

# Structure similarity measure with penalty for close non-equivalent residues

Ruslan I. Sadreyev<sup>1</sup>, ShuoYong Shi<sup>2</sup>, David Baker<sup>3</sup>, and Nick V. Grishin<sup>1,2,\*</sup>

<sup>1</sup>Howard Hughes Medical Institute

<sup>2</sup>Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390-9050, USA

<sup>3</sup>Department of Biochemistry, University of Washington, J Wing, Health Sciences Building, Box 357350, Seattle, WA 98195  
Associate Editor: Prof. Burkhard Rost

## ABSTRACT

**Motivation:** Recent improvement in homology-based structure modeling emphasizes the importance of sensitive evaluation measures that help identify and correct modest distortions in models compared to the target structures. GDT\_TS, otherwise a very powerful and effective measure for model evaluation, is still insensitive to and can even reward such distortions, as observed for remote homology modeling in the latest Comparative Assessment of Structure Prediction, CASP8.

**Results:** We develop a new measure that balances GDT\_TS reward for the closeness of equivalent model and target residues (“attraction” term) with the penalty for the closeness of non-equivalent residues (“repulsion” term). Compared to GDT\_TS, the resulting score, TR, is much more sensitive to structure compression both in real remote homologs and in CASP models. TR is correlated yet different from other measures of structure similarity. The largest difference from GDT\_TS is observed in models of mid-range quality based on remote homology modeling.

**Availability:** The script for TR calculation is included in supplementary data. TR scores for all server models in CASP8 are available at <http://prodata.swmed.edu/CASP8>.

**Contact:** [grishin@chop.swmed.edu](mailto:grishin@chop.swmed.edu)

**Supplementary information:** All scripts and numerical data are available for download at [ftp://iole.swmed.edu/pub/tr\\_score/](ftp://iole.swmed.edu/pub/tr_score/)

## 1 INTRODUCTION

The current improvement of methods for sequence-based structure prediction (Kopp, et al., 2007; Shi, et al., 2008) leads to the increasing importance of accurate and biologically relevant measures of structural model quality. These measures are essential not only as criteria for benchmarking of method performance, but also as standards for the development of new, more powerful approaches to protein structure modeling. Any measure rewards certain aspects of model quality. Using the measure to guide method’s design inevitably emphasizes these aspects in the produced models.

Among various measures of model quality, GDT\_TS (Global Distance Test Total Score) is widely considered one of the most informative and robust. Proposed in a seminal paper (Zemla, et al., 2001) as a benchmarking criterion at early stages of CASP (Comparative Assessment of Structure Prediction), this measure has been successfully used by many independent assessor groups (Kinch, et al., 2003; Kopp, et al., 2007; Wang, et al., 2005; Zemla, et al., 2001). In brief, GDT\_TS is based on several target-model superpositions, each maximizing the number of equivalent target and model residues separated by a distance shorter than a given distance cutoff. These maximized numbers are normalized by the chain length and then averaged over the superpositions, producing the final score. Thus, GDT\_TS is a reward for placing as many model residues as possible in a vicinity of their target counterparts. This approach, although extremely effective in most cases, may still be insensitive to more modest deviations of models from native-like structure. As an example, in the models based on remote sequence homologs GDT\_TS does not penalize and may even reward unrealistic placement of several model residues in a close vicinity of a single target residue (Aloy, et al., 2003). When target and its closest homolog of known structure share only remote similarity, it is hard to predict the correct position of model residues based on the homolog as a template. In this situation, placing model residues closer to each other may improve the chances of some of them being rewarded for closeness to the target. Noted in previous CASPs (Aloy, et al., 2003), this effect has become more pronounced among recent models (Shi, et al., 2008) (see Fig.1). The last years have brought a significant improvement in the accuracy of remote homology modeling. As demonstrated by recent CASPs (Kopp, et al., 2007; Shi, et al., 2008), the current methods produce much more accurate models whose quality often differs only by a small GDT\_TS margin; thus optimizing a method for higher GDT\_TS might improve the apparent method’s standing, even at the cost of less realistic models. As a consequence, more fine-grained evaluation is often needed to distinguish between alternative models of reasonable quality.

