



Sujet: Repliement d'un modèle simplifié de protéine par un algorithme de Monte-Carlo et échange de répliques

Objectif: Réaliser un programme reprenant la méthode décrite dans l'article. Dans un premier temps l'algorithme de Monte-Carlo simple sera implémenté, ensuite le RE. Une séquence arbitraire de protéine pourra être soumise au programme afin de calculer son repliement par ce modèle.



Contact : Jean-Christophe Gelly (jean-christophe.gelly@univ-paris-diderot.fr)

Thachuk C, Shmygelska A, Hoos HH. A replica exchange Monte Carlo algorithm for protein folding in the HP model. BMC Bioinformatics. 2007 Sep 17;8:342. PubMed PMID: 17875212; PubMed Central PMCID: PMC2071922.

Sujet: Calcul de la surface accessible au solvant d'une protéine

Objectif : Réalisez un programme permettant de calculer la surface accessible au solvant (absolue et relative) à partir des coordonnées d'une protéine issue d'un fichier PDB.

Etapes :

- Extraction des coordonnées de chaque atome
- A partir de chaque atome contenu caractérisant la protéine, créez un nuage de points uniformément sur la surface d'une sphère centrée sur l'atome. La sphère aura pour valeur de rayon le rayon de Van der Waals de l'atome + rayon de la sonde (solvant = rayon d'un atome d'oxygène) (l'algorithme de Saff et Kuijlaars pourra être utilisé).
- A partir de chaque point de la sphère recherchez s'il existe un point appartenant à une autre sphère à une distance inférieure au rayon de la sonde
- A partir des points sans contact, calculez la surface accessible totale en terme de points puis la convertir en \AA^2 de chaque résidu de la protéine. Calculez ensuite la surface accessible relative, et enfin le pourcentage d'accessibilité au solvant.
- Comparez et évaluez la méthode par rapport à NACCESS

Shrake, A; Rupley, JA. (1973). "Environment and exposure to solvent of protein atoms. Lysozyme and insulin". *J Mol Biol* **79** (2): 351–71. doi: [10.1016/0022-2836\(73\)90011-9](https://doi.org/10.1016/0022-2836(73)90011-9).

Contact : Jean-Christophe Gelly (jean-christophe.gelly@univ-paris-diderot.fr)

Sujet : Assignment des structures secondaire de protéines

Objectif: Implémentation de la méthode DSSP pour l'assignation des structures secondaires. Dans un premier temps, il est possible de s'intéresser qu'aux structures secondaires régulières (alpha Hélice, et brin bêta).

Contact : Alexandre de Brevern (alexandre.debrevern@univ-paris-diderot.fr) & Jean-Christophe Gelly (jean-christophe.gelly@univ-paris-diderot.fr)

Référence :

Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Kabsch W, Sander C, Biopolymers. 1983 22 2577-2637. PMID: 6667333; UI: 84128824.

Sujet: Assignment et détection des parties transmembranaires d'une protéine

Objectif: Etablir un outil permettant de déterminer les zones transmembranaires d'une protéine.

Etapes:

- Choisir des protéines membranaires et globulaires de référence (simple à analyser, voir banque OPM)
- Calcul de la surface accessible au solvant de chaque acide aminé avec NACCESS
- Lecture du PDB et extraction des coordonnées des Calpha
- Calcul du centre de masse
- Détermination des droites passant par le centre de masse et quadrillant avec suffisamment de précision toutes les directions
- Déplacement d'une tranche de 1 Angstrom normale à une droite et calcul de l'hydrophobicité relative des résidus dans les zones exposées dans la tranche
- Calcul de la position de la membrane par la moyenne de l'hydrophobicité relative et en comparant ces valeurs selon les différentes droites

Contacts: Jean-Christophe Gelly

(jean-christophe.gelly@univ-paris-diderot.fr)

Référence :

Tusnády GE, Dosztányi Z, Simon I. Transmembrane proteins in the Protein Data Bank: identification and classification. Bioinformatics. 2004 Nov

22;20(17):2964-72. Epub 2004 Jun 4. PubMed PMID: 15180935.

Remarque

- Il n'est pas demandé de vérifier si la protéine est de taille suffisante ou si le fichier contient effectivement une protéine (on prendra donc soin de vérifier son programme sur des chaîne unique de protéines membranaire et de protéines globulaire)
- Il n'est pas demandé de reconstruire le modèle biologique de la molécule (à partir du champ BIOMOLECULE)
- Le programme NACCESS ou DSSP sera utilisé pour calculer la surface accessible au solvant

Sujet: Conception d'un programme de *threading* par double programmation dynamique

Objectif: Réaliser un programme reprenant la méthode décrite dans l'article 3 basé sur la double programmation dynamique (pour plus d'information 4). Le *threading* (enfilage) [1,2,3] est une stratégie pour rechercher des séquences compatibles avec une structure. Seul les carbones α de la protéine seront considérés. Vous utiliserez les potentiels statistiques DOPE [5].

