

DEA ANALYSE DE GÉNOMES ET MODÉLISATION MOLÉCULAIRE
UNIVERSITÉ PARIS VII - DENIS DIDEROT

STAGE RÉALISÉ AU LABORATOIRE DE BIOCHIMIE THÉORIQUE
INSTITUT DE BIOLOGIE PHYSICO-CHIMIQUE

*Etude des interactions ADN-protéines
d'un point de vue énergétique*

Stéphane Teletchéa

Stage réalisé de janvier à juin 2000
sous la direction de Richard Lavery

Remerciements

Je tiens à remercier le Docteur Richard Lavery pour m'avoir accueilli au sein de son laboratoire. Ses conseils avisés et les réflexions qu'il m'a incité à mener m'ont permis de découvrir la structure de l'ADN telle qu'elle est connue et étudiée.

Parallèlement, j'ai pu appréhender la portée, les promesses mais aussi les limites de la modélisation moléculaire.

Hormis les conseils et critiques avisés de Richard, il y avait bien sûr Alexey, Anne, Brigitte, Chantal, Edna, Krystina, Thérèse, Youri, les docteurs présents dans le laboratoire, qui m'ont souvent été d'un grand secours pour trouver rapidement quelqu'un pouvant répondre à mes interrogations.

J'ai eu la chance de côtoyer des thésards à différents stades de leur étude, ce qui m'a donné un bon point de vue sur le déroulement de cet exercice de longue haleine. Merci donc à Elodie, Emmanuel, Ingrid, Guillaume et Raphaël (« fais 'man' ... »).

Enfin, merci à Daniel et Roger pour leur chaleureux accueil.

Je dois dire que l'atmosphère du laboratoire, tout en étant sérieuse, dégageait une sérénité qui rend propice la réflexion, et rend le travail encore plus agréable.

Résumé

Former et optimiser la structure d'un complexe protéine-ADN est un exercice inaccessible dans le cadre d'une simulation par dynamique moléculaire. Pourtant, de nombreuses informations pourraient être exploitées s'il était possible de calculer et d'analyser la partie énergétique d'une telle interaction. Grâce à l'emploi de coordonnées internes et d'une représentation implicite du solvant, JUMNA permet d'approcher un élément de ce problème, à savoir l'énergétique de la déformation de l'ADN au sein de ces complexes. Durant ce stage, nous nous sommes attachés à atteindre ce but, par la conception d'un moule moléculaire sur lequel l'ADN est contraint à se déformer. Nous avons ainsi pu reproduire fidèlement la déformation induite par différentes protéines au sein des structures cristallographiques. Cette méthodologie nous a permis d'obtenir de nombreux résultats sur l'énergie de déformation de l'ADN et peut par la suite faire partie d'une approche globale pour obtenir l'énergie libre de complexation (collaboration en cours avec le groupe de David Beveridge).

Abstract

Forming and optimising the structure of a protein-DNA complex by molecular dynamics simulation is not feasible. However, much useful information could be obtained if we could calculate and analyse the energetics of such interactions. By using internal coordinates and an implicit solvent model, JUMNA is able to treat one element of this problem, namely the energetics of DNA deformation within such complexes. During this research project, we have attempted to reach this goal by creating "molecular moulds" with which DNA can be deformed. In this way we have accurately reproduced the deformation induced by different proteins within crystallographic complexes. This methodology enabled us to obtain many results on the deformation energy of DNA and can subsequently be included in a global approach to obtaining free energies of complexation (ongoing collaboration with the group of David Beveridge).

Table des matières

1	Introduction	1
2	Méthodologie	2
2.1	CONTACT	2
2.2	JUMNA	4
2.3	CURVES	6
2.4	Comment décrire l'interaction?	6
2.4.1	La constante de force	6
2.4.2	La précision de positionnement des atomes	7
2.4.3	La distance seuil définissant les atomes du moule	7
2.4.4	Structure des brins	8
2.4.5	Validation	8
3	Résultats	9
3.1	Déformation de l'ADN	9
3.2	Composantes de l'énergie de déformation	10
3.3	Cas des protéines dimériques	12
3.4	Mutation	14
4	Conclusion	15

1 Introduction

Ce projet porte sur l'étude théorique de la reconnaissance entre les protéines et l'ADN. Malgré le nombre croissant de complexes dans les bases de données, les mécanismes d'interaction expliquant celle-ci ne sont toujours pas connus. Comme ces mécanismes sont à la base de la régulation des gènes, leur compréhension devient de plus en plus nécessaire afin d'exploiter les informations issues du décryptage de génomes entiers.

Il est d'ores et déjà possible de caractériser les propriétés structurales de plusieurs familles de protéines qui reconnaissent l'ADN avec une spécificité inégale. [1, 2]. Cependant comme dans la plupart des cas l'énergétique de ces interactions n'est pas accessible, il faut identifier ses différentes composantes. Ainsi interviennent l'enthalpie d'interaction protéine-ADN, la perte d'entropie associée, le réarrangement des molécules du solvant et des contre-ions, des changements de conformations des deux molécules.

Malheureusement, les simulations par dynamique moléculaire – qui permettent en principe les calculs de différence d'énergie libre – sont inefficaces dans ce cas puisque la formation d'un complexe macromoléculaire est à la fois trop lente et implique des mouvements trop importants pour être simulée durant un temps raisonnable. Pour pallier à ces inconvénients, il est nécessaire de recourir à des approximations permettant de diminuer la quantité de calculs nécessaires. Ainsi en adoptant une représentation implicite du solvant et des contre-ions, il devient possible d'estimer les différents facteurs constitutifs de l'énergie libre de formation du complexe.

Beveridge et ses collaborateurs [3, 4] ont tenté une telle approche, il en ressort des informations sur l'importance relative des différentes composantes de la complexation avec parfois des surprises (par exemple, la faible importance des interactions électrostatiques). Néanmoins certains facteurs sont difficiles à calculer, en particulier l'enthalpie de déformation de l'ADN et de la protéine lors de leur interaction.

Lors de ce stage, je me suis consacré à calculer les diverses contributions intervenant uniquement du côté de l'ADN en mettant au point une méthodologie originale. Les résultats obtenus nous permettent de caractériser la déformation subie par l'ADN dans différentes familles de complexes et confirment que la flexibilité de celui-ci en fonction de sa séquence joue un rôle significatif dans la reconnaissance par la protéine. Cette approche – qui peut être étendue à l'énergie de déformation du côté protéique – est destinée à être intégrée dans l'analyse des complexes entreprise par Beveridge et ses collaborateurs.

A terme, la méthodologie ainsi élaborée doit permettre de prédire l'énergie libre lors de la fixation d'une protéine sur sa séquence cible ainsi que les conséquences d'une mutation de paires de bases du côté nucléique ou d'acides aminés du côté protéique sur la stabilité du complexe ainsi modifié.

