PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

CAD-score: A new contact area difference-based function for evaluation of protein structural models

Kliment Olechnovič, Eleonora Kulberkytė, and Česlovas Venclovas*

Institute of Biotechnology, Vilnius University, Graičiūno 8, LT-02241 Vilnius, Lithuania

ABSTRACT

Evaluation of protein models against the native structure is essential for the development and benchmarking of protein structure prediction methods. Although a number of evaluation scores have been proposed to date, many aspects of model assessment still lack desired robustness. In this study we present CAD-score, a new evaluation function quantifying differences between physical contacts in a model and the reference structure. The new score uses the concept of residue–residue contact area difference (CAD) introduced by Abagyan and Totrov (J Mol Biol 1997; 268:678–685). Contact areas, the underlying basis of the score, are derived using the Voronoi tessellation of protein structure. The newly introduced CAD-score is a continuous function, confined within fixed limits, free of any arbitrary thresholds or parameters. The built-in logic for treatment of missing residues allows consistent ranking of models of any degree of completeness. We tested CAD-score on a large set of diverse models and compared it to GDT-TS, a widely accepted measure of model accuracy. Similarly to GDT-TS, CAD-score showed a robust performance on single-domain proteins, but displayed a stronger preference for physically more realistic models. Unlike GDT-TS, the new score revealed a balanced assessment of domain rearrangement, removing the necessity for different treatment of single-domain, multi-domain, and multi-subunit structures. Moreover, CAD-score makes it possible to assess the accuracy of inter-domain or inter-subunit interfaces directly. In addition, the approach offers an alternative to the superposition-based model clustering. The CAD-score implementation is available both as a web server and a standalone software package at http://www.ibt.lt/bioinformatics/cad-score/.

Proteins 2013; 81:149–162.

Key words: protein structure; model accuracy; residue–residue contacts; physical realism; domain rearrangement; multi-subunit structures; CASP; GDT.

INTRODUCTION

Effective assessment of protein structural models against the experimentally determined protein structure (the reference) is at the heart of development and objective comparison of protein structure prediction methods. It may seem that one-to-one correspondence of amino acids in a model and the reference structure should make such a task trivial. However, this impression is misleading. The task is complex and, despite the fact that many evaluation scores have been devised over the years, it continues to be an active area of research.

One of the earliest and best known scores is root mean square deviation (RMSD).[1] RMSD indicates the mean distance between the corresponding atoms in the two protein structures after their optimal rigid-body superposition. It is typically calculated for Cα atoms, but it can be applied to any subset of residue atoms. Although RMSD is a popular score, it is informative only if the differences are reasonably small and fairly equally distributed. The main disadvantage of RMSD is its sensitivity to large local deviations. Even few poorly modeled residues, which may be of little structural and/or biological importance (e.g., poorly structured protein termini or a flexible loop), may have a large impact on the resulting RMSD score. If different models include different number of residues, corresponding RMSD values may be entirely misleading as to the true accuracy of models. In particular, the inadequacy of RMSD for evaluation and ranking of very different and often incomplete protein

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: Howard Hughes Medical Institute; Grant number: 55005627.
*Correspondence to: Česlovas Venclovas, Institute of Biotechnology, Vilnius University, Graičiūno 8, LT-02241 Vilnius, Lithuania. E-mail: venclovas@ibt.lt
Received 5 April 2012; Revised 9 August 2012; Accepted 25 August 2012
Published online 30 August 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24172

models became apparent during early CASP experiments,[2,3] established for monitoring the state-of-the-art in protein structure prediction.

Thus, CASP experiments revealed a need for scores that would be robust in a wide range of model accuracy and completeness. Global distance test (GDT)[4] was one of the scores developed to overcome shortcomings of RMSD. GDT identifies the largest subset of model residues (represented by their Cα atoms) that can be superimposed with the corresponding residues in the reference structure under a specific distance threshold. The overall model accuracy is summarized by GDT total score (GDT-TS), a single value derived by averaging the fractions of residues obtained in the four independent superpositions under 1, 2, 4, and 8 Å distance thresholds.[5] Due to multiple superpositions of different stringency, GDT-TS is able to rank models quite effectively in a wide range of accuracy. Unlike RMSD, GDT-TS rewards the good bits of the model without adding a penalty for the inaccurately modeled regions. As a result, GDT-based benchmarking promotes methods that attempt to construct not only the most accurate, but also the most complete structural models. Other scores, similar to GDT-TS, include MaxSub[6] and TM-score.[7] MaxSub, just like GDT-TS, aims at identifying the largest subset of residues that can be superimposed under specific distance threshold. However, in contrast to GDT-TS, MaxSub uses only a single 3.5 Å distance threshold. This makes MaxSub somewhat less robust in ranking models, in particular those of lower accuracy.[7] TM-score considers all the corresponding residue pairs. It uses the distance-dependent weighting scheme, which reduces the contribution from significantly deviating residue pairs. In addition, the distance-dependent down-weighting varies with the protein size, making the score less size-dependent in comparison with either GDT-TS or MaxSub. Yet, similar to MaxSub, TM-score is derived from a single superposition. When size dependence is not an issue (e.g., evaluating models against the same reference) multiple superpositions as implemented in GDT-TS offer an obvious advantage.

Not surprisingly, GDT-TS has *de facto* become the central score in the automated reference-based model evaluation during CASP experiments.[8] However, despite its common use, GDT-based scoring is not without weaknesses. Since GDT-TS is based on the rigid-body superposition, it performs poorly on multi-domain proteins. A slight change in the mutual domain orientation may be biologically irrelevant, yet it may strongly affect the GDT-TS score. Another GDT weakness is that it uses only Cα atoms and therefore lacks information about the correctness of residue side chain modeling. However, this is an important component in benchmarking high accuracy comparative modeling or protein structure refinement methods. One additional and perhaps the most disconcerting issue is the lack of direct relationship between the GDT-TS score and the physicochemical characteristics of a protein model. A model having unrealistic features such as extensive interatomic clashes or systematic structural distortions may still receive a favorable GDT-TS score.[9–11] The same limitations are characteristic of similar scores, MaxSub and TM-score.

In attempt to address some of these issues, a number of modifications to the GDT-TS score have been proposed. Some of them were directed at a better resolution of higher accuracy models. Thus, GDT-HA,[12] a more stringent version of GDT-TS, uses distance thresholds half the size of those for GDT-TS. GDC, another modified score, is capable of including different thresholds and different subsets of residue atoms.[13,14] To make the score mindful of steric clashes, the inclusion of repulsion term into GDT-TS was proposed.[9] However, each modification addresses only one of several limitations of the GDT-TS score.

Therefore, there is a clear need to have the best features of GDT-TS (robustness over the wide model accuracy range and the ability to compare models of different degree of completeness) combined with a more physically meaningful representation of protein structure. Globular proteins fold into specific 3D structures that are defined by residue–residue interactions, which are reflected by physical contacts. Therefore, it seems that contacts might be well-suited for quantifying deviations in a model with respect to the reference structure. Besides, the comparison of contacts does not require structure superposition with all the associated caveats. Indeed, a number of scores that use the concept of residue–residue contacts have been proposed.[11,14–17] However, typically, "contacts" in these scores are represented by distances between Cα, Cβ, or all atoms within the arbitrarily specified threshold. Obviously, the physical meaning of contacts in such scores is lost. If only a single atom per residue (e.g. Cα) is used, important structural details are lost as well.

An interesting idea of using the explicit description of physical residue–residue contacts for model evaluation was introduced by Abagyan and Totrov.[18] They proposed to use the residue–residue contact area as the basis for comparing a model and the reference structure. Furthermore, they introduced a single-number score, contact area difference (CAD), as a measure of the overall model accuracy. CAD, as defined by Abagyan and Totrov, has a number of appealing features. It is continuous and threshold-free, works in a wide range of model accuracies, adequately penalizes domain, fragment, and side-chain rearrangements and captures essential geometrical characteristics of protein structure.[18]

However, the original CAD has some properties that make its use for evaluation of methods on a large scale (e.g. CASP experiments) problematic. First, CAD considers only residues common for both the model and the reference structure. It means that a complete model would be evaluated against the complete reference structure, while a modeled short fragment would be evaluated against the corresponding reference fragment. In other words, the exact choice of the reference depends on the completeness of the model. This can hardly be considered an objective mode for benchmarking different methods. Second, the normalizing CAD term includes inter-residue contact areas not only of the reference structure but also of the model. Although this is not expected to have a large impact on the total score, it nevertheless makes the CAD normalization model-specific.