Contact : Jean-Christophe Gelly (jean-christophe.gelly@univ-paris-diderot.fr)

1- Jones, D.T., Taylor, W.R. & Thornton, J.M. (1992) A new approach to protein fold recognition. *Nature*. 358, 86-89.

2- Jones, D.T., Miller, R.T. & Thornton, J.M. (1995) Successful protein fold recognition by optimal sequence threading validated by rigorous blind testing. *Proteins*. 23, 387-397.

3- Jones, D.T. (1998) THREADER : Protein Sequence Threading by Double Dynamic Programming. (in) *Computational Methods in Molecular Biology*. Steven Salzberg, David Searls, and Simon Kasif, Eds. Elsevier Science. Chapter 13.

4- [Protein Structure Comparison Using SAP - Springer](#)

5- <http://www.dsimb.inserm.fr/~gelly/doc/dope.par>

Sujet : *Protein Interactions Calculator* implémentation.

Objectif: *Protein Interactions Calculator* (PIC; <http://pic.mbu.iisc.ernet.in>) est un serveur web particulièrement intéressant qui permet le calcul des différents types d'interactions importantes pour la structure protéique : les liaisons disulfure, les interactions hydrophobes, les interactions ioniques, les liaisons hydrogènes, les interactions aromatiques-aromatiques, les interactions aromatiques-soufre et les interactions cation- π (Tina *et al*, *Nucleic Acid Res*, 2007). Il a aussi d'autres propriétés mais cela est en dehors de ce projet.

Le projet est de ré-implémenter ces calculs à partir d'une structure correcte (pas de problème spécifique à gérer), et générer une sortie similaire (et utilisable localement) mettant en avant les interactions entre atomes / résidus.

Contacts : Alexandre de Brevern (alexandre.debrevern@univ-paris-diderot.fr),

Aline Floch (aline.floch@efs.sante.fr) &

Jean-Christophe Gelly (jean-christophe.gelly@univ-paris-diderot.fr).

Référence: □K. G. Tina, R. Bhadra and N. Srinivasan, PIC: Protein Interactions Calculator, *Nucleic Acids Research*, 2007, Vol. 35, Web Server issue W473–W476.

Sujet : Analyse de la communication allostérique à l'aide d'un alphabet structural

Objectif: En 2013, le groupe de Fraternalli propose l'implémentation d'une approche d'analyse des trajectoires de dynamiques moléculaires nommées GSAtools (Pandani *et al*, *Bioinformatics*, 2013, <http://mathbio.nimr.mrc.ac.uk/wiki/GSATools>, <http://people.brunel.ac.uk/~csstaap2/gsatools.html>). Cette approche a été précédemment utilisée sur des ensembles de dynamiques moléculaires (Baussand *et al*, 2012 ; Pandini *et al*, 2012 ; Kleinjung *et al*, 2012). Cette approche est basée sur un ensemble successif d'analyses statistiques (souvent du R). Malheureusement, GSA_tools est dédié à une version ancienne de Gromacs (et n'était d'ailleurs pas réellement utilisable).

Le principe de ce projet est de reproduire les différentes étapes du programme GSAtools, en se basant sur un autre alphabet structural, les Blocs Protéiques. Un logiciel nommé PBxplore a été mis à disposition de la communauté pour traiter des Dynamiques Moléculaires (Barnoud *et al*, bioarxiv preprint ; <https://github.com/pierrepo/PBxplore>). Des simulations de dynamiques moléculaires venant du domaine Calf-1 seront proposés pour tester l'approche (Goguet, Tarwani *et al*, 2017)

Contacts : Alexandre de Brevern (alexandre.debrevern@univ-paris-diderot.fr),

Tarun J. Narwani (tarun.narwani@inserm.fr) &

Jean-Christophe Gelly (jean-christophe.gelly@univ-paris-diderot.fr)

Références: □

Pandini A, Fornili A, Fraternali F, Kleinjung J "GSATools: analysis of allosteric communication and functional local motions using a Structural Alphabet" *Bioinformatics* 29(16):2053-2055, 2013

Références Complémentaires :

Baussand J, Kleinjung J. Specific Conformational States of Ras GTPase upon Effector Binding. *J Chem Theory Comput.* 2013 Jan 8;9(1):738-749.

Pandini A, Fornili A, Fraternali F, Kleinjung J. Detection of allosteric signal transmission by information-theoretic analysis of protein dynamics. *FASEB J.* 2012 Feb;26(2):868-81.

Kleinjung J, Scott WR, Allison JR, van Gunsteren WF, Fraternali F. Implicit Solvation Parameters Derived from Explicit Water Forces in Large-Scale

Molecular Dynamics Simulations. J Chem Theory Comput. 2012 Jul 10;8(7):2391-2403.

Jonathan Barnoud, Hubert Santuz, Pierrick Craveur, Agnel Praveen Joseph, Vincent Jallu, Alexandre G. de Brevern, Pierre Poulain, PBxplorer: A Tool To Analyze Local Protein Structure And Deformability With Protein Blocks (bioRxiv 136408; doi: <https://doi.org/10.1101/136408>).

