

Protein structure prediction using residue- and fragment-environment potentials in CASP11

Hyungrae Kim¹ and Daisuke Kihara^{1,2*}

¹Department of Biological Sciences, Purdue University, West Lafayette, Indiana, 47906

²Department of Computer Science, Purdue University, West Lafayette, Indiana, 47907

ABSTRACT

An accurate scoring function that can select near-native structure models from a pool of alternative models is key for successful protein structure prediction. For the critical assessment of techniques for protein structure prediction (CASP) 11, we have built a protocol of protein structure prediction that has novel coarse-grained scoring functions for selecting decoys as the heart of its pipeline. The score named PRESCO (Protein Residue Environment SCORe) developed recently by our group evaluates the native-likeness of local structural environment of residues in a structure decoy considering positions and the depth of side-chains of spatially neighboring residues. We also introduced a helix interaction potential as an additional scoring function for selecting decoys. The best models selected by PRESCO and the helix interaction potential underwent structure refinement, which includes side-chain modeling and relaxation with a short molecular dynamics simulation. Our protocol was successful, achieving the top rank in the free modeling category with a significant margin of the accumulated Z-score to the subsequent groups when the top 1 models were considered.

Proteins 2015; 00:000–000.
© 2015 Wiley Periodicals, Inc.

Key words: protein structure prediction; CASP11; decoy selection; scoring functions; residue environments; knowledge-based potential; helix interaction.

INTRODUCTION

Due to the increased number of deposited structures in the Protein Data Bank (PDB)¹ and the technical advancement of structure prediction algorithms, many recent methods are able to produce moderate to highly accurate models when appropriate template structures can be found in PDB. However, challenges remain for modeling a novel fold; that is, where appropriate template structures that cover a large portion of a target protein do not exist. Structure prediction methods that predict novel folds without relying on availability of template structures, often called *ab initio* or de novo folding methods, are also very important for designing artificial proteins.² In CASP11 (<http://predictioncenter.org/casp11/>), held in 2014, performance of prediction methods for novel folds was evaluated under the category of “free modeling.”

In structure prediction, particularly in an *ab initio* approach, it is key to develop an accurate scoring function for guiding the structure building process or for selecting near-native models from a pool of decoy struc-

tures. Many scoring functions have been developed over the past two decades, including physics-based functions and knowledge-based functions, which are based on statistics of geometric features of native proteins in PDB.¹ One well-studied and important class of knowledge-based scoring functions is contact potentials, which capture the propensities that residues or atoms interact with each other in protein structures.^{3–5} Contact potentials differ in various aspects, including contacting centers,⁶ additional geometric features considered (e.g., angles^{7,8}),

Abbreviations: CASP, Critical assessment of techniques for protein structure prediction; FM, Free modeling; MD, Molecular dynamics; PDB, Protein Data Bank; RMSD, Root-mean-square deviation; SCP, Screened coulomb potential; SDE, Side-chain depth environment

Grant sponsor: National Institute of General Medical Sciences of the National Institutes of Health; Grant number: R01GM097528; Grant sponsor: National Science Foundation; Grant numbers: IIS1319551, DBI1262189, IOS1127027.

*Correspondence to: Daisuke Kihara, Department of Biological Sciences/Computer Science, Purdue University, West Lafayette, IN 47906. E-mail: dkihara@purdue.edu

Received 27 May 2015; Revised 3 August 2015; Accepted 31 August 2015
Published online 7 September 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24920

the reference state used to normalize observed counts of contacts in the database,⁹ and the number of residues/atoms considered.^{10–12}

We have recently developed a new knowledge-based scoring function for selecting near-native structures called PRESCO (Protein Residue Environment SCORE).¹³ PRESCO compares the local structural environment of each residue in a structure model to those in native structures to quantify native-likeness of the residue environment. The local structural environment considers the coordinates and the residue depth¹⁴ of neighboring amino acids of the target residue. A target residue environment is rewarded if similar environments found in the reference structure dataset are of identical or similar residue types to the target, which indicates that the target environment is specific to the residue type. Residue similarity is defined by a combination of several amino acid similarity matrices.¹⁵ PRESCO is designed to capture the interactions of multiple residues, which distinguishes itself from the pairwise nature of existing knowledge-based potentials. PRESCO was benchmarked on decoy sets that are commonly used to test scoring functions' native and near-native recognition ability and showed better performance than several existing scoring functions.¹³

We have also recently developed another coarse-grained statistical potential that captures the preferred spatial arrangement of helices in protein structures. The characteristic distribution of the crossing angles of two helices has been reported earlier^{16–21} and knowledge-based potentials have been developed for helix packing preferences.^{22,23} We implemented a helix-helix interaction potential in CABS, a coarse-grained protein folding program,²⁴ as a scoring term to supplement the other CABS potentials.

In CASP11, we used PRESCO and the helix-helix interaction potential for selecting decoys from server models that were made available for human predictors. Selected models were refined by side-chain modeling and structure relaxation with a short molecular dynamics (MD) simulation. The whole procedure turned out to be successful. Our group was ranked first among all the participants in the free modeling (FM) category when the top 1 models were considered (http://www.predictioncenter.org/casp11/zscores_final.cgi?formula=assessors). Using the assessors' formula for model evaluation, which considers six scores, GDT-TS,²⁵ IDDT,²⁶ TenS,²⁷ QCS,²⁸ contS,²⁹ and Molprobity,³⁰ our group (Kiharalab, group code 333) obtained the Sum Z-score of 44.2048, which has a relatively large gap of 7.5064 to the second group with 36.6984. When the models with the best scores among five submitted models were considered, our group was ranked second. In this case, the Sum Z-score gap to the first group was 5.452. Now that the native structures of targets and group rankings have been released, we perform a retrospective examination of each

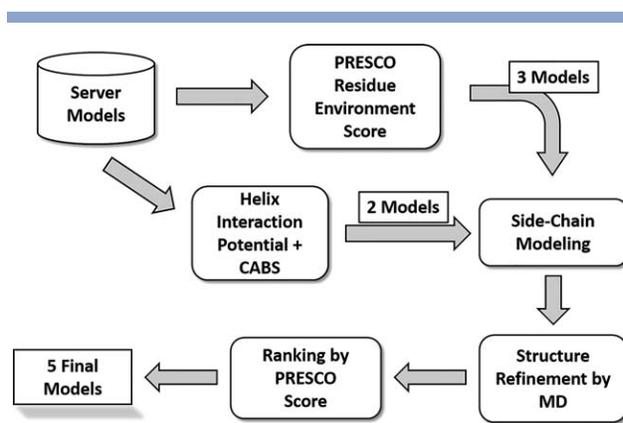


Figure 1

The schematic diagram of the prediction protocol we used in CASP11.

step of our prediction procedure to discuss its successes as well as potential areas for improvement.

MATERIALS AND METHODS

Overall structure prediction procedure

Our structure prediction procedure is summarized in Figure 1. PRESCO played the main role in selecting initial models from server models that were made available to human predictors. Three models were selected with PRESCO and the other two were selected by our recently developed helix-helix interaction potential that is implemented in CABS. Side-chains of selected models were removed and rebuilt with Oscar-Star.³¹ Finally, models underwent structure relaxation with a short MD simulation. The five models were ranked by PRESCO again and submitted. Below, each step is explained in more detail.

