WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Contact: Assessment

Assessment of intramolecular contact predictions for CASP7

José M. G. Izarzugaza,[1] Osvaldo Graña,[1] Michael L. Tress,[1] Alfonso Valencia,[1] and Neil D. Clarke[2]*

[1] Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre, 3E-28029 Madrid, Spain

[2] Computational and Systems Biology, Genome Institute of Singapore, 138672 Singapore

ABSTRACT

Predictions of intramolecular residue–residue contacts were assessed as part of the seventh community-wide Critical Assessment of Structure Prediction experiment (CASP7). As in past assessments, we focused on contacts that lie far apart in sequence as these are likely to be more informative in predicting protein structure. One lab did somewhat better than others according to our assessment, and there is some reason to think that this lab's results represent progress over CASP6. In general, contacts inferred from 3D structural predictions are similar in accuracy to those predicted by contact prediction methods. However, contact prediction methods were more accurate for some targets.

Proteins 2007; 69(Suppl 8):152–158.

Key words: contact prediction; CASP7; intramolecular contacts.

INTRODUCTION

The prediction of intramolecular contacts could be helpful in predicting the three-dimensional structures of proteins. Structures might even be inferred directly if it were possible to predict a sufficiently large number of contacts with sufficient accuracy. It is not clear how many contact predictions would be required for structure prediction, but one estimate is that as few as one contact on average for every seven residues might be sufficient.[1] Of course, not all contact pairs are equally useful in constraining 3D models. Also unclear is what accuracy of contact prediction would be required for de novo structure prediction. However, contact predictions might be useful for selecting among alternative structures or for constraining conformational searches even if they are too few in number, or too inaccurate, to be used for de novo prediction.

The utility of contact prediction is not restricted to protein structure prediction. Techniques that use strictly evolutionary information (in the form of multiple sequence alignments) have been used to infer correlated mutations[2–12] and tree-determinant positions.[13] Interprotein-correlated mutations have been interpreted as physical sites of interaction.[14] It has been shown that these predictions can distinguish between correct and incorrect docking models.[15] Tree-determinant positions are those residues that are conserved between subfamilies of proteins, and these can be used to predict the interacting surface between protein and substrates, or between different proteins.

The numerical assessment of contact predictions in Critical Assessment of Structure Prediction experiment (CASP7) followed closely the methods established by some of the authors in past CASPs and CAFASPs, and have been implemented in the EVA contact evaluation server.[16,17] The only substantive change was in the way contact predictions were inferred from 3D predictions for comparison to the contact prediction accuracy of explicit contact prediction methods.

For the assessment of contact predictors, we restricted our analysis to Free Modeling targets (FM) and to targets that were on the borderline between FM and Template Based Modeling (FM/TBM). The exclusion of pure TBM targets eliminates contact pre-

The authors state no conflict of interest.

*Correspondence to: Neil D. Clarke, Computational and Systems Biology, Genome Institute of Singapore, 60 Biopolis Street, 138672 Singapore. E-mail: clarken@gis.a-star.edu.sg

Received 28 February 2007; Revised 22 May 2007; Accepted 26 May 2007
Published online 1 August 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21637

dictions that might be based purely on fold recognition. However, numerical assessments were performed for all targets, including TBM targets, and the data are available at http://casp.bioinfo.cnio.es.

METHODS

Target selection

The target classification used in the other CASP7 prediction categories was used here as well. All 15 FM targets and all 4 FM/TBM targets were used in the assessment. Pure TBM targets were not used to avoid the possibility of predictions that were based on comparative modeling. As one measure of target difficulty, we determined the number of homologs to the original full-length target sequences using BLAST/PSI-BLAST and the EBI 90% nrdb sequence database. Four FM targets lacked any homologs by this criterion at the time of assessment (T0309, T0287, T0314, T0353).

Contact prediction format and definition

Contact prediction groups submitted lists of residue pairs, as well as a probability estimate that each pair is in contact. The CASP contact format also asked predictors to specify the contact distance range for each pair, but this was not used in the assessment. Instead, contacts were defined as a Cβ–Cβ pair (Cα in the case of Gly) less than or equal to 8 Å apart. This follows the precedent of recent CASP contact prediction assessments, and was noted in the contact prediction format specification on the CASP7 website. For reasons of brevity and clarity we have used the group numbers assigned to predictors in all figures; the group names that correspond to these group numbers are shown in Table I.

