New strategy for optimizing knowledge-based docking parameters: application to ssRNA-protein docking

Anna Kravchenko, Malika Smaïl-Tabbone, Isaure Chauvot de Beauchêne, Sjoerd Jacob de Vries

New strategy for optimizing knowledge-based docking parameters: application to ssRNA-protein docking

Anna KRAVCHENKO[1], Malika SMAIL-TABBONE[1], Isaure CHAUVOT DE BEAUCHENE[1] and Sjoerd Jacob DE VRIES[2]

[1] LORIA,campus scientific, 54000, Nancy, France.
[2] RPBS, 35 rue Hélène Brion, 75013, Paris, France

Corresponding Author: anna.kravchenko@loria.fr

Computational prediction of a 3D structure of a molecular complex, also known as *docking*, is essential in modern biological research. It can complement MD, provide working directions to experimentalists, etc. We are invested in fragment-based docking, specifically for the single-stranded RNA-protein complexes. Why ssRNA specifically? Generally speaking, molecular flexibility is a scourge of docking: it increases its complexity and decreases results' reliability. High flexibility leads to a near-infinite number of docking models, the processing of which is too expensive computationally. Hence, highly flexible ssRNA is a challenging target to work on.

A fragment-based docking approach was developed to tackle this high flexibility issue [1]. Its core idea is to split the ligand into overlapping fragments, and dock them onto the rigid receptor separately, assembling the fragments back into the whole ligand afterwards. For the full procedure to succeed, each fragment must return at least one correct pose (so-called near-native): this is the *sampling* problem. The poses are obtained by minimisation using a differentiable *energy function*. Then, before assembling, docked fragments must be filtered, keeping a high percentage of near-natives. Otherwise, the assembly task once again becomes too expensive computationally: this is the *scoring* problem. The filtration is done using a *scoring function*. We are working with the ATTRACT docking engine, where the same function is used both for sampling and scoring. It has the shape of a Lennard-Jones potential, and 2 parameters per atom type pair (1054 in total). The current parameters were obtained in 2010 by extraction of the statistical potentials from RNA-protein crystal structures and were optimized by a random Monte Carlo-like strategy [2].

These parameters were not initially tailored to ssRNA and their performance is not flawless. Our goal is to optimize docking parameters, improving both sampling and scoring performance. To achieve it, we created an up-to-date dataset of ssRNA-protein complexes and set up a novel histogram-based approach. For each pair of interacting atom types, we (a) convert the current energy function into a log-odds histogram of the expected occurrences of atom-atom distances (discretized into bins) in native/non-native poses, using the Boltzmann equation; (b) obtain the corresponding histogram on a benchmark-wide docking test, which corresponds to the residual error of the energy function; (c) sum the predicted and real histograms; (d) analytically fit the energy parameters to the resulting histogram. Repeat until convergence - until the residual histogram is flat.

Our newly created dataset of ssRNA-protein complexes is expected to be sufficient for optimization. The proposed approach for the ssRNA-protein fragment-based docking parameter optimization is expected to be more robust in terms of the fitness of the initial parameters compared to the previous approach. It has potential to benefit both the sampling and the scoring problems.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 813239.

References

[1] Isaure Chauvot de Beauchene, Sjoerd Jacob de Vries, Martin Zacharias. Fragment-based modelling of single stranded RNA bound to RNA recognition motif containing proteins. Nucleic Acids Res., 44(10):4565-4580, 2016.

[2] Piotr Setny, Martin Zacharias. A coarse-grained force field for Protein-RNA docking. Nucleic Acids Res., Vol. 39, No. 21, 9118–9129, 2011.