2 Méthodologie

Les complexes protéine-acide nucléique étudiés sont issus de données cristallographiques tirées de la NDB (Nucleic acid Data Bank). Ils ont été choisis afin de constituer une base de données non redondante, avec une résolution inférieure à 3 Å. Leur description est présentée dans le tableau 1. Aux fins de comparaison, seules les structures étudiées également par Beveridge et ses collaborateurs [4] ont été analysées. Les 20 complexes appartiennent aux familles suivantes :

- ADN polymérases: eco RV, terminateur de réplication, RAP I.
- facteurs de transcription procaryotes: répresseur du phage lambda, répresseur cro du phage 434, répresseur tryptophane (trp), répresseur méthionine (met).
- facteurs de transcription eucaryotes:
 1. homéodomains: mat $\alpha 1$ et mat $\alpha 2$, paired dimer, even-skipped domain, 'engrailed', POU, domaine myod.
 2. protéine liant le zinc: ZIF 268, tramtrack, five-finger gli, régulateur I de la voie des pyrimidines.
 3. protéine se liant sur la TATA box: TBP.
 4. Récepteur d'hormones: récepteur des œstrogènes.
- recombinase: recombinase HIN.

2.1 CONTACT

Pour décrire efficacement l'interface entre une protéine et sa séquence cible d'ADN, un important travail a déjà été fait par Janin et ses collaborateurs [1]. Il en ressort des données sur la surface d'interaction, les caractéristiques des liaisons hydrogènes et ponts salins, la densité atomique à l'interface. Suite à ces analyses, Janin et ses collaborateurs en déduisent que l'interaction est essentiellement due à une reconnaissance fine de l'anatomie entre l'ADN et la protéine.

Partant de cette constatation, nous avons défini une zone de contact entre les deux molécules. Le programme CONTACT nous a permis de définir celle-ci.

En effet il compare la distance séparant chaque paire d'atomes (un pour la protéine, un pour l'ADN). Si celle-ci est inférieure ou égale à une distance seuil (4,5 Å) cette paire est retenue comme faisant partie de l'interface. La zone de contact ainsi créée par les atomes de l'ADN forme un moule dont une illustration est présentée figure 1. Pour contraindre la structure de l'ADN libre à s'adapter au moule, nous imposons des contraintes de distance entre l'atome de l'ADN et sa position au sein du moule (après une superposition préalable du moule sur la conformation initiale de l'ADN). Si un atome de l'ADN est éloigné de la

No	Réf. NDB ^a	Nom du complexe	M/D	ΔE_{def}	ΔC	RMSD
1	e008	répresseur met	D	14	5	0.9
2	t011	tramtrack DNA : binding domain	M	17	7	1.8
3	t039	peptide en doigts de zinc ZIF 268	M	17	6	0.6
4	r001	répresseur cro	D	18	0	1.0
5	t016	domaine myod BHLH	D	20	-7	1.2
6	rc03	récepteur aux œstrogènes	D	21	-21	0.7
7	r009	répresseur trp	D	22	6	0.8
8	r018	paired dimer	M	23	-2	0.8
9	t019	domaine POU	M	26	11	1.0
10	r010	répresseur lambda	D	34	33	0.7
11	t028	mat $\alpha 1$ & mat $\alpha 2$ homéodomaines	D	37	36	0.7
12	t031	even-skipped domain	D	38	-1	0.7
13	t017	régulateur I de la voie des pyrimidines	D	42	-18	0.9
14	t004	domaine 'Engrailed'	D	42	-15	1.3
15	t035	RAP I, ligand des télomères	M	55	24	0.8
16	e003	eco RV	D	57	38	0.5
17	t009	TATA box binding protein (TBP)	M	64	66	1.1
18	te01	terminateur de réplication	M	68	26	0.7
19	e009	recombinase HIN	M	78	9	0.6
20	t008	five-finger gli	M	102	3	1.5

TAB. 1 – Complexes analysés lors de cette étude. Les données correspondant à la partie nucléique du complexe. (ΔE_{def} : énergie de déformation en kcal.mole⁻¹; ΔC : variation de la courbure lors de la complexation, en degrés). RMSD écart quadratique moyen entre l'ADN contraint et la structure de l'ADN au sein du complexe cristallographique. M/D : la protéine dans le complexe est monomérique/dimérique.

^a Dans la NDB, chaque nom de complexe a pour préfixe «pd»

position qu'il occupe dans le moule, une pénalité quadratique en fonction de la distance est appliquée.

