

# Modéliser l'ADN avec Graphite-MicroMégas

Modeling DNA with Graphite-MicroMégas

Samuel Hornus and Damien Larivière

**English Abstract**—Biologists need simple spatial modeling tools to help in understanding the role of objects' relative position in the functioning of the cell. In this context, we report on a tool for DNA modeling. An english version of this text is available on the authors' webpages.



## 1 LA MODÉLISATION 3D : UN MOTEUR D'HYPOTHÈSES ET DE QUESTIONS NOUVELLES

L'arrivée en 2005 des instruments de séquençage dits de nouvelle génération a provoqué un déluge de données biologiques. Les sciences de la vie se sont ainsi retrouvées en chef de file du développement d'infrastructures pour le stockage, l'analyse, l'interprétation et la visualisation de petabytes de données [5].

Les avancées récentes en biologie montrent que c'est moins la structure des objets qui importe dans le fonctionnement d'une cellule que la place que ces objets occupent dans la cellule [1]. La tendance des protéines à former des assemblages est un élément majeur de l'organisation cellulaire interne. Si la *Protein Data Bank* contient près de 60 000 structures, les assemblages de protéines sont souvent réfractaires à toute détermination structurale.

Il est de fait indispensable de développer des outils de modélisation 3D permettant de reconstruire virtuellement les assemblages de protéines et d'ADN. Il ne s'agit pas uniquement d'illustrer en 3D la connaissance biologique mais d'aider à la formulation d'hypothèses expérimentalement testables sur l'organisation spatiale des composants cellulaires à mesure que le processus de modélisation progresse: modélisation et visualisation 3D sont des techniques de raisonnement à partir de données quantitatives [6].

## 2 PRÉSENTATION DE MICROMÉGAS

MicroMégas est une contribution à cet effort visant à la production d'outils de modélisation 3D. Nous présentons ici la partie de MicroMégas dédiée à la modélisation de l'ADN. Techniquement, MicroMégas est un plugin du logiciel Graphite, qui est une plateforme de recherche pour la géométrie numérique

développée par les membres de l'équipe ALICE de l'INRIA.<sup>1</sup> Une page web explique l'installation et l'utilisation de MicroMégas.<sup>2</sup>

### 2.1 Modélisation de l'ADN

La modélisation débute généralement par la création ou l'import d'une surface de protéine qui servira de guide à la modélisation du brin d'ADN avec lequel elle est en interaction. S'ensuit une phase classique de modélisation d'une courbe dans l'espace (courbe de Bézier quadratique, cubique et courbe d'interpolation).

À n'importe quel moment, la courbe modélisée peut être visualisée de plusieurs façons différentes : comme une simple ligne, un tube ou de l'ADN à l'échelle atomique. Cette visualisation peut être partiellement transparente pour pouvoir continuer la modélisation tout en voyant le brin d'ADN se déformer en temps réel.

Quand la forme globale de l'ADN est fixée, il est possible d'ajuster finement la position de chaque paire de bases, dont la rotation autour de la courbe peut-être spécifiée. Si la position angulaire d'une paire de bases n'est pas spécifiée (ce qui est le cas pour la vaste majorité des paires), alors celle-ci est calculée automatiquement de façon à se rapprocher le plus possible de la forme « en hélice » de l'ADN. L'hélice du brin d'ADN peut être ainsi « dévissée » localement pour prendre en compte les effets d'une déformation plus ou moins localisée de l'ADN par une protéine (Figure 5).

Enfin, il est possible d'exporter le brin d'ADN ainsi modélisé au format PDB, pour vérifier l'intégrité physique du brin, ajuster la position des atomes par simulation, ou à des fins d'illustration scientifiques par exemple.

À notre connaissance, MicroMégas est le premier outil dédié à la modélisation simple et intuitive de l'ADN. Les outils de graphisme 3D professionnels tels que Autodesk Maya permettent aussi de modéliser

• Samuel Hornus: INRIA  
E-mail: samuel.hornus@inria.fr.  
• Damien Larivière: Fondation Fourmentin-Guilbert  
E-mail: damien.lariviere@fourmentinguilbert.org.

1. [alice.loria.fr/index.php/software.html](http://alice.loria.fr/index.php/software.html)  
2. [www.loria.fr/~shornus/FFG/micromegas.html](http://www.loria.fr/~shornus/FFG/micromegas.html)

des assemblages formés de protéines et d'ADN mais plusieurs années sont nécessaires pour une prise en main efficace.

