

Propriétés Structurales et dynamiques de fragments d'ADN reconnus préférentiellement par des protéines (Intégrase du VIH-1, Topoisomérase II). Etudes par RMN et modélisation moléculaire.

Cosquer, S., Renisio, J-G., El Antri, S., René, B., & Mauffret, O.

Conférence de Olivier MAUFFRET UMR 8113 CNRS LBPA ENS CACHAN

Résumé

Les fragments des génomes reconnus par des protéines se caractérisent souvent par des propriétés structurales et dynamiques originales. L'étude de ces dernières se fait en solution en utilisant principalement les techniques de Résonance magnétique nucléaire. Les progrès réalisés ces dernières années dans ces méthodes se situent au niveau 1) de la mise au point de techniques d'enrichissement des molécules en ^{13}C et ^{15}N qui permettent d'augmenter le nombre d'informations sur celles-ci ; 2) de la mesure de paramètres d'un nouveau type : les couplages dipolaires résiduels qui constituent un nouvel outil pour les études structurales. L'utilisation de ces méthodes sur les systèmes que nous étudions au laboratoire : séquences terminales du génome du VIH-1 reconnues spécifiquement par l'intégrase du VIH-1 et séquences d'ADN préférentiellement reconnues et clivées par la topoisomérase II, nous a permis de progresser de façon importante dans la compréhension des mécanismes fins de reconnaissance mais aussi de coupure de l'ADN par ces enzymes. Dans le cas des séquences LTR (Long Terminal Repeat) du HIV-1 clivées très spécifiquement par l'intégrase du VIH, nous avons pu faire apparaître une pré-organisation de l'ADN qui le rend parfaitement apte à être traité de façon optimale par l'enzyme. La dynamique moléculaire en eau et ions explicites utilisant le programme AMBER permet d'avancer dans le détail des interactions et apportent des informations sur ces partenaires essentiels de la coupure que sont les ions (notamment divalents) et les molécules d'eau. Pour les ADN reconnus par les topoisomérases, des mouvements moléculaires au niveau de certains sites semblent pouvoir être corrélés avec la reconnaissance de ceux-ci par l'enzyme et sont perceptibles à la fois par les méthodes de RMN et de dynamique moléculaire.

Mots-Clés : Interactions ADN-protéines, RMN, dynamique moléculaire, Intégrase du VIH-1, Topoisomérase II

Introduction

La reconnaissance ADN-protéine est un processus fondamental dans l'expression et la transmission de l'information génétique. Ses bases moléculaires et les lois physicochimiques auxquelles elle est soumise sont encore insuffisamment décrites et comprises. Dans cette optique, il nous semble nécessaire de préciser l'image non seulement du complexe mais aussi celles des partenaires isolés. Notons aussi que la compréhension du fonctionnement de protéines comme la topoisomérase II ou l'ADN polymérase qui sont de véritables nanomachines moléculaires, impose de connaître les conformations à différentes étapes du cycle catalytique. Dans le cadre de nos travaux nous étudions plus particulièrement les propriétés structurales et dynamiques de fragments d'ADN afin de percevoir comment l'information est disponible pour la reconnaissance protéique. La comparaison des ADN nus et complexés est à notre sens nécessaire pour décrypter les principes de cette dernière.

La méthode que nous utilisons au laboratoire est la résonance magnétique nucléaire (RMN) qui permet d'étudier les molécules en solution contrairement à la cristallographie aux rayons X qui les étudie en milieu cristallin.

Nous avons utilisé plusieurs développements méthodologiques récents de RMN et de calculs et les avons adaptés à nos thématiques. Les applications ont porté sur deux systèmes : celui de l'intégrase du VIH-1 (Virus de l'immunodéficience humaine) et de la topoisomérase II.

