PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

DOMAIN AND CONTACTS: ASSESSMENT

Assessment of domain boundary predictions and the prediction of intramolecular contacts in CASP8

Iakes Ezkurdia, Osvaldo Graña, José M. G. Izarzugaza, and Michael L. Tress*

Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

ABSTRACT

This article details the assessment process and evaluation results for two categories in the 8th Critical Assessment of Protein Structure Prediction experiment (CASP8). The domain prediction category was evaluated with a range of scores including the Normalized Domain Overlap score and a domain boundary distance measure. Residue-residue contact predictions were evaluated with standard CASP measures, prediction accuracy, and *Xd*. In the domain boundary prediction category, prediction methods still make reliable predictions for targets that have structural templates, but continue to struggle to make good predictions for the few *ab initio* targets in CASP. There was little indication of improvement in the domain prediction category. The contact prediction category demonstrated that there was renewed interest among predictors and despite the small sample size the results suggested that there had been an increase in prediction accuracy. In contrast to CASP7 contact specialists predicted contacts more accurately than the majority of tertiary structure predictors. Despite this small success, the lack of free modeling targets makes it unlikely that either category will be included in their present form in CASP9.

Proteins 2009; 77(Suppl 9):196–209.

Key words: protein structure; structural domains; residue-residue contacts.

INTRODUCTION

Many protein structures can be sub-divided into semi-autonomous, compact folding units with separate hydrophobic cores.[1] The identification of these structural domains is a crucial first step in many processes such as the prediction of protein structure,[2,3] experimental and theoretical studies on the function of individual proteins,[4,5] and target selection for structural genomics.[6]

The definition of a protein domain boundary will depend to a large extent on the purpose for which the boundaries are being defined. For example, structural domains do not always coincide with functional domains; structural domains are usually defined by their hydrophobic cores and relative independence, while functional domains are more influenced by evolutionary relationships. Domain definitions that are useful for structure prediction are not always the same as those that would be useful for crystallizing a protein. For many proteins, these different definitions of domain boundaries might more or less intersect, but for many others, the task of defining domain boundaries is a complex one. In this experiment, we lean towards a definition of structural domains that are useful for structure prediction. Predicting domains as a means of target selection for structural genomics is not feasible unless predictors are presented with the sequence of the whole protein in CASP the sequence that is provided for prediction is not always the whole protein and often already includes the expression tags used to clone the protein for crystallization.

Domain boundary prediction has been part of CASP since CASP6.[7] There was a clear general improvement in prediction accuracy in CASP7,[8] in particular for those targets where it was possible to model

Additional Supporting Information may be found in the online version of this article.
The authors state no conflict of interest.
*Correspondence to: Michael Tress, Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), c./Melchor Fernandez Almagro, 28029 Madrid, Spain.
E-mail: mtress@cnio.es
Received 1 May 2009; Revised 10 June 2009; Accepted 24 June 2009
Published online 24 July 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22554

a structure based on a template in advance of the domain definition. CASP7 showed that several groups were able to make reliable, good quality template-based domain predictions. However, where domain boundaries fell in regions that had to be modeled *ab initio*, the prediction of domain boundaries was less good.

Here we report on the assessment of the CASP8 domain prediction category. We have assessed predictions using a range of scoring schemes. Once again many predicting groups were able to make good template-based predictions of domain boundaries and the differences between the better groups and the rest of the predictors were statistically significant. However, there was no evidence to suggest a general improvement in performance over CASP7.

We have also assessed the prediction of residue-residue contacts in this article. Contact predictions are usually most valuable for targets in the free modeling regime and although previous CASP experiments[9,10] have suggested that residue–residue contact predictions are not yet accurate enough to be used for *de novo* prediction, reliably predicted contacts do have the potential to be a valuable aid in protein structure prediction. Indeed, it has been suggested that predicting just a few important residue–residue contacts should be enough to allow the construction of approximate 3D model structures for many small proteins.[11,12] The actual number of reliably predicted contacts necessary to fold a protein *de novo* depends on the size of the protein, the accuracy of the predictions and the importance of the predicted pair of interactions to the overall fold.

Although it may not yet be possible to use predicted contacts to generate 3D models directly, contact predictions may still be useful as a means of guiding a 3D structure prediction protocol or in directly selecting from a range of alternative structural models. It may also be possible to use less reliable predictions to help validate *de novo* modeled loop regions. There has been renewed interest in contact prediction since CASP7, as can be seen from the number of groups that have published new methods since the experiment.[13–17]

For the contact prediction evaluation section, we concentrated on assessment units that were defined as free modeling (FM) or template-based modeling/free modeling overlap (TBM/FM), because contacts from template-based modeling target domains are trivial to predict from the templates themselves. The numerical criteria used in the assessment are essentially the same as they were in CASP6 and CASP7. Although the small sample size meant that we were not able to draw any firm conclusions, the assessment did suggest that there had been a general improvement in prediction accuracy.

DOMAINS AND DOMAIN BOUNDARY PREDICTION

Domain assignment

The official assessment units defined by the assessors for the evaluation of the structure prediction experiments formed the basis of the domain prediction experiment. That definition process is described in detail in the domain definition and categorization paper in this issue.[18] The majority of the official domain definitions were retained in the domain prediction experiment, but four targets that were treated as single target domains in the structure prediction experiment were defined as two-domain proteins for the domain boundary prediction experiment. These included two anomalies from the domain definition process (T0483 and T0494) that were not split into domains even though two other domains with the same fold were split into domains (T0430 and T0456).

In addition to these changes, residues that were excluded from the structure prediction assessment but that clearly interacted with the surface of a domain were defined as being part of that domain. Target T0362 illustrates this change. The C-terminal residues interact with the surface of domain 2, but were left out of the structure prediction assessment. Because these residues clearly interact with domain 2, predictors should be penalized for predicting these C-terminal residues as part of domain 1. Therefore, the C-terminus was redefined as part of domain 2 for the domain prediction experiment.

Domain definition is not straightforward. Given a set of target proteins with known structures automatic domain definition algorithms and human experts may converge on a single domain definition in a number of cases, but for many structures there is likely to be disagreement.[19] For several targets, we felt that there were two or more valid domain definitions for the domain prediction assessment. For example, T0424 might be split into three or four domains depending on whether the domain is defined from a structural or evolutionary point of view (Fig. 1).

Twelve targets were allowed two equally valid alternative definitions and three targets (T0391, T0450, and T0443) had three equally valid alternative domain definitions. For targets with multiple alternative definitions, scores were calculated for each of the alternative definitions and the best score was kept for the evaluation. This applied to all scoring schemes.

In CASP7, there were no multidomain free modeling targets. In principle, the CASP8 targets were a little more difficult because three of them could not easily be split into domains using templates. T0405 is a two domain free modeling target, T0443 a multiple domain target with templates that proved very difficult to find, and T0496 a two-domain target that had one free modeling domain and a two helical domain extension. All domain definitions and alternative definitions are in Supporting Information Table I.

Predictors

A total of 18 predicting groups made predictions for the domain prediction category. This was eight fewer than in CASP7. Eleven of the groups were server predic-

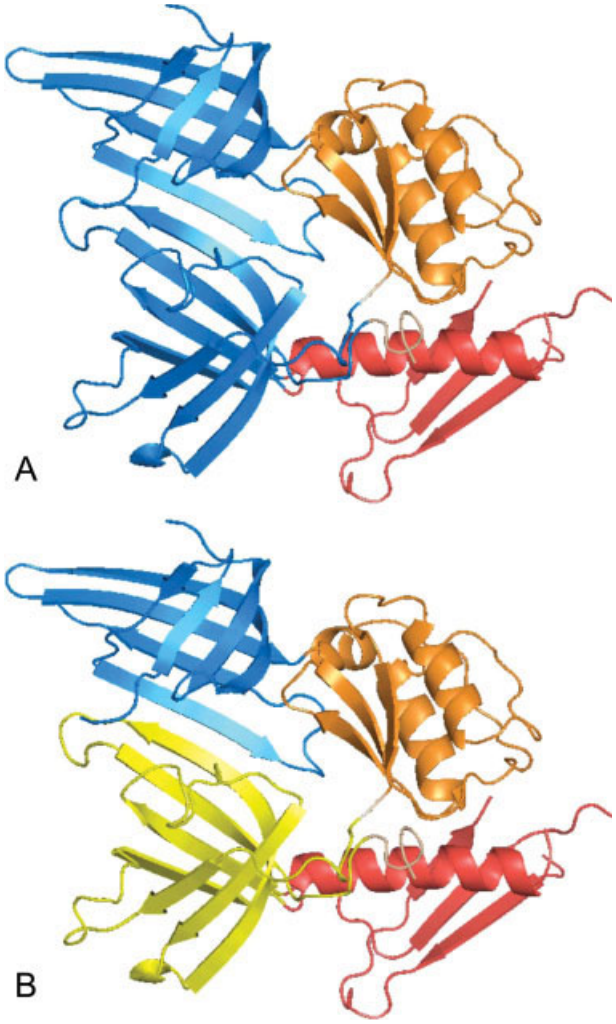

Figure 1

Target T0424 had two possible domain divisions. For this target, we allowed predictions of three domains or four domains.

tors. The 18 predictors came from just 12 distinct research groups. Although not all the predictors were the same, 17 of the predictors were from groups or participants that took part in the domain prediction experiment in CASP7. We have included one extra predictor (*DOMSERV_H&E*, DP136) whose results were not assessed at the Cagliari meeting due to a miscommunication. In the article, predictors are generally referred to by group number. Group names and numbers for both categories are listed in Table I.