Most of the existing approaches to structure comparison reward similar regions but do not explicitly penalize dissimilar regions. Taking an analogy with physical forces, these approaches concentrate on “attraction” but not on “repulsion”. It might have been

\*To whom correspondence should be addressed.

©2009 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

reasonable a few years ago, when the quality of structure predictions was much poorer, and rewarding for the spatial closeness of structure regions to the target would suffice to discriminate between models. It was important to detect any positive feature of a model, since there were more negatives about a model than positives. Today, many models reflect structures well. When the positives start to outweigh the negatives, it becomes important to pay attention to the negatives. Here we introduce an explicit "repulsion" component into the GDT\_TS score, which penalizes for incorrect pairing of non-equivalent residues in the compared structures. We show that unlike GDT\_TS, the new measure is sensitive to mild structure compression and thus may be a valuable tool in discriminating biologically unrealistic structure predictions.

## 2 METHODS

**Compression of real homologous domains.** We choose pairs of SCOP domain representatives that are confident homologs but do not have a strong structure similarity. From the total of ~100,000 ASTRAL40 domain pairs that share the same superfamily, we select two subsets based on different definitions of marginal structure similarity. First set includes ~27,000 pairs with lower-range but still significant DALI Z-scores, DALI Z between 2.0 and 5.0. Second set includes the lower third of all domain pairs ranked by DALI Z-score (~30,000 pairs, DALI Z between 2.0 and 5.8). In each pair, the set of equivalent residues is chosen according to the blocks (capital letters) of confidently aligned positions in DALI alignment. In separate experiments, each of the domains was compressed (compression ratio of 0.9 to 1.0), followed by the calculation of GDT\_TS and TR on the set of equivalent residues.

**Compression of the CASP fold recognition models.** CASP targets and server models are downloaded from <http://predictioncenter.org/casp8>. We further processed target and model structures, split them into domains, and assigned them to standard categories by prediction difficulty as described at <http://prodata.swmed.edu/CASP8>. We use target-model pairs for 25 target domains in the fold recognition category, subject each model to varying degrees of compression, and calculate corresponding GDT\_TS and TR scores. As an example in Fig. 3B, we show the plot for models by Robetta server, which does not introduce significant distortions of local geometry.

**Perturbation of the torsion angles.** We perturb torsion angles phi, psi in fold recognition models with no chain breaks. We chose Robetta models as they have very few chain breaks: among 125 FR models by Robetta, only five models for the same target have a break. To every original phi/psi value we add a randomly distributed Gaussian term with a mean of 0 and a certain standard deviation. The original structure is modified by consequent rotation of backbone segments according to the new angle values, keeping all other geometric parameters intact.

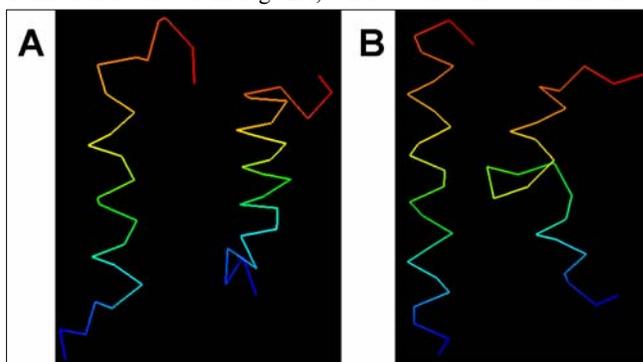
**Correlation between TR and the existing scores.** For each CASP8 target, we perform GDT\_TS-based ranking of first models submitted by servers and select top 10 models. We average target-model similarity scores for those models and plot TR against other measures, with each target represented by a single point. To make the scale of DALI Z scores compatible to TR, we normalize them by DALI Z-scores for the comparison of target domain to itself.