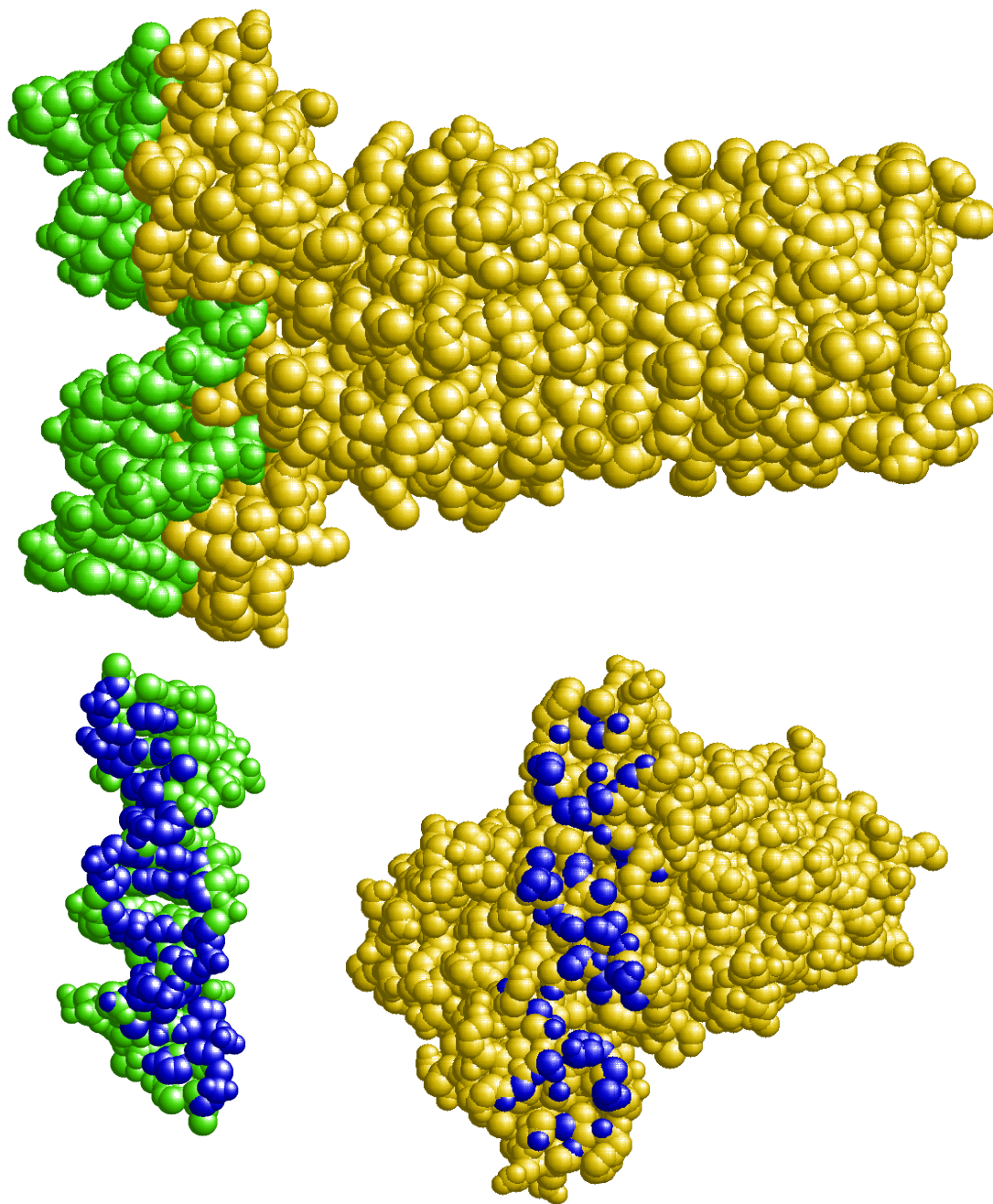


FIG. 1 – Atomes impliqués dans l'interaction ADN-Protéine pour un répresseur de la voie des purines. En jaune : atomes de la protéine ; en vert : atomes de l'ADN ; en bleu : ceux formant le moule.

2.2 JUMNA

JUMNA (JUNCTION Minimization of Nucleic Acids) [5, 6, 7] est un logiciel de minimisation utilisant un système de coordonnées internes comprenant les angles dièdres et de valence ainsi que des variables représentant les paramètres hélicoïdaux de l'ADN.

Chaque nucléotide est relié au suivant par la reconstruction du brin phosphodiester entre le monophosphate en 3' de la base n et le méthylène C5' de la base $n + 1$. Le dernier nucléotide ne comporte pas de groupe phosphate en 3'.

L'énergie totale de l'ADN est ensuite minimisée par un algorithme de type gradient conjugué. L'énergie est obtenue en appliquant le champ de force FLEX [7], développé au laboratoire.

$$\begin{aligned}
E &= \sum_{i,j} \frac{Q_i Q_j}{\epsilon_{(R)} R_{ij}} + \sum_{i,j} \left(\frac{-A_{ij}}{R_{ij}^6} + \frac{B_{ij}}{R_{ij}^{12}} \right) \\
&+ \sum_{i,j} \left(\cos \theta \left(\frac{-A_{ij}^{HB}}{R_{ij}^6} + \frac{B_{ij}^{HB}}{R_{ij}^{12}} \right) + (1 - \cos \theta) \left(\frac{-A_{ij}}{R_{ij}^6} + \frac{B_{ij}}{R_{ij}^{12}} \right) \right) \\
&+ \sum_{i,j} \frac{V_s}{2} (1 \pm \cos N_s \theta_s) + \sum_{i,j} V_a (\sigma_a - \sigma_{a^0})^2
\end{aligned} \tag{1}$$

Il est à noter le traitement implicite des effets du solvant et des contre-ions. En effet, traditionnellement en dynamique moléculaire, les molécules d'eau et contre-ions d'un système sont inclus de manière explicite dans le calcul, ce qui revient à multiplier le nombre de variables du système d'un facteur d'environ 16. Dans le cas d'un ADN de 15 paires de bases (1000 atomes), il y a 5000 molécules d'eau à ajouter afin d'obtenir une bonne hydratation de l'ADN.

Afin de diminuer le nombre d'atomes à traiter, il est possible de considérer le solvant comme un milieu diélectrique continu. Ce traitement rudimentaire implique une fonction diélectrique à dépendance sigmoïdale :

$$\begin{aligned}
\epsilon_{(R)} &= D - \frac{(D-1)}{2} \left((RS)^2 + 2RS + 2 \right) \exp^{(-RS)} \\
D &: \text{valeur de la constante diélectrique au plateau (78)} \\
S &: \text{pente de la courbe (0.356)} \\
R &: \text{distance entre les atomes considérés.}
\end{aligned} \tag{2}$$

L'effet des contre-ions est pris en compte par la réduction de la charge totale des phosphates à -0,5.

A l'avenir, il sera important d'améliorer ce modèle. Malheureusement, des calculs Poisson-Boltzmann précis ne sont pas possibles pour des systèmes aussi grands que nos complexes. Nous sommes donc en train de développer une approche de type « Generalized Born » [8] qui pourrait être intégrée dans JUMNA.

Il est important d'ajouter que cette étude de la déformation de l'ADN est possible uniquement à cause de la représentation en coordonnées internes adoptée dans JUMNA. Ce choix, en contraste avec l'emploi des coordonnées cartésiennes [9, 10] autorise l'étude de grands changements de conformation par simple minimisation d'énergie.

2.3 CURVES

C'est un logiciel qui permet de définir un axe idéal autorisant une représentation unique de l'ADN selon ses paramètres hélicoïdaux. [11, 12, 13] Grâce aux nombreuses informations fournies, il m'a permis de déterminer quantitativement les déformations induites dans la double hélice.

2.4 Comment décrire l'interaction ?

Pour créer un moule capable de déformer l'ADN correctement, il est nécessaire de trouver les paramètres optimaux qui le définissent. Ils y a trois paramètres à étudier :

1. la constante de force appliquée aux contraintes ;
2. la précision avec laquelle les atomes de l'ADN doivent reproduire la position des atomes au sein du complexe cristallographique ;
3. la distance seuil entre les atomes de l'ADN et de la protéine qui discrimine les atomes de l'ADN appartenant au moule.

2.4.1 La constante de force

Les premiers tests ont permis de montrer qu'une constante de force d'au moins 50 kcal.mole⁻¹.Å⁻² est suffisante pour respecter parfaitement les contraintes imposées sur l'ADN cible (figure 2). Nous avons adopté une valeur de 100 kcal.mole⁻¹.Å⁻² pour l'ensemble de nos études, mais pour éviter de déformer l'ADN trop brusquement nous avons appliqué les contraintes en deux étapes : 5 kcal.mole⁻¹.Å⁻² pour la première minimisation, 100 kcal.mole⁻¹.Å⁻² pour la deuxième.