Most groups made predictions for the vast majority of the targets and eight groups predicted all 122 targets (Fig. 2). The breakdown of the predictions into single and multiple domain predictions shows that predictors have different prediction strategies. Four groups (DP105, DP059, DP317, and DP355) are markedly more conservative in their predictions and predict that over 80% of the target proteins form single domains. At the other end of the scale are the three predictors from the Baker group (DP051, DP333, and DP350) that actually predict more multiple domains than single domains (Fig. 2).

As a first approximation, we counted the number of times that the predictors agreed with the number of domains assigned to each of the target structures by the assessors. For those targets where there was more than one possible answer (in other words where the assessors felt that one or two, or two or three domains were equally possible), we allowed predictors to choose either definition. The results can be seen in Figure 3. The accuracy of the prediction of domain number ranges from 89.3% (DP136) to 66%.

Scoring

The domain prediction category was assessed in a similar way to the CASP7 assessment. This time predictions were analyzed using three separate scoring schemes, one that measures the sensitivity and specificity of domain break predictions, one that converts the precision of

Table I

Groups and Codes for the Domain and Contact Prediction Experiments

Predictor	Type	Domains	Contacts
SMEG-CCP	Human	—	RR014
AK_RF_2	Human	—	RR032
BAKER-ROSETTADOM	**Server**	**DP051**	—
BHSAI	Human	—	RR059
MULTICOM-CMFR	**Server**	**DP069**	**RR069**
Infobiotics	Human	—	RR072
Distill	**Server**	**DP073**	**RR073**
Pairings	**Server**	—	**RR077**
ProtAnG_s	**Server**	**DP105**	—
Oka	Human	DP118	—
MULTICOM-RANK	**Server**	—	**RR131**
DOMSERVE_H&E	**Server**	**DP136**	—
Distill_domains	Human	DP144	—
3Dpro	**Server**	**DP157**	**RR157**
LCBContacts	Human	—	RR158
SAM-T08-2stage	**Server**	—	**RR197**
CBRC-DP_DR	Human	DP229	—
FLOUDAS	Human	—	RR236
RR_Fang_1	**Server**	—	**RR249**
SAM-T08-server	**Server**	—	**RR256**
LEE-SERVER	**Server**	**DP293**	**RR293**
DomFOLD	**Server**	**DP317**	—
RR_Fang_2	Human	—	RR327
BAKER-DP_HYBRID	**Server**	**DP333**	—
BAKER-GINZU	**Server**	**DP350**	—
DomPred	**Server**	**DP355**	—
LEE	Human	DP407	RR407
SVMSEQ	**Server**	—	**RR413**
Hamilton-Torda-Huber	**Server**	—	**RR424**
MUProt	**Server**	**DP443**	**RR443**
MULTICOM	Human	DP453	RR453
SAM-T06-server	**Server**	—	**RR477**
SPINE-2DA-Zhou	**Server**	—	**RR487**

Server groups in bold. Those codes with a grey background did not have a participating group in CASP7.

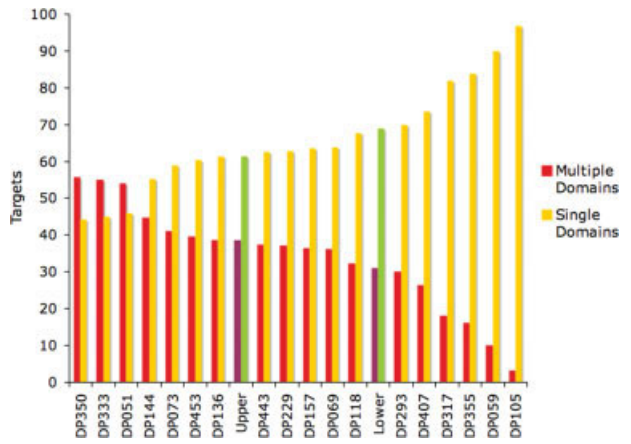

Figure 2

The percentage of targets predicted as single or multiple domains. The percentage of targets predicted as single or multiple domains by each group, ordered by increasing proportion of single domains. The upper and lower limits of the assessor-defined domains (several targets were allowed to be single or double in this assessment) are also shown in green (single) and magenta (double) bars. Four groups predict a very high proportion of single domains. Three predictors predict more multiple domains than single domains. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

domain sub-division into a single normalized score (Normalized Domain Overlap), and one that assesses the structural integrity of the predictions.

The Normalized Domain Overlap (NDO) scoring system was introduced in the CASP6 domain assessment.[7] The advantage of this scoring scheme is that it reduces the scoring of the prediction to a single normalized score and it penalizes both under prediction and over prediction of domains. However, it does not take into account the prediction of interdomain linkers and does not explicitly penalize predictors for predicting domain breaks that would probably destabilize the fold of the real target structure.

The scoring scheme is explained in more detail in the CASP6 and CASP7 assessment papers.[7,8] The NDO score is calculated by counting the number of residues that overlap between predicted domains and correct domains. A total overlap score is calculated from the matrix of the numbers generated from the overlap, and the overlap score is normalized by the number of residues in the target. For those targets that had more than one possible domain division, NDO scores were calculated for each domain definition and the best score was taken as predictor score.

We calculated NDO scores for three subsets of targets. The "all" subset comprised those targets that were defined as either multiple domain targets or single domain targets. This subset excluded the 10 targets where we felt that it was equally possible to predict multiple domains or single domains. For these targets, it is much easier to score full marks by predicting a single domain target and this gives a slight advantage to conservative predictors. The "multiple domain" subset included all those targets that were defined exclusively as multiple domain targets. There were 38 targets in this subset. "Hard" targets were those multiple domain targets where at least one of the domains was categorized as "free modeling" in the structure assessment or where none of the domains were classified in the "high accuracy" subset. There were 23 targets in this subset.

The results can be seen in Figure 4. The results of the NDO comparison resemble those of Figure 3—the better groups at predicting the correct number of domains are also those that have the higher NDO scores. The group with the highest NDO scores in all three subsets is

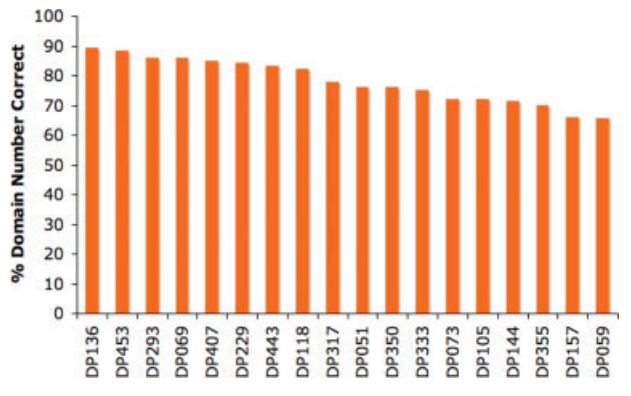

Figure 3

Predicting the correct number of domains. Predicting the correct number of domains is part of predicting the correct domain boundaries. Here, we show the percentage of cases in which predicting groups predicted the same number of domains as the assessors. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

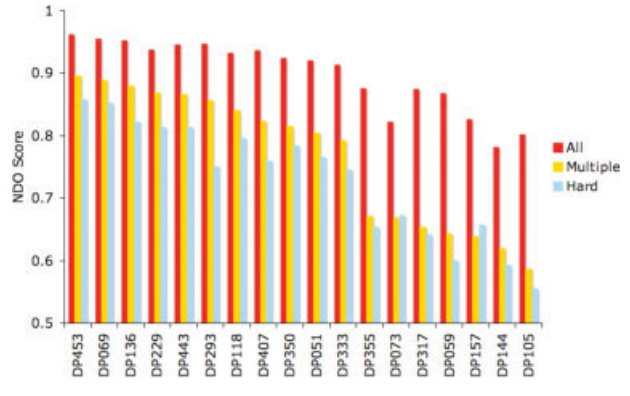

Figure 4

Mean NDO scores for a range of target subsets. The mean NDO scores for all groups for all the targets, for the subset of multiple domain targets and the subset of hard multiple domain targets. Bars ordered by the NDO scores for the multiple domains targets. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

