



HAL
open science

Confrontation entre modèles et données réelles pour la structure spatiale des molécules d'ADN par le biais de la réalité augmentée

Mouna Essabbah, Rachid Gherbi, Malik Mallem

► To cite this version:

Mouna Essabbah, Rachid Gherbi, Malik Mallem. Confrontation entre modèles et données réelles pour la structure spatiale des molécules d'ADN par le biais de la réalité augmentée. 3e Workshop Applications Médicales de l'Informatique: Nouvelles Approches (AMINA 2006), Nov 2006, Monastir, Tunisie. pp.n.a. hal-00633202

HAL Id: hal-00633202

<https://hal.science/hal-00633202v1>

Submitted on 17 Oct 2011

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.

CONFRONTATION ENTRE MODÈLES ET DONNÉES RÉELLES POUR LA STRUCTURE SPATIALE DES MOLÉCULES D'ADN PAR LE BIAIS DE LA RÉALITÉ AUGMENTÉE

ESSABBAH Mouna ¹, GHERBI Rachid^{1,2}, MALLEM Malik¹

IBISC-CNRS, Université d'Evry Val d'Essonne ¹
40 Rue du Pelvoux, CE1455 Courcouronnes
91020 Evry Cédex, France
{ Mouna.Essabbah@ibisc.fr, Rachid.Gherbi@ibisc.fr, Malik.Mallem@ibisc.fr }

LIMSI-CNRS²
BP 133, Bât. 508, Université Paris-Sud
91403 Orsay cedex, France
gherbi@limsi.fr

RESUME

Si les séquences génomiques sont d'abord connues sous leur forme linéaire, elles ont aussi une structure tridimensionnelle potentiellement utilisable pour l'analyse des génomes. Cette représentation de la structure 3D apporte un point de vue nouveau pour l'analyse des séquences. Les propriétés géométriques de l'ADN sont très peu étudiées, en particulier pour de grands segments d'ADN. C'est justement sur ce dernier verrou que la synergie entre informaticiens et biologistes peut en exploitant la pluridisciplinarité être extrêmement fructueuse. Un logiciel complet, ADN-Viewer, a été développé pour aborder la modélisation 3D de l'ADN. Il est basé sur un modèle de conformation local. Notre objectif est de confronter les résultats prédits par ce logiciel avec des données (images) réelles issues d'expérimentations biologiques.

Il s'agit donc d'essayer d'affiner le modèle 3D par appariement d'images réelles et prédites. L'analyse des molécules en microscopie de force atomique (AFM) nous permettra de faire une analyse structurale de différentes régions intéressantes du génome. Ces images seront comparées, appariées et recalées avec les images prédites avec des techniques d'analyse et de traitement d'images. Il serait pertinent de quantifier l'erreur de prédiction et de corriger, par retour vers le modèle, les règles de conformation tri-dimensionnelle.

MOTS CLES

Structure spatiale de l'ADN, Réalité augmentée, Images microscopiques, Extraction de trajectoire, Mise en correspondance, Analyse de forme.

1. Introduction

La modélisation 3D de l'ADN est de plus en plus abordée en recherche biologique et bioinformatique. Les chercheurs s'intéressent à l'étude du comportement spatial des séquences génomiques par la visualisation et l'analyse de ces données tridimensionnelles. Cette approche permet d'avoir une vision globale de la structure spatiale de l'ADN ainsi que de percevoir les possibilités d'interactions avec d'autres molécules (protéines). Le logiciel ADN-Viewer a justement été développé pour exploiter la puissance de ces données (Fig.1.a) [3,5]. Pour optimiser le modèle 3D nous le confrontons à des données réelles* (Fig.1.b) (images microscopiques).

*Merci au Pr. ZOZIME, de l'Université d'Evry-Val-d'Essonne, Laboratoire Multicouches Nanométriques, et son équipe pour nous avoir fourni les images AFM.

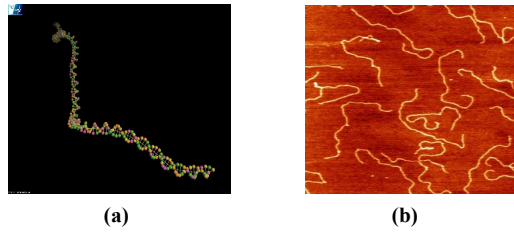


Fig.1: (a) Modélisation moléculaire 3D d'un gène, donnée par ADN-Viewer (b) Image à microscopie AFM d'un échantillon d'ADN.

2. Une approche guidée par le modèle

Sachant que dans un échantillon d'ADN imagé il existe différentes trajectoires possibles. On pourrait imaginer une base de donnée regroupant toutes ces trajectoires comme les formes possibles que peut prendre l'ADN. A ce moment là, on recherchera le modèle 2D prédit dans cet ensemble selon un certain coefficient de ressemblance [1]. Initialement, nous disposons d'un modèle 3D prédit [3] et de son image AFM correspondante. Ces deux modèles sont donc représentés dans deux espaces différents. Rappelons que notre but est de comparer les deux modèles (réel et prédit). Ainsi nous projetons le modèle prédit 3D selon un certain angle et sur un certain plan. Cette approche est dite guidée par le modèle, car on fait l'hypothèse que le modèle 3D est la référence qu'il faut comparer avec les trajectoires réelles. Elle se présente sous forme de trois processus. Le premier processus est l'extraction de la trajectoire réelle de l'ADN à partir des images microscopiques. Le deuxième est l'extraction de la trajectoire prédite à partir du modèle 3D de notre séquence génomique. La troisième étapes repose sur la comparaison des deux trajectoires précédemment extraites.

