CASP6 Assessment of Contact Prediction

Osvaldo Graña,[1] David Baker,[2] Robert M. MacCallum,[3] Jens Meiler,[4] Marco Punta,[5] Burkhard Rost,[5] Michael L. Tress,[1] and Alfonso Valencia[1]*

[1]*Protein Design Group, Centro Nacional de Biotecnologia (CNB-CSIC), C/Darwin 3, Cantoblanco, Madrid, Spain*
[2]*University of Washington, Department of Biochemistry, Seattle, Washington*
[3]*Stockholm Bioinformatics Center, Stockholm University, Stockholm, Sweden*
[4]*Department of Chemistry, Center for Structural Biology, Vanderbilt University, Nashville, Tennessee*
[5]*CUBIC, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York*

ABSTRACT **Here we present the evaluation results of the Critical Assessment of Protein Structure Prediction (CASP6) contact prediction category. Contact prediction was assessed with standard measures well known in the field and the performance of specialist groups was evaluated alongside groups that submitted models with 3D coordinates. The evaluation was mainly focused on long range contact predictions for the set of new fold targets, although we analyzed predictions for all targets. Three groups with similar levels of accuracy and coverage performed a little better than the others. Comparisons of the predictions of the three best methods with those of CASP5/CAFASP3 suggested some improvement, although there were not enough targets in the comparisons to make this statistically significant. Proteins 2005;Suppl 7:214–224.**
© 2005 Wiley-Liss, Inc.

Key words: contact prediction evaluation; interresidue distances; CASP6 assessment

INTRODUCTION

Contact prediction has traditionally been considered an avenue for the prediction of protein structures. The classical view considers that methods commonly used in NMR such as distance geometry and simulation techniques can be used to predict protein structures using predicted contacts as restraints. It has been suggested that for small proteins just one correctly predicted contact for every seven residues would be enough to predict approximate models.[1] Something that is less clear is what level of error in predicted contacts can be tolerated in the prediction of reasonable models.

Another option for the use of predicted contacts is to use them to filter potential models generated with threading methods,[2] or their direct inclusion in prediction strategies to restrict the initial search space.[3]

In recent years different methods have been developed for the direct prediction of protein three-dimensional contacts, including approaches based on the study of variations in multiple sequence alignments,[4–12] molecular dynamics,[13,14] and the application of various machine learning techniques trained on real contact maps.[15–22]

In contact prediction as well as in other fields of structural bioinformatics the evaluation of current methods is essential to spark progress and concentrate resources in the most interesting new developments. As in previous CASP[23–26] editions, contact prediction methods have been evaluated as an independent category in CASP6.

For the evaluation of contact prediction we have used the definition of targets common to all the CASP6 categories,[27] and analyzed structures corresponding to the well-defined structural domains obtained by expert analysis of the targets.

The numerical criteria for the evaluation of the results are essentially the same as those used in past editions of CAFASP,[28,29] and are available on the EVA contact evaluation server.[30]

To gain perspective on the significance of the contact prediction methods we have compared the results of the contact prediction specialists with those of other CASP6 structure prediction methods (i.e., homology modeling, fold recognition, and ab initio methods), using the same contact prediction evaluation criteria.

Note that it is not common in CASP publications to have a combined paper with assessment and methods together, but in this case the organizers felt that as there were three groups with similar levels of performance representing three different methodologies, there should be only one paper, and it should include both the assessment and brief descriptions from the three groups.

MATERIALS AND METHODS

Input Information

The contact specialist methods provided as output a file with a list of pairs of residues predicted to be in contact, the distance range between the C-beta atoms for each

Grant sponsor: BIOSAPIENS; Grant number: LSHC-CT-2003-505265; Grant sponsor: GENEFUN; Grant number: LSHG-CT-2004-503567 (to O.G., M.T., and A.V.); Grant sponsor: National Institutes of Health (NIH); Grant number: R01-GM64633; Grant sponsor: the National Library of Medicine (NLM); Grant number: R01-LM07329 (supporting the PROFcon).

*Correspondence to: Alfonso Valencia, Protein Design Group, Centro Nacional de Biotechnologia (CNB-CSIC), c/Darwin 3, Cantoblanco, 28049 Madrid, Spain. E-mail: Valencia@cnb.uam.es

Received 22 April 2005; Accepted 1 July 2005

Published online 26 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20739

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version.

© 2005 WILEY-LISS, INC.

residue pair and the probability of each assigned distance range.

For the other methods, contacts were directly obtained from the 3D coordinates. In those cases in which only the C-alpha trace was available, the side-chain positions were reproduced with the MaxSprout[31] software.

Specialist contact groups submitted long lists of contacts. To equivalence the predictions we sorted them by the probability values and we considered various levels of prediction proportional to the length of the protein sequence (L—number of residues in the target), including L/10, L/5, L/2, L, and 2L top contacts, taken from the list of predicted contacts. As a comparison, L/5 in this data set was 26.4% of native contacts. However, this is only an average; in the targets used in the data set L/5 ranged from 13.4 to 90.4% of native contacts.

As predictions with coordinates do not have assigned contact probability values, we sampled randomly an equivalent number of contacts when considering contacts from 3D model structures.

Contact Definition

Pairs of residues in the experimental structure were considered to be in contact if the distance between their C-beta atoms (C-alpha for Gly) was less than or equal to 8 Å.

Sequence Separation

We have defined three different regions of the contact map for the evaluation of the methods: short-range contacts, as those in which the members of the pair were separated by at least six residues in sequence, medium-range contacts as those separated by at least 12, and long-range contacts, separated by at least 24 residues along the sequence.

Evaluation Criteria

Predictions were evaluated with the following measures: accuracy (Acc), coverage (Cov), improvement over random (Imp), and Xd.

Accuracy is calculated as TP/(TP + FP), and coverage as TP/(native contacts), where TP = true positive and FP = false positive.

Improvement over random is calculated as the ratio between accuracy and the accuracy of a random prediction. We obtain the accuracy of a random prediction assuming that all the possible pairs of the experimental structure at the corresponding ranges of sequence distance separation are contacts (C); therefore, AccRand =(native contacts)/C.

$$\mathrm{Xd} = \sum_{i=1}^{i=15} \frac{(\mathrm{P}ip - \mathrm{Pia})}{(di * 15)} \qquad (1)$$

Xd[32] represents an evaluation of the proximity of the predicted contacts rather than a direct evaluation of the physical contacts, that is, the difference between the distance distribution of the predicted contacts and the all-pairs distance distribution in the 3D target structure.

