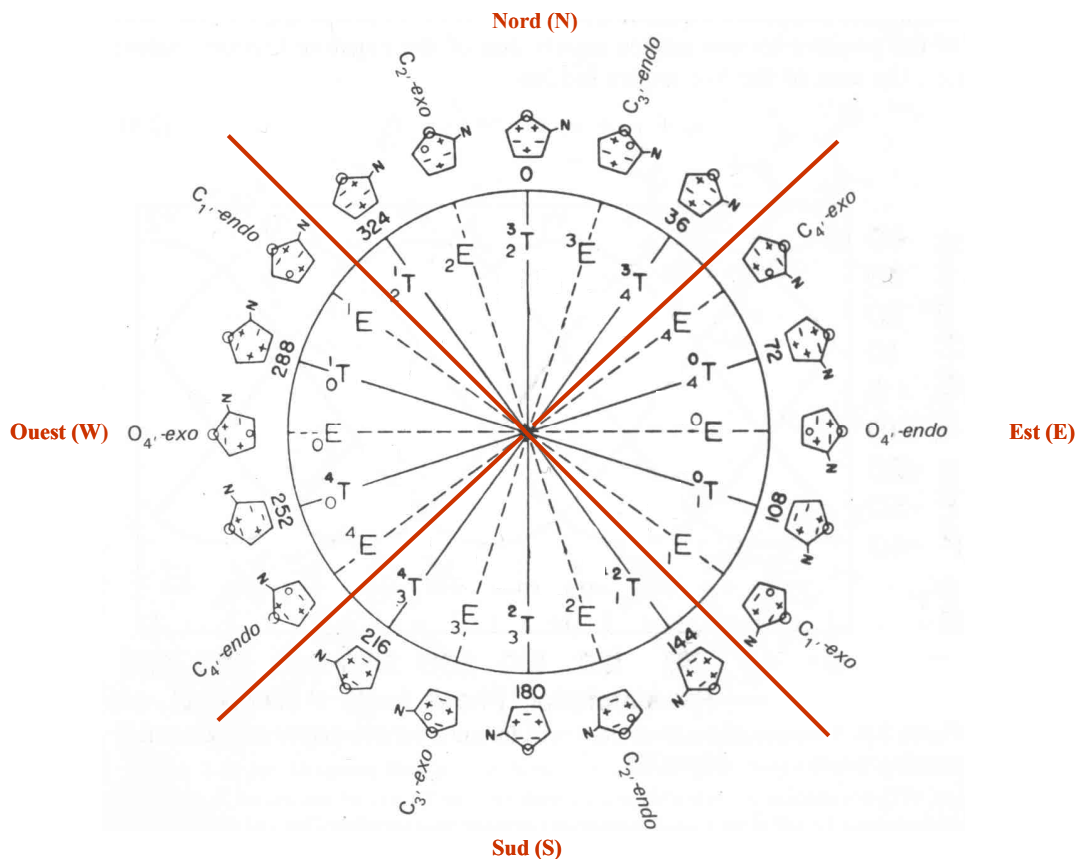
$$A = 2/5 \sum_{j=0}^4 \tau_j \cos[4\pi(j-2)/5] \quad (2)$$

$$B = -2/5 \sum_{j=0}^4 \tau_j \sin[4\pi(j-2)/5] \quad (3)$$

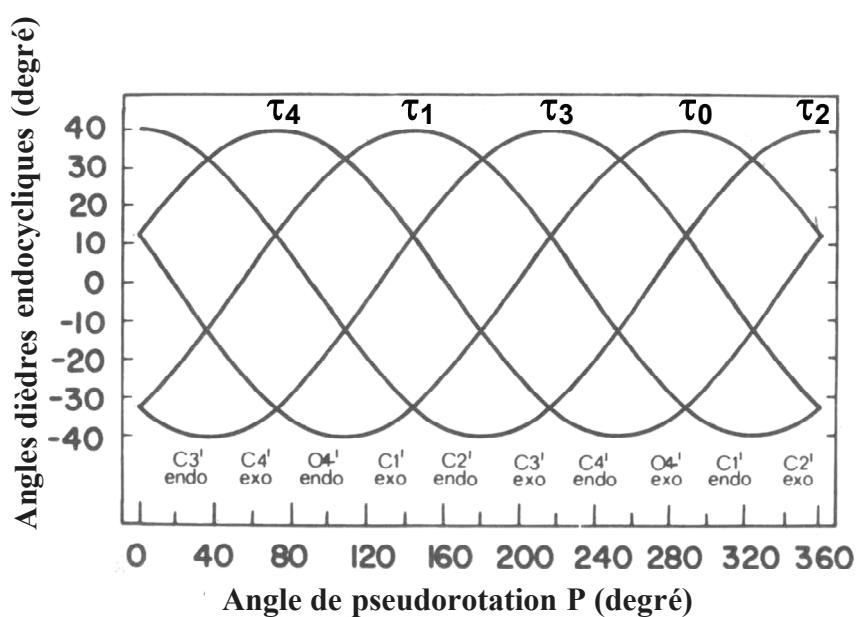
L'amplitude maximale  $\tau$  et l'angle de pseudorotation P sont alors donnés par :

$$\tau = (A^2 + B^2)^{1/2} \quad (4)$$

$$P = \tan^{-1} B/A \quad (5)$$



**Figure 9-** Cycle de pseudorotation indiquant la relation entre l'angle de pseudorotation P et les formes, enveloppe E et twist T (Altona and Sundaralingam, 1972). Les formes E et T s'alternent tous les 18°. A la périphérie du cycle, les sucres avec les signes des angles dièdres endocycliques sont indiqués : (+) positif, (-) négatif et (0) pour 0°. L'espace de pseudorotation est divisé en quatre cadrans de taille égale centrés autour de P=0°, P=90°, P=180° et P=270° qui sont respectivement appelés Nord (N), Est (E), Sud (S) et Ouest (W).



**Figure 10-** Les changements théoriques des cinq angles dièdres endocycliques du furanose au cours d'un cycle complet de pseudorotation (Levitt M. and Warshel A., 1978).

Les formes T et E sont représentées dans l'espace de pseudorotation ( $\tau$ , P) sur la **figure 9**.

La **figure 10** indique les changements théoriques des cinq angles dièdres endocycliques durant un cycle complet de pseudorotation (Levitt and Warshel, 1978). Pour chaque angle de pseudorotation, la somme des angles dièdres positifs est égale à la somme des angles

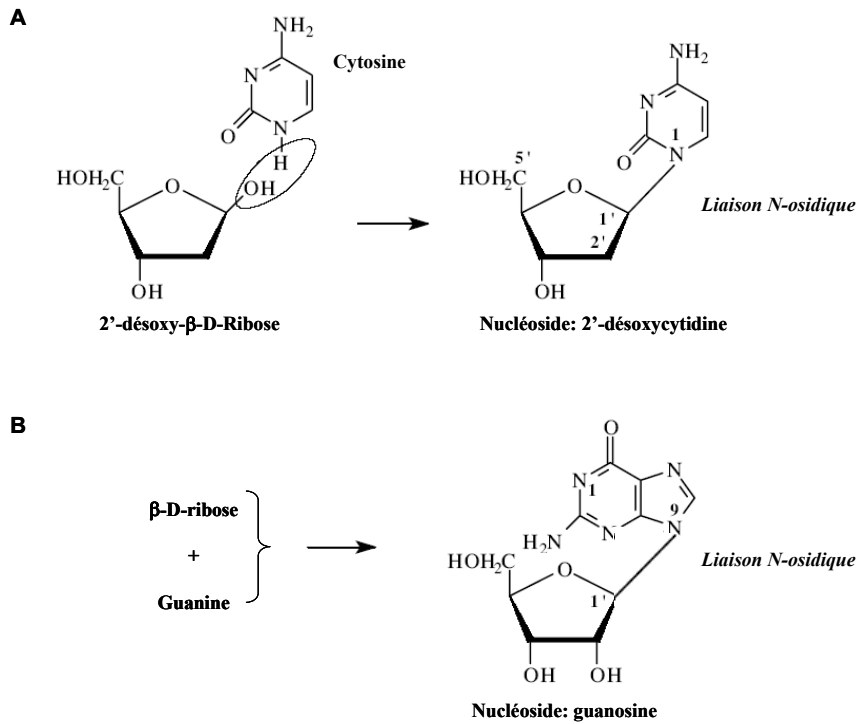
Dans les structures de nucléosides, de nucléotides (Altona and Sundralingam, 1973; Davies and Danyluk, 1974; Davies and Danyluk, 1975; Gelbin et al., 1996) et d'oligonucléotides (Berman et al., 1992; Dickerson et al., 1982), le sucre a été trouvé essentiellement dans deux types de plissement : C3' endo (Nord) et C2' endo (Sud). Le plissement Nord est associé à l'ARN et à l'ADN de forme A tandis que la conformation Sud est associée à la forme B de l'ADN. Dans la forme Z de l'ADN, le sucre des nucléotides puriques est en Sud et celui des nucléotides pyrimidiques est en Nord (Berman et al., 1992; Dickerson et al., 1982). En solution, les sucres de l'ADN sont donc en équilibre dynamique entre les deux plissements Nord et Sud. Des transitions Sud→Nord de certains nucléotides de l'ADN B sont observées dans les complexes ADN-protéine (Lu et al., 2000; Tolstorukov et al., 2001; Tolstorukov et al., 2004). Par ailleurs, dans les structures d'oligonucléotides d'ADN B, l'angle de pseudorotation a été trouvé plus élevé lorsque le sucre (en conformation Sud) est lié à une purine (Berman et al., 1992). Dans l'ARN, le plissement du sucre apparaît plus rigide que dans l'ADN (Auffinger and Westhof, 2000; Cheatham and Kollman, 1997; Pan and MacKerell, 2003).

Plusieurs théories ont été utilisées pour prédire le plissement des sucres à partir de la formule du nucléoside isolé : l'effet gauche, l'effet anomérique et l'électronégativité du substituant en 2'.

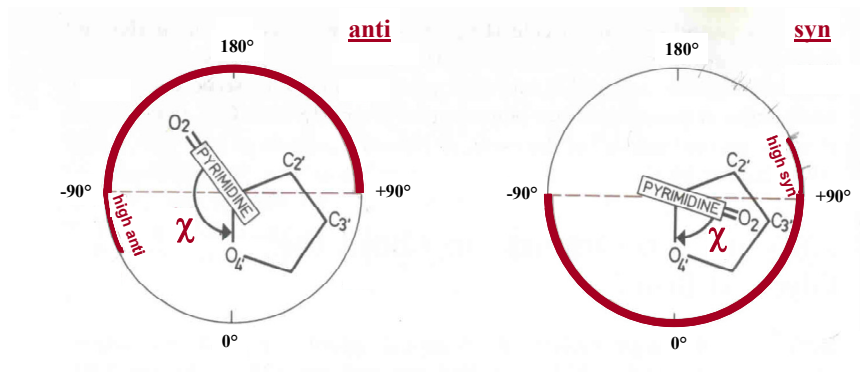
Selon la théorie de l'effet gauche un dièdre de type X-C-C-Y où X et Y sont des atomes électronégatifs, O ou N, évite la conformation trans et préfère être gauche. Le dièdre O4'-C4'-C3'-O3' ( $= \tau_3 - 120^\circ$ ) tire donc le sucre vers le Sud ( $\tau_3 = 40^\circ$ ). Dans un arabinonucléoside, le dièdre O4'-C1'-C2'-O2' ( $= \tau_1 - 120^\circ$ ) a le même effet. Dans un ribonucléoside, le dièdre O4'-C1'-C2'-O2' ( $= \tau_1 + 120^\circ$ ) tire le sucre vers le Nord.

L'effet anomérique favorise la forme Nord. Il provient de l'interaction d'un doublet libre de O4' avec l'orbitale antiliante de la liaison glycosidique.

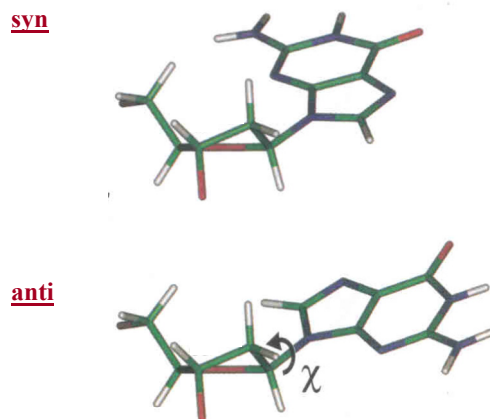
Plus le substituant en 2' est électronégatif, plus le sucre est en Nord.



**Figure 11-** Formation de la liaison N-osidique entre un sucre et une base pyrimidique (A) et une base purique (B).



**Figure 12-** Définition des orientations anti (à gauche) et syn (à droite), avec l'exemple d'un nucléoside pyrimidique ( $\chi = O_4 - C_1 - N_1 - C_2$ ) (1983). La région high anti avec  $\chi \sim -90^\circ$  appartient à syn. La région high syn avec  $\chi \sim +90^\circ$  appartient à anti.



**Figure 13-** Orientations syn et anti de la base purique de la 2'-désoxyguanosine.

### III- Les nucléosides et les nucléotides

Les nucléosides résultent de l'association d'une base et d'un sucre par formation d'une liaison N-osidique (glycosidique) entre le C1' (carbone anomérique) du furanose en conformation  $\beta$ , et le N1 des bases pyrimidiques ou le N9 des bases puriques (**figure 11**).

Par rapport au sucre, la base peut adopter deux orientations principales, appelées *syn* et *anti*. Ces orientations sont définies par l'angle dièdre glycosidique  $\chi$  :

- $\chi$  (O<sub>4'</sub> – C<sub>1'</sub> – N<sub>9</sub> – C<sub>4</sub>) pour les purines
- $\chi$  (O<sub>4'</sub> – C<sub>1'</sub> – N<sub>1</sub> – C<sub>2</sub>) pour les pyrimidines.

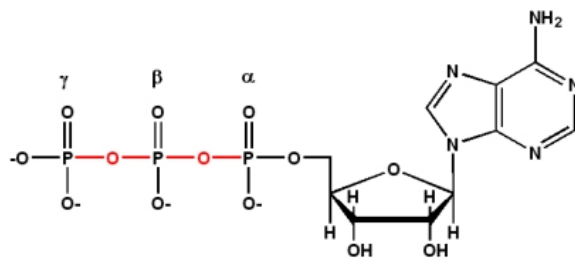
La gamme de valeurs de  $\chi$  qui est associée à une orientation *syn* est centrée autour de 0° et celle définissant une orientation *anti*, autour de 180° (**figure 12**).

Dans une orientation *syn*, le cycle à 6 atomes des purines et l'O2 des pyrimidines sont placés au-dessus du sucre. Par contre, dans une orientation *anti*, ils sont placés à l'opposé du sucre (**figures 12 et 13**).

Dans les acides nucléiques hélicoïdaux, la conformation *anti* est dominante car elle minimise l'encombrement stérique entre les atomes de la base et ceux du sucre. La conformation *syn* est observée dans des structures particulières comme l'ADN Z, généralement pour des purines où les gênes stériques sont moins importantes.

Les nucléotides sont des esters-phosphates des nucléosides. Un ou plusieurs phosphates peuvent être liés en un ou plusieurs sites hydroxylés du sucre des nucléosides. Le site le plus habituel d'estérification dans les nucléotides naturels est le groupe hydroxyl en C5' du sucre. Un composé phosphaté au niveau du C5' d'un sucre nucléosidique est appelé nucléoside 5'-phosphate ou 5'-nucléotide. Les nucléosides 5'-monophosphates sont les monomères constituant les acides nucléiques (**tableau 1**). Avant que ces nucléotides soient intégrés dans les acides nucléiques, ils sont sous forme triphosphatés (5'-dATP, 5'-ATP, 5'-dGTP, 5'-GTP, 5'-dCTP, 5'-CTP, 5'-UTP, 5'-dTTP). Le premier phosphate fixé en C5' du sucre nucléosidique est nommé  $\alpha$ , le second  $\beta$  et le troisième  $\gamma$  (**figure 14**).





