DERIVING PROTEIN STRUCTURE TOPOLOGY FROM THE HELIX SKELETON IN LOW RESOLUTION DENSITY MAP USING ROSETTA

YONGGANG LU, JING HE[1]
Dept. Computer Science, New Mexico State University
Las Cruces, NM 88003-8001, USA

CHARLIE E. M. STRAUSS
Bioscience Division, M888, Los Alamos National Laboratory
Los Alamos, NM 87545, USA

Electron cryo-microscopy (cryo-EM) is an experimental technique to determine the 3-dimensional structure for large protein complexes. Currently this technique is able to generate protein density maps at 6 to 9 Å resolution. Although secondary structures such as α-helix and β-sheet can be visualized from these maps, there is no mature approach to deduce their tertiary topology, the linear order of the secondary structures on the sequence. The problem is challenging because given N secondary structure elements, the number of possible orders is $(2^N)*N!$. We have developed a method to predict the topology of the secondary structures using *ab initio* structure prediction. The Rosetta structure prediction algorithm was used to make purely sequence based structure predictions for the protein. We produced 1000 of these *ab initio* models, and then screened the models produced by Rosetta for agreement with the helix skeleton derived from the density map. The method was benchmarked on 60 mainly alpha helical proteins, finding that for about 3/4 of all the proteins, the majority of the helices in the skeleton were correctly assigned by one of the top 10 suggested topologies from the method, while for about 1/3 of all the proteins the best topology assignment without errors was ranked the first. This approach also provides an estimate of the sequence alignment of the skeleton. For most of those true-positive assignments, the alignment was accurate to within +/- 2 amino acids in the sequence.

1. Introduction

Electron cryo-microscopy (cryo-EM) is an attractive method for structure determination because it can work with proteins that are poorly soluble or otherwise fail to crystallize, and is amenable to the structure determination of large protein complexes as well [1-7]. Although the cryo-EM can generate structures in the form of electron density maps at 6 to 9 Å resolution, it is currently not sufficient to determine the atomic structure directly since the side-chains cannot be resolved at this low-to-intermediate resolution [7, 8]. However, the location of secondary structures (SS), such as helices and β-sheets, can be visually and computationally identified [8-12]. It has been an emerging question about how to combine the low resolution density map with structure prediction techniques in order to derive the 3-dimensional structure of the protein [7, 8, 13]. The identified SS are often the major components of a protein and they form its skeleton (Fig. 1). Although the

This work is supported by LANL-NMSU MOU, NSF-HRD-0420407, LANL-LDRD-DR and DOE: Genomics: GTL Carbon sequestration.

[1] Email address of the corresponding author: jinghe@cs.nmsu.edu

skeleton contains the geometrical location of SS, it does not provide the information about where the SS are aligned with the protein sequence. The order of the SS with respect to the sequence, the topology, is also unknown. In this paper, we used helix skeleton, which is composed of the helices computationally identified by our software HelixTracer [12]. Given a protein density map at 6-10 Å resolution, HelixTracer can output the locations of helices represented by their central axial lines which can be potentially curved (Fig. 1).

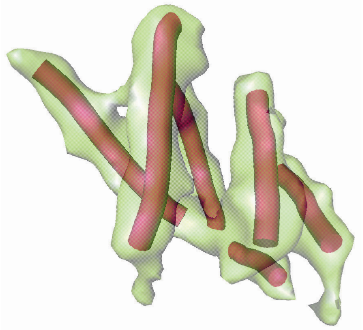

Fig. 1. Surface representation of the simulated density map (green) and the helix skeleton (purple cylinder-like sticks) found by HelixTracer for protein 1abv

Rosetta is one of the most successful *ab initio* structure prediction methods [14-18]. Unlike comparative modeling, *ab initio* methods do not require a structural homology to start with. For small protein domains, Rosetta is frequently able to produce low resolution models with correct topologies spanning the majority of the protein sequence. Previous work has shown that Rosetta is useful in refining NMR data [15, 18].

We have developed a method to derive the topology of a protein by combining its helix skeleton information with the predicted models obtained from Rosetta. Our method involves two components: MatchHelices and consensus analysis.