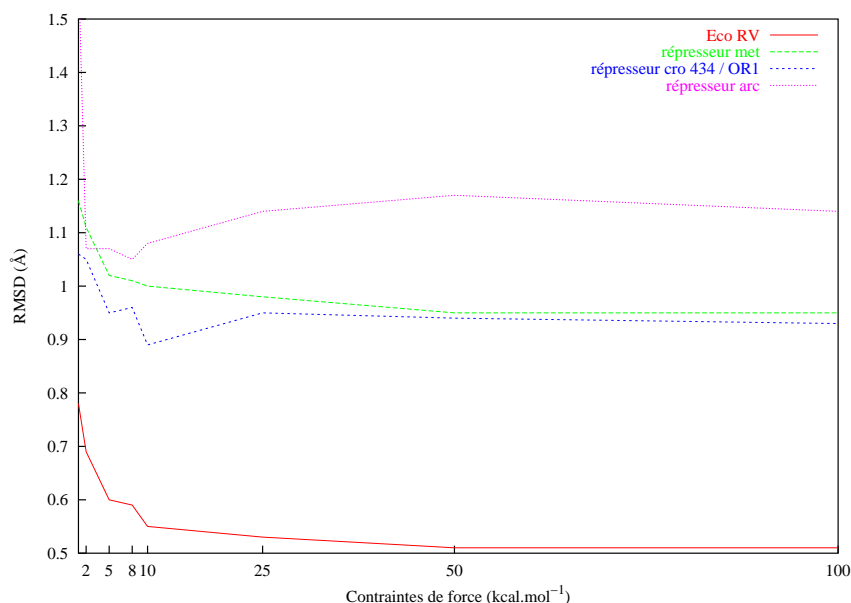


FIG. 2 – *RMSD en fonction des contraintes de force appliquées.*

2.4.2 La précision de positionnement des atomes

Il y a deux raisons pour ne pas vouloir reproduire exactement les positions occupées par les atomes au sein des complexes cristallographiques. Premièrement, compte-tenu de leur résolution moyenne (la majorité des complexes a une résolution supérieure à 2,5 Å), il y a une incertitude expérimentale de l'ordre de quelques dixièmes d'angstroems. Deuxièmement, compte-tenu de l'utilisation des coordonnées internes dans JUMNA et, par conséquent, de la fixation de toutes les longueurs de liaison et de la plupart des angles de valence, il n'y a pas suffisamment de variables pour reproduire exactement les structures expérimentales. Les résultats présentés figure 3 montrent cette limitation à travers l'augmentation de l'énergie dite «de pénalité», associée aux contraintes qui assurent la satisfaction des jonctions entre les nucléotides d'un même brin et au sein des cycles désoxyribose. Cette énergie commence à augmenter rapidement quand des tentatives de positionnement des atomes avec une précision supérieure à 0,4 Å sont initiées. Nous avons donc fixé cette valeur comme seuil pour l'ensemble de nos études.

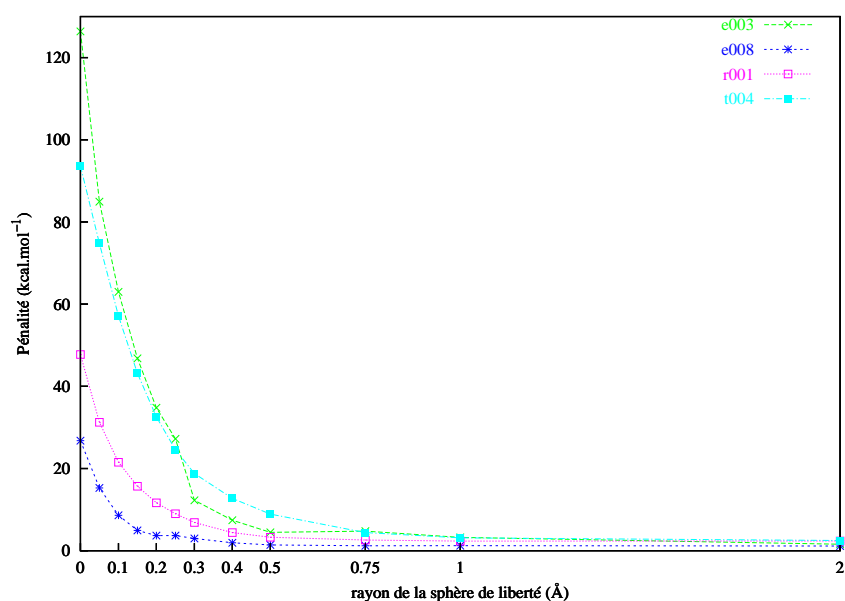


FIG. 3 – *Energie de pénalité en fonction des contraintes de force appliquées.*

2.4.3 La distance seuil définissant les atomes du moule

Des tests effectués sur un ensemble de complexes ont permis de montrer qu'une distance seuil de 4.5 Å entre les atomes de l'ADN et de la protéine génère des moules capables de reproduire la conformation de l'ADN au sein des complexes avec un RMSD d'environ 1 Å dans la grande majorité des cas (voir tableau 1, page 3). Au vu de la précision des calculs que nous effectuons, il n'est pas utile d'augmenter cette tolérance. De surcroît, quand la distance qui définit le moule augmente, on prend en compte les atomes de l'ADN qui ne sont plus en contact direct avec la protéine et donc sont plus

susceptibles d'être influencés par les effets d'agencement des molécules au sein de la maille cristallographique.

2.4.4 Structure des brins

En plus de l'application de l'ADN sur le moule comme décrit ci-dessus, nous avons introduit des contraintes pour éviter que les angles dièdres α (O3'-P-O5'-C5') et γ (O5'-C5'-C4'-C3') adoptent des valeurs non canoniques dans les ADN que nous étudions (c'est à dire autres que $\alpha = g^-$ et $\gamma = g^+$). Des conformations non canoniques de ces angles sont souvent observées dans les structures cristallographiques mais rarement en solution (bien qu'il reste des incertitudes sur la meilleure façon de les déterminer par spectroscopie RMN). [14, 15]

2.4.5 Validation

Nous pouvons redouter que la simple minimisation sous contraintes ne sera pas capable de déformer un fragment d'ADN sans être bloquée dans des conformations correspondant à des minima locaux. L'étude des complexes impliquant des protéines dimériques permet de tester ce comportement. En effet pour ces complexes, nous pouvons déformer l'ADN avec le monomère 1 puis avec le monomère 2, dans l'ordre inverse ou avec les deux monomères ensemble. Ainsi, nous pouvons produire un ADN déformé par trois chemins différents. Les tests effectués sur un ensemble de complexes de types différents montrent que les trois chemins mènent à des énergies de déformation totale qui diffèrent par moins d'une kcal.mole⁻¹.

3 Résultats

La méthodologie mise au point permet de reproduire efficacement la structure de l'ADN observée au sein des complexes protéine-ADN que nous avons sélectionnés. Nous sommes ainsi en mesure d'obtenir des informations sur l'énergie totale de déformation, sur sa distribution le long des segments de l'ADN et sur la contribution relative des différentes composantes énergétiques.

