University of Massachusetts Amherst

Program Number: 374

Search Improvements Lead to Increased Accuracy in Protein Structure Prediction

TJ Brunette, Oliver Brock
Robotics & Biology Laboratory, University of Massachusetts Amherst

Robotics & Biology Laboratory

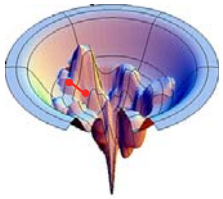
Protein Structure Prediction

The key to increase accuracy in protein structure prediction is search. Since exhaustive search is computationally intractable, our search method focuses exploration on the most promising regions of the search space.

Our two main hypothesis

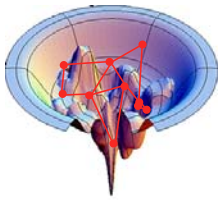
1. The selection of appropriate regions in which to focus exploration is the key to improving conformational space search.
2. This selection should be informed by as much information as possible.

Using Information To Guide Search



Monte Carlo retains information only from the previous search step. A single energy evaluation does not provide sufficient information to evaluate a region of conformation space.

Conformation Space Annealing (CSA) maintains a set of samples and the distances among them. This information is sufficient to prevent search from collapsing into a single minimum but does not accurately capture properties of the underlying landscape.



We believe domain information can help interpret information from samples. This information can then be used to direct resources toward promising regions.

Relevant Publications

Brunette, TJ and Oliver Brock. *Improving Protein Structure Prediction with Model-Based Search.* Bioinformatics 21(Suppl. 1):66-74, June 2005.Special Issue for the International Conference on Intelligent Systems for Molecular Biology (ISMB), Detroit, USA

Rohl, Strauss, Misura, and Baker. *Protein Structure Prediction Using Rosetta.* Methods in Enzymology. 383:66-93 2004

Funding

NIGMS 1R01GM076706 in the CST program; Predicting Protein Structure with Guided Conformation Space Search.

CSE 0551500 in the CRI program; Computational Biology Facility for Western Massachusetts.

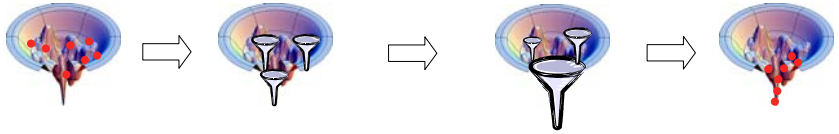
Model-Based Search

Acquire information by sampling

Select regions by interpreting the information to find spatial features shown as funnels

Acquire additional information; funnel size illustrates the likelihood the region contains the native state

Leverage information to focus resources on the most promising regions



1:Select Regions for Exploration

Benefits of interpreting groups of samples as energy wells

1. The shape of a well corresponds to actual physical properties of the landscape.
2. These properties can be used to guide computational resources.
3. This enables efficient acquisition of higher quality information

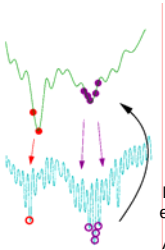
Finding energy wells
The distribution of samples captures features of the energy landscape

Starting from a low-energy sample, we examine neighboring conformations to find the ridge of the well associated with that sample. The depth and shape of the well reveal properties of the energy landscape that are used to guide search.

2: Extract Higher Quality Information

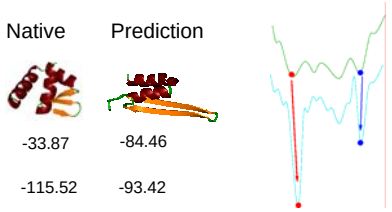
To improve our ability to identify appropriate regions of the search space we apply a highly accurate quality function to acquire additional information.

Our quality function acquires information about each funnel by using Monte Carlo trajectories on a subset of samples. The final state of each Monte Carlo trajectory is evaluated with the all-atom energy function.



Our quality function can help identify rank inversions and local minima

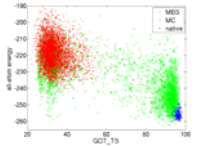
	Native	Prediction
Low-resolution energy function	-33.87	-84.46
All-atom energy function	-115.52	-93.42



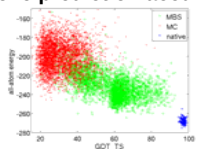
Results

We integrate Model-Based Search (MBS) with Rosetta and compare the efficiency with simulated annealing Monte Carlo. Both search methods search the same energy function with equivalent amounts of computational resources.

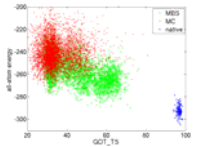
MBS uses information to improve prediction accuracy



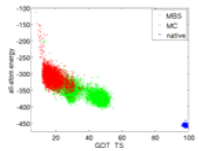
109aa EF Hand Parvalbumin PDB:1pva



123aa NTF2 M84E Mutant PDB:1jb2

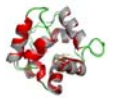


136aa Retinoic Acid Binding PDB:1cbr

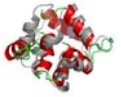


195aa Anthanilate Synthase PDB:1qdl

EF Hand Parvalbumin

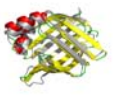


MBS 1.1 RMSD



MC 5.3 RMSD

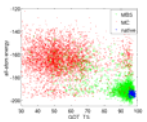
Retinoic Acid Binding Protein



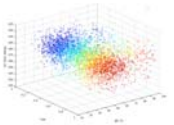
MBS 2.7 RMSD

MC 6.7 RMSD

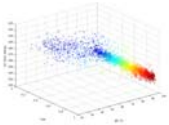
MBS helps choose the most accurate prediction



88aa IF3-C PDB 1tig

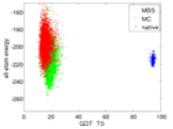


MC Sampling

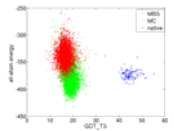


MBS Sampling

MBS identifies inaccuracies in the energy function



139aa Human CD40 Ligand PDB: 1aly



213aa Chloramphenicol Acetyltransferase:3cla