Adénosine 5'-triphosphate (5'-ATP)

**Figure 14-** Un exemple de nucléoside 5'-triphosphate : l'adénosine 5'-triphosphate

**Tableau 1-** Les principaux nucléotides constituant les unités monomériques des acides nucléiques

Bases	Sucres	5'-nucléotides	
cytosine	$\beta$ -D-ribose	Cytidine 5'-monophosphate =acide cytidylique = 5'-CMP	
cytosine	2'-désoxy- $\beta$ -D-ribose	Désoxycytidine 5'-monophosphate = acide désoxycytidylique =5'-dCMP	
uracile	$\beta$ -D-ribose	Uridine 5'-monophosphate = acide uridylique =5'-UMP	
thymine	2'-désoxy- $\beta$ -D-ribose	Désoxythymine 5'-monophosphate = acide désoxythymidylique = 5'-dTMP	
adénine	$\beta$ -D-ribose	Adénosine 5'-monophosphate = acide adénylique = 5'-AMP	
adénine	2'-désoxy- $\beta$ -D-ribose	Désoxyadénosine 5'-monophosphate = acide désoxyadénylique = 5'-dAMP	
guanine	$\beta$ -D-ribose	Guanosine 5'-monophosphate = acide guanylique = 5'-GMP	
guanine	2'-désoxy- $\beta$ -D-ribose	Désoxyguanosine 5'-monophosphate = acide désoxyguanylique = 5'-dGMP	



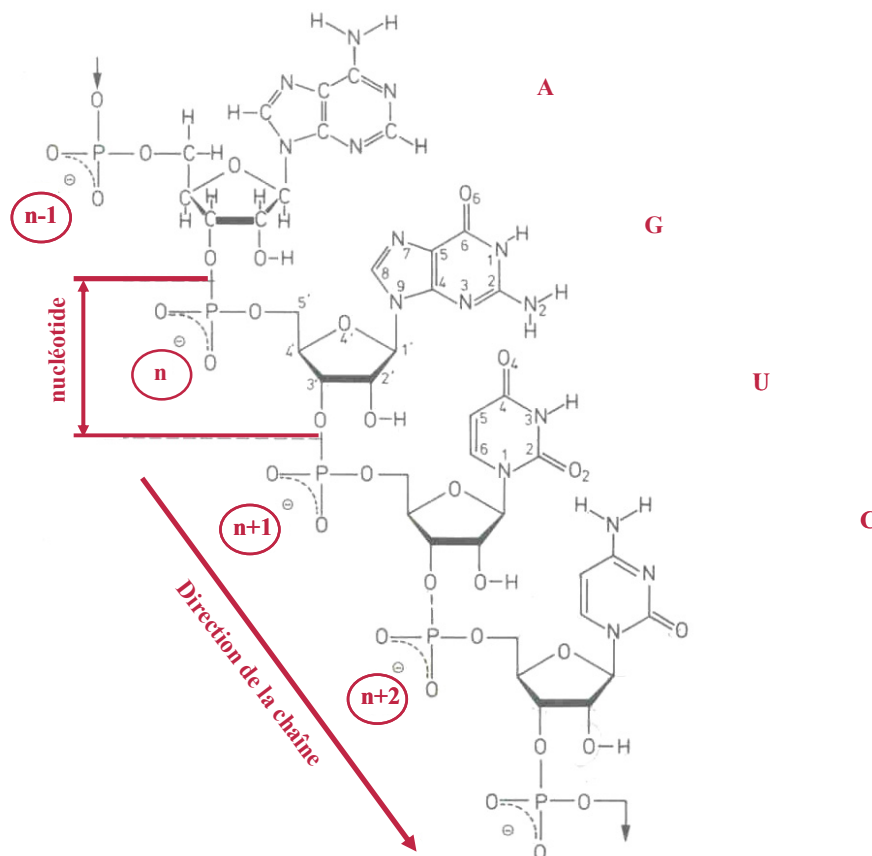


#### IV- Les chaînes polynucléotidiques

Toutes les macromolécules d'acides nucléiques sont formées par l'enchaînement de nucléosides 5'-monophosphates (**figure 15**). Cet assemblage est réalisé par des liaisons phosphodiesters 3'-5'. Ces liaisons se forment par attaque nucléophile du 3'-OH du sucre d'un nucléotide sur le phosphate  $\alpha$  d'un nucléoside 5'-triphosphate.

La chaîne polynucléotidique ou structure primaire des acides nucléiques s'écrit conventionnellement dans la direction 5'  $\rightarrow$  3'. C'est le sens dans lequel les séquences d'acides nucléiques sont utilisées comme molécules informationnelles (transcription, traduction).

Alors qu'un nucléotide libre peut comporter deux charges négatives sur le phosphate (avec des pK d'environ 1 et 6), il ne reste qu'une charge par nucléotide dans les acides nucléiques. Les acides nucléiques sont donc des polyélectrolytes fortement anioniques.



**Figure 15-** Exemple d'une chaîne polynucléotidique : un fragment d'ARN simple brin. La direction de la chaîne est montrée : de l'extrémité 5' vers l'extrémité 3'. Les numéros des atomes sont indiqués pour une unité nucléotidique, le 5'-GMP. Seulement pour le nucléotide n-1, tous les atomes d'hydrogène sont représentés. Ce fragment peut être noté : pApGpUpCp ou pAGUCp. Dans une chaîne polynucléotidique d'ADN, le groupe hydroxyl attaché au C2' est remplacé par un hydrogène et l'uracile par la thymine.

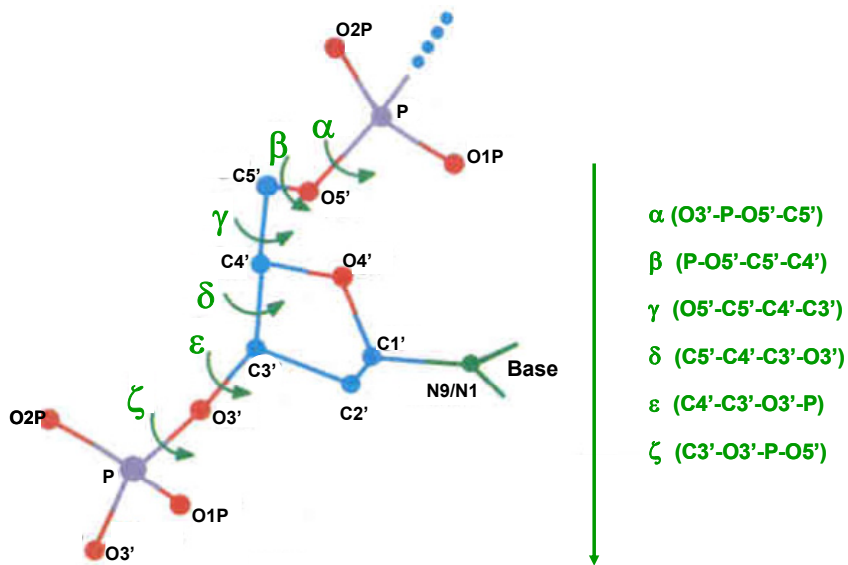


Figure 16- Définition des angles dièdres du squelette sucre-phosphate d'une chaîne polynucléotidique.

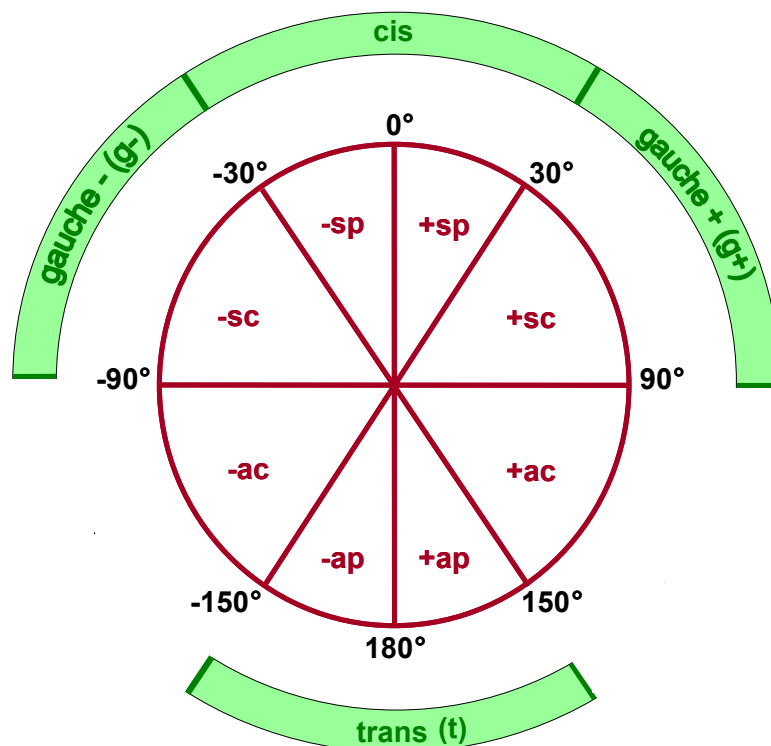


Figure 17- Nomenclatures des gammes de valeurs des angles dièdres du squelette sucre-phosphate : celle de Klyne et Prelog (Klyne W. and Prelog V., 1960) est montrée en rouge et celle utilisée en spectroscopie et cristallographie, en vert.

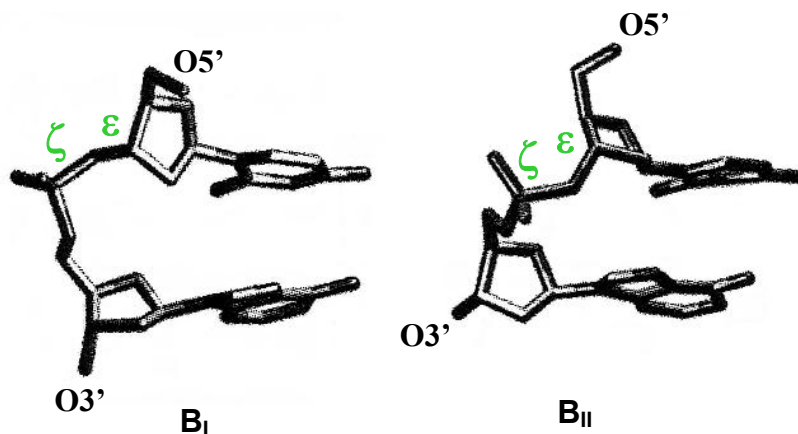
La conformation du squelette sucre-phosphate est caractérisée par six angles dièdres (**figure 16**). Il existe deux nomenclatures pour décrire les gammes de valeurs de ces angles : celle définie par Klyne et Prelog (Klyne and Prelog, 1960) et celle utilisée en spectroscopie et cristallographie (**figure 17**). Dans la nomenclature de Klyne et Prelog, l'espace conformationnel est divisé en deux régions synpériplanares (+sp et -sp), deux régions synclinales (+sc et -sc), deux régions anticlinales (+ac et -ac) et deux régions antipériplanares (+ap et -ap) (**figure 17**, en rouge). La seconde nomenclature définit quatre régions : cis (c), trans (t), gauche+ (g+) et gauche- (g-) (**figure 17**, en vert).

Ces six paramètres angulaires laissent supposer un grand polymorphisme du squelette sucre-phosphate des acides nucléiques. En réalité, les contraintes stériques et les barrières de potentiel de rotation limitent partiellement les gammes de valeurs accessibles de chacun des angles. Globalement, trois gammes de valeurs sont observées : t, g+ et g-.

## V- Corrélation entre les différents paramètres angulaires

Les variations des angles dièdres s'effectuent souvent de manières corrélées.

Dans l'ADN B, la modification du couple  $\varepsilon/\zeta$  de t/g- vers g-/t entraîne une rotation du groupe phosphate de sa conformation préférentielle B<sub>I</sub> à une seconde conformation B<sub>II</sub>, en équilibre dynamique avec la première (Gupta et al., 1980; Prive et al., 1987). Dans la conformation B<sub>I</sub>, le phosphate se trouve dans une position rigoureusement symétrique par rapport aux deux sillons alors que la conformation B<sub>II</sub> déplace le phosphate dans le petit sillon (**figure 18**). Des interconversions B<sub>I</sub>→B<sub>II</sub> de certains nucléotides interviennent dans les interactions ADN/protéine (Tisne et al., 1998; Winger et al., 1998).



**Figure 18-** Conformations B<sub>I</sub> (à gauche) et B<sub>II</sub> (à droite) du squelette phosphodiester de l'ADN.



Toute variation de  $\delta$  s'accompagne d'une déformation du cycle du sucre et donc d'une variation du couple de valeur  $\tau$  et  $P$  (ces variations étant plus perceptibles en terme de phase qu'en terme d'amplitude). La contrainte de fermeture du cycle furanose impose des valeurs de  $\delta$  comprises entre  $70^\circ$  et  $170^\circ$ .

Les angles  $\delta$  et  $\chi$  sont corrélés. En anti, les valeurs de  $\chi$  sont en  $-ac$  pour un sucre en Sud et en  $-ap$  pour un sucre en Nord.

Les angles  $\delta$  et  $\zeta$  sont également corrélés :  $\zeta$  est en  $-ac$  pour un sucre en Sud et en  $-sc$  pour un sucre en Nord.

Les angles  $\alpha$  et  $\gamma$  sont généralement corrélés et l'on peut observer des transitions conformationnelles de  $g^-/g^+$  à  $t/t$  (Prive et al., 1987; Venkateswarlu et al., 1999).  $\alpha$  et  $\gamma$  sont aussi couplés, dans un degré moindre, à l'angle  $\beta$  qui, dans ce cas varie de  $t$  vers  $g^+$ .

## **VI- Les doubles hélices**

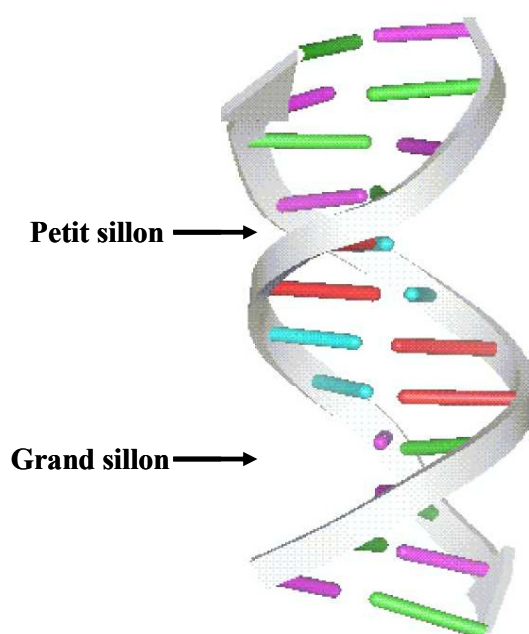
Dans le modèle structural de J. D. Watson et F. Crick (Watson and Crick, 1953), la molécule d'ADN est composée de deux chaînes polynucléotidiques enroulées autour d'un axe commun pour former une double hélice. L'enroulement des chaînes a lieu dans le sens des aiguilles d'une montre : l'hélice est dite droite. Les liaisons phosphodiesteres  $3'-5'$  des deux chaînes sont orientées en sens inverse si bien que les chaînes se déroulent dans des directions opposées : elles sont antiparallèles. Les enchaînements sucres-phosphates forment les squelettes hélicoïdaux extérieurs, parallèles. Les bases appariées complémentaires (A/T ou A/U et G/C) se trouvent à l'intérieur du cylindre central de l'hélice.

Depuis le modèle de Watson et Crick, la double hélice d'ADN a suscité de nombreux travaux. L'ADN en double hélice peut adopter différentes conformations selon son environnement : force ionique, degré d'hydratation, température, nature du solvant mais aussi selon sa composition en bases. A partir de nombreuses études expérimentales, les formes allomorphes de l'ADN ont pu être classées en plusieurs familles ; les trois principales sont A, B et Z. Les structures des formes A et B ont été déterminées par diffraction aux rayons X de fibres d'ADN naturelles. La forme B de l'ADN est la plus stable en solution aqueuse fortement hydratée et apparaît comme la conformation typique de l'ADN *in vivo*. Dans des conditions d'hydratation plus faible comme dans le cas d'un mélange éthanol-eau, la forme B de l'ADN fait place à la forme A. Les doubles hélices d'ARN en solution ont également une forme A.