Model selection by PRESCO

Here we overview the essence of the PRESCO algorithm. Please refer to the original paper¹³ for further details. PRESCO evaluates how much each residue in a structure model is native-like by comparing the neighboring residues of the target residue to those in a reference structure database. The comparison is performed in the following steps. First, the main-chain conformation of the nine-residue-long fragment centered at the target residue is compared against the structures in the reference database and the 500 lowest root-mean-square deviation (RMSD) fragments of the same size from the database are pre-selected. Then, the neighborhood of residues are compared between the target residue and those in the center of the pre-selected fragments, where the neighborhood is defined with a sphere of 8.0 and 6.0 Å radius (the side-chain depth environment, SDE). SDEs that have a different number of side-chain centroids in

the sphere are discarded. To compute similarity of two SDEs, first the side-chain centroids of residues in the spheres are superimposed to pair residues in the two SDEs. Then, the similarity of two SDEs is defined by the RMSD of the residue depth¹⁴ of the amino acids in the two spheres. Residue depth quantifies the distance from the residue position to the protein surface.¹⁴ For a target residue, the 40 most similar (that is, smallest depth RMSD) SDEs to the query SDE are selected and ranked according to their depth RMSD. Subsequently, a score is computed for the target residue, which is a weighted sum of an amino acid similarity score computed for each retrieved residue, where the weight reflects the rank of the retrieved residue. Thus, if a retrieved residue is similar to the target residue according to the amino acid similarity matrix used, the target residue will receive a high score. We explain in more detail below. Finally, the score of a model is the sum of the score given to each residue in the model:

$$\text{SDE_based_Score} = \sum_{i=1}^L \sum_{j=1}^{N \times 40} w_j S_{a_i - a_j} \quad (1)$$

where L is the length of the protein model, N is the number of times that the residue i appears in SDEs of the protein model, 40 is the number of SDEs retrieved from the database for a query SDE in the model, and $S_{a_i - a_j}$ is the amino acid similarity score taken from a matrix S for residue i in the query and residue j retrieved from the database.

One of the keys of this scoring function is the choice of the weighting scheme and the amino acid similarity matrix. During the course of developing PRESCO, we benchmarked a dozen amino acid similarity matrices obtained from the AAIndex database^{32,33} and identified several weighted combinations of matrices that performed well in identified near-native decoy structures. Combinations of matrices and weights were tested in near-native decoy recognition on the Rykunov & Fiser dataset³⁴ which consists of decoys of 143 proteins that were used as prediction targets in rounds 5 to 8 of CASP. On average, there are 18.3 models per target.

We explored combinations of two matrices and found the following five combinations gave good performance:

CC80 matrix¹⁵ with a weight $1/(\text{RMSD})^{0.01}$ and BLOSUM30³⁵ with a weight $1/\{\lfloor n/2 \rfloor + 1\}$

QU_C1 matrix³⁶ with a weight $1/\{\lfloor n/5 \rfloor + 1\}$ and QUIB matrix³⁷ with a weight $1/\{\lfloor n/5 \rfloor + 1\}$

CCPC matrix¹⁵ with a weight $1/(\text{RMSD})^{0.01}$ and BLOSUM30 with a weight $1/\{\lfloor n/2 \rfloor + 1\}$

QU_C2 matrix³⁶ with a weight $1/(\text{RMSD})^{0.1}$ and BLOSUM30 with a weight $1/\{\lfloor n/2 \rfloor + 1\}$

QUIB matrix³⁷ with a weight $1/\{\lfloor n/2 \rfloor + 1\}$ and QU_C1 with a weight $1/\{\lfloor n/5 \rfloor + 1\}$

In the weight equations, n is the rank of the SDE among 40 retrieved from the reference dataset and $\lfloor \cdot \rfloor$ is the floor function (which returns the largest integer that does not exceed the provided value). The weight decreases as the rank of the retrieved residue decreases. CCPC is a matrix that is based on the correlation coefficients of an amino acid residue contact potential, while CC80 is a linear combination of CCPC and another matrix (AAIndex ID: KOLA920101),³⁸ which is based on the similarity of the dihedral angles of amino acids. QUIB (AAIndex ID: QUIB020101) is a numerically optimized amino acid matrix to minimize the average RMSD of aligned proteins in benchmark databases.³⁷ QU_C1 (AAIndex ID: QU_C930101) and QU_C2 (AAIndex ID: QU_C930102) are matrices that capture amino acid residue contact propensities.³⁶ BLOSUM30 was computed from observed mutation frequency in multiple sequence alignments and is commonly used for sequence alignments and sequence database searches.³⁵

The idea of developing scores for evaluating residue environments is not new. Following the observations that a residue in a structure affects to the structure and distribution of surrounding residues,^{39–44} several multi-body contact potentials^{11,12,45} and scoring functions that consider residue environments have been developed.^{46–48} The Levitt group has developed a hydrophobic score that consider the number of residues within a 10 Å radius sphere and the number of interactions with surrounding hydrophobic residues and applied it to threading.⁴⁷ DeGrado and his co-workers developed a statistical potential for atomic environments where an atom environment is specified by the number of other atoms in the environment and types of contacting atoms.⁴⁶ They used the potential to select native and near-native structures from decoy sets. Mooney and his colleagues used a residue environment representation that captures atoms within concentric spheres around a C β atom of a residue to recognize functional sites of proteins.⁴⁸

Compared to the environment scores mentioned above, PRESCO have two novel aspects. First, PRESCO judges the similarity of residue environments with multiple structural criteria by considering the similarity in main-chain conformation, the number of residues in the probe sphere, and the depth of residues from the protein surface. Second, PRESCO employs multiple amino acid similarity matrices, which reflect different aspects of amino acids, to define amino acid similarity between the query residue and residues with similar environments that are retrieved from the reference database. Thus, PRESCO examines the environment of residues more thoroughly from various different angles than the existing environment scores.

The near-native model recognition performance of these five score combinations on the Rykunov & Fiser set is summarized in Table I. In the table we also list the

Table I

Performance of SDE Pair Combinations of PRESCO on the Rykunov & Fiser Decoy Set

Scoring function	Average rank ^a	Ranked 1 ^b
SDE (QUIB)	2.89	56
CC80 + BLSM30	2.82	66
QU_C2 + BLSM30	3.24	66
QU_C1 + QUIB	3.13	66
QUIB + QU_C1	3.03	66
CCPC + BLSM30	3.04	66
QMEAN6	2.87	85
RWplus	2.97	57
RW	3.08	51
DOPE	5.77	54
DFIRE	6.03	50
OPUS_PSP	5.39	54
Random	9.72	13.9

The values for QMEAN6 and below are taken from Zhang & Zhang (2010).

^aThe average rank of the lowest energy (or the highest scored for PRESCO) decoy by each score in the absence of the native structure in the decoy set. The decoys were ranked in terms of their GDT-TS score to the native.

^bThe number of decoy sets when the best scoring model by each score was the closest to the native. The native structure was excluded in the decoy sets.

performance of six existing scores on this dataset taken from Zhang & Zhang (2010).⁸ Among the scores compared, the combination of CC80 and BLSM30 showed the best performance in the average rank of the best decoys (2.82). In terms of the number of decoy sets where a scoring function selected the closest-to-native model with the best score, QMEAN6 showed the best results with 85 decoy sets. The PRESCO score pair combinations came next with 66 decoys, which were better than the other five existing scores.

Server models downloaded from the CASP website were ranked by the Z-score sum of the five score combinations, and three best scoring models were selected from the pool. Two more models were selected by the helix-helix interaction potential as we describe in the next section. After the refinement step, the five refined models were ranked with PRESCO again to decide the final order of the models. The first three were always the models that were originally selected with PRESCO, and the first model (Model 1) was always the best model by PRESCO. The order of the second and the third models were decided by visual inspection. The fourth and the fifth models were those originally selected by helix potential.