Matthieu Goguet, Tarun Narwani, Rachel Petermann, Vincent Jallu, Alexandre G. de Brevern (2017) Scientific Reports, sous presse.

Sujet : Regroupement par les k-means de structures à l'aide d'un alphabet structural

Objectif: Le logiciel PBxplore a été mis à disposition pour traiter des Dynamiques Moléculaires à l'aide d'un alphabet structural, les Blocs Protéiques (Barnoud *et al*, bioarxiv preprint ; <https://github.com/pierrepo/PBxplore>). Il permet de (i) coder chaque snapshots (structures) à l'aide des Blocs Protéiques, puis (ii) propose une analyse statistique des variations conformationnelles.

Ce projet portera donc sur le regroupement par les k-means de structures, à l'aide d'un alphabet structural. Dans Gromacs, il existe une approche nommée `g_cluster` qui peut regrouper des structures en utilisant plusieurs méthodes différentes. Ici le principe est de créer k groupes à l'aide de l'encodage des Blocs Protéiques. Deux niveaux de difficulté peuvent être vus (a) la distance est l'identité ou non, (b) avec l'utilisation d'une similarité (les Blocs pouvant se ressembler).

Des simulations de dynamiques moléculaires venant du domaine Calf-1 seront proposés pour tester l'approche (Goguet, Tarwani *et al*, 2017) avec une structure référence (Wild-Type) et/ou des variants associés aux pathologies.

Contacts: Alexandre de Brevern (alexandre.debrevern@univ-paris-diderot.fr),

Jonathan Barnoud (jonathan@barnoud.net) &

Jean-Christophe Gelly (jean-christophe.gelly@univ-paris-diderot.fr)

Référence : □Jonathan Barnoud, Hubert Santuz, Pierrick Craveur, Agnel Praveen Joseph, Vincent Jallu, Alexandre G. de Brevern, Pierre Poulain, PBxplore: A Tool To Analyze Local Protein Structure And Deformability With Protein Blocks (bioRxiv 136408; doi: <https://doi.org/10.1101/136408>).

Matthieu Goguet, Tarun Narwani, Rachel Petermann, Vincent Jallu, Alexandre G. de Brevern (2017) Scientific Reports, sous presse.

Sujet : Comparaison de la dynamique de variants associés à des pathologies à l'aide d'un alphabet structural

Objectif: Le logiciel PBxplore a été mis à disposition pour traiter des Dynamiques Moléculaires à l'aide d'un alphabet structural, les Blocs Protéiques (Barnoud et al, bioarxiv preprint ; <https://github.com/pierrepo/PBxplore>). Il permet de (i) coder chaque *snapshots* (structures) à l'aide des Blocs Protéiques, puis (ii) propose une analyse statistique des variations conformationnelles.

Ce projet portera donc sur la comparaison de résultats de dynamique moléculaire d'une protéine et de variants ponctuels (donc même longueur de séquence) associés à une pathologie. Nous avons comparé par des petits scripts les sorties de PBxplore. Il faudrait avoir une approche rationnelle et correcte d'un point de vue informatique.

Des simulations de dynamiques moléculaires venant du domaine Calf-1 seront proposés pour tester l'approche (Goguet, Tarwani *et al*, 2017) avec une structure référence (Wild-Type) et des variants associés à la Thrombasthénie de Glanzmann.

Contact : Alexandre de Brevern (alexandre.debrevern@univ-paris-diderot.fr),

Vincent Jallu (vjallu@ints.fr) &

Jean-Christophe Gelly (jean-christophe.gelly@univ-paris-diderot.fr)

Référence : □Jonathan Barnoud, Hubert Santuz, Pierrick Craveur, Agnel Praveen Joseph, Vincent Jallu, Alexandre G. de Brevern, Pierre Poulain, PBxplore: A Tool To Analyze Local Protein Structure And Deformability With Protein Blocks (bioRxiv 136408; doi: <https://doi.org/10.1101/136408>).

Matthieu Goguet, Tarun Narwani, Rachel Petermann, Vincent Jallu, Alexandre G. de Brevern (2017) Scientific Reports, sous presse.

Sujet : Assemblage *de novo* d'un génome circulaire

Objectif: Réalisez un programme permettant l'assemblage *de novo* d'un petit génome circulaire en utilisant un graphe de de Bruijn. Pour la reconstruction du génome, vous reprendrez une des deux stratégies décrites dans l'article, cycle hamiltonien **ou** cycle d'Euler et vous justifierez votre choix. Vous commencerez par générer aléatoirement le génome à reconstruire ainsi que les reads.

Contact : Charlotte Périn (charlotte.perin@univ-paris-diderot.fr) &
Jean-Christophe Gelly (jean-christophe.gelly@univ-paris-diderot.fr)

Référence :

Compeau, P. E., Pevzner, P. A., & Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nature biotechnology*, 29(11), 987.