Here, we introduce a new contact area difference-based score (CAD-score) for model evaluation against the reference structure. It combines the original CAD concept[18] with some of the ideas underlying the design of the GDT-TS score.[4,5] The new CAD-score uses a more accurate algorithm for the contact area calculations. Furthermore, the new algorithm resolves residue–residue contacts at the level of atoms, making it possible to consider subsets of residue atoms (main chain, side chain) separately. CAD-score has a way of treating missing residues in the model, and, therefore, similarly to GDT, may efficiently rank both complete and incomplete models. The new CAD-score is normalized against the contacts in the reference structure and always falls between 0 and 1. The software for the calculation and visual representation of CAD-score, contact maps, and the local deviation are available both as a web server and as a standalone package.

MATERIALS AND METHODS

Computation of residue–residue contact areas

Residue contact areas are derived using the protein structure tessellation approach. We use the tessellation referred to as the Voronoi diagram of 3D spheres[19] (also known as the additively weighted Voronoi diagram or the Apollonius diagram). Spheres in such a tessellation correspond to heavy atoms of van der Waals radii. Here we used van der Waals radii for protein heavy atoms derived by Li and Nussinov.[20] For each atom we can define the Voronoi cell, a set of all points closer to this particular atom than to any other atom. The Voronoi diagram of the protein structure is the set of Voronoi cells of all the heavy atoms of the protein. Two atoms are said to be Voronoi neighbors if their Voronoi cells share a common subset of points.

Interatomic contacts are derived from the Voronoi diagram of atoms based on the idea proposed by McConkey *et al.*[21] Neighboring protein atoms are defined as contacting each other if a water molecule cannot fit between them. Thus, the complete contact surface of an atom is represented by the sphere of the radius equal to the sum of van der Waals radius of the atom and the standard radius (1.4 Å) of a water molecule. We term it a contact sphere. Point p on the contact sphere of atom i belongs to the contact surface with atom j, if the following two conditions are satisfied: (1) i and j are Voronoi neighbors; (2) p is closer to j than to i or any other neighbor of i. If p is closer to i than to any neighbor of i, then it belongs to the solvent-accessible surface.

For a given atom, we construct its contact surfaces by intersecting a triangulated representation of the atom contact sphere with hyperboloids that correspond to the junctures of neighboring Voronoi cells. As a result, the entire contact sphere of an atom is unambiguously partitioned into contact areas and solvent accessible areas. Residue–residue contacts are constructed by simply grouping contacts between atoms of corresponding residues. The contact between C and N atoms forming the peptide bond of residues adjacent in sequence is not considered. Since contacts are resolved at the level of atoms, we can define contacts not only for the entire residue but also for various subsets of its atoms (e.g. main chain and side chain). Figure 1 illustrates how the combination of Voronoi cells and contact spheres is used to construct contact surfaces for an atom [Fig. 1(A)] and for a residue [Fig. 1(B)].

The tessellation of protein structure and interatomic contact areas are calculated using a modification of the algorithm developed for Voroprot.[22] The modified algorithm (details to be described elsewhere) features a significant increase in speed, achieved by the optimization of spatial search operations. In addition, the improved treatment of various input abnormalities allows the modified algorithm to handle even protein structures with a share of physically unrealistic features.

The CAD-score definition

We defined CAD-score based on the three main considerations: (1) contacts in the model should be evaluated according to the contacts in the reference structure (target); (2) any missing residues in the model should be treated in the same way as if none of their contacts were correctly predicted; (3) strong over-prediction (nonphysical overlap) of a particular contact should be equivalent to missing that contact entirely. The mathematical definition of CAD-score is presented below.

Let G denote the set of all the pairs of residues (i,j) that have a nonzero contact area $T_{(i,j)}$ in the target structure. Then for every residue pair $(i,j)\in G$ we calculate the contact area $M_{(i,j)}$ in the model. If the model has additional residues not present in the target, these residues are excluded from the calculation of contact areas. If some residue is present in the target, but is missing from the model, all the contact areas for that residue in the model are assigned zeroes.

For every residue pair $(i,j)\in G$, we can then define contact area difference as the absolute difference of contact

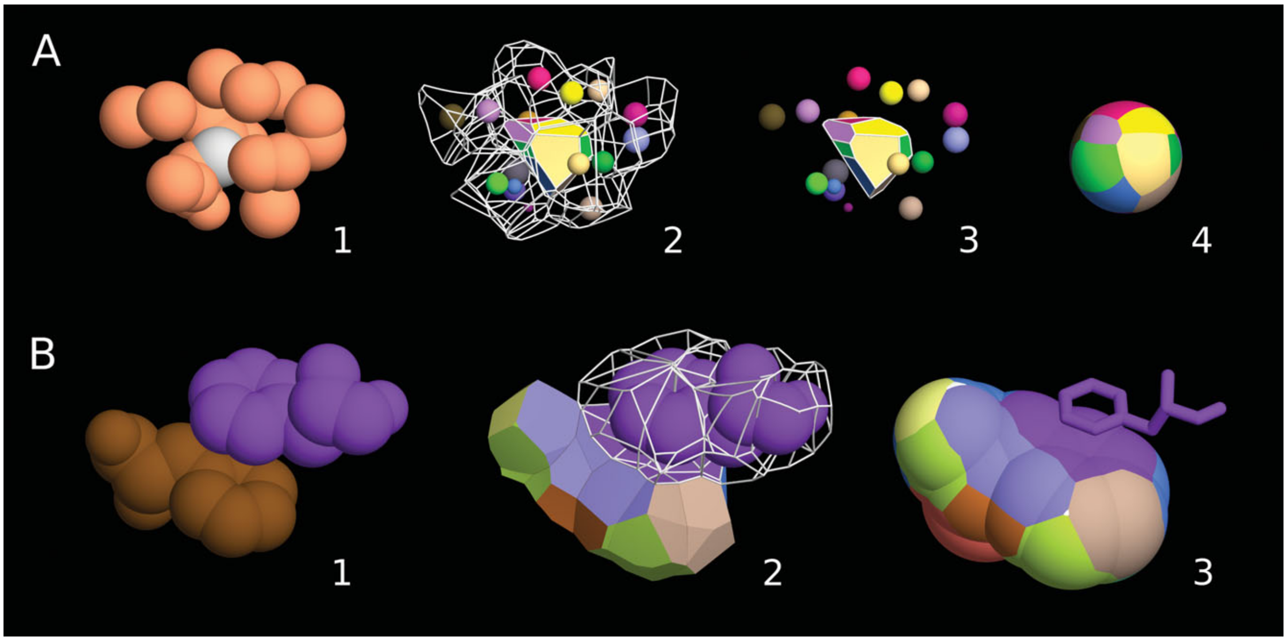

Figure 1

Illustration of the procedure for deriving contact surfaces for atoms (A) and residues (B). (**A**) Interatomic contacts: 1, the considered atom (grey) surrounded by neighboring atoms; 2, the Voronoi cell of the considered atom (solid) and neighboring Voronoi cells (wireframe); small colored spheres correspond to the same neighboring atoms shown as large spheres in 1; 3, the Voronoi cell with its faces colored according to the color of neighboring atoms; 4, interatomic contact surfaces mapped onto the contact sphere of the atom. (**B**) Inter-residue contacts: 1, two interacting phenylalanine residues in the space-filling representation; 2, Voronoi cells of the same residues; faces of one of the residues are colored according to the color of neighboring residues; 3, the map of inter-residue contact surfaces for one of the interacting residues.

areas between residues i and j in target T and in model M:

$$\mathrm{CAD}_{(i,j)} = |T_{(i,j)} - M_{(i,j)}| \tag{1}$$

To impose symmetrical treatment of over-prediction and under-prediction of the contact area, instead of the raw $\mathrm{CAD}_{(i,j)}$ value, we use bounded $\mathrm{CAD}_{(i,j)}$ defined as follows:

$$\mathrm{CAD}^{\mathrm{bounded}}_{(i,j)} = \min(\mathrm{CAD}_{(i,j)}, T_{(i,j)}) \tag{2}$$

CAD-score for the whole model is then defined as:

$$\text{CAD-score} = 1 - \frac{\sum_{(i,j)\in G} \mathrm{CAD}^{\mathrm{bounded}}_{(i,j)}}{\sum_{(i,j)\in G} T_{(i,j)}} \tag{3}$$

The sum in the numerator of Eq. (3) never exceeds the sum of all contact areas $T_{(i,j)}$ in the target structure. In other words, CAD-score defined by Eq. (3) is always within the [0,1] range. If model and target structures are identical, CAD-score = 1. At the other extreme, if not a single contact is reproduced with sufficient accuracy (there are no cases satisfying the condition: $CAD_{(i,j)} < T_{(i,j)}$), CAD-score = 0.