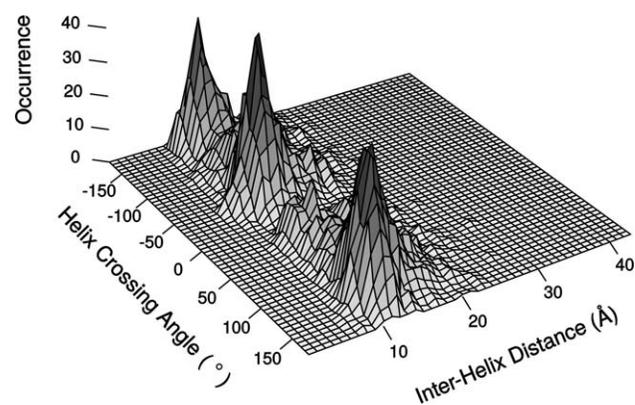
Helix-helix interaction potential and CABS

We have developed a new knowledge-based helix-helix interaction potential based on the observed frequency of helix pair interactions depending on the angle and distance between them. This potential is still under development and details will be described elsewhere. In this section, we outline the nature of the potential and show benchmark results on a small dataset.

The statistics of helix interactions were taken from a set of 2536 globular proteins obtained from the PISCES server⁴⁹ with a 25% sequence identity cutoff. Figure 2 shows the observed frequencies of crossing angles and distance between two helices. The angle distribution is consistent with previous studies.^{19,20} The observed distribution was normalized by the expected number of counts of angles²⁰ and distance to compute a knowledge-based potential. The new helix-helix interaction potential was implemented into the CABS protein structure modeling program²⁴ as an additional potential term among the other knowledge-based scoring terms in CABS. The helix potential was only computed for input protein structure models with two or more helices. Otherwise, structures were evaluated with the original potentials implemented in CABS. CABS is a coarse-grained protein folding program that can fold a protein structure for *ab initio* structure prediction; however, we limited movement of the initial structure and mainly used the framework for evaluating models.

In Table II, we show benchmark results of the helix potential on selecting the native structure for 19 CASP9 targets out of an average of 47.9 server decoys. All of these targets were helical or α/β proteins. Three scoring schemes were compared: the helix potential, the original CABS potential, and the integrated CABS with the helix potential. The average native rank of the 19 targets was 12.2 by the CABS potential, 10.9 by the helix potential, and 8.7 for the integrated CABS with the helix potential. In a head-to-head comparison in terms of the native structure rank, the helix potential outperformed CABS for 10 targets while the integrated helix and CABS won over CABS for 12 among the 19 targets (with one tie).

The advantage of the helix potential over CABS became clearer for difficult targets when none of the decoys were close to the native. When 10 targets were

**Figure 2**

The distribution of crossing angles of helices was shown as a function of inter-helices distance. The distance was measured between the centers of helices.

Table II

CASP 9 Native Structure Rank Scored by Helix and CABS Potential

Helix bundles CASP9 target domains	Top5 models mean RMSD to Native (Å)	Native rank by helix potential	Native rank by CABS potential	Native rank By helix and CABS potential
T0516-D1	2.6	3	19	5
T0534-D1	23.5	1	4	1
T0534-D2	17.7	1	5	1
T0538-D1	2.1	17	3	10
T0544-D1	11.2	6	28	4
T0547-D3	9.4	3	1	2
T0548-D2	2.7	5	12	3
T0553-D2	6.3	10	20	10
T0555-D1	11.4	9	8	12
T0575-D1	4.7	20	16	14
T0575-D2	3.2	23	8	21
T0586-D2	2.9	28	17	7
T0602-D1	1.5	16	24	18
T0608-D1	12.1	3	4	4
T0611-D1	5.5	1	4	1
T0615-D1	4.8	9	7	5
T0619-D1	1.6	34	20	27
T0627-D1	5.7	7	24	9
T0637-D1	19.9	11	8	11
Average Rank of Native Structures for 10 high-mean-RMSD Targets		10.9	12.2	8.7
Average Rank of Native Structures for 10 high-mean-RMSD Targets		4.0	11.5	5.5

considered where the RMSD of the best five models was larger than 5.0 Å, the average rank of the native structure by CABS, the helix potential, and the integrated helix and CABS were 11.5, 4.0, and 5.5, respectively. These results suggest that the helix potential is able to capture coarse-grained native-like features of helical proteins.

Side-chain remodeling

The refinement steps consist of side-chain remodeling and structure relaxation by MD. We used Oscar-Star to rebuild the side-chains³¹ of a model after having removed them. Oscar-star was chosen because it performed best in our recent benchmark study in which we compared eight side-chain prediction programs in building side-chains of different structural environments.⁵⁰

Structure minimization by MD

Next, the structure of a model was relaxed with MD with the CHARMM molecular mechanics potential.⁵¹ An implicit solvent with screened coulomb potential (SCP) was used. After a 50 step initial minimization, MD was run for 100 steps (0.2 picoseconds) at 100 K while restraining the C α atoms to their initial position.

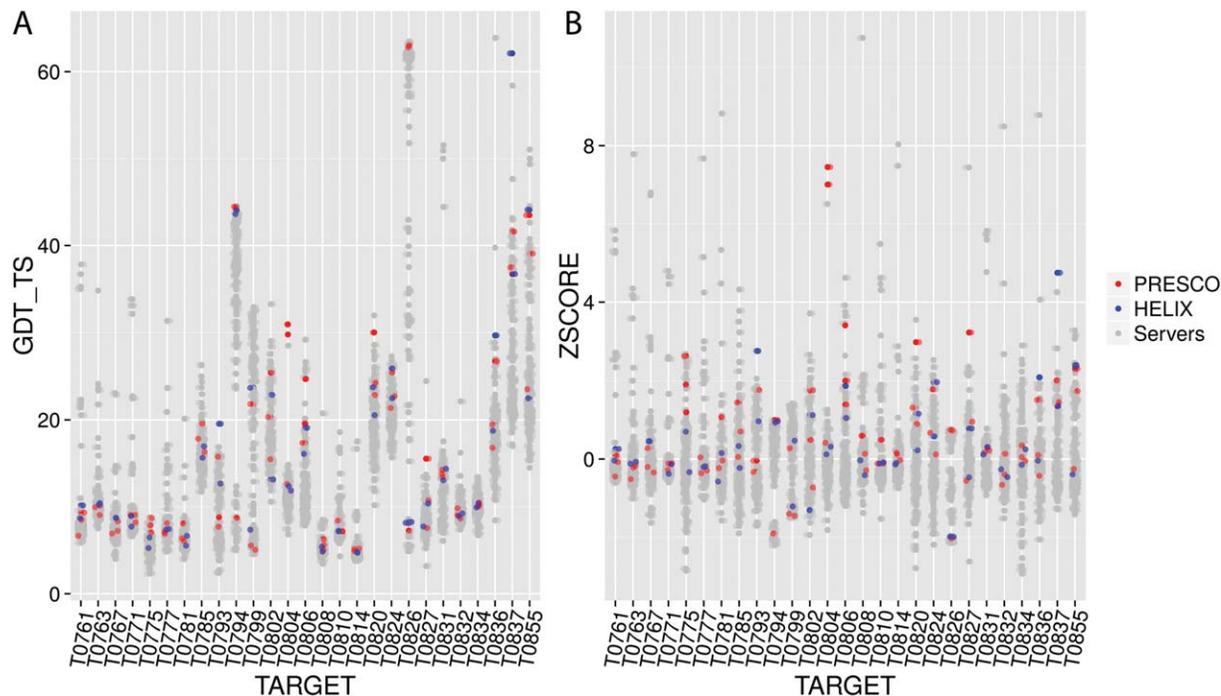
Scores for model evaluation

In CASP11, six scores were used by the assessors to evaluate and rank submitted models. They were GDT-TS (Global Distance Test Total Score),²⁵ IDDT (local Dis-

tance Difference Test score),²⁶ TenS (Ten scores used in CASP5),²⁷ QCS (Quality Control Score),²⁸ contS (Contact score),²⁹ and Molprobit.³⁰ GDT-TS evaluates the fraction of residues in a model that are placed within four cutoff distances to the native structure after superimposition. IDDT is an alignment-free local structure evaluation for a model which considers the fraction of preserved atom pair distances between a model and native. TenS integrates ten scores which include GDT-TS, secondary structure overlap, structure alignment scores, and sequence-level alignment scores. QCS aims to capture global features of a model by evaluating the mutual arrangement of secondary structure elements. contS evaluates the similarity of C α distances between a model and native. Molprobit was designed to validate structures solved by X-ray crystallography and thus evaluates the most detailed structural features of a model including atom contacts, hydrogen bonds, bond angles and lengths, side-chain rotamers, and main-chain dihedral angles. With the exception of Molprobit, a larger score indicates a better quality model.

