

Structural bioinformatics

FoldNucleus: web server for the prediction of RNA and protein folding nuclei from their 3D structures

Leonid B. Pereyaslavets, Igor V. Sokolovsky and Oxana V. Galzitskaya*

Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russia

*To whom correspondence should be addressed.

Associate Editor: Burkhard Rost

Received on March 23, 2015; revised on May 25, 2015; accepted on June 10, 2015

Abstract

Motivation: To gain insight into how biopolymers fold as quickly as they do, it is useful to determine which structural elements limit the rate of RNA/protein folding.

Summary: We have created a new web server, FoldNucleus. Using this server, it is possible to calculate the folding nucleus for RNA molecules with known 3D structures—including pseudoknots, tRNAs, hairpins and ribozymes—and for protein molecules with known 3D structures, as long as they are smaller than 200 amino acid residues. Researchers can determine and understand which elements of the structure limit the folding process for various types of RNAs and protein molecules. Experimental Φ values for 21 proteins can be found and compared with those determined by our method: http://bioinfo.protres.ru/resources/phi_values.htm.

Availability and implementation: <http://bioinfo.protres.ru/foldnucleus/>.

Contact: ogalzit@vega.protres.ru

1 Introduction

In theoretical studies of the structure of RNA molecules, the focus is often on the prediction of secondary and tertiary structures of RNA and on RNA folding kinetics, which can be described as a free energy landscape. Because the function of RNA depends on its conformation, which is analogous to the relationship between the function and folding structure of proteins, researchers have successfully applied methods developed for proteins, such as the Φ analysis (Matouschek *et al.*, 1990). In the folding process, the RNA strand, like a protein globule, passes through numerous intermediate states able to play a key role in the kinetics of the process. The problem of how a biopolymer chooses its native fold between a huge number of alternative folds is crucial for both RNA and protein molecules (Levinthal, 1968). Computer experiments with RNA- and protein-like model chains have shown that all can reach their lowest-energy fold without an exhaustive search over all possible folds (Thirumalai and Hyeon, 2005). It is obvious that not all nucleotides/residues play a decisive role in the RNA/protein folding. Thus, the knowledge of folding nuclei (a folding nucleus is the structured part of the molecule in a transient state) makes it

possible to reveal the structural elements that limit the rate of RNA/protein folding. The theoretical prediction of the RNA nucleotides/residues important for the formation of a folding nucleus would define the most probable kinetic pathway of folding. This, in turn, makes it possible to base RNA/protein-engineering efforts on the experimental detection of the nucleus of folding for an RNA/protein structure.

It has recently been demonstrated for tRNA^{Asp} that secondary and tertiary interactions are generated simultaneously (Wilkinson *et al.*, 2005) but not in a hierarchical process as was supposed for a long time. This allows us to adopt and apply an algorithm that was previously developed for the prediction of protein folding nuclei (Galzitskaya and Finkelstein, 1999) for RNA structures. The calculated Φ values for tRNA structures agree with the previously obtained experimental data (Pereyaslavets *et al.*, 2011). According to the experiment, the nucleotides of the D and T hairpin loops are the last to become involved in the tRNA tertiary structure (Maglott *et al.*, 1999). The purpose of this work is the creation of a server for the prediction of the structure of folding nuclei for proteins and RNA molecules, starting from known 3D structures.

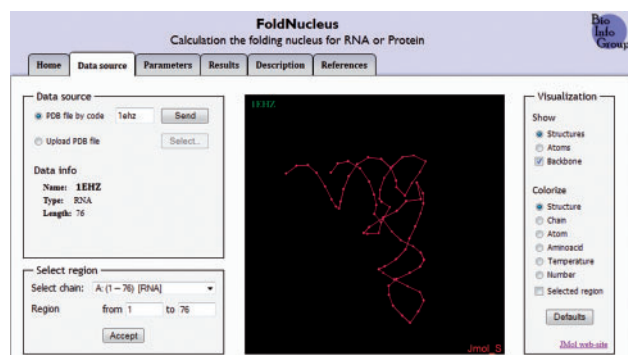


Fig. 1. Example of folding nuclei prediction (PDB:1ehz.ent) with the help of the FoldNucleus server

2 Algorithm

The algorithm, previously developed for the prediction of protein folding nuclei, has been successfully adopted and applied to RNA structures (Pereyaslavets *et al.*, 2011). The protein or RNA folding/unfolding process is modeled using the technique of dynamic programming as the reversible unfolding of its native structure. The construction of a network of protein/RNA unfolding pathways and the point of thermodynamic equilibrium, estimation of the free energy for proteins/RNAs and calculation of Φ values are described at http://bioinfo.protres.ru/foldnucleus/dscr/FoldNucleus_RNA.pdf and http://bioinfo.protres.ru/foldnucleus/dscr/FoldNucleus_Protein.pdf.

3 The FoldNucleus server

The FoldNucleus webserver is available at <http://bioinfo.protres.ru/foldnucleus/>. For the prediction of Φ values using this server, one should specify RNA or a protein structure for which the prediction is to be made. For this purpose, one should specify the corresponding PDB entry (in a standard 4-symbol format, e.g. the PDB entry is 1evv). For a protein or RNA in which more than one chain (or a protein-RNA complex) exists in the PDB file to be used, one should also specify which chain should be used (in the current version, the program simulates unfolding of a single-chain protein or RNA molecule).