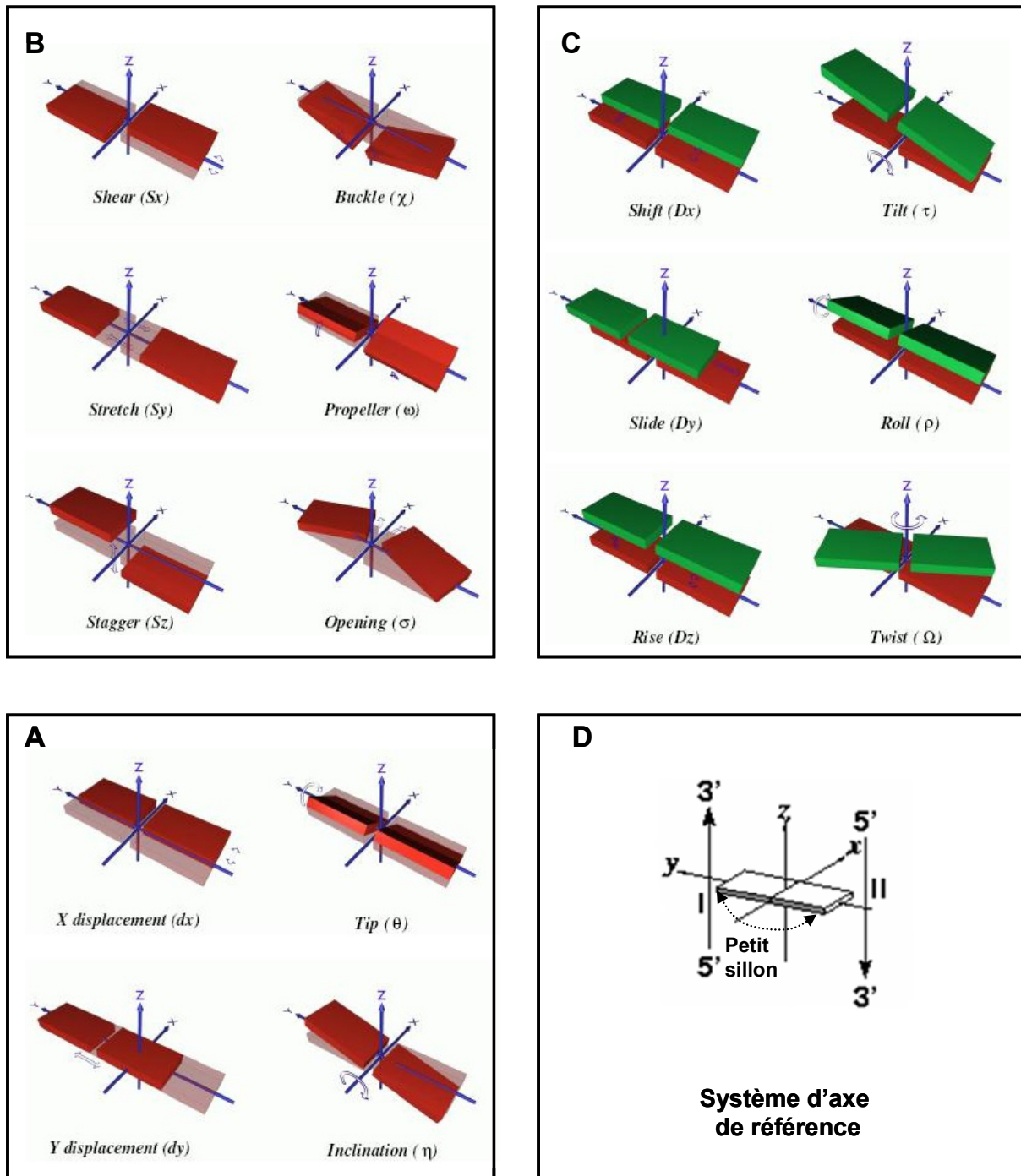
Par la suite, l'étude de cristaux de fragments d'ADN a permis de distinguer la troisième famille principale, la forme Z qui n'a pu être stabilisée que par une forte concentration en sel, en présence d'ions multivalents et pour des séquences particulières telles que des poly(dG-dC). Nous ne présentons pas en détail cette dernière conformation.

Du fait que les deux liaisons glycosidiques des paires de bases sont du même côté, deux sillons de largeurs et de profondeurs différentes s'alternent sur les flancs de la double hélice (**figure 19**). La géométrie des sillons a une grande importance pour les interactions entre l'acide nucléique et les molécules qui l'entourent (eau, contre-ions, protéines) : chaque sillon forme un environnement local différent du point de vue électrostatique et stérique. Entre les différentes formes de doubles hélices, la géométrie des sillons varie de manière importante. La séquence joue aussi un rôle dans l'aspect des sillons. Par exemple, les suites d'adénine (A-tracts) entraînent un rétrécissement du petit sillon dans la direction 5'→3' (Bostock-Smith et al., 1999).



**Figure 19-** Représentation schématique des sillons d'un ADN.





**Figure 20-** Représentation schématique : des paramètres hélicoïdaux (A), des paramètres de bases (B) et des paramètres de pas (C). Les schémas représentent des valeurs positives des paramètres. Le système d'axe de référence est montré (D). Nomenclature établie lors du colloque de Cambridge 1988 (Dickerson, 1989).

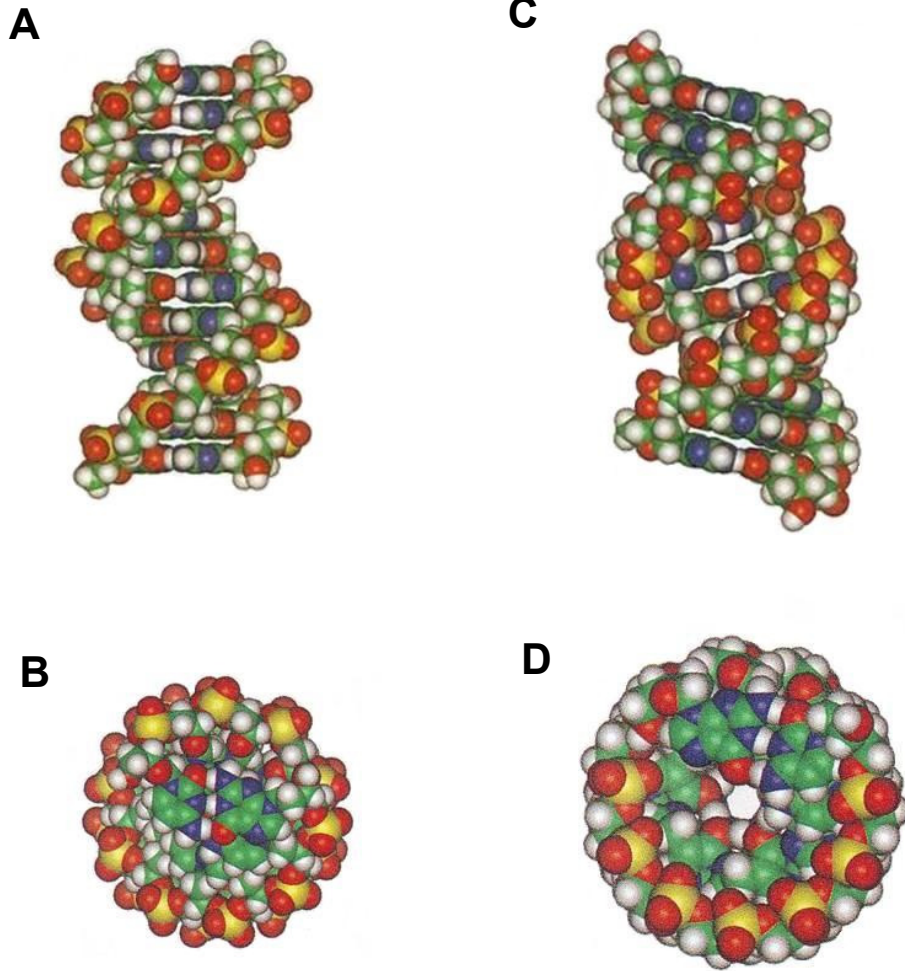
Les paramètres angulaires (plissements du sucre, angles dièdres du squelette sucre-phosphate et angles glycosidiques) permettent de caractériser la structure interne de chaque chaîne d'un acide nucléique. Cependant, pour avoir un meilleur aperçu de la structure des acides nucléiques, un certain nombre de paramètres ont été définis de façon à caractériser le positionnement et l'orientation des bases (ou paires de bases) entre elles ou par rapport à l'axe hélicoïdal. L'ensemble de ces paramètres est réparti en trois groupes :

1. Quatre paramètres hélicoïdaux qui décrivent le positionnement et l'orientation d'une paire de bases par rapport à l'axe hélicoïdal (deux rotations et deux translations) (**figure 20-A**).
2. Six paramètres de bases qui décrivent le positionnement et l'orientation des deux bases d'une paire, l'une par rapport à l'autre (trois translations et trois rotations) (**figure 20-B**).
3. Six paramètres de pas qui décrivent le positionnement et l'orientation des deux paires de bases successives, l'une par rapport à l'autre (trois translations et trois rotations) (**figure 20-C**).

Les noms et abréviations de chaque paramètre, la définition du système d'axe de référence (**figure 20-D**) et les signes des translations et rotations ont été standardisés lors du colloque de Cambridge en 1988 (Dickerson, 1989). Le sens positif des angles de rotation est donné par la « règle de la main droite » autour des axes associés. Les translations sont positives si elles ont lieu dans les directions positives des axes  $x$ ,  $y$  et  $z$ .  $x$  est positif dans la direction du grand sillon ;  $y$  est positif dans la direction allant du brin II vers le brin I ;  $z$  est positif dans la direction  $5' \rightarrow 3'$  du brin I. Les schémas de la **figure 20** représentent des valeurs positives des paramètres.

### **A- La double hélice de forme B**

La forme B est une hélice droite. Elle comporte en moyenne dix paires de bases par tour, espacées d'environ 3,3 Å (**tableau 2**). Les paires de bases sont quasi perpendiculaires à l'axe hélicoïdal et ce dernier passe près du centre de chaque paire de bases. Les sucres sont en Sud, les angles glycosidiques en anti et les groupements phosphate en conformation  $B_1$ . Les groupements phosphates sont diamétralement opposés et leurs oxygènes anioniques pointent vers l'extérieur de l'hélice. Ces derniers sont stabilisés par des interactions avec les molécules d'eau et les contre-ions positifs, réduisant ainsi la répulsion électrostatique entre les deux



**Figure 21-** Deux vues de l'ADN B canonique (A et B) et deux vues de l'ADN A canonique (C et D)

brins. La forme générale de la double hélice se caractérise par un grand sillon large et un petit sillon étroit de profondeur comparable (**figure 21**). Les valeurs modales des paramètres structuraux de la forme B sont regroupées dans le **tableau 2**.

**Tableau 2-** Caractéristiques structurales des doubles hélices de formes B et A

	<b>Forme B</b>	<b>Forme A</b>
<b>Description</b>		
Sens de l'hélice	Droit	Droit
Nombre de paires de bases/tour d'hélice	10-10,5	11
Diamètre de l'hélice	≈ 20 Å	≈ 26 Å
Plissement du sucre	Sud	Nord
Grand sillon	Large (≈ 12 Å) Profond (≈ 8,5 Å)	Étroit (≈ 2,7 Å) Profond (≈ 13,5 Å)
Petit sillon	Étroit (≈ -5,7 Å) Profond (≈ 7,5 Å)	Large (≈ 11 Å) Peu profond (≈ 2,8 Å)
<b>Paramètres angulaires (°)*</b>		
$\alpha$	-62 (19)	-69 (14)
$\beta$	168 (25)	175 (15)
$\gamma$	51 (16)	57 (10)
$\varepsilon$	-173 (20)	-155 (14)
$\zeta$	-98 (23)	-73 (9)
$\chi$	-108 (23)	-161 (11)
<b>Paramètres de bases*</b>		
Shear (Å)	0,00 (0,21)	0,01 (0,23)
Stretch (Å)	-0,15 (0,12)	-0,18 (0,10)
Stagger (Å)	0,09 (0,19)	0,02 (0,25)
Buckle (°)	0,53 (6,74)	-0,13 (7,77)
Propeller (°)	-11,35 (5,26)	-11,79 (4,14)
Opening (°)	0,63 (3,05)	0,57 (2,80)
<b>Paramètres de pas*</b>		
Shift (Å)	-0,02 (0,45)	-0,00 (0,54)
Slide (Å)	0,23 (0,81)	-1,53 (0,34)
Rise (Å)	3,32 (0,19)	3,31 (0,20)
Tilt (°)	-0,05 (2,52)	0,06 (2,81)
Roll (°)	0,60 (5,21)	7,95 (3,87)
Twist (°)	35,99 (6,80)	31,11 (4,01)
<b>Paramètres hélicoïdaux*</b>		
X-disp (Å)	0,05 (1,28)	-4,17 (1,22)
Y-disp (Å)	0,02 (0,87)	0,01 (0,89)
Inclination (°)	2,10 (9,15)	14,66 (7,30)
Tip (°)	0,02 (4,33)	-0,13 (5,16)

\* Valeurs modales observées dans les structures cristallographiques à haute résolution ( $\leq 2$  Å) d'ADN de formes A et B (Berman et al., 1992). Entre parenthèses est indiquée la dispersion.



## **B- La double hélice de forme A**

La forme A est une hélice droite dont le diamètre et le nombre de résidus par tour d'hélice est plus grand que celui de la forme B (**tableau 2**). Les angles glycosidiques restent en anti mais ils sont en  $-ap$  alors qu'ils sont en  $-ac$  dans la forme B. Les angles  $\varepsilon$  et  $\zeta$  sont légèrement plus élevés. Le roll est beaucoup plus élevé et le twist est réduit. Les paires de bases sont inclinées d'environ  $15^\circ$  par rapport à l'axe de l'hélice. Elles sont également décalées vers l'extérieur de l'hélice et l'axe principal passe par le grand sillon. Le déplacement des paires de bases a une conséquence sur l'aspect des sillons : le petit sillon est large et peu profond et le grand sillon est étroit et profond (**figure 21**). Contrairement à la forme B, les sucres sont en Nord. Ces changements structuraux induisent le rapprochement des groupements phosphates vers le grand sillon : le potentiel électro-négatif aux abords de ce dernier est ainsi augmenté, ce qui favorise l'insertion de cations dans le grand sillon. Les oxygènes anioniques se retrouvent face à face et sont fortement hydratés. Ils peuvent ainsi former entre eux, du fait de leur proximité, des liaisons inter et intra-brin par l'intermédiaire des molécules d'eau (Saenger et al., 1986).

Les données structurales dont la qualité ne cesse de s'améliorer montrent que la conformation de l'ADN est en réalité très complexe. L'ADN est une molécule relativement flexible qui présente des conformations locales diverses, dépendantes de la séquence en bases et de son environnement. Les trois formes canoniques, A, B et Z, doivent donc être vues comme trois familles de structures et non comme des structures uniques.

Il est généralement considéré par la communauté scientifique que la double hélice d'ARN est plus rigide que celle d'ADN. Cela est soutenu par des études expérimentales (Hagerman, 1988; Hagerman, 1997; Shindo et al., 1985) et théoriques (Auffinger and Westhof, 2000; Auffinger and Westhof, 2001b; Cheatham and Kollman, 1997). Les analyses des angles dièdres  $\alpha$ ,  $\chi$  et  $\gamma$  et du plissement des sucres durant des simulations de dynamique moléculaire de 2 ns ont amené les auteurs à conclure que l'ARN-A est plus rigide que l'ADN-B (Cheatham and Kollman, 1997). Les auteurs ont suggéré que les molécules d'eau très ordonnées autour de l'ARN réduisent sa flexibilité. Des calculs similaires réalisés par Auffinger et Westhof (Auffinger and Westhof, 2000; Auffinger and Westhof, 2001b) ont conduit à la même conclusion. Toutefois, des dynamiques moléculaires réalisées récemment (Noy et al., 2004; Pan and MacKerell, 2003) ainsi qu'une comparaison des structures



cristallographiques d'ADN-B et d'ARN-A (Pan and MacKerell, 2003) soulignent que la flexibilité relative de l'ADN et de l'ARN peut dépendre des déformations considérées. Ainsi, on pourrait plutôt dire que l'ADN et l'ARN ont des propriétés de flexibilité différentes.