RESULTS

The results will be shown in two parts. First, we discuss how well our decoy selection procedure performed. Subsequently, the effect of the refinement steps will be examined.

**Figure 3**

Quality of selected models by PRESCO and the helix potential. 3 models were selected with PRESCO (red) and two more models were selected with the helix-helix interaction potential (blue). A, GDT-TS; B, the Z-score of GDT-TS; of the selected models among server models made available. 81 models selected with PRESCO and 54 those which were selected with the helix potential.

Quality of selected models by PRESCO and the helix potential

Selecting good quality models from available server models was a key for success in our protocol. In Figure 3(A,B), we show the distribution of GDT-TS and the Z-score (computed among all the server models) of GDT-TS of the selected models among server models that were made available for human predictors. The number of submitted server models for a target ranges from 184 to 199 models with an average of 191.69. Even though six of them (T0775, T0793, T0799, T0802, T0804, T0826) do not have their crystal structure available as of writing of this article, we discuss all 27 targets released for prediction for FM targets based on the released assessment from the prediction center. Three models were selected with PRESCO [shown in red Fig. 3(A,B)] and two additional models were selected by the helix potential [shown in blue Fig. 3(A,B)].

The average GDT-TS score of all the server models (the grand mean of the target means) was 14.6. The average GDT-TS score of PRESCO-selected models was higher, 16.47, while that of the models selected by the helix potential was 15.42. If we consider the best GDT-TS model selected by PRESCO and the helix potential for each target, the margin between PRESCO and the helix potential increased slightly to 20.18 and 17.52 for

PRESCO and the helix potential, respectively. In terms of the average Z-score [Fig. 3(B)], PRESCO also showed better performance than the helix potential. The average Z-scores of the selected models were 0.61 and 0.32 for PRESCO and the helix potential, respectively. When the best Z-score model was considered for each target, again the advantage of PRESCO over the helix potential increased, with the average Z-scores of the selected models being 1.39 and 0.75, for PRESCO and the helix potential, respectively.

Although our scoring functions did not always select the top models among the available server models, there are notable cases where the selection was very successful. Among the 27 targets, there were 8 and 5 cases where PRESCO and the helix potential, respectively, selected a model among the top 5 server models available. Specifically, PRESCO selected the best model out of 192 server models for T0804, the second best model out of 192 server models for T0775, T0804 (thus both the best and the second best models were selected for this target by PRESCO), T0820, T0827 and the third best model for T0794. On the other hand, the helix potential selected the best model from 192 server models for T0793 and T0837 and the third best model for the targets T0836. T0837 and T0836 have an α -helix bundle structure and T0793 is an α/β class protein.

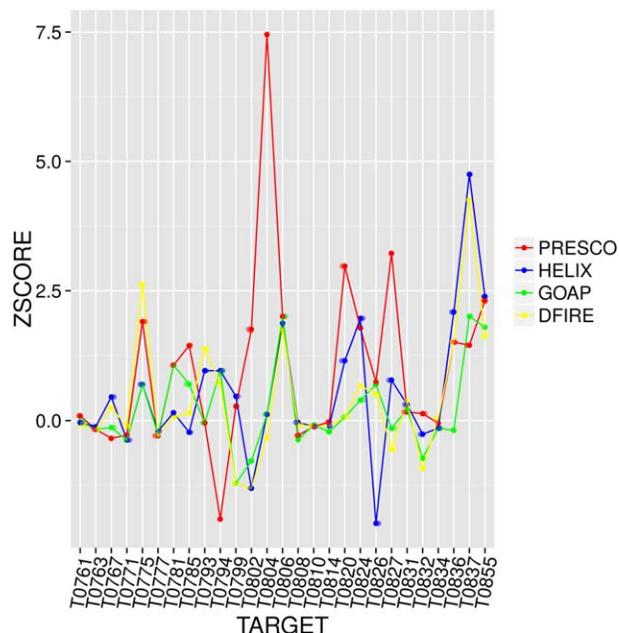


Figure 4

Comparison of selected models by PRESCO, the helix potential, GOAP, and DFIRE. Colors are PRESCO: red, the helix potential: blue, GOAP: green, and DFIRE: yellow. The GDT-TS Z-score of models that were ranked the best by each score among the available server models were plotted.

In Figure 4, we compared the performance of the two scoring functions with two existing ones, DFIRE⁹ and GOAP.⁵² The average GDT-TS Z-score of the top-selected models with PRESCO, the helix potential, DFIRE, and GOAP were 0.99, 0.53, 0.40, and 0.22, respectively. When the best models among between the PRESCO and the helix potential was considered, the PRESCO/helix potential showed an average Z-score of 1.58 while DFIRE and GOAP's values were 1.11 and 0.87, respectively. Examples of targets for which PRESCO and the helix potential outperformed GOAP and DFIRE and opposite cases are shown in Figure 5 and the associated Table III. As shown in Figure 4, PRESCO selected better models than DFIRE and GOPE for most of the targets. Those targets include α -class proteins, such as T0804, T0802, and T0785 [Fig. 5(A–C)] and α -class proteins, including T0827 and T0820 [Fig. 5(D,E)]. But for some α -helical proteins were better selected by the helix potential than PRESCO. T0836 and T0837 are such examples [Fig. 5(F,G)]. The last two panels, T0775 and T0793 [Fig. 5(H,I)], show the opposite cases, where DFIRE performed better than PRESCO in selecting decoys. These are relatively large proteins with long loops.

Overall, PRESCO and the helix potential performed fairly well in selecting good quality models with notable success in several cases. In this evaluation, our scoring