3.1 Déformation de l'ADN

La figure 4 illustre l'énergie de déformation des différents complexes en fonction de la courbure induite lors de la complexation. L'énergie totale est comprise entre 10 et 100 kcal.mole⁻¹, tandis que le changement de courbure varie de 75° à -20°. En général la formation du complexe avec une protéine augmente la courbure de l'ADN, typiquement vers le grand sillon, mais pour quelques protéines la courbure intrinsèque de l'ADN cible diminue (rc03, t017, t004).

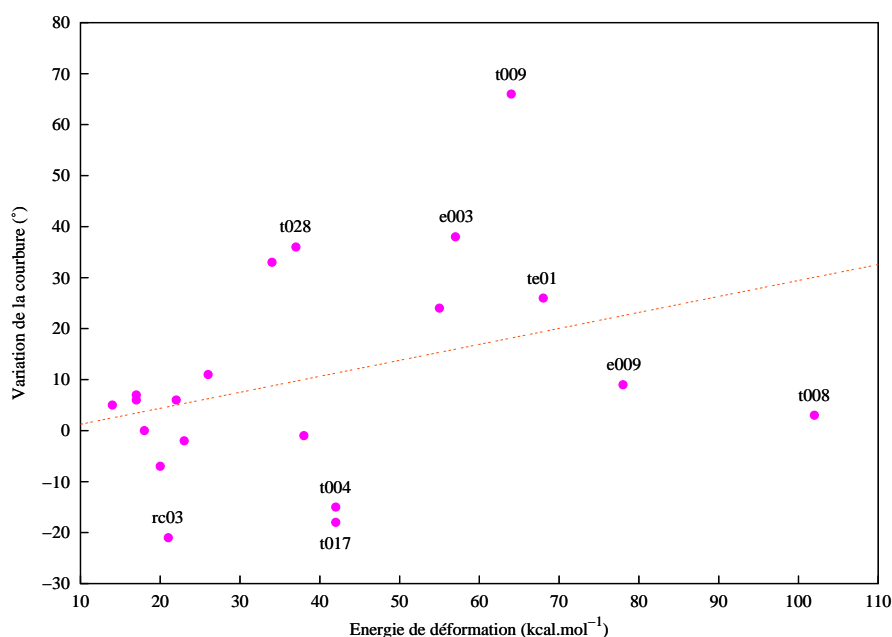


FIG. 4 – Variation de la courbure de l'ADN en fonction de son énergie de déformation.

La courbure est peu corrélée avec l'énergie de déformation. L'énergie nécessaire pour courber l'ADN est donc relativement faible par rapport à d'autres types de déformations (perturbation des interactions des bases et modifications locales de la conformation). Ainsi, certains fragments d'ADN se courbent beaucoup sous l'influence d'une protéine, tel que dans le complexe t028, bien que l'énergie de déformation soit faible, tandis que d'autres, comme te01, ont des fortes énergies de déformation mais sont moins courbés. Pour te01, une analyse détaillée du changement de conformation induit par la protéine

(tableau 2) permet de constater que l'agencement de plusieurs paires de bases est perturbé (fortes valeurs de gauchissement (*buckle*), roulis (*roll*), torsion hélicoïdale (*propeller*), torsion(*twist*)) et que plusieurs déformations importantes ont lieu au sein des brins phosphodiester (notamment la déformation de plusieurs sucres qui adoptent des plissement de très faible amplitude). La somme de ces modifications conduit à une perte d'environ 68 kcal.mole⁻¹ (dont 15 kcal.mole⁻¹ associés uniquement à la modification de l'interaction des bases), mais moins de 30° de courbure.

	Nucléotides	Δ Buckle	Δ Propeller	Δ Roll	Δ Twist	ΔA_1	ΔA_2
1)	A	-32	-2			-2	0
2)	G	14	13	-6	-4	-1	1
3)	T	20	-19	-6	-8	-4	-10
4)	T	2	4	2	5	1	6
5)	A	4	5	25	-17	-9	-1
6)	C	-2	12	5	-7	7	4
7)	A	3	5	21	-17	1	1
8)	A	14	13	-6	-4	-3	1
9)	C	3	14	-3	-4	-15	-1
10)	A	0	11	0	-2	-1	7
11)	T	4	3	0	-3	-5	-1
12)	A	6	-5	-1	-1	0	0
13)	C	2	-11	-2	0	0	-3
14)	T	-24	-27	8	0	1	4

TAB. 2 – Décomposition des perturbations observées pour *te01*. ΔA_n : amplitude des sucres sur le brin *n*. Toutes les valeurs sont exprimées en degrés.

3.2 Composantes de l'énergie de déformation

La figure 5 présente pour tous les complexes (numérotés de 1 à 20, voir tableau 1, page 3) la répartition des différentes composantes de l'énergie de déformation. Il est intéressant de constater que celle-ci est dominée par le terme Lennard-Jones (c'est à dire les interactions répulsion-dispersion) et par le terme torsion (modification des angles dièdres) pour l'ensemble des complexes étudiés. L'énergie électrostatique est toujours faible. L'énergie associée avec les angles de valence montre une tendance à diminuer faiblement lors de la complexation sauf dans le cas des complexes les plus déformés où elle signale des contraintes locales fortes sur les brins phosphodiester. L'énergie de déformation inter-brins est souvent supérieure à celle intra-brins.

Pour illustrer la répartition de l'énergie entre les différents résidus de l'ADN le tableau 3 donne les détails pour deux complexes, *te01* discuté ci-dessus, et *e003*, un complexe dimérique qui sera traité dans la section 3.3.