CAD-score variants

Our algorithm computes inter-residue contact areas at the resolution of individual atoms. Therefore, we can define contact area as well as contact area difference not only for the entire residue, but also for any subset of its atoms. In all cases contact areas are calculated with all atoms present, but if a subset of residue atoms is considered, only contact areas corresponding to this subset are retained. Here we use all residue atoms (A) and two standard subsets of atoms, main chain (M) and side chain (S), resulting in nine CAD-score variants (Table I). Three pairs of CAD-score variants (A-S and S-A, A-M and M-A, and S-M and M-S; gray background in Table I) are not entirely symmetric. For example, glycine does not have a side chain and therefore cannot form any S-A contacts, but it can form A-S contacts. Nevertheless, for

Table I

CAD-score Variants Based on Standard Subsets of Residue Atoms

	All atoms (A)	Side chain (S)	Main chain (M)
All atoms (A)	A-A	S-A	M-A
Side chain (S)	A-S	S-S	M-S
Main chain (M)	A-M	S-M	M-M

practical purposes these three pairs of CAD-score variants may be considered to be redundant. As a result, for standard subsets of residue atoms there are six non-redundant CAD-score variants that can be used to address different questions in evaluating models against the reference structure.

RESULTS

To test the properties of CAD-score and its effectiveness in evaluating and ranking models, we applied it to models obtained during CASP9, the ninth community-wide assessment of protein structure prediction methods.[23] CASP models are generated by a large array of different methods and, therefore, represent a wide range of accuracies. In addition, the set contains models of different degree of completeness including complete models, those missing a few residues and only short structural fragments. Moreover, models differ greatly by their physical plausibility. Some of them feature structural characteristics reminiscent of high resolution experimental structures, while some others have a number of unrealistic features such as steric clashes and strongly deviating covalent bond geometries. All these aspects of CASP models make them an excellent test set for an automatic reference-based model evaluation score as the set presents a serious challenge for objective and fair model ranking.[24] To have a representative and least redundant set, we only considered CASP9 models generated by automatic methods (servers) taking a single most confident (first) model per method for a given prediction target. Since CAD-score is an all-atom measure, we excluded from our analysis models produced by methods representing amino acid residues in a simplified or incomplete form. Models for one of the targets (T0629; the long tail fiber protein gp37 of the T4 bacteriophage) were also excluded. T0629 forms the needle-shaped parallel homotrimer, and considering the isolated single chain is both structurally and biologically meaningless.[25]

CAD-score is a robust measure for evaluating and ranking single-domain models

As a first step, we decided to compare CAD-score with GDT-TS, a standard CASP score that withstood the test of time and is generally recognized as the single most effective reference-based score.[17] To make an overall comparison of CAD-score and GDT-TS, we selected CASP9 models for individual domains ("assessment units" to be more precise) of prediction targets as defined by the assessors.[26] For the resulting diverse set of 8429 models, we compiled both GDT-TS and CAD scores. GDT-TS values were taken from the data archive of the Prediction Center (http://www.predictioncenter.org/) while different variants of CAD-score were calculated as described in Materials and Methods. The plots displaying the relationship between GDT-TS and six non-redundant CAD-score variants are shown in Figure 2.

It is evident that there is a strong correlation between GDT-TS and CAD-score values, which is surprising considering the different nature of scores. Notably, this is true not only for Pearson's correlation coefficient, which depends on the linear relationship between the two scores. Even better values in all cases are obtained for Spearman's rank correlation, which indicates the extent to which ranking by GDT-TS agrees with ranking by CAD-score without the assumption of the linear relationship between the two scores. In particular, three types of CAD-score ('all atoms–side chain' (A-S), 'side chain–side chain' (S-S), and 'all atoms–all atoms' (A-A)) show the strongest correlation [Fig. 2(A–C)]. For these three CAD-score variants Pearson's correlation coefficients are in the (0.91–0.94) range, and Spearman's rank correlation values are in the (0.93–0.95) range. The other three types of CAD-score, in particular, the variant based on 'main chain–main chain' (M-M) contacts, correlate somewhat weaker. We reasoned that the lower correlation to a large degree might be determined by the abundance of local M-M contacts that are not linked to the global topology of the structure. If this is true, the type of secondary structure should be a major factor. Indeed, when analyzed separately, the correlation for proteins rich in β-strands (many non-local M-M contacts) improved, while for α-helical proteins (mostly local M-M contacts) it decreased further (Supporting Information Fig. S1).

We also looked at the correlation between CAD-score and GDT-HA,[12] a more stringent variant of GDT-TS. GDT-HA is similarly derived from four independent superpositions, but their threshold distances (0.5, 1, 2, and 4 Å) are half the size of those used for standard GDT-TS. Therefore, GDT-HA can provide a better resolution for models of higher accuracy. The best correlating CAD-score variants are the same (A-S, S-S and A-A) and their correlation values remain very similar. Namely, the ranges for Pearson's and Spearman's correlation coefficients are (0.91–0.95) and (0.92–0.95), respectively (Supporting Information Fig. S2).

The only adjustable parameter used in CAD-score is the values of van der Waals (VDW) radii of protein atoms. Since different VDW radii sets have been reported in the literature we asked whether the results are sensitive to the choice of a particular set. To this end, in addition to the assessment of CASP9 models using standard VDW radii reported by Li and Nussinov,[20] we repeated the analysis using the set of minimal VDW radii derived by the same authors.[20] Although differences between the two VDW sets are variable and some are fairly significant (up to 0.45 Å), we observed only negligible differences in CAD-score values and their correlation with either GDT-TS or GDT-HA (Supporting Information Table S1). This finding should not be too surprising after all, since