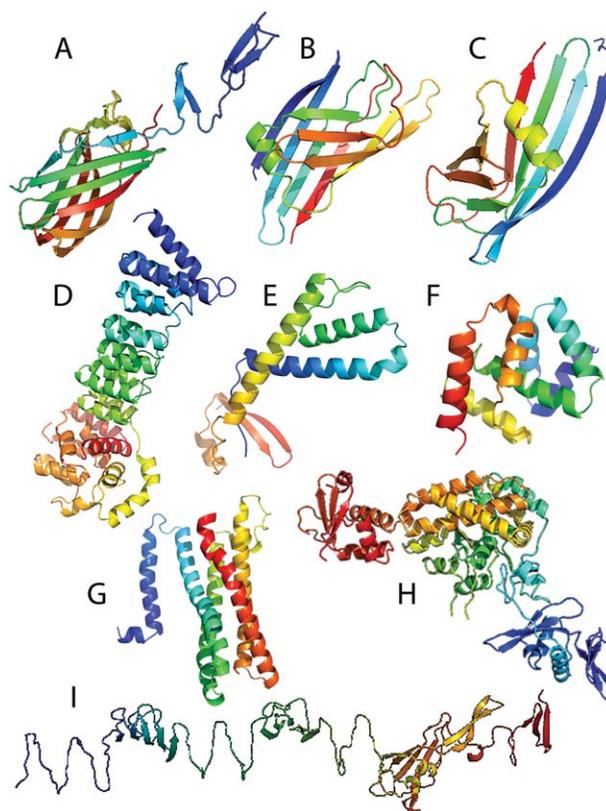


Figure 5

Examples of targets for which a score performed better than other scores. Four scores, PRESCO, the helix potential, DFIRE, and GOAP were compared. A, T0804; B, T0802; C, T0785; D, T0827; E, T0820; F, T0837; G, T0836; H, T0775; I, T0793. A, B, C, are examples of targets of β -class folds for which PRESCO outperformed the other scores. D and E are examples of α -class fold targets for which PRESCO outperformed. F and G are helix bundle protein targets for which the helix potential's selections were better than the other three scores. H and I are cases that DFIRE performed better than PRESCO and the helix potential. The Z-scores of the models selected by the four scores for these targets are listed in Table III.

Table III

Z-Scores of the Selected Models for Representative Targets by the Four Scores

Targets	Fig. 5 Panel a)	PRESCO	Helix	DFIRE	GOAP
T0804	A	7.45	0.12	-0.32	0.12
T0802	B	1.76	-1.31	-1.31	0.12
T0785	C	1.45	-0.23	0.15	0.70
T0827	D	3.22	0.78	-0.56	-0.15
T0820	E	2.98	1.16	0.07	0.07
T0837	F	1.48	4.74	4.25	2.01
T0836	G	1.51	2.09	-0.19	1.51
T0775	H	1.90	0.70	2.62	0.70
T0793	I	-0.04	0.96	1.38	-0.04

The Z-score of the top choice model by the four scores are listed. The largest Z-score for each target among the four selected models is shown in bold. A Z-score of a model was computed for the model's GDT-TS score relative to all the server models.

^aCorresponding panels in Figure 5 are indicated.

Table IV
Servers From Which Models Were Selected By Our Scoring Functions

A. The number of models selected from each servers by PRESCO.

Servers	TOP1	Within TOP3
myprotein-me	6	14
BAKER-ROSETTA-Server	6	11
Zhang-Server	5	19
RBO_Aleph	4	7
QUARK	3	13
nns	1	4
FFAS-3D	1	2
SAM-T08-server	1	1
TASSER-VMT	0	4
RaptorX-FM	0	2
MULTICOM-NOVEL	0	1
Seok-server	0	1
BioSerf	0	1
STRINGS	0	1
Total	27	81

Three out of five models were selected by PRESCO.

B. The number of models selected from each servers by the helix potential.

Servers	TOP1	Within TOP2
Zhang-Server	5	7
Baker-ROSETTA-Server	4	8
Seok-server	4	6
myprotein-me	3	8
nns	2	5
QUARK	2	4
Pcons-net	2	2
RBO_Aleph	2	3
SAM-T08-server	1	1
TASSER-VMT	1	1
BioSerf	1	1
RaptorX-FM	0	1
MULTICOM-REFINE	0	1
MULTICOM-CLUSTER	0	1
FFAS-3D	0	1
eThread	0	1
Distill	0	1
PSF	0	1
Atome2_CBS	0	1
Total	27	54

Two models were selected by the helix potential.

function performed better than the two existing potentials.

In Table IV, we provide a list of servers from which PRESCO and the helix potential from which models were selected. The majority (88.9% when only Model 1 models were considered, and 79.0% when three models were considered) of PRESCO's choices were from five servers (Table IV A). The helix potential selected models from more diverse servers (Table IV B). When the top choices of the helix potential were considered, 88.9% were comprised of eight servers.

Refinement of selected models

Selected models underwent the two refinement steps (Fig. 1), side-chain rebuilding and structure relaxation by

MD. We analyzed how much the six evaluation scores changed due to the two refinement steps applied to the models. Figure 6 shows the Molprobity score (lower is better) for 135 models submitted for the 27 targets. It is shown that the Molprobity score improved or showed no change for 92.6% (73 cases improved out of 135, there were no change in 50 cases, 12 cases become worse) targets by the side-chain rebuilding with Oscar-Star [Fig. 6(A)]. In many cases the improvement is substantial with a change of over 0.5. The average decrease of Molprobity score was 0.682.

However, it turned out that the subsequent structure relaxation step deteriorated many models [Fig. 6(B)]. Indeed, the Molprobity score of 74.8% of the models was made worse by structure relaxation. Particularly, the score of 19 models showed an adverse change of 2 to 3. Because of this unsuccessful structure relaxation step, the overall post-processing procedure decreased the effect of refinement [Fig. 6(C)]. At the end, the number of models with an improved or unchanged score after the entire refinement procedure reduced to 60 (44.4%) from the 123 that were improved after the side-chain rebuilding step. We also examined changes of the other five scores, GDT-TS, IDDT, ContS, QCS, and TenS, which evaluate larger structural differences of models, but only minor changes were observed (data not shown).

During CASP11, we used DFIRE energy to evaluate the effect of the refinement procedure as the native structures of targets were not known. Figure 7 shows the change in DFIRE energy (lower is better) of the submitted 135 models for the 27 FM targets. Improvement of DFIRE was observed for the majority of the models. The average decrease was -1586.60 . The most significant decrease of DFIRE was observed for a model for target T0793, whose energy improved from -48381.36 to -56840.32 by -8458.96 .

To summarize the results in Figures 6 and 7, the applied refinement procedure improved the DFIRE energy of the majority of the models, but did not impact the evaluation scores with the exception of Molprobity. Molprobity was improved substantially by the side-chain rebuilding but was worsened by the subsequent structure relaxation with MD, which weakened the effect of the entire refinement effort.

Quality of our submitted models

Figure 8 shows six scores, GDT-TS, IDDT, TenS, QCS, contS, and Molprobity, of our submitted first model (Model 1) in comparison with Model 1 models of all human and server groups. The average rank of our Model 1 models for the 38 domains was 20.9 for GDT-TS, 16.9 for IDDT, 17.4 for tenS, 21.4 for QCS, 18.0 for contS, and 24.9 for Molprobity. Our models were ranked within the top 5 by GDT-TS 6 times, and 9, 9, 8, 6, 8 times, by IDDT, tenS, QCS, contS, and Molprobity,

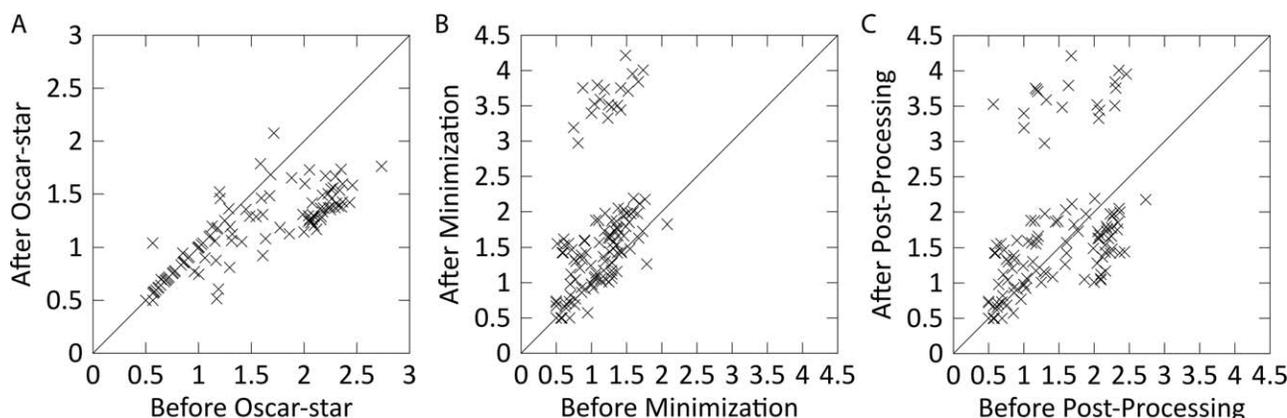


Figure 6

Change of the Molprobity score in the model refinement. **A**, Molprobity of models before and after the side-chain rebuilding with Oscar-Star. **B**, Molprobity of models before and after structural relaxation with short energy minimization by MD. Our submitted 135 models for 27 FM targets were analyzed. **C**, Molprobity of models before and after the whole refinement procedure that consists of the side-chain rebuilding and the short energy minimization.