	453	136	69	229	443	293	407	118	350	51	333	73	355	317	59	157	144	105
453	NA	0.84	0.46	0.57	0.19	0.08	0.07	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
136	0.84	NA	0.74	0.55	0.62	0.21	0.03	0.03	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
69	0.46	0.74	NA	0.51	0.18	0.11	0.06	0.07	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
229	0.57	0.55	0.51	NA	0.94	0.40	0.14	0.11	0.11	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
443	0.19	0.62	0.18	0.94	NA	0.16	0.13	0.09	0.16	0.09	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
293	0.08	0.21	0.11	0.40	0.16	NA	0.32	0.74	0.46	0.37	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00
407	0.07	0.03	0.06	0.14	0.13	0.32	NA	0.81	0.97	0.74	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
118	0.03	0.03	0.07	0.11	0.09	0.74	0.81	NA	0.51	0.35	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00
350	0.01	0.05	0.02	0.11	0.16	0.46	0.97	0.51	NA	0.17	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
51	0.01	0.02	0.01	0.06	0.09	0.37	0.74	0.35	0.17	NA	0.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00
333	0.00	0.01	0.00	0.01	0.04	0.26	0.50	0.26	0.12	0.45	NA	0.00	0.00	0.00	0.00	0.00	0.00	0.00
73	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	0.79	0.61	0.76	0.23	0.24	0.01
355	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.79	NA	0.83	0.53	0.36	0.16	0.00
317	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.61	0.83	NA	0.57	0.60	0.54	0.03
59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.76	0.53	0.57	NA	0.56	0.53	0.01
157	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.36	0.60	0.56	NA	0.68	0.07
144	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.16	0.54	0.53	0.68	NA	0.94
105	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.03	0.01	0.07	0.94	NA

Figure 5

Paired *t*-test evaluation of NDO scores. Predictors were compared head-to-head over multiple domain targets predicted by both. Significant differences between the predictions for each group are indicated by those squares with a white background. The darkest squares indicate where two groups should not be considered statistically different. The results clearly show that there two separate groups of predictors with significant differences between the two groups.

DP453, while server DP069 from the same group is a close second. The two subsets of multiple domain targets strongly suggest that the predictors can be divided into two groups. This is backed up by head-to-head comparisons of NDO scores over common subsets of multiple domain targets (Fig. 5). The *P*-values from these comparisons shows that the 11 predictors with higher NDO scores have significantly higher NDO scores than the other seven predictors in this experiment.

In addition to calculating NDO, we also calculated scores for domain boundary prediction. This score is the distance of the predicted domain boundary from the assessor-defined domain boundary. The domain boundary distance score measures how close predictors get to the "correct" domain boundary (or boundaries). The score is simple, but for some proteins whether the predicted domain boundary is, say, four or six residues from the true domain boundary is essentially meaningless because both would be equally incorrect.

Domain boundary distance scores are calculated in a similar way to the GDT-TS scoring system. Predictions are given one point for agreeing with the assessor defined boundary, another point if they are within one residue, a further point if they are within two, and so on up to seven residues. A prediction three residues away from the correct boundary would, therefore, have five points. If the official domain boundary has a linker, the whole linker is regarded as the domain boundary. Scores are calculated based on all distances between the predicted and correct domain boundaries for a target and the total of all predicted boundary scores is normalized by dividing by eight and the total number of domain boundaries. The number of domain boundaries comes from either the target or the prediction, whichever is higher. This penalizes over-prediction of domain boundaries.

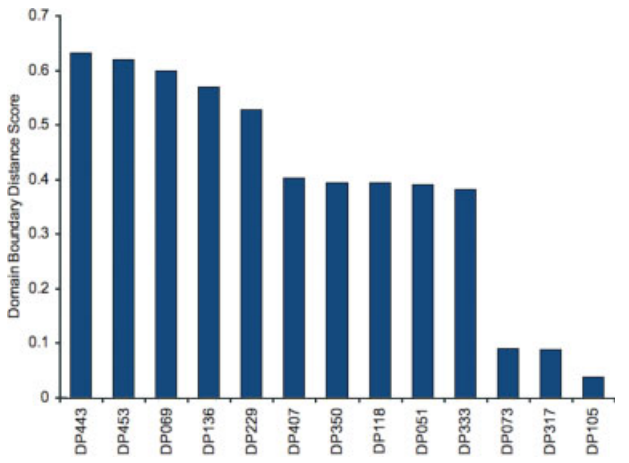

Figure 6

Mean domain boundary distance scores for the targets with multiple domains. Mean per target domain boundary distance scores for each predictor (over a common subset of 35 multiple domain targets). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

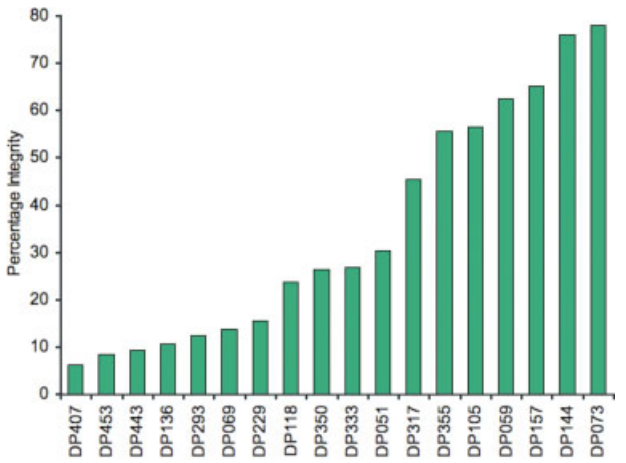

Figure 7

Percentage integrity scores. The proportion of predicted multiple domains where the proposed domain boundaries were likely to generate structures that would be missing essential elements of the core of the protein and would be unlikely to fold correctly. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

The scores for those groups that predicted a common subset of 35 multiple domain targets are shown in Figure 6. The groups that have high NDO scores also score well with the domain boundary distance score. The domain boundary distance score can also be dissected and viewed from the distinct perspectives of accuracy and sensitivity in the Supporting Information.

One further measure was used to differentiate the predicting groups. The measure evaluates the structural integrity of the predicted domains. Here we looked at all predictions by eye to determine whether the predicted domain split was likely to leave a domain with a disrupted hydrophobic core. Here, it was not important whether or not the prediction was correct, but whether the domain split would remove important structural residues. A prediction that missed the correct domain boundary but did not cut crucial secondary structure or core elements was not considered as disruptive. However, we did consider domains that were likely to be too short to fold as disruptive.

The total number of cases where the predictions would have disrupted the structural integrity were calculated for each group and this score divided by the number of predicted multiple domains (single domain predictions cannot disrupt the structure). The results are shown in Figure 7 and show that the groups that predict multiple domains with the lowest proportion of disrupted structures were the same groups that had the highest NDO scores and the highest domain boundary prediction scores.

Statistical comparisons

It is clear from all the measures that eleven groups perform somewhat better than the other seven over all target subsets and all measures. We carried out statistical comparisons to test whether the differences between all

	453	443	69	136	229	293	350	407	51	118	333	317	73	144	59	157	355	105
453	-	0.91	0.73	0.20	0.09	0.03	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
443	0.91	-	0.84	0.04	0.09	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
69	0.73	0.84	-	0.08	0.09	0.07	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
136	0.20	0.04	0.08	-	0.34	0.15	0.09	0.09	0.08	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
229	0.09	0.09	0.09	0.34	-	0.97	0.41	0.42	0.39	0.34	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00
293	0.03	0.01	0.07	0.15	0.97	-	0.51	0.28	0.48	0.26	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
350	0.00	0.01	0.00	0.09	0.41	0.51	-	0.87	0.32	0.93	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00
407	0.02	0.01	0.02	0.09	0.42	0.28	0.87	-	0.90	0.90	0.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00
51	0.00	0.00	0.00	0.08	0.39	0.48	0.32	0.90	-	0.89	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00
118	0.00	0.00	0.00	0.02	0.34	0.26	0.93	0.90	0.89	-	0.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00
333	0.00	0.00	0.00	0.01	0.14	0.20	0.18	0.71	0.23	0.80	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
317	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-	0.77	0.54	0.30	0.40	0.05	0.02
73	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.77	-	0.33	0.88	0.34	0.28	0.00
144	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.54	0.33	-	0.24	0.79	0.19	0.04
59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.88	0.24	-	0.40	0.88	0.16
157	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.40	0.34	0.79	0.40	-	0.66	0.13
355	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.28	0.19	0.88	0.66	-	0.17
105	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.04	0.16	0.13	0.17	-

Figure 8

Paired *t*-test evaluation of the domain boundary scores. Predictors were compared head-to-head over multiple domain targets predicted by both groups. Significant differences between the predictions for each group are indicated by squares with a white background. The darkest squares indicate where two groups are not considered statistically different. The results suggest that the top scoring set of predictors may in fact be two separate subsets of predictors where the differences between the two groups are significant or close to significant in many cases.