There are 15 distance bins covering the range from 0 to 60 Å. The sum runs for all the distance bins. *di* is the distance representing each bin, its upper limit (normalized to 60). *Pip* is the percentage of predicted pairs whose distance is included in the *i* bin. Pia is the same for all the pairs.

When the average distance between predicted residue pairs is less than the average distance between all residue pairs in the structure, Xd is >0. Xd = 0 shows no separations between the distance populations.

We also implemented the evaluation of contacts proposed by Ortiz et al.,[3] where we measure the percentage of predicted contacts that are within a certain sequence separation (delta) of the correct experimental contact. Although for simplicity all the results are presented for delta = 0 (corresponding to real contacts), we have evaluated predictions for values of delta ranging from 0 to 5 (data not shown).

RESULTS

Evaluation Scope

We have evaluated the performance of all the groups over all the targets (see additional data on Web, http://www.pdg.cnb.uam.es/CASP6_ContactPredictionEvaluation/index.html). The most useful comparative set corresponds to the NF targets for which additional sequence or structural information was not available, that is, T0201, T0202_2, T0209_2, T0216_1, T0216_2, T0238, T0241_1, T0241_2, T0242, T0248_2, and T0273. These were the NF targets at the time of the Gaeta meeting in December 2004 and are described elsewhere in the issue.[27]

The general ranking of the prediction methods is shown in Figure 2. We have represented the results of those methods (6 out of 16) that submitted predictions for the complete set of 11 NF targets (the number of targets predicted by each group is available on the Web pages).

As in CASP rounds 2, 4, and 5, contact specialists were encouraged to predict contacts for residues within 8 Å. Two groups used a different cutoff: RR089 (4.5 Å) and RR348 (12 Å). To simplify the analysis their predictions were evaluated as if they had been predicted with a cutoff of 8 Å.

For those groups that predicted more than one model per target, we evaluated all of them (see Web), but only the first model was considered for the comparison of the performance among the different groups.

We focused on long range contacts, because these better define the essential characteristics of a fold and should be more useful for the prediction of three-dimensional structures. We took for the comparative study a number of pairs equivalent to L/5 as the different contact predictors tended to predict a different number of pairs and we wanted to be able to make comparisons with the predicted 3D models. Results for non long-range contacts are available on the Web.

Different Number of Predicted Contacts

It is important to note that there are big differences in the number of contacts predicted by the various groups

TABLE I. Total Number of Pairs Predicted by the Specialists That Submitted Contact Predictions for all the NF Targets

Targets	Target length	Total number of pairs predicted by the specialists											
		RR011		RR012		RR089		RR100		RR301		RR348	
		> = 24	*> = 12*	*> = 24*	*> = 12*	*> = 24*	*> = 12*	*> = 24*	*> = 12*	*> = 24*	*> = 12*	*> = 24*	*> = 12*
201	94	94	94	8	59	1	17	1432	1844	122	152	359	599
202_2	124	148	148	50	127	8	24	1056	1571	71	158	892	1350
209_2	61	16	16	3	7	6	9	661	932	11	41	52	126
216_1	209	140	140	30	91	22	47	3159	4381	179	388	979	1334
216_2	213	208	208	66	182	61	77	1893	2851	95	192	2582	3037
238	181	183	183	48	156	32	38	369	510	109	187	862	1185
241_1	117	61	61	45	98	15	25	952	1441	64	108	338	622
241_2	119	41	41	5	30	13	29	979	1378	34	147	552	690
242	115	116	116	16	78	39	51	1809	2431	114	164	706	1081
248_2	87	49	49	27	95	7	12	533	951	23	74	182	325
273	186	183	183	62	136	60	76	3485	4779	173	262	1240	1818
Total		1239	1239	360	1059	264	405	16328	23069	995	1873	8744	12167

Numbers are shown for long-range and medium-range contacts.

(Table I). In particular group RR100 predicted many more contacts than the others.

Figure 1 shows three examples of long-range contact predictions mapped as a contact network onto the 3D structures of the targets to illustrate the contrast between the full set of predictions and the top L/5 predictions for each target. In the first example [Fig. 1(a) and (b)] group RR018 predicted 11 contacts correctly out of a total set of 32 (with an accuracy of 0.347) for target T0201, but predicted 7 out of 18 correctly for the list of the top L/5 contacts (Acc = 0.388). The second example shows the prediction of group RR301 for target T0273. Here they predicted 27 out of 173 contacts correctly with their full list and 8 out of 37 when only the L/5 selection was taken into account (Acc = 0.156 and Acc =0.216, respectively). The last example shows the prediction from group RR100 for target T0273, where the number of predicted contacts in the full list was very high (3485) and the accuracy very low, but where the accuracy improved more than 10-fold when the top L/5 were evaluated.

As suggested in Figure 1 the proportion of correctly predicted contact pairs is generally higher at L/5, indicating a relation between the probability values of the various methods and the accuracy of the predictions.

The prediction by RR018 for target T0201 [Fig. 1(b)] identified a contact between two beta-strands separated by 79 residues in sequence. The prediction by RR301 for the target T0273 was able to reflect the contact between two pairs of beta-strands [Fig. 1(d)]. One of the pairs of the beta-strands separated by 24 and the other by 31 residues in sequence. These contacts between beta-strands were also present in the prediction by group RR100 for the same target [Fig. 1(f)].

Comparing Performances

Three groups, MacCallum-GPCpred (RR012), Baker (RR100), and Rost-PROFcon (RR301), presented results of similar quality, considering the balance between accuracy and coverage. They also used different and interesting technical approaches.

For the set of NF targets, their average accuracies ranged from 15.93 to 22.56%, and their coverage of predicted experimental contacts ranged from 3.26 to 5.73% for long-range contacts, with six to eight points of improvement over random (Fig. 2). We observed that the Xd values correlate fairly well with accuracy values.

Other groups might also have been considered for selection because the differences between the top scoring groups were not always statistically significant. Two groups in particular, RR011 and RR089, have levels of accuracy and Xd close to those of the representative groups. Several groups did not submit predictions for all the NF targets. For each one of them we compared their results with the other groups using a common subset of predicted targets. Only one group, RR018, was able to perform better in accuracy than the three best groups in the corresponding subset (two NF targets, see supplementary data on Web) but these results were not statistically significant.

