The packing attribute of alpha helices as an advisor to evaluate proposed protein structures

Robert Fraser [i], Mireille Gomes [ii], and Janice Glasgow [ii]

[i] David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada, N2L3G1
[ii] School of Computing, Queen’s University, Kingston, ON, Canada, K7L3N6

The packing attribute of alpha helices as an advisor to evaluate proposed protein structures

Protein Structure Prediction

- Primary sequence
- Contact map
- Tertiary Structure

The protein chain is chain A from the Protein Data Bank (PDB) file 1A0A.pdb.

Contact map

The contact map is an N × M boolean matrix, entry D_{ij} indicates whether the distance from amino acid i to amino acid j is within some threshold. Contact maps can be predicted from 1st sequence, currently the accuracy is less than 66%.

Tertiary Structure

The 3rd structure of the protein is the coordinates of all of the atoms of the protein in three dimensional space. This includes secondary structures, which are local structures like the helices shown in this cartoon representation [2],

- the goal of this case-based reasoning (CBR) approach to prediction is to be able to predict the structure of a protein given the contact map.

Evaluation of Predictions

The evaluation step is the heart of the CBR framework.

Predicted structures are evaluated by a series of advisors which computationally rank the quality of each prediction based on different bio-chemical and physical properties of a protein.

A neural network assigns weights to the knowledge-based advisors so that the most useful advisors have the greatest influence on the final score (the weighted sum of individual advisor rankings) of the prediction.

- The more advisors that are present, the more robust the CBR system will be.
- Let’s look at pairs of α-helices...

α-Helix Pairs

α-helices are significant for the stability of a protein. The 3D structure of a protein is primarily determined by groups of secondary structures. α-helices are the most common type of secondary structure.

- Evaluating the correctness of the α-helix pairs in a predicted protein structure would be a very useful advisor, contributing to the bottom-up approach.
- Let’s consider finding α-helix pairs within a protein’s contact map...

α-Helix Pairs’ Contact Maps

a) The yellow box highlights the contact map for a pair of α-helices, which is extracted to produce b).
- The interface between the two helices is given by the smallest square enclosing all of the contacts.
- The interface pattern for the pair.
- α-helix pair packing

Contact Map Classes

The α-helix pair contact maps are divided into several classes: corner, edge, and central contacts. Examples of each are shown below.

- Corner contacts have contacts in a 2 × 2 corner of the map. Edge contacts have contacts along the perimeter of the map, but not in a corner. Central contacts have no contacts in the outer two rows or columns of the map [3].

Clustering contact maps

Contact maps in each class can be clustered, using each contact entry as a dimension.

Each corner map is transformed to a 15 × 15 map, where the origin is in the corner. Each map corresponds to a 225 character binary string.

Clustering Results (see [4] for details):

A database of clustered contact maps was created based on 1078 alpha helix pairs belonging to 171 all alpha helical proteins from the PDB.

References and Notes

[1] The protein chain is chain A from the Protein Data Bank (PDB) file 1A0A.pdb.

This research was supported by grants from the PRECARN Institute for Robotics and Intelligent Systems (IRIS), the Natural Science and Engineering Research Council (NSERC) and the CRA-W Canadian Distributed Mentor Project.

For additional information, please contact:
Mireille Gomes
School of Computing
Queen’s University
3ming2@qlink.queensu.ca

Contact Map Classes

Corner contacts have contacts in a 2 × 2 corner of the map. Edge contacts have contacts along the perimeter of the map, but not in a corner. Central contacts have no contacts in the outer two rows or columns of the map [3].

Clustering contact maps

Contact maps in each class can be clustered, using each contact entry as a dimension.

Each corner map is transformed to a 15 x 15 map, where the origin is in the corner. Each map corresponds to a 225 character binary string.

Clustering Results (see [4] for details):

A database of clustered contact maps was created based on 1078 alpha helix pairs belonging to 171 all alpha helical proteins from the PDB.

Contact Map Classes

Corner contacts have contacts in a 2 × 2 corner of the map. Edge contacts have contacts along the perimeter of the map, but not in a corner. Central contacts have no contacts in the outer two rows or columns of the map [3].

Clustering contact maps

Contact maps in each class can be clustered, using each contact entry as a dimension.

Each corner map is transformed to a 15 x 15 map, where the origin is in the corner. Each map corresponds to a 225 character binary string.

Clustering Results (see [4] for details):

A database of clustered contact maps was created based on 1078 alpha helix pairs belonging to 171 all alpha helical proteins from the PDB.

Contact Map Classes

Corner contacts have contacts in a 2 × 2 corner of the map. Edge contacts have contacts along the perimeter of the map, but not in a corner. Central contacts have no contacts in the outer two rows or columns of the map [3].

Clustering contact maps

Contact maps in each class can be clustered, using each contact entry as a dimension.

Each corner map is transformed to a 15 x 15 map, where the origin is in the corner. Each map corresponds to a 225 character binary string.

Clustering Results (see [4] for details):

A database of clustered contact maps was created based on 1078 alpha helix pairs belonging to 171 all alpha helical proteins from the PDB.

Contact Map Classes

Corner contacts have contacts in a 2 × 2 corner of the map. Edge contacts have contacts along the perimeter of the map, but not in a corner. Central contacts have no contacts in the outer two rows or columns of the map [3].

Clustering contact maps

Contact maps in each class can be clustered, using each contact entry as a dimension.

Each corner map is transformed to a 15 x 15 map, where the origin is in the corner. Each map corresponds to a 225 character binary string.

Clustering Results (see [4] for details):

A database of clustered contact maps was created based on 1078 alpha helix pairs belonging to 171 all alpha helical proteins from the PDB.

Contact Map Classes

Corner contacts have contacts in a 2 × 2 corner of the map. Edge contacts have contacts along the perimeter of the map, but not in a corner. Central contacts have no contacts in the outer two rows or columns of the map [3].

Clustering contact maps

Contact maps in each class can be clustered, using each contact entry as a dimension.

Each corner map is transformed to a 15 x 15 map, where the origin is in the corner. Each map corresponds to a 225 character binary string.

Clustering Results (see [4] for details):

A database of clustered contact maps was created based on 1078 alpha helix pairs belonging to 171 all alpha helical proteins from the PDB.

Contact Map Classes

Corner contacts have contacts in a 2 × 2 corner of the map. Edge contacts have contacts along the perimeter of the map, but not in a corner. Central contacts have no contacts in the outer two rows or columns of the map [3].

Clustering contact maps

Contact maps in each class can be clustered, using each contact entry as a dimension.

Each corner map is transformed to a 15 x 15 map, where the origin is in the corner. Each map corresponds to a 225 character binary string.

Clustering Results (see [4] for details):

A database of clustered contact maps was created based on 1078 alpha helix pairs belonging to 171 all alpha helical proteins from the PDB.