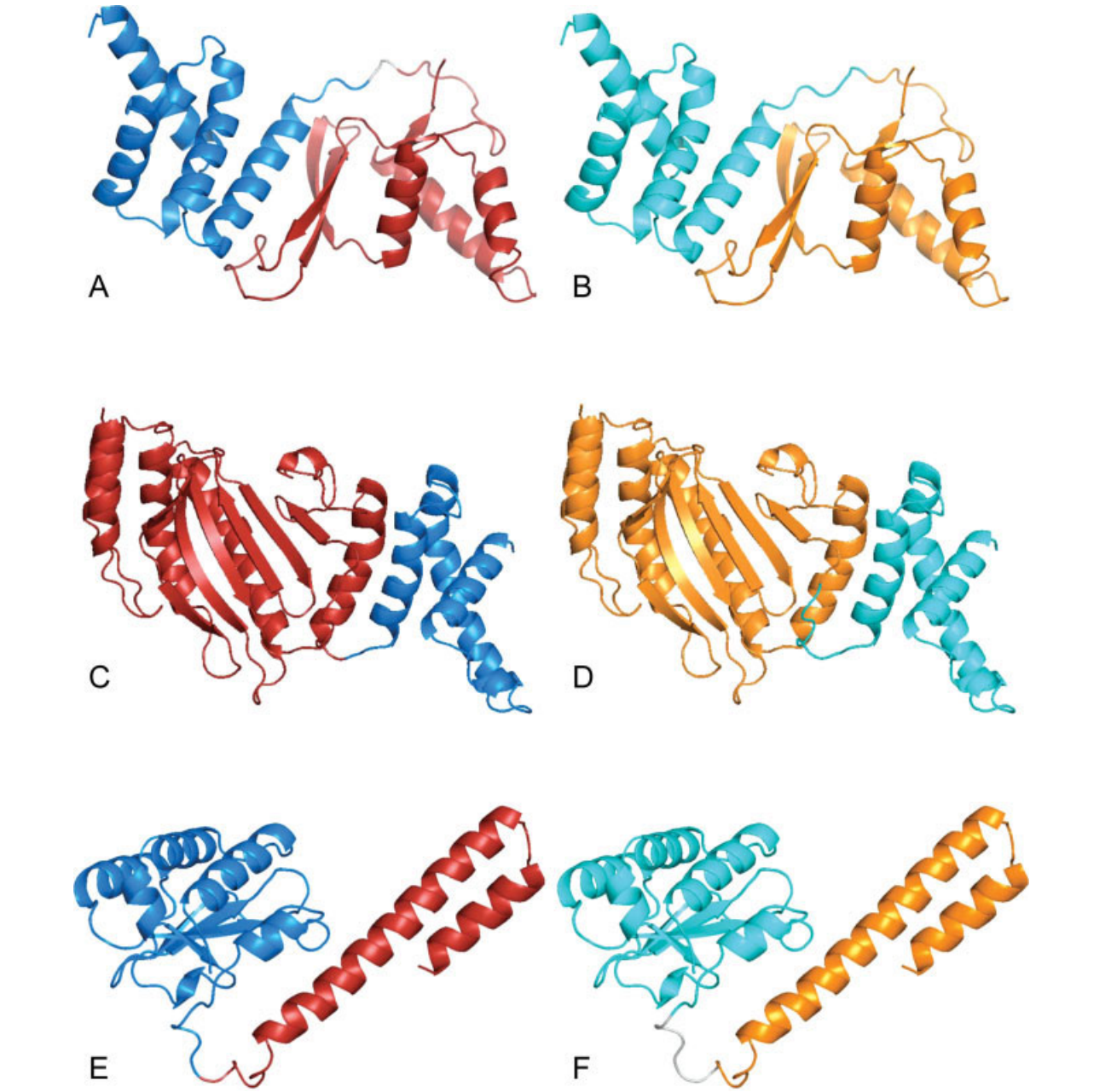

Figure 9

Outstanding predictions. Three outstanding predictions for difficult multiple domain proteins. The official domains are on the left in blue and red, the predictions on the right in orange and cyan. **A** and **B** show target T0443. The assessor-defined domain (A) is one of three possible alternative domains; the prediction is from group DP136 and has an NDO score of 0.98. **C** and **D** show the free modeling target T0405. The assessor-defined domain (C) and the prediction from group DP453 (NDO score of 0.96) differ in only a few residues. **E** and **F** are from the target T0496. The assessor-defined domain (E) and the prediction from group DP229 (NDO score of 0.93) are identical except that DP229 defines a linker, a perfectly plausible domain split in this case.

the groups, and in particular these two subsets of predictors, were significant.

We carried out paired *t*-tests between each pair of groups head-to-head over common subsets of predicted multiple domain targets for both NDO and domain boundary distance scores. The *P*-values from the statistical comparisons performed with the NDO scores are in Figure 5, those from the domain boundary distance scores in Figure 8. The results show that the two groups of predictors are statistically distinguishable from each other and that the predictors inside each group are statistically similar. The results from the domain bound-

ary distance scores suggest that the best 11 predictors may actually form two separate groups with predictors DP0453, DP0443, DP069, DP136, and DP229 in the first group. As in CASP7, the *P*-values from the NDO scores are less decisive.

The standard of template-based prediction was universally high, but there were some outstanding predictions, including several for the three targets that might be considered as *ab initio*. DP136 and DP229 made predictions for T0443 and T0496 that were perfectly possible domain splits (Fig. 9), and the prediction from DP453 for target T0405 was only a few residues away from the assessor-determined domain split.

Comparison to CASP7

As in CASP7, we also used the automated tool PDP[20] to parse the targets into domains. PDP is a well-used and robust method, indeed several of the predicting groups used PDP to parse their models into domains for this experiment. However, it tends to overcut domains[21]; in CASP8, PDP split 47 targets into multiple domains. We calculated NDO scores for each of the predictions against the PDP domain parse, and these scores can be seen in the Supporting Information. As expected, the NDO scores of those predictors that used PDP to split their models did improve slightly, while predictors that predicted more single domains were most penalized.

We also used PDP as a control to give us an idea as to whether there has been progress between CASP7 and CASP8. For the comparison, the PDP parsed domains were considered as just another predictor and the NDO scores for the PDP parsed domains were calculated against the assessor defined domains. For each of the last three CASP experiments, the NDO scores of the predictors and the PDP parsed domains were compared head to head using paired *t*-tests. The *P*-values between the predictors and the PDP "predictors" can be seen in Figure 10(a).

Predictors improved with respect to the PDP "predictor" between CASP6 and CASP7, and in CASP7 six groups (DP722, DP581, DP497, DP136, DP556, and DP105) could not be distinguished statistically from PDP. This trend seems to have been somewhat reversed in CASP8. Comparison between experiments is not simple because the targets are different each year and the relative ease of predicting the domain boundaries for each set of targets needs to be taken into consideration. However, it seems highly unlikely that there have been substantial improvements since CASP7.

One explanation for the reverse in CASP8 can be seen in Figure 10(b). Seven of the top 10 predictors from CASP7 repeated as predictors in CASP8 with identical or equivalent servers and it was possible to plot the evolution of their *P*-values from CASP7 to CASP8. From this figure, it can be seen that one of the top groups is still statistically indistinguishable from PDP (DP069) and two have improved with respect to the PDP "predictor" (DP229 and DP118). However, four other groups are not performing as well as they were in CASP7. The same four groups (DP350, DP051, DP407, and DP333) are also part of the second subset of the top scoring predictors in the domain boundary distance scores (Fig. 9), when they were among the best predictors in CASP7.