Description of the Three Representative Methods

RR012 (MacCallum-GPCpred)

The server version of GPCpred (RR012) is based on a published contact prediction method,[20] with some modifications.

An important component of the method is the initial preprocessing of sequence profiles generated for the target sequence using PSI-BLAST.[33] Windows of sequence profiles, for instance as used in secondary structure prediction, contain a lot of information: for example 300 values for a window of 15 residues. We have used Kohonen-style Self-Organizing Maps (SOMs) to reduce this high-dimensional data into a more manageable form: a discrete location on a 3D SOM grid. We found that this dimension-reduced sequence profile window information was particularly informative when converted into an RGB color code and displayed on (known) 3D structures in an interactive viewer. Most noticeable was the near-identical coloring of some neighboring parallel strands, giving a striking striped appearance. A few instances of correlated color patterns in antiparallel strands were also observed. We suggested

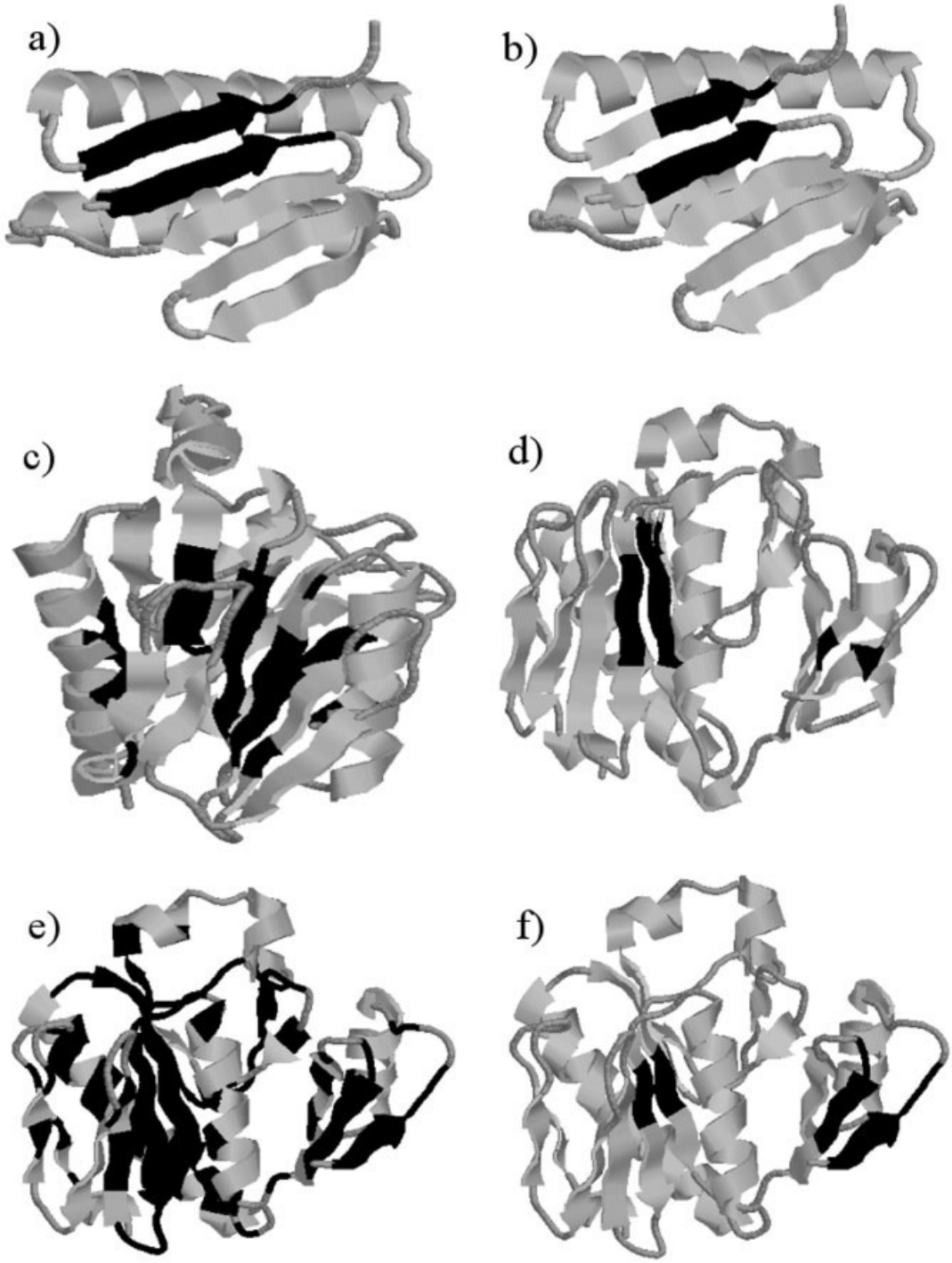

Fig. 1. Three examples of 3D target structures with the full list of contact predictions (left column) and the L/5 predictions (right column) mapped onto the structures. In (**a**) and (**b**) the predictions from group RR018 are mapped onto target T0201. In (**c**) and (**d**) group RR301 predictions are mapped onto T0273. In (**e**) and (**f**) the predictions from group RR100 are mapped onto T0273.

that these observations are explained by fact that two neighboring strands will quite often pass through similar structural environments. Because structure is a major constraint on sequence evolution, this could explain why similar sequence patterns are seen in certain strand neighbors.

