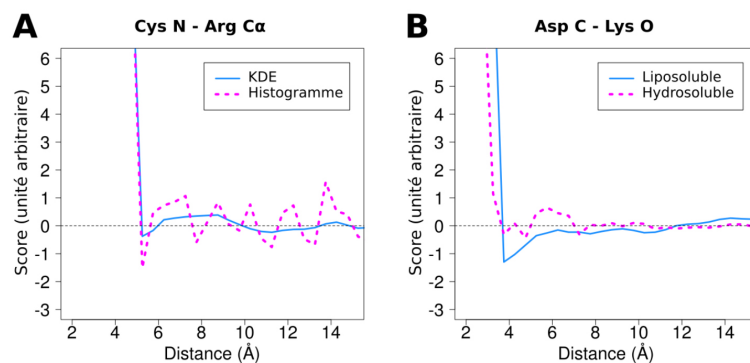


## Un logiciel libre dédié au classement des modèles protéiques

**La modélisation moléculaire permet d'accéder à la structure 3D des protéines, porteuse de leurs fonctions biologiques, de manière beaucoup moins coûteuse que les méthodes expérimentales, telles que la cristallographie aux rayons X, la RMN ou la cryo-microscopie électronique. La qualité des structures ainsi prédites nécessite toutefois d'être évaluée, et de nombreuses méthodes informatiques ont été développées à cet effet. Cependant, aucune ne prend en compte l'environnement cellulaire dans lequel les protéines sont fonctionnelles, alors même que les propriétés physico-chimiques de celui-ci influencent grandement la conformation des protéines. Ainsi, ces méthodes génériques échouent à distinguer entre les prédictions structurales correctes et incorrectes, car elles ignorent les spécificités des différents types de protéines. Des chercheurs de l'équipe Bioinformatique et Biophysique de l'IMPMC, associés à un chercheur de l'université de Copenhague, ont développé un programme d'évaluation de la qualité des modèles qui est adaptable à tous types de structures protéiques.**

Pour être utile, un modèle moléculaire de protéine doit s'approcher autant que possible de la structure native, qui est le repliement adopté par la chaîne polypeptidique dans les conditions natives. Selon l'hypothèse thermodynamique d'Anfinsen (prix Nobel en 1972), cette structure native correspond à la conformation de plus basse énergie. Ainsi, l'évaluation de la qualité des modèles se fait au moyen de fonctions dites de scores dont le minimum global correspond à la conformation native. Le plus souvent, ces fonctions de scores sont des potentiels interatomiques construits à partir d'un ensemble de structures protéiques déterminées expérimentalement. Les statistiques des distances interatomiques de ces structures sont utilisées pour calculer une approximation de l'énergie (le score) via la réciproque de la loi de distribution de Boltzmann : c'est la méthode des potentiels statistiques de force moyenne (abrégiés PMFs en anglais).

L'équipe de l'IMPMC a développé MyPMFs, un programme permettant à l'utilisateur de créer une fonction de score à partir d'un ensemble de protéines de son choix. Pour démontrer que la construction d'une fonction de score précise repose sur une sélection pertinente de structures natives, un banc d'essais de plus de 15 000 structures prédites à partir de 54 protéines membranaires différentes est utilisé. Il apparaît ainsi qu'une fonction de score élaborée à partir d'un ensemble de protéines hydrosolubles ne parvient pas à évaluer correctement la qualité des modèles de protéines issues de l'environnement liposoluble de la bicouche lipidique. En revanche, la fonction de score construite sur des protéines membranaires grâce à MyPMFs produit de bons résultats en termes de classement des modèles de protéines membranaires. Ceux-là s'expliquent en partie par la possibilité qu'offre le programme d'utiliser des traitements statistiques permettant de pallier le manque de données cristallographiques inhérent à l'étude d'un type particulier de protéines, comme celles de la membrane (Figure 1A).



**Figure 1**

(A) Effets de l'utilisation d'une estimation de densité par noyau (KDE) ou d'un histogramme sur les fluctuations artificielles du potentiel statistique causées par le manque de données. (B) Comparaison de deux potentiels statistiques (interaction du carbone de l'aspartate avec l'oxygène de la lysine) construits à partir d'ensembles de protéines liposolubles ou hydrosolubles.

Hormis le classement de modèles protéiques selon leur qualité, MyPMFs permet également d'étudier, par comparaison entre les potentiels statistiques, les règles qui régissent le positionnement relatif des acides aminés et qui sont spécifiques à chaque environnement cellulaire. Par exemple, le potentiel d'interaction des atomes de deux acides aminés chargés diffère selon qu'il est construit à partir de protéines hydrosolubles du cytosol ou bien liposolubles de la membrane (Figure 1B). Cette observation peut être attribuée à la différence de permittivité entre ces deux milieux. L'absence de pont disulfure au sein de la bicouche lipidique est également une observation facilement vérifiée grâce au programme MyPMFs et pourtant d'un grand intérêt car très peu documentée dans la littérature.

Enfin, si les performances des PMFs pour évaluer la qualité des modèles moléculaires ne souffrent aucune contestation depuis presque trois décennies, l'explication de ces performances fait, elle, débat dans la communauté, l'approximation de l'enthalpie libre étant jugée insatisfaisante. Ce travail fournit également un bref éclairage sur le lien entre l'efficacité de ces potentiels statistiques et une forme de raisonnement probabiliste appelée « cinématique des probabilités ».

Ces travaux ont permis le développement et la mise à la libre disposition de la communauté (code source ouvert) d'un logiciel utile à la modélisation des structures de protéines et à l'étude des interactions entre acides aminés. Les futurs développements prévoient l'extension aux interactions impliquant trois atomes ou plus, ainsi que l'intégration des potentiels interatomiques produits par MyPMFs dans des fonctions de scores composites dont les termes sont pondérés par apprentissage automatique.

#### Référence

"MyPMFs: a simple tool for creating statistical potentials to assess protein structural models"  
Guillaume Postic, Thomas Hamelryck, Jacques Chomilier, Dirk Stratmann  
Biochimie, 151, 37-41 (2018)

#### Contacts

guillaume.postic@sorbonne-universite.fr  
dirk.stratmann@sorbonne-universite.fr