## 3 RESULTS

Analyzing structure predictions produced by automated servers in the recent CASP, CASP8 (Shi, et al., 2008), we observe a number of models that are assigned high GDT\_TS scores but include unrealistically compressed regions (Fig. 1). These regions often show

violations of secondary structure geometry, shorter distances between C-alpha atoms, sidechain clashes, and distortion of hydrogen bond patterns.

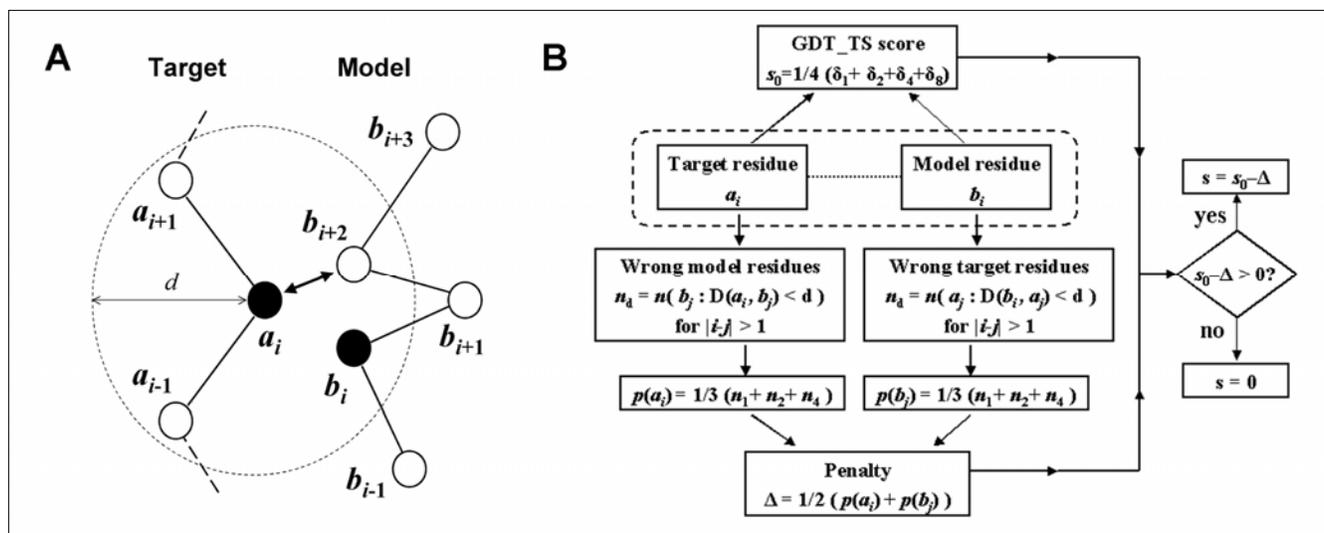
In fold recognition modeling, when the closest homolog with 3D structure is relatively distant from the target, it is often difficult to predict the correct location of target residues, even if the general positioning of secondary structure elements can be inferred from the sequence-based alignment. Such situations typically involve ambiguity of angles between secondary structure elements, register shifts in alpha helices and beta strands, and unknown loop conformations. By definition, GDT\_TS rewards relatively close positioning of equivalent C<sub>α</sub> atoms in model and target, but does not penalize situations where two or more model residues are close to the same target residue. Therefore, a GDT\_TS-trained automatic predictor may in some cases choose to concentrate C<sub>α</sub> atoms in a smaller volume, which increases the probability for the target residue to have the correct model residue in the vicinity, even if the correct residue is unknown. A schematic example of such a conformation is shown in Fig. 2A, where both correct and incorrect



model C<sub>α</sub> atoms are located closely to the target C<sub>α</sub>.