Table I
Contact Group Numbers and Names

Group number	Group name
RR010	SAM-T06
RR042	Frishman
RR066	UF_GATORS
RR103	Huber_Torda
RR138	SVMcon
RR141	BETApro
RR154	GPCPRED
RR168	Distill
RR169	Meiler
RR176	CYKAY-AT-NTU-AND-III
RR230	Possum
RR271	BIME@NTU
RR296	PROFcon-Rost
RR389	SAM_T06_server
RR393	Distill_human
RR618	GajdaPairings
RR763	DISTILLFM

Contact lists and filters

Prediction groups submitted contact pairs based on the entire target sequence, but only contact pairs that lie within the official domain definition for the target were used in the analysis. Short range contacts have less value than long range contacts in constraining 3D structure predictions, so sequence separations were used to filter contact predictions. The criteria used for filtering contacts by sequence separation were slightly different than in past CASP assessments in that upper limits were specified so that different sequence separation thresholds do not include the same contacts. The three separation ranges used, and for which results are available at http://casp.bioinfo.cnio.es, are $6 \leq x < 12$, $12 \leq x < 24$, and $x \geq 24$. For all of the analyses discussed here, except those shown in Figure 5, a sequence separation of 24 or greater was used. This threshold was used extensively in the assessment of contact predictions in CASP6 and was chosen because it highlights sequence-distant contacts that are likely to be most useful for structure prediction.[16] The CASP7 assessment also followed the precedent of earlier CASP assessments in using target sequence length to define how many predicted contacts were evaluated for each target. For each target of length L, we separately calculated quality measures for sets of L/x predictions, where a variety of values of x were used in the range of 0.5–20. Predictions were sorted according to the predictor's probability estimates, and secondarily by residue number. After eliminating residue pairs whose sequence separation was less than the threshold, the top L/x predictions were used in the analysis. Figure 5 shows accuracy results for several groups at a variety of values of x and for two sequence separation thresholds. For all other analyses described here, we used an $L/5$ cutoff with a sequence separation of $\geq$24 for evaluating the performance of prediction groups. Values of L/x smaller than $L/5$ imply fewer predicted contacts and, as a result, less confidence in the analyses. For values higher than $L/5$, there are fewer groups who predicted a sufficient number of contact pairs. Numerical evaluation data for other combinations of parameters are available at http://casp.bioinfo.cnio.es.

Comparison to contacts inferred from 3D models

Results for contact predictors were compared to the predictions that might have been made by 3D structure predictors had they made contact predictions based on their 3D models. Residue pairs in 3D models were ranked by Cβ–Cβ distance (Cα for glycines) and the closest L/x contacts were considered to be contact predictions, provided the contact distance was less than or equal to 8 Å. This method differs from past CASP assessments which sampled randomly from amongst the full set of contacts that were 8 Å apart or less. We used only the "model 1" prediction from each group, since the contact predictors only submitted one contact list per target. For 3D models

that contained only Cα or backbone atoms, the Jackal package was used to calculate Cβ positions.[18]

Numerical evaluation criteria

All numerical evaluations were performed using a fixed number of contact predictions based on target sequence length, as described above. If the number of predicted contacts that met the sequence separation criterion was less than the length-defined threshold, then no assessment of that contact model was made. Predictions were evaluated using two measures described in previous CASP prediction assessment papers. One is accuracy (Acc), which is the fraction of predicted contacts that are found in the experimentally determined target structure. Acc = TP/(TP + FP), where TP = true positive and FP = false positive. The second metric, X_d, is a measure of how the distribution of distances that are observed for predicted contacts differs from the distance distribution expected by chance. It is defined as $X_d = \sum_{i=1}^{i=15} (Pip - Pia)/(di \times 15)$. There are 15 distance bins, starting from 0–4 Å ($i = 1$), and increasing in 4-Å increments up to 56–60 Å ($i = 15$) *Pip* is the fraction of predicted contacts that are in bin *i* and *Pia* is the fraction of all residue pairs that are in bin *i*. *di* is the upper limit of the distance range for bin *i*, normalized to 60. Note that the existence of *di* in the denominator weights the summation towards shorter contact distances. Larger values of X_d represent greater skewing of the predicted contacts towards shorter observed distances.