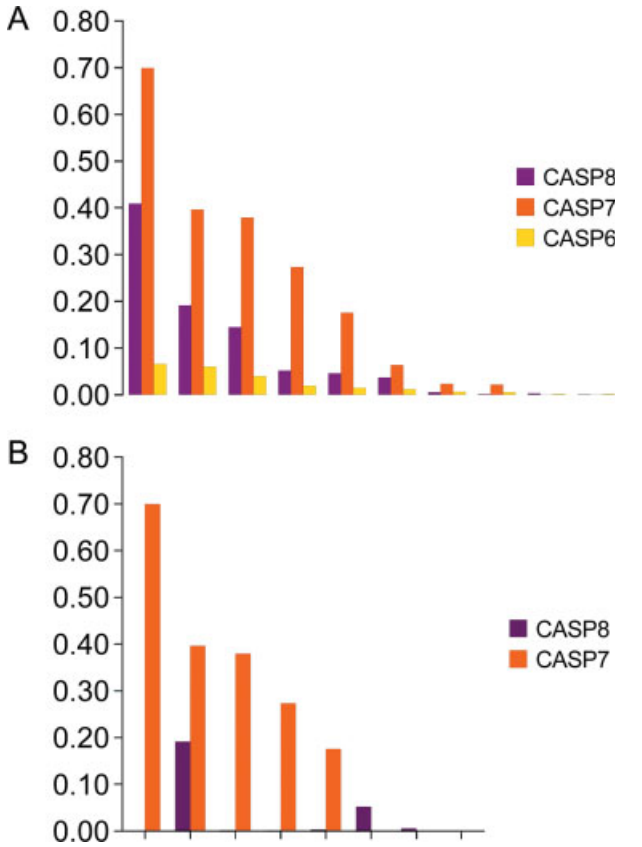

Figure 10

Paired *t*-tests between PDP and predicting groups from CASP6, CASP7, and CASP8. Part **A** shows the *P*-values from the paired *t*-tests for NDO scores between each of the CASP6, CASP7, and CASP8 participating groups and the predicted domains from the structure-based domain predictor, PDP. The *P*-value 0.05 is the cut-off for significance. Groups with *P*-values higher than the cut-off are not significantly different from the PDP predictor. There are fewer groups with *P*-values higher than 0.05 in CASP8. Part **B** shows the same *P*-values, but this time grouped by predictor. Scores are only shown for those predictors that took part in CASP7 and CASP8 with equivalent servers. For example, the second set of bars with *P*-values of 0.4 in CASP7 and 0.2 in CASP8 is for server group DP069, which is a development of the server DomPRO in CASP7. Four of the five groups that could not be distinguished from the PDP predictor in CASP7 are now statistically worse. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

CONTACT PREDICTION

Predictors

There were 22 participating groups in the CASP contact prediction category. This compares to 17 in CASP7, so there was a slight increase in predictors. In addition, eleven of the submitting groups were not involved in the CASP7 contact prediction experiment, so half the predicting groups were new. As in the domain assessment, the predicting groups are referred to in this article by their CASP predictor codes. See Table I for the equivalences between group names and predictor codes.

Contact prediction groups submitted lists of residue pairs that were estimated to be in contact (for the purposes of the experiment within 8Å), as well as a probability estimate that each pair is in contact. Where possible the predictions were sorted by this reliability score before assessment. Group RR158 did not provide a probability estimate and we took their contact list without reordering. Several groups sent in more than one prediction per target, but the assessors based their assessment on just the first set of predicted contacts.

Target selection

Predictors predicted contacts for every target but were evaluated against a set of eight FM and four TBM/FM overlap target domains. There was one fewer FM target domain than in the structure prediction assessment because we considered the two T0405 domains as a single target for this assessment (both domains were categorized as FM). In addition, one extra target has been added to this assessment—target T0460 was assigned to the TBM/FM overlap category by the assessors after the Cagliari meeting. There were seven fewer target domains than in the CASP7 contact prediction assessment.

Assessment

Only residues in the official domain definitions for each target were considered in the analysis. Residues were considered to be in contact within the assessment units if their Cb atoms (Ca for glycines) were within a distance of 8Å. The length of the target domain sequence was used to allow us to compare predicted contacts over a fixed number of residues. In previous CASP experiments, we used the length of the whole target sequence. For target domains of length L, we evaluated the top ranked $L/5$ and $L/10$ predictions according to the predictor probability scores. To be assessed, predictors had to have at least $L/5$ or $L/10$ contacts at each of the different sequence separation limits. Contact predictions were not assessed if they failed to meet the minimum number of predicted contacts for the sequence separation criterion.

These requirements meant that a number of predictions failed to reach sufficient contacts for assessment, particularly for the smaller target domains in multidomain targets. At the $L/5$ cut-off, only 15 groups were evaluated over 50% or more of the targets. To include as many predictions as possible, we also assessed the accuracy of the first five predicted pairs for each target domain.

We concentrated on just two sequence separation ranges, $12 \leq x < 24$ and $x \geq 24$. The majority of the analyses used a sequence separation of 24 or greater. The same threshold was used in all evaluations in the CASP6[9] and CASP7[10] contact prediction assessments. A cut-off of greater than 24 residues distance was used because long-range contacts are much more valuable as structure constraints in 3D structure prediction.[9]

Predictions were evaluated using two different scores, accuracy [true positives/(true positives + false positives)], and *Xd*, which measures how the distribution of distances for predicted contact pairs differs from the distribution of all pairs of residues in the target domain structure. *Xd* was calculated in the same way as in CASP6 and CASP7, and the formula is explained in more detail in those papers.[9,10]

The specialist residue contact predictors were also compared against the contacts inferred from the 3D structural models predicted in the structure prediction experiment. To infer contacts from the 3D structural models, we collected all the distances between Cb-Cb atoms (Ca for glycines) for all the residues in each model structure and ranked the pairs (the inferred contacts) by their Cb-Cb distance. The closest L/x pairs were taken as the "predictions" from the models.

RESULTS

Most of the groups that submitted predictions for the contact prediction category did so for all 12 of the targets, but a third of the predictions failed to meet the $L/5$ threshold at a sequence separation of greater than 24 residues. As a result, it was not always possible to compare predictors over all 12 targets.

It was suggested in CASP7[10] that the length-dependent cut-off was the factor that made the comparison between groups difficult because predictors have no idea how multiple domain targets will be split. In this CASP, we defined the length L as the length of the target domain, not the length of the whole target sequence and as a result the cut-off was reduced. In Figure 11, we show the number of predictors that reach eligibility at three different cut-offs, $L/5$, $L/10$ and just five contacts. As can be seen from the figure, reducing the number of contacts needed for eligibility beyond $L/5$ has little effect on the number of eligible predictions. Groups had noticeably fewer eligible predictions for two targets, T0416-D2 (just eight predictors were eligible at $L/5$) and T0513-D3 (six predictors were eligible at $L/5$). The reason that so few groups reached eligibility for these targets

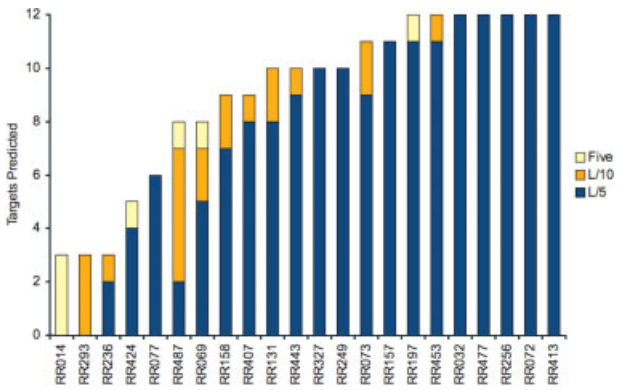

Figure 11

The number of targets predicted at different cut-offs. The figure shows the number of eligible predictions for each group at three different cut-offs. Two are dependent on the length of the target domain (*L*/5 and *L*/10) and one uses just the five highest ranked predicted pairs (Five). Bars are superimposed with the *L*/5 (the most stringent cut-off superimposed on the rest). Predictors reach eligibility for very few extra targets as the cut-off is relaxed (*L*/10, five residues). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

is that both these targets are smaller free modeling target domains that are attached to much larger template-based modeling assessment units. The highest scoring predicted pairs would all be in the template-based domain. So to reach an eligible number of predictions for the FM target domains and to be included in the comparisons, predictors would have to include predictions for as many contacts as possible, even if the reliability scores are very low when compared with those in the template-based modeling region of the target.