To test the generality of this phenomenon, we decided to produce a general purpose contact predictor (as opposed to a specialized strand pair predictor) that uses only our simplified sequence profile window information as input. Because protein architecture and folding are not well understood, and because our input data is abstract and

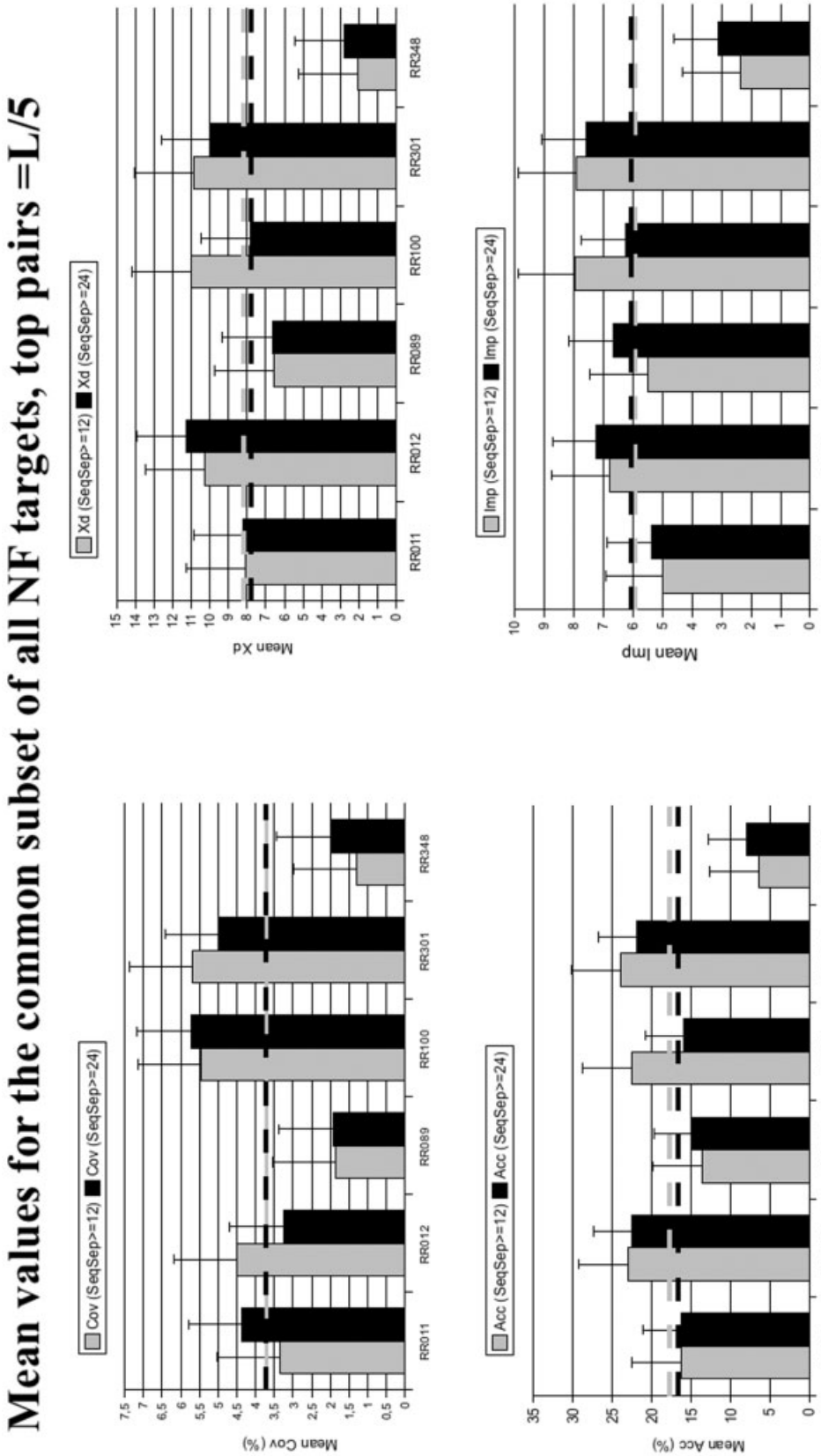

Fig. 2. Mean values of accuracy, coverage, Xd, and improvement over random for those contact specialists that made predictions for the whole set of NF targets. Dashed lines represent average values.

nonphysical, we use a machine learning method to make the decision about which residue pairs are in contact. We chose genetic programming (GP) because it seeks a rule-based solution rather than an optimal set of parameters to a predefined model. Our GP approach to this problem starts with a population of random "contact predictor" functions, which take as input the SOM-processed sequence profile information and output a contact map (via an intermediate residue–residue distance matrix). The individuals that produce more accurate predictions on our training set (measured with the standard Acc measure for a specified fraction of contacts and residue separation) get more opportunities to reproduce in the genetic algorithm. After many CPU hours, a competitive contact predictor is evolved.

Server method details

For a target sequence of length L, the inputs to the contact predictor function are two $L \times 3$ matrices of values corresponding to 3D SOM coordinates obtained by mapping overlapping windows of sequence profiles using window sizes 1 and 15. Using a grammar-based GP system (http://perlgp.org) we can ensure that all the evolved functions return a $L \times L$ "distance" matrix. This is made easy using the high level matrix manipulation functions of the Perl Data Language (PDL, http://pdl.perl.org). The functions allowed in the evolving code include the standard arithmetic operators and the functions abs, log, and sin. It is also possible for the evolved code to rotate matrices along their long edges and to perform simple statistics, such as mean, median, min, and max. Residue separation information is also available implicitly through the provision of "residue number vector." This all-at-once calculation of residue distances is the biggest change from the original GPCpred approach, in which pairwise "distance" calculations were only performed for a subset of residues.

The evolutionary algorithm was run on 20 machines for just over 4 weeks with occasional migration between populations. The fitness function (Acc) was calculated using L/2 predicted contacts (those with the smallest "distance") and a sequence separation of 8. After this, the 20 "best ever" and 20 "best of the final generation" predictors were saved. The fitness function was then changed so that it used a sequence separation of 24, and the runs were then continued for seven more days after which another 20 + 20 predictors were saved, giving a total of 80 evolved contact predictors. To avoid the arbitrary choice of one predictor over another, we produced a simple consensus predictor from all 80 predictors. Each component predictor assigns a distance-based rank to each residue pair (the closest pair is ranked 1). These ranks are then averaged over the 80 predictors.

The lowest mean-ranked residue pairs (with separation = 8 residues) are sent back to the user in CASP6 text format and a contact map image is available via the Web. The URL for the GPCpred server is: http://www.sbc.su.se/~maccallr/contactmaps.

RR100 (Baker)

Metaservers[34,35] outperform all primary automated fold recognition methods, in particular, when it comes to hard fold recognition targets. Although many of the primary methods might fail on these targets, a metaserver is capable of detecting the best answer from the remaining servers by applying a new scoring function and analyzing the consensus. However, comparative models built on such distant templates do frequently lack quality.