Two other measures used in previous CASPs were calculated and are available at the web site. One is the improvement in accuracy over a random prediction, and the other is coverage, defined as the fraction of structurally observed contacts that are predicted (TP/observed contacts). We did not use these values in the assessment because they can be determined directly from the accuracy of a prediction and the sequence length of the target, and are thus redundant for this purpose.

RESULTS

Prediction data sets

Seventeen groups participated in contact prediction in CASP7. Most groups submitted predictions for all, or nearly all, of the 19 FM and TBM/FM targets. However, roughly half of the predictions failed to meet the threshold number of contacts required for assessment. As a result, the number of targets used to assess contact predictors varied widely [Fig. 1(A)]. For the analyses presented here, *L*/5 contacts were required with a sequence separation of at least 24 amino acids (see Methods). This threshold has been used in past CASP assessments as well. For simplicity, we refer to contact predictions that meet these criteria as "eligible predictions" and predictions that fail to meet these criteria as "ineligible predictions." However these terms are not meant to imply that the predictors did anything incorrect. Although predictors could have anticipated our use of this threshold, there is no way for them to know for sure what the threshold number of contacts will be in advance. For example, in a target sequence that has been divided into domains for the purposes of CASP assessments, predictors need to meet the contact number threshold for the individual domain even though they do not know the domain boundaries in advance. For this reason, domain targets tend to have fewer eligible predictions [Fig. 1(B)].

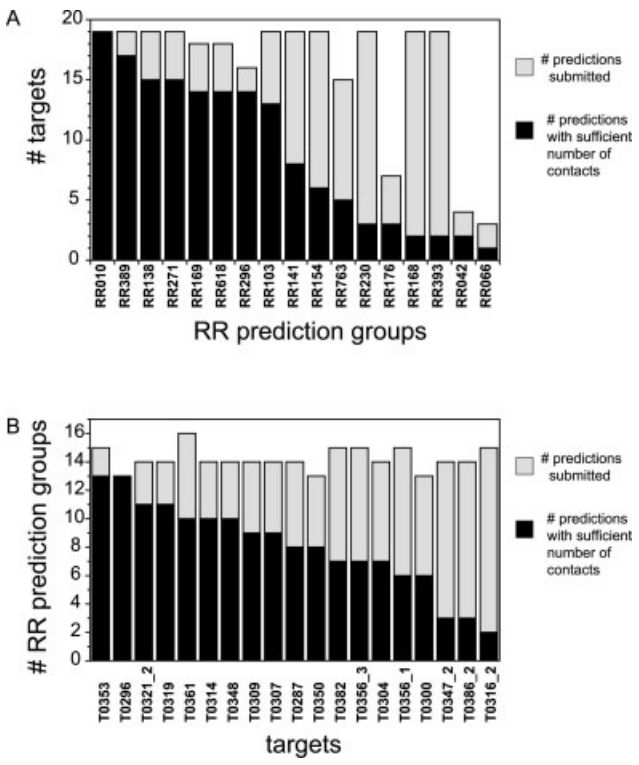

Figure 1

*Coverage of contact prediction targets by contact prediction groups (**A**). Number of targets for which each group submitted predictions (total height of histogram bars), and the number of predictions which met the L/5 threshold for contacts with a spacing of at least 24 amino acids (black bar). (**B**) Same as A, except the number of prediction groups who made predictions is plotted for each target.*

Prediction accuracy by target

Contact prediction accuracy varies from target to target, but is typically quite low (mean of 0.13 across all specialist (RR group) predictors and targets) (Fig. 2). For the four targets that appear to lack homologs (Methods), the average prediction accuracy is substantially lower (0.04). The 15 targets that have readily identified homologs have a mean contact prediction accuracy of 0.15.