Prediction accuracy

The mean prediction accuracy at *L*/5 and 24-residue separation in CASP7 was 13% across all targets. In CASP8, the mean target accuracy was 21.5%. Indeed, two targets (T0397-D1 and T0510-D3) had mean accuracies of over 30%. We were able to show that as in CASP7, the accuracy of the predictions tended to improve as the number of contacts went down (Supporting Information). The increase in accuracy could demonstrate a real improvement in prediction accuracy, but could also be because the targets were easier than in CASP7.

It is possible to compare the accuracy and *Xd* scores across all targets, but the results are not very revealing because there are so few targets and because groups rarely reach eligibility for all targets. There are two ways to deal with the difference in the numbers of targets predicted by each predictor. The first is to look at the accuracy for those groups that predict a subset of targets [Fig. 12(a), 9 targets *L5* and 10 targets *L10*]. The disadvantage of this comparison is that it leaves out several targets and also several groups do not form part of the comparison. So we also calculated the *Z*-scores of the accuracy of the contact predictions for each target at both *L*5 and *L*10. The *Z*-scores go some way towards normalizing for target difficulty and so allow comparisons over all eligible predictions. Individual negative *Z*-scores were converted to zero before calculating the mean in order not to over-weight the less accurate predictions. The mean *Z*-scores are shown in Figure 12(b).

The results from the two comparisons are more or less in agreement. At *L*/5, group RR157 has the highest accuracy with both the common subset and the *Z*-scores. At *L*/10, group RR069, which did not have enough targets to be considered in the common subset, has mean *Z*-scores comparable to RR157, but over just seven targets (RR157 predicted all 12 targets). RR453, RR477, RR197, and RR072 also have high scores in both comparisons, while RR131 and RR073 have higher *Z*-scores, but not enough targets to be eligible for the common subset comparison. RR158 (not shown) also scored well in the

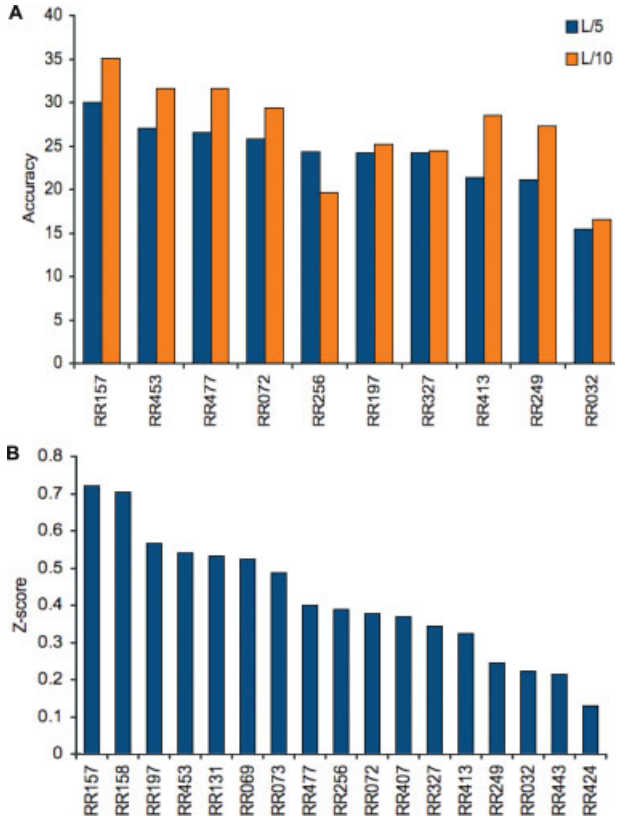

Figure 12

Predictor accuracy. Part **A** shows mean accuracy for those predictors that predicted a common subset of targets. There were nine targets in the common subset at *L*/5, 10 targets in the *L*/10 subset. In **B** the mean *Z*-scores for all the groups irrespective of the number of targets predicted. Only positive *Z*-scores are shown. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

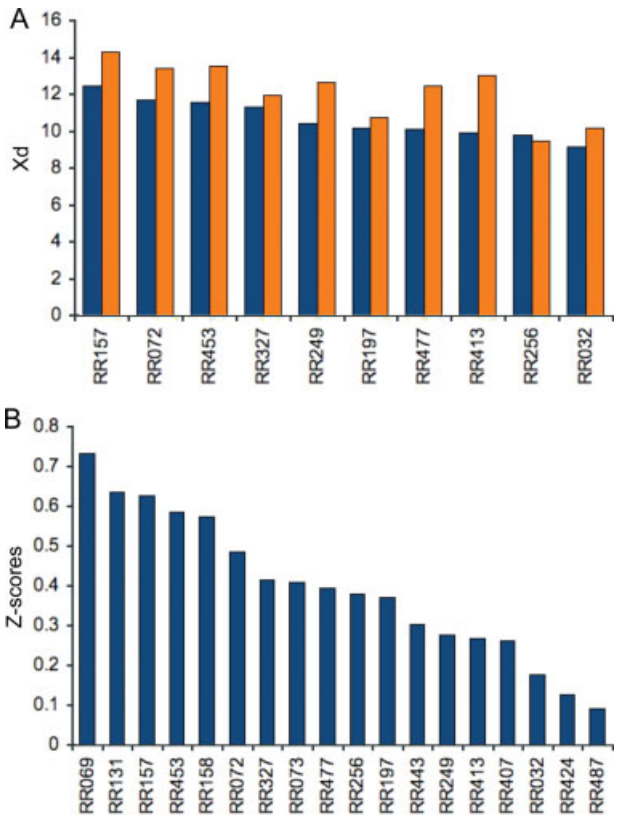

Figure 13

Predictor *Xd*. Part **A** shows mean *Xd* for those predictors that predicted a common subset of targets. There were nine targets in the common subset at *L*/5, 10 targets in the *L*/10 subset. In **B**, the mean *Z*-scores for all the groups irrespective of the number of targets predicted. Only positive *Z*-scores are shown. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Z-score comparisons, but did not provide reliability scores for their predictions so could not be compared in an equivalent fashion.

Similar comparisons with the *Xd* scores are shown in Figure 13. Again RR157 is the top predictor from the common subset. RR069 and RR131 (*L*/5 only) have slightly better *Z*-scores with fewer eligible targets. However, the results in Figures 12 and 13 should not be read as a ranking, head-to-head comparisons over common targets showed that none of the differences between groups were significant for either accuracy or *Xd* (figure not shown).

Contact predictors versus structure predictors

The comparison between structure predictors and contact specialists in CASP7 suggested that the best structure predictors were still better at predicting contacts than the best contact specialists.[10] This seems not to be the case in CASP8, despite following the same model contact selection procedure as in CASP7. We calculated *Z*-scores for each target at various length cut-offs and sequence separations using the predictions from the specialist groups and the inferred predictions from the 3D models. The plot of mean *Z*-scores for the accuracy of the predictions for the 12 targets in the evaluation (Fig. 14) shows that only two of the top twelve groups were structure predictors, TS489 (DBAKER) and TS453 (MULTICOM). This is an interesting figure, but the lack of targets means that the differences between the groups are again not statistically significant. For example, the difference between TS489 and the other top scoring groups is caused entirely by one target (T0405), while the advantage enjoyed by RR069 is in part because the group only had eligible predictions for five targets. Despite this the better RR groups had significantly better accuracy and *Xd* scores than approximately half of the structure prediction groups.

If *L*/10 is used to make the comparison, the contact specialists perform slightly better than they do when the *L*/5 cut-off is used. If the 12 to 23 sequence separation is used the structure predictors perform slightly better than when the greater residue sequence separation is used. In all cases, if the comparison is made with *Xd*, the contact specialists perform substantially better.

