

Projet de Recherche Doctoral Concours IPV 2021

Intitulé du Projet de Recherche Doctoral : « Chaînon manquant » pour la modélisation moléculaire et mécanique des protéines et des acides nucléiques

Directeur de Thèse porteur du projet (titulaire d'une HDR) :

NOM : **COGNET** Prénom : **Jean**

Titre : **Professeur des Universités**

e-mail : jean.cognet@sorbonne-universite.fr

Adresse professionnelle : Laboratoire Jean Perrin LJP

UMR 8237 T32-33 4^e case courrier 114,

Sorbonne Université CNRS 4 place Jussieu 75252 PARIS cedex 05

Unité de Recherche :

Intitulé : Laboratoire Jean Perrin LJP

Code : UMR 8237

Equipe de Recherche (au sein de l'unité) :

Intitulé : **Modélisation mésoscopique des biopolymères**

Thématique de recherche : Modélisation des biomolécules à différentes échelles

Responsable d'équipe :

NOM : **COGNET** Prénom : **Jean**

Ecole Doctorale de rattachement de l'équipe &

d'inscription du doctorant : ED564-Physique en IdF

Doctorants actuellement encadrés par le directeur de thèse (préciser le nombre de doctorants, leur année de 1^{ere} inscription et la quotité d'encadrement) : 0 (aucun)

CO-DIRECTION (obligatoire)

Co-Directeur de Thèse (titulaire d'une HDR) :

NOM : **NEUKIRCH** Prénom : **Sébastien**

Titre : **Directeur de recherche CNRS** HDR

e-mail : sebastien.neukirch@sorbonne-universite.fr

Unité de Recherche :

Intitulé : Institut Jean Le Rond d'Alembert (IJLRDA)

Code : UMR 7190

Equipe de Recherche (au sein de l'unité) :

Intitulé : **Mécanique et Ingénierie des Solides Et des Structures (MISES)**

Thématique de recherche : Mécanique des solides

Responsable d'équipe :

NOM : **BRENNER** Prénom : **Renald**

Ecole Doctorale de rattachement : ED 391 SMAER

Ou si ED non SU :

Doctorants actuellement encadrés par le co-directeur de thèse (préciser le nombre de doctorants, leur année de 1^{ere} inscription et la quotité d'encadrement) : 1, 2018-21, à 50% :

Il s'agit de :

Raphaël CHARRONDIÈRE : 1^{ère} année : 19 Sept. 2018 – 3^{ème} année Sept. 2021 à l'INRIA

Grenoble en co-encadrement avec S. Neukirch (50%), et Mme Florence Bertails-Descoubes (INRIA) (50%).

Cotutelle internationale : Non Oui, précisez Pays et Université :

Résumé (2 000 caractères maximum) :

Le problème du repliement des Protéines (P) et des Acides Nucléiques (AN) est au cœur de la machinerie de la vie. Or, les méthodes actuelles de modélisation moléculaire ne suffisent pas pour résoudre ce problème primordial. Une difficulté majeure est qu'il n'existe aucune connaissance détaillée de la façon d'enchaîner les fragments de biopolymères, et de la façon de concevoir les fragments à assembler, sauf dans le contexte spécial d'alphabet structural.

Nous proposons une vision novatrice de ces macromolécules, en les décrivant comme un assemblage mécanique de Tiges Elastiques (TE) (inextensibles). Cette description de géométrie différentielle est très naturelle pour les P et les AN, puisque ces biopolymères sont des chaînes linéaires assimilables à des fils ou tiges rigides à l'échelle de plusieurs résidus. Elle a fourni et fournira de nouvelles relations mathématiques, mécaniques, et physiques qui relient les fragments de chaînes de biopolymères.

Pour les AN, cette approche quantifiera pour la première fois à l'échelle des nucléotides, les forces et les énergies mécaniques dans des conformations de boucles d'ADN de référence. Pour les protéines, nous validerons le modèle développé en décrivant toutes les lettres d'un des meilleurs alphabets structuraux. Dans les deux cas, nous montrerons comment les assembler géométriquement et mécaniquement. Nous utiliserons cette description pour comprendre la protéine ribosomale, eL42, qui a un rôle catalytique vital pour la synthèse de toutes les protéines, et qui est classée comme Protéine Intrinsèquement Désordonnée.

Un outil général de simulation fondé sur ces approches de mécanique continue et discrète est alors envisageable pour compléter la résolution et la conception des biopolymères, car ils pourront être soumis à des interactions à la disposition des utilisateurs pour toutes les questions de modélisations à différentes échelles : reconnaissance, interaction, docking, simulations haptiques.

