Poster I-26

Using Knowledge-Based Constraints To Optimise Comparative Models Of Protein Structure

Authors:
Braddon K. Lance (*Macquarie University, Department of Statistics*)
Graham R. Wood (*Macquarie University, Department of Statistics*)
Charlotte M. Deane (*University of Oxford, Department of Statistics*)

Short Abstract: Comparative modelling can be improved by better refinement of the inherited template backbone. To facilitate structure refinement, we generate local backbone conformations that are consistent with the observed structural variation by estimating the multivariate distribution of fragment atoms. We use this distribution in a simulated annealing algorithm for structure refinement.

Long Abstract:

Comparative modelling is currently the most successful method of protein structure prediction, with the potential to produce models that closely resemble the true protein structure. In comparative modelling, prediction of an unknown structure begins with the identification of homologous proteins that are likely to be similar to the unknown structure. Following sequence alignment, the prediction is made by attribution of atom coordinates from the homologues to the unknown structure. Current methods, however, often produce models that are little or no better than the closest template structure [1], and structure refinement usually worsens the prediction [2]. For structure predictions to be consistently and substantially better than their template structures, improved methods of structure refinement are needed [2-4]. The purpose of this work is to explore knowledge-based constraints on local backbone variation in comparatively modelled structures, with a view to developing a complete comparative modelling methodology that facilitates efficient optimisation of the backbone structure.

Fragment based modelling generates a structure prediction for an amino-acid sequence by assembly of the best matching fragments from its homologues [5]. When predicting a new structure in this way, some regions of the backbone may be considered exchangeable because the root mean squared deviation between them is small. In contrast, some regions of the backbone may not have a similar or topologically equivalent region in their homologues because of an insertion or deletion in the amino-acid sequence. Between these extremes, there may be regions of some homologues that are structurally similar, but not close enough that the fragment of one homologue could be approximated by another without introducing considerable error into the prediction. In this case, there is the possibility of exploiting our knowledge of the local structural variation within the homologues to guide the predictions we make for the local backbone structure of our unknown structure.

The major challenges in structure refinement are, firstly, that protein structures exist in a very high-dimensional space, meaning it is difficult to comprehensively sample conformations from structure space. Secondly, imperfect objective functions calculate rugged energy

landscapes with an abundance of local minima [7-8], meaning that even comprehensive sampling does not clearly indicate what directions of movement in structure space constitute a real improvement in the structure prediction.

A possible solution we explore in this poster is to sample backbone conformations in a way that follows the observed fragment variation, and thus samples are more likely to resemble the true fragment backbone compared to a random fragment. We achieve this by defining a multivariate distribution for C? coordinates of fragments in conformation space, using the variation in corresponding fragments of homologous proteins. To identify the set of corresponding fragments from homologues, we used HOMSTRAD [6], a database of structure-based alignments for homologous protein structure families. Using aligned residue pairs as an indicator of topological equivalence within fragments, we calculated the covariance matrix of the atom coordinates in the fragment C? backbone. The covariance structure of the distribution of fragments in conformation space is estimated by this sample covariance matrix. To find the optimal fragment structure in conformation space, we use a simulated annealing algorithm. The key components of the simulated annealing algorithm are the proposal distribution, the scoring function, and the acceptance probability. During optimisation, each new fragment shape (or 'state') is drawn from the proposal distribution, which we initially consider to be multivariate normal, with the covariance matrix defined as above. We use a standard scoring function and the Boltzmann acceptance probability. Preliminary results will be presented.

Literature Cited

1. Tress, M.L., et al., Assessment of predictions submitted for the CASP6 comparative modelling category. Proteins, 2005. In Press.
2. Valencia, A., Protein refinement: A new challenge for CASP in its 10th anniversary. Bioinformatics, 2005. 21(3): p. 277.
3. Contreras-Moreira, B., et al., Empirical limits for template-based protein structure prediction: the CASP5 example. FEBS, 2005. 579: p. 1203-1207.
4. Qian, B., A.R. Oritz, and D. Baker, Improvement of comparative model accuracy by free-energy optimisation along principal components of natural structural variation. PNAS, 2004. 101(43): p. 15346-15351.
5. Ginalski, K., Comparative modeling for protein structure prediction. Curr. Opin. Struc. Biol., 2006. 16: p. 1-6.
6. Mizuguchi, K., et al., HOMSTRAD: a database of protein structure alignments for homologous families. Prot. Sci., 1998. 7: p. 2469-2471.