Concernant le plissement du sucre, l'ARN apparaît toujours plus rigide que l'ADN (Auffinger and Westhof, 2000; Auffinger and Westhof, 2001b; Cheatham and Kollman, 1997; Pan and MacKerell, 2003).





# ***REFERENCES BIBLIOGRAPHIQUES***

---

## REFERENCES BIBLIOGRAPHIQUES

- Abd-Elazem, I.S., Chen, H.S., Bates, R.B. and Huang, R.C. (2002) Isolation of two highly potent and non-toxic inhibitors of human immunodeficiency virus type 1 (HIV-1) integrase from *Salvia miltiorrhiza*. *Antiviral Res.*, **55**, 91-106.
- Adamiak, D.A., Milecki, J., Popena, M., Adamiak, R.W., Dauter, Z. and Rypniewski, W.R. (1997) Crystal structure of 2'-O-Me(CGCGCG)<sub>2</sub>, an RNA duplex at 1.30 Å resolution. Hydration pattern of 2'-O-methylated RNA. *Nucleic Acids Res.*, **25**, 4599-4607.
- Adesokan, A.A., Roberts, V.A., Lee, K.W., Lins, R.D. and Briggs, J.M. (2004) Prediction of HIV-1 integrase/viral DNA interactions in the catalytic domain by fast molecular docking. *J.Med.Chem.*, **47**, 821-828.
- Agapkina, J., Smolov, M., Barbe, S., Zubin, E., Zatsepin, T., Deprez, E., Le Bret, M., Mouscadet, J.F. and Gottikh, M. (2006) Probing of HIV-1 integrase-DNA interactions using novel analogs of viral DNA. *J.Biol.Chem.*
- Aiken, C. and Trono, D. (1995) Nef stimulates human immunodeficiency virus type 1 proviral DNA synthesis 1. *J.Virol.*, **69**, 5048-5056.
- Altona, C. and Sundralingam, M. (1973) Conformational analysis of the sugar ring in nucleosides and nucleotides. Improved method for the interpretation of proton magnetic resonance coupling constants. *J.Am.Chem.Soc.*, **95**, 2333-2334.
- Arora, K. and Schlick, T. (2003) Deoxyadenosine sugar puckering pathway simulated by the stochastic difference equation algorithm. *Chemical Physics Letters*, **378**, 1-8.
- Asante-Appiah, E., Seeholzer, S.H. and Skalka, A.M. (1998) Structural determinants of metal-induced conformational changes in HIV-1 integrase. *J.Biol.Chem.*, **273**, 35078-35087.
- Asante-Appiah, E. and Skalka, A.M. (1999) HIV-1 integrase: structural organization, conformational changes, and catalysis. *Adv.Virus Res.*, **52**, 351-369.
- Auffinger, P. and Westhof, E. (1997) Rules governing the orientation of the 2'-hydroxyl group in RNA. *J.Mol.Biol.*, **274**, 54-63.
- Auffinger, P. and Westhof, E. (2000) Water and ion binding around RNA and DNA (C,G) oligomers. *J Mol Biol*, **300**, 1113-1131.
- Auffinger, P. and Westhof, E. (2001a) Hydrophobic groups stabilize the hydration shell of 2'-O-methylated RNA duplexes. *Angewandte Chemie-International Edition*, **40**, 4648-+.
- Auffinger, P. and Westhof, E. (2001b) Water and ion binding around r(UpA)<sub>12</sub> and d(TpA)<sub>12</sub> oligomers--comparison with RNA and DNA (CpG)<sub>12</sub> duplexes. *J Mol Biol*, **305**, 1057-1072.
- Aurup, H., Tuschl, T., Benseler, F., Ludwig, J. and Eckstein, F. (1994) Oligonucleotide duplexes containing 2'-amino-2'-deoxycytidines: thermal stability and chemical reactivity. *Nucleic Acids Res.*, **22**, 20-24.
- Balakrishnan, M. and Jonsson, C.B. (1997) Functional identification of nucleotides conferring substrate specificity to retroviral integrase reactions. *J.Virol.*, **71**, 1025-1035.
- Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. and Montagnier, L. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **220**, 868-871.
- Beese, L.S. and Steitz, T.A. (1991) Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.*, **10**, 25-33.

- Benard, C., Zouhiri, F., Normand-Bayle, M., Danet, M., Desmaele, D., Leh, H., Mouscadet, J.F., Mbemba, G., Thomas, C.M., Bonnenfant, S., Le Bret, M. and d'Angelo, J. (2004) Linker-modified quinoline derivatives targeting HIV-1 integrase: synthesis and biological activity. *Bioorg.Med.Chem.Lett.*, **14**, 2473-2476.
- Berendsen, H.J., Postma, J.P., Van Gunsteren, W.F., Di Nola, A. and Haak, J.R. (1984) Molecular dynamic with coupling to an external bath. *J.Chem Phys*, 3684-3690.
- Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.H., Srinivasan, A.R. and Schneider, B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys.J.*, **63**, 751-759.
- Bernardi, F., Bottoni, A., De Vivo, M., Garavelli, M., Keseru, G. and Naray-Szabo, G. (2002) A hypothetical mechanism for HIV-1 integrase catalytic action: DFT modelling of a bio-mimetic environment. *Chemical Physics Letters*, **362**, 1-7.
- Billich, A. (2003) S-1360 Shionogi-GlaxoSmithKline. *Curr.Opin.Investig.Drugs*, **4**, 206-209.
- Bonnenfant, S., Thomas, C.M., Vita, C., Subra, F., Deprez, E., Zouhiri, F., Desmaele, D., d'Angelo, J., Mouscadet, J.F. and Leh, H. (2004) Styrylquinolines, integrase inhibitors acting prior to integration: a new mechanism of action for anti-integrase agents. *J.Virol.*, **78**, 5728-5736.
- Bostock-Smith, C.E., Laughton, C.A. and Searle, M.S. (1999) Solution structure and dynamics of the A-T tract DNA decamer duplex d(GGTAATTACC)<sub>2</sub>: implications for recognition by minor groove binding drugs. *Biochem J*, **342 ( Pt 1)**, 125-132.
- Brameld, K.A. and Goddard, W.A. (1999) Ab initio quantum mechanical study of the structures and energies for the pseudorotation of 5'-dehydroxy analogues of 2'-deoxyribose and ribose sugars. *Journal of the American Chemical Society*, **121**, 985-993.
- Brodin, P., Pinskaya, M., Buckle, M., Parsch, U., Romanova, E., Engels, J., Gottikh, M. and Mouscadet, J.F. (2002) Disruption of HIV-1 integrase-DNA complexes by short 6-oxocytosine-containing oligonucleotides. *Biochemistry*, **41**, 1529-1538.
- Brodin, P., Pinskaya, M., Parsch, U., Bischerour, J., Leh, H., Romanova, E., Engels, J.W., Gottikh, M. and Mouscadet, J.F. (2001) 6-oxocytidine containing oligonucleotides inhibit the HIV-1 integrase in vitro. *Nucleosides Nucleotides Nucleic Acids*, **20**, 481-486.
- Brown, P.O., Coffin, J.M., Hughes, S.H. and Varmus, H.E. (1998) *Retroviruses*. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
- Bujacz, G., Alexandratos, J., Qing, Z.L., Clement-Mella, C. and Wlodawer, A. (1996a) The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant. *FEBS Lett.*, **398**, 175-178.
- Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrade, M., Katz, R.A. and Skalka, A.M. (1997) Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J.Biol.Chem.*, **272**, 18161-18168.
- Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A. and Skalka, A.M. (1995) High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J.Mol.Biol.*, **253**, 333-346.
- Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A. and Skalka, A.M. (1996b) The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure*, **4**, 89-96.
- Bukrinsky, M.I., Sharova, N., Dempsey, M.P., Stanwick, T.L., Bukrinskaya, A.G., Haggerty, S. and Stevenson, M. (1992) Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc.Natl.Acad.Sci.U.S.A*, **89**, 6580-6584.

- Burke, C.J., Sanyal, G., Bruner, M.W., Ryan, J.A., LaFemina, R.L., Robbins, H.L., Zeft, A.S., Middaugh, C.R. and Cordingley, M.G. (1992) Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *J.Biol.Chem.*, **267**, 9639-9644.
- Burkhoff, A.M. and Tullius, T.D. (1987) The unusual conformation adopted by the adenine tracts in kinetoplast DNA. *Cell*, **48**, 935-943.
- Bushman, F.D. and Craigie, R. (1991) Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. *Proc.Natl.Acad.Sci.U.S.A*, **88**, 1339-1343.
- Bushman, F.D., Engelman, A., Palmer, I., Wingfield, P. and Craigie, R. (1993) Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc.Natl.Acad.Sci.U.S.A*, **90**, 3428-3432.
- Bushman, F.D. and Wang, B. (1994) Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding. *J.Virol.*, **68**, 2215-2223.
- Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G.M. and Gronenborn, A.M. (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat.Struct.Biol.*, **4**, 567-577.
- Calmels, C., de Soultrait, V.R., Caumont, A., Desjobert, C., Faure, A., Fournier, M., Tarrago-Litvak, L. and Parissi, V. (2004) Biochemical and random mutagenesis analysis of the region carrying the catalytic E152 amino acid of HIV-1 integrase. *Nucleic Acids Res.*, **32**, 1527-1538.
- Carteau, S., Hoffmann, C. and Bushman, F. (1998) Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. *J.Virol.*, **72**, 4005-4014.
- Case, D.A., Darden, T.E., Cheatham, I.T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Wang, B., Pearlman, D.A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J.W., Ross, W.S. and Kollman, P.A. (2004) AMBER 8, University of California, San Francisco.
- Chan, D.C. and Kim, P.S. (1998) HIV entry and its inhibition. *Cell*, **93**, 681-684.
- Chargaff, E. (1951) Structure and function of nucleic acids as cell constituents. *Fed.Proc.*, **10**, 654-659.
- Cheatham, T.E. and Kollman, P.A. (1997) Molecular dynamics simulations highlights the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid duplexes. *J Am Chem Soc*, **119**, 4805-4825.
- Chen, A., Weber, I.T., Harrison, R.W. and Leis, J. (2006) Identification of amino acids in HIV-1 and avian sarcoma virus integrase subsites required for specific recognition of the long terminal repeat Ends. *J.Biol.Chem.*, **281**, 4173-4182.
- Chen, H. and Engelman, A. (1998) The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc.Natl.Acad.Sci.U.S.A*, **95**, 15270-15274.
- Chen, J.C., Krucinski, J., Miercke, L.J., Finer-Moore, J.S., Tang, A.H., Leavitt, A.D. and Stroud, R.M. (2000a) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc.Natl.Acad.Sci.U.S.A*, **97**, 8233-8238.
- Chen, Z., Yan, Y., Munshi, S., Li, Y., Zugay-Murphy, J., Xu, B., Witmer, M., Felock, P., Wolfe, A., Sardana, V., Emini, E.A., Hazuda, D. and Kuo, L.C. (2000b) X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50-293)--an initial glance of the viral DNA binding platform. *J.Mol.Biol.*, **296**, 521-533.

- Cherepanov, P., Ambrosio, A.L., Rahman, S., Ellenberger, T. and Engelman, A. (2005a) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc.Natl.Acad.Sci.U.S.A*, **102**, 17308-17313.
- Cherepanov, P., Este, J.A., Rando, R.F., Ojwang, J.O., Reekmans, G., Steinfeld, R., David, G., De Clercq, E. and Debyser, Z. (1997) Mode of interaction of G-quartets with the integrase of human immunodeficiency virus type 1. *Mol.Pharmacol.*, **52**, 771-780.
- Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E. and Debyser, Z. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J.Biol.Chem.*, **278**, 372-381.
- Cherepanov, P., Sun, Z.Y., Rahman, S., Maertens, G., Wagner, G. and Engelman, A. (2005b) Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nat.Struct.Mol.Biol.*, **12**, 526-532.
- Chow, S.A. and Brown, P.O. (1994) Substrate features important for recognition and catalysis by human immunodeficiency virus type 1 integrase identified by using novel DNA substrates. *J.Virol.*, **68**, 3896-3907.
- Chow, S.A., Vincent, K.A., Ellison, V. and Brown, P.O. (1992) Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science*, **255**, 723-726.
- Chrencik, J.E., Burgin, A.B., Pommier, Y., Stewart, L. and Redinbo, M.R. (2003) Structural impact of the leukemia drug 1-beta-D-arabinofuranosylcytosine (Ara-C) on the covalent human topoisomerase I-DNA complex. *J.Biol.Chem.*, **278**, 12461-12466.
- Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M.A., Santos-Ferreira, M.O., Laurent, A.G., Dauguet, C., Katlama, C. and Rouzioux, C. (1986) Isolation of a new human retrovirus from West African patients with AIDS. *Science*, **233**, 343-346.
- Clubb, R.T., Mizuuchi, M., Huth, J.R., Omichinski, J.G., Savilahti, H., Mizuuchi, K., Clore, G.M. and Gronenborn, A.M. (1996) The wing of the enhancer-binding domain of Mu phage transposase is flexible and is essential for efficient transposition. *Proc.Natl.Acad.Sci.U.S.A*, **93**, 1146-1150.
- Cummins, L.L., Owens, S.R., Risen, L.M., Lesnik, E.A., Freier, S.M., McGee, D., Guinasso, C.J. and Cook, P.D. (1995) Characterization of fully 2'-modified oligoribonucleotide hetero- and homoduplex hybridization and nuclease sensitivity. *Nucleic Acids Res.*, **23**, 2019-2024.
- Daniel, R., Greger, J.G., Katz, R.A., Taganov, K.D., Wu, X., Kappes, J.C. and Skalka, A.M. (2004) Evidence that stable retroviral transduction and cell survival following DNA integration depend on components of the nonhomologous end joining repair pathway. *J.Virol.*, **78**, 8573-8581.
- Davies, D.B. and Danyluk, S.S. (1974) Nuclear magnetic resonance studies of 5'-ribo- and deoxyribonucleotide structures in solution. *Biochemistry*, **13**, 4417-4434.
- Davies, D.B. and Danyluk, S.S. (1975) Nuclear magnetic resonance studies of 2'- and 3'-ribonucleotide structures in solution. *Biochemistry*, **14**, 543-554.
- Davies, D.R., Goryshin, I.Y., Reznikoff, W.S. and Rayment, I. (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science*, **289**, 77-85.
- Davies, J.F., Hostomska, Z., Hostomsky, Z., Jordan, S.R. and Matthews, D.A. (1991) Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science*, **252**, 88-95.
- De Luca, L., Pedretti, A., Vistoli, G., Barreca, M.L., Villa, L., Monforte, P. and Chimirri, A. (2003) Analysis of the full-length integrase-DNA complex by a modified approach for DNA docking. *Biochem.Biophys.Res.Commun.*, **310**, 1083-1088.