There were two impetuses to develop MatchHelices. First, we were not aware of an existing efficient approach to matching predicted model structures to a skeleton without any sequence information. Second, it is a waypoint towards fully integrating density maps into constrained *ab initio* modeling that rapidly reduces the search space. An alternative approach for comparison is used in “Foldhunter” from EMAN [9, 19] which uses correlation of density maps to align the skeleton with a structure. This approach is slow because of the grid search for translation and rotation parameters. Since MatchHelices only searches through the possible orientations suggested by the helix skeleton, the computation is significantly less. For a typical alignment between two structures of 100 amino acids, Foldhunter needs several minutes while MatchHelices takes a few seconds on a 3GHz machine.

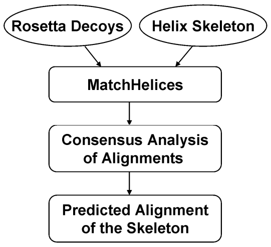
Rosetta Decoys
Helix Skeleton
MatchHelices
Consensus Analysis of Alignments
Predicted Alignment of the Skeleton

Fig. 2. The overall approach

2. Methods

2.1 The Overall Approach

Given a protein sequence, Rosetta can generate protein-like conformations known as decoys. The decoys are predicted possible conformations of the protein. Different decoys may have quite different conformations. The idea is to use helix skeleton to group the decoys and to derive the topology from each group. The overall procedure is shown in Fig. 2. It was tested on 60 mainly alpha-helical single domain proteins ranging in size of 50 to 150 residues. For each protein, Rosetta was used to generate 1000 decoys, and "pdb2mrc" was used to generate its density map at 8Å resolution [19]. Then the helix skeleton was identified from the density map by HelixTracer [12], and all the decoys were aligned with the skeleton by MatchHelices. Finally a consensus analysis was employed to identify the ten most popular alignments of the skeleton with the sequence.

2.2 MatchHelices

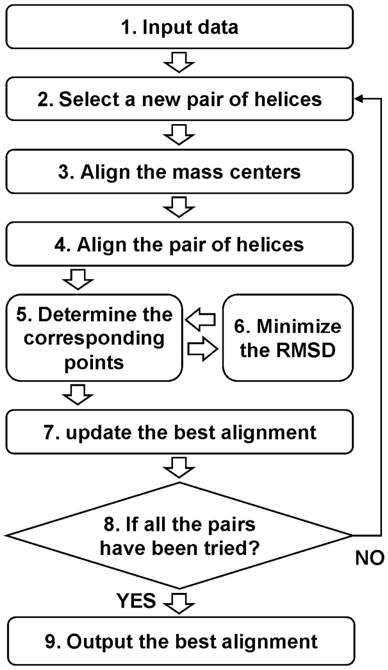
1. Input data
2. Select a new pair of helices
3. Align the mass centers
4. Align the pair of helices
5. Determine the corresponding points
6. Minimize the RMSD
7. update the best alignment
8. If all the pairs have been tried?
NO
YES
9. Output the best alignment

Fig. 3. The flowchart of MatchHelices

MatchHelices is a method to align a decoy with a helix skeleton (Fig. 3). The input is composed of a decoy and the helix skeleton found by HelixTracer (step 1). MatchHelices is a greedy and iterative method which first constructs seed alignment and then refines the alignment. A seed alignment is an initial trial of alignment that satisfies the following two criteria. The first requires the alignment of the two mass centers, one from the decoy and the other from the density map (step 3). The second requires that a pair of helices, one from the decoy and the other from the skeleton (selected in step 2), is positioned in a way so that the two helices are as close as possible while keeping the mass centers aligned. The second requirement is satisfied by a rotation of the decoy model around its mass center to maximally superimpose the two helices in a pair (step 4). The seed alignment is then refined by allowing certain level of mismatch between the mass centers. Since the seed alignment roughly positions the two helices in a pair, the corresponding points can be assigned between the two helices. If the C_α atoms of the helix in the decoy are within 5 Å distance from the helix axis in the skeleton, they are assigned to the nearest corresponding points on the skeleton (step 5). By doing this, a set of corresponding points are determined between the two helices in a pair. During the refinement step, the decoy is rotated and translated to minimize the RMS deviation between the corresponding points of a pair of helices (step 6). The step 5 and step 6 are repeated iteratively finding atoms within the cut-off and re-superposition to convergence. In step 7, the alignment score is calculated and the best alignment is updated. The alignment score is an ad hoc combination of the number of overlapped C-alpha atoms and the number of helices matched. The process (step 2 to step 7) is repeated for all the permutations of helix pairs and their relative directionalities to determine the best alignment between the decoy and the skeleton. In case of ambiguity, in which different parts of a decoy helix overlap two or more skeleton helices, the assignment is made according to the center residue of the helical segment of the model. Two example decoys aligned to the skeleton by MatchHelices are shown in Fig. 4.

