

« Alphabet mécanique » pour la modélisation moléculaire interactive de biopolymères à différentes échelles

Présentation générale Les biopolymères, Protéines (P) et Acides Nucléiques (AN=ADN, ARN), sont des micromachines universelles à tout faire extrêmement efficaces. Un des problèmes fondamentaux est de comprendre comment une séquence particulière donne lieu à un repliement et à une ou des fonctions spécifiques comme tentent de le faire les programmes phares « [FoldIt](#) » ou « [Rosetta](#) ». Ce sujet vise à participer à la résolution de ce problème fondamental, et à développer un outil de modélisation et de simulation interactif des biopolymères, à l'interface de la physique et de la biologie (équipe 1), à l'aide d'approches de mécanique et de robotique (équipe 2). Il propose de construire un nouveau langage géométrique, qui devrait compléter nos connaissances non seulement sur les boucles des biopolymères structurés, sur lesquelles on dispose peu d'outils spécifiques théoriques, mais aussi sur celles des [Protéines Intrinsèquement Désordonnées \(PID\)](#). Les PID constituent une catégorie très importante en nombre et par leurs fonctions, et sont par définition quasi absentes de la banque mondiale de données des structures résolues de biopolymères ([PDB](#)).

Stratégie générale Il s'agit d'établir une véritable approche interdisciplinaire entre la mécanique et la biochimie structurale, pour la description et construction des biopolymères, afin de réconcilier à long terme le comportement mécanique et la réactivité. L'approche proposée consiste dans un premier temps à décrire la conformation des biopolymères comme une chaîne cinématique discrète formée de liens ou liaisons atomiques, mais aussi comme la déformée d'une tige solide flexible en trois dimensions, au moyen de la théorie d'élasticité non-linéaire des poutres. C'est une description multi-échelles des molécules comme des chaînes articulées à l'échelle atomique, mais aussi comme des objets modulaires à des échelles intermédiaires, et continus réguliers à des échelles plus grandes.

Contexte et enjeux En 2018, la PDB s'est accrue de 11247 P pour un total de 136924 P, et seulement de 119 AN pour un total de 3221 AN. Pourtant depuis ~2000, il est admis que les conformations des AN pourraient être aussi riches que celle des protéines. Aussi, cette richesse structurale et les faibles connaissances actuelles sur les AN rendent difficile la démarche bioinformatique qui consiste à inférer une conformation à partir de celles déjà connues. D'un autre côté, les protéines structurées constituent une référence complète, puisqu'on estime que la majorité, sinon la quasi totalité des repliements protéiques est connue depuis 2008 (cf. bases de données SCOP et CATH). Toutefois, les PID représentent un champ relativement vierge.

Pour avancer vers une compréhension des AN et des PID, un problème central est de parvenir à faire le lien hiérarchique physique entre les différentes échelles moléculaires (exemple des AN : atomes, angles de torsion, résidus (phosphate, sucre, bases), nucléotides, boucles, hélices). Le squelette de la chaîne macromoléculaire en est le lien naturel. L'équipe 1 a développé une approche de modélisation moléculaire [1] appelée *Biopolymer Chain Elasticity* (BCE), fondée sur l'observation que la chaîne sucre-phosphate des AN se comporte aux échelles mésoscopiques comme une tige flexible continue. Avec ce protocole de résolution [2], nous avons résolu la structure d'un aptamère en épingle à cheveux d'ADN anti-MUC1 [3]. Les résultats sont remarquables car ils reproduisent toutes les données RMN, et ils correspondent à la fois à un minimum global, intermédiaire, et local, i.e. à l'échelle de plusieurs nucléotides, du nucléotide, et de la liaison atomique. Entre 2014 et 2018, nous sommes parvenus à lever deux verrous majeurs dans le domaine de la mécanique des tiges par une collaboration étroite entre le LJP et l'ISIR avec la participation de M. Sébastien Neukirch à l'IJLRDA. Avec la thèse d'Olivier Ameline (2014-18) nous avons montré que trois paramètres physiques (λ , t_p , a), suffisent pour décrire et classer toutes les

déformations statiquement admissibles des tiges infinies de manière très compacte, comme par exemple les boucles, et les hélices. Nous avons aussi démontré que toute déformée s'inscrit dans un tube hélicoïdal paramétrable (cf. Fig 1) [4].

