

# Fragment-based modeling of protein-bound ssRNA

Isaure Chauvot de Beauchêne, Sjoerd J de Vries, Martin Zacharias

## ▶ To cite this version:

Isaure Chauvot de Beauchêne, Sjoerd J de Vries, Martin Zacharias. Fragment-based modeling of protein-bound ssRNA. ECCB 2016: The 15th European Conference on Computational Biology, Sep 2016, Den Haag, Netherlands. 2016. hal-01573352

HAL Id: hal-01573352

https://hal.science/hal-01573352

Submitted on 9 Aug 2017

**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.

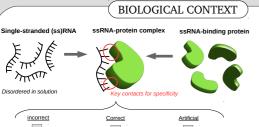


# Fragment-based modeling of protein-bound ssRNA



Isaure Chauvot de Beauchêne, Sjoerd J. De Vries, Martin Zacharias.

Technische Universität München (Germany)



Disease

Experimentaly known structures

SSRNA

Biological function

Transport of mRNA Maturation of mRNA Regulation of transla

Classical Docking

Sampling of relative

± ligand deformations) Scoring of the final poses

Ab initio Modeling

Modeling the RNA in situ,

12 DOF to explore per nucleotide

Drug

RATIONALE

Growth factor inhibitor
 Anti-virus (C-hepatitis)
 In vitro diagnosis

(i) understand its function or

(ii) modulate or create it, for medicine or biotechnology

Experimental methods to obtain such structure (X-ray, NMR) are costly, time-consuming or limited to some complexes.

Therefore, it often requires computational modeling methods. Such methods exist for RNA-protein complexes, but fail at modeling single-stranded RNA because of their flexibility.

Solving experimentally the

or RNA is easier than of a

On can use those isolate structures to model the complex via docking methods

structure of an isolate protein

But if the structure of the free

ligand is unknown, it has to be modelled from sequence.

For highly flexible objects like

numerous to be all modelled Our alternative approach consists in modeling RNA local conformations and

long ssRNA, possible

conformations are too

assemble them on the protein surface.

### **METHODS**

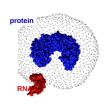
### Docking

1/~ 107 random starting states (position \* orientation \* conformation)

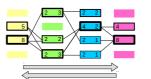
- 2/ Energy minimisation of bead-bead interactions in an empirical force field
- 3/ Elimination of redundant poses (converged on same local minima)
- 4/ Ranking of poses by score ( = pseudo-energy)

ATTRACT docking engin [1,2]

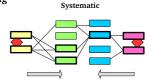
=> For each fragment : best pose at 1 - 3 Å from X-ray structure among ~105 - 106 poses





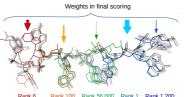


Forward – backward paths count => Selection of the most connected poses (i.e. most probably part of the correct chain)



Pruning from each anchoring contact => Enumeration of all possible chains

### Scoring



Chains are scored by the geometric mean of the ranking\* of the poses.

This enhances the weight of very well-ranked poses, to account for hot-spots\*\* in the RNA.

Score = (  $\prod$  rank(i))  $^{1/N}$ 

low score = good chain (hopefully)

- by pseudo-energy of protein-RNA intera-
- \*\* Fragments that bind with high energy (Often key parts for specific recognition)

# STRATEGY

Fewer correlations (constraints) than in double helix => combinatorial explosion for > 5-6 nucl. RNA adopt discrete local conformations (= rotamers) => they can be represented by a finit nur structural fragments



Divided into overlaping trimers

# Fragments library



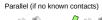
AUG UGG GGU GUC

All possible local conformations Extracted from structure databases

### Protein structure



### Fragments docking



Sequential (if known contacts)











### Assembling

spatially overlapping poses



### CONCLUSION

### Achievements

Our fragment-based approach to model protein-bound ssRNA proved effective to sample fragment poses at the surface of the protein. This permits to **predict the RNA** binding site with same sensitivity and higher specificity than all other binding-site prediction methods based on protein charge. protein structure [3].

We can predict the orientation of nucleotides binding to RNA-protein conserved contacts, in the most abundant RNA-binding domains of proteins (RRM and PUF) [4].

With those anchoring nucleotides, we could model bound ssRNA up to 12-nucleotides long, with a resolution comparable to X-ray structures [4].

### Perspectives

We considered so far that we know which part of our RNA is single-stranded (ss) and binds the protein. In many real-cases, the RNA is partially structured (e.g. double-stranded, ds) or parts of it oscillate between ss and ds state. Moreover, the ss part binds the protein by only some of the

Therefore, we will include RNA secondary structure prediction methods together with in vitro data (e.g. SHAPE) to evaluate the likelihood of protein-binding for each nucleotide in the RNA of interest, before or within the docking process.

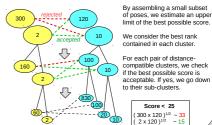
### Hierarchical clustering for efficient pruning

# By distance

By computing the distance between the centers of two clusters, we asses if they could conta overlapping poses

evaluated pairwise and so on

This spares pose pose comparisons for pairs of poses belonging to distant clusters.



By best ranks

limit of the best possible score We consider the best rank

contained in each cluster

For each pair of distance-compatible clusters, we check if the best possible score is acceptable. If yes, we go down to their sub-clusters.

( 300 x 120 )<sup>1/2</sup> ~ 33 ( 2 x 120 )<sup>1/2</sup> ~ 15

### RESULTS

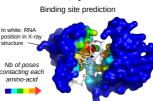
### Without predicted contacts

We blind-tested this approach on 2 complexes of

Filtering the poses by their chain-forming propensity (connectivity) enriches the pool in correct poses (RMSD <5Å) more effectively than the docking score alone :

Correct poses Total

Docking score	1% - 4 %	$6 . 10^5 - 1 . 10^6$
Connectivity	10% - 13 %	$3.10^3 - 8.10^3$
X-ray Docking More than 10% of poses are correct		
Binding site prediction		



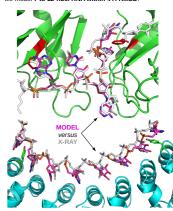
### With predicted contacts

We blind-tested this approach on 8 complexes of known structure [4].

We predicted the position and orientation of nucleotides establishing conserved contact, within 0.8 - 3.2  $\hbox{\normalfont\AA}$  RMSD from the real structure (average 1.4  $\hbox{\normalfont\AA}$ ).

From those contacts, we modelled **5 to 7-nucl RNA within 2 Å RMSD** from the real structure, for 7 out of 8 complexes, among 130 to 270 proposed models.

By prolonging the RNA chain beyond the predicted contacts, we model 7 to 12-nucl RNA within 4 Å RMSD.



Predictions (pink) versus experimental structure (X-ray, white) of RNA bound to either a RRM domain (green, pdbcode 2CVJ) or a PUF domain (cyan, pdb-code 3BX3)