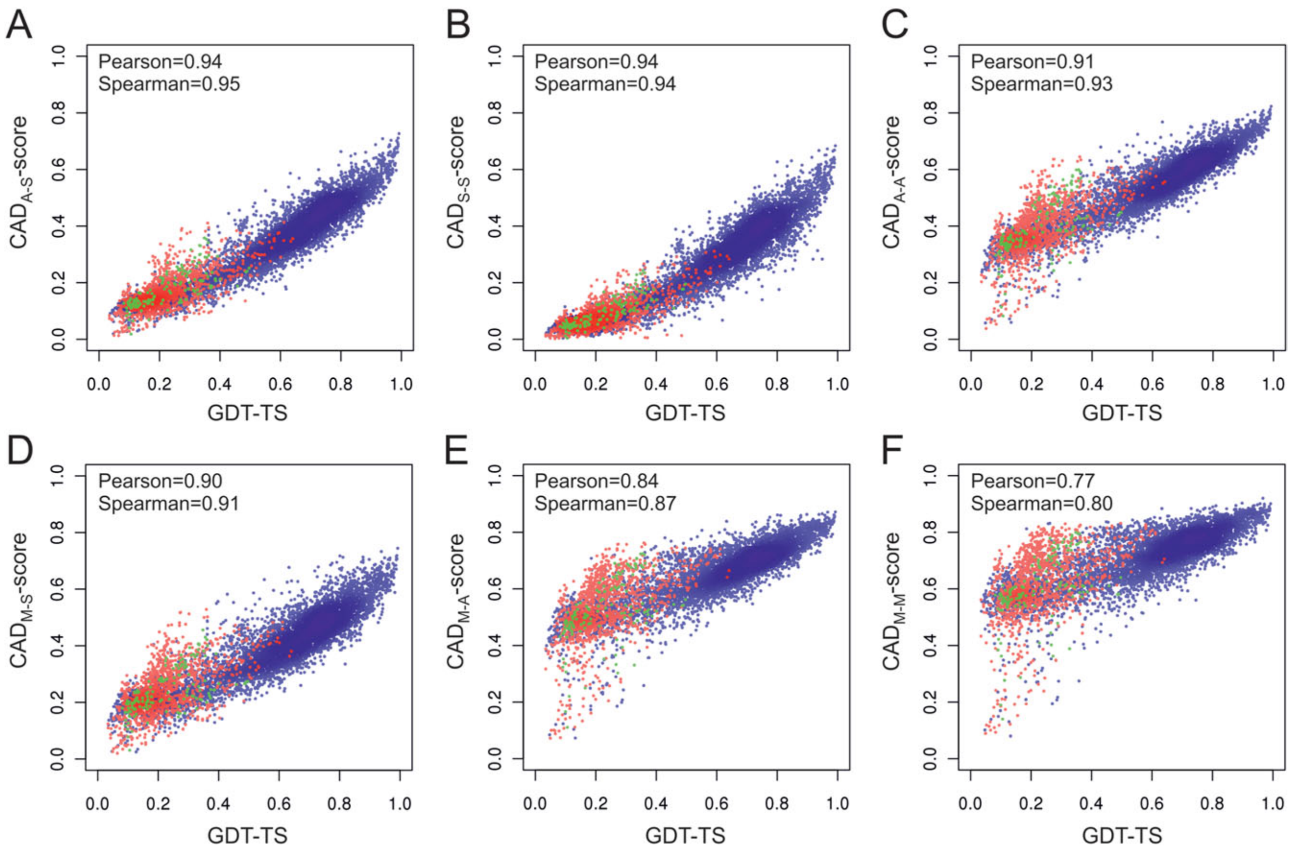

Figure 2

Relationship between GDT-TS (horizontal axis) and different variants of CAD-score (vertical axis) for CASP9 models. CAD-score variants (**A**–**F**) are arranged in the order of their decreasing correlation with GDT-TS. Blue, red, and green colors represent models assessed in template-based (TBM), free modeling (FM) and unresolved (TBM/FM) categories respectively. Higher color intensity reflects higher density of models. Pearson's correlation coefficients and Spearman's rank correlation coefficients are indicated for each plot.

CAD-score is based on contact area differences rather than the absolute contact area sizes.

Taken together, these analyses revealed a robust performance of CAD-score on single-domain proteins. In particular, the three CAD-score types (A-S, S-S and A-A) stand out. They provide some of the highest resolution and the best correlation with GDT-TS/GDT-HA. Therefore, we will further focus mostly on the properties of these three CAD-score variants.

CAD-score promotes the physical realism of structural models

It is generally assumed that the better model score indicates a more accurate representation of the reference structure. However, it has been noticed that some model evaluation scores including GDT-TS are fairly insensitive to unrealistic structural features such as steric clashes or deviations in residue geometries.[9] Therefore, an improvement according to a particular score may come at the expense of physical realism of structural models. In other words, some protein structure prediction methods, especially if they are optimized against a particular score, may seemingly "improve" their performance according to that score without real improvement in model accuracy.

What about CAD-score? How the improvement of models according to CAD-score relates to their physical realism? Since CAD-score is highly correlated with GDT-TS (Fig. 2), how does it fare in comparison to GDT-TS in this regard? To answer these questions, we analyzed pairs of models for which CAD-score and GDT-TS rankings were in conflict, namely, CAD-score and GDT-TS assigned better values to different models within the considered pair. We asked which score in those cases is more consistent with the physical realism of models. We chose the MolProbity score[27] as a measure of physical realism. MolProbity is one of the widely used structure quality evaluation suites. The MolProbity score is a single number that represents the central MolProbity protein statistics collected from a large number of high quality protein crystal structures. The score takes into account clashes between non-bonded atoms, backbone Ramachandran

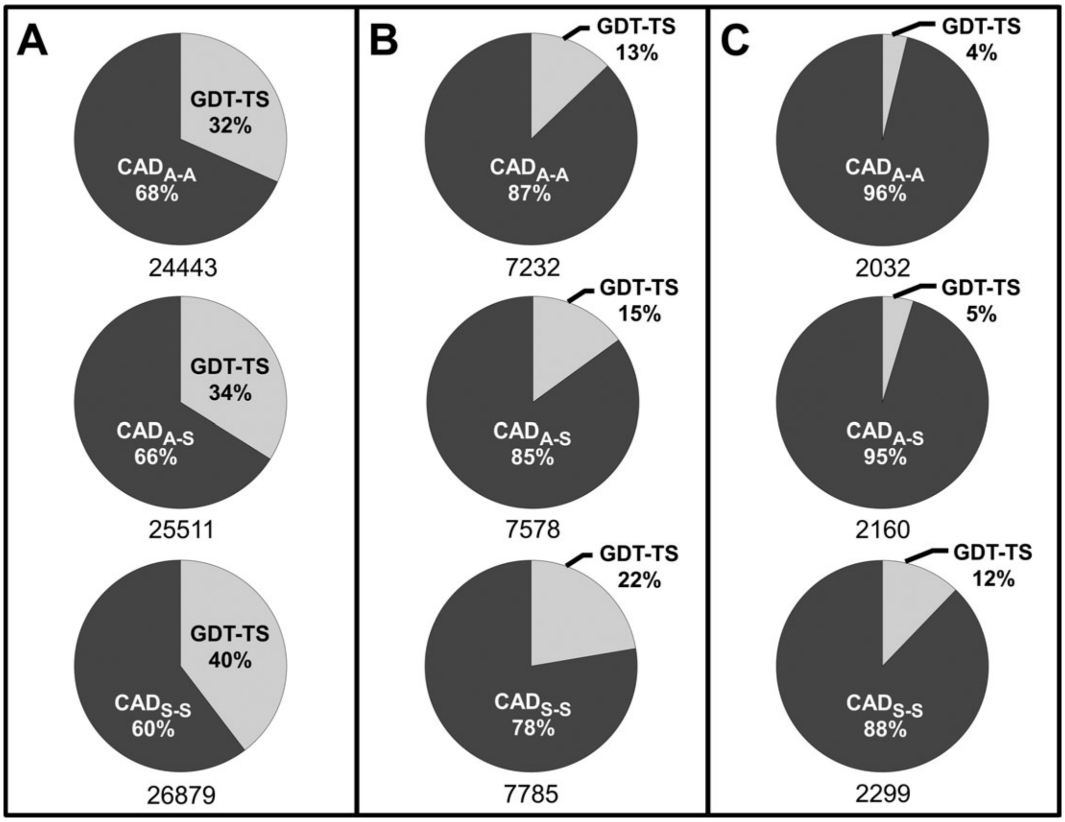

Figure 3

Pairs of CASP9 models with conflicting ranking by GDT-TS and CAD-score. Only models with GDT-TS over 0.6 (60%) were considered. Pie charts represent the MolProbity score agreement with rankings by GDT-TS and each of the three variants of CAD-score. Numbers of analyzed model pairs are indicated below each chart. (**A**) Complete MolProbity score data. (**B**) Data for model pairs with the absolute MolProbity score difference greater than the standard deviation (0.9). (**C**) Data for model pairs derived as in (B) with the additional requirement that the absolute difference of either GDT-TS or CAD-score would be greater than the corresponding standard deviation [0.06 (6%) for GDT-TS, 0.05 for $\mathrm{CAD}_{\mathrm{A\text{-}A}}$, 0.06 for $\mathrm{CAD}_{\mathrm{A\text{-}S}}$, and 0.07 for $\mathrm{CAD}_{\mathrm{S\text{-}S}}$].

conformations outside the favored regions, and side chain rotamer outliers.[27] Unlike GDT-TS and CAD-score, the MolProbity score is not a reference-centric measure. It does not tell how close the model is to the native structure. Instead, it reports how "protein-like" the model is. Therefore, MolProbity may be considered as an independent "judge" for resolving ranking conflicts between the two reference-based scores.