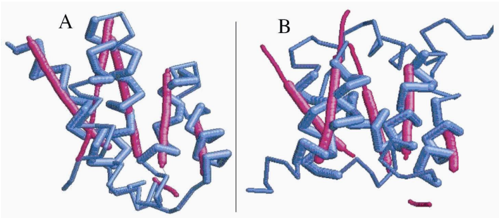

Fig. 4. Examples of MatchHelices Results: Rosetta decoy backbone (blue) and it's superposition with the helix skeleton (red) for protein 1abv. The thick regions show where the superposition criteria are satisfied. A. Corresponds to one of the decoys in first ranked topology prediction (a good prediction) in Fig. 6C, and B. corresponds to second ranked (a poor prediction) in Fig. 6C.

Since we are hunting for partial alignments over subsets of the skeleton, it is in principle possible that there could be multiple alignments of the skeleton to a given decoy that are equally good. In practice, several attributes of our approach seem to avoid this issue. First, because we require the decoy center of mass to be concentric with the density center, this forces an overall overlap beyond just the helices we are pairing, and breaks most degenerate cases. Second, HelixTracer produces skeletons with curved segments, not straight lines, and this too breaks the degeneracy. Third, we simply discard skeleton matches that are insufficiently complex (e.g. if only a single segment is matched while there are more than two helix segments in the skeleton). Lastly, even if this does occur in a particular decoy, we use many decoys and form a consensus.

2.3 Consensus Analysis

After all the decoys have been aligned with the helix skeleton, they are clustered based on where the skeleton helices are aligned on the protein sequence (Fig. 5). We used the secondary structure locations ("estimated helix positions") generated by Rosetta as a guide to assist the clustering process. For each residue, if more than half of the Rosetta decoys assign helix secondary structure to this residue, the residue is labeled to be within an estimated helix (indicated by H's below the protein sequence in Fig. 5). Within the regions of the sequence where the estimated helices reside, the nature of the alignment is examined. Two alignments are grouped into the same cluster if the skeleton helices they align roughly have the same location on the sequence. Particularly, each center amino acid of the skeleton helix has to be within the corresponding "estimated" helix. Therefore, the decoys in the same cluster are those with the same number of aligned skeleton helices. Moreover, their order, their directions on the sequence, and the corresponding "estimated" helices must be the same. For example, in Fig. 5, "Decoy A", "Decoy B" and "Decoy C" are grouped into one cluster while "Decoy E" and "Decoy F" are grouped into another cluster. "Decoy D" has only 3 skeleton helices aligned, so it

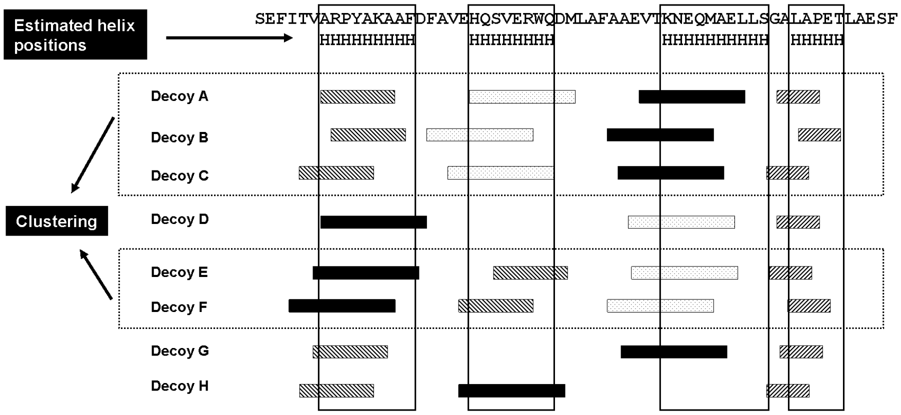
Estimated helix positions
SEFITVARPYAKAAFDFAVEHQSVERWQDMLAFAAEVTKNEQMAELLSGALAPETLAESF
HHHHHHHHH HHHHHHHH HHHHHHHHHH HHHHH
Decoy A
Decoy B
Decoy C
Clustering
Decoy D
Decoy E
Decoy F
Decoy G
Decoy H

