

Nom – Prénom Chauvot de Beauchene
Laboratoire de rattachement LORIA
Intitulé du diplôme HDR Informatique

Titre de l'HDR Modélisation 3D des interactions protéine-ARN par assemblage de fragments

Abstract (français)

L'amarrage (*docking*) de l'ARNsb aux protéines nécessite un échantillonnage exhaustif et précis de ses conformations 3D, ce que sa très grande flexibilité rend difficile : l'ensemble des conformations est trop grand pour le docking classiques. Mon approche par par assemblage combinatoire de fragments permet de resserrer l'espace de recherche autour de conformations globales compatibles avec une liaison à la protéine cible. La séquence d'ARN est découpée en fragments chevauchants suffisamment petits pour que toutes leurs conformations soient représentées par un ensemble de prototypes de faible cardinalité. Ces ensembles sont amarrés sur la protéine, les poses de meilleur score sont conservées, et les chaînes de poses géométriquement compatibles sont assemblées en autant de modèles finaux. Cet assemblage consiste en la recherche, par programmation dynamique, des meilleurs chemins de longueur fixe dans le graphe de connectivité des poses pondérées de leur score de docking. Je présente d'abord cette approche et ses limites, puis décrit mes avancées sur différents aspects de cette approche globale : un outil de création de bibliothèques de fragments d'ARN et leur analyse statistique; un algorithme de clustering qui crée le plus petit ensemble de prototypes représentatif d'un ensemble fini de structure 3D; une approche hiérarchique *branch-and-bound* de sampling exhaustif de poses de fragment dans une poche protéique; le développement analytique d'un potentiel gros-grain protéine-ARNsb. Je présente enfin des variantes pour applications particulières : aux RNARecognition Motifs (RRM), aux boucles d'épingles d'ARN, aux ADNsb et aux glycsaminoglycans.

Abstract (anglais)

The docking of ssRNA on proteins requires an exhaustive and precise sampling of its 3D conformations, which its great flexibility makes difficult: the ensemble of conformations is too large for classical docking. My combinatorial fragment assembly approach narrows the search space to global conformations compatible with binding to the target protein. The RNA sequence is cut into overlapping fragments small enough that all their conformations are represented by a set of prototypes of low cardinality. These sets are docked to the protein, the best-scoring poses are retained, and chains of geometrically compatible poses are assembled into as many final models. This assembly consists in searching, by dynamic programming, for the best paths of fixed length in the connectivity graph of poses weighted by their docking score. I first present this approach and its limitations, then describe my progress on various aspects of this global approach: a tool for creating RNA fragment libraries and their statistical analysis; a clustering algorithm that creates the smallest set of prototypes representative of a finite set of 3D structures; a hierarchical branch-and-bound approach to exhaustive sampling of fragment poses in a protein pocket; the analytical development of a coarse-grained protein-ssRNA scoring function. Finally, I present variants for specific applications: to RNA-Recognition Motifs (RRMs), RNA hairpin loops, ssDNA and glycsaminoglycans.