Matériel et Méthodes

- Le marquage des oligonucléotides ADN en isotopes stables ^{15}N et ^{13}C en utilisant les méthodes de PCR (Polymerase Chain Reaction)
- L'utilisation des couplages dipolaires résiduels à la détermination de la structure tridimensionnelle des ADN (Figure 1).
- Dynamiques moléculaires en eau et ions explicites avec le programme AMBER

Les techniques de RMN sont efficacement complétées par les techniques de simulation de dynamique moléculaire *in silico*. Le programme AMBER permet l'obtention de trajectoires stables à l'échelle de plusieurs dizaines de nanosecondes et l'investigation de la dynamique de la molécule. L'impact de l'eau et des contre-ions peut aussi être

examinée. Les données obtenues par cette approche peuvent dans un deuxième temps être confrontées à celles issues de la RMN et constituer ainsi une vision très complète de la dynamique de la molécule, cette dernière ayant un rôle évident dans la reconnaissance ADN-protéine selon des règles et des mécanismes qu'il nous appartient maintenant d'explorer.

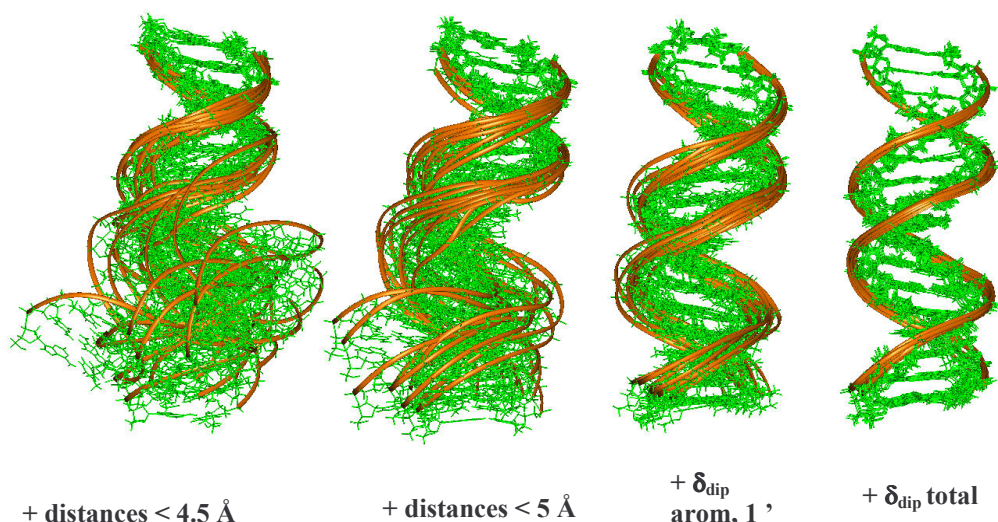


Figure 1 : Simulations sur ordinateur de la reconstruction de la structure d'un oligonucléotide de 17 paires de bases correspondant à l'extrémité des séquences LTR (Long terminal Repeat) du génome du VIH-1. La reconstruction est faite en utilisant plusieurs sous-ensembles de contraintes : avec les distances uniquement (579 distances) les deux groupes de conformères les plus à gauche, en ajoutant les couplages dipolaires (50 valeurs, troisième à partir de la gauche et 120 valeurs, quatrième à partir de la gauche). Chaque groupe présente les dix meilleurs conformères. Ces résultats illustrent l'intérêt des CDR, même en nombre limité, pour la détermination de la structure des ADN. (Mauffret *et al.*, 2002)

Résultats et discussion

a) séquences LTR (Long Terminal Repeat) reconnues par l'intégrase du VIH-1 (Figure 2).

L'intégrase du VIH-1 est responsable de l'intégration du génome viral dans le génome cellulaire. La réaction se fait en deux étapes : dans un premier temps (étape de maturation) l'ADN viral rétrotranscrit est reconnu par l'enzyme au niveau de ses extrémités terminales qui portent les séquences consensus CAGT.

Le dinucléotide GT est excisé de chacune des extrémités du génome le rendant ainsi apte à effectuer la deuxième étape : le transfert de brin qui constitue l'étape d'intégration proprement dite.