Some of the outstanding predictions in the CASP8 experiment are shown in Figure 15. There were many predictions with high accuracy in addition to those highlighted, but a number of the predictions, in particular for those targets that were mainly beta sheet targets, tended to cluster most of their predictions around a single contact (an example is shown in Figure 16). Although this tends to drive up the accuracy and *Xd* scores, the

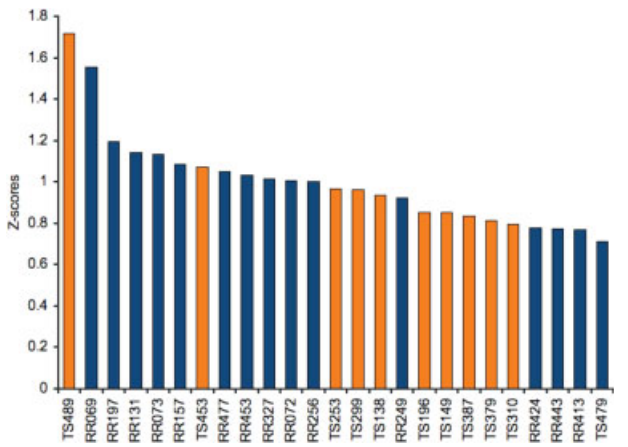

Figure 14

Accuracy for contact specialists and tertiary structure prediction groups. Mean *Z*-scores for the accuracy of contact predictions from the specialist contact groups (RRxxx in blue in the figure) and those inferred from the structure prediction groups (TSxxx, in orange). The eligibility cut off here is *L*/5, groups must have a minimum of four predicted targets. Just the 25 groups (TS or RR) with the best *Z*-scores are shown. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

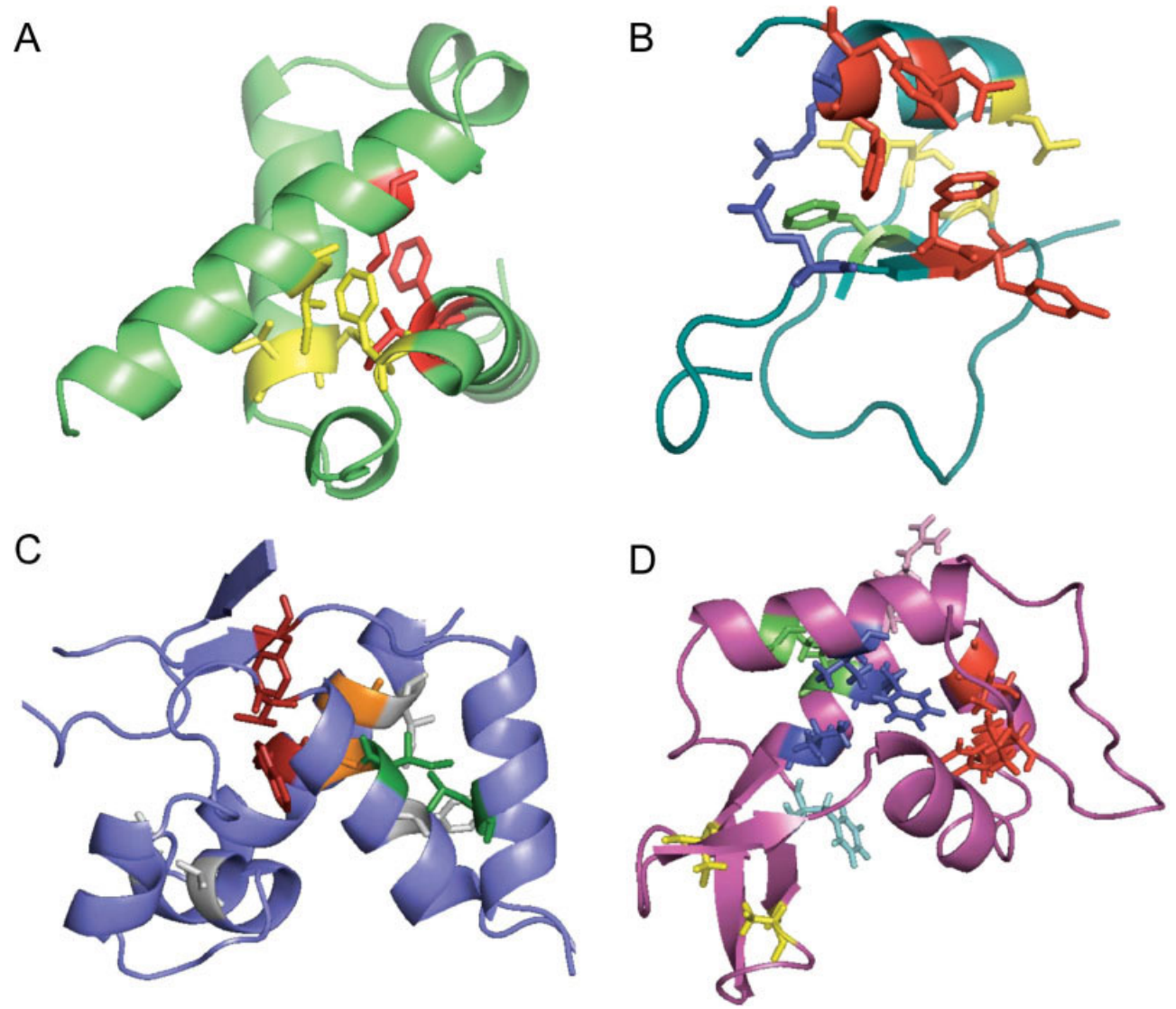

Figure 15

Four outstanding predictions. Predicted residues are mapped onto the target domains as sticks, residues predicted to be in contact are in the same color. If a predictor predicts several residues to be in contact together, the whole cluster is colored identically. False positive predictions are shown as nonpaired colors. Some false positives may be obscured because they also form true positives as part of a second predicted pair. Part **A** shows the prediction of RR072 at *L*/10 for target domain T0443-D1 (66.7% accuracy). Part **B** shows the prediction of RR032 for target domain T0510-D3 (50% accuracy at *L*/5). Part **C** shows the prediction of RR477 at *L*/10 for target domain T0465 (36.4% accuracy). Part **D** shows the prediction of RR131 for target domain T0443-D1 (62.5% accuracy at *L*/10).

predicted contacts are not so useful for prediction purposes as contacts that are predicted for the whole protein, particularly if the reason for the clustering is that the predictor is simply predicting all beta sheets to be sticky.

Progress with respect to CASP7

Because of the small sample sizes (there were even fewer targets in CASP8 than in CASP7) and potential differences in target difficulty, it is impossible to state with any certainty whether there has been any improvement with respect to CASP7. However, two separate pieces of evidence suggest that there has been some improvement. Firstly, the comparison between the specialist contact predictors and the contacts inferred from the structure prediction groups appear to suggest that a number of predictors have improved dramatically with respect to structure prediction groups. The equivalent of Figure 14 from CASP7 shows just two specialist contact groups among the top 25 in contact prediction accuracy (both Karplus groups). However, the improvement in CASP8 is so extreme that it suggests that much of the ground gained by the specialist contact predictors is likely to be a side effect of the small sample sizes.

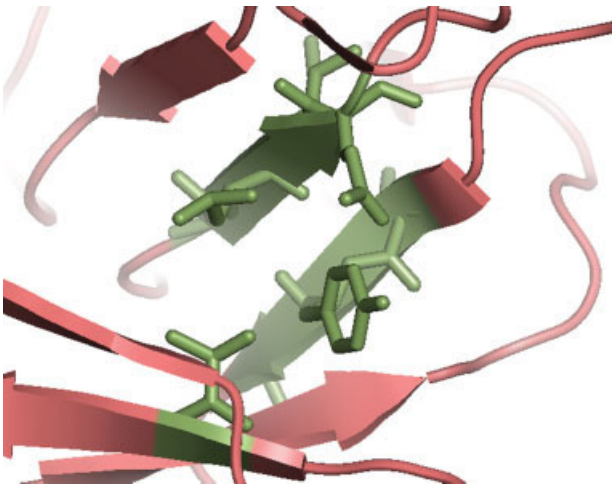

Figure 16

A typical prediction for a β-sheet structure. The first seven predicted residue contacts from group RR157 for target domain T0397-D1. Predicted residues are mapped onto the target domains as sticks. Most of the predictions are true positives, but most predictions are for the same beta strand pair. In fact 11 of the top 12 predicted pairs are between the two strands in the figure. In this particular case, this is a useful prediction, in spite of the redundancy of the predictions, because contact is also predicted with the strand at the bottom of the figure. The target has an open barrel conformation that no structure prediction groups predicted.

The second piece of evidence is that the SAM T06 server from the Karplus group[14] that was the best server in the CASP7 evaluation was maintained as a server in the CASP8 experiment. In CASP8, several predicting groups had consistently higher accuracy and *Xd* scores than the SAM T06 server, albeit over the small subset of targets, and these included groups RR157, RR158, RR131, RR072, RR453, and RR069.

All the better scoring groups use classifiers based on different machine learning methods with sequence information and predicted structural information. Group RR477 uses local structure predictions and regularized amino acid composition. RR072 extracts a rule set from sequence information data, predicted secondary structure and solvent accessibility. RR157 uses 2D profiles and predicted secondary structure and solvent accessibility. RR158 uses hidden Markov models, secondary structure predictions, and local descriptors from a protein structure backbone library. RR453 is a consensus predictor for servers RR069, RR443, and RR131 that use classifiers to predict contacts from a range of sequence features including profiles, secondary structure prediction, and contact potentials.