**Joindre en annexe un descriptif du PRD avec références au format pdf
(« NOM_2_IPV_2021 » / 3 pages maximum, taille police 11)**

AVIS et VALIDATION de l'ECOLE DOCTORALE :

Favorable



École Doctorale
Physique en Île de France
EDPIF - ED n° 564
P^r Maria Chamarro
Directeur-adjoint SU

**à envoyer simultanément par e-mail à l'ED de rattachement et au programme :
interfaces_pour_le_vivant@listes.upmc.fr avant le lundi 15 février minuit.**

« Chaînon manquant » pour la modélisation moléculaire et mécanique des protéines et des acides nucléiques

Présentation générale Le problème du repliement des Protéines (P) et des Acides Nucléiques (AN=ADN, ARN), est au cœur de la machinerie de la vie et des fonctions cellulaires. Or, les méthodes actuelles de modélisation de ces biopolymères ne suffisent pas pour résoudre ce problème primordial. L'Intelligence Artificielle connaît maintenant un très beau succès avec Alpha Fold 2. Mais l'IA repose sur les informations structurales disponibles, qui sont rares pour les AN, et très rares pour les [Protéines Intrinsèquement Désordonnées \(PID\)](#). Celles-ci, très importantes par leur nombre et leurs fonctions, sont, par définition, quasi absentes de la banque mondiale de données des structures résolues de biopolymères ([PDB](#)). 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'aide d'approches de biophysique moléculaire (équipe 1), et de mécanique (équipe 2), à l'interface avec la biologie. Il propose de construire un nouveau système géométrique, analytique et informatique (les splines élastiques) à la frontière entre la mécanique, la physique des biopolymères, et l'ingénierie qui devrait fournir un nouveau point de vue mécanique pour comprendre non seulement les boucles des biopolymères structurés, pour lesquelles on dispose de peu d'outils, mais aussi les PID.

Contexte et enjeux Classiquement, les structures macromoléculaires sont considérées comme des nuages de points dans l'espace 3D cartésien. Chaque point représente un atome, et les forces interatomiques peuvent être décrites par une fonction paramétrable (champ de force). Dans les simulations biomoléculaires, à chaque étape de temps, les atomes sont déplacés en fonction de ces forces. Cette représentation mécanique et physique a permis de simuler avec Anton, un superordinateur massivement parallèle, le repliement spontané de plusieurs protéines dans de l'eau explicite (Piana *et al.*, 2012). Cependant, elles restent trop coûteuses en temps de calcul pour bien caractériser l'espace de repliement des macromolécules. Cette limitation peut être surmontée par l'analyse des modes normaux, qui résout analytiquement la loi du mouvement de Newton en utilisant une approximation quadratique de l'énergie potentielle. Mais elle est trop limitée par la géométrie de départ et ne peut pas modifier les structures secondaires. Comme alternative aux coordonnées cartésiennes, des angles de torsion définissant la conformation des macromolécules peuvent être utilisés. Cependant, de petites variations de ces coordonnées internes entraînent une instabilité structurale.

Alternativement, les méthodes fondées sur les connaissances, c-à-d sur des bases de données structurales, calculent des conformations 3D en assemblant des hélices et plus généralement des motifs extraits des structures connues des Protéines et des AN. Cette stratégie est utilisée par des programmes de modélisation très populaires, comme Rosetta et FoldIt, pour la prédiction et la conception de la structure protéique. FARFAR2 étend actuellement cette approche à l'ARN (Watkins *et al.*, 2020). Une des limites de ces méthodes est qu'elles reposent en grande partie sur les informations structurales disponibles. Mais, la quantité relativement faible de structures d'AN ou P/AN (seulement ~11303 sur les ~158966 entrées de la PDB en 2019, <https://www.rcsb.org>) empêche une description précise de leur espace de repliement.

Tous les outils de modélisation actuellement disponibles ont du mal à prédire avec précision les structures et le comportement des petits fragments de P et d'AN tels que les boucles. Cette observation a mené à l'identification des questions majeures qu'il faut résoudre à l'échelle de quelques résidus (Das, 2011). Les boucles jouent souvent des rôles cruciaux dans le fonctionnement et les interactions des macromolécules. Une description plus précise de leurs structures et dynamiques est donc nécessaire. De façon plus générale, il n'existe aucune connaissance détaillée de la façon d'enchaîner les fragments de biopolymères et de la façon de concevoir les fragments à assembler, sauf dans le contexte spécial d'alphabet structural (Camproux *et al.* 2004).

