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, Isaaure Chauvo 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 atomistic structural description is crucial to understand the recognition mechanism, but the experimental resolution of their structure is arduous. 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 assembly 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) [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.

ATTRACT

Search space : 10^6 states (position, orientation, conformer)
Energy mininimization in empirical force field
Select 10^4 poses of lowest energy

deepATTRACT

Disordered = w. conformations

B. Fragments assembly by stochastic backtracking

We then search for chains of compatible docking poses with low total energy (approximating poses energy as additive). 10^4 docking poses have to be kept to ensure retrieving enough correct solutions at each sequence position. => 10^2 possible assemblies !

How can we retrieve the most probable assemblies while avoiding a brute-force enumeration?

By an unbiased sampling of paths: For each position, at each position in sequence, the next pose is chosen among all poses with a probability given by the partition function Z, which defines the probability of each pose to participate in a downstream path. Z is pre-calculated by dynamic programming.

Conclusion

We present an improved method capable of modeling protein-bound ssRNAs 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. This shows improved sampling quality (lower RMSDs) 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 x162 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

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 conformational and use all of them for docking, one can apply to the protein the same principle of decoupled local sampling as for RNA.

Co-assembly phase : protein-ligand complexes reside domain, check compatible pathways and pick one as ensemble.

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 probability in the protein-assembly.

Combine probabilities

Weight the edges and nodes of the graph by :

\[ \text{Connectivity} = P_\text{SS} \times P_\text{SS} \times P_\text{SS} \]

RMSD criteria

\[ \text{RMSD} < 2.4 \]

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

Loop conformations : Explore local conformations inside domain, check compatible pathways and pick one as ensemble.

Appendix

<table>
<thead>
<tr>
<th>Nuc 4c</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>ab</th>
<th>ac</th>
<th>abc</th>
<th>RMD &lt; 5 Å</th>
<th>RMD &lt; 4 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuc 3c</td>
<td>10%</td>
<td>1%</td>
<td>0.1%</td>
<td>10%</td>
<td>10%</td>
<td>10%</td>
<td>RMD &lt; 5 Å</td>
<td>RMD &lt; 4 Å</td>
</tr>
<tr>
<td>Nuc 2c</td>
<td>1%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>RMD &lt; 5 Å</td>
<td>RMD &lt; 4 Å</td>
</tr>
<tr>
<td>Nuc 1c</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>RMD &lt; 5 Å</td>
<td>RMD &lt; 4 Å</td>
</tr>
</tbody>
</table>

Aromatic rings are known for establishing stacking interactions with RNA bases. One is present at the entrance of the protein pocket – we impose the 4 Å base to be at < 4 Å from the aromatic ring.

B. Mg^2+ ions are known for establishing ion-ionic interactions with RNA phosphate groups. One such ion is present in the bottom of the protein pocket – We impose the last phosphate to be at < 7 Å from the Mg^2+.

C. The single-stranded RNA as linear – nucleotides not in same fragment are at < 6 Å from each other.

Applying each filter separately leads to an enrichment of x 1.6 to x34 in correct solutions. Combining two filters leads to an enrichment of x 2.7 to x 563. Combining the three filters retains only one model, which is at 4.0 Å from the reference position (Fig above).

References

2 | Martin Zacharias (2003) Proteins