- De Luca, L., Vistoli, G., Pedretti, A., Barreca, M.L. and Chimirri, A. (2005) Molecular dynamics studies of the full-length integrase-DNA complex. *Biochem.Biophys.Res.Commun.*, **336**, 1010-1016.
- de Soultrait, V.R., Lozach, P.Y., Altmeyer, R., Tarrago-Litvak, L., Litvak, S. and Andreola, M.L. (2002) DNA aptamers derived from HIV-1 RNase H inhibitors are strong anti-integrase agents. *J.Mol.Biol.*, **324**, 195-203.
- Dennis, E.A. and Westheimer, F.H. (1966) The geometry of the transition state in the hydrolysis of phosphate esters. *J.Am.Chem.Soc.*, **88**, 3432-3433.
- Deprez, E., Tauc, P., Leh, H., Mouscadet, J.F., Auclair, C. and Brochon, J.C. (2000) Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. *Biochemistry*, **39**, 9275-9284.
- Deprez, E., Tauc, P., Leh, H., Mouscadet, J.F., Auclair, C., Hawkins, M.E. and Brochon, J.C. (2001) DNA binding induces dissociation of the multimeric form of HIV-1 integrase: a time-resolved fluorescence anisotropy study. *Proc.Natl.Acad.Sci.U.S.A.*, **98**, 10090-10095.
- Dickerson, R.E. (1989) Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res*, **17**, 1797-1803.
- Dickerson, R.E., Drew, H.R., Conner, B.N., Wing, R.M., Fratini, A.V. and Kopka, M.L. (1982) The anatomy of A-, B- and Z- DNA. *Science*, **216**, 475-485.
- Dirac, A.M. and Kjems, J. (2001) Mapping DNA-binding sites of HIV-1 integrase by protein footprinting. *Eur.J.Biochem.*, **268**, 743-751.
- Dixit, S.B. and Beveridge, D.L. (2005) Axis curvature and ligand induced bending in the CAP-DNA oligomers. *Biophys.J.*, **88**, L04-L06.
- Donzella, G.A., Jonsson, C.B. and Roth, M.J. (1993) Influence of substrate structure on disintegration activity of Moloney murine leukemia virus integrase. *J.Virol.*, **67**, 7077-7087.
- Drelich, M., Wilhelm, R. and Mous, J. (1992) Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein in vitro. *Virology*, **188**, 459-468.
- Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. and Davies, D.R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981-1986.
- Egli, M., Portmann, S. and Usman, N. (1996) RNA hydration: a detailed look. *Biochemistry*, **35**, 8489-8494.
- Eijkelenboom, A.P., Sprangers, R., Hard, K., Puras Lutzke, R.A., Plasterk, R.H., Boelens, R. and Kaptein, R. (1999) Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase. *Proteins*, **36**, 556-564.
- Ellison, V. and Brown, P.O. (1994) A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro. *Proc.Natl.Acad.Sci.U.S.A.*, **91**, 7316-7320.
- Ellison, V., Gerton, J., Vincent, K.A. and Brown, P.O. (1995) An essential interaction between distinct domains of HIV-1 integrase mediates assembly of the active multimer. *J.Biol.Chem.*, **270**, 3320-3326.
- Engelman, A., Bushman, F.D. and Craigie, R. (1993) Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.*, **12**, 3269-3275.
- Engelman, A. and Craigie, R. (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J.Virol.*, **66**, 6361-6369.

- Engelman, A., Hickman, A.B. and Craigie, R. (1994) The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J.Virol.*, **68**, 5911-5917.
- Engelman, A., Liu, Y., Chen, H., Farzan, M. and Dyda, F. (1997) Structure-based mutagenesis of the catalytic domain of human immunodeficiency virus type 1 integrase. *J.Virol.*, **71**, 3507-3514.
- Eom, S.H., Wang, J. and Steitz, T.A. (1996) Structure of Taq polymerase with DNA at the polymerase active site. *Nature*, **382**, 278-281.
- Espeseth, A.S., Felock, P., Wolfe, A., Witmer, M., Grobler, J., Anthony, N., Egbertson, M., Melamed, J.Y., Young, S., Hamill, T., Cole, J.L. and Hazuda, D.J. (2000) HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc.Natl.Acad.Sci.U.S.A.*, **97**, 11244-11249.
- Esposito, D. and Craigie, R. (1998) Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.*, **17**, 5832-5843.
- Essman, U., Perera, L., Berkowitz, M.L., Dardeb, T., Lee, H. and Pedersen, L.G. (1995) Asmooth particle mesh Ewald method. *J.Chem Phys.*, **103**, 8577-8593.
- Farnet, C.M. and Bushman, F.D. (1997) HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell*, **88**, 483-492.
- Farnet, C.M., Wang, B., Hansen, M., Lipford, J.R., Zalkow, L., Robinson, W.E., Jr., Siegel, J. and Bushman, F. (1998) Human immunodeficiency virus type 1 cDNA integration: new aromatic hydroxylated inhibitors and studies of the inhibition mechanism. *Antimicrob.Agents Chemother.*, **42**, 2245-2253.
- Farnet, C.M., Wang, B., Lipford, J.R. and Bushman, F.D. (1996) Differential inhibition of HIV-1 preintegration complexes and purified integrase protein by small molecules. *Proc.Natl.Acad.Sci.U.S.A.*, **93**, 9742-9747.
- Faure, A., Calmels, C., Desjobert, C., Castroviejo, M., Caumont-Sarcos, A., Tarrago-Litvak, L., Litvak, S. and Parissi, V. (2005) HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res.*, **33**, 977-986.
- Fesen, M.R., Kohn, K.W., Leteurtre, F. and Pommier, Y. (1993) Inhibitors of human immunodeficiency virus integrase. *Proc.Natl.Acad.Sci.U.S.A.*, **90**, 2399-2403.
- Fesen, M.R., Pommier, Y., Leteurtre, F., Hiroguchi, S., Yung, J. and Kohn, K.W. (1994) Inhibition of HIV-1 integrase by flavones, caffeic acid phenethyl ester (CAPE) and related compounds. *Biochem.Pharmacol.*, **48**, 595-608.
- Fikkert, V., Van Maele, B., Vercammen, J., Hantson, A., Van Remoortel, B., Michiels, M., Gurnari, C., Pannecouque, C., De Maeyer, M., Engelborghs, Y., De Clercq, E., Debyser, Z. and Witvrouw, M. (2003) Development of resistance against diketo derivatives of human immunodeficiency virus type 1 by progressive accumulation of integrase mutations. *J.Virol.*, **77**, 11459-11470.
- Firley, D., Courcot, B., Gillet, J.M., Fraisse, B., Zouhiri, F., Desmaele, D., d'Angelo, J. and Ghermani, N.E. (2006) Experimental/theoretical electrostatic properties of a styrylquinoline-type HIV-1 integrase inhibitor and its progenitors. *J.Phys.Chem.B Condens.Matter Mater.Surf.Interfaces.Biophys.*, **110**, 537-547.
- Flatters, D. and Lavery, R. (1998) Sequence-dependent dynamics of TATA-Box binding sites. *Biophys J*, **75**, 372-381.
- Flatters, D., Young, M., Beveridge, D.L. and Lavery, R. (1997) Conformational properties of the TATA-box binding sequence of DNA. *J Biomol Struct Dyn*, **14**, 757-765.
- Flick, K.E., Jurica, M.S., Monnat, R.J., Jr. and Stoddard, B.L. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*, **394**, 96-101.



- Freed, E.O. (1998) HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology*, **251**, 1-15.
- Freed, E.O., Orenstein, J.M., Buckler-White, A.J. and Martin, M.A. (1994) Single amino acid changes in the human immunodeficiency virus type 1 matrix protein block virus particle production. *J.Virol.*, **68**, 5311-5320.
- Fujiwara, T. and Shindo, H. (1985) Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 2. Molecular motions in hydrated DNA. *Biochemistry*, **24**, 896-902.
- Gabarro-Arpa, J., Cognet, J.A. and Le, B.M. (1992) Object Command Language: a formalism to build molecular models and to analyze structural parameters in macromolecules, with applications to nucleic acids. *J.Mol.Graph.*, **10**, 166-173, 162.
- Galburt, E.A., Chevalier, B., Tang, W., Jurica, M.S., Flick, K.E., Monnat, R.J., Jr. and Stoddard, B.L. (1999) A novel endonuclease mechanism directly visualized for I-PpoI. *Nat Struct Biol*, **6**, 1096-1099.
- Gallo, R.C., Sarin, P.S., Gelmann, E.P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V.S., Mann, D., Sidhu, G.D., Stahl, R.E., Zolla-Pazner, S., Leibowitch, J. and Popovic, M. (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science*, **220**, 865-867.
- Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M. and Hahn, B.H. (1999) Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature*, **397**, 436-441.
- Gao, K., Butler, S.L. and Bushman, F. (2001) Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes. *EMBO J.*, **20**, 3565-3576.
- Gao, K., Wong, S. and Bushman, F. (2004) Metal binding by the D,DX35E motif of human immunodeficiency virus type 1 integrase: selective rescue of Cys substitutions by Mn<sup>2+</sup> in vitro. *J.Virol.*, **78**, 6715-6722.
- Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.H., Olson, W.K. and Berman, H.M. (1996) Geometric parameters in nucleic acids: sugar and phosphate constituents. *J.Am.Chem.Soc.*, **118** 519-528.
- Gerton, J.L. and Brown, P.O. (1997) The core domain of HIV-1 integrase recognizes key features of its DNA substrates. *J.Biol.Chem.*, **272**, 25809-25815.
- Gerton, J.L., Herschlag, D. and Brown, P.O. (1999) Stereospecificity of reactions catalyzed by HIV-1 integrase. *J.Biol.Chem.*, **274**, 33480-33487.
- Gerton, J.L., Ohgi, S., Olsen, M., DeRisi, J. and Brown, P.O. (1998) Effects of mutations in residues near the active site of human immunodeficiency virus type 1 integrase on specific enzyme-substrate interactions. *J.Virol.*, **72**, 5046-5055.
- Gervais, V., Cognet, J.A., Le, B.M., Sowers, L.C. and Fazakerley, G.V. (1995) Solution structure of two mismatches A.A and T.T in the K-ras gene context by nuclear magnetic resonance and molecular dynamics. *Eur.J.Biochem.*, **228**, 279-290.
- Goedken, E.R. and Marqusee, S. (2001) Co-crystal of Escherichia coli RNase HI with Mn<sup>2+</sup> ions reveals two divalent metals bound in the active site. *J.Biol.Chem.*, **276**, 7266-7271.
- Goldgur, Y., Craigie, R., Cohen, G.H., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H. and Davies, D.R. (1999) Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. *Proc.Natl.Acad.Sci.U.S.A*, **96**, 13040-13043.
- Goldgur, Y., Dyda, F., Hickman, A.B., Jenkins, T.M., Craigie, R. and Davies, D.R. (1998) Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc.Natl.Acad.Sci.U.S.A*, **95**, 9150-9154.

- Goncalves, J., Korin, Y., Zack, J. and Gabuzda, D. (1996) Role of Vif in human immunodeficiency virus type 1 reverse transcription 1. *J.Virol.*, **70**, 8701-8709.
- Gottlieb, M.S., Schroff, R., Schanker, H.M., Weisman, J.D., Fan, P.T., Wolf, R.A. and Saxon, A. (1981) Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N.Engl.J.Med.*, **305**, 1425-1431.
- Greenwald, J., Le, V., Butler, S.L., Bushman, F.D. and Choe, S. (1999) The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry*, **38**, 8892-8898.
- Grobler, J.A., Stillmock, K., Hu, B., Witmer, M., Felock, P., Espeseth, A.S., Wolfe, A., Egbertson, M., Bourgeois, M., Melamed, J., Wai, J.S., Young, S., Vacca, J. and Hazuda, D.J. (2002) Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. *Proc.Natl.Acad.Sci.U.S.A*, **99**, 6661-6666.
- Gupta, G., Bansal, M. and Sasisekharan, V. (1980) Conformational flexibility of DNA: polymorphism and handedness. *Proc Natl Acad Sci U S A*, **77**, 6486-6490.
- Hagerman, P.J. (1988) Flexibility of DNA. *Annu Rev Biophys Biophys Chem*, **17**, 265-286.
- Hagerman, P.J. (1997) Flexibility of RNA. *Annu Rev Biophys Biomol Struct*, **26**, 139-156.
- Harding, M.M. (2001) Geometry of metal-ligand interactions in proteins. *Acta Crystallogr.D Biol.Crystallogr.*, **57**, 401-411.
- Harding, M.M. (2002) Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallogr.D Biol.Crystallogr.*, **58**, 872-874.
- Harper, A.L., Skinner, L.M., Sudol, M. and Katzman, M. (2001) Use of patient-derived human immunodeficiency virus type 1 integrases to identify a protein residue that affects target site selection. *J.Virol.*, **75**, 7756-7762.
- Harrich, D., Ulich, C., Garcia-Martinez, L.F. and Gaynor, R.B. (1997) Tat is required for efficient HIV-1 reverse transcription 1. *EMBO J.*, **16**, 1224-1235.
- Harris, D. and Engelman, A. (2000) Both the structure and DNA binding function of the barrier-to-autointegration factor contribute to reconstitution of HIV type 1 integration in vitro. *J.Biol.Chem.*, **275**, 39671-39677.
- Harvey, S.C., Tan, R.K.Z. and Cheatham, I.T.E. (1998) The flying ice cube : velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J.Comput.Chem.*, **19**, 726-740.
- Hazuda, D.J., Anthony, N.J., Gomez, R.P., Jolly, S.M., Wai, J.S., Zhuang, L., Fisher, T.E., Embrey, M., Guare, J.P., Jr., Egbertson, M.S., Vacca, J.P., Huff, J.R., Felock, P.J., Witmer, M.V., Stillmock, K.A., Danovich, R., Grobler, J., Miller, M.D., Espeseth, A.S., Jin, L., Chen, I.W., Lin, J.H., Kassahun, K., Ellis, J.D., Wong, B.K., Xu, W., Pearson, P.G., Schleif, W.A., Cortese, R., Emini, E., Summa, V., Holloway, M.K. and Young, S.D. (2004a) A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proc.Natl.Acad.Sci.U.S.A*, **101**, 11233-11238.
- Hazuda, D.J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J.A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C. and Miller, M.D. (2000) Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science*, **287**, 646-650.
- Hazuda, D.J., Felock, P.J., Hastings, J.C., Pramanik, B. and Wolfe, A.L. (1997) Differential divalent cation requirements uncouple the assembly and catalytic reactions of human immunodeficiency virus type 1 integrase. *J.Virol.*, **71**, 7005-7011.
- Hazuda, D.J., Young, S.D., Guare, J.P., Anthony, N.J., Gomez, R.P., Wai, J.S., Vacca, J.P., Handt, L., Motzel, S.L., Klein, H.J., Dornadula, G., Danovich, R.M., Witmer, M.V.,

- Wilson, K.A., Tussey, L., Schleif, W.A., Gabryelski, L.S., Jin, L., Miller, M.D., Casimiro, D.R., Emini, E.A. and Shiver, J.W. (2004b) Integrase inhibitors and cellular immunity suppress retroviral replication in rhesus macaques. *Science*, **305**, 528-532.
- Heuer, T.S. and Brown, P.O. (1997) Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry*, **36**, 10655-10665.
- Heuer, T.S. and Brown, P.O. (1998) Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry*, **37**, 6667-6678.
- Hickman, A.B., Palmer, I., Engelman, A., Craigie, R. and Wingfield, P. (1994) Biophysical and enzymatic properties of the catalytic domain of HIV-1 integrase. *J.Biol.Chem.*, **269**, 29279-29287.
- Hindmarsh, P., Ridky, T., Reeves, R., Andrade, M., Skalka, A.M. and Leis, J. (1999) HMG protein family members stimulate human immunodeficiency virus type 1 and avian sarcoma virus concerted DNA integration in vitro. *J.Virol.*, **73**, 2994-3003.
- Hocquet, A., Leulliot, N. and Ghomi, M. (2000) Ground-state properties of nucleic acid constituents studied by density functional calculations. 3. Role of sugar puckering and base orientation on the energetics and geometry of 2'-deoxyribonucleosides and ribonucleosides. *Journal of Physical Chemistry B*, **104**, 4560-4568.
- Hong, H., Neamati, N., Wang, S., Nicklaus, M.C., Mazumder, A., Zhao, H., Burke, T.R., Jr., Pommier, Y. and Milne, G.W. (1997) Discovery of HIV-1 integrase inhibitors by pharmacophore searching. *J.Med.Chem.*, **40**, 930-936.
- Hymes, K.B., Cheung, T., Greene, J.B., Prose, N.S., Marcus, A., Ballard, H., William, D.C. and Laubenstein, L.J. (1981) Kaposi's sarcoma in homosexual men—a report of eight cases. *Lancet*, **2**, 598-600.
- Jacks, T., Madhani, H.D., Masiarz, F.R. and Varmus, H.E. (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell*, **55**, 447-458.
- Jaroszewski, J.W., Syi, J.L., Maizel, J. and Cohen, J.S. (1992) Towards rational design of antisense DNA: molecular modelling of phosphorothioate DNA analogues. *Anticancer Drug Des*, **7**, 253-262.
- Jenkins, T.M., Engelman, A., Ghirlando, R. and Craigie, R. (1996) A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. *J.Biol.Chem.*, **271**, 7712-7718.
- Jenkins, T.M., Esposito, D., Engelman, A. and Craigie, R. (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J.*, **16**, 6849-6859.
- Jenkins, T.M., Hickman, A.B., Dyda, F., Ghirlando, R., Davies, D.R. and Craigie, R. (1995) Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc.Natl.Acad.Sci.U.S.A*, **92**, 6057-6061.
- Jing, N., Marchand, C., Liu, J., Mitra, R., Hogan, M.E. and Pommier, Y. (2000) Mechanism of inhibition of HIV-1 integrase by G-tetrad-forming oligonucleotides in Vitro. *J.Biol.Chem.*, **275**, 21460-21467.
- Jing, N., Xiong, W., Guan, Y., Pallansch, L. and Wang, S. (2002) Potassium-dependent folding: a key to intracellular delivery of G-quartet oligonucleotides as HIV inhibitors. *Biochemistry*, **41**, 5397-5403.
- Johnson, A.A., Santos, W., Pais, G.C., Marchand, C., Amin, R., Burke, T.R., Jr., Verdine, G. and Pommier, Y. (2006a) Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. *J.Biol.Chem.*, **281**, 461-467.