Fig. 5. Illustration of the clustering process of the decoy alignments. Horizontal bars following the label "Decoy A"…"Decoy H" are the alignment results of different decoys. Different patterns of the bars represent different helices in the skeleton.

cannot be grouped into the same cluster with "Decoy E" and "Decoy F". Although "Decoy G" and "Decoy H" have the same set of skeleton helices aligned to the sequence, they cannot be grouped together since the center of the skeleton helix shown as the black bar is aligned to two different estimated helices in the two alignments. After the clustering, the resulting clusters are ranked by the number of decoys which they contain.

3. Results and Discussion

Fig. 6 shows the top ranked four topologies for protein 1abv. The top ranked topology correctly predicted the order of the six skeleton helices identified by HelixTracer (Fig. 6C). Besides the correct topology, our method also roughly aligned the skeleton helices to their correct locations on the sequence for the top ranked result (Fig. 6C). Each topology is derived from a cluster of decoys (Fig. 6C). The two decoys fitted to the skeleton of protein 1abv in Fig. 4 are the members of the two clusters corresponding to the first and the second topologies diagrammed in Fig. 6C. It can be seen in Fig. 6C that the topology inferred from the alignment in Fig. 4A is the correct one and the topology inferred from Fig. 4B is incorrect.

In the case of 1abv, HelixTracer correctly identified all of the six helices. But this is not always the case. Sometimes it misses or over-predicts some of the helices. Therefore, there could be an error in the result of HelixTracer. Similarly, there is often an error in the predicted model of Rosetta. However, the error sometimes can be partially compensated in the consensus analysis step, which will be shown later in this section.

We tested our method on 60 mainly alpha helical proteins. In Table 1, we only listed 44 of them. These 44 proteins have majority of the skeleton helices correctly assigned by one of the top 10 predicted topologies, judging by the column "Correct TH" and the column "Assigned Helices" (Table 1). From the last column "Alignment Offsets", we can see that for most of the alignments with the sequence, the offsets of the centers are within

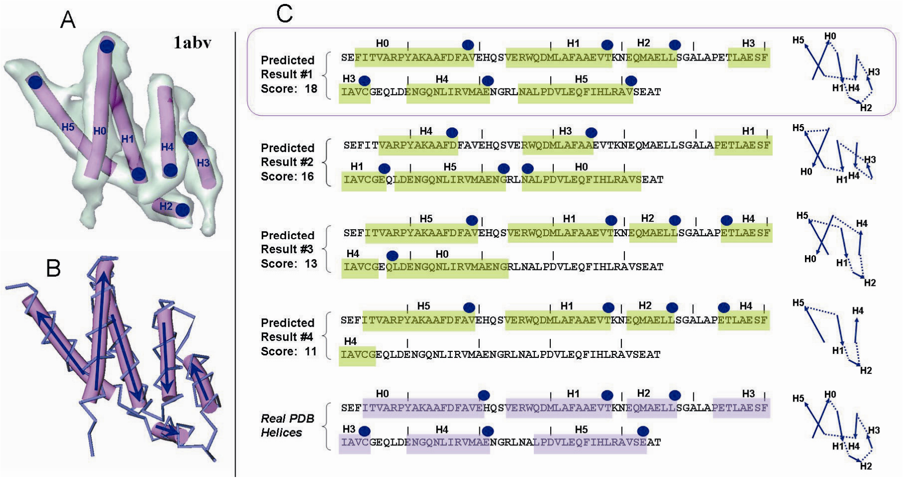
A
1abv
H5
H0
H1
H4
H3
H2
B
C
Predicted Result #1 Score: 18
SEFITVARPYAKAAFDFAVEHQSVERWQDMLAFAAEVTKNEQMAELLSGALAPETLAESF
IAVCGEQLDENGQNLIRVMAENGRLNALPDVLEQFIHLRAVSEAT
H0 H1 H2 H3 H4 H5
Predicted Result #2 Score: 16
Predicted Result #3 Score: 13
Predicted Result #4 Score: 11
Real PDB Helices

Fig. 6. Predicted sequence alignment of the skeleton for protein 1abv.
A: Simulated density of protein 1abv at 8Å resolution, with inset helix skeleton (purple sticks).
B: The helix skeleton (purple sticks) overlapped on the backbone of the protein from the PDB.
C: Top 4 predicted topologies and the true topology of 1abv. Green: predicted sequence alignments of the helices. Purple: the true sequence alignment of the helices. Right: diagrams of the topologies. Blue dots indicate true C-termini of helices labeled in A and C.