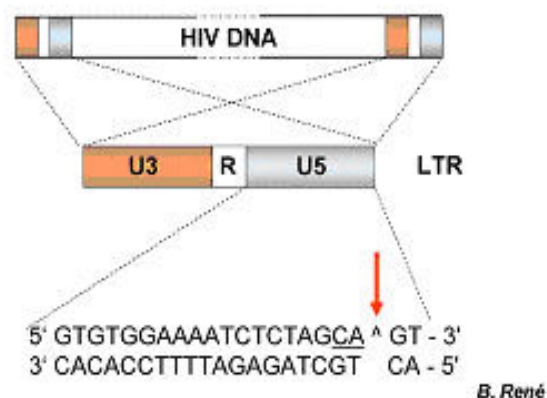


Figure 2 : Extrémités des séquences LTR U5 (Long Terminal Repeat) présentes à l'extrémité du génome du VIH-1 reconnues et coupées spécifiquement par l'intégrase du rétrovirus

Nous nous intéressons plus particulièrement ici à la première étape et à la reconnaissance spécifique de l'extrémité du génome par un sous domaine de l'enzyme situé dans le cœur catalytique. Il a pu être montré que les deux partenaires : 1) les 17 dernières paires de bases du génome et 2) un motif extrait du cœur catalytique (l'hélice $\alpha 4$) interagissent l'un avec l'autre avec une forte affinité. Nous menons des études structurales afin de mieux comprendre les raisons de cette spécificité puisque ces séquences sont reconnues et coupées par l'enzyme avec une efficacité de 5 à 6 ordres de grandeur par rapport à une autre séquence. Nous disposons de la structure de l'hélice $\alpha 4$ qui est le principal partenaire de l'interaction du côté protéique. La structure d'un oligonucléotide de 17 paires de bases a été examinée en utilisant conjointement les techniques de RMN et de dynamique moléculaires. Nos calculs avec les constantes de couplages dipolaires résiduels (CDR) associées aux contraintes de type classiques, effets Overhauser Nucléaires et constantes de couplages scalaires, ont permis d'obtenir des structures à très haute résolution. Les structures sont améliorées au niveau global mais aussi local. Les paramètres hélicoïdaux décrivant la structure fine de la double

hélice sont déterminés avec une bien meilleure précision.

Il apparaît que le site de clivage, relativement au reste de la molécule, est caractérisé par une structure très hétérogène avec des déviations importantes par rapport à la géométrie de l'ADN-B canonique. Le site de clivage CAGT présente une courbure importante vers le grand sillon ouvrant le petit sillon avec des conformations locales se rapprochant de l'ADN-A. Ce type de conformation a été retrouvé dans de nombreux ADN reconnus et coupés du côté du petit sillon par des protéines. Il est à noter que ces caractéristiques étaient difficilement perceptibles en considérant les structures calculées sans les couplages dipolaires. Une autre particularité intéressante est la faible distance entre le phosphate coupé et l'un de ses voisins immédiats sur le même brin, elle nous paraît apte à favoriser la catalyse selon des mécanismes subtils faisant notamment intervenir des molécules d'eau. Ces différentes propriétés sont liées à un empilement interbrin des bases G21 et A15 ce qui est tout à fait atypique dans l'ADN B (Figure 3). Cet empilement interbrin est corrélatif d'un empilement intrabrin très faible au niveau du site de clivage.

