From: ploscompbiol@plos.org
Subject: **PLoS Computational Biology Decision 08-PLCB-RA-0738**
Date: December 8, 2008 9:26:50 PM PST
To: yuany@stanford.edu

Dear Dr. Yao,

Thank you very much for submitting your manuscript "Topological Methods for Exploring Low-density States on Biomolecular Folding Pathways" for review by PLoS Computational Biology. As with all papers submitted to the journal, yours was fully evaluated by myself in consultation with the PLoS Computational Biology Editorial Team. In this case, your article was also evaluated by 2 independent reviewers. The reviewers appreciated the attention to an important topic, but they raised substantial concerns about the paper. Based on the reviews and editorial discussions, I regret that we will not be able to accept this manuscript for publication in the journal.

Specifically, we apologize for the delay as a result of having some difficulty in obtaining a second review. Subsequently, after discussion, we feel that this paper provides biological insights that are limited and in some question since they have not validated. As such the paper is outside of our current scope.

The reviews are attached, and we hope they may help you should you decide to revise the manuscript for submission elsewhere. I am sorry that we cannot be more positive on this occasion, but hope that you appreciate the reasons for this decision and that you will consider PLoS Computational Biology for other submissions in the future.

While we cannot consider your manuscript for publication in PLoS Computational Biology for the reasons outlined above, we very much appreciate your wish to present your work in an Open Access publication. We therefore want to alert you to an alternative that you may find attractive. We have recently launched a new publication called PLoS ONE (www.plosone.org). PLoS ONE is a swift, high-volume, efficient and economical system for the publication of peer-reviewed research in all areas of science and medicine; a unique publishing forum that exploits the full potential of the web to make the most of every piece of research. This journal published its first online issue in December 2006.

If you would like to submit your work to PLoS ONE we can transfer your files directly into PLoS ONE's manuscript handling system; please contact the PLoS ONE publication staff (plosone@plos.org) now, citing your manuscript tracking number. If you would like more information about submitting to PLoS ONE please either visit its website or email plosone@plos.org.

Thank you again for your support of PLoS Computational Biology and open access publishing. Please don't hesitate to get in touch if I can provide any further assistance.

Sincerely,

Eugene Shakhnovich
Associate Editor
PLoS Computational Biology

Philip E. Bourne
Editor-in-Chief
PLoS Computational Biology

Reviewer #1 (Remarks for the Author):

This paper addresses an important and challenging problem that is to identify transition states or intermediate states in biological processes such as protein and RNA folding, which are, in general, difficult to do experimentally due to their transient nature. The authors presented a new method, Mapper, based on topological data analysis to attack this problem, and then applied this method to an RNA-hairpin folding. The authors successfully identified two intermediate low-density clusters in the RNA hairpin folding pathway. The paper is well-written and key ideas clearly presented, thus I would recommend publishing it in PLoS Computational Biology, provided the following minor change can be addressed.

The only weakness I see is that these intermediates identified in the paper can not be easily verified. How do we know these intermediates are real? Could they be artifacts from simulations, topological mapping, or both? Can authors compare with some other methods such as ISOMAP for a cross-validation? Of course, this might be beyond the scope of the current paper, but some discussion and comments along these lines might be useful for readers, particularly the beginners in this field.

Reviewer #3 (Remarks for the Author):

The manuscript by Yao et.al. presents a computational method for analysis of folding/unfolding trajectories. More specifically, presented study

constitutes an application of a recently developed and published algorithm for analysis of multidimensional data to refolding trajectories of a short (12nt) RNA. While a new application of an existing algorithm to a well-studied problem may constitute a potential step forward in Computer Science, the limited scope of new biological insights obtained through this application makes me question the appropriateness of the manuscript for PLoS Computational Biology. There are also a number of serious technical issues that I outline below.

1. Authors represent a structure of 12nt RNA by the contact map of
formed base-pairs, i.e. by the contact map of its secondary structure.
The most complicated and novel step of proposed algorithm is to find the best representation of the multi-deminsional data as a graph which keeps topology of the space unperturbed. Authors claim that analysis of obatined contact maps is a challneging problem in R^55. I was perlexed by this claim, as the total number of possble secondary structures for a short RNA is very small and mapping from this space to a graph is trivial. RNA secondary structure without pseudoknots (and those are sterically impossible for a short RNA) was shown to map one-to-one to an expression with matched parenteces (a dot-parenthesis representation). The number of possible dot-parenthesis strings of length 12nt is easy to count and my back of the envelope estimate, assuming the central 4nts unpaired, gives
1+2choose8+2*4choose8+4*6choose8+8*1=289 possible secondary structures
for studied RNA, where NchooseK is a binomial coefficient. Thus the whole analysis of RNA refolding trajectories maps to analysis of paths on a graph with 289 nodes. Thus mapping from the space of structures to a graph, a problem that is accomplished by presented algorithms is rather trivial. I suggest authors to demonstrate scaling of their approach to a larger system, and take advantage of this studied toy example as in comparing obtained graphs with the exact one.

2. Critical for proposed algorithm is the choise of the filter
function, i.e. a function over the space of structures that is used 'to slice' the space. Authors claim that the choice they made 'approximates the free energy near the folded state', bringing no evidences to suport this claim. I cannot see how presented genomterical function choosen by the authors can approximate the free energy, as it is missing the energy term and the temperature. In fact the energy calcualted by molecular dynamics is not used in its construction at all.

3. The manuscript focuses on analysis of structures obtained from
SREMD simulations. It is unclear how kinetic information is being used in this geometrical analysis, i.e. how transitions between the structures, rates of these transtions and the dwell times are being used (if used at all).

4. The central result of the study is that folding and unfoding
trajecotries follow different routes and have different intermediates.
The careful reading shows that this intriguing proposal is based on the analysis where folding and unfolding trajectories have been treated very differently. The manuscript reads:
In the study of folding pathways, we take configurations from refolding events, and then weight heavily a neighborhood around the native states. However in the study of unfolding pathways, we sample from unfolding events, and focus on a neighborhood of the unfolded states.
It's important to demonstarte that obtained intermediates and transitions between them are not affected by the bias introduced by this procedure.

5. Authors provide comparison with K-means clustering technique,
which as they rightfuly point, has a tendency to idnetify spherical clusters. Recent progress in clustering analysis however yieled a number of methods that are robust to metrics used and can identify clsuters of any shape, e.g. a method call Super-Paramagnetic Clustering developed by Domany et al have been successful in a number of challenging applications to gene expression and protein structure classification. I suggest authors to compare their topological method with the SPC algorithm.