This prediction accuracy is not significantly better, on average, than the accuracy of contact predictions inferred

DOI 10.1002/prot

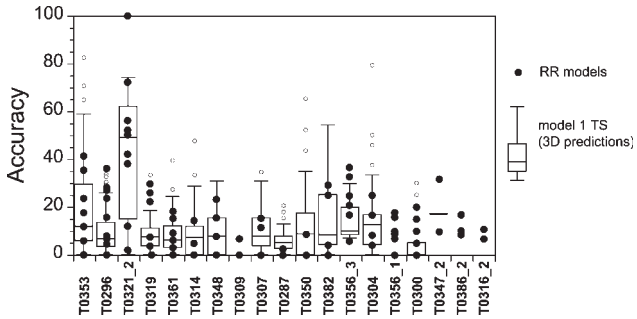

Figure 2

Accuracy of predictions for each contact prediction target. Accuracy values for specialist contact prediction groups (RR) are shown as black dots. In some cases, more than one group predicted contacts with the same accuracy. Accuracy values for contacts inferred from 3D models are shown as box plots, with outliers shown as open circles. As described more explicitly in Methods, accuracy is the fraction of predictions that are true positives.

from 3D structural models (model 1 predictions; mean of 0.14 across all targets and predictions). However, this average value obscures a notable difference among two classes of targets: those that are derived from target sequences that were split into multiple target domains by the assessors, and those for which the target sequence resulted in a single target structure. There were six targets in the FM and TBM/FM sets that were derived from multidomain target sequences, and for each of these targets the predictions inferred from 3D models were worse than the best contact predictions by specialist groups. In fact, very few 3D models were even evaluated for these targets because most of the models did not meet the $L/5$ criterion for number of contacts. Presumably this reflects difficulties in the 3D structure prediction of target sequences that are subsequently split into multiple target domains. In contrast, for the 13 targets derived from a singe-domain target sequence, 12 were predicted best by 3D models.

Predictions for T0321_2 are anomalous in their exceptional accuracy. This target was classified as a TBM/FM target because there is a template that covers part of the domain, and many of the 3D modeling groups did well on that part of the structure. This explains the exceptionally high median accuracy for contacts inferred from 3D models. It is less clear why the contact predictors did so well on this target. One group (RR618) had 100% accuracy on the top $L/5$ contacts they predicted (50 in this case). Although this was exceptional, relatively high accuracy was the rule for most groups. Of the 11 contact prediction groups that had eligible predictions for T0321_2, 9 had higher accuracies for this target than for any other. T0321_2 consists, in part, of a six-stranded beta sheet, five strands of which are parallel. Perhaps contacts for this kind of structure are more easily predicted. Some support for this comes from T0296, which is the only single domain protein for which the best contact predictions beat the best 3D model derived predictions. T0296 also has extensive parallel beta strand structure. Alternatively, the fact that T0321_2 was not only exceptionally well predicted but was also the only target that had a clear (albeit partial) template, suggests that prediction groups were able to take advantage of this, either implicitly or explicitly.

Prediction quality by group

Since the number of eligible contact predictions varies widely from group to group, we adopted two methods for assessing which groups did best. The first was to compare separately each pair of groups using only the targets for which those groups had predictions in common. A graphical representation of this analysis shows that group RR010 typically predicted more targets with higher accuracy than did other groups, followed very closely by group RR389 (Fig. 3). These two groups are both from the K. Karplus lab, group RR010 being the "human" group and RR389 the server.

We also asked which groups did best when averaging over all their eligible predictions, regardless how much of