We limited our analysis to reasonably accurate models of single domains (assessment units) as it would be meaningless to consider the physical realism of grossly incorrect models. Thus, we selected models above the GDT-TS threshold of 0.6 (60%) and compiled pairs of models with the conflicting rankings between GDT-TS and each of the three CAD-score variants. We then looked at how the MolProbity score would rank models within the same pairs. The results of this analysis show that in conflicting rankings, CAD-score is supported by the MolProbity score much stronger than GDT-TS [Fig. 3(A)]. Among the three CAD-score variants, $\mathrm{CAD}_{\mathrm{A\text{-}A}}$ received the greatest MolProbity support, followed by $\mathrm{CAD}_{\mathrm{A\text{-}S}}$ and then by the most stringent variant, $\mathrm{CAD}_{\mathrm{S\text{-}S}}$. However, model pairs with small differences of MolProbity, GDT-TS, or CAD-score values might be expected to contribute a certain level of noise to the results. Therefore, we performed two additional tests aimed at the progressive elimination of the impact of noise. First, we looked only at those conflicting rankings, for which the absolute MolProbity score difference is greater than the standard deviation of the MolProbity score distribution on all considered models [Supporting Information Fig. S3(A)]. As a result, the CAD-score agreement with MolProbity increased dramatically [Fig. 3(B)]. For the second test, in addition to the constraint on the MolProbity score difference, we asked that either GDT-TS or CAD-score values would also differ more than the corresponding standard deviation [Supporting Information Fig. S3(B–E)]. The second test has further emphasized the overwhelming MolProbity support for CAD-score [Fig. 3(C)]. For example, the ranking by $\mathrm{CAD}_{\mathrm{A\text{-}A}}$ agreed with the MolProbity score in 24 out of 25 cases, and only in one case this was true for GDT-TS. Collectively, these

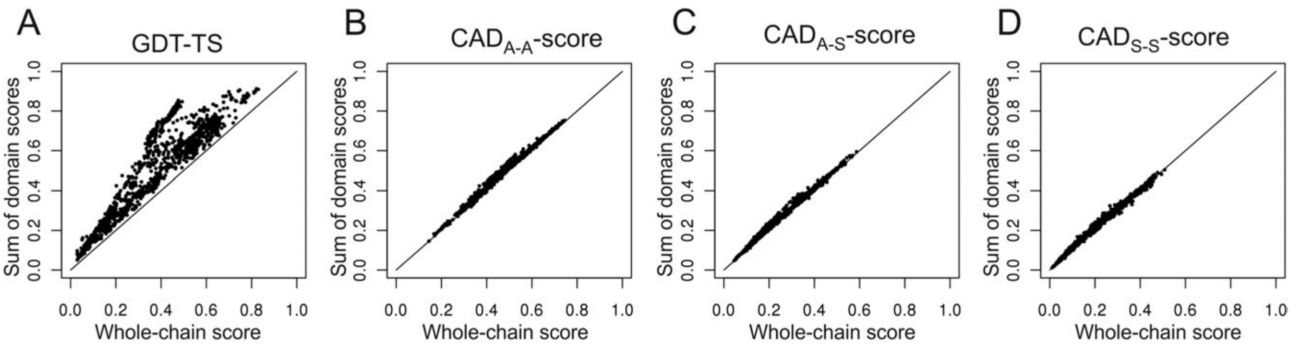

Figure 4

Correlation between the model scores for the whole-chain (horizontal axis) and the weighted sum of domain scores (vertical axis) for CASP9 multi-domain targets. Different plots represent the analysis of the same models using different scores: (**A**) GDT-TS, (**B**) CAD$_{A-A}$, (**C**) CAD$_{A-S}$, and (**D**) CAD$_{S-S}$.

analyses indicate that if there is a disagreement about the relative ranking of models, CAD-score assigns a better score to the physically more realistic model much more often than does GDT-TS. This CAD-score property might be especially relevant for tasks such as ranking models of higher accuracy and assessing model refinement, because the better performance according to CAD-score would strongly imply the improvement in physical realism as well.

CAD-score removes the necessity to split multi-domain proteins into domains for model evaluation purposes

Many proteins are composed of multiple structural domains. However, GDT-TS and other scores based on the rigid-body superposition (e.g. TM-score, RMSD) are sensitive to even small differences in domain orientation. As a result, the score of the model for the entire structure may often be disconnected from the scores of the models for individual domains. This problem can be alleviated by splitting the target structure into domains and performing domain-based evaluation. However, as there are no universal criteria for domain definition, it is often impossible to unequivocally define both the number of domains and their exact boundaries. Moreover, it is not always clear whether it is necessary to split the multi-domain target structure into domains for evaluation purposes. A simple method for helping to decide whether or not the splitting into domains is required was recently introduced by Grishin and colleagues.[17] Their method, used in the "official" CASP9 evaluation,[26] is based on the analysis of correlation between GDT-TS scores of the whole-chain models and the weighted sum of GDT-TS for individual domains. The weighted sum is defined as follows: GDT-TS scores for each individual domain, multiplied by its length, are summed up and divided by the sum of the domain lengths.[17] The main idea is that if the scores for the whole-chain models are systematically lower (or higher) than the weighted sum of domain scores, then the splitting into domains should be considered. Since this idea is quite general, we decided to perform a similar analysis based on CAD-score and to compare the results with those obtained for GDT-TS. However, some CASP9 whole-chain target structures have additional residues compared to the sum of individual domains. To make the analysis entirely objective, we removed these additional residues from multi-domain whole-chain target structures, so that the whole-chain structure and the sum of domains would have exactly the same residues. We then assessed models against these whole-chain targets by both CAD-score and GDT-TS. The latter data was recalculated using LGA.[28] The resulting analysis of 1287 models for 24 multi-domain targets is presented in Figure 4. There is a stark difference between the GDT-TS plot [Fig. 4(A)] and those based on CAD-score [Fig. 4(B–D)]. In the case of GDT-TS, essentially for all the models the weighted sum of domain scores is higher than the score for the entire structure. This reaffirms the choice made by the assessors to parse these CASP9 targets into domains (assessment units) for performing robust model evaluation using GDT-TS. In contrast to GDT-TS, all three CAD-score variants [Fig. 4(B–D)] show at most only small differences between scores of the whole-chain structure and the combined scores of domains. In other words, the evaluation based on CAD-score allows the objective comparison of models for multi-domain proteins even without parsing the structures into domains.

CAD-score provides a balanced assessment of the inter-domain arrangement accuracy in models for multi-domain proteins

Although CAD-score shows little or no difference between domain-based and whole-chain evaluation (Fig. 4), the important question is whether or not this reflects an adequate scoring of domain rearrangement. In our view, the accuracy of predicting mutual domain arrangement should not be judged by simple error in the directional orientation between the domains. If domains are kept together only by a connecting linker, any fixed mutual orientation might be structurally and/or biologically irrelevant (especially if the linker is flexible). In such case, the penalty for not predicting a particular orientation observed in the crystal structure would be unfair. In contrast, if domains share extensive interface, their specific arrangement suggests structural and/or biological importance and therefore should contribute to the evaluation score more significantly. In other words, the larger is the fraction of protein surface area buried at the domain interface, the larger potential impact (positive or negative) it should be able to exert on the total score of the model. Following this logic, we analyzed the expected and the observed contributions of the domain arrangement to the total model score. We defined the expected contribution as the fraction of solvent accessible surface (SAS) buried at the domain–domain interface(s) of a target corrected for the accuracy of a given whole-chain model. The correction was performed by simply multiplying the SAS fraction buried at the interface by the whole-chain score. Buried SAS was determined by subtracting SAS of the whole-chain structure from the sum of SAS for individual domains and dividing by two. Of course, the definition of expected contribution of the domain arrangement is simplistic, as we consider the accuracy of the interface prediction to be the same as the average accuracy of all domains. Nevertheless, this concept is useful for exploring the relationship between the expected and the observed contributions. The observed contribution was defined as the difference between the whole-chain scores and the weighted sum of domain scores (as shown in Fig. 4).

We analyzed the relationship between the expected and the observed contributions of the inter-domain interface prediction component to the total score of the model for both CAD-score and GDT-TS. The results are presented in Figure 5. We only included data for those multi-domain protein models, for which all individual domains had GDT-TS values over 0.4 (40%) and therefore were expected to represent at least a correct structural fold. Despite some data noisiness, the figure reveals a strikingly different behavior of GDT-TS and CAD-score.

Based on the data for GDT-TS [Fig. 5(A)], two important observations can be made. Firstly, the largest observed contributions to the total score are several times that of the largest expected contributions. This is the result of the GDT-TS property to strongly exaggerate the domain rearrangement making the domain-based evaluation a necessity. Secondly, this exaggeration is most strongly pronounced for models with some of the smallest expected values. In other words, given similar average quality of individual domains, models for targets having the smallest inter-domain interface are more likely to produce poor scores for the whole-chain structure.