### 3 EXEMPLE D'APPLICATIONS: LE PROCESSUS DE RÉPARATION DE L'ADN BACTÉRIEN

La réplication du chromosome dans une bactérie n'est pas sans erreur : Le taux d'erreur de l'ADN polymérase atteint 1 pour  $10^6$  de paires de bases synthétisées [4]. Le processus de réparation est coordonné par les protéines MutS, MutL and MutH. La structure du complexe qu'elles forment, avec ou sans ADN, est inconnue, mais des modèles structuraux ont été créés à partir de la structure de chacune des protéines [8].

Graphite-MicroMégas permet d'intégrer facilement un brin d'ADN dans ce type de modèle (Figure 3) en dessinant sa trajectoire reliant plusieurs sites de liaison, hypothétiques ou avérés, à la surface des protéines.

Cela permet de repérer de nouveaux sites de liaison ADN-protéine, et de concevoir des expériences de *crosslink* ou de FRET dans le but de corroborer ou invalider des arrangements de protéines lorsque la forme très globale de l'ADN (une boucle ou une portion droite par exemple) est déjà connue expérimentalement.

#### 3.1 Autres exemples

Graphite-MicroMégas permet l'étude de l'assemblage de nucléosomes [2] (Figure 4) et d'explorer le compactage de longs brins d'ADN [3] (Figure 1).

## 4 IMPLÉMENTATION

Nous décrivons ici quelques détails techniques sur l'implémentation de l'outil de modélisation de l'ADN.

Nous avons naturellement choisi de modéliser un brin d'ADN par une courbe lisse  $C$  : en chaque point de la courbe, le vecteur tangent doit être défini, ainsi que la longueur d'arc à partir d'une extrémité de  $C$ .

Nous utilisons deux échantillons de  $C$  : un échantillon *adaptatif* qui consiste en une série de points formant une approximation linéaire de la courbe telle que deux segments consécutifs fassent un angle en dessous d'un seuil fixé. Cet échantillon adaptatif sert à l'affichage d'une courbe lisse visuellement, tout en minimisant le nombre de segments.

Nous utilisons un autre échantillon *uniforme*  $S_u$  de  $C$ , dans lequel toutes les parties de  $C$  entre deux points consécutifs ont la même longueur d'arc. Nous choisissons un espacement proche de  $3,4 \text{ \AA}$  pour que les points de l'échantillon  $S_u$  puissent servir de point d'ancrage aux paires de bases de l'ADN.

#### 4.1 Un repère mobile sans torsion

Après avoir obtenu un échantillon uniforme  $S_u$  de  $C$  ainsi que le vecteur tangent à chaque point de  $S_u$ , nous devons maintenant orienter chaque paire

de bases correctement le long de la courbe, et donc calculer un repère orthonormal en chaque point de  $S_u$ . Nous ne pouvons pas utiliser de méthode classique pour ce faire (comme le repère de Frenet) car elles produisent des discontinuités. De plus, nous souhaitons un repère mobile le long de  $C$  mais aussi rigide que possible (sans torsion), qui puisse servir de base pour l'orientation de l'ADN.

Le travail de Wang *et al.* montre comment obtenir rapidement une bonne approximation d'un tel repère mobile [7]. Nous obtenons ainsi une suite de triplets  $\{f_i = (o_i, t_i, n_i), i \in [1, N]\}$  où  $t_i$  est la tangente et  $n_i$  la normale à la courbe  $C$  au point  $o_i$  de  $S_u$  (Figure 2).

#### 4.2 Un modèle 3D du brin d'ADN

MicroMégas ne permet pas encore de spécifier la séquence précise de nucléotides et utilise implicitement la séquence AAA... Le dessin du brin d'ADN complet s'effectue en instanciant un modèle atomique 3D de la paire de bases AT dans chaque repère  $f_i$  de  $S_u$ . Pour reproduire la forme en hélice, le modèle est tourné de  $\frac{2i\pi}{10}$  dans le repère  $f_i$ . Le coût mémoire est ainsi minimal : on peut ainsi modéliser un brin comportant plusieurs millions de paires de bases et envisager la visualisation d'un chromosome entier.

#### 4.3 Torsion du brin d'ADN

Pour mieux ajuster le brin d'ADN aux protéines avec lesquelles il interagit, on peut spécifier l'angle  $a_i$ , autour de la tangente, de n'importe quelle paire de bases. Dans le repère  $f_i$ , le modèle atomique sera tourné de  $\frac{2i\pi}{10} + a_i$  radians. Les valeurs  $a_i$  non définies par l'utilisateur sont simplement calculées par interpolation des valeurs définies (Figure 5).