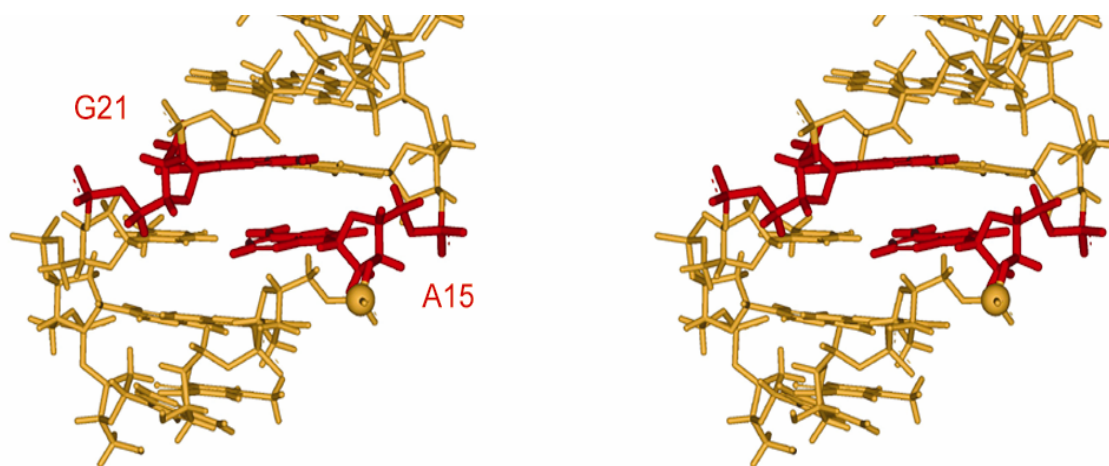


Figure 3 : Vue en stéréo de la région du site clivé de l'extrémité terminale du LTR-U5 déterminée par RMN. (d G₁GAAA₅ATCTC₁₀TAGC₁₄A₁₅GT₁₇. A₁₈CTG₂₁CTAG₂₅AGATT₃₀TTCC). Le phosphate clivé est indiqué par une boule et les bases en empilement interbrin en rouge et en mode « batonnet ». (Renisio *et al.*, 2005).

Il est intéressant de noter que ces empilements intra-brin faibles au site clivé ont été également perçus par des techniques biochimiques. Les méthodes RMN permettent l'étude de la conformation du squelette phosphodiester en utilisant la RMN du ³¹P. Les spectres réalisés établissent que les sites A15pG16 et G21pC22 se caractérisent par des proportions de forme BII élevées. La conformation des phosphates dans un brin d'ADN possèdent deux états : l'état BI, majoritaire et l'état BII minoritaire; ils se

caractérisent par les conformations des angles ϵ et ζ du squelette et sont en interconversion rapide sur des temps de l'ordre de la ps-ns. Cette échelle est compatible avec les simulations de dynamique moléculaire et ces mouvements du squelette sont effectivement observés. Les résultats des deux approches convergent et établissent que le brin coupé est particulièrement riche en conformation BII au voisinage du site de clivage, le brin non coupé est

caractérisé par un unique site très riche en forme BII : l'enchaînement G21pC22.

Nous avons tenté d'opérer une synthèse de ces informations en utilisant les structures du cristal du cœur catalytique de l'intégrase, la structure de notre oligonucléotide obtenue par RMN et la structure du complexe de la Transposase Tn5 avec l'ADN, cette dernière protéine constituant un système enzymatique très proche de l'intégrase.

Cette analogie nous a permis de construire un modèle d'interaction qui est en excellent accord avec les données expérimentales et théoriques (Figure 4). L'hélice reconnaît l'ADN dans le petit sillon qui est précisément élargi au niveau du site de clivage, les points de contact avec le squelette phosphodiester

sont ceux qui présentent des conformations du squelette phosphodiester atypiques. Le « désempilement » des bases dans le brin au niveau du site coupé donne une flexibilité supplémentaire qui permet la bonne adaptation des deux partenaires. Enfin le rapprochement des phosphates au site CAG favorise la catalyse par l'action conjuguée du E152 et du K159 deux résidus très conservés dans la protéine. On voit ainsi comment l'apport des données RMN et notamment des couplages dipolaires permet la visualisation de détails fins qui sont nécessaires pour saisir le mécanisme de la reconnaissance et de la catalyse. Nous étudions maintenant les complexes entre les deux partenaires ainsi que les effets des ions magnésium sur l'interaction.