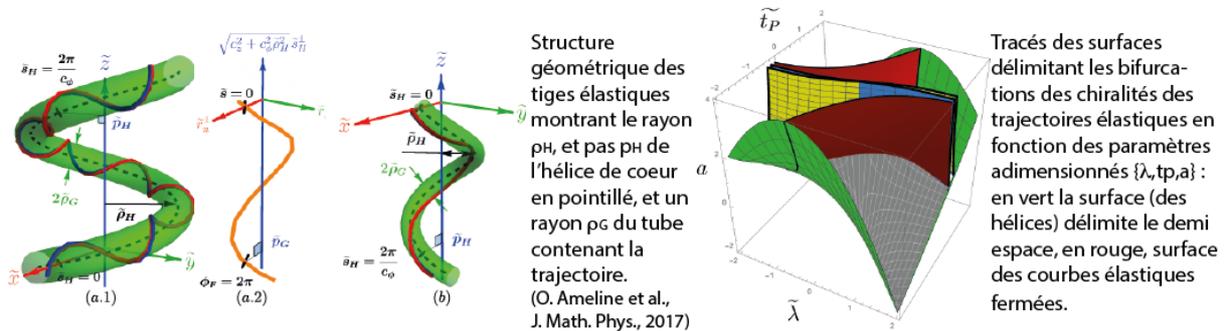


Fig. 1 : Paramètres géométriques, et Classification des trajectoires élastiques

Nous avons ainsi à notre disposition des descriptions directes en fonction de paramètres physiques, ou en fonction de paramètres de caractéristiques géométriques. De plus, nous disposons d'un système analytique interactif robuste capable de générer des solutions du problème de repliement des poutres flexibles, ainsi que d'une méthode inverse pour rechercher des segments de trajectoires analytiques en fonction de conditions spécifiques d'encastrement des deux extrémités de la tige [5]. Ces tiges ou poutres en grande déformation correspondent à des objets solides avec orientation des sections ou repères de Darboux, pour repérer la torsion physique, t_p , le long de la tige. Cette modélisation continue et ce paramètre offrent à la fois une richesse suffisante de formes pour reproduire celles des biopolymères, et une propriété naturelle de rabotage des modules de plusieurs monomères ou des formes isolées.

Parallèlement à cette démarche, nous avons aussi montré que les chaînes cinématiques infinies possédaient les mêmes caractéristiques géométriques que celles que nous avons décrites ci-dessus pour les tiges infinies.

Projet : Nos découvertes sur la classification des formes de tiges infinies sont initialement inspirées des formes observables dans la PDB. En retour, il faut maintenant formaliser ce langage paramétré avec des formes élémentaires. En effet, à l'heure actuelle, les « [splines](#) », ou courbes de Bézier, ou les rubans sont l'un des modes les plus utilisés de représentations graphiques du squelette des biopolymères. Ils sont calculés avec des splines ordinaires (B-splines) qui sont des courbes polynomiales mathématiques très utiles graphiquement en CAO, mais sans signification physique. À l'inverse, ces formes élémentaires, correspondent à des « vraies splines élastiques », et seront des objets géométriques et mécaniques précis qui minimisent l'énergie, pour représenter et manipuler continûment toutes les chaînes de biopolymères. Ces objets représenteront aussi les orientations solides correctes des chaînes latérales attachées à des sections particulières.

Pour atteindre ce premier objectif, il reste plusieurs points théoriques à résoudre et notamment la définition du paramètre, t_p , commun aux descriptions des chaînes cinématiques discrètes et des tiges continues.

Une preuve de concept devra être validée en installant ce moteur générateur de formes dans le programme de modélisation moléculaire du laboratoire, Smol, ou encore [FoldIt](#), ou [Chimera](#).

Implications attendues : Par définition, nos splines élastiques peuvent être soumises à des interactions à la disposition des utilisateurs pour toutes les questions de modélisations de toutes tailles: reconnaissance, interaction, docking. Depuis Unger *et al.* (1989) **Proteins**, 5, 355-373, le concept d'alphabets structuraux s'est progressivement imposé, notamment avec les travaux de plusieurs équipes parisiennes (A.-C. Camproux, A. de Brevern, P. Tufféry et

Projet de Recherche Doctoral Concours IPV 2020

site RPBS...). On peut donc restreindre l'étude générale des protéines de la PDB aux lettres des alphabets structuraux, et aux éléments les plus importants des classifications établies (hélices α et feuillettes β , et classifications de boucles de J. Chomilier, IMPMC, SU). Les splines élastiques de l'approche proposée ici doivent être identifiées dans les protéines, afin de faire apparaître un alphabet avec un vrai sens physique. Il s'agit dès lors d'une nouvelle façon d'aborder les molécules. Pour les AN, cette approche doit donner pour la première fois un moyen de quantifier à une échelle mésoscopique les forces et les énergies mécaniques à l'œuvre dans les conformations des deux boucles d'ADN de référence très bien résolues [1-3]. Les boucles d'AN sont typiquement des agencements de ces formes élémentaires.