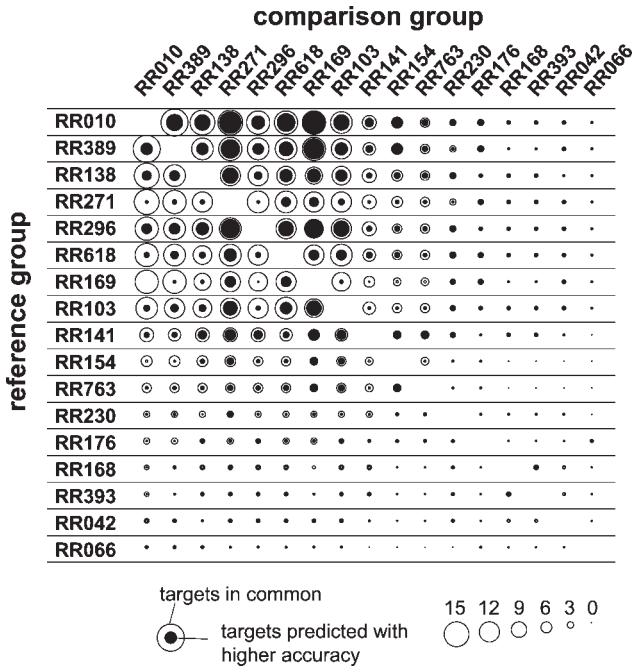

Figure 3

Pairwise comparison of RR groups. Each row compares a reference group with each of the other RR groups (comparison groups). For each cell in the table, the size of the outer circle reflects the number of predicted targets in common, and the inner, black circle reflects the number of targets for which the reference group had a higher prediction accuracy than the comparison group; note the scale at the bottom of the figure. Ties were counted as half a win. The rows (reference groups) have been sorted by the total number of targets in common with other groups.

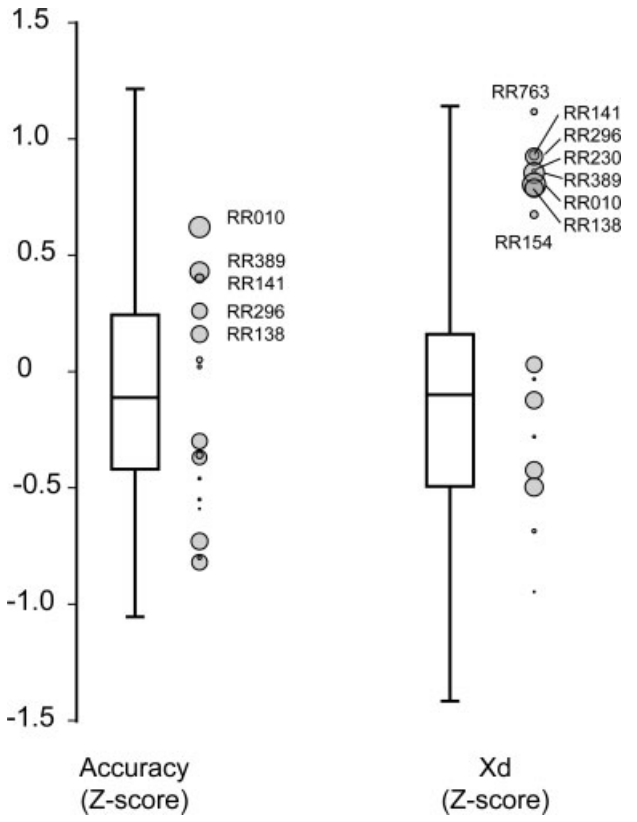

Figure 4

Accuracy and X_d Z-scores averaged over all targets for which a group had eligible predictions. RR groups are shown as circles, with the size of the circle reflecting the number of targets for which the group had eligible predictions. The box plots show the values for 3D structural models; for clarity, outliers are not shown. As noted in Methods, accuracy and X_d values were calculated for the top L/5 predictions with sequence spacing of 24 residues or greater.

an overlap in targets there was with other groups. To normalize for target difficulty we converted accuracy values for each target to *Z*-scores based on the mean and standard deviation of scores for all predictions for that target (RR groups and 3D models). We then determined, for each predictor, the average *Z*-score across all targets for which the group had eligible predictions. Figure 4 shows that groups RR010 and RR389 performed best by this criterion, as they did in the pairwise comparison. It is also notable that RR010 had eligible predictions for more targets (19/19) than did any other group, and that RR389 was similar (17) (Fig. 4). Groups RR141 (8 targets), RR296 (14 targets), and RR138 (15 targets) round out the set of top scoring predictors. The small sample size and large variances preclude any of the top groups from being shown to be best with statistical confidence.