Projet : Nous proposons une vision vraiment novatrice des macromolécules, en les décrivant comme des Tiges Élastiques (TE), et plus précisément comme des tiges inextensibles

élastiques en flexion et en torsion. Cette représentation est à la fois géométriquement complémentaire et plus intuitive que les coordonnées cartésiennes ou internes classiques. Dans l'espace cartésien, déplacer des atomes équivaut à translater ces atomes. Tout en préservant la structure globale de la chaîne des biopolymères, il en résulte la rupture des liaisons atomiques et la génération d'encombrements stériques. Dans l'espace de coordonnées internes, le déplacement des atomes équivaut à la rotation des atomes autour des liaisons atomiques. Tout en préservant la géométrie locale des liaisons atomiques, cela détruit l'orientation globale de la chaîne de polymères et dégrade la plupart des interactions stabilisatrices. La description des atomes le long d'une tige inextensible (avec une abscisse curviligne 1D fixe dans l'espace 3D) résout ce dilemme en permettant de grandes déformations tout en préservant la géométrie locale. Cette description (dans les repères de la géométrie différentielle) est très naturelle pour les protéines et les acides nucléiques puisque ces biopolymères sont des chaînes linéaires assimilables à des fils ou des tiges rigides à l'échelle de plusieurs résidus (Rose *et al.*, 2006). Associée aux autres représentations classiques des biopolymères, elle ajoute un nouveau niveau de description physique pertinent à l'échelle intermédiaire de plusieurs résidus (acides aminés ou nucléotides). En conséquence, cela fournit de nouvelles relations mathématiques et physiques pour relier les différents fragments de chaînes de biopolymères.

Stratégie générale L'approche proposée consiste à décrire la conformation des biopolymères de deux façons différentes, d'une part classiquement comme une Chaîne Articulée (CA) discrète formée de liens ou liaisons atomiques, et d'autre part comme la déformée d'une Tige Élastique (TE) solide flexible en trois dimensions, au moyen de la théorie d'élasticité non-linéaire des poutres. C'est donc une description multi-échelles des molécules vues comme des CA à l'échelle des atomes et des acides aminés (ou d'AN), mais aussi comme des objets modulaires à ces échelles intermédiaires, et continus réguliers à des échelles plus grandes, représentés mécaniquement par des TE. Il s'agit d'établir une véritable approche interdisciplinaire entre la mécanique et la biophysique structurale, pour la description et construction des biopolymères, afin de réconcilier à long terme le comportement mécanique et la réactivité [1].

NB : les « [splines](#) », ou les rubans sont l'un des modes les plus utilisés de représentations graphiques du squelette des biopolymères. Ce sont des courbes polynomiales mathématiques très utiles graphiquement en CAO, mais sans signification physique. À l'inverse, l'équivalence entre CA et TE que nous voulons construire correspond à de « vraies splines élastiques ». Elles seront des objets géométriques et mécaniques précis qui minimisent l'énergie, pour représenter et manipuler continûment toutes les chaînes de biopolymères en 3D.

Méthodologie Pour mener à bien ce projet, nous nous appuyerons sur un corpus de preuves formelles et d'outils mathématiques récemment développés dans le domaine de l'élasticité. Plus précisément, l'équipe 1 a montré que le squelette sucre-phosphate des AN se comporte à l'échelle mésoscopique (quelques nucléotides) comme une tige élastique. Les modèles ont été confirmés expérimentalement à la résolution atomique [2-4]. Nous avons récemment constaté, que trois paramètres physiques sont suffisants pour décrire et classer de façon analytique toutes les déformations admissibles de tiges infinies [5]. Selon ce modèle, la trajectoire de la tige s'enroule autour d'une hélice, appelée « hélice de cœur », avec un rayon et un pas hélicoïdaux donnés, entre deux tubes concentriques autour de cette hélice, avec des rayons intérieurs et extérieurs donnés (*cf.* Fig.). Plus récemment, nous avons observé que ces paramètres géométriques peuvent décrire des CA infinies générées à partir d'un motif répété. C'est un résultat remarquable car un court fragment de biopolymère (objet discret) peut être traité comme une tige élastique continue. Le problème inverse (dérivation de la forme de la tige à partir des conditions d'ancrage) a également été résolu [6]. L'équipe 2 est experte dans les problèmes de mécanique très complexes liés aux biomolécules [7], ainsi qu'aux questions topologiques posées par les tiges [8].