- Johnson, A.A., Sayer, J.M., Yagi, H., Kalena, G.P., Amin, R., Jerina, D.M. and Pommier, Y. (2004) Position-specific suppression and enhancement of HIV-1 integrase reactions by minor groove benzo[a]pyrene diol epoxide deoxyguanine adducts: implications for molecular interactions between integrase and substrates. *J.Biol.Chem.*, **279**, 7947-7955.
- Johnson, A.A., Sayer, J.M., Yagi, H., Patil, S.S., Debart, F., Maier, M.A., Corey, D.R., Vasseur, J.J., Burke, T.R., Jr., Marquez, V.E., Jerina, D.M. and Pommier, Y. (2006b) Effect of DNA modifications on DNA processing by HIV-1 integrase and inhibitor binding: role of DNA backbone flexibility and an open catalytic site. *J Biol Chem.*
- Jonsson, C.B., Donzella, G.A., Gaucan, E., Smith, C.M. and Roth, M.J. (1996) Functional domains of Moloney murine leukemia virus integrase defined by mutation and complementation analysis. *J.Virol.*, **70**, 4585-4597.
- Jonsson, C.B., Donzella, G.A. and Roth, M.J. (1993) Characterization of the forward and reverse integration reactions of the Moloney murine leukemia virus integrase protein purified from *Escherichia coli*. *J.Biol.Chem.*, **268**, 1462-1469.
- Jorgensen, W.L. (1981) Transferable intermolecular potential functions for water, alcohols and ethers, application to liquid water. *J.Am.Chem.Soc.*, **103**, 335-340.
- Kao, S.Y., Calman, A.F., Luciw, P.A. and Peterlin, B.M. (1987) Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature*, **330**, 489-493.
- Karki, R.G., Tang, Y., Burke, T.R., Jr. and Nicklaus, M.C. (2004) Model of full-length HIV-1 integrase complexed with viral DNA as template for anti-HIV drug design. *J.Comput.Aided Mol.Des.*, **18**, 739-760.
- Karn, J. (1999) Tackling Tat. *J.Mol.Biol.*, **293**, 235-254.
- Katayanagi, K., Okumura, M. and Morikawa, K. (1993) Crystal structure of *Escherichia coli* RNase HI in complex with Mg<sup>2+</sup> at 2.8 Å resolution: proof for a single Mg(2<sup>+</sup>)-binding site. *Proteins*, **17**, 337-346.
- Katz, R.A., Merkel, G. and Skalka, A.M. (1996) Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: in vitro activities and incorporation of a fusion protein into viral particles. *Virology*, **217**, 178-190.
- Katzman, M., Mack, J.P., Skalka, A.M. and Leis, J. (1991) A covalent complex between retroviral integrase and nicked substrate DNA. *Proc.Natl.Acad.Sci.U.S.A.*, **88**, 4695-4699.
- Katzman, M. and Sudol, M. (1995) Mapping domains of retroviral integrase responsible for viral DNA specificity and target site selection by analysis of chimeras between human immunodeficiency virus type 1 and visna virus integrases. *J.Virol.*, **69**, 5687-5696.
- Katzman, M. and Sudol, M. (1996a) Influence of subterminal viral DNA nucleotides on differential susceptibility to cleavage by human immunodeficiency virus type 1 and visna virus integrases. *J.Virol.*, **70**, 9069-9073.
- Katzman, M. and Sudol, M. (1996b) Nonspecific alcoholysis, a novel endonuclease activity of human immunodeficiency virus type 1 and other retroviral integrases. *J.Virol.*, **70**, 2598-2604.
- Keele, B.F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-Ruche, F., Chen, Y., Wain, L.V., Liegeois, F., Loul, S., Mpoudi, N.E., Bienvenue, Y., Delaporte, E., Brookfield, J.F., Sharp, P.M., Shaw, G.M., Peeters, M. and Hahn, B.H. (2006) Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1. *Science*.
- Khan, E., Mack, J.P., Katz, R.A., Kulkosky, J. and Skalka, A.M. (1991) Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.*, **19**, 851-860.

- Kiernan, R.E., Ono, A., Englund, G. and Freed, E.O. (1998) Role of matrix in an early postentry step in the human immunodeficiency virus type 1 life cycle 2. *J.Virol.*, **72**, 4116-4126.
- Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J.C. and Montagnier, L. (1984) T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV 1. *Nature*, **312**, 767-768.
- Klyne, W. and Prelog, V. (1960) Description of Steric Relationships Across Single Bonds. *Experientia*, **16**, 521-568.
- Knowles, J.R. (1980) Enzyme-catalyzed phosphoryl transfer reactions. *Annu.Rev.Biochem.*, **49**, 877-919.
- Kowalski, M., Potz, J., Basiripour, L., Dorfman, T., Goh, W.C., Terwilliger, E., Dayton, A., Rosen, C., Haseltine, W. and Sodroski, J. (1987) Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1 3. *Science*, **237**, 1351-1355.
- Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P. and Skalka, A.M. (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol.Cell Biol.*, **12**, 2331-2338.
- Kulkosky, J., Katz, R.A., Merkel, G. and Skalka, A.M. (1995) Activities and substrate specificity of the evolutionarily conserved central domain of retroviral integrase. *Virology*, **206**, 448-456.
- Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J. and Hendrickson, W.A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, **393**, 648-659.
- Laboulais, C., Deprez, E., Leh, H., Mouscadet, J.F., Brochon, J.C. and Le Bret, M. (2001) HIV-1 integrase catalytic core: molecular dynamics and simulated fluorescence decays. *Biophys.J.*, **81**, 473-489.
- LaFemina, R.L., Callahan, P.L. and Cordingley, M.G. (1991) Substrate specificity of recombinant human immunodeficiency virus integrase protein. *J.Virol.*, **65**, 5624-5630.
- Le Bret, M., Gabarro-Arpa, J., Gilbert, J.C. and Lemarechal, C. (1991) MORCAD, an object-oriented molecular modelling package running on IBM RS/6000 and SGI 4Dxxx workstations. *J.Chim.Phys.*, **88**, 2489-2496.
- Leavitt, A.D., Rose, R.B. and Varmus, H.E. (1992) Both substrate and target oligonucleotide sequences affect in vitro integration mediated by human immunodeficiency virus type 1 integrase protein produced in *Saccharomyces cerevisiae*. *J.Virol.*, **66**, 2359-2368.
- Leavitt, A.D., Shiue, L. and Varmus, H.E. (1993) Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. *J.Biol.Chem.*, **268**, 2113-2119.
- Lee, S.P., Xiao, J., Knutson, J.R., Lewis, M.S. and Han, M.K. (1997) Zn<sup>2+</sup> promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. *Biochemistry*, **36**, 173-180.
- Leh, H., Brodin, P., Bischerour, J., Deprez, E., Tauc, P., Brochon, J.C., LeCam, E., Coulaud, D., Auclair, C. and Mouscadet, J.F. (2000) Determinants of Mg<sup>2+</sup>-dependent activities of recombinant human immunodeficiency virus type 1 integrase. *Biochemistry*, **39**, 9285-9294.
- Lesnik, E.A. and Freier, S.M. (1998) What affects the effect of 2'-alkoxy modifications? 1. Stabilization effect of 2'-methoxy substitutions in uniformly modified DNA oligonucleotides. *Biochemistry*, **37**, 6991-6997.
- Leulliot, N., Ghomi, M., Scalmani, G. and Berthier, G. (1999) Ground state properties of the nucleic acid constituents studied by density functional calculations. I. Conformational

- features of ribose, dimethyl phosphate, uridine, cytidine, 5'-methyl phosphate-uridine, and 3'-methyl phosphate-uridine. *Journal of Physical Chemistry A*, **103**, 8716-8724.
- Levitt, M. and Warshel, A. (1978) Extreme conformational flexibility of the furanose ring in DNA and RNA. *J.Am.Chem.Soc*, **100**, 2607-2613.
- Li, X., Quan, Y., Arts, E.J., Li, Z., Preston, B.D., de Rocquigny, H., Roques, B.P., Darlix, J.L., Kleiman, L., Parniak, M.A. and Wainberg, M.A. (1996) Human immunodeficiency virus Type 1 nucleocapsid protein (NCp7) directs specific initiation of minus-strand DNA synthesis primed by human tRNA(Lys3) in vitro: studies of viral RNA molecules mutated in regions that flank the primer binding site 2. *J.Virol.*, **70**, 4996-5004.
- Lin, C.W. and Engelman, A. (2003) The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes. *J.Virol.*, **77**, 5030-5036.
- Lins, R.D., Straatsma, T.P. and Briggs, J.M. (2000) Similarities in the HIV-1 and ASV integrase active sites upon metal cofactor binding. *Biopolymers*, **53**, 308-315.
- Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A. and Landau, N.R. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell*, **86**, 367-377.
- Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M. and Poeschla, E.M. (2004) LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J.Virol.*, **78**, 9524-9537.
- Lodi, P.J., Ernst, J.A., Kuszewski, J., Hickman, A.B., Engelman, A., Craigie, R., Clore, G.M. and Gronenborn, A.M. (1995) Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry*, **34**, 9826-9833.
- Lu, X.J., Shakked, Z. and Olson, W.K. (2000) A-form conformational motifs in ligand-bound DNA structures 445. *J.Mol.Biol.*, **300**, 819-840.
- Lubini, P., Zurcher, W. and Egli, M. (1994) Stabilizing effects of the RNA 2'-substituent: crystal structure of an oligodeoxynucleotide duplex containing 2'-O-methylated adenosines. *Chem.Biol.*, **1**, 39-45.
- Lubkowski, J., Dauter, Z., Yang, F., Alexandratos, J., Merkel, G., Skalka, A.M. and Wlodawer, A. (1999) Atomic resolution structures of the core domain of avian sarcoma virus integrase and its D64N mutant. *Biochemistry*, **38**, 13512-13522.
- Lubkowski, J., Yang, F., Alexandratos, J., Merkel, G., Katz, R.A., Gravuer, K., Skalka, A.M. and Wlodawer, A. (1998a) Structural basis for inactivating mutations and pH-dependent activity of avian sarcoma virus integrase. *J.Biol.Chem.*, **273**, 32685-32689.
- Lubkowski, J., Yang, F., Alexandratos, J., Wlodawer, A., Zhao, H., Burke, T.R., Jr., Neamati, N., Pommier, Y., Merkel, G. and Skalka, A.M. (1998b) Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase-targeted inhibitor. *Proc.Natl.Acad.Sci.U.S.A*, **95**, 4831-4836.
- Lutzke, R.A. and Plasterk, R.H. (1998) Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. *J.Virol.*, **72**, 4841-4848.
- Lutzke, R.A., Vink, C. and Plasterk, R.H. (1994) Characterization of the minimal DNA-binding domain of the HIV integrase protein. *Nucleic Acids Res.*, **22**, 4125-4131.
- Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z. and Engelborghs, Y. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J.Biol.Chem.*, **278**, 33528-33539.

- Maignan, S., Guilloteau, J.P., Zhou-Liu, Q., Clement-Mella, C. and Mikol, V. (1998) Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *J.Mol.Biol.*, **282**, 359-368.
- Marchand, C., Johnson, A.A., Karki, R.G., Pais, G.C., Zhang, X., Cowansage, K., Patel, T.A., Nicklaus, M.C., Burke, T.R., Jr. and Pommier, Y. (2003) Metal-dependent inhibition of HIV-1 integrase by beta-diketo acids and resistance of the soluble double-mutant (F185K/C280S). *Mol.Pharmacol.*, **64**, 600-609.
- Marchand, C., Neamati, N. and Pommier, Y. (2001) In vitro human immunodeficiency virus type 1 integrase assays. *Methods Enzymol.*, **340**, 624-633.
- Marchand, C., Zhang, X., Pais, G.C., Cowansage, K., Neamati, N., Burke, T.R., Jr. and Pommier, Y. (2002) Structural determinants for HIV-1 integrase inhibition by beta-diketo acids. *J.Biol.Chem.*, **277**, 12596-12603.
- Marky, L.A. and Breslauer, K.J. (1982) Calorimetric determination of base-stacking enthalpies in double-helical DNA molecules. *Biopolymers*, **21**, 2185-2194.
- Maroun, R.G., Gayet, S., Benleulmi, M.S., Porumb, H., Zargarian, L., Merad, H., Leh, H., Mouscadet, J.F., Troalen, F. and Femandjian, S. (2001) Peptide inhibitors of HIV-1 integrase dissociate the enzyme oligomers. *Biochemistry*, **40**, 13840-13848.
- Martinson, J.J., Chapman, N.H., Rees, D.C., Liu, Y.T. and Clegg, J.B. (1997) Global distribution of the CCR5 gene 32-basepair deletion 1. *Nat.Genet.*, **16**, 100-103.
- Mazumder, A., Neamati, N., Ojwang, J.O., Sunder, S., Rando, R.F. and Pommier, Y. (1996a) Inhibition of the human immunodeficiency virus type 1 integrase by guanosine quartet structures. *Biochemistry*, **35**, 13762-13771.
- Mazumder, A., Neamati, N., Pilon, A.A., Sunder, S. and Pommier, Y. (1996b) Chemical trapping of ternary complexes of human immunodeficiency virus type 1 integrase, divalent metal, and DNA substrates containing an abasic site. Implications for the role of lysine 136 in DNA binding. *J.Biol.Chem.*, **271**, 27330-27338.
- Mazumder, A., Neamati, N., Sunder, S., Schulz, J., Pertz, H., Eich, E. and Pommier, Y. (1997a) Curcumin analogs with altered potencies against HIV-1 integrase as probes for biochemical mechanisms of drug action. *J.Med.Chem.*, **40**, 3057-3063.
- Mazumder, A. and Pommier, Y. (1995) Processing of deoxyuridine mismatches and abasic sites by human immunodeficiency virus type-1 integrase. *Nucleic Acids Res.*, **23**, 2865-2871.
- Mazumder, A., Uchida, H., Neamati, N., Sunder, S., Jaworska-Maslanka, M., Wickstrom, E., Zeng, F., Jones, R.A., Mandes, R.F., Chenault, H.K. and Pommier, Y. (1997b) Probing interactions between viral DNA and human immunodeficiency virus type 1 integrase using dinucleotides 641. *Mol.Pharmacol.*, **51**, 567-575.
- McConnell, K.J. and Beveridge, D.L. (2001) Molecular dynamics simulations of B'-DNA: sequence effects on A-tract-induced bending and flexibility. *J Mol Biol*, **314**, 23-40.
- McDougall, B., King, P.J., Wu, B.W., Hostomsky, Z., Reinecke, M.G. and Robinson, W.E., Jr. (1998) Dicafeoylquinic and dicafeoyltartaric acids are selective inhibitors of human immunodeficiency virus type 1 integrase. *Antimicrob.Agents Chemother.*, **42**, 140-146.
- 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**, 137-160.
- McLachlan, A.D. (1979) Least squares fitting of two structures. *J.Mol.Biol.*, **128**, 49-79.
- Mekouar, K., Mouscadet, J.F., Desmaele, D., Subra, F., Leh, H., Savoure, D., Auclair, C. and d'Angelo, J. (1998) Styrylquinoline derivatives: a new class of potent HIV-1 integrase inhibitors that block HIV-1 replication in CEM cells. *J.Med.Chem.*, **41**, 2846-2857.