## REFERENCES

- [1] A. DANCHIN : Bacteria as computers making computers. *FEMS Microbiology Reviews*, 33(1):3–26, jan. 2009.
- [2] C. A. DAVEY, D. F. SARGENT, K. LUGER, A. W. MAEDER, et T. J. RICHMOND : Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *Journal of Molecular Biology*, 319(5):1097–1113, juin 2002.
- [3] I. JUNIER, O. MARTIN et F. KÉPÈS : Spatial and topological organization of DNA chains induced by gene co-localization. *PLoS Computational Biology*, 6(2):e1000678, fév. 2010.
- [4] T. A. KUNKEL : DNA replication fidelity. *Journal of Biological Chemistry*, 279(17):16895–16898, 2004.
- [5] C. SOUTHAN et G. CAMERON : *The Fourth Paradigm: Data-Intensive Scientific Discovery*, chap. Beyond the Tsunami: Developing the Infrastructure to Deal with Life Sciences Data. Microsoft Research, 2009.
- [6] T. WALTER, D. W. SHATTUCK, R. BALDOCK, M. E. BASTIN, A. E. CARPENTER, S. DUCE, J. ELLENBERG, A. FRASER, N. HAMILTON, S. PIEPER, M. A. RAGAN, J. E. SCHNEIDER, P. TOMANCAK et J.-K. HÉRICHÉ : Visualization of image data from cells to organisms. *Nature Methods*, 7(3):26–41, 2010.
- [7] W. WANG, B. JÜTTLER, D. ZHENG et Y. LIU : Computation of rotation minimizing frames. *ACM Transactions on Graphics*, 27(1):2:1–18, mars 2008.
- [8] I. WINKLER, A. D. MARX, D. LARIVIÈRE, L. MANELYTE, L. GIRON-MONZON, R. J. HEINZE, M. CRISTOVAO, A. REUMER, U. CURTH, T. K. SIXMA et P. FRIEDHOFF : Chemical trapping of the dynamic MutS-MutL complex formed in DNA mismatch repair in *Escherichia coli*. *Journal of Biological Chemistry*, 286(19):17326–17337, mai 2011.

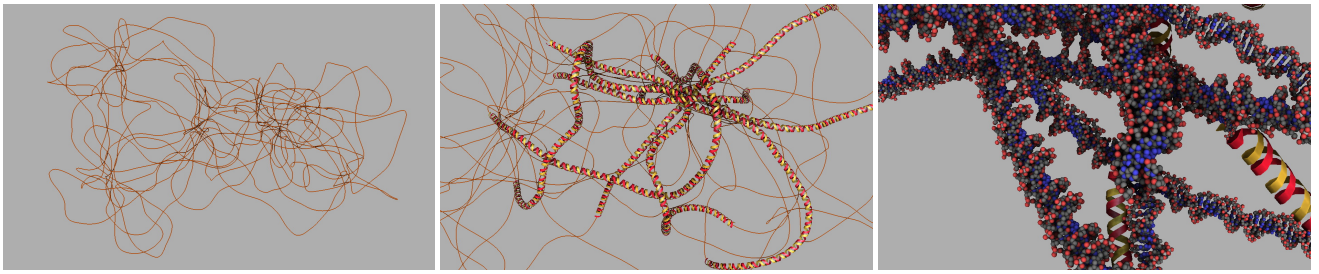


Fig. 1. Brin d'ADN de 28 000 paires de bases visualisé avec différents niveaux de détail. Pour la simulation de la compaction du brin, voir [3].

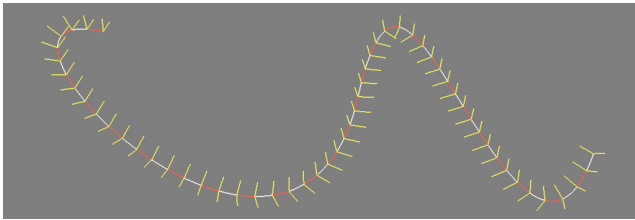


Fig. 2. Une suite de repères sans torsion le long d'une courbe de Bézier cubique. La normale et la binormale ont en jaune.