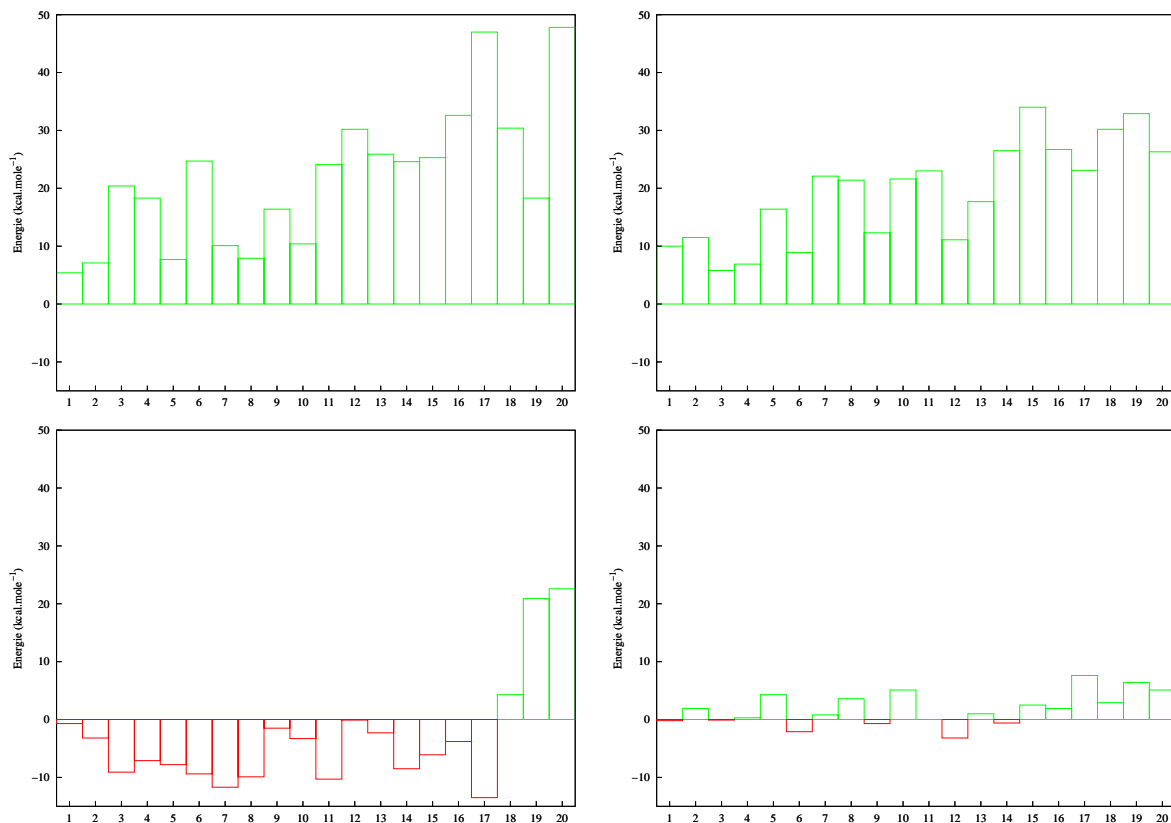


FIG. 5 – Composantes de l'énergie de déformation. De gauche à droite et de haut en bas: Lennard-Jones (Lj), angles dièdres (torsion, $Torg$), angles de valence (Ang), électrostatique ($Elec$).

te01											
	PP	PS	PB	SS	SB	BB	Lj	Elec	Ang	Torg	Tot
1-1	-2.4	11.5	-8	12.8	-8	.2	7.0	-1.5	-1.6	16.6	20.5
1-2	-3.9	7.2	-1.3	.1	4.1	10.8	13.1	4.0	.0	.0	17.1
2-2	-.1	11.4	.7	14.3	.4	3.6	10.3	.4	5.9	13.6	30.2
Tot	-6.4	30.1	-1.4	27.2	3.7	14.6	30.4	2.9	4.3	30.2	67.8
e003											
	PP	PS	PB	SS	SB	BB	Lj	Elec	Ang	Torg	Tot
1-1	-.6	9.2	.6	6.8	6.7	-6.1	2.7	2.2	-2.9	14.5	16.5
1-2	-6.1	15.7	-.1	3.4	8.3	2.0	27.0	-4.1	.0	.0	23.0
2-2	-.4	6.7	.8	9.2	5.4	-3.6	2.9	3.8	-.9	12.2	18.0
Tot	-7.1	31.6	1.3	19.4	20.4	-7.7	32.6	1.9	-3.8	26.7	57.5

TAB. 3 – Décomposition des variations de l'énergie de déformation. P : phosphate, S : sucre, B : base. La combinaison de deux lettres désigne l'interaction (PP : interaction phosphate-phosphate). $n-m$: interaction brin n - brin m .

L'énergie de déformation pour te01 est répartie équitablement entre chaque brin (20 et 30 kcal.mole⁻¹ respectivement) et sur l'interaction entre ces brins (17 kcal.mole⁻¹). Les termes qui sont les plus affectés sont les interactions des bases (15 kcal.mole⁻¹, voir discussion ci-dessus), les interactions phosphate-sucre (30 kcal.mole⁻¹) et sucre-sucre (27 kcal.mole⁻¹). Pour e003 les résultats sont globalement similaires mais nous constatons que par rapport à te01, l'énergie de déformation inter-brin devient plus importante. Les termes intra-brins et l'énergie de déformation base-base sont favorables tandis que le terme sucre-base devient fortement positif. Ces changements sont en accord avec le fait que le complexe te01 comprend de fortes déformations locales qui sont absentes pour e003 où l'ADN subit une déformation plus uniforme (accompagnée d'une augmentation plus nette de sa courbure).

3.3 Cas des protéines dimériques

Dans le cas de complexes dimériques, nous avons désormais la possibilité de voir si les deux monomères coopèrent ou se gênent pour déformer l'ADN cible. La nature de leurs interactions découle de l'étude de la déformation induite par chacun des monomères et de la comparaison entre la somme de ces énergies et l'énergie de déformation calculée pour le dimère (l'empreinte des protéines sur l'ADN est présentée tableau 5).

Le tableau 4 résume cette information. Nous constatons que dans la majorité des cas, l'énergie de déformation du dimère est très proche de la somme des énergies associées à chaque monomère. Néanmoins pour rc03, r010, t017, t031 et e003 nous observons des différences significatives. Pour les trois premiers de ces complexes la différence est positive et donc l'agencement relatif des deux monomères nécessite une déformation supplémentaire de l'ADN. Dans ce cas nous pouvons prédire que la complexation dépendra de la séquence en bases séparant les monomères, même si celle-ci n'est pas directement contactée par la protéine. Pour t031 et e003 la différence est négative et donc la complexation d'un monomère prépare l'ADN à recevoir le deuxième monomère. Afin d'analyser plus finement le rôle énergétique de chaque monomère sur la déformation de l'ADN, nous décomposons l'énergie de déformation par paire de nucléotides. Elle est calculée en faisant la somme de l'énergie interne d'une paire de nucléotides et de la moitié de l'énergie d'interaction de cette paire de nucléotides avec le reste de la molécule.

Ce calcul définit un profil énergétique de l'interaction, ou énergie *environnementale*. Il est à noter que la somme de toutes ces composantes est égale à l'énergie de déformation totale. La figure 6 en donne une représentation pour rc03 et e003. En ce qui concerne rc03, chaque monomère ne déforme que la moitié de l'ADN cible, à gauche pour le premier monomère (courbe rouge) et à droite pour le deuxième (courbe bleue). La courbe pour le dimère (noire) suit les courbes des deux monomères de très près sauf au milieu de

Nom	ΔE_{def} dimère	ΔE_{def} monomère I	ΔE_{def} monomère II	Différence
e008	14	6	8	0
r001	18	9	8	1
t016	20	10	12	-2
rc03	21	6	9	6
r009	22	13	12	-3
r010	34	8	21	5
t028	37	15	23	-1
t031	38	25	22	-9
t017	42	23	14	5
e003	57	42	40	-25

TAB. 4 – *Energie de déformation associée aux monomères et comparaison avec celle du dimère (ΔE_{def} : énergie de déformation de l'ADN en kcal.mole^{-1}).*