We performed a similar analysis using the X_d metric (Fig. 4). X_d is a measure of how skewed the predicted contacts are towards shorter distances compared to the distribution of distances for all residue pairs that meet the sequence separation criterion (Methods). Compared to accuracy, X_d shows less discrimination amongst the top contact prediction groups, with the five groups mentioned above all having average *Z*-values between 0.80 and 0.93. The top scoring group by this criterion is actually a different group, RR763, with a *Z*-score of 1.11. However this group had eligible predictions for only 5 targets. Other groups near the top were RR230 ($Z = 0.86$; 3 predictions) and RR154 ($Z = 0.67$; 6 predictions). In general, contact predictors do better relative to 3D modelers using the X_d metric than they do using accuracy (compare panels A and B in Fig. 4). However, even by X_d, the best contact predictions inferred from 3D models outperform the best predictions from contact prediction groups.

Effect of alternative minimum spacing and assessed contact numbers

The analyses presented here and at the CASP7 meeting were based on a sequence separation of at least 24 amino acids, using the *L*/5 top predictions from each group for each target. To get some sense for how these parameters might have affected the analysis, we determined the average prediction accuracy for each group using two different definitions of sequence-remote contacts ($\geq$24 and $\geq$12 amino acids) and six different cutoffs for the number of contact predictions assessed (Fig. 5). On the

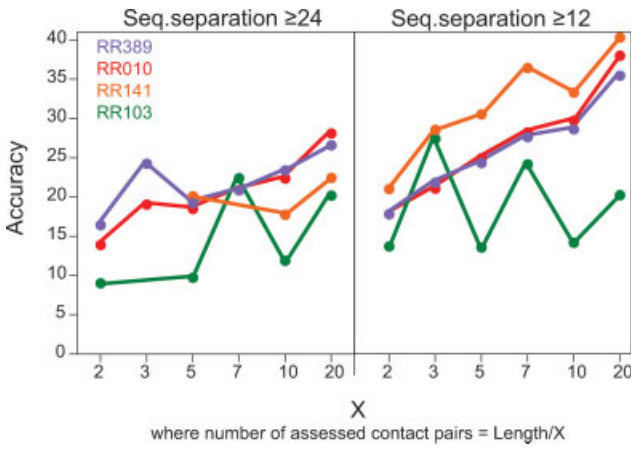

Figure 5

Mean accuracy for two different sequence separation criteria and for six different numbers of assessed contacts. For purposes of this analysis, only groups with eligible predictions for at least a third of the targets were considered. Of the 17 groups, 4 had the highest mean accuracy under at least one of the 12 combinations of sequence separation and L/x value. Just those four groups are shown here. Circles that are missing from this plot indicate that there were an insufficient number of targets with eligible predictions when evaluated with the indicated parameters. (For example, missing points for RR103 and RR141 at a sequence separation of 24 and an x value of 3 indicates that there were eligible predictions for fewer than a third of the targets).

DOI 10.1002/prot

whole, at a sequence separation of 24 or greater, the Karplus lab groups (RR010 and RR389) perform best across the full range of contact number thresholds, consistent with the more detailed analyses we have presented here using a contact number of *L*/5. However, when a sequence separation of 12 amino acids or more is used, group RR141 (BETAPro) performs better than all others (Fig. 5). This shorter sequence separation appears to favor the prediction of beta sheet contacts, which is what BETAPro is designed to do.

For example, at a sequence separation of 12 amino acids or more and a contact number threshold of *L*/7, RR141 has eligible predictions for 12 of the 19 targets. These 12 targets include only one all-helical protein. In contrast, among the seven targets for which RR141 does not have an eligible prediction, most are all helical. In general contact prediction methods perform better for sheet-containing structures rather than all-helical structures. Group RR010, which had eligible predictions for all 19 targets at the *L*/7 cutoff, had an average accuracy of 32% for the targets predicted in common with RR141 (i.e., predominantly beta-sheet), but an accuracy of only 22% for the other targets (predominantly helical). In this particular case the bias towards sheet-containing proteins means that we are only able to compare methods over a subset of what might be considered "easier" targets. Clearly, the choice of whether to use 12 or 24 residue spacing, or some other number altogether, can have a substantial effect on the assessment.

Has there been progress?