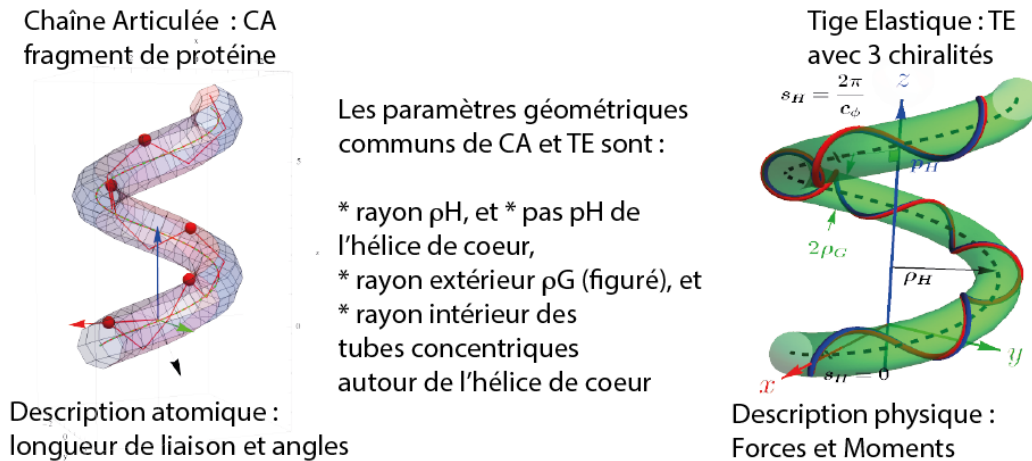


Fig. 1 : Présentation des paramètres géométriques communs aux Chaînes Articulées (CA) et aux Tiges Élastiques (TE), et suffisants pour les décrire : l'hélice centrale ou « hélice de coeur » est figurée par la ligne centrale rouge (à gauche), et par une ligne épaisse en tirets verts (à droite). Les trois chiralités d'une TE infinie sont : (1) l'hélice de coeur (enroulement droit) ; (2) la trajectoire de la tige s'enroule autour de l'hélice de coeur (enroulement gauche) et oscille entre les tubes concentriques extérieur (figuré en vert) et intérieur (non figuré) ; (3) la TE est physiquement tordue comme montrée par les couleurs bleue et rouge de la tige (Ameline *et al*, 2017) [5].

Validations et Implications attendues Pour les AN, cette approche quantifiera pour la première fois à l'échelle des nucléotides, les forces et les énergies mécaniques dans des conformations de boucles d'ADN de référence. Pour les protéines, nous validerons le modèle développé en décrivant toutes les lettres d'un des meilleurs alphabets structuraux (avec P. Tufféry RPBS), et en les testant dans une banque de boucles (avec J. Chomilier & D. Stratmann, IMPMC, SU). Dans les deux cas, nous montrerons comment les assembler géométriquement et mécaniquement. Nous utiliserons enfin cette description pour comprendre la protéine ribosomale, eL42, qui a un rôle catalytique vital pour la synthèse de toutes les protéines [9] et qui est classée comme PID.

Un outil général de simulation fondé sur ces approches de mécanique continue et discrète est alors envisageable pour compléter la résolution et la conception des biopolymères, car ils pourront être soumis à des interactions à la disposition des utilisateurs pour toutes les questions de modélisations à différentes échelles : reconnaissance, interaction, docking. Cette approche de modélisation par des tiges flexibles / chaînes articulées est particulièrement pertinente pour la virtualisation moléculaire et pour la simulation interactive avec retour haptique à l'ISIR.

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

- 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 : mécanique des tiges élastiques : étude théorique et numérique, topologie des solutions, applications biologiques.

Bibliographie <http://www.labojeanperrin.fr/ljp/?article12> : références & liens téléchargeables

- [1] S. Neukirch, A. Goriely, et A. Hausrath, (2008) *Int. J. of Non-Linear Mechanics*, **43**, 1064-1073.
- [2] Thèses de G. Santini (2005), et C. Pakleza (2002) sur Biopolymer Chain Elasticity (équipe 1).
- [3] G.P.H. Santini, J.A.H. Cognet, et al. (2009) *J. Phys. Chem. B*, **113**, 6881-6887.
- [4] M. Baouendi, J.A.H. Cognet et al. (2012) *FEBS J.*, **279**, 479-490.
- [5] O. Ameline, S Haliyo, X. Huang, J.A.H. Cognet (2017), *J. Math. Phys.*, **58**, 062902, 1-27.
- [6] O. Ameline, S Haliyo, X. Huang, J.A.H. Cognet (2018), *J. Comp. Phys.*, **373**, 736-749.
- [7] A. Goriely, A. Hausrath, et S. Neukirch, (2008) *Biophysical Reviews and Letters*, **3**, 77-101.
- [8] C. B. Prior et S. Neukirch (2016) *J. Phys. A: Math. Theo.*, **49**, 215201.
- [9] C. Hountondji (J. Cognet) (2017) *J. Biochem.*, **162**, 6, 437-448, (2019) *Biochimie*, **158**, 20-33.