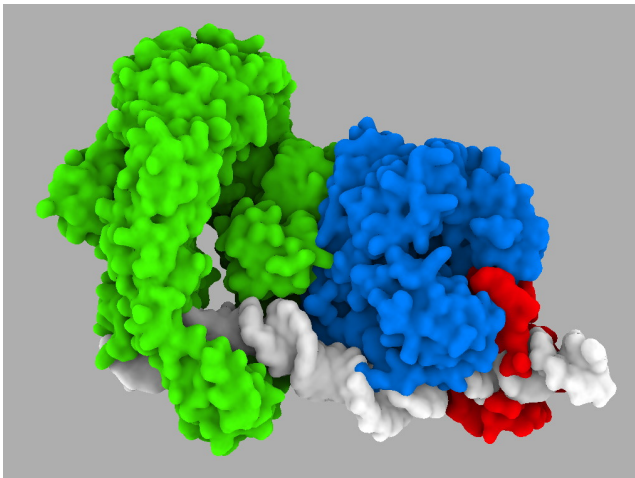


Fig. 3. Arrangement modélisé d'un brin d'ADN en interaction avec des protéines de réparation [8].

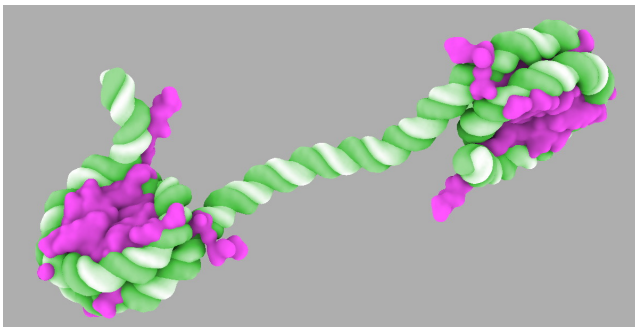


Fig. 4. Un modèle structural de l'enroulement de l'ADN autour de deux protéines nucléosomales [2].

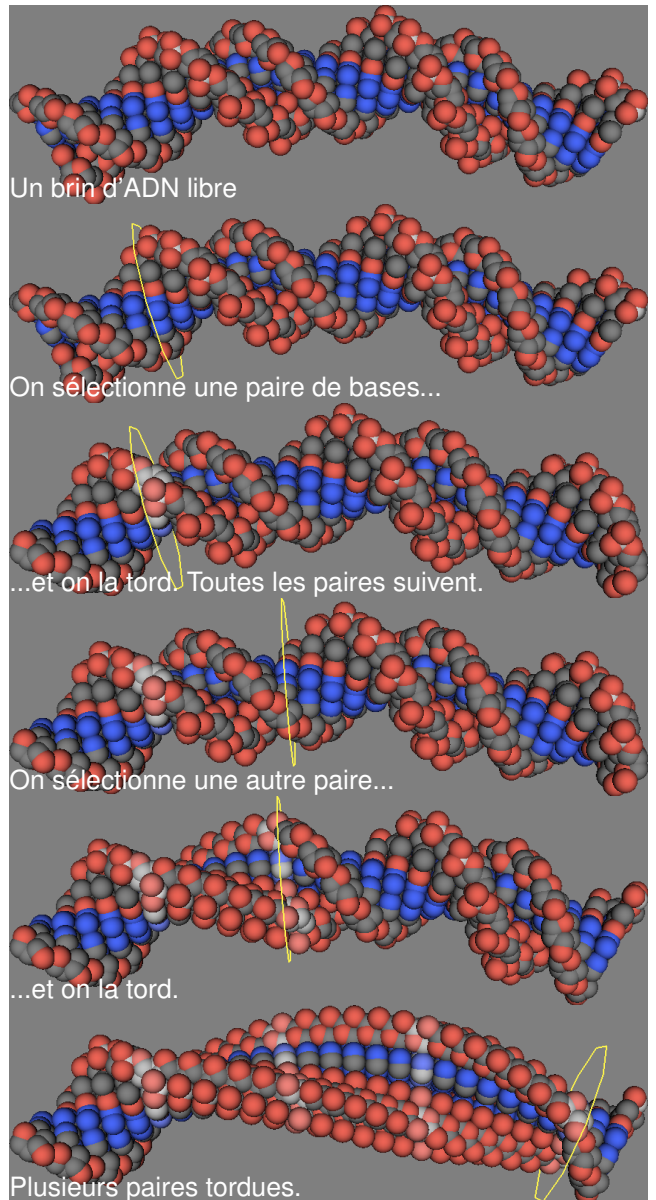


Fig. 5. Torsion de l'ADN.