Concernant les PID, certaines parties des protéines du ribosome sont considérées comme « déstructurées ». Or, nous avons montré que l'une d'entre elles, eL42, a un rôle catalytique vital pour la synthèse de toutes les protéines [6,7]. eL42 est donc une protéine qui pose un paradoxe. Notre nouvelle description mécanique devrait fournir des moyens de mieux rendre compte de ses propriétés mécaniques extraordinaires qui lui servent à synthétiser toutes les protéines structurées ou non, ainsi qu'elle-même.

Ces approches de modélisation ouvrent de nouvelles perspectives pour la manipulation interactive des biopolymères par la simulation avec retour haptique. La description mécanique, couplée à des champs de force, permettra à un opérateur de manipuler des biopolymères en réalité virtuelle et de ressentir leurs interactions. Toucher et manipuler les molécules est le rêve de la simulation moléculaire haptique interactive.¹ Or l'approche chaîne de polymère flexible, BCE, et les splines élastiques présentent l'avantage remarquable de reproduire le comportement de tige flexible comme le mode de déformation physique pertinent à l'échelle (mésoscopique) de la rigidité de quelques nucléotides et à différentes échelles. C'est pourquoi nous voulons développer la simulation moléculaire haptique de ces molécules, avec un outil de simulation immersive en utilisant les compétences et le matériel de l'**ISIR** [8,9].

Une retombée potentielle intéressante est de concevoir des systèmes robotiques modulaires inspirés des biopolymères pour la conception et la commande de robots, ou encore qui remplaceraient une chaîne articulée d'éléments solides rigides par des éléments déformables.

Rôle respectif des équipes engagées Chaque équipe apporte des expertises uniques, très différentes et complémentaires, dont le doctorant pourra bénéficier selon son origine :

1 : biophysique moléculaire et structurale (conformations des AN, à l'origine d'effets biologiques, par mécanique et dynamique moléculaire ; développement de l'approche BCE),
2 : interfaces, notamment haptiques (pilotage intuitif de l'ordinateur, manipulation d'un système mécanique) : asservissement et réalité virtuelle en modélisation moléculaire interactive et les équipements existants : salle de modélisation virtuelle de l'**ISIR**

Bibliographie (<http://www.labos.upmc.fr/ljp/?article12> : références et liens téléchargeables)

[1] Thèses de G. Santini¹ (2005), et C. Pakleza¹ (2002) sur Biopolymer Chain Elasticity (équipe 1).

[2] G.P.H. Santini¹, J.A.H. Cognet¹, et al. (2009) **J. Phys. Chem. B**, **113**, 6881-6887.

[3] M. Baouendi¹, J.A.H. Cognet¹ et al. (2012) **FEBS J.**, **279**, 479-490.

[4] O. Ameline^{1,2}, S Haliyo², X. Huang^{1,2}, J.A.H. Cognet¹ (2017), **J. Math. Phys.**, **58**, 062902, 1-27.

[5] O. Ameline^{1,2}, S Haliyo², X. Huang^{1,2}, J.A.H. Cognet¹ (2018), **J. Comp. Phys.**, **373**, 736-749.

[6] C. Hountondji et al. (avec J.A.H. Cognet¹) (2017) **J. Biochem.**, **162**, 6, 437-448.

[7] C. Hountondji et al. (avec J.A.H. Cognet¹) (2019) **Biochimie**, **158**, 20-33.

[8] A. Bolopion², et al. & S. Régnier² (2011), **IEEE WorldHaptics**, 469-474.

[9] A. Bolopion², B. Cagneau², D. S. Haliyo² & S. Régnier² (2008) **J. of Micro-Nano Mechatronics**, **4**, 145-158.

¹ Une interface haptique, ou à retour d'effort, est un dispositif robotique (poignée) qui reçoit un déplacement imposé par l'utilisateur et qui est capable de générer des forces et des couples envoyés à l'utilisateur.