- Minchenkova, L.E., Schyolkina, A.K., Chernov, B.K. and Ivanov, V.I. (1986) CC/GG contacts facilitate the B to A transition of DNA in solution. *J Biomol Struct Dyn*, **4**, 463-476.
- Mizuuchi, K. (1992) Polynucleotidyl transfer reactions in transpositional DNA recombination. *J.Biol.Chem.*, **267**, 21273-21276.
- Mizuuchi, K. (1997) Polynucleotidyl transfer reactions in site-specific DNA recombination 1. *Genes Cells*, **2**, 1-12.
- Molteni, V., Greenwald, J., Rhodes, D., Hwang, Y., Kwiatkowski, W., Bushman, F.D., Siegel, J.S. and Choe, S. (2001) Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain. *Acta Crystallogr.D Biol.Crystallogr.*, **57**, 536-544.
- Mondor, I., Ugolini, S. and Sattentau, Q.J. (1998) Human immunodeficiency virus type 1 attachment to HeLa CD4 cells is CD4 independent and gp120 dependent and requires cell surface heparans 2. *J.Virol.*, **72**, 3623-3634.
- Mousnier, A., Leh, H., Mouscadet, J.F. and Dargemont, C. (2004) Nuclear import of HIV-1 integrase is inhibited in vitro by styrylquinoline derivatives. *Mol.Pharmacol.*, **66**, 783-788.
- Mumm, S.R. and Grandgenett, D.P. (1991) Defining nucleic acid-binding properties of avian retrovirus integrase by deletion analysis. *J.Virol.*, **65**, 1160-1167.
- Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J. and Brierley, I. (2006) A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*, **441**, 244-247.
- Neamati, N., Hong, H., Owen, J.M., Sunder, S., Winslow, H.E., Christensen, J.L., Zhao, H., Burke, T.R., Jr., Milne, G.W. and Pommier, Y. (1998) Salicylhydrazine-containing inhibitors of HIV-1 integrase: implication for a selective chelation in the integrase active site. *J.Med.Chem.*, **41**, 3202-3209.
- Neamati, N., Lin, Z., Karki, R.G., Orr, A., Cowansage, K., Strumberg, D., Pais, G.C., Voigt, J.H., Nicklaus, M.C., Winslow, H.E., Zhao, H., Turpin, J.A., Yi, J., Skalka, A.M., Burke, T.R., Jr. and Pommier, Y. (2002) Metal-dependent inhibition of HIV-1 integrase. *J.Med.Chem.*, **45**, 5661-5670.
- Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1995) Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, **269**, 656-663.
- Newmark, R.A. and Cantor, C.R. (1968) Nuclear magnetic resonance study of the interactions of guanosine and cytidine in dimethyl sulfoxide. *J Am Chem Soc*, **90**, 5010-5017.
- Nguyen, D.H. and Hildreth, J.E. (2000) Evidence for budding of human immunodeficiency virus type 1 selectively from glycolipid-enriched membrane lipid rafts. *J.Virol.*, **74**, 3264-3272.
- Ni, H., Sotriffer, C.A. and McCammon, J.A. (2001) Ordered water and ligand mobility in the HIV-1 integrase-5CITEP complex: a molecular dynamics study. *J.Med.Chem.*, **44**, 3043-3047.
- Nicklaus, M.C., Neamati, N., Hong, H., Mazumder, A., Sunder, S., Chen, J., Milne, G.W. and Pommier, Y. (1997) HIV-1 integrase pharmacophore: discovery of inhibitors through three-dimensional database searching. *J.Med.Chem.*, **40**, 920-929.
- Normand-Bayle, M., Benard, C., Zouhiri, F., Mouscadet, J.F., Leh, H., Thomas, C.M., Mbemba, G., Desmaele, D. and d'Angelo, J. (2005) New HIV-1 replication inhibitors of the styrylquinoline class bearing aroyl/acyl groups at the C-7 position: synthesis and biological activity. *Bioorg.Med.Chem.Lett.*, **15**, 4019-4022.
- Noy, A., Perez, A., Lankas, F., Javier Luque, F. and Orozco, M. (2004) Relative flexibility of DNA and RNA: a molecular dynamics study. *J Mol Biol*, **343**, 627-638.



- Olson, W.K., Bansal, M., Burley, S.K., Dickerson, R.E., Gerstein, M., Harvey, S.C., Heinemann, U., Lu, X.J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C.S., Westhof, E., Wolberger, C. and Berman, H.M. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J.Mol.Biol.*, **313**, 229-237.
- Ono, A. and Freed, E.O. (2001) Plasma membrane rafts play a critical role in HIV-1 assembly and release. *Proc.Natl.Acad.Sci.U.S.A*, **98**, 13925-13930.
- Ono, A., Huang, M. and Freed, E.O. (1997) Characterization of human immunodeficiency virus type 1 matrix revertants: effects on virus assembly, Gag processing, and Env incorporation into virions. *J.Virol.*, **71**, 4409-4418.
- Ono, A., Orenstein, J.M. and Freed, E.O. (2000) Role of the Gag matrix domain in targeting human immunodeficiency virus type 1 assembly. *J.Virol.*, **74**, 2855-2866.
- Ott, D.E. (2002) Potential roles of cellular proteins in HIV-1. *Rev.Med.Virol.*, **12**, 359-374.
- Ott, D.E., Coren, L.V., Copeland, T.D., Kane, B.P., Johnson, D.G., Sowder, R.C., Yoshinaka, Y., Oroszlan, S., Arthur, L.O. and Henderson, L.E. (1998) Ubiquitin is covalently attached to the p6Gag proteins of human immunodeficiency virus type 1 and simian immunodeficiency virus and to the p12Gag protein of Moloney murine leukemia virus. *J.Virol.*, **72**, 2962-2968.
- Ouali, M., Laboulais, C., Leh, H., Gill, D., Desmaele, D., Mekouar, K., Zouhiri, F., d'Angelo, J., Auclair, C., Mouscadet, J.F. and Le Bret, M. (2000) Modeling of the inhibition of retroviral integrases by styrylquinoline derivatives. *J.Med.Chem.*, **43**, 1949-1957.
- Pahl, A. and Flugel, R.M. (1995) Characterization of the human spuma retrovirus integrase by site-directed mutagenesis, by complementation analysis, and by swapping the zinc finger domain of HIV-1 747. *J.Biol.Chem.*, **270**, 2957-2966.
- Pan, Y. and MacKerell, A.D., Jr. (2003) Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. *Nucleic Acids Res*, **31**, 7131-7140.
- Pannecouque, C., Pluymers, W., Van Maele, B., Tetz, V., Cherepanov, P., De Clercq, E., Witvrouw, M. and Debyser, Z. (2002) New class of HIV integrase inhibitors that block viral replication in cell culture. *Curr.Biol.*, **12**, 1169-1177.
- Parissi, V., Calmels, C., de Soultrait, V.R., Caumont, A., Fournier, M., Chaignepain, S. and Litvak, S. (2001) Functional interactions of human immunodeficiency virus type 1 integrase with human and yeast HSP60. *J.Virol.*, **75**, 11344-11353.
- Parkin, N.T., Chamorro, M. and Varmus, H.E. (1992) Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. *J.Virol.*, **66**, 5147-5151.
- Pauwels, R., Andries, K., Desmyter, J., Schols, D., Kukla, M.J., Breslin, H.J., Raeymaeckers, A., Van Gelder, J., Woestenborghs, R. and Heykants, J. (1990) Potent and selective inhibition of HIV-1 replication in vitro by a novel series of TIBO derivatives. *Nature*, **343**, 470-474.
- Perez, A., Noy, A., Lankas, F., Luque, F.J. and Orozco, M. (2004) The relative flexibility of B-DNA and A-RNA duplexes: database analysis. *Nucleic Acids Res*, **32**, 6144-6151.
- Perryman, A.L. and McCammon, J.A. (2002) AutoDocking dinucleotides to the HIV-1 integrase core domain: exploring possible binding sites for viral and genomic DNA. *J.Med.Chem.*, **45**, 5624-5627.
- Pluymers, W., Neamati, N., Pannecouque, C., Fikkert, V., Marchand, C., Burke, T.R., Jr., Pommier, Y., Schols, D., De Clercq, E., Debyser, Z. and Witvrouw, M. (2000) Viral entry as the primary target for the anti-HIV activity of chicoric acid and its tetra-acetyl esters. *Mol.Pharmacol.*, **58**, 641-648.

- Podtelezhnikov, A.A., Gao, K., Bushman, F.D. and McCammon, J.A. (2003) Modeling HIV-1 integrase complexes based on their hydrodynamic properties. *Biopolymers*, **68**, 110-120.
- Polanski, J., Zouhiri, F., Jeanson, L., Desmaele, D., d'Angelo, J., Mouscadet, J.F., Gieleciak, R., Gasteiger, J. and Le Bret, M. (2002) Use of the Kohonen neural network for rapid screening of ex vivo anti-HIV activity of styrylquinolines. *J.Med.Chem.*, **45**, 4647-4654.
- Pommier, Y., Johnson, A.A. and Marchand, C. (2005) Integrase inhibitors to treat HIV/AIDS. *Nat Rev Drug Discov*, **4**, 236-248.
- Popenda, M., Biala, E., Milecki, J. and Adamiak, R.W. (1997) Solution structure of RNA duplexes containing alternating CG base pairs: NMR study of r(CGCGCG)<sub>2</sub> and 2'-O-Me(CGCGCG)<sub>2</sub> under low salt conditions. *Nucleic Acids Res.*, **25**, 4589-4598.
- Price, D.H. (2000) P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol.Cell Biol.*, **20**, 2629-2634.
- Prive, G.G., Heinemann, U., Chandrasegaran, S., Kan, L.S., Kopka, M.L. and Dickerson, R.E. (1987) Helix geometry, hydration, and G.A mismatch in a B-DNA decamer. *Science*, **238**, 498-504.
- Puras Lutzke, R.A., Eppens, N.A., Weber, P.A., Houghten, R.A. and Plasterk, R.H. (1995) Identification of a hexapeptide inhibitor of the human immunodeficiency virus integrase protein by using a combinatorial chemical library. *Proc.Natl.Acad.Sci.U.S.A*, **92**, 11456-11460.
- Rao, S., Westhof, E. and Sundaralingam, M. (1981) Exact method for the calculation of pseudorotation parameters  $P$ ,  $[\tau]$  and their errors. A comparison of the Altona-Sundaralingam and Cremer-Pople treatment of puckering of five-membered rings. *Acta Crystallogr.A*, **37**, 421-425.
- Reicin, A.S., Kalpana, G., Paik, S., Marmon, S. and Goff, S. (1995) Sequences in the human immunodeficiency virus type 1 U3 region required for in vivo and in vitro integration. *J.Virol.*, **69**, 5904-5907.
- Reinke, R.A., King, P.J., Victoria, J.G., McDougall, B.R., Ma, G., Mao, Y., Reinecke, M.G. and Robinson, W.E., Jr. (2002) Dicafeoyltartaric acid analogues inhibit human immunodeficiency virus type 1 (HIV-1) integrase and HIV-1 replication at nontoxic concentrations. *J.Med.Chem.*, **45**, 3669-3683.
- Reinke, R.A., Lee, D.J., McDougall, B.R., King, P.J., Victoria, J., Mao, Y., Lei, X., Reinecke, M.G. and Robinson, W.E., Jr. (2004) L-chicoric acid inhibits human immunodeficiency virus type 1 integration in vivo and is a noncompetitive but reversible inhibitor of HIV-1 integrase in vitro. *Virology*, **326**, 203-219.
- Rice-Evans, C.E. and Packer, L. (1998) *Flavonoids in health and disease*. Marcel Dekker Inc., New York.
- Rice, P., Craigie, R. and Davies, D.R. (1996) Retroviral integrases and their cousins. *Curr.Opin.Struct.Biol.*, **6**, 76-83.
- Rice, P. and Mizuuchi, K. (1995) Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell*, **82**, 209-220.
- Rice, P.A. and Steitz, T.A. (1994) Refinement of gamma delta resolvase reveals a strikingly flexible molecule. *Structure*, **2**, 371-384.
- Robinson, W.E., Jr., Cordeiro, M., Abdel-Malek, S., Jia, Q., Chow, S.A., Reinecke, M.G. and Mitchell, W.M. (1996a) Dicafeoylquinic acid inhibitors of human immunodeficiency virus integrase: inhibition of the core catalytic domain of human immunodeficiency virus integrase. *Mol.Pharmacol.*, **50**, 846-855.

- Robinson, W.E., Jr., Reinecke, M.G., Abdel-Malek, S., Jia, Q. and Chow, S.A. (1996b) Inhibitors of HIV-1 replication [corrected; erratum to be published] that inhibit HIV integrase. *Proc.Natl.Acad.Sci.U.S.A.*, **93**, 6326-6331.
- Rowley, D.C., Hansen, M.S., Rhodes, D., Sotriffer, C.A., Ni, H., McCammon, J.A., Bushman, F.D. and Fenical, W. (2002) Thalassiolins A-C: new marine-derived inhibitors of HIV cDNA integrase. *Bioorg.Med.Chem.*, **10**, 3619-3625.
- Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints : molecular dynamics of n-alkanes. *J.Comput.Phys.*, **23**, 327-336.
- Saenger, W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.
- Saenger, W., Hunter, W.N. and Kennard, O. (1986) DNA conformation is determined by economics in the hydration of phosphate groups. *Nature*, **324**, 385-388.
- Scarlata, S. and Carter, C. (2003) Role of HIV-1 Gag domains in viral assembly. *Biochim.Biophys.Acta*, **1614**, 62-72.
- Schauer, M. and Billich, A. (1992) The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA-binding. *Biochem.Biophys.Res.Commun.*, **185**, 874-880.
- Schneider, B., Patel, K. and Berman, H.M. (1998) Hydration of the phosphate group in double-helical DNA. *Biophys J*, **75**, 2422-2434.
- Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R. and Bushman, F. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, **110**, 521-529.
- Scottoline, B.P., Chow, S., Ellison, V. and Brown, P.O. (1997) Disruption of the terminal base pairs of retroviral DNA during integration. *Genes Dev.*, **11**, 371-382.
- Sechi, M., Bacchi, A., Carcelli, M., Compari, C., Duce, E., Fisicaro, E., Rogolino, D., Gates, P., Derudas, M., Al-Mawsawi, L.Q. and Neamati, N. (2006) From ligand to complexes: inhibition of human immunodeficiency virus type 1 integrase by beta-diketo acid metal complexes. *J.Med.Chem.*, **49**, 4248-4260.
- Sherman, M.P. and Greene, W.C. (2002) Slipping through the door: HIV entry into the nucleus. *Microbes.Infect.*, **4**, 67-73.
- Sherman, P.A., Dickson, M.L. and Fyfe, J.A. (1992) Human immunodeficiency virus type 1 integration protein: DNA sequence requirements for cleaving and joining reactions. *J.Virol.*, **66**, 3593-3601.
- Sherman, P.A. and Fyfe, J.A. (1990) Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc.Natl.Acad.Sci.U.S.A.*, **87**, 5119-5123.
- Shibagaki, Y., Holmes, M.L., Appa, R.S. and Chow, S.A. (1997) Characterization of feline immunodeficiency virus integrase and analysis of functional domains. *Virology*, **230**, 1-10.
- Shindo, H., Fujiwara, T., Akutsu, H., Matsumoto, U. and Kyogoku, Y. (1985) Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 1. Static geometry of DNA double helices. *Biochemistry*, **24**, 887-895.
- Singh, S.B., Jayasuriya, H., Salituro, G.M., Zink, D.L., Shafiee, A., Heimbuch, B., Silverman, K.C., Lingham, R.B., Genilloud, O., Teran, A., Vilella, D., Felock, P. and Hazuda, D. (2001) The complestatins as HIV-1 integrase inhibitors. Efficient isolation, structure elucidation, and inhibitory activities of isocomplestatin, chloropectin I, new complestatins, A and B, and acid-hydrolysis products of chloropectin I. *J. Nat. Prod.*, **64**, 874-882.
- Singh, S.B., Zink, D.L., Bills, G.F., F., P., Teran, A., Collado, J., Silverman, K., Lingham, R., Felok, P. and Hazuda, D. (2002a) Discovery, structure and HIV-1 integrase inhibitory