**Fig. 1.** Examples of non-native structure geometries in CASP8 models. In A and B, left is a helix from a target structure, right is the same region modeled by one of the servers. A. Unrealistic compression of a helix along the axis. (model helix is rotated for clarity). In the model, the helix is positioned at a different angle; its compression improves GDT\_TS. B. "Smart" compression: C-alpha distance restrictions are not violated, but helix distortion concentrates residues around the same point.

Such deviations from native protein geometry may be loosely viewed as an implementation of the minimax principle from game theory: select the conformation that minimizes the maximal possible loss in the case of failure. For example, for an ambiguous helix register, putting residues closer to each other may produce a positive contribution to GDT\_TS even if the register is wrong (Fig. 1). In the case of unknown loop conformation, predicting an unrealistic collapsed loop bears less risk of large GDT\_TS loss than predicting an extended (and locally correct) conformation that may deviate further from the target if overall orientation is wrong. More surprisingly, even a simplistic uniform contraction of the model can in many cases produce higher GDT\_TS (Fig. 2, see section 3.3).



**Fig. 2.** Calculation of penalty for close positioning of non-equivalent residues in target and model. A. Schema of superposition for a pair of equivalent target and model residues ( $a_i$  and  $b_i$ , black circles). To calculate the penalty, we count model residues located close to the target residue  $a_i$ , excluding its counterpart  $b_i$  and adjacent residues  $b_{i-1}$ ,  $b_{i+1}$ . In this example, there is one residue located within a given distance  $d=4\text{\AA}$  from the target residue (marked with an arrow). B. We average the numbers of such “incorrect” residues for three different distance cutoffs  $d=1, 2, 4\text{\AA}$ , and combine these averages for both target and model residues. The resulting penalty is subtracted from the pair’s GDT\_TS score. If the penalty is higher than the GDT\_TS, the pair’s contribution is set to zero.

### 3.1 Score for model quality with explicit repulsion term

To develop a measure sensitive to the observed structure compression, we complement GDT\_TS with an explicit “repulsion” term that penalizes for close spatial placement of incorrect residue pairs (Fig. 2A). We call this score TR (Total score with Repulsion). The score is calculated as follows (Fig. 2B).

- (1) Superimpose model with target using LGA (Zemla, 2003) in the sequence-dependent mode, maximizing the number of aligned residue pairs within distance cutoffs of  $d = 1, 2, 4$ , and  $8\text{\AA}$ .
- (2) For each pair of target and model residues ( $a_i, b_i$ ), calculate a GDT\_TS-like score:  $s_0 = 1/4 (\delta_1 + \delta_2 + \delta_4 + \delta_8)$ , where  $\delta_d = 1$  if  $C_\alpha-C_\alpha$  distance  $< d\text{\AA}$  and  $\delta_d = 0$  otherwise.
- (3) In LGA superposition for  $d = 4\text{\AA}$ , consider individual aligned residues in both structures. For each residue  $R$ , choose residues in the other structure that are spatially close to  $R$ , excluding the residue aligned with  $R$  and its immediate neighbors in the chain. Count numbers of such residues with  $C_\alpha-C_\alpha$  distance to  $R$  within cutoffs of 1, 2, and  $4\text{\AA}$ . (As opposed to GDT\_TS, we do not use the cutoff of  $8\text{\AA}$  as too inclusive: in native proteins, this distance is not prohibited for different residues of the same chain.)
- (4) The average of these counts defines the penalty assigned to a given residue  $R$ :  $p(R) = 1/3 (n_1 + n_2 + n_4)$ .
- (5) Finally, for each aligned residue pair ( $a_i, b_i$ ), the average of penalties for each residue  $\Delta = 1/2 [p(a_i) + p(b_i)]$  is weighted and subtracted from the GDT\_TS score for this pair. The final score is prohibited from being negative, in order to avoid rewarding shorter models limited to only the confident structure core:

$$s = \max(s_0 - w\Delta, 0)$$

Among tested values of weight  $w$ , we find that  $w=1.0$  produces the scores that are most consistent with the evaluation of model abnormalities by human experts.