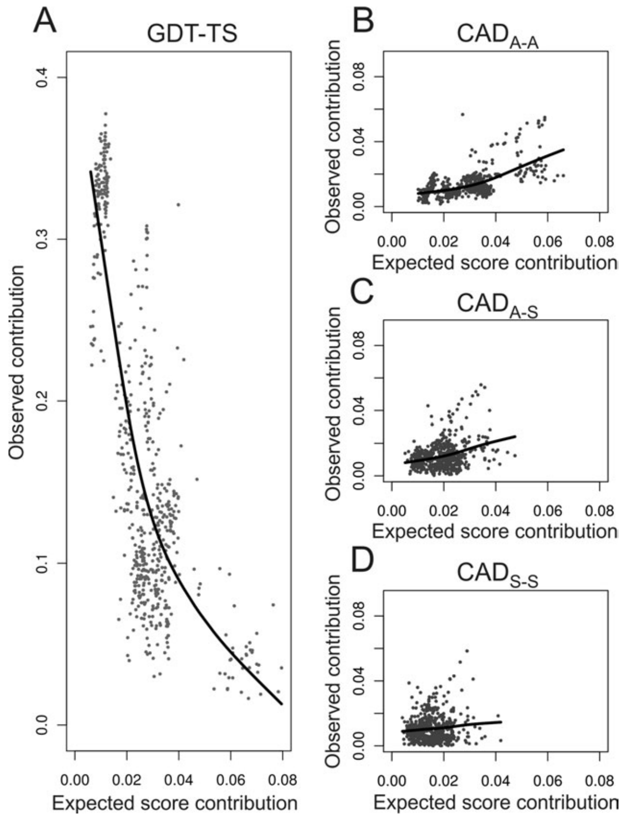

Figure 5

Relationship between the absolute values of expected (horizontal axis) and observed (vertical axis) contributions of the domain rearrangement to the total model score. For definitions of expected and observed contributions see the main text. Only data for models with GDT-TS > 0.4 (40%) for any individual domain are included. General trends for each plot are indicated by a cubic spline applied to the data (solid line). (**A**) GDT-TS, (**B**) CAD_{A-A}, (**C**) CAD_{A-S}, and (**D**) CAD_{S-S} data.

In contrast, for CAD-score [Fig. 5(B–D)] the observed contribution of the domain arrangement score to the total score tends to increase as the expected contribution increases. The best agreement is displayed by CAD_{A-A}-score followed by CAD_{A-S} and CAD_{S-S} scores. Although the relationship is somewhat noisy, the observed contributions almost never exceed the expected ones, indicating the balanced impact of domain arrangement errors to the total score.

An illustrative example of GDT-TS problems upon evaluation of models for multi-domain targets that disappear with the application of CAD-score is provided in Figure 6. GDT-TS scores for both domains of CASP9 model TS453 [Fig. 6(B)] are better than those for TS245 [Fig. 6(C)]. However, despite the visually very similar mutual domain arrangement in both models [Fig. 6(A)], TS453 is assigned a worse full-chain GDT-TS value. Obviously, this cannot be considered a fair assessment. In contrast, CAD-score

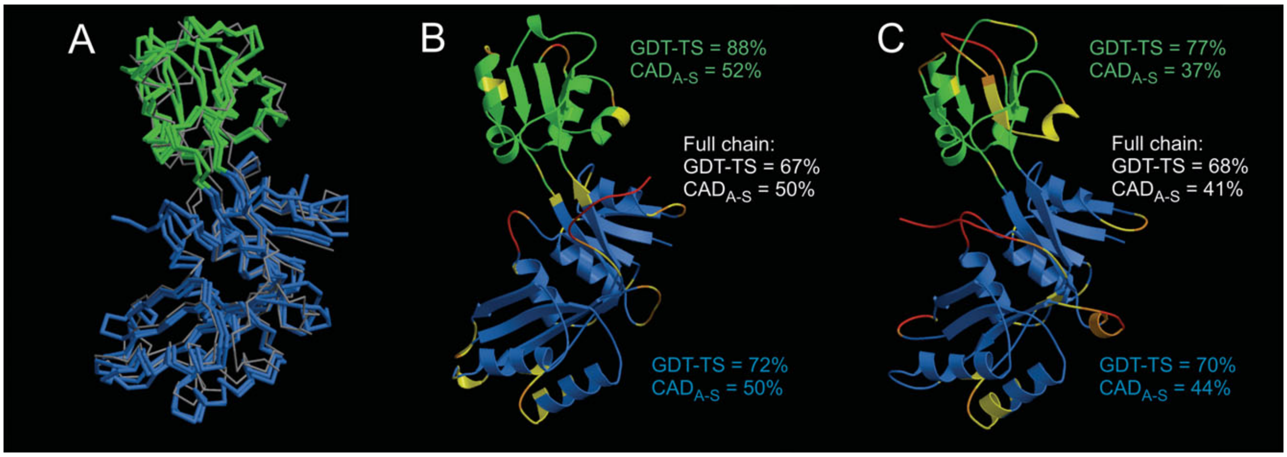

Figure 6

An example of multi-domain structure evaluation by GDT-TS and CAD-score. (**A**) Two models, TS453 and TS245, colored by domains (blue and green) are superimposed with the target T0533 structure (grey). Cartoon representations show models TS453 (**B**) and TS245 (**C**). Increasingly larger deviations of Cα-atoms are indicated by yellow, orange, and red colors, respectively. GDT-TS and CAD_{A-S}-score values in blue and green are for the corresponding domains, white, for the entire model.

assigns better scores not only for individual domains of TS453, but, as might be expected, also for the full-chain model. The tendency of GDT-TS to overestimate tiny differences in mutual domain arrangement is apparent even within the same model. It would be reasonable to expect the accuracy for a full-chain model to be in between the worst-scoring and the best-scoring domains. However, according to GDT-TS, both full-chain models in Figure 6 are worse than their least accurate domain. Again, this problem is non-existent for CAD-score.

Since the mutual arrangement of domains is not conceptually different from the arrangement of protein chains, CAD-score can also be used to evaluate the accuracy of models for protein complexes. The larger is the inter-subunit interface, the bigger impact of its prediction accuracy on the total CAD-score of the protein complex may be expected.

CAD-score can directly evaluate the accuracy of inter-domain or inter-subunit interfaces

In addition to scoring models for entire multi-domain or multi-subunit structures, CAD-score provides a direct way for assessing the accuracy of the interface prediction. The only difference is the reference against which the model is evaluated. In this case the reference would be defined as contact areas between residues originating from either different protein domains (inter-domain interface) or different protein subunits (inter-subunit interface). Figure 7 provides specific examples of inter-domain and inter-subunit interfaces of different accuracy. The first example [Fig. 7(A)] illustrates the accuracy of the inter-domain interface for two models of target T0533 that have been analyzed in detail above. It confirms once again that model TS453 has a more accurate inter-domain interface than TS245. Another example [Fig. 7(B)] features inter-subunit interfaces of different accuracy within two oligomeric predictions for target T0576. One of the two models, TS458, was identified by CAD-score as having the most accurate interface for this target. This CAD-score assignment completely agrees with the CASP9 assessment of oligomeric predictions.[14]

CAD-score web server and standalone software

To make the method for calculating CAD-scores easily accessible, we implemented it as a web server available at http://www.ibt.lt/bioinformatics/cad-score/. The server features a simple and intuitive interface. There are several main functionalities. The CAD-score server can evaluate the accuracy of single-chain protein models as well as models of protein complexes (multi-chain structures) against the reference structure. In addition, the server may be used to specifically evaluate the accuracy of interface prediction. The interface can be defined in a very flexible manner. It can be defined either between different protein chains or between any user-defined ranges of residues (as in the interface between protein domains).