Although the *de novo* structure prediction technique ROSETTA[36] outperforms comparative models from distant fold recognition templates for structures of low complexity, it fails to fold protein of higher complexity.[37] To support ROSETTA in building more complex folds a consensus contact prediction technique has been developed using an approach analogous to metaservers. Utilizing a few non-local contact predictions can greatly improve ROSETTA predictions for complex folds, as was shown in the case of T0272.[38]

Here we present an artificial neural network (ANN) for consensus contact prediction based on the protein structure predictions of 24 servers that participated in the LIVEBENCH 7 and LIVEBENCH 8 experiments, where structures for 357 targets where predicted.[39]

The network is set up to predict a potential contact between two amino acids. By sweeping over all pairs of amino acids the whole contact map can be predicted. For training all amino acid pairs having their C-alpha atoms in the native structure closer than 11 Å were considered as being in contact. To focus on non-local contacts, amino acid pairs were excluded if they were separated by less than 10 amino acids in sequence.

Inputs to the neural network are the position of the two amino acids in sequence and the total chain length (three numbers), the ratio of servers that predicted this specific contact and the total number of servers with predictions (two numbers), JUFO secondary structure prediction[40] (www.jens-meiler.de/jufo.html, three numbers), amino acid property profiles[41] (seven numbers), as well as position specific scoring matrices from PSIBLAST[33] (20 numbers) for two windows of five amino acids around the two amino acids. In addition, the contacts predicted in the top five models of the 24 servers are used together with the respective scores. Thus, the neural network had $5 + (20 + 7 + 3) \times 5 \times 2 + 24 \times 5 \times 2 = 545$ inputs, 32 hidden neurons, and one output neuron. The output range is [0,1] with 0 being "no contact" and 1 being "contact." After withholding 12% of all proteins for testing, 10 independent networks were trained with 90% of all remaining data while 10% were used for monitoring in a crossvalidating fashion. Non-contacts were removed randomly from the datasets until the ratio of contacts and non-contacts was balanced. The training algorithm was back-propagation of errors. The networks were trained until the square deviation of the monitoring dataset was minimized (approximately 10,000 periods). For prediction the average of all 10 networks and the standard deviation is reported.

If output levels above 0.5 are considered as contacts, 70% of all contacts are correctly predicted. However, in

this case 13% of all non-contacts would also be classified as contacts. This leads to a large number of false positives due to the overrepresentation of non-contacts by a factor of approximately 13 in protein structures. At output levels of 0.7 the network classifies approximately half the contacts correctly and mispredicts only 3% of the non-contacts. The number of false positives can be further reduced if the output level filter is increased to 0.8 or 0.9, although less true contacts are detected. The results obtained for the independent dataset are summarized on the Web page.

In CASP6 the consensus contact prediction was used as indicator for potential non-local contacts that should be present in *de novo* predicted models for distant fold recognition targets.[38] Figure 3 shows domain 1 of target 272. The non-local beta-sheet topology was only predicted correctly by few fold recognition methods. Nevertheless, the neural network detects these correct contact predictions (signal) among a large number of incorrect predictions (noise). Therefore, it increases the signal-to-noise ratio significantly. The server is available for academic use at www.jens-meiler.de/contact.html.

RR301 (Rost-PROFcon)

The PROFcon method is described in detail elsewhere.[22] Here, we sketched some of the major aspects and illustrated the performance with one representative example.

Standard feed-forward neural network

We trained standard feed-forward neural networks with back-propagation and momentum term.[42] We addressed the extremely unequal distribution of true (contact) and false (non-contact) samples by balanced training.[42] Symmetry between the contact probabilities for the prediction between *ij* and *ji* were enforced through a simple postprocessing average over both raw output values.[17] In total, we used 738 input, 100 hidden, and 2 output units (contact, non-contact).

Information used for input

For each residue pair *ij*, the network incorporated information from all residues in two windows of size 9 centered around *i* and *j* and in a third window of size 5 spanning the central region of the segment connecting *i* and *j*. Each residue position was characterized by the evolutionary profile, the predicted secondary structure, the predicted solvent accessibility, and the conservation weight.[43] Alignments were obtained through PSI-BLAST.[33] We used PROFphd[44–46] to predict secondary structure and solvent accessibility. We also introduced additional features to better characterize the central residues, namely a coarse-grained biophysical classification[47] (hydrophobic–hydrophobic, polar–polar, charged–polar, opposite charge, same charge, aromatic–aromatic, other), and a classification according to "residue complexity" (taken from SEG[48]). Connecting segments were globally described by their length, amino acid, and secondary structure composition. Finally, we used average amino acid and secondary structure composition of the entire protein chain, and the protein length.

CASP-independent estimate of performance

Of all the different sources of information used, the most important bits are the evolutionary information, sequence separation, and the additional global information about the connecting segment.[22] Overall, performance is best for proteins with a mixture of helices and strands and worst for all-alpha proteins (according to SCOP classification[49]); predictions are better for short than for long proteins.[49,50] Quoting single numbers for accuracy and coverage is problematic as such numbers are extremely sensitive to the particular data set chosen and to the particular definition of "long-range contact." However, to provide a ballpark figure over a data set representative for PDB: about 30% of the predicted contacts are correct (accuracy) at a prediction threshold at which about 10% of the observed contacts are predicted if we consider all contacts between residue pairs that are separated by at least six residues.

Example T0248_2 from CASP6 (Fig. 4)

T0248_2 (PDB[51] identifier: 1td6) is the second of three domains; it has 87 residues; it was classified as NF in CASP6. It has an alpha + beta fold constituted of a helical bundle (with five short helices) and a beta-sheet with two long strands (lower right triangle in Fig. 4). PROFcon correctly predicted most of the interactions between the regular secondary structure segments (e.g., between helices 1–2 and 4–3 and between the two strands, upper left triangle in Fig. 4). However, it incorrectly overpredicted contact between helices 2 and 3 that was not observed. Most interesting is the correct prediction that the sequence-distant helices 2 and 4–5 are parallel. The overall accuracy for this target was 20%, that is, close to the average performance of PROFcon for 2*L predictions and sequence separation ≥ 6.

PROFcon is available as part of the PredictProtein server,[44] as well as through a PROFcon-specific submission form http://www.predictprotein.org/submit_profcon.html.