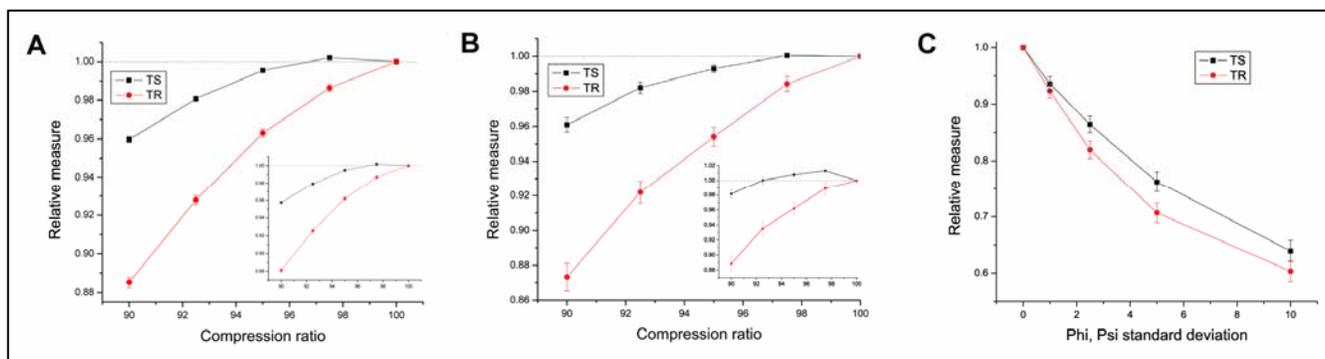
The score for the two compared structures is calculated as the sum of scores for individual residue pairs, normalized by target chain length  $L$ :  $TR = (1/L) \sum s_i$ .

The cutoff of  $4\text{\AA}$  for the used LGA superposition was set after testing multiple values around typical GDT\_TS distance cutoffs. This value produced the scoring of CASP8 models that was most consistent with the manual expert assessment of the model quality.

### 3.2 The new score has improved sensitivity to model distortions

As an example of effects of mild structure distortion on GDT\_TS and TR, we observe behavior of these scores in two simplistic experiments. In each experiment we calculate the scores on a pair of structures and then perturb one of the structures. We vary the degree of the perturbation and observe corresponding changes in GDT\_TS and TR scores. In the first experiment, we uniformly contract one of the structures towards its center of mass, so that its radius of gyration decreases and its residues become slightly closer to each other (Fig. 3A,B). In the second experiment, we perturb torsion angles  $\phi, \psi$  by adding a random Gaussian term to each angle in the structure (Fig. 3C).

By design, TR score is expected to bring improvement in sensitivity at medium levels of modeling difficulty, where sequence alignment to a homolog template is possible yet non-trivial. TR should not be better than GDT\_TS either for models based on close homology where model-target misalignments are



**Fig. 3.** Effects of structure distortion on GDT\_TS and TR scores. A,B: One of the two compared proteins is uniformly compressed, and dependency of the scores on the degree of compression is shown relative to the score for intact structures. A. Pairs of remotely homologous protein domains. Average relative scores are based on the set domain pairs that share the same SCOP superfamily and have DALI Z-score between 2 and 5. Inset: the same dependency for a different definition of remote homologs: lower third (by DALI Z score) of all domain pairs that share the same SCOP superfamily and have DALI Z > 2.0. B. Targets and models in the fold recognition category of the recent CASP, CASP8. As an example, average relative scores are based on models by Robetta server. The dependency is shown for the whole set of FR domain models by Robetta (125 models). Inset: dependency for the subset of FR domain models whose GDT\_TS grows with compression (48 models). C. Random perturbation of torsion angles (phi, psi) in Robetta models of FR domains. Dependency of relative score values on the standard deviation of added random Gaussian perturbations.

rare, or for template-free models where even a general fold prediction is a challenge. Therefore, in our experiments we concentrate on cases of clear yet distant template-target homology.

