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 is crucial to understanding the recognition mechanism, but the experimental resolution of their structure is often limited. Computational docking methods aim to model 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 pre-assembled overlapping fragments, we could select a pool of most-connected poses that delineates the RNA-binding site (Fig. right) [1]. 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). 109 docking poses have to be kept to ensure retaining enough correct solutions at each sequence position.

**Conclusion**

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

- A new docking protocol for docking RNA fragments inside deep pockets with systematic sampling. This shows improved sampling quality (lower RMSD) for the nucleotide fragments buried inside the binding pocket.
- 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 x163 after filtering.

As a proof of principle, we applied this method on a 12-mer poly-U ssRNA embedded in the deep cavity of an exonuclease.

**Perspectives**

**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 their possible global conformations and use all of them for docking, one can apply to the protein the same principle of decoupled approximating poses energy as additive).

**Fold the RNA on the protein**

From RNA sequence, predict a Boltzmann ensemble of structures, then incorporate the 3D structure probabilistically in the protein assembly.

**Combine probabilities**

Weigh the edges and nodes of the graph by:

- Contourmational energy
- Connectivity
- RMD ≤ 1 Å
- ≤ 0.5
- ≤ 0.5

**Domains orientation**

Disentangle the ensemble of possible pairwise orientations, then assess which are geometrically compatible.

**Loop conformations**

Enumerate local conformations inside domain, check compatibilities pairwise and with all others to ensemble docking programs.

---