CONCLUSIONS

The two experiments detailed in this article both suffered from a similar problem, the lack of suitable targets. In the case of the residue–residue contact prediction experiment, the lack of FM and TBM/FM targets meant that the assessors could not reach any real conclusions. The numbers of targets in the experiment dropped by over 50%, and if trends from recent CASP experiments continue,[18] it looks as if the residue–residue contact prediction experiment will be extremely difficult to assess in CASP9.

There were sufficient targets in the domain prediction experiments to draw some conclusions about the efficacy of structural domain prediction. However, these conclusions were already clear from the CASP7 experiment and the results suggest that little has changed since then. Domain prediction methods are reliable when a template structure can be found that covers the domain boundary or something close to the domain boundary. Many prediction groups make reliable, good quality template-based domain predictions.

It is much less clear how well domain definition methods function in those cases where a domain prediction is more important, that is to say those cases where the domain boundary cannot be predicted from a template, in particular for those proteins that need to be split into their constituent domains in order to be modeled *de novo*. There were simply too few larger free modeling targets in the CASP7 and CASP8 experiments to draw any conclusions at all. Although there does seem to be room for improvement, there were some examples of successful *ab initio* predictions.

The statistical tests suggest that several groups (for example, DP118 and DP229) may have improved. However, the comparison with PDP as a predictor suggested that there have not been any great advances since CASP7 and that some predictors may even be performing worse than they were in CASP7.

The better scoring methods in CASP7 and CASP8 used some form of hybrid prediction that was based on structural templates and on sequence-based predictions of domains or of domain linkers. Where a reliable template was found the template took precedence over the sequence-based predictions. Regions that could be modeled based on templates were divided into domains by a range of automatic and nonautomatic strategies, for example, the Baker group servers used a variation on Taylor's method,[22] the Cheng group servers used PDP[20] and the CBRC-DP_DR human group split the models after visual inspection.

The strategy chosen to split the model structures into domains—to split or not to split—had some bearing on the NDO and domain boundary scores as assessor domain definitions are inevitably subjective. But we feel that the scores in this assessment do reflect the value of the predictions by the server groups. For example, those groups that tended to split the target into more domains than the assessors tended to do so by splitting core and secondary structure regions of the target structure, and

those groups that had a strategy of splitting few targets consistently failed to split targets such as T0445, T0457, and T0501 that were indubitably multiple domain targets.

In our summary of the CASP7 domain prediction experiment,[8] we suggested that CASP was not the right format to assess domain prediction and in particular the *ab initio* prediction of domain boundaries. We still believe that. Improving the accuracy of *ab initio* domain prediction is an important challenge, but there are too few difficult multiple domain targets in CASP.

There was an increase in the number of predicting groups in the residue–residue contact prediction experiment in this CASP, and despite the lack of targets in the experiment, there were indications that there has been some improvement in prediction accuracy for the second CASP running. Contact prediction accuracy is still quite low, but it appears that several groups may have made improved predictions in CASP8. It is difficult to pick out one single group as performing better than the rest but several groups appear at or near the top in all the different evaluations. However, the scarcity of targets in this category makes dependable comparisons almost impossible. Unfortunately, with such a small sample size, even the differences between the best scoring and worst scoring groups are not statistically significant.

In CASP7, it was suggested that certain types of structure were more amenable to prediction by contact prediction groups; there seemed to no similar correlation in this CASP.

The increase in the numbers of predictors coupled with the suggestions at the Cagliari meeting that at least one group had successfully used predicted contacts to aid in structure selection almost certainly means that there will be an increase in interest in using predicted contacts to aid *de novo* structure prediction or to score models. If that is the case, there are good reasons to have a continuous assessment of contact predictions in the style of EVA[23] or LiveBench.[24] It seems unlikely that contact prediction can continue in CASP unless more free modeling targets can be found.

REFERENCES

1. Richardson JS. The anatomy and taxonomy of protein structure. Adv Protein Chem 1981;34:167–339.
2. Kaminska KH, Baraniak U, Boniecki M, Nowaczyk K, Czerwoniec A, Bujnicki JM. Structural bioinformatics analysis of enzymes involved in the biosynthesis pathway of the hypermodified nucleoside ms(2)io(6)A37 in tRNA. Proteins 2008;70:1–18.
3. Chivian D, Kim DE, Malmstrom L, Bradley P, Robertson T, Murphy P, Strauss CE, Bonneau R, Rohl CA, Baker D. Automated prediction of CASP-5 structures using the Robetta server. Proteins 2005;53: 524–533.
4. Shiozawa K, Maita N, Tomii K, Seto A, Goda N, Akiyama Y, Shimizu T, Shirakawa M, Hiroaki H. Structure of the N-terminal domain of PEX1 AAA-ATPase. Characterization of a putative adaptor-binding domain. J Biol Chem 2004;279:50060–50068.
5. Sanchez-Pulido L, Valencia A, Rojas AM. Are promyelocytic leukaemia protein nuclear bodies a scaffold for caspase-2 programmed cell death? Trends Biochem Sci 2007;32:400–406.
6. Fan E, Baker D, Fields S, Gelb MH, Buckner FS, Van Voorhis WC, Phizicky E, Dumont M, Mehlin C, Grayhack E, Sullivan M, Verlinde C, Detitta G, Meldrum DR, Merritt EA, Earnest T, Soltis M, Zucker F, Myler PJ, Schoenfeld L, Kim D, Worthey L, Lacount D, Vignali M, Li J, Mondal S, Massey A, Carroll B, Gulde S, Luft J, Desoto L, Holl M, Caruthers J, Bosch J, Robien M, Arakaki T, Holmes M, Le Trong I, Hol WG. Structural genomics of pathogenic protozoa: an overview. Methods Mol Biol 2008;426:497–513.
7. Tai CH, Lee WJ, Vincent JJ, Lee BK. Evaluation of domain prediction in CASP7. Proteins 2005;61:183–192.
8. Tress M, Cheng J, Baldi P, Joo K, Lee J, Seo JH, Lee J, Baker D, Chivian D, Kim D, Ezkurdia I. Assessment of predictions submitted for the CASP7 domain prediction category. Proteins 2007;69:137–51.
9. Grana O, Baker D, MacCallum RM, Meiler J, Punta M, Rost B, Tress ML, Valencia A. CASP6 assessment of contact prediction. Proteins 2005;61:214–224.
10. Izarzugaza JMG, Graña O, Tress ML, Valencia A, Clarke ND. Assessment of intramolecular contact predictions for CASP7. Proteins 2007;69:152–158.
11. Skolnick J, Kolinski A, Ortiz AR. MONSSTER: a method for folding globular proteins with a small number of distance restraints. J Mol Biol 1997;265:217–241.
12. Li W, Zhang Y, Skolnik J. Application of sparse NMR restraints to large-scale protein structure prediction. Biophys J 2004;87:1241–1248.
13. Cheng J, Baldi P. Improved residue contact prediction using support vector machines and a large feature set. BMC Bioinformatics 2007;8:113.
14. Shackelford G, Karplus K. Contact prediction using mutual information and neural nets. Proteins 2007;69:159–164.
15. Wu S, Zhang Y. A comprehensive assessment of sequence-based and template-based methods for protein contact prediction. Bioinformatics 2008;24:924–931.
16. Rajgaria R, McAllister SR, Floudas CA. Towards accurate residue-residue hydrophobic contact prediction for alpha helical proteins via integer linear optimization. Proteins 2009;74:929–947.
17. Walsh I, Baù D, Martin AJ, Mooney C, Vullo A, Pollastri G. Ab initio and template-based prediction of multi-class distance maps by two-dimensional recursive neural networks. BMC Struct Biol 2009;9:5.
18. Tress ML, Ezkurdia I, Richardson JS. Domain definition and target classification paper, CASP8. Proteins 2009;77(Suppl 9):196–209.
19. Hadley C, Jones DT. A systematic comparison of protein structure classifications: SCOP, CATH and FSSP. Structure 1999;7:1099–1112.
20. Alexandrov N, Shindyalov IN. PDP: protein domain parser. Bioinformatics 2003;19:429–430.
21. Holland T, Veretnik S, Shindyalov IN, Bourne PE. A benchmark for domain assignment from protein 3-dimensional structure and its applications. J Mol Biol 2006;361:562–590.
22. Taylor WR. Protein structural domain identification. Protein Eng 1999;12:203–216.
23. Grana O, Eyrich VA, Pazos F, Rost B, Valencia A. EVAcon: a protein contact prediction evaluation service. Nucleic Acids Res 2005;33: W347–W351.
24. Bujnicki JM, Elofsson A, Fischer D, Rychlewski L. LiveBench-1: continuous benchmarking of protein structure prediction servers. Protein Sci 2001;10:352–361.