The input to the server is either a single model or multiple models and the reference structure to be evaluated against. The CAD-score server calculates all CAD-score variants, but by default reports only the three that correlate best with GDT-TS, namely CAD_{A-A}-score, CAD_{A-S}-score, and CAD_{S-S}-score. As a summary of model evaluation results, the server provides not only CAD-score variants, but also TM-score, GDT-TS, and

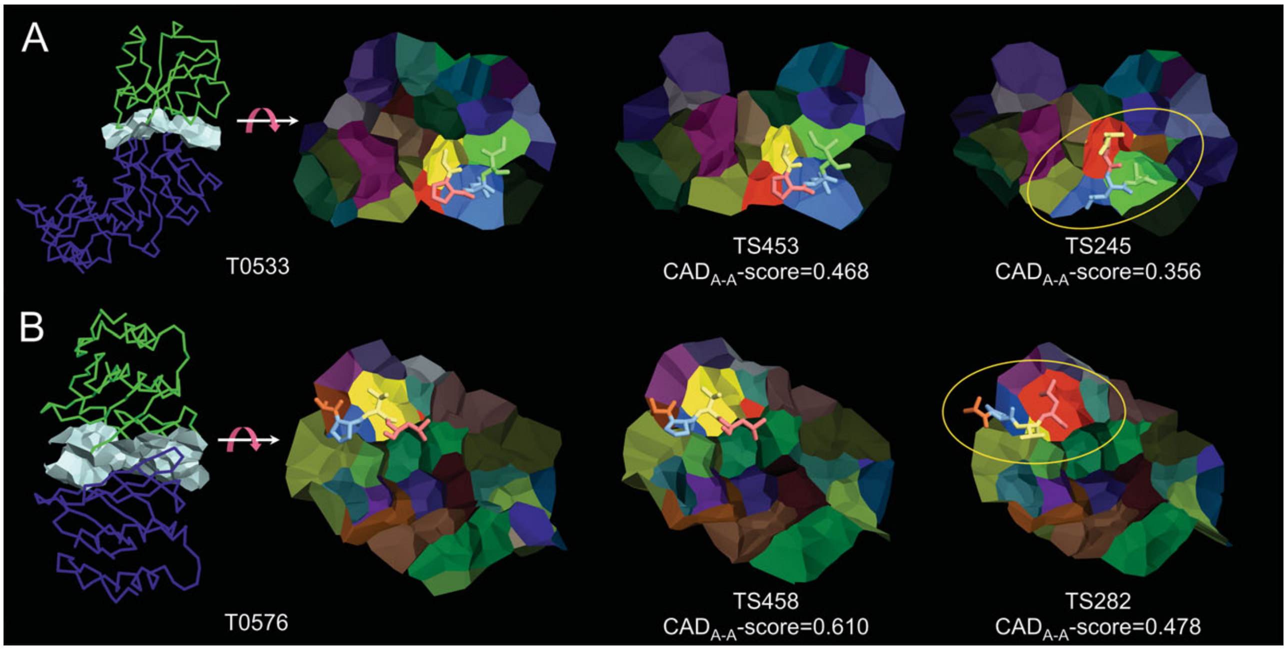

Figure 7

Examples of direct evaluation of the interface between domains (A) and subunits (B). (**A**) The inter-domain interface within the two-domain target T0533 (left) is compared with interfaces in two models, TS453 and TS245. For ease of comparison, interfaces are represented as sets of colored faces of Voronoi cells in the same orientation. Different colors correspond to different residues at the interface. Interface CAD$_{A-A}$-score values are indicated for each model. Major errors within the less accurate interface are indicated with ellipse. The corresponding protein chain fragment is shown as sticks in the target and both models. (**B**) The inter-subunit interface within the dimeric structure of target T0576 (left) is compared with interfaces in two multi-chain models, TS458 and TS282. Notations are the same as in (A).

GDT-HA values produced by the TM-score software.[7] Models can be sorted by any of these scores. In addition to the summary, the server generates position dependent color-coded CAD error profiles for protein models. These profiles are particularly useful for visual identification of either unique or common patterns in a particular region of a protein chain in a group of models. The user can also explore individual models in detail. The detailed analysis includes superimposed contact maps for the model and the reference, the error profile with different smoothing windows, and the Jmol visualization of model and reference structures color-coded according to the local CAD error.

Along with the CAD-score web server, the corresponding standalone software package is available for local use. The standalone CAD-score software computes the same data as the server but provides additional flexibility, in particular, if large-scale evaluation or clustering of models is needed. The software can be downloaded from the CAD-score web server address.

DISCUSSION

The development of protein structure prediction methods and scores used for their benchmarking are interdependent. Robust and effective scores promote improvements in protein structure prediction methods. On the other hand, the overall improvement in model accuracy necessitates a more sensitive and more comprehensive evaluation. At present, due to both the improvement of structure prediction methods and the dominance of template-based models, the focus is shifting toward the accuracy of structural features beyond the backbone. More emphasis is put on the physical plausibility of computational models. The ability to evaluate the accuracy of mutual domain arrangement in models for multi-domain proteins and the arrangement of subunits within protein complexes is also becoming increasingly important.

In this study we present CAD-score, a new model scoring function for comprehensive evaluation of structural models. CAD-score builds upon the concept of contact area difference (CAD) originally introduced by Abagyan and Totrov.[18] However, the new score differs significantly in its design and algorithmic implementation.

One of the key differences is the treatment of missing residues in the model. The original CAD only takes into account the subset of residues that are common for both target and model. In this regard, it is reminiscent of RMSD, which can be calculated only on a common set of residues. In the newly defined CAD-score both the failure to include the residue into the model and the failure to predict all of its contacts are treated identically. To put it differently, CAD-score encourages the construction of the complete model. Incorrectly modeled regions can make at most only negligible improvements to the score; however, even grossly incorrect regions of the model cannot make the score worse compared to the situation when they are not modeled at all. In this respect, the design of CAD-score is similar to that of GDT-TS, which does not reward, but at the same time does not penalize grossly inaccurate regions. We believe that this is a very positive feature of a reference-based model evaluation score, as it allows testing of new bold ideas in protein structure prediction without being penalized for large local errors.

The second difference is the normalization procedure. The normalizing factor in the CAD number as proposed by Abagyan and Totrov is different for different models of the same reference structure (target). This makes the ranking of models for a given target problematic. In our case, the normalizing term is constant for a given target, no matter how unusual or how different evaluated models are.

Yet another difference is the range of values. The originally proposed CAD number is not always guaranteed to fall within the range from 0 to 1 (0–100%). In contrast, the newly defined CAD-score can never be outside of the [0,1] range. This is assured by "symmetric" boundaries of a maximal contact area difference for a given residue pair. We treat the failure to predict an existing contact in the same way as its "strong" over-prediction. The "strong" over-prediction is defined as the case when the absolute contact area difference is larger than the reference contact area itself. In both extremes we consider the prediction to be equally wrong, and therefore the contact area difference is bounded by the reference contact area. As a result, the sum of bounded contact area differences for the model can never exceed the sum of contact areas of the target.

Algorithms for deriving contact areas in our case and the original CAD study are also substantially different. We derive contact areas using a protein structure tessellation approach. It allows us to take into account the influence of other residues surrounding the considered residue pair. In the original CAD study, the contact area for a pair of residues is calculated in isolation, thereby tending to overestimate the size of contact area. In addition, the resolution of contact areas is different in the two methods. In contrast to Abagyan and Totrov, we calculate contact areas at the level of heavy atoms, and that allows us to derive contact areas not only for entire residues, but also for subsets of residue atoms such as main chain and side chain. In turn, this allows us to define a number of CAD-score variants, addressing different aspects of model accuracy and providing different degrees of sensitivity.

In this study, we explored properties of the newly introduced CAD-score and compared it primarily with GDT-TS, a widely accepted score for reference-based model evaluation. We found that for single structural domains CAD-score shows a strong correlation with GDT-TS (Fig. 2) and GDT-HA (Supporting Information Fig. S2). In both cases the strongest correlation is obtained for those CAD-score variants that include either all residue atoms or side chain in any combination. It may seem somewhat surprising that contacts between all atoms and side chains (A-S) and even those between side chains (S-S) correlate with GDT-TS better than all atom to all atom (A-A) contacts. However, side chains make up about two thirds of the protein structure and apparently their packing is what gives rise to a specific folding pattern. CAD-score variants that include only main chain atoms on at least one side of the contact show somewhat weaker correlation, with the main chain to main chain (M-M) variant occupying the lower end. The character of main chain to main chain contacts differs significantly depending on the secondary structure type. While in β-sheets these contacts are defined by the global topology, for α-helices they are local and are mostly defined by the accuracy of secondary structure assignment. Apparently, the lack of non-local contacts within α-helical structures is a major factor in making the M-M variant least correlated with GDT-TS (Supporting Information Fig. S1).