- activities of integracins, novel dimeric alkyl aromatics from *Cytonaema* sp. *Tetrahedron Letters*, **43**, 1617.
- Singh, S.B., Zink, D.L., Heimbach, B., Genilloud, O., Teran, A., Silverman, K.C., Lingham, R.B., Felock, P. and Hazuda, D.J. (2002b) Structure, stereochemistry, and biological activity of integramycin, a novel hexacyclic natural product produced by *Actinoplanes* sp. that inhibits HIV-1 integrase. *Org Lett*, **4**, 1123-1126.
- Singh, S.B., Zink, D.L., Quamina, D.S., Pelaez, F., Teran, A., Felock, P. and Hazuda, D. (2002c) Integrastatins: structure and HIV-1 integrase inhibitory activities of two novel racemic tetracyclic aromatic heterocycles produced by two fungal species. *Tetrahedron Letters*, **24**, 2351-2354.
- Sotriffer, C.A., Ni, H. and McCammon, J.A. (2000) Active site binding modes of HIV-1 integrase inhibitors. *J.Med.Chem.*, **43**, 4109-4117.
- Stark, L.A. and Hay, R.T. (1998) Human immunodeficiency virus type 1 (HIV-1) viral protein R (Vpr) interacts with Lys-tRNA synthetase: implications for priming of HIV-1 reverse transcription 1. *J.Virol.*, **72**, 3037-3044.
- Stofer, E., Chipot, C. and Lavery, R. (1999) Free Energy calculations of Watson-Crick base pairing in aqueous solution. *J. Am. Chem. Soc.*, **121**, 9503-9508.
- Strebel, K. (2003) Virus-host interactions: role of HIV proteins Vif, Tat, and Rev. *AIDS*, **17 Suppl 4**, S25-S34.
- Takahashi, H., Matsuda, M., Kojima, A., Sata, T., Andoh, T., Kurata, T., Nagashima, K. and Hall, W.W. (1995) Human immunodeficiency virus type 1 reverse transcriptase: enhancement of activity by interaction with cellular topoisomerase I. *Proc.Natl.Acad.Sci.U.S.A*, **92**, 5694-5698.
- Taktakishvili, M., Neamati, N., Pommier, Y. and Nair, V. (2000) Recognition and inhibition of HIV integrase by a novel dinucleotide. *Bioorg.Med.Chem.Lett.*, **10**, 249-251.
- Taktakishvili, M., Neamati, N., Pommier, Y. and Nair, V. (2001) Discovery of a nuclease-resistant, non-natural dinucleotide that inhibits HIV-1 integrase. *Bioorg.Med.Chem.Lett.*, **11**, 1433-1435.
- Tang, C., Loeliger, E., Luncsford, P., Kinde, I., Beckett, D. and Summers, M.F. (2004) Entropic switch regulates myristate exposure in the HIV-1 matrix protein. *Proc.Natl.Acad.Sci.U.S.A*, **101**, 517-522.
- Thali, M., Bukovsky, A., Kondo, E., Rosenwirth, B., Walsh, C.T., Sodroski, J. and Gottlinger, H.G. (1994) Functional association of cyclophilin A with HIV-1 virions 1. *Nature*, **372**, 363-365.
- Tisne, C., Hantz, E., Hartmann, B. and Delepierre, M. (1998) Solution structure of a non-palindromic 16 base-pair DNA related to the HIV-1 kappa B site: evidence for BI-BII equilibrium inducing a global dynamic curvature of the duplex. *J Mol Biol*, **279**, 127-142.
- Tock, M.R., Frary, E., Sayers, J.R. and Grasby, J.A. (2003) Dynamic evidence for metal ion catalysis in the reaction mediated by a flap endonuclease. *EMBO J.*, **22**, 995-1004.
- Tolstorukov, M.Y., Ivanov, V.I., Malenkov, G.G., Jernigan, R.L. and Zhurkin, V.B. (2001) Sequence-dependent B<-->A transition in DNA evaluated with dimeric and trimeric scales. *Biophys.J.*, **81**, 3409-3421.
- Tolstorukov, M.Y., Jernigan, R.L. and Zhurkin, V.B. (2004) Protein-DNA hydrophobic recognition in the minor groove is facilitated by sugar switching. *J.Mol.Biol.*, **337**, 65-76.
- Trotta, E. and Paci, M. (1998) Solution structure of DAPI selectively bound in the minor groove of a DNA T.T mismatch-containing site: NMR and molecular dynamics studies. *Nucleic Acids Res.*, **26**, 4706-4713.

- Tsurutani, N., Kubo, M., Maeda, Y., Ohashi, T., Yamamoto, N., Kannagi, M. and Masuda, T. (2000) Identification of critical amino acid residues in human immunodeficiency virus type 1 IN required for efficient proviral DNA formation at steps prior to integration in dividing and nondividing cells. *J.Virol.*, **74**, 4795-4806.
- Turelli, P., Doucas, V., Craig, E., Mangeat, B., Klages, N., Evans, R., Kalpana, G. and Trono, D. (2001) Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. *Mol.Cell*, **7**, 1245-1254.
- Uesugi, S., Miki, H., Ikehara, M., Iwahashi, H. and Kyogoku, Y. (1979) A linear relationship between electronegativity of 2'-substituents and conformation of adenine nucleosides. *Tetrahedron Letters*, **42**, 4073-4076.
- Ugolini, S., Mondor, I. and Sattentau, Q.J. (1999) HIV-1 attachment: another look 3. *Trends Microbiol.*, **7**, 144-149.
- van den Ent, F.M., Vink, C. and Plasterk, R.H. (1994) DNA substrate requirements for different activities of the human immunodeficiency virus type 1 integrase protein. *J.Virol.*, **68**, 7825-7832.
- van den Ent, F.M., Vos, A. and Plasterk, R.H. (1999) Dissecting the role of the N-terminal domain of human immunodeficiency virus integrase by trans-complementation analysis. *J.Virol.*, **73**, 3176-3183.
- van Gent, D.C., Groeneger, A.A. and Plasterk, R.H. (1992) Mutational analysis of the integrase protein of human immunodeficiency virus type 2. *Proc.Natl.Acad.Sci.U.S.A.*, **89**, 9598-9602.
- van Gent, D.C., Vink, C., Groeneger, A.A. and Plasterk, R.H. (1993) Complementation between HIV integrase proteins mutated in different domains. *EMBO J.*, **12**, 3261-3267.
- Venkateswarlu, D., Lind, K.E., Mohan, V., Manoharan, M. and Ferguson, D.M. (1999) Structural properties of DNA:RNA duplexes containing 2'-O-methyl and 2'-S-methyl substitutions: a molecular dynamics investigation. *Nucleic Acids Res.*, **27**, 2189-2195.
- Ventura, M., Tarrago-Litvak, L., Dolle, V., Nguyen, C.H., Legraverend, M., Fleury, H.J. and Litvak, S. (1999) Effect of nucleoside analogs and non-nucleoside inhibitors of HIV-1 reverse transcriptase on cell-free virions. *Arch.Virol.*, **144**, 513-523.
- Vercammen, J., Maertens, G., Gerard, M., De Clercq, E., Debyser, Z. and Engelborghs, Y. (2002) DNA-induced polymerization of HIV-1 integrase analyzed with fluorescence fluctuation spectroscopy. *J.Biol.Chem.*, **277**, 38045-38052.
- VerPlank, L., Bouamr, F., LaGrassa, T.J., Agresta, B., Kikonyogo, A., Leis, J. and Carter, C.A. (2001) Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). *Proc.Natl.Acad.Sci.U.S.A.*, **98**, 7724-7729.
- Vincent, K.A., Ellison, V., Chow, S.A. and Brown, P.O. (1993) Characterization of human immunodeficiency virus type 1 integrase expressed in *Escherichia coli* and analysis of variants with amino-terminal mutations. *J.Virol.*, **67**, 425-437.
- Vink, C., Lutzke, R.A. and Plasterk, R.H. (1994) Formation of a stable complex between the human immunodeficiency virus integrase protein and viral DNA. *Nucleic Acids Res.*, **22**, 4103-4110.
- Vink, C., Oude Groeneger, A.M. and Plasterk, R.H. (1993) Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucleic Acids Res.*, **21**, 1419-1425.
- Vink, C., van Gent, D.C., Elgersma, Y. and Plasterk, R.H. (1991a) Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. *J.Virol.*, **65**, 4636-4644.

- Vink, C., Yeheskiely, E., van der Marel, G.A., van Boom, J.H. and Plasterk, R.H. (1991b) Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. *Nucleic Acids Res.*, **19**, 6691-6698.
- Violot, S., Hong, S.S., Rakotobe, D., Petit, C., Gay, B., Moreau, K., Billaud, G., Priet, S., Sire, J., Schwartz, O., Mouscadet, J.F. and Boulanger, P. (2003) The human polycomb group EED protein interacts with the integrase of human immunodeficiency virus type 1. *J.Virol.*, **77**, 12507-12522.
- Vodicka, M.A. (2001) Determinants for lentiviral infection of non-dividing cells. *Somat.Cell Mol.Genet.*, **26**, 35-49.
- Voet, D. and Rich, A. (1970) The crystal structures of purines, pyrimidines and their intermolecular complexes. *Prog.Nucleic Acid Res.Mol.Biol.*, **10**, 183-265.
- Wang, J.Y., Ling, H., Yang, W. and Craigie, R. (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.*, **20**, 7333-7343.
- Wang, L.D., Liu, C.L., Chen, W.Z. and Wang, C.X. (2005) Constructing HIV-1 integrase tetramer and exploring influences of metal ions on forming integrase-DNA complex. *Biochem.Biophys.Res.Commun.*, **337**, 313-319.
- Wang, T., Balakrishnan, M. and Jonsson, C.B. (1999) Major and minor groove contacts in retroviral integrase-LTR interactions 541. *Biochemistry*, **38**, 3624-3632.
- Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
- Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H. and Jones, K.A. (1998) A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell*, **92**, 451-462.
- Weston, S.A., Lahm, A. and Suck, D. (1992) X-ray structure of the DNase I-d(GGTATACC)<sub>2</sub> complex at 2.3 Å resolution. *J.Mol.Biol.*, **226**, 1237-1256.
- Wielens, J., Crosby, I.T. and Chalmers, D.K. (2005) A three-dimensional model of the human immunodeficiency virus type 1 integration complex. *J.Comput.Aided Mol.Des*, **19**, 301-317.
- Winger, R.H., Liedl, K.R., Rudisser, S., Pichler, A., Hallbrucker, A. and Mayer, E. (1998) B-DNA's B-I -> B-II conformer substate dynamics is coupled with water migration. *Journal of Physical Chemistry B*, **102**, 8934-8940.
- Woerner, A.M. and Marcus-Sekura, C.J. (1993) Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. *Nucleic Acids Res.*, **21**, 3507-3511.
- Wolfe, A.L., Felock, P.J., Hastings, J.C., Blau, C.U. and Hazuda, D.J. (1996) The role of manganese in promoting multimerization and assembly of human immunodeficiency virus type 1 integrase as a catalytically active complex on immobilized long terminal repeat substrates. *J.Virol.*, **70**, 1424-1432.
- Wu, H.M. and Crothers, D.M. (1984) The locus of sequence-directed and protein-induced DNA bending. *Nature*, **308**, 509-513.
- Wu, X., Li, Y., Crise, B. and Burgess, S.M. (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science*, **300**, 1749-1751.
- Wu, X., Liu, H., Xiao, H., Conway, J.A., Hehl, E., Kalpana, G.V., Prasad, V. and Kappes, J.C. (1999) Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through specific interactions with the nucleoprotein reverse transcription complex. *J.Virol.*, **73**, 2126-2135.
- Yang, Z.N., Mueser, T.C., Bushman, F.D. and Hyde, C.C. (2000) Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J.Mol.Biol.*, **296**, 535-548.

- Yoder, K.E. and Bushman, F.D. (2000) Repair of gaps in retroviral DNA integration intermediates 1. *J.Virol.*, **74**, 11191-11200.
- Yoshinaga, T. and Fujiwara, T. (1995) Different roles of bases within the integration signal sequence of human immunodeficiency virus type 1 in vitro. *J.Virol.*, **69**, 3233-3236.
- Yoshinaga, T., Kimura-Ohtani, Y. and Fujiwara, T. (1994) Detection and characterization of a functional complex of human immunodeficiency virus type 1 integrase and its DNA substrate by UV cross-linking. *J.Virol.*, **68**, 5690-5697.
- Yung, E., Sorin, M., Wang, E.J., Perumal, S., Ott, D. and Kalpana, G.V. (2004) Specificity of interaction of INI1/hSNF5 with retroviral integrases and its functional significance. *J.Virol.*, **78**, 2222-2231.
- Zennou, V., Petit, C., Guetard, D., Nerhbass, U., Montagnier, L. and Charneau, P. (2000) HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell*, **101**, 173-185.
- Zhao, H., Neamati, N., Sunder, S., Hong, H., Wang, S., Milne, G.W., Pommier, Y. and Burke, T.R., Jr. (1997) Hydrazide-containing inhibitors of HIV-1 integrase. *J.Med.Chem.*, **40**, 937-941.
- Zhao, L., O'Reilly, M.K., Shultz, M.D. and Chmielewski, J. (2003) Interfacial peptide inhibitors of HIV-1 integrase activity and dimerization. *Bioorg.Med.Chem.Lett.*, **13**, 1175-1177.
- Zheng, R., Jenkins, T.M. and Craigie, R. (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc.Natl.Acad.Sci.U.S.A*, **93**, 13659-13664.
- Zhu, H.M., Chen, W.Z. and Wang, C.X. (2005) Docking dinucleotides to HIV-1 integrase carboxyl-terminal domain to find possible DNA binding sites. *Bioorg.Med.Chem.Lett.*, **15**, 475-477.
- Zhu, K., Dobard, C. and Chow, S.A. (2004) Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase. *J.Virol.*, **78**, 5045-5055.
- Zhuang, L., Wai, J.S., Embrey, M.W., Fisher, T.E., Egbertson, M.S., Payne, L.S., Guare, J.P., Jr., Vacca, J.P., Hazuda, D.J., Felock, P.J., Wolfe, A.L., Stillmock, K.A., Witmer, M.V., Moyer, G., Schleif, W.A., Gabryelski, L.J., Leonard, Y.M., Lynch, J.J., Jr., Michelson, S.R. and Young, S.D. (2003) Design and synthesis of 8-hydroxy-[1,6]naphthyridines as novel inhibitors of HIV-1 integrase in vitro and in infected cells. *J.Med.Chem.*, **46**, 453-456.
- Zouhiri, F., Mouscadet, J.F., Mekouar, K., Desmaele, D., Savoure, D., Leh, H., Subra, F., Le Bret, M., Auclair, C. and d'Angelo, J. (2000) Structure-activity relationships and binding mode of styrylquinolines as potent inhibitors of HIV-1 integrase and replication of HIV-1 in cell culture. *J.Med.Chem.*, **43**, 1533-1540.