respectively. The average Z-score of our Model 1 models for the 38 domains was 1.15 for GDT-TS, 1.48 for IDDT, 1.30 for tenS, 1.03 for QCS, 1.15 for contS, and 1.16 for Molprobity. Thus, among the six scores, our models were evaluated better on average by IDDT and tenS relative to the other groups' submissions.

9 of our Model 1 models were ranked within the top 5 by two or more measures, and our model for T0761-D2 was selected among the top 5 models in terms of five scores, GDT-TS, IDDT, tenS, QCS, and contS. T0775-D5, T0804-D1, T0804-D2, and T0834-D1 were ranked among the top 5 models by GDT-TS, IDDT, and tenS. T0785-D1, T0793-D1, and T0793-D5 were ranked among the top 5 by IDDT and QCS. T0826-D1 was ranked among the top 5 by IDDT and Molprobity. T0855-D1 was ranked among the top 5 by contS and Molprobity.

Examples of submitted models

In Figure 9, three examples of our models are shown. The first example is the Model 1 model for T0804-D2 [Fig. 9(A)], which is a domain of residues 46–197 of murine adenovirus fibre head (PDB structure not yet released). This is the best model among all submissions for this target. The GDT-TS score of this model is 38.82. There are two other groups (Boniecki_pred and Skwark), who produced models with a similar GDT-TS (38.65, 37.83, respectively), but all the rest of the submitted models have substantially worse GDT-TS of lower than 21.0. Compared to its native structure [Fig. 9(B)], the β -structure of this protein is not perfectly modelled, but the topology of the main-chain is essentially the same as native. Our model has a substantially better Molprobity score of 1.38 than the Boniecki_pred and Skwark models, whose scores are 2.63 and 2.98, respectively. This indi-

cates that the structure refinement worked for this model.

The second example is the Model 1 model of T0799-D1 [Fig. 9(C)], which is a domain of residue 1 to 141 of a 408 residue-long protein, pb1 plus chaperone domain (PDB structure not yet released). Together with other two groups (MUFOLD-R and SHORTLE), our first model for this domain has the best GDT-TS of 19.86. This is a difficult target as indicated in the average GDT-TS of 14.19 by all human and server models. Compared to the native [Fig. 9(D)], the structure of the core of the domain with three strands and a flanking helix is captured by our model, although the model failed to predict

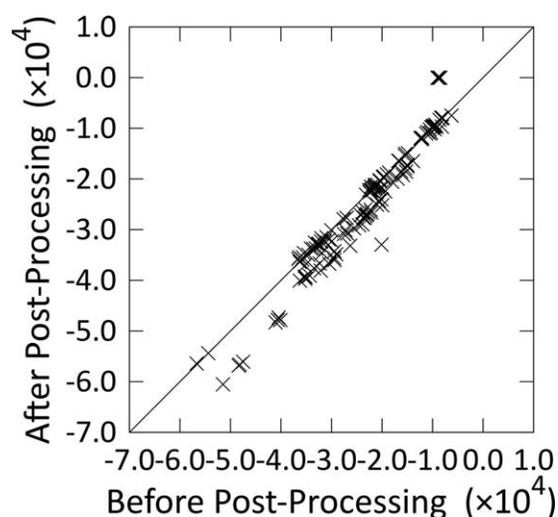


Figure 7

Change of DFIRE energy by applying the refinement procedure to models. 135 submitted models for all 27 FM targets were plotted.

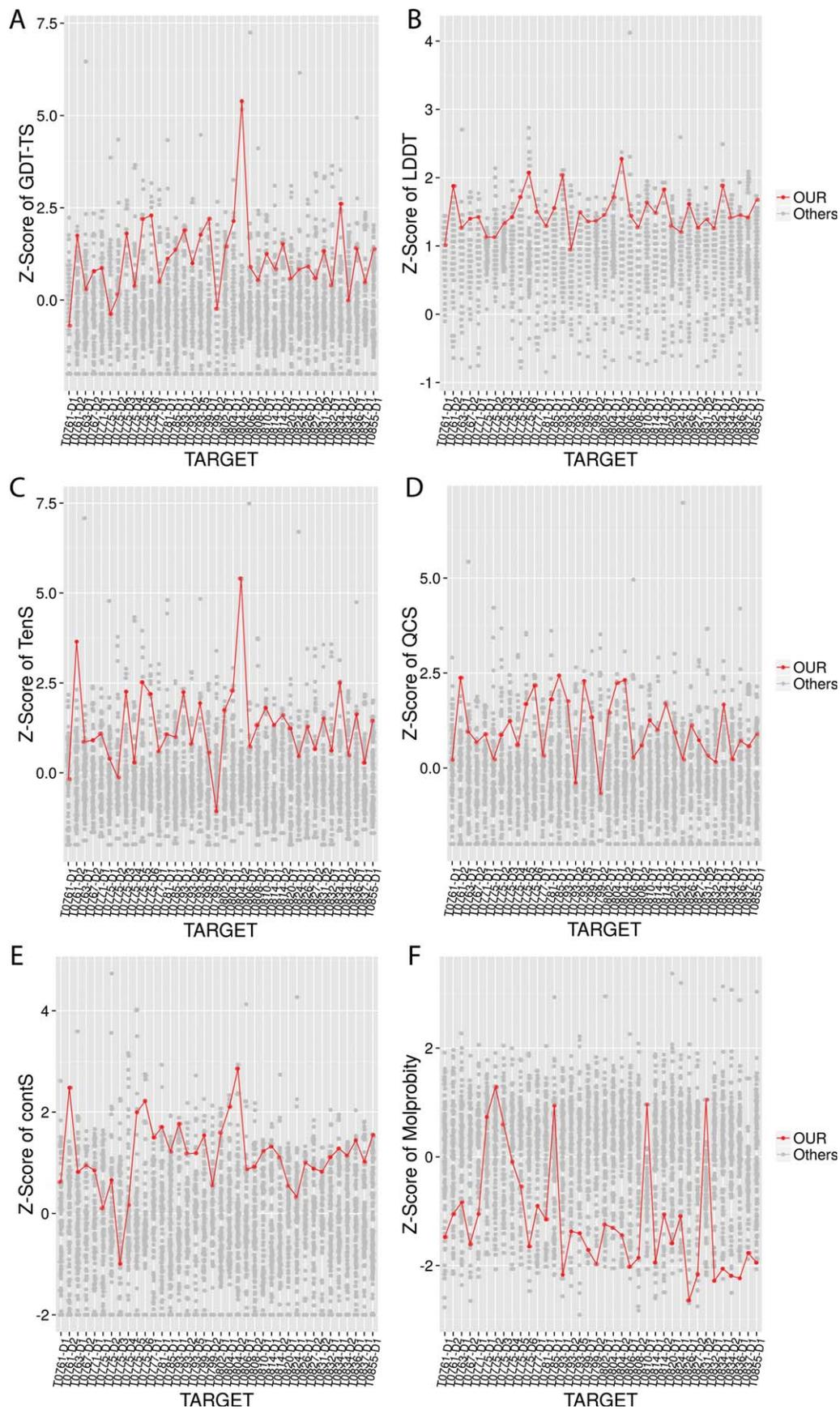
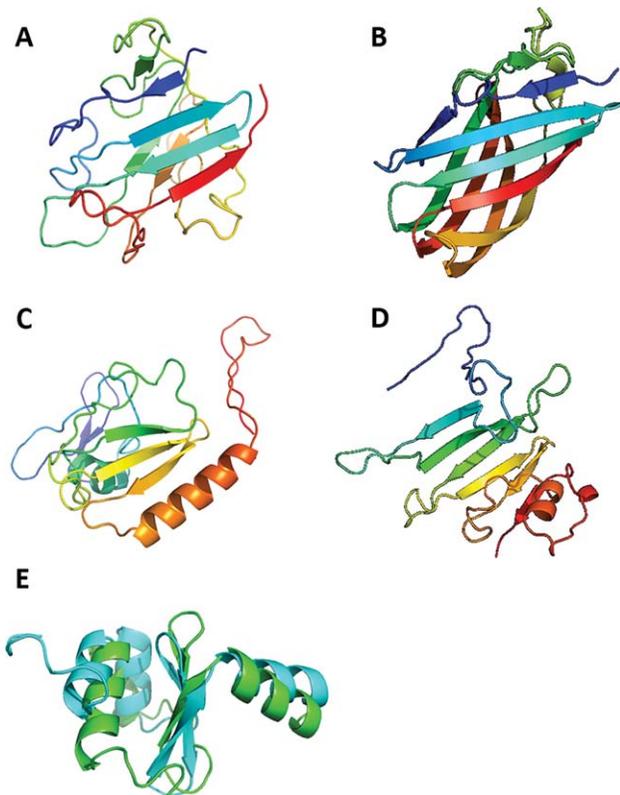