Extraction de la trajectoire 2D réelle à partir de l'image

La première étape consiste à traiter les images réelles de sorte à en extraire les différentes formes possibles que peut adopter l'ADN. Une sélection est ensuite faite pour ne garder que les trajectoires complètes et indépendantes des autres (Fig.2.b). Sachant que l'ADN est perçue à travers un bruit assez important, nous procédons à un filtrage Gaussien afin de lisser l'image (Fig.2.c). Nous séparons ensuite le fond de l'image de la forme par une segmentation avant de supprimer tous les points isolés (Fig.2.e). Une squelettisation nous permettra d'obtenir pour chaque forme un seul brin continu sur un fond unifié (Fig.2.f). Le squelette représente la trajectoire de l'ADN. Il décrit ainsi sa forme et sa conformation dans le plan. Désormais cette donnée constituera la base de notre comparaison avec le modèle prédit.

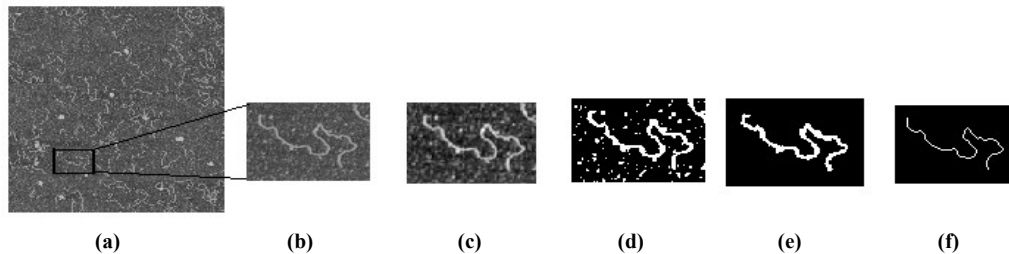


Fig.2: Étapes de l'extraction des trajectoires d'ADN à partir d'une image AFM

Projection du modèle 3D en 2D

Le logiciel ADN-Viewer permet, à partir d'une séquence textuelle d'ADN, de modéliser en 3D la molécule par des méthodes de prédiction [3]. Il sauvegarde les coordonnées 3D de chaque nucléotide de la séquence dans un fichier de coordonnées tridimensionnelles [5]. Nous pouvons ainsi restituer la forme spatiale de l'ADN (Fig.3.a) pour la projeter, selon un certain plan de projection, et obtenir le modèle 2D correspondant (Fig.3.b).

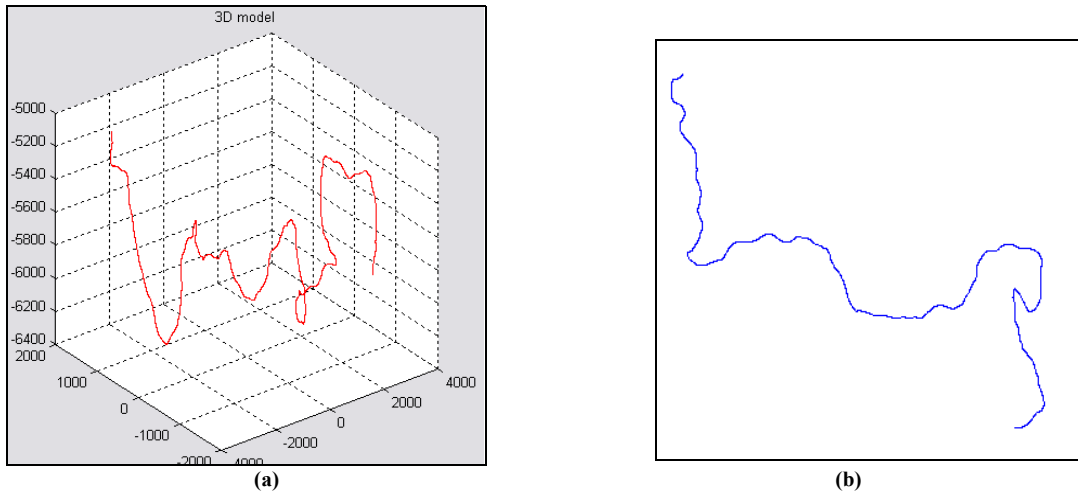


Fig.3: (a) Modèle 3D restitué (b) sa projection sur le plan XY.