For example (Fig. 1), if one writes B: in the corresponding field (region:) (default prediction), the server will use chain B of the protein or RNA. In addition, there is the option to make a prediction for a fragment of the chain: e.g. B:120-220 according to the numbering of the corresponding PDB entry, which means that the server will simulate only fragment 120–220 of chain B.

In the current version, the simulation is complete when the size of a chain link is one nucleotide for short RNA molecules (<30 nucleotide residues), two nucleotides for RNA less than 60 nucleotide residues and four (or three) nucleotides for larger RNAs. The server allows the user to choose the fragment length. In the default mode, the program uses the size of the fragment determined by the equation $N/50$, where N is the chain length. The cutoff values for atom-atom contacts are 6 Å for heavy atom-heavy atom contacts, 4 Å for hydrogen atom-hydrogen atom contacts and 5 Å for heavy atom-hydrogen atom contacts. For protein molecules, there is only one method, the method of contacts. For RNA molecules, the user can choose one of two methods to calculate the free energy: the method of contacts or the coarse-grained model adapted from Doholyan's model (Ding *et al.*, 2008). The predictions of Φ values are made upon consideration of mutations of all residues to glycine or alanine residues. In the case of RNA, all nucleotides will be deleted similar

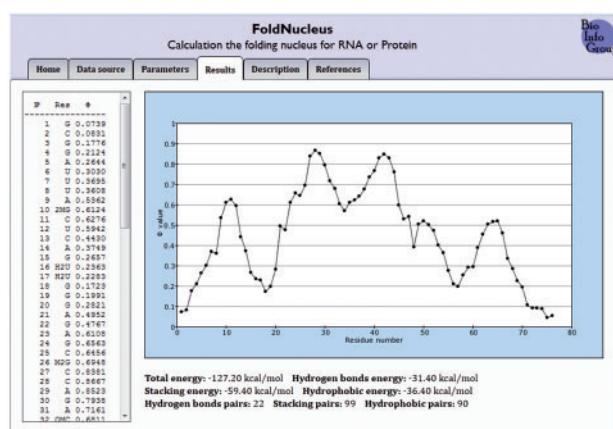


Fig. 2. Profile of Φ values for tRNA molecule (1ehz.ent)

to the substitutions for glycine in proteins. When the prediction is completed, the user obtains Φ values for each residue/nucleotide, along with a plot of these values (Fig. 2).

4 Implementation

The server was used to determine the folding nuclei for 21 proteins. A comparison of the calculation results with the experimental data shows that the model provides good Φ -value predictions for protein structures determined by X-ray analysis with consideration of hydrogen atoms and, less successfully, for structures determined by nuclear magnetic resonance (see Table 1 at http://bioinfo.protres.ru/foldnucleus/dscr/FoldNucleus_Protein.pdf). For protein structures, we consider all possible versions for the prediction of folding nuclei. They are as follows. The size of a link is five residues, and the link has the minimum size. Only the optimal folding nucleus and the entire ensemble of folding nuclei without hydrogen atoms and with hydrogen atoms are taken into account. All of these versions are listed in Table 1. One can see that the correlation is the best for the case when we consider the entire ensemble of folding nuclei with hydrogen atoms. When we consider the minimum size of a link, the correlation coefficient for 12 proteins is 0.56 ± 0.08 . In the case in which the size of a link is five residues, the correlation coefficient is 0.58 ± 0.09 .

The tertiary unfolding transition state of unmodified yeast tRNA^{Phe} has been studied (Maglott *et al.*, 1999). The authors concluded that the D/T-loop junction is formed last during the tRNA^{Phe} folding. The calculated Φ values for tRNA structures (~150 PDB structures) agree with the previously obtained experimental data. According to the experiment, the nucleotides of the D and T hairpin loops are the last to be involved in the tRNA tertiary structure. One of the advantages of our method is that it allows us to calculate the energy and folding nucleus in the structures with nontrivial RNA motifs, such as pseudoknots and tRNAs, if their spatial structures are in the PDB or NDB databases (see Tables 2 and 3 at http://bioinfo.protres.ru/foldnucleus/dscr/FoldNucleus_RNA.pdf).

Acknowledgements

We are grateful to Dr D. Reifsnnyder Hickey and T.B. Kuvshinkina for assistance in the preparation of the paper.

Funding

This study was supported by the Russian Science Foundation grant no. 14-14-00536 for OVG and Molecular and Cell Biology program (grant 01201353567) for IVS.

Conflict of interest: none declared.

References

- Ding, F. *et al.* (2008) Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA*, **14**, 1164–1173.
- Galitskaya, O.V. and Finkelstein A.V. (1999). A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl. Acad. Sci. USA*, **96**, 11299–11304.
- Levinthal, C. (1968) Are there pathways for protein folding? *J. Chem. Phys.*, **65**, 44–45.
- Maglott, E.J., *et al.* (1999) Probing the structure of an RNA tertiary unfolding transition state. *J. Am. Chem. Soc.*, **121**, 7461–7462.
- Matouschek, A. *et al.* (1990). Transient folding intermediates characterized by protein engineering. *Nature*, **346**, 440–445.
- Pereyaslavets, L.B. *et al.* (2011). Prediction of folding nuclei in tRNA molecules. *Biochemistry*, **76**, 236–244.
- Thirumalai, D. and Hyeon, C. (2005). RNA and protein folding: common themes and variations. *Biochemistry*, **44**, 4957–4970.
- Wilkinson, K.A. *et al.* (2005) RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *J. Am. Chem. Soc.*, **127**, 4659–4667.