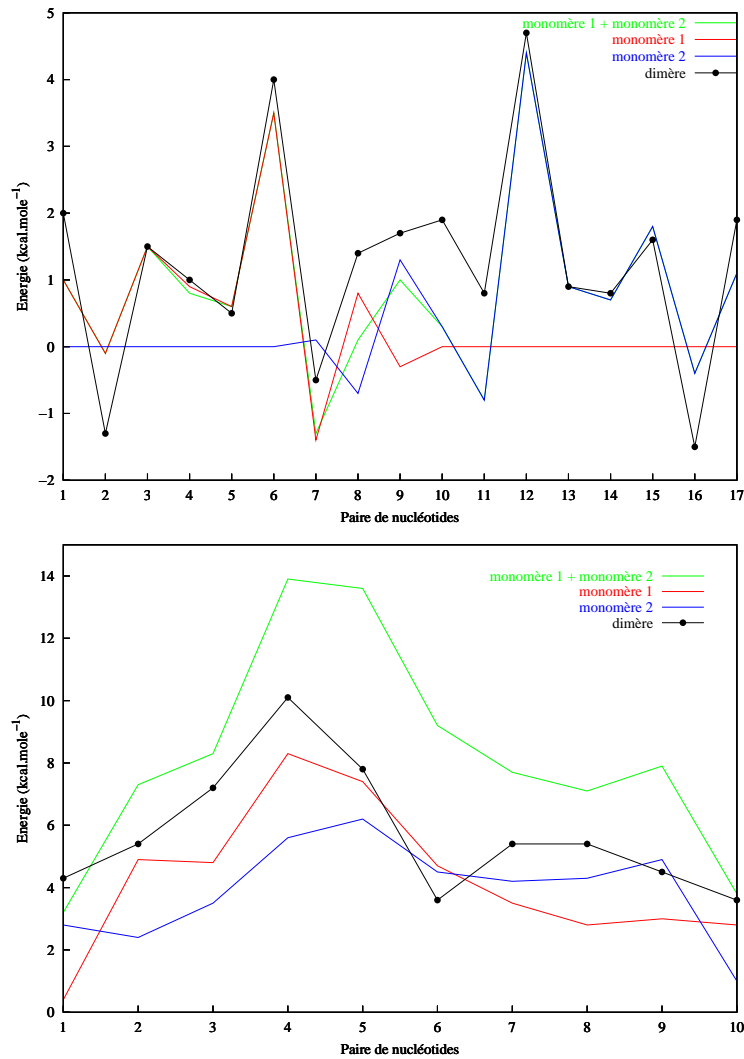


FIG. 6 – *Energie environnementale en fonction de la séquence. En haut : rc03, en bas : e003.*

sa séquence cible où elle est au-dessus de la somme des deux dimères (verte) indiquant qu'une déformation supplémentaire accompagne la formation du complexe dimérique. Pour e003 la situation est inversée. Chaque monomère agit sur pratiquement tout l'ADN cible. Par conséquent un monomère prépare l'arrivée du deuxième. C'est pourquoi la courbe correspondant au complexe dimérique (noire) est presque toujours en dessous de la courbe correspondant à la somme de l'énergie des monomères (verte).

3.4 Mutation

A plus long terme notre but est de pouvoir prédire l'impact de mutations (de paires de bases ou d'acides aminés) sur la stabilité d'un complexe protéine-ADN. Ceci nécessite le calcul des changements d'énergie libre associés à de telles mutations. Mais si nous choisissons un cas où la déformation de l'ADN semble jouer un rôle prédominant, nous pouvons espérer que les variations de la simple énergie de déformation seront suffisantes pour prédire les conséquences de la mutation d'une paire de bases. La TBP est un bon cas pour ce test puisque plusieurs études expérimentales et théoriques montrent que la conformation et la flexibilité de l'ADN cible influencent fortement la stabilité et la durée de vie du complexe, ainsi que sa capacité à initier la transcription [16, 17, 18]. Nous avons donc utilisé CONTACT pour créer un moule pour la TBP (t009) après avoir introduit une mutation T→A au sein de la TATA-box. Cette mutation est connue pour diminuer *in vitro* la durée de vie du complexe d'un facteur 100 et le taux de transcription d'un facteur 10.

Séquence initiale: GCTATAAAAGGGCA $\Delta E_{def} = 64 \text{ kcal.mole.}^{-1}$
CGATATTTTCCCGT
Séquence finale: GCTAAAAAAGGGCA $\Delta E_{def} = 84 \text{ kcal.mole.}^{-1}$
CGATTTTTTCCCGT

Après la déformation de l'ADN cible muté, nous constatons une augmentation de l'énergie de déformation de $20 \text{ kcal.mole}^{-1}$. Ce résultat très prometteur est en accord avec les observations expérimentales et nous encourage à croire que notre approche sera applicable à la prédiction des effets de telles mutations dans d'autres complexes.

4 Conclusion

Nous avons élaboré une méthodologie qui permet de calculer l'énergie de déformation de l'ADN lors de la formation d'un complexe protéique à partir de la conformation expérimentale du complexe par l'emploi d'un «moule moléculaire». Cette méthodologie originale permet d'apporter de nombreuses informations qui étaient jusqu'alors inaccessibles. Ainsi, nous avons pu observer une gamme d'énergie de déformation allant de 10 à 100 kcal.mole⁻¹, constater qu'il n'existe qu'un faible rapport entre ces énergies et la courbure de l'ADN et enfin identifier les composantes les plus importantes de la déformation (termes Lennard-Jones et angles dièdres). Cette approche permet aussi d'analyser en détail l'importance de la déformation de l'ADN le long d'un oligomère. Dans le cas des protéines dimériques cette possibilité permet de déterminer si les deux monomères coopèrent ou se gênent dans la déformation de l'ADN cible. Des améliorations de notre méthodologie incluent un meilleur traitement des effets du solvant et des contre ions. Ainsi nous travaillons sur une approche de type «Generalized Born» [8] qui pourrait être employée lors de la minimisation d'énergie et qui, en contraste avec les calculs Poisson-Boltzmann numériques, pourrait être appliquée à des complexes de grandes dimensions.

Notre but est d'intégrer cette approche dans un calcul de l'ensemble des composants constituant l'énergie libre de complexation protéine-acide nucléique afin de pouvoir prédire non seulement la stabilité des complexes existant mais aussi l'impact sur ces complexes des mutations au sein de l'un ou l'autre des partenaires. Il est à noter que nous avons déjà obtenu un résultat encourageant dans cette direction avec la TBP en étudiant l'effet d'une mutation d'une base sur l'énergie de déformation de l'ADN. En effet, il est raisonnable de supposer que dans ce cas ce composant peut dominer l'énergie libre de complexation.