Comparaison des deux trajectoires extraites

Après une brève étude des différentes techniques d'analyse de forme [6], nous choisissons de considérer la forme comme un ensemble de points caractéristiques aux quels nous associons un descripteur d'image locale [4]. En effet, à cause de l'aspect tridimensionnel de notre objet nous avons besoin d'un modèle plus général invariant à toute transformation rigide, affine ou projective. Nous adoptons donc un algorithme pour la reconnaissance de forme basé sur les ensembles de distance [1,2]. Il s'agit d'un algorithme qui, à chaque primitive, associe un ensemble de distance (euclidienne) entre ce point et ses n-plus proches voisins (Fig.4). L'association entre deux formes est faite lorsque le coefficient de dissimilarité(distance) entre deux ensembles de distances est proche de zéro.

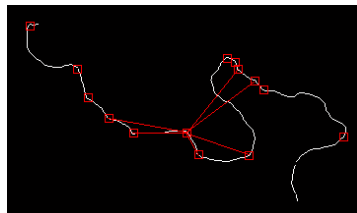


Fig.4: Distances d'une primitive à ses 5-plus proches voisins

Résultats

Le classement donné par l'algorithme de reconnaissance de forme basé sur les ensembles de distances correspond donc à notre choix visuel. Nous nous intéressons aux deux trajectoires les plus proches de la trajectoire de référence, données par la figure ci-dessous.



Fig.5.a Trajectoire de référence



Fig.5.b Im9, $D_{S1,N1,S2,N2} = 0.1165$



Fig.5.c Im3, $D_{S1,N1,S2,N2} = 0.1484$

Les testes ont révélé une similitude satisfaisante et avec un taux « d'erreur » pouvant être quantifié. Cependant, nous remarquons que l'algorithme détecte quelques similitudes entre des trajectoires qui ont des conformation très différentes.

Nous pensons que le problème réside dans le fait que cet algorithme, bien qu'il soit invariant aux transformations géométriques, ne peut pas déceler l'orientation de la trajectoire. La distance euclidienne ne prend pas en considération la position des points dans le plan (Fig.6), c'est bien ce qui fait l'invariance de cet algorithme. En plus d'être son point fort, cette propriété est aussi son point faible car elle limite ses performances.



Fig. 6.a Trajectoire de référence



Fig. 6.b Im6, $D_{SI,NI,S2,N2} = 0.1538$

En conclusion, l'algorithme de reconnaissance de forme par ensembles de distances est efficace car il retrouvera dans la plus part des cas les formes les plus proches. Il est facile à implémenter. Seulement, il faut faire attention aux formes intruses qu'il décèle. Toutefois, une augmentation du nombre de voisins suffit à limiter ces cas de figures.

3. Conclusion

L'intérêt de cette approche, dite guidée par le modèle, est de se baser sur le modèle prédit, donnée par ADN-Viewer. Dans ce cas, nous nous référons à un modèle déjà existant dans le but de simplifier le traitement. De plus, le passage du modèle 3D vers le modèle 2D, par projection géométrique, est simple à mettre en place.

Nous avons alors privilégié l'analyse d'image et la reconnaissance de forme 2D car nous avons une forme d'origine à retrouver parmi un ensemble de formes différentes. Par ailleurs, l'étude que nous avons menée nous a révélé que ce domaine regorge de travaux sur l'appariement, le recalage, l'identification d'objet et la reconnaissance de forme en imagerie bidimensionnelle. Cette approche, guidée par le modèle, évite une trop grande dépendance des données (images réelles) que nous ne maîtrisons pas totalement. Elle nous a permis d'aboutir à un premier résultat assez satisfaisant.

Toutefois, les possibilités de correspondances trajectoires réelles/trajectoires prédites sont énormes (prenant compte du nombre de formes différentes dans une même image microscopique). Il serait judicieux d'apprendre aux système à classer les trajectoires selon leurs degrés de similarité et de n'en retenir que les plus pertinentes, sachant que notre prochain objectif sera de quantifier la différence des trajectoires et d'améliorer le modèle 3D prédictif.

Références

- [1] A. Yali: "2D Object Detection and Recognition Models, Algorithms, and Networks", *MIT Press*, ISBN 0-262-01194-8, 2002.
- [2] C. Grigorescu, N. Petkov: "Distance sets for shape filters and shape recognition", *IEEE Transactions on Image Processing* 12(10): 1274-1286, 2003.
- [3] J. Herisson, R. Gherbi: « Model-Based Prediction of the 3D Trajectory of Huge DNA Sequences Interactive Visualization and Exploration », *2nd IEEE International Symposium on Bioinformatics and Bioengineering (BIBE'01)* p. 263, 2001.
- [4] K. Mikolajczk and C. Schmid: "A performance evaluation of local descriptors," *IEEE Conf. On Computer Vision and Pattern Recognition*, Juin 2003.
- [5] R. Gherbi, J. Hérisson: « 3D modeling tools for spatial-based in silico analysis of DNA ». *International Electronic Journal on Computer Graphics and Geometry*. Vol. 3, N^o1, 2001.
- [6] S. Loncaric, "A survey of shape analysis techniques," *Pattern Recognition*, vol. 31, no. 8, pp. 983–1001, 1998.