It is difficult to be sure whether progress has been made in contact prediction since CASP6, due to small sample sizes and large variances. CASP7 targets may also have been more difficult than in CASP6. Several CASP participants had the sense that targets in CASP7 were more likely to have few or no homologs, or to have homologs that were less informative. It is certainly the case, of course, that the number and diversity of homologs will have an effect on target difficulty. This is reflected in the low accuracy scores for the four FM targets that lack homologs (0.04 vs. 0.15 for targets with homologs). In an attempt to validate the view that targets were in general more difficult in CASP7, we compared the number of PSI-BLAST hits per target obtained during the CASP6 analysis with those obtained for CASP7. We were not able to conclude from this analysis that targets were harder, in part because somewhat different protocols were used for the homology searches in the two CASPs. More important, though, is an inability to objectively define how difficulty varies with the number and diversity of homologs. The relative difficulty of a target for a particular predictor depends in part on how that predictor uses the information in homologous sequences.

In an attempt to normalize across the CASP6 and CASP7 targets, we used results obtained from the ProfCON server to predict contacts for the CASP7 targets.[19] The ProfCON contact prediction algorithm has not changed since CASP6, so its prediction accuracies on CASP6 and CASP7 targets can be used as a rough measure of target difficulty. In CASP6, the ProfCON server was used by the Rost group, which was deemed to be one of three groups that were a bit better than others, and in CASP7 it was used by group RR296. Using a subset of New Fold (NF) targets that formed the basis for much of the assessment in CASP6, ProfCON has a mean accuracy of 21.8, quite close to the best mean accuracy of any group (22.6). ProfCON did not perform as well on the 15 FM targets from CASP7 (mean accuracy of 11.3) suggesting that the targets in CASP7 were harder. In addition, group RR010 did quite a bit better on the CASP7 targets than did ProfCON (15.5 vs. 11.3). Thus, it can be argued that RR010, and perhaps other groups, may be doing better, when corrected for target difficulty, than anyone did in CASP6. It must be kept in mind, of course, that the variances in accuracy values across targets are very large, so that a different set of criteria for selecting targets for comparison could have produced a different result. In addition ProfCON, like any server, has different strengths and weaknesses for different types of targets. To reduce the effects of these biases, it would be helpful in the future to have multiple servers that are left unchanged from one CASP to the next.

DISCUSSION

The total number of groups participating in contact prediction in CASP7 (17) was about the same as in CASP6 (16). It is difficult to know exactly how much new interest there is in contact prediction because some labs participate as multiple groups, and at least one group in CASP7 was headed by a former member of a CASP6 group. Nevertheless, there do appear to be a few new groups, which presumably injected new ideas into the problem. In addition, while the Karplus lab (RR010/RR389) did participate in contact prediction in CASP6, they tried different approaches this time and were much more successful. This should encourage other groups who may be considering participation in future CASPs.

As in some other CASP categories, we are unable to demonstrate progress with statistical confidence due to small sample sizes, differences in target difficulty, and a lack of frozen servers and data sets with which to compare new methods. Nevertheless, there is reason for encouragement as RR010 appears to outperform the Rost ProfCON program, one of the three predictors considered to be tied for best at CASP6.

A more provocative question is how much contact prediction needs to progress to make a contribution to

de novo structure prediction. Perhaps the utility of contact prediction will lie in choosing among alternative models. There do seem to be categories of targets in which the best contact predictions are better than the best predictions inferred from 3D models. Based on the targets and predictions analyzed in CASP7, contact predictions are perhaps most likely to add value for those targets that are derived from multidomain proteins, and perhaps for those targets that have parallel beta sheets.

As in all the CASP prediction categories, changes in thinking about how predictions might be used in the real world require a re-thinking of the metrics that are used to evaluate CASP predictions. Accuracy and X_d seem to be good overall measures of prediction quality, and have served contact prediction assessments well. In addition, contact predictors are by now familiar with the use of a length-dependent criterion for the number of contacts to be assessed. In principle, this criterion allows predictors to estimate how many contact pairs may need to be predicted. Unfortunately, without knowing the exact domain boundaries for the target, it is impossible to know exactly how many contacts need to be predicted, and what residue range to use. As a result, almost half of the submitted prediction lists were rejected for having too few predictions. A second disadvantage of the length-dependent threshold is that it has different effects on short and long proteins. Relative to protein size, short proteins have fewer structurally observed contacts that meet the 24 amino acid sequence separation criterion. Therefore, to achieve the same level of prediction accuracy, predictions for small proteins need to have greater coverage of observed contacts than do predictions for larger proteins.