One of the important advantages of CAD-score compared to GDT-TS and other structure superposition-based methods is the robust evaluation of models for multi-domain proteins and protein complexes. Our analysis showed that in contrast to GDT-TS, CAD-scores of individual domains and the whole-chain structure are tightly connected (Fig. 4). Moreover, the accuracy of the inter-domain or inter-subunit interface is an integral part of the total score. The more extensive is the interface, the more potential improvement or deterioration to the total score it may contribute (Fig. 5). Although the domain-based model evaluation is perfectly possible, CAD-score removes the necessity to chop the structure into domains to get meaningful results. Moreover, even if the structure is split into domains, the performance of CAD-score cannot be strongly affected by imprecise or even outright wrong domain boundary definition, which would have a large impact in the GDT-TS-based evaluation.

According to CAD-score, the accuracy of the model depends only on how closely the contact areas between residues (or subsets of residue atoms) correspond to those in the reference structure. However, what may seem a simplistic definition of model accuracy in fact incorporates many structural features such as interatomic distances, dihedral angles, hydrogen bonds, and bond lengths. Protein structure prediction methods trained using a particular model evaluation score, in some cases may "improve" their performance by optimizing some of the model structural parameters at the expense of others.

Here, we showed that CAD-score is associated with physical realism of models much stronger than GDT-TS (Fig. 3). In particular, this property of CAD-score may be relevant for assessing model refinement, which turns out to be a surprisingly hard problem.[29]

Although we developed the new CAD-score with the reference-based model evaluation in mind, the approach may be a valuable tool for other tasks such as clustering of structural models. Model clustering is one of the steps employed by many current protein structure prediction approaches, especially if there are no suitable structural templates. The clustering step is used for the identification of near-native structures from a large set of candidate structures (decoys). Since contact areas between residues directly reflect the strength of physical interactions, CAD-score values may be more suitable for grouping models with similar energies compared with Cartesian distance-based approaches such as RMSD or GDT. As clustering typically involves large numbers of models, the clustering method needs to be fast. In CAD-based clustering, the slowest step is the computation of contact areas between residues in individual models. However, once it is done, subsequent calculation of pairwise CAD-scores is very fast. An example of model clustering results using CAD-score is presented in Supporting Information Figure S4.

CAD-score is based on interatomic contacts and as such it is not exclusively restricted to protein structures. A similar approach could be applied for evaluation of models of other biomolecules forming complex 3D structures such as RNA. Similarly, evaluation of the protein–protein interface (inter-domain or inter-subunit) accuracy could be easily extended to the more general case of protein–ligand interfaces. Obviously, the CAD-score based evaluation would be most appropriate for large interfaces such as those in protein-nucleic acids complexes, but perhaps it may be sufficiently informative even for interfaces between proteins and small molecules.

CONCLUSION

The newly introduced CAD-score has a number of attractive properties. It is based on physical contacts between residues, thereby directly reflecting interactions within the protein structure. It is a continuous, threshold-free function that returns quantitative accuracy scores within the strictly defined boundaries. The definition of CAD-score does not contain any arbitrary parameters. CAD-score provides a single uniform framework for assessing single-domain, multi-domain, and even multi-subunit protein structural models of varying degree of accuracy and completeness. While being highly correlated with GDT-TS on single-domain structures, CAD-score displays a stronger emphasis on the physical realism of models. We believe that all these attractive properties make CAD-score a valuable tool for the development and assessment of protein structure prediction and refinement methods as well as for clustering models based on their mutual similarity.

ACKNOWLEDGMENTS

The authors thank Andriy Kryshtafovych for providing CASP9 GDT-HA data, Adam Zemla for making the standalone version of the LGA program available and Mindaugas Margelevičius for critically reading the manuscript. They also acknowledge Yang Zhang and Jeff Skolnick for the TM-score software.

REFERENCES

1. Kabsch W. A solution for the best rotation to relate two sets of vectors. Acta Crystallograp Sec A 1976;32:922–923.
2. Moult J, Pedersen JT, Judson R, Fidelis K. A large-scale experiment to assess protein structure prediction methods. Proteins 1995;23:ii–v.
3. Moult J, Hubbard T, Bryant SH, Fidelis K, Pedersen JT. Critical assessment of methods of protein structure prediction (CASP): round II. Proteins 1997;Suppl 1:2–6.
4. Zemla A, Venclovas Č, Moult J, Fidelis K. Processing and analysis of CASP3 protein structure predictions. Proteins 1999;Suppl 3:22–29.
5. Zemla A, Venclovas Č, Moult J, Fidelis K. Processing and evaluation of predictions in CASP4. Proteins 2001;Suppl 5:13–21.
6. Siew N, Elofsson A, Rychlewski L, Fischer D. MaxSub: an automated measure for the assessment of protein structure prediction quality. Bioinformatics 2000;16:776–785.
7. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins 2004;57:702–710.
8. Cozzetto D, Kryshtafovych A, Fidelis K, Moult J, Rost B, Tramontano A. Evaluation of template-based models in CASP8 with standard measures. Proteins 2009;77:18–28.
9. Sadreyev RI, Shi S, Baker D, Grishin NV. Structure similarity measure with penalty for close non-equivalent residues. Bioinformatics 2009;25:1259–1263.
10. Aloy P, Stark A, Hadley C, Russell RB. Predictions without templates: new folds, secondary structure, and contacts in CASP5. Proteins 2003;53:436–456.
11. 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;53:395–409.
12. Kopp J, Bordoli L, Battey JN, Kiefer F, Schwede T. Assessment of CASP7 predictions for template-based modeling targets. Proteins 2007;69:38–56.
13. Keedy DA, Williams CJ, Headd JJ, Arendall WB, III, Chen VB, Kapral GJ, Gillespie RA, Block JN, Zemla A, Richardson DC, Richardson JS. The other 90% of the protein: assessment beyond the Calphas for CASP8 template-based and high-accuracy models. Proteins 2009;77:29–49.
14. Mariani V, Kiefer F, Schmidt T, Haas J, Schwede T. Assessment of template based protein structure predictions in CASP9. Proteins 2011;79:37–58.
15. Jauch R, Yeo HC, Kolatkar PR, Clarke ND. Assessment of CASP7 structure predictions for template free targets. Proteins 2007;69: 57–67.
16. Ben-David M, Noivirt-Brik O, Paz A, Prilusky J, Sussman JL, Levy Y. Assessment of CASP8 structure predictions for template free targets. Proteins 2009;77:50–65.
17. 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 (Oxford) 2009; bap003.

18. Abagyan RA, Totrov MM. Contact area difference (CAD): a robust measure to evaluate accuracy of protein models. J Mol Biol 1997;268:678–685.
19. Kim DS, Cho Y, Kim D. Euclidean Voronoi diagram of 3D balls and its computation via tracing edges. Comput Aided Design 2005;37:1412–1424.
20. Li AJ, Nussinov R. A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. Proteins 1998;32:111–127.
21. McConkey BJ, Sobolev V, Edelman M. Quantification of protein surfaces, volumes and atom-atom contacts using a constrained Voronoi procedure. Bioinformatics 2002;18:1365–1373.
22. Olechnovič K, Margelevičius M, Venclovas Č. Voroprot: an interactive tool for the analysis and visualization of complex geometric features of protein structure. Bioinformatics 2011;27:723–724.
23. Moult J, Fidelis K, Kryshtafovych A, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)—round IX. Proteins 2011;79:1–5.
24. Tramontano A, Cozzetto D, Giorgetti A, Raimondo D. The assessment of methods for protein structure prediction. Methods Mol Biol 2008;413:43–57.
25. Kryshtafovych A, Moult J, Bartual SG, Bazan JF, Berman H, Casteel DE, Christodoulou E, Everett JK, Hausmann J, Heidebrecht T, Hills T, Hui R, Hunt JF, Seetharaman J, Joachimiak A, Kennedy MA, Kim C, Lingel A, Michalska K, Montelione GT, Otero JM, Perrakis A, Pizarro JC, van Raaij MJ, Ramelot TA, Rousseau F, Tong L, Wernimont AK, Young J, Schwede T. Target highlights in CASP9: Experimental target structures for the critical assessment of techniques for protein structure prediction. Proteins 2011;79:6–20.
26. Kinch LN, Shi S, Cheng H, Cong Q, Pei J, Mariani V, Schwede T, Grishin NV. CASP9 target classification. Proteins 2011;79:21–36.
27. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 2010;66: 12–21.
28. Zemla A. LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res 2003;31:3370–3374.
29. MacCallum JL, Perez A, Schnieders MJ, Hua L, Jacobson MP, Dill KA. Assessment of protein structure refinement in CASP9. Proteins 2011;79:74–90.