+/-2 residues. From the column "Rank", it is noticeable that for 17 proteins, about 1/3 of the total proteins tested, the best assignment is ranked the first. Column "Cluster Size" lists the number of the decoys in the cluster that was used to produce the correct assignment. In general, a larger value in the "Cluster Size" column makes the result more reliable. However, for some proteins, the "Cluster Size" value is very small and it still produces good results. It can also be noticed that when the value of the "Total Clusters" is larger, the value of the "Cluster Size" is usually smaller. This is reasonable, because the number of decoys is the same for each protein, and fewer decoys will usually be in a cluster if more clusters are formed.

The other 16 proteins (with PDB ID 1ag2_, 1bgf_, 1bo9A, 1d8bA, 1eo0A, 1ey1A, 1jhgA, 1jli_, 1jw2A, 1k04A, 1klxA, 1koyA, 1l9lA, 1lre_, 1lriA and 1qqvA) failed to generate correct topologies from the top 10 predictions. These 16 proteins are not listed in the table. Although for the 16 proteins we didn't get satisfied results in the top ten topologies, we noticed that a subset of the assignments of the skeleton helices may still be correct. They were not qualified to be "good results" because we used a very strict criterion that all of the assignments in a result including the topological ordering of the helical segments and their directionalities should be correct. So it is possible that useful information still can be acquired from the top 10 results for the 16 proteins.

In some cases only a subset of the skeleton helices were aligned (see Fig. 6C), thus there could be more than one result having correct topologies but using different helical subsets within the top 10 predictions. In such cases, we only listed the one that has the most skeleton helices involved in the assignment.

Table 1. Results of the 44 out of 60 proteins which have the majority of the helices in the skeleton correctly assigned by one of the top 10 suggested topologies