Comparing Contact and 3D Prediction Methods by Their Capacity to Predict 3D Contacts

For those groups (contact specialists and structure prediction servers) that submitted predictions for the complete subset of 11 NF targets, we calculated the mean accuracy and mean Xd for long-range contacts at L/5 (Fig. 5). For contact specialists the highest scoring L/5 predictions were used, whereas for the structure prediction servers the L/5 predictions had to be selected randomly as there is no corresponding score. The comparison suggests that on average contact specialists are superior in contact prediction to the 3D structural methods for these targets. Additional head-to-head comparisons of the groups over the 11 NF targets confirmed this tendency, and a paired *t*-test showed that the difference is significant for the best contact prediction groups (see supplementary material). Because of the difference in sampling the L/5 predictions from the two types of prediction, this is not a perfectly fair

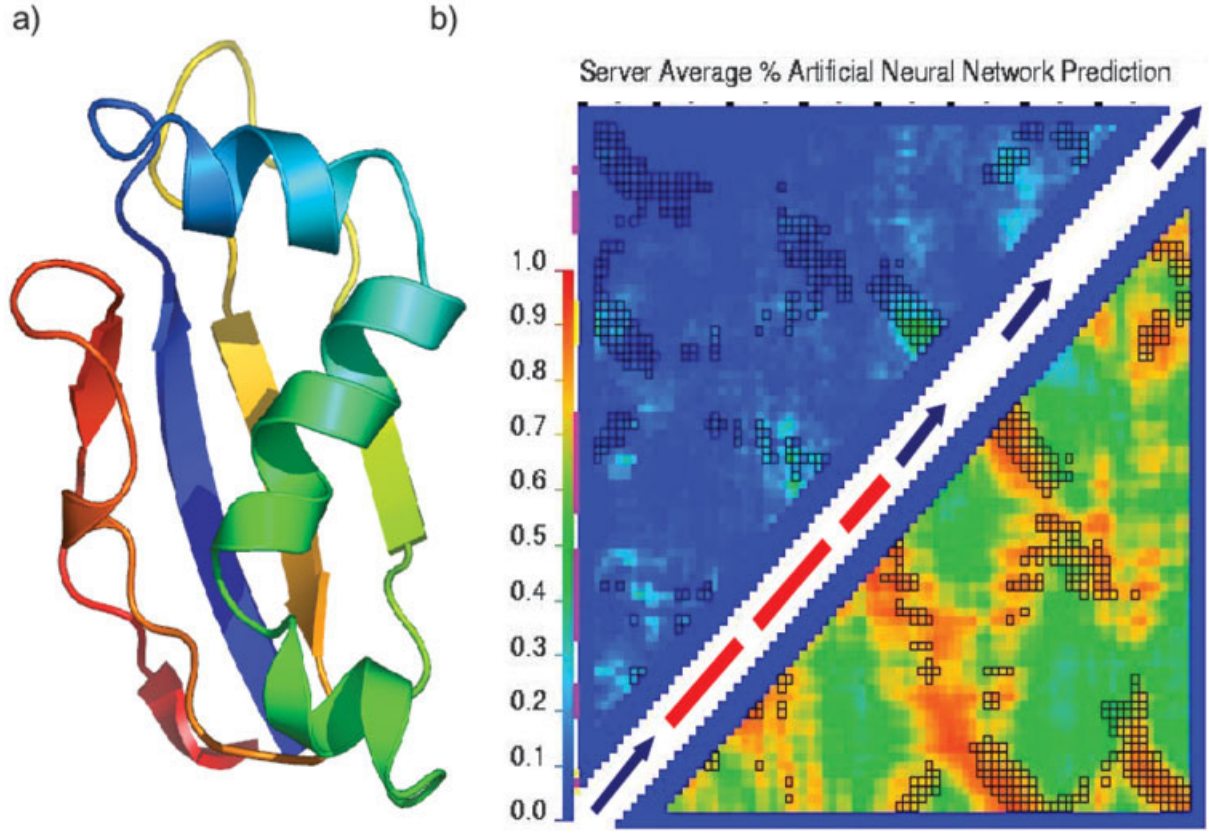

Fig. 3. The backbone structure of domain 1 in target T0272 colored from N- to C-terminus according to sequence position blue to red is shown in (**a**). (**b**) The frequency with which fold recognition servers predicted the presence of contacts (upper left triangle) and the neural network output (lower right triangle). Secondary structure elements of the native fold are indicated on the diagonal. The signal-to-noise ratio in the prediction of strand pairings between strands 1–3 and 1–4 is significantly increased with respect to the server output. Native contacts are indicated as black-bordered squares. This figure was supplied by group RR100.

