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 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 assemble them on the protein surface. By docking then re-assembling overlapping fragments, we could select a pool of most-connected poses that defines 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] todock the recognition mechanism 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.

1. ATTRACT

- Search space: 10^7 states (position, orientation, conformer)
- Energy minimization in empirical force field
- Select top 10 sets of lowest energy

2. deepATTRACT

- Detection of pocket regions
- Selection of points cloud centers
- Clustering fragment library to keep low representative

3. Local filters

- Sample orientations * 10^4 * few conformers
- Return low-energy poses

4. Search space

- Top 10^5 lowest energy conformers clustered w. 2 Å cutoff
- Keep low-energy poses, minimize further, keep 10^4 lowest energy

B. Fragments assembly by stochastic backtracking

We then search for chains of compatible docking poses with low total energy (approximating poses energy as additive). 10^8 docking poses have to be kept to ensure retrieving enough correct solutions at each sequence position.

10^8 possible assemblies !

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

By an unbiased sampling of paths: For each sample, 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.

1. Partition function Z

\[
Z_{\text{path}}(k) = \sum_{i=0}^{k} Z(i, j)
\]

2. Dynamic programming

\[
Z(k) = Z(k-1) + \exp(-\Delta E/k_BT)
\]

3. Backtracking algorithm

\[
Z(\text{path}) = \prod_{i} Z(i, j)
\]

C. Enrichment by combinable filters

The stochastic backtracking provided 100,000 models containing 0.3% correct solutions, and a best solution at 2 Å from the reference position (Xray).

The scoring function of ATTRACT is not precise enough to select the best models. Instead, we used some knowledge on the system to define geometric constraints used as filters:

1. Aromatic rings are known for establishing stacking interactions with RNA bases. One is present at the entrance of the protein - pocket - we impose the 4th base to be at < 3 Å from the aromatic ring.
2. Mg²⁺ 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 Mg²⁺
3. We consider the single-stranded RNA as linear – nucleotides not in same fragment are at > 3 Å from each other.

D. Results

Here, we present an improved method capable of modeling the ssRNA solely based on the protein structure and ssRNA sequence. Improvements include (A) a new docking protocol for systematic sampling in deep pockets, (B) a stochastic backtracking algorithm for unbiased sampling of assemblies, and (C) a combination of filters of the ensemble of sampled models based on biophysical characteristics of the binding site.

As a proof-of-principle, we applied this method on a 11-mer poly-U ssRNA inset into the deep cavity of an exonuclease

Perspectives

1. Account for protein flexibility
2. Sample multiple conformations
3. Achieve higher sampling

Co-assemble w. protein conformations

- Protein flexibility
- Diffusion trajectories
- Resample the trajectory

Values: 1 for contact and 0 for none