Références

- [1] Structural features of protein-acid nucleic recognition sites K. Nadassy, S.J. Wodak, J. Janin. *Biochemistry* 1999, 38, 1999-2017.
- [2] DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. W.K. Olson, A.A. Gorin, X.-J. Lu, L.M. Hock, V.B. Zhurkin. *PNAS*, 1998, 95, 11163-11168.
- [3] Free analysis of protein-DNA binding : the ECO RI Endonuclease-DNA complex. B. Jayaram, K.J. McConnel, S.B. Dixit, D.L. Beveridge. *J. Comput. Phys.* 1999, 151, 333-357
- [4] Free energy component analysis of 40 protein-DNA complexes : a consensus view of the nature of binding at the molecular level. B. Jayaram, K. McConnel, S.B. Dixit, D.L. Beveridge (communication personnelle).
- [5] Junctions and bends in nucleic acids : a new theoretical modeling approach. R. Lavery. *DNA Bending and Curvature. Structure and Expression* (W.K. Olson, M.H. Sarma, R.H., M. Sundaralingam, eds), 1988, vol 3, 191-211.
- [6] DNA flexibility under control : the JUMNA algorithm and its application to BZ junctions. R. Lavery. *Unusual DNA structures* (R.D. Wells, Harvey S.C, eds), 1988, 189-206, Springer-Verlag, Berlin, Heidelberg, New York.
- [7] JUMNA (Junction Minimisation of Nucleic Acids). R. Lavery, K. Zakrzewska, H. Sklenar. *Comput. Phys. Commun.* 1995, 91, 135-158.
- [8] Molecular dynamics simulation of nucleic acids with a Generalized Born solvation model. V. Tsui and D.A. Case. *J. Am. Chem. Soc.* 2000, 122, 2489-2498.
- [9] Molecular dynamics anneals large-scale deformations of model macromolecules : stretching the DNA double helix to form an intercalation site. M. Prabhakaran, S.C. Harvey. *J. Phys. Chem* 1985, 89, 5767-5769.
- [10] Modeling nucleic acids : fine structure, flexibility and conformational transitions. R. Lavery. *Adv. Comput. Chem.* 1994, 1, 69-145.
- [11] The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. R. Lavery, H. Sklenar. *J. Biomol. Struct. Dyn.* 1988, 63-91.
- [12] Defining the structure of irregular nucleic acids : conventions and principles. R. Lavery, H. Sklenar. *J. Biomol. Struct. Dyn.* 1989, 6, 655-667.
- [13] A quantitative description of the conformation of biomolecules. R. Lavery, H. Sklenar. *Proceedings of the 6th conversation in biomolecular stereodynamics*, eds R.H. Sarma, Adenine press, New York, 1990, 215-235.
- [14] ³¹P NMR analysis of the DNA conformation induced by protein binding SRY/DNA complexes. C. Castagné, E.C. Murphy, A.M. Gronenborn, M. Delepierre. *Eur. J. Biochem.* 2000, 267, 1223-1229.
- [15] Chemical shift and structure relationship in nucleic acids : correlation of backbone torsion angles γ and α with ¹³C chemical shifts. X.-P. Xu, W.-L.A.K. Chiu, S.C.F. Au-Yeung. *J. Am. Chem. Soc.* 1998, 120, 4230-4231.
- [16] Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. C.R. Wobbe, K. Struhl. *Mol. Cell. Biol.* 1990, 10, 3859-3867.
- [17] DNA bending is an important component of site-specific recognition by the TATA binding protein. D.B. Starr, B.C. Hoopes, D.K. Hawley. *J. Mol. Biol.* 1995, 250, 434-446.
- [18] Sequence-dependent dynamics of TATA-box binding sites. D. Flatters, R. Lavery. *Biophys. J.* 1998, 75, 372-381.

e008	T • A • G • A • C • G • T • C • T • • A • T • C • T • G • C • A • G • A
t011	T • A • A • T • A • A • G • G • A • T • A • A • C • G • T • C • C • G • • A • T • T • A • T • T • C • C • T • A • T • T • G • C • A • G • G • C
t039	G • C • G • T • G • G • G • C • G • T • • C • G • C • A • C • C • C • G • C • A
r001	A • G • T • A • C • A • A • A • C • T • T • T • C • T • T • G • T • A • T • • T • C • A • T • G • T • T • T • G • A • A • A • G • A • A • C • A • T • A
t016	T • C • A • A • C • A • G • C • T • G • T • T • G • A • • A • G • T • T • G • T • C • G • A • C • A • A • C • T
rc03	C • A • G • G • T • C • A • C • A • G • T • G • A • C • C • T • G • • G • T • C • C • A • G • T • G • T • C • A • C • T • G • G • A • C
r009	G • T • A • C • T • A • G • T • T • A • A • C • T • A • G • T • A • C • • C • A • T • G • A • T • C • A • A • T • T • G • A • T • C • A • T • G
r018	C • G • T • C • A • C • G • G • T • T • G • A • C • • G • C • A • G • T • G • C • C • A • A • C • T • G
t019	G • T • A • T • G • C • A • A • A • T • A • A • G • G • • C • A • T • A • C • G • T • T • T • A • T • T • C • C
r010	A • T • A • C • C • A • C • T • G • G • C • G • G • T • G • A • T • A • T • • T • A • T • G • G • T • G • A • C • C • G • C • C • A • C • T • A • T • A
t028	C • A • T • G • T • A • A • T • T • A • T • T • A • C • A • T • C • A • • G • T • A • C • A • T • T • A • A • A • T • A • A • T • G • T • A • G • T
t031	T • A • A • T • T • G • A • A • T • T • • A • T • T • A • A • C • T • T • A • A
t017	T • C • G • G • C • A • A • T • T • G • C • C • G • A • • A • G • C • C • G • T • T • A • A • C • G • G • C • T
t004	T • T • T • G • C • C • A • T • G • T • A • A • T • T • A • C • C • T • A • A • • A • A • A • C • G • G • T • A • C • A • T • T • A • A • T • G • G • A • T • T
t035	C • G • C • A • C • A • C • C • C • A • C • A • C • A • C • C • A • G • • G • C • G • T • G • T • G • G • G • T • G • T • G • T • G • G • T • C
e003	G • G • G • A • T • A • T • C • C • C • • C • C • C • T • A • T • A • G • G • G
t009	G • C • T • A • T • A • A • A • A • G • G • G • C • A • • C • G • A • T • A • T • T • T • C • C • C • G • T
te01	A • G • T • T • A • C • A • A • C • A • T • A • C • T • • T • C • A • A • T • G • T • T • G • T • A • T • G • A
e009	T • T • T • T • T • G • A • T • A • A • G • A • • A • A • A • A • A • C • T • A • T • T • C • T
t008	T • T • C • G • T • C • T • T • G • G • G • T • G • G • T • C • C • A • C • G • • A • A • G • C • A • G • A • A • C • C • C • A • C • C • A • G • G • T • G • C

TAB. 5 – Zones de contact entre l'ADN et les protéines. • : squelette phosphodiester, A, T, C, G : bases. Si la protéine est dimérique, • et A, T, C, G pour le deuxième monomère. Si les deux monomères sont en contact au même endroit, •, A, T, C, G