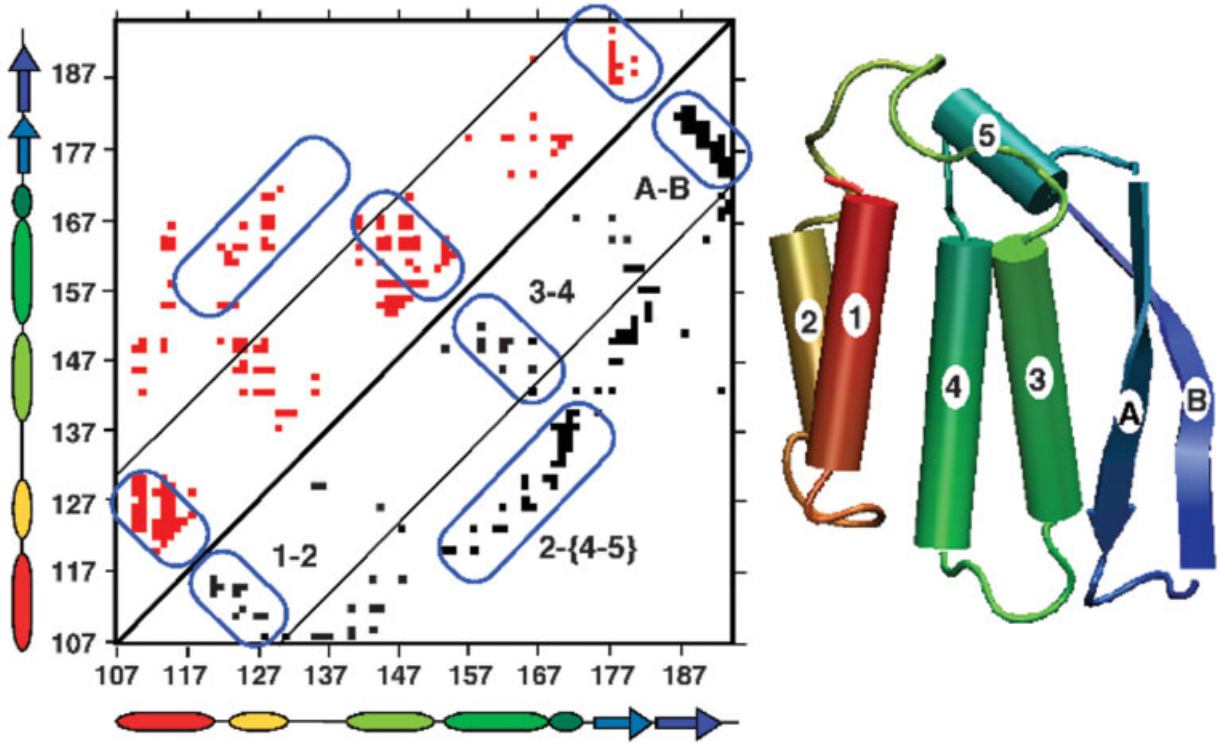

Fig. 4. PROFcon for T0248_2. Lower right triangle: experimentally observed contacts (black dots), upper left triangle: contacts predicted by PROFcon (red dots; contact for C-beta $\leq$ 8 Å and separation $\geq$ 6) when considering the best 2*L (here 2*94 =188) predictions. The two lines parallel to the main diagonal mark a sequence separation of 24 residues. Blue boxes label specific clusters of interactions in the experimental contact map. The left and bottom axes give the residue numbers and the corresponding regular secondary structures (arrows for strands, boxes for helices). Secondary structure colors reflect the position along the chain (from red at the N-terminal begin of the protein to blue at the C-terminal end of the protein). The right-hand panel sketches the structure of T0248_2 (generated with VMD[52]). This figure was supplied by group RR301.

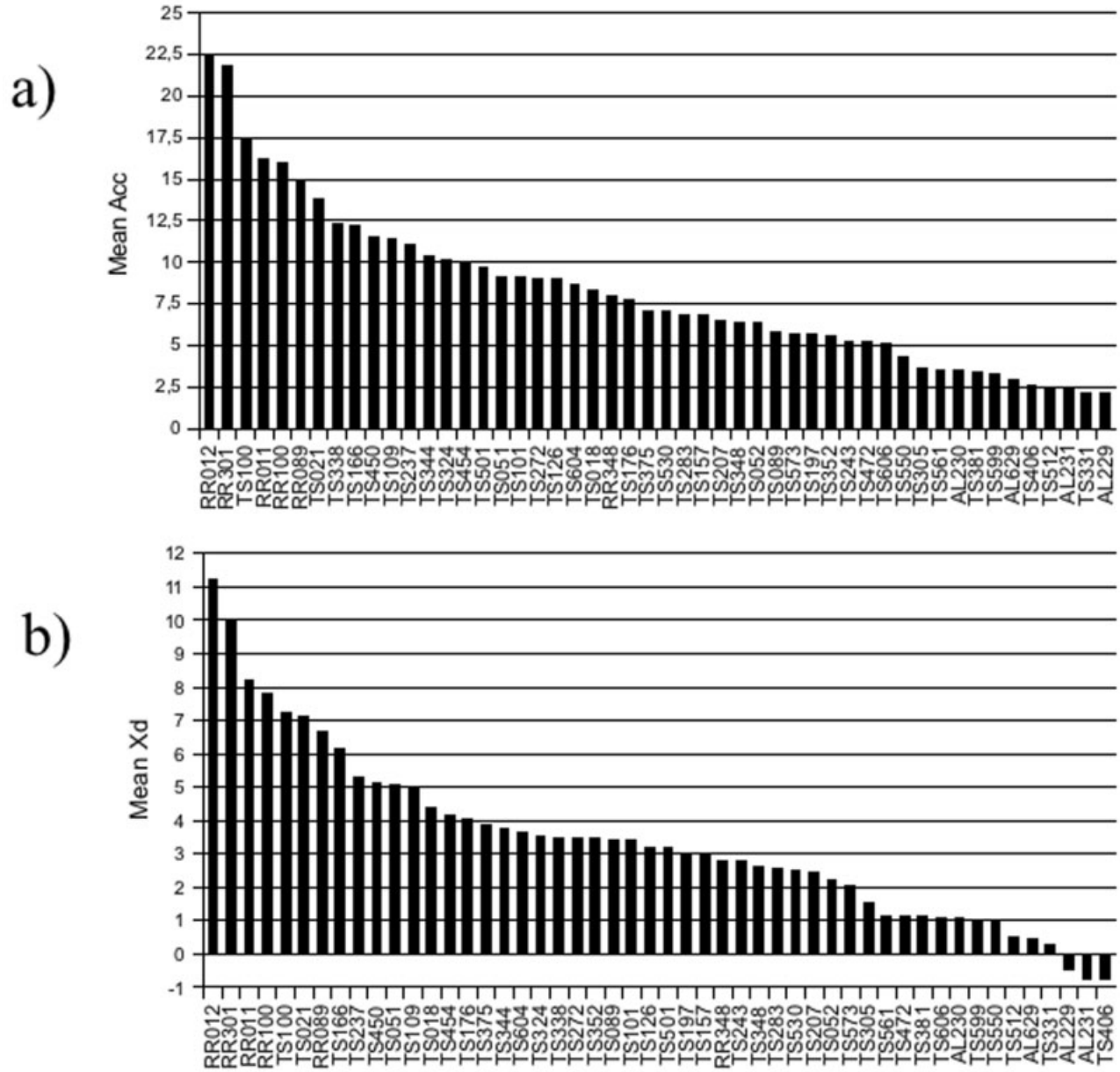

Fig. 5. Mean values of accuracy and Xd for all groups that made predictions for all 11 NF targets. (**a**) mean accuracy and (**b**) mean Xd. Groups labeled "RR" are contact prediction groups, while those labeled "AL" and "TS" are structure prediction groups.

comparison, but does show the effectiveness of the score function of the contact specialists.