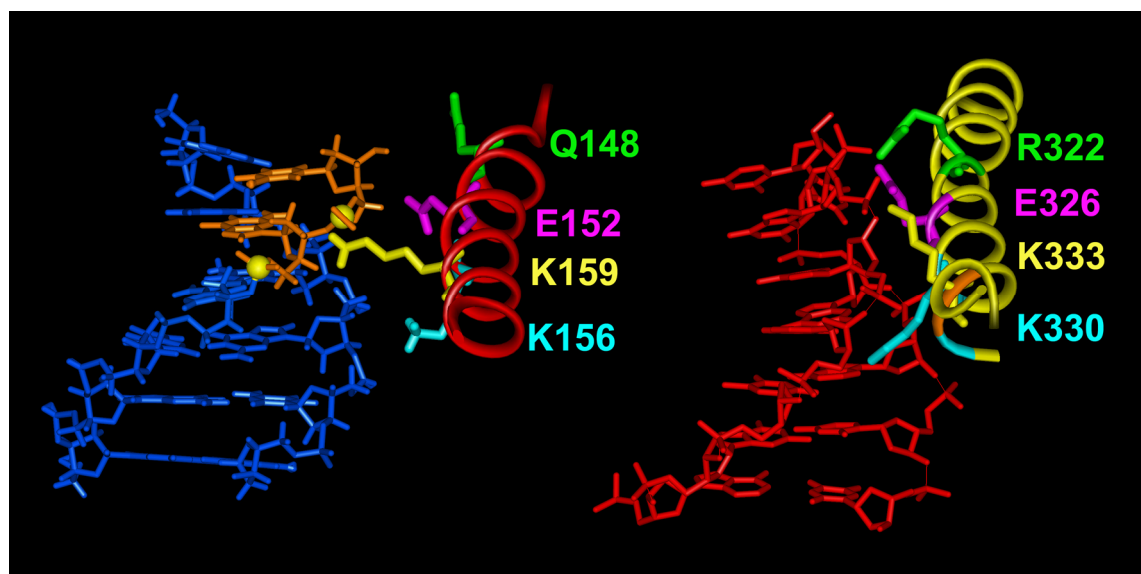


Figure 4 : Modèle d'interaction entre l'extrémité de la séquence LTR de l'ADN retroviral et l'hélice $\alpha 4$ de l'intégrase du VIH-1 responsable de la reconnaissance spécifique de l'ADN (Gauche). Structure du complexe entre l'hélice correspondante dans la transposase et l'ADN (Davies *et al.* 2000).

b) Séquences d'ADN reconnues préférentiellement par l'ADN topoisomérase II

L'ADN topoisomérase II est une enzyme qui contrôle la topologie de l'ADN et qui est de ce fait indispensable aux opérations de transcription, réplication, recombinaison et réparation. Elle procède en effectuant une coupure double brin dans l'ADN puis en y faisant passer un autre double brin après quoi l'ADN est ressoudé. Nous avons identifié dans le plasmide pbR322 un site de coupure fort de l'enzyme et avons étudié les propriétés de cette séquence par les méthodes décrites ci-dessus.

Le couplage des méthodes de RMN et de dynamique moléculaire a permis de montrer que le site de clivage de la topoisomérase sur l'ADN se caractérise par une exceptionnelle flexibilité visualisée notamment au niveau du squelette phosphodiester. Les paramètres de la double hélice sont à ce niveau très hétérogènes

avec des amplitudes importantes dans leur variation. Nous avons pu analyser la relation de ces mouvements avec l'eau et les contre-ions ; nos données révèlent notamment le rôle important des cations sur les mouvements de l'ADN. Le fait que l'ADN soit clivé précisément au site où des mouvements importants sont détectés indique la « fragilité » de cette séquence. La détermination des causes exactes du clivage préférentiel nécessite évidemment des expériences complémentaires effectuées avec le partenaire protéique sous forme d'un fragment de topoisomérase plus particulièrement impliqué dans la reconnaissance de l'ADN. La sélection d'un tel site peut se faire à différents niveaux : l'étape de reconnaissance et de clivage de l'ADN par l'enzyme peut être favorisée par les mouvements au site de clivage ; on peut aussi supposer que l'étape de religation de l'ADN soit

rendue problématique du fait des mouvements importants qui gêneraient l'appariement des deux brins. Les deux types de mécanismes conduiront à un taux de coupure accrue.

Conclusion

Nos études sur les fragments d'ADN reconnus et coupés par l'intégrase et la topoisomérase II nous donnent une nouvelle image de la reconnaissance ADN-protéine notamment en nous donnant accès à l'aspect dynamique de cette opération. L'étude des mêmes oligonucléotides avec leurs partenaires protéiques constituera la prochaine étape de ce travail.

Bibliographie

- Davies, DR, Goryshin, IY, Reznikoff, WS., Rayment, I. (2000) *Science*. 289: 77-85.
- Mauffret, O. Tevanian, G. & Femandjian, S. (2002) *J.Biomolecular NMR* 24 : 317-328.
- Renisio, JG, Cosquer, S., Cherrak, I., El Antri, S., Mauffret, O., Femandjian, S. (2005) *Nucl.Acids.Res.* 33 :1970-1981.