Protein ID	Total Residues	Possible Clusters[a]	Total Clusters[b]	Rank[c]	Cluster Size[d]	PDB Helices[e]	Total TH[f]	Correct TH[g]	**Assigned Helices[h]**	Alignment Offsets[i]
1a43_	72	62700	315	3rd	23	6	5	5	**4**	{0,2.5,0,0.5 }
1a6s_	87	360	213	9th	19	4	3	3	**3**	{0,2.5,1.5 }
1a7w_	68	138	80	3rd	56	3	3	3	**2**	{0,0.5 }
1abv_	105	291792	611	1st	18	6	6	6	**6**	{1,0,0,1,0,2 }
1bby_	69	138	71	1st	346	3	3	3	**3**	{0.5,0.5,0 }
1bkrA	108	378640	607	2nd	5	8	5	5	**3**	{1,0.5,0 }
1c3yA	108	62700	670	1st	12	6	5	5	**4**	{0.5,0.5,2,0 }
1daqA	71	36	18	4th	95	3	2	2	**1**	{1 }
1dgnA	89	1044204	578	8th	8	7	6	6	**4**	{1,0,0,0.5 }
1dk8A	147	1432180	725	3rd	5	10	5	5	**3**	{1,0,0.5 }
1dlwA	116	39040	334	1st	34	8	4	4	**4**	{0.5,0,0.5,1.5 }
1doqA	69	4360	87	1st	234	5	4	4	**4**	{0,0,1,1 }
1dp3A	55	138	35	1st	194	3	3	3	**2**	{1,0 }
1du2A	76	64	61	7th	36	4	2	2	**2**	{0,1 }
1dxsA	57	4360	234	5th	21	5	4	4	**4**	{0,1,0,2.5 }
1ef4A	55	138	47	1st	518	3	3	3	**3**	{0.5,0,2.5 }
1eyhA	144	1.69E+08	823	1st	15	10	7	7	**4**	{0,0.5,1,4.5 }
1f68A	103	62700	427	2nd	44	6	5	5	**3**	{0.5,0.5,0.5 }
1f6vA	91	1356	235	7th	21	6	3	3	**2**	{0.5,2.5 }
1fe5A	118	750	182	7th	16	5	3	3	**2**	{1.5,2 }
1g03A	134	291792	635	5th	8	6	6	6	**3**	{2,1,5.5 }
1g7dA	106	19090	338	2nd	53	5	5	5	**5**	{1,1,1.5,0,1.5 }
1gab_	53	138	116	4th	57	3	3	3	**3**	{1.5,0.5,0.5 }
1gd6A	119	166390	347	1st	14	7	5	4	**3**	{0.5,0.5,1 }
1gxgA	85	360	149	2nd	62	4	3	3	**3**	{1,1,0.5 }
1gxqA	105	1472	95	1st	68	4	4	4	**3**	{0.5,0.5,1 }
1hb6A	86	4360	235	1st	35	5	4	4	**4**	{1,1,0,0.5 }
1hbkA	89	1472	218	3rd	17	4	4	4	**4**	{0,1,0,0.5 }
1hdj_	77	360	76	1st	532	4	3	3	**3**	{1.5,1,2 }
1hp8_	68	360	57	1st	253	4	3	3	**2**	{1.5,1.5 }
1ig6A	107	166390	517	7th	5	7	5	5	**5**	{1,0,2,2,9 }
1iizA	120	62700	313	1st	36	6	5	3	**3**	{0.5,1.5,2 }
1iygA	133	1044204	501	6th	18	7	6	5	**4**	{1,0,0,0.5 }
1ji8A	111	166390	316	7th	17	7	5	5	**5**	{1,2,2,0,0.5 }
1kr7A	110	1044204	602	6th	9	7	6	5	**4**	{0.5,0,0,0.5 }
1myo_	118	19748176	816	2nd	6	8	7	7	**4**	{0,0,1,1 }
1ngr_	85	1044204	752	8th	4	6	7	6	**3**	{0,1,0.5 }
1nkl_	78	19090	303	6th	17	5	5	5	**3**	{0.5,0.5,4 }
1pru_	56	138	62	1st	320	3	3	3	**3**	{0.5,0,0 }
1utg_	70	1472	111	1st	79	4	4	4	**4**	{0,1,0.5,1 }
2asr_	142	19090	447	7th	4	5	5	5	**3**	{3,0.5,0.5 }
2end_	137	138	204	8th	14	3	3	3	**2**	{1,5 }
2lisA	131	1044204	695	7th	3	6	7	5	**3**	{2.5,0.5,2 }
2mhr_	118	4360	258	1st	84	5	4	4	**4**	{4,0.5,1,0.5 }

[a] The total number of all possible clusters in the search space.
[b] The total number of clusters produced in the consensus analysis step.
[c] The rank of the cluster from which the correct topology assignment was produced.
[d] The size (the number of the decoy members in the cluster) of the cluster from which the correct topology assignment was produced.
[e] The total number of helices in the crystal structure.
[f] The total number of helices (in the skeleton) identified by HelixTracer.
[g] The number of correctly identified helices (in the skeleton) by HelixTracer.
[h] The number of helices (in the skeleton) assigned in the correct topology assignment.
[i] The offsets of the centers of helices (in the skeleton) from the actual positions in the alignment corresponding to the correct topology assignment. The offsets are separated by commas.

The column “Possible Clusters” lists the total number of all possible clusters in the search space, N_P, which can be calculated by:

$$N_p = \sum_{k=1}^{\min(Ht,Hp)} C_{Ht}^{k} C_{Hp}^{k} \left(k! \times 2^{k} \right)$$

in which Ht is the total number of helices identified by HelixTracer, and Hp is the total number of helices in the crystal structure.

It can be seen from the table that for protein “1iizA” with 6 helices, HelixTracer only correctly identified 3 of them, with 3 helices missed and 2 helices over-predicted. With so many errors contained in the input, the correct topology was still found and it was ranked the first in the final results. Other examples for which HelixTracer over-predicted helices include protein “igd6A”, “1iygA”, “1kr7A”, “1ngr_” and “2lisA”. So our method can sometimes compensate the errors produced in the skeleton identification step.

4. Conclusion

This work is a preliminary study to see if we can quickly collapse the factorial complexity of the topology assignment and sequence alignment problem to a small number of possibilities that include an assured true positive. The results showed that our method was capable of assign majorities of the helices in the skeleton correctly within top 10 assignments for most of the proteins tested. For about 1/3 of all the proteins, the best result with the correct topology was ranked the first. Our method also showed robustness in working with bad inputs, namely the false positive and false negative identifications of the true helices by HelixTracer. The predicted topologies for the helix skeleton can be very helpful for structure determination with cryo-EM method. And it also can become the basis of a more careful constrained search using Rosetta.