Given the accuracy and coverage that might be required to achieve complete de novo structure prediction, it seems unlikely that contact prediction will compete successfully with fragment assembly methods and other techniques that may be developed in the future. However, having a smaller number of contacts that are predicted with greater confidence could still add value to other methods for structure prediction. If that is the case, it may be useful in future CASPs to consider measures of accuracy and coverage that do not require the predictor to predict a certain number of contacts. Perhaps contact predictors could even be asked to use their predictions to participate in quality assessment of other groups' 3D predictions.

ACKNOWLEDGMENTS

We are grateful to the organizers and participants of CASP, and to the CASP Prediction Center, particularly Andriy Kryshtafovych. The contribution of the CNIO team was partially funded by the GeneFun project (LSHG-CT-2004-503567). N.D.C. was supported by the Agency for Science, Technology and Research (Singapore).

REFERENCES

1. 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.
2. Altschuh D, Lesk AM, Bloomer AC, Klug A. Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. J Mol Biol 1987;193:693–707.
3. Altschuh D, Vernet T, Berti P, Moras D, Nagai K. Coordinated amino acid changes in homologous protein families. Protein Eng 1988;2:193–199.
4. Gobel U, Sander C, Schneider R, Valencia A. Correlated mutations and residue contacts in proteins. Proteins 1994;18:309–317.
5. Taylor WR, Hatrick K. Compensating changes in protein multiple sequence alignments. Protein Eng 1994;7:341–348.
6. Pollock DD, Taylor WR, Goldman N. Coevolving protein residues: Maximum likelihood identification and relationship to structure. J Mol Biol 1999;287:187–198.
7. Tuffery P, Darlu P. Exploring a phylogenetic approach for the detection of correlated substitutions in proteins. Mol Biol Evol 2000; 17:1753–1759.
8. Fukami-Kobayashi K, Schreiber DR, Benner SA. Detecting compensatory covariation signals in protein evolution using reconstructed ancestral sequences. J Mol Biol 2002;319:729–743.
9. Dekker JP, Fodor A, Aldrich RW, Yellen G. A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments. Bioinformatics 2004;20: 1565–1572.
10. Fares MA, Travers SA. A novel method for detecting intramolecular coevolution: Adding a further dimension to selective constraints analyses. Genetics 2006;173:9–23.
11. Kundrotas PJ, Alexov EG. Predicting residue contacts using pragmatic correlated mutations method: Reducing the false positives. BMC Bioinformatics 2006;7:503.
12. Fodor AA, Aldrich RW. Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. Proteins 2004;56:211–221.
13. Casari G, Sander C, Valencia A. A method to predict functional residues in proteins. Nat Struct Biol 1995;2:171–178.
14. Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. Correlated mutations contain information about protein-protein interaction. J Mol Biol 1997;271:511–523.
15. Tress M, de Juan D, Grana O, Gomez MJ, Gomez-Puertas P, Gonzalez JM, Lopez G, Valencia A. Scoring docking models with evolutionary information. Proteins 2005;60:275–280.
16. Grana O, Baker D, MacCallum RM, Meiler J, Punta M, Rost B, Tress ML, Valencia A. CASP6 assessment of contact prediction. Proteins 2005;61(Suppl 7):214–224.
17. Grana O, Eyrich VA, Pazos F, Rost B, Valencia A. EVAcon: A protein contact prediction evaluation service. Nucleic Acids Res 2005; 33:W347–W351. Online issue.
18. Petrey D, Xiang Z, Tang CL, Xie L, Gimpelev M, Mitros T, Soto CS, Goldsmith-Fischman S, Kernytsky A, Schlessinger A, Koh IY, Alexov E, Honig B. Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. Proteins 2003;53(Suppl 6):430–435.
19. Punta M, Rost B. PROFcon: Novel prediction of long-range contacts. Bioinformatics 2005;21:2960–2968.

DOI 10.1002/prot