In the first experiment (Fig. 3A) we use real pairs of remotely homologous protein domains. Among all pairs of ASTRAL40 (Chandonia, et al., 2004) representatives that share the same superfamily in SCOP classification (Andreeva, et al., 2008), we choose ~27,000 pairs with marginal structure similarity according to DALI Z-scores (Holm and Sander, 1993) and compress one of the domains in each pair. On average, GDT\_TS does not decrease until compression reaches ~5% (Fig. 3A). Moreover, in 40% cases GDT\_TS actually increases at mild compression levels. In contrast, TR is consistently reduced on compressed structures, with the rate of reduction much higher than for GDT\_TS.

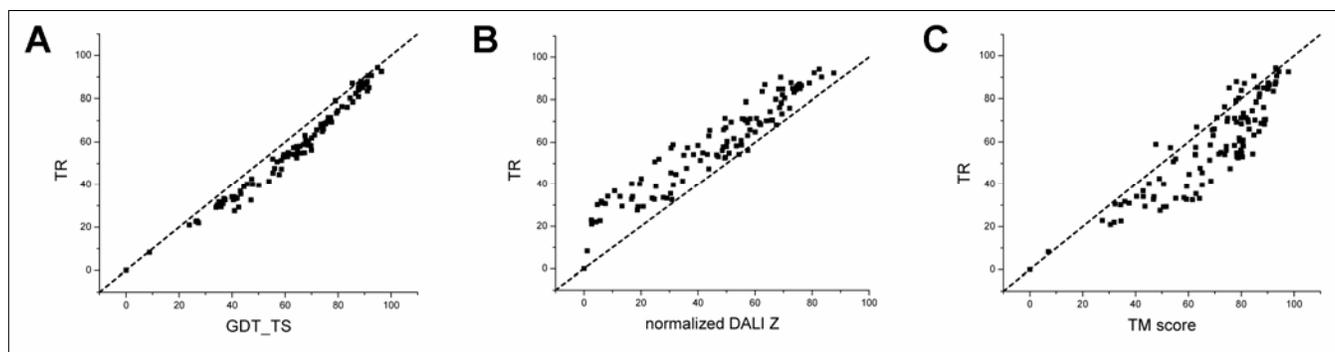
We perform the same experiment (Fig. 3B) on pairs of target and model domains from automatic CASP8 server predictions in the fold recognition category, as defined by our analysis (Shi, et al., 2008). Similar to the first dataset, mild compression leads to GDT\_TS increase in 40% cases. Fig. 3B shows the average degree of GDT\_TS gain for these models, as well as for the whole model set. At the same time, TR penalizes the compression much more effectively. Full sets of GDT\_TS and TR scores for all server models of CASP8 are available at <http://prodata.swmed.edu/CASP8>. When we perturb torsion angles in the same set of models, TR decreases with the variance of the perturbation significantly faster than GDT\_TS, although the difference is somewhat less pronounced (Fig. 3C).

### 3.3 Correlation with existing measures for structure similarity

We compared the new score to other similarity scores based on the set of first server models produced for all CASP8 targets. Fig. 4 shows correlation plots for TR vs GDT\_TS, DALI Z-scores, and TM scores (Zhang and Skolnick, 2005). The table of correlation coefficients of TR with these and other measures is included in the supplement.

