A hybrid combinatorial method for docking single stranded RNA on proteins at the thermodynamic equilibrium
Chinmay Singhal, Yann Ponty, Isaure Chauvot de Beauchêne

To cite this version:
Chinmay Singhal, Yann Ponty, Isaure Chauvot de Beauchêne. A hybrid combinatorial method for docking single stranded RNA on proteins at the thermodynamic equilibrium. RECOMB 2018 - 22nd Annual International Conference on Research in Computational Molecular Biology, Apr 2018, Paris, France. hal-01925083

HAL Id: hal-01925083
https://inria.hal.science/hal-01925083
Submitted on 16 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A hybrid combinatorial method for docking ssRNA on proteins at the thermodynamic equilibrium

Chinmay Singhal 1, Yann Ponty 2, Isaure Chauvot de Beauchene 3
1 AMBio team, Inria Saclay 2 CNRS - LIX 3 CNRS - LORIA

Introduction

Protein-RNA complexes participate in many aspects of cell regulation. Their atomic structure description is crucial to understand the recognition mechanisms, but the experimental resolution of their structure is arduous. Computational docking methods aim at modeling a 3D assembly, by assembling structures of each isolated constituent. Yet for highly flexible objects like single-stranded RNA (ssRNA), the isolated structure of the whole molecule can adopt an ensemble of conformations too large to be experimentally solved or computationally modeled.

We recently proposed a fragment-based approach to model ssRNA local conformations and assemble them on the protein surface. By docking then re-assembling overlapping fragments, we could select a pool of most-connected poses that delineates the RNA-binding site (Fig. right). But the number of possible assemblies is beyond the reach of brute force approaches.

A. RNA fragment docking w. deepATTRACT

For the fragment assembly to succeed, all positions in the sequence must have been sampled correctly. We previously used the ATTRACT software [2] to dock RNA on the protein surface. Here we adapted ATTRACT for docking inside deep cavities. As we deal with a poly-U, one single docking of a UUU fragment was performed.

B. Fragments assembly by stochastic backtracking

We then search for chains of compatible docking poses with low total energy (approximating poses energy as additive). 10^9 docking poses have to be kept to ensure retrieving enough correct solutions at each sequence position.

We previously used the ATTRACT software [3] to dock RNA on the protein surface. Here we could select a pool of most-connected poses that delineates the RNA-binding site (Fig. right).

Conclusion

We present an improved method capable of modeling the ssRNA solely based on the protein structure and ssRNA sequence. The main improvements include:

A new docking protocol for docking RNA fragments inside deep pockets with systematic sampling.
A stochastic backtracking algorithm that can perform unbiased sampling from the connectivity graph of the docked fragments. This algorithm generates near-native RNA chains among 100,000 samples.
An efficient and effective filtering procedure to incorporate experimental knowledge on the protein-ssRNA system. The filters can be combined, leading to an enrichment of up to x63 after filtering.
As a first proof of principle, the method could model ab into a protein-bound ssRNA with a length of 10 nucleotides, an unprecedented length far beyond the reach of standard small-molecule docking programs.

Perspectives

Account for protein flexibility: Proteins also change conformation upon binding, albeit less drastically than RNAs. To avoid sampling many of the possible global conformations and use of all of them for docking, one can apply the protein the same principle of decoupled local sampling as for RNA.

Fold the RNA on the protein

From RNA sequence, predict a Boltzmann ensemble of secondary structures (SS), dock all fragment of possible structure at each position in sequence, then incorporate the 3D structure probabilistically in the protein-assembled structure.

Co-assemble w. protein conformations

Account for protein flexibility: Proteins also change conformation upon binding, albeit less drastically than RNAs. To avoid sampling many of the possible global conformations and use of all of them for docking, one can apply the protein the same principle of decoupled local sampling as for RNA.

Co-assembly of fragments libraries

Each “local” set of conformations can be used for docking, then the compatibility of adjacent conformations can be assessed together with the connectivity of RNA poses.

2 Martin Zacharias (2003) Proteins 1-75