Figure 8

Z-score distribution of six scores for Model 1 from all human and the server groups. Our models are colored in red. **A**, GDT-TS; **B**, LDDT; **C**, TenS; **D**, QCS; **E**, contS; and **F**, Molprobability. GDT_TS, ContS, tenS, QCS results were provided by the organizer upon our requests and LDDT and Molprobability were computed by us for models downloaded from the CASP11 website.

**Figure 9**

Examples of our successful models relative to the other submissions. **A**, Model 1 of our group for T0804-D2. **B**, the native structure of T0804-D2. **C**, Our Model 1 for T0799-D1. **D**, the native structure of T0799-D1. **E**, Superposition of our first model (green) and the native structure (blue) of the residue 130 to 192 of T0834-D1. This model has a TM-score of 0.45 and a GDT-TS score of 0.59.

the N-terminal region of the protein. Similar to the first example, our model had a better Molprobrity score (0.80) than the two models with the same GDT-TS (1.17 and 2.29).

The last example is the first model for T0834-D1, which consists of two separated regions of the sensor domain of histidine kinase (PDB ID: 4r7q) residues, residue 2 to 37 and another region of residue 130 to 192 [Fig. 9(E)]. Our Model 1 model produced the second part of the domain well, with a TM-score of 0.45 and a

GDT-TS of 0.59 and ranked the third among all submissions. Again our model had a better Molprobrity score (0.69) than the two other models that had a higher GDT-TS score than our model (0.96, 1.15).

Computational time of PRESCO

In Table V, we compared the computational time needed by PRESCO with three other scores, GOAP, RWplus,⁸ and dDFire.⁵³ In the current naïve implementation of PRESCO, it takes significantly longer time to compute a score for a structure model compared to the other three scores. This is because residue environments, MRE and SDE, of 2536 reference structures in the database are not precomputed but newly computed again when each of residues from a model is compared against. We are in the process of improving the computational speed by precomputing and storing the MREs and SDEs of reference structures and by using an efficient searching method.

DISCUSSION

Here we investigated the effectiveness of each step in the structure prediction procedure we employed in CASP11. We limited the targets to examine only those categorized for FM since our group performed well for FM targets. The new concepts we applied in CASP11 were coarse-grained residue-environment and helix-helix interaction potentials, which performed better than existing residue-pair or atom-pair knowledge-based potentials in considering multi-body interactions. Multi-body contact potentials, such as four-body potentials, have been developed in the past; however, PRESCO has technical advantages over such multi-body potentials. While previous multi-body contact potentials are limited to a single number of residues (e.g., four), PRESCO considers residue interactions of various different numbers in the reference sphere. Furthermore, a typical four-body potential requires interaction statistics of every four-residue combination; therefore, rare combinations may have an insufficient sample size. In contrast, PRESCO is based on pairwise amino acids found in similar residue environments, which allow for sufficient sampling of each residue type.

Table V

Computational Time of PRESCO and Other Scoring Functions

CASP11 Targets	Length (Residues)	GOAP	RWplus	dDFire	MRE	SDE
T0824	110	19.038s	0.972s	0.914s	12m20.59s	17m54.82s
T0804	202	19.845s	1.007s	0.984s	13m20.51s	26m24.04s
T0767	318	19.632s	1.201s	1.114s	20m45.83s	42m43.92s
T0827	407	19.452s	1.338s	1.217s	26m29.40s	56m33.27s

The times shown are for processing one structure model of the CASP targets. The computational times were measured on a Linux machine with Intel Core i7-920 2.67 GHz CPU and 20 GB RAM.

The model selection step went very well, for which we employed the PRESCO residue environment score and the helix interaction potential. According to the current analysis, it was shown that these two scores performed better than two existing scores, DFIRE and GOAP. In particular, we were surprised to see that the helix potential worked with a level of accuracy comparable to PRESCO.

The overall refinement step did not work as well. During CASP11, we believed that the models were refined because improvement of DFIRE energy was observed. However, it turned out that in many cases the improvement was small in terms of the evaluation scores used by the assessors. The lone exception was Molprobit, which was improved by the side-chain rebuilding with Oscar-Star for many models and remained as improved or no-change for 44.4% of the models after the structure relaxation.

The structure relaxation step by MD did not work well. In CASP11, our group was ranked among the best in the model refinement category according to the assessors' presentation in the CASP11 evaluation meeting (<http://www.predictioncenter.org/casp11/docs.cgi?view=presentations>). For the refinement category, our group employed an MD-based refinement procedure that is similar to what the Feig group has used,⁵⁴ except that we used an implicit solvent model in running MD to reduce the computational cost. In retrospect, we should have applied the same refinement procedure for both the FM and template based modeling (TBM) as we used on the refinement targets. Thus, development of an effective and computationally reasonable refinement procedure remains as an important future goal for our group.

ACKNOWLEDGMENT

The authors are grateful to Andriy Kryshchak for providing us with the GDT_TS, ContS, tenS, and QCS scores for the CASP11 models. The authors thank Andrzej Kolinski for providing the CABS protein modeling codes. The authors thank Lenna Peterson and Joshua McGraw for proofreading the manuscript.