Previously Rosetta has been used to refine NMR data [15, 18]. Here we are using it only as a consensus screen and we are not (as yet) incorporating the cryo-EM data into the prediction algorithm as restraints as it was done for NMR. Our previous work with NMR data showed that only a few constraints are needed to achieve very high accuracy, however when false positive constraints are included in a constraint set, the prediction quality rapidly deteriorates. Thus our screening protocol was biased away from making complete assignments of all the SS elements, and towards predictions with few false positives over the majority of the SS elements at the top of its ranked list.

References

1. Chiu, W., M. L. Baker, W. Jiang, and Z. H. Zhou. 2002. Deriving folds of macromolecular complexes through electron cryomicroscopy and bioinformatics approaches. Curr Opin Struct Biol 12: 263-9.
2. Topf, M., and A. Sali. 2005. Combining electron microscopy and comparative protein structure modeling. Curr Opin Struct Biol 15: 578-85.

3. Yonekura, K., S. Maki-Yonekura, and K. Namba. 2005. Building the atomic model for the bacterial flagellar filament by electron cryomicroscopy and image analysis. Structure 13: 407-12.
4. Zhou, Z. H., M. Dougherty, J. Jakana, J. He, F. J. Rixon, and W. Chiu. 2000. Seeing the herpesvirus capsid at 8.5 A. Science 288: 877-80.
5. Zhou, Z. H., M. L. Baker, W. Jiang, M. Dougherty, J. Jakana, G. Dong, G. Lu, and W. Chiu. 2001. Electron cryomicroscopy and bioinformatics suggest protein fold models for rice dwarf virus. Nat Struct Biol 8: 868-73.
6. Chiu, W., M. L. Baker, W. Jiang, M. Dougherty, and M. F. Schmid. 2005. Electron cryomicroscopy of biological machines at subnanometer resolution. Structure 13: 363-72.
7. Topf, M., M. L. Baker, B. John, W. Chiu, and A. Sali. 2005. Structural characterization of components of protein assemblies by comparative modeling and electron cryo-microscopy. J Struct Biol 149: 191-203.
8. He, J., Y. Lu, and E. Pontelli. 2004. A Parallel Algorithm for Helix Mapping between 3-D and 1-D Protein Structure using the Length Constraints. Lecture Notes in Computer Science 3358: 746-756.
9. Jiang, W., M. L. Baker, S. J. Ludtke, and W. Chiu. 2001. Bridging the information gap: computational tools for intermediate resolution structure interpretation. J Mol Biol 308: 1033-44.
10. Kong, Y., and J. Ma. 2003. A structural-informatics approach for mining beta-sheets: locating sheets in intermediate-resolution density maps. J Mol Biol 332: 399-413.
11. Kong, Y., X. Zhang, T. S. Baker, and J. Ma. 2004. A Structural-informatics approach for tracing beta-sheets: building pseudo-C(alpha) traces for beta-strands in intermediate-resolution density maps. J Mol Biol 339: 117-30.
12. Del Palu, A., J. He, E. Pontelli, and Y. Lu. 2006. (accepted) Identification of Alpha-Helices from Low Resolution Protein Density Maps. CSB Computational Systems Bioinformatics.
13. Wu, Y., M. Chen, M. Lu, Q. Wang, and J. Ma. 2005. Determining protein topology from skeletons of secondary structures. J Mol Biol 350: 571-86.
14. Rohl, C. A., C. E. Strauss, K. M. Misura, and D. Baker. 2004. Protein structure prediction using Rosetta. Methods Enzymol 383: 66-93.
15. Kim, D. E., D. Chivian, and D. Baker. 2004. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32: W526-31.
16. Simons, K. T., I. Ruczinski, C. Kooperberg, B. A. Fox, C. Bystroff, and D. Baker. 1999. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 34: 82-95.
17. Simons, K. T., C. Kooperberg, E. Huang, and D. Baker. 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 268: 209-25.
18. Bowers, P. M., C. E. Strauss, and D. Baker. 2000. De novo protein structure determination using sparse NMR data. J Biomol NMR 18: 311-8.
19. Ludtke, S. J., P. R. Baldwin, and W. Chiu. 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol 128: 82-97.