A Contact-assisted Approach to Protein Structure Prediction and Its Assessment in CASP10

Badri Adhikari, Xin Deng, Jilong Li, Debswapna Bhattacharyya, and Jianlin Cheng
Department of Computer Science, University of Missouri, Columbia, MO 65211 USA

ABSTRACT

Among different approaches to predict the 3D structure of a protein, one important idea is to predict a protein residue-residue contact map and then construct a full 3D structure from the contact-map. Instead of building a structure purely from contacts information, here we describe a contact-assisted structure prediction approach that uses only a few known contacts to improve the quality of already predicted models. Our approach for contact assisted structure prediction uses a novel method for selecting and refining protein structural models. With input test data as the predicted structures for 15 protein targets used in the contact-assisted prediction category in the 10th Critical Assessment of Techniques for Protein Structure Prediction (CASP10), we demonstrate that weighted contacts satisfaction score along with other established model quality assessment scores is a promising technique for selecting good structures and ultimately for better structure prediction.

Availability: http://protein.rnet.missouri.edu/contact_assisted/index.html

RESULTS

Evaluation of the scoring function of ranking models. Y-axis denotes real GDT-TS scores and X-axis indices of the models. Each group of models represents the models for a target, ordered according to their real GDT-TS scores. In each group, the top models selected by the total score, the APOLLO component, and the contact components were marked by three legends, respectively. The second group of models does not have a highest contact scoring model because only no-contacts were provided for this target.

<table>
<thead>
<tr>
<th>#</th>
<th>Target</th>
<th>Selected Model</th>
<th>Refined Model</th>
<th>Models Model</th>
<th>Total Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tc653</td>
<td>13.080 0.3206 0.3421</td>
<td>13.700 0.3029</td>
<td>13.700 0.3029</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>2</td>
<td>Tc653</td>
<td>14.780 0.2180 0.2230</td>
<td>14.720 0.2180</td>
<td>14.720 0.2180</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>3</td>
<td>Tc529</td>
<td>14.220 0.2278 0.2302</td>
<td>14.220 0.2278</td>
<td>14.220 0.2278</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>4</td>
<td>Tc656</td>
<td>9.850 0.2389 0.2427</td>
<td>9.850 0.2389</td>
<td>9.850 0.2389</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>5</td>
<td>Tc271</td>
<td>9.584 0.2586 0.2622</td>
<td>9.584 0.2586</td>
<td>9.584 0.2586</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>6</td>
<td>Tc657</td>
<td>11.480 0.2425 0.2463</td>
<td>11.480 0.2425</td>
<td>11.480 0.2425</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>7</td>
<td>Tc705</td>
<td>9.500 0.2240 0.2270</td>
<td>9.500 0.2240</td>
<td>9.500 0.2240</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>8</td>
<td>Tc542</td>
<td>10.710 0.2770 0.2805</td>
<td>10.710 0.2770</td>
<td>10.710 0.2770</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>9</td>
<td>Tc645</td>
<td>10.460 0.2120 0.2150</td>
<td>10.460 0.2120</td>
<td>10.460 0.2120</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>10</td>
<td>Tc656</td>
<td>9.620 0.2558 0.2597</td>
<td>9.620 0.2558</td>
<td>9.620 0.2558</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>11</td>
<td>Tc657</td>
<td>9.500 0.2240 0.2270</td>
<td>9.500 0.2240</td>
<td>9.500 0.2240</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>12</td>
<td>Tc271</td>
<td>10.500 0.2200 0.2225</td>
<td>10.500 0.2200</td>
<td>10.500 0.2200</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>13</td>
<td>Tc705</td>
<td>10.650 0.1767 0.1805</td>
<td>10.650 0.1767</td>
<td>10.650 0.1767</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>14</td>
<td>Tc656</td>
<td>9.760 0.2503 0.2543</td>
<td>9.760 0.2503</td>
<td>9.760 0.2503</td>
<td>0.282 0.072</td>
</tr>
<tr>
<td>15</td>
<td>Tc657</td>
<td>9.760 0.2503 0.2543</td>
<td>9.760 0.2503</td>
<td>9.760 0.2503</td>
<td>0.282 0.072</td>
</tr>
</tbody>
</table>

Evaluation of the top 1 prediction for all targets. Selected Models column shows the RMSD and GDT-TS score of the top 1 ranked model, selected by our Total Score formula, compared with the native structure. Refined Models column shows the RMSD and GDT-TS score of the top 1 ranked model after refinement. Final Improvement column shows the improvement in RMSD and GDT-TS after remodeling with MODELLER. Highlighted models are the models sent to the CASP10 competition. Re-modeling was not performed for targets Tc653 because no contacts were provided for this target. Targets Tc705, Tc271 and Tc734 were missed by mistake during the CASP10 experiment and so were not sent to CASP10.

EXAMPLE

Prediction of first domain of target Tc735 using our method. (a) model ranked top 1 by our Total Score formula in orange superimposed with native structure in green (b) Same structure after refinement in red superimposed with native with in green (c) Same structure re-folded using MODELLER with contacts as distance restraints superimposed with native in green.

ACKNOWLEDGEMENT

The work was partially supported by a NIH NIGMS grant (R01GM093123) to JC.