It is clear that GDT\_TS and TR scores are well correlated (Fig. 4A), with Pearson correlation coefficient of 0.99. By design, TR is always equal or lower than GDT\_TS. Notably, the trend curve of the correlation is concave, so that the difference is more pronounced around the mid-range GDT\_TS. This range roughly corresponds to the targets for the categories of harder

**Fig. 4.** Correlation of the new score with existing scores for structure similarity. For each CASP8 target, average scores for the top 10 first server models are plotted. A. TR vs GDT\_TS. B. TR vs DALI Z-score (scaled by self Z-score). C. TR vs TM score.



comparative modeling and fold recognition, where models become less similar to targets and modeled residues are frequently placed nearby non-equivalent residues, which results in higher penalty by TR. For very low model quality (GDT\_TS below 30%) the reward is much lower, and penalty drops as well (Fig. 4A).

TR also shows general correlation with other similarity measures, although to a lesser degree. Correlation coefficient with conceptually different contact-based DALI Z-scores (normalized to have a compatible scale) is relatively high (0.95). Interestingly, TM score (Zhang and Skolnick, 2005) based on a concept similar to GDT\_TS is less correlated with TR than DALI Z-scores (Fig. 4C, correlation coefficient of 0.88), but the trend curve's concave shape is similar to the plot of TR against GDT\_TS.

In conclusion, we develop a new measure for protein structure similarity with explicit repulsion term that penalizes for spatially close positioning of non-equivalent residues. This measure improves sensitivity of GDT\_TS to moderate structure distortion and has a potential value for the assessment of structure similarity and homology-based structure modeling.

## ACKNOWLEDGEMENTS

We would like to thank Yang Zhang for stimulating discussions and critical reading of the manuscript and Bong-Hyun Kim for sharing DALI Z-scores for ASTRAL domain pairs. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high-performance computing resources.

*Funding:* This work was supported by NIH grant GM67165 to NVG.

## REFERENCES

- Aloy, P., Stark, A., Hadley, C. and Russell, R.B. (2003) Predictions without templates: new folds, secondary structure, and contacts in CASP5, *Proteins*, **53 Suppl 6**, 436-456.
- Andreeva, A., Howorth, D., Chandonia, J.M., Brenner, S.E., Hubbard, T.J., Chothia, C. and Murzin, A.G. (2008) Data growth and its impact on the SCOP database: new developments, *Nucleic Acids Res*, **36**, D419-425.
- Chandonia, J.M., Hon, G., Walker, N.S., Lo Conte, L., Koehl, P., Levitt, M. and Brenner, S.E. (2004) The ASTRAL Compendium in 2004, *Nucleic Acids Res*, **32**, D189-192.
- Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J Mol Biol*, **233**, 123-138.
- Kinch, L.N., Wrabl, J.O., Krishna, S.S., Majumdar, I., Sadreyev, R.I., Qi, Y., Pei, J., Cheng, H. and Grishin, N.V. (2003) CASP5 assessment of fold recognition target predictions, *Proteins*, **53 Suppl 6**, 395-409.
- Kopp, J., Bordoli, L., Battey, J.N., Kiefer, F. and Schwede, T. (2007) Assessment of CASP7 predictions for template-based modeling targets, *Proteins*, **69 Suppl 8**, 38-56.
- Shi, S., Pei, J., Sadreyev, R.I., Kinch, L.N., Majumdar, I., Tong, J., Cheng, H., Kim, B.-H. and Grishin, N.V. (2008) Analysis of

CASP8 targets and server predictions by the Grishin lab (2008) on World Wide Web <http://prodata.swmed.edu/CASP8>.

Wang, G., Jin, Y. and Dunbrack, R.L., Jr. (2005) Assessment of fold recognition predictions in CASP6, *Proteins*, **61 Suppl 7**, 46-66.

Zemla, A. (2003) LGA: A method for finding 3D similarities in protein structures, *Nucleic Acids Res*, **31**, 3370-3374.

Zemla, A., Venclovas, M., Moulton, J. and Fidelis, K. (2001) Processing and evaluation of predictions in CASP4, *Proteins*, **Suppl 5**, 13-21.

Zhang, Y. and Skolnick, J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score, *Nucleic Acids Res*, **33**, 2302-2309.