REFERENCES

- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res* 2000;28:235–242.
- Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, Baker D. Principles for designing ideal protein structures. *Nature* 2012;491:222–227.
- Sippl MJ. Knowledge-based potentials for proteins. *Curr Opin Struct Biol* 1995;5:229–235.
- Skolnick J, Jaroszewski L, Kolinski A, Godzik A. Derivation and testing of pair potentials for protein folding. When Is the Quasi-chemical Approximation Correct? *Protein Sci* 1997;6:676
- Skolnick J. In quest of an empirical potential for protein structure prediction. *Curr Opin Struct Biol* 2006;16:166–171.
- Yuan C, Chen H, Kihara D. Effective inter-residue contact definitions for accurate protein fold recognition. *BMC Bioinformatics* 2012;13:292
- Lu M, Dousis AD, Ma J. OPUS-PSP: an orientation-dependent statistical all-atom potential derived from side-chain packing. *J Mol Biol* 2008;376:288–301.
- Zhang J, Zhang Y. A novel side-chain orientation dependent potential derived from random-walk reference state for protein fold selection and structure prediction. *PloS One* 2010;5:e15386.
- Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci* 2002;11:2714–2726.
- Gniewek P, Leelananda SP, Kolinski A, Jernigan RL, Kloczkowski A. Multibody coarse-grained potentials for native structure recognition and quality assessment of protein models. *Proteins* 2011;79:1923–1929.
- Munson PJ, Singh RK. Statistical significance of hierarchical multibody potentials based on Delaunay tessellation and their application in sequence-structure alignment. *Protein Sci* 1997;6:1467–1481.
- Sanchez-Gonzalez G, Kim JK, Kim DS, Garduno-Juarez R. A beta-complex statistical four body contact potential combined with a hydrogen bond statistical potential recognizes the correct native structure from protein decoy sets. *Proteins* 2013; 81:1420–1433.
- Kim H, Kihara D. Detecting local residue environment similarity for recognizing near-native structure models. *Proteins* 2014; 82: 3255–3272.
- Chakravarty S, Varadarajan R. Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure* 1999; 7:723–732.
- Tan YH, Huang H, Kihara D. Statistical potential-based amino acid similarity matrices for aligning distantly related protein sequences. *Proteins* 2006; 64:587.
- Engel DE, DeGrado WF. Alpha-alpha linking motifs and interhelical orientations. *Proteins* 2005;61:325–337.
- Kurochkina N. Helix-helix interactions and their impact on protein motifs and assemblies. *J Theoretical Biol* 2010;264:585–592.
- Dalton JA, Michalopoulos I, Westhead DR. Calculation of helix packing angles in protein structures. *Bioinformatics* 2003;19:1298–1299.
- Walther D, Springer C, Cohen FE. Helix-helix packing angle preferences for finite helix axes. *Proteins* 1998;33:457–459.
- Bowie JU. Helix packing angle preferences. *Nat Struct Biol* 1997;4: 915–917.
- Kallblad P, Dean PM. Backbone-backbone geometry of tertiary contacts between alpha-helices. *Proteins* 2004;56:693–703.
- Lee J, Im W. Implementation and application of helix-helix distance and crossing angle restraint potentials. *J Comput Chem* 2007;28: 669–680.
- Woetzel N, Karakas M, Staritzbichler R, Muller R, Weiner BE, Meiler J. BCL::Score-knowledge based energy potentials for ranking protein models represented by idealized secondary structure elements. *PloS One* 2012;7:e49242.
- Kolinski A. Protein modeling and structure prediction with a reduced representation. *Biochim Polym Acta* 2004;51:349.
- Zemla A. LGA: a method for finding 3D similarities in protein structures. *Nucleic Acids Res* 2003;31:3370.
- Mariani V, Biasini M, Barbato A, Schwede T. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* 2013;29:2722–2728.
- Kinch LN, Wrabl JO, Krishna SS, Majumdar I, Sadreyev RI, Qi Y, Pei J, Cheng H, Grishin NV. CASP5 assessment of fold recognition target predictions. *Proteins* 2003;53(Suppl 6):395
- Cong Q, Kinch LN, Pei J, Shi S, Grishin VN, Li W, Grishin NV. An automatic method for CASP9 free modeling structure prediction assessment. *Bioinformatics* 2011;27:3371–3378.

29. Shi S, Pei J, Sadreyev RI, Kinch LN, Majumdar I, Tong J, Cheng H, Kim BH, Grishin NV. Analysis of CASP8 targets, predictions and assessment methods. *Database* 2009;2009:bap003.
30. Chen VB, Arendall WB, III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Section D* 2010;66(Pt 1):12–21.
31. Liang S, Zhou Y, Grishin N, Standley DM. Protein side chain modeling with orientation-dependent atomic force fields derived by series expansions. *J Comput Chem* 2011;32:1680–1686.
32. Tomii K, Kanehisa M. Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins. *Protein Eng* 1996;9:27–36.
33. Kawashima S, Kanehisa M. AAindex: amino acid index database. *Nucleic Acids Res* 2000;28:374.
34. Rykunov D, Fiser A. New statistical potential for quality assessment of protein models and a survey of energy functions. *BMC Bioinform* 2010;11:128.
35. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 1992;89:10915–10919.
36. Qu CX, Lai LH, Xu XJ, Tang YQ. Phyletic relationships of protein structures based on spatial preference of residues. *J Mol Evol* 1993;36:67.
37. Qian B, Goldstein RA. Optimization of a new score function for the generation of accurate alignments. *Proteins* 2002;48:605.
38. Kolaskar AS, Kulkarni-Kale U. Sequence alignment approach to pick up conformationally similar protein fragments. *J Mol Biol* 1992;223:1053.
39. Manavalan P, Ponnuswamy PK. Hydrophobic character of amino acid residues in globular proteins. *Nature* 1978;275:673–674.
40. Manavalan P, Ponnuswamy PK. A study of the preferred environment of amino acid residues in globular proteins. *Arch Biochem Biophys* 1977;184:476–487.
41. Karlin S, Zhu ZY, Baud F. Atom density in protein structures. *Proc Natl Acad Sci U S A* 1999;96:12500–12505.
42. Kihara D. The effect of long-range interactions on the secondary structure formation of proteins. *Protein Sci* 2005;14:1955.
43. Zhong L, Johnson WC, Jr. Environment affects amino acid preference for secondary structure. *Proc Natl Acad Sci U S A* 1992;89:4462–4465.
44. Minor DL, Jr, Kim PS. Context-dependent secondary structure formation of a designed protein sequence. *Nature* 1996;380:730–734.
45. Feng Y, Kloczkowski A, Jernigan RL. Four-body contact potentials derived from two protein datasets to discriminate native structures from decoys. *Proteins* 2007;68:57–66.
46. Summa CM, Levitt M, Degrado WF. An atomic environment potential for use in protein structure prediction. *J Mol Biol* 2005;352:986–1001.
47. Huang ES, Subbiah S, Levitt M. Recognizing native folds by the arrangement of hydrophobic and polar residues. *J Mol Biol* 1995; 252:709–720.
48. Mooney SD, Liang MH, DeConde R, Altman RB. Structural characterization of proteins using residue environments. *Proteins* 2005; 61: 741–747.
49. Wang G, Dunbrack RL, Jr. PISCES: recent improvements to a PDB sequence culling server. *Nucleic Acids Res* 2005;33(Web Server issue):W94–W98.
50. Peterson LX, Kang X, Kihara D. Assessment of protein side-chain conformation prediction methods in different residue environments. *Proteins* 2014;82:1971–1984.
51. Brooks BR, Brooks CL, III, Mackerell AD, Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. *J Comput Chem* 2009;30:1545–1614.
52. Zhou H, Skolnick J. GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. *Biophys J* 2011;101:2043–2052.
53. Yang Y, Zhou Y. Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins* 2008;72:793–803.
54. Mirjalili V, Noyes K, Feig M. Physics-based protein structure refinement through multiple molecular dynamics trajectories and structure averaging. *Proteins* 2014;82(Suppl 2):196–207.