Consensus Predictions

One thing that might be interesting is to see the extent to which the best long-range contact predictions overlap. We took the best L/5 predictions at 24 residues for groups RR012, RR100, and RR301 and compared the accuracy of the consensus predictions. There was little overlap between the predictions of RR100 and the other two groups, making evaluation impossible. Groups RR012 and RR301 did overlap in 11.6% of their contact predictions and these predictions had a slightly improved accuracy of 27.7% over the 11 targets. Unsurprisingly, however, coverage was very low.

Comparison with Previous Rounds of CASP

Because we performed the assessment for CAFASP3 using a similar methodology, we decided that the most homogeneous way of evaluating the performances was to compare the three best methods described above with those in the previous CAFASP3. The CAFASP3 experiment occurred in conjunction with the CASP5 experiment.

As there were only five NF targets in CASP5/CAFASP3, we randomly selected common subsets of 15 NF + FR/A targets for both sides. Performance in terms of accuracy for

long-range contacts and L/5 is better for the three best predictors in CASP6, some 6–7% points higher than the best groups in CAFASP3. Although we see an improvement in this comparison it is not possible to know if this is significant due to the limited number of targets in the samples and the differences in the difficulty of the targets between different rounds of CASP.

DISCUSSION

The 16 participating teams in CASP6 represent a considerable increase in the number of participating contact prediction groups over previous CASPs. This reflects the increase in the number of publications in the field.

We have evaluated all the models predicted by all groups in all categories in terms of contact prediction accuracy. The presentation focused on long range contacts, that is, those contacts with at least 24 residues in sequence between the members of the residue pair, and the L/5 most reliable pairs predicted. The measurements correspond to the standards accepted in the field: accuracy, coverage, and also Xd, a score that represents average distances more than strict contacts.

Three groups had better predictions with respect to the rest. The methods they use are based on different concepts: one uses genetic programming while the other two use neural networks trained with different types of input information. They illustrate the flourishing of approaches in the field during the last few years. It is interesting that these methods and others have similar performances even if they treat input contact information in very different ways, propose different learning methods, and implement different algorithms. One of the important observations is that even the most confidently predicted contacts contain a large number of incorrect predictions. This suggests that one important task in contact prediction is to improve the assignment of probability values to predictions of contacting residues.

The CASP6 experiment is too small (11 NF targets) to know if these methods are reaching a natural limit of the information contained in contact maps, or if such limit exists. It would now be interesting to monitor how these and other methods perform in the EVA continuous evaluation, for which they would have to be implemented as publicly available servers (GPCpred, PROFcon, and CMAPpro are already inscribed in EVA).

The number of cases is also too small to provide a statistically valid comparison with previous experiments. If one compares directly the average accuracy results for a common subset of 15 CASP6 NF + FR/A targets it is possible to see some improvement compared to a similar subset of CAFASP3/CASP5.

What is interesting is that in CASP6 contact prediction methods are performing better on average than 3D prediction methods when applied to difficult targets (in other words, when the 3D prediction methods cannot use information from homologous sequences of known structure). In a sense, these results contradict the extended notion that contact prediction methods were performing worse than other prediction methods.

Supplemental information

http://www.pdg.cnb.uam.es/CASP6_ContactPrediction-Evaluation/index.html.

Authors contribution

O.G., M.T., and A.V. were responsible of the contact prediction assessment of CASP6. R.M.M. contributed with the GPCpred server description (RR012). J.M. and D.B. contributed with the description of the methodology used by group RR100. M.P. and B.R. contributed with the PROFcon server description (RR301).

ACKNOWLEDGMENTS

The PDG authors (O.G., M.T., and A.V.) thank the PDG members for their suggestions and the interesting discussions, in particular to Florencio Pazos and David Juan. They also thank the team at the Lawrence Livermore Laboratory for all their practical help, in particular to Andriy Kryshtafovych. Finally, thanks to the organizers of CASP. The PROFcon authors (M.P. and B.R.) thank Guy Yachdav (Columbia) for the implementation of the PROFcon server and the following for insightful discussions: Yanay Ofran and Dariusz Przybylski (both Columbia), Søren Brunak (CBS Copenhagen), Piero Fariselli and Rita Casadio (both Bologna Univ.), Reinhard Schneider (EMBL Heidelberg), and Chris Sander (Sloan Kettering, NYC).

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. Olmea O, Rost B, Valencia A. Effective use of sequence correlation and conservation in fold recognition. J Mol Biol 1999;293:1221–1239.
3. Ortiz AR, Kolinski A, Rotkiewicz P, Ilkowski B, Skolnick J. Ab initio folding of proteins using restraints derived from evolutionary information. Proteins Suppl 1999;3:177–185.
4. 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.
5. Altschuh D, Vernet T, Berti P, Moras D, Nagai K. Coordinated amino acid changes in homologous protein families. Protein Eng 1988;2:193–199.
6. Shindyalov IN, Kolchanov NA, Sander C. Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? Protein Eng 1994;7:349–358.
7. Gobel U, Sander C, Schneider R, Valencia A. Correlated mutations and residue contacts in proteins. Proteins 1994;18:309–317.
8. Taylor WR, Hatrick K. Compensating changes in protein multiple sequence alignments. Protein Eng 1994;7:341–348.
9. Olmea O, Valencia A. Improving contact predictions by the combination of correlated mutations and other sources of sequence information. Fold Des Suppl 1997;2:25–32.
10. Lapedes AS, Giraud BG, Liu LC, Stormo GD. Correlated mutations in protein sequences: phylogenetic and structural effects. PASCSMB 1997;1–22.
11. Chelvanayagam G, Eggenschwiler A, Knecht L, Gonnet GH, Benner SA. An analysis of simultaneous variation in protein structures. Protein Eng 1997;10:307–316.
12. Dekker JP, Fodor A, Aldrich RW, Yellen G. A perturbation-based method for calculating explicit likelihood of evolutionary covariance in multiple sequence alignments. Bioinformatics 2004;20:1565–1572.
13. Zhang C, Hou J, Kim SH. Fold prediction of helical proteins using torsion angle dynamics and predicted restraints. Proc Natl Acad Sci USA 2002;99:3581–3585.
14. Zhu J, Zhu Q, Shi Y, Liu H. How well can we predict native

contacts in proteins based on decoy structures and their energies? Proteins 2003;52:598–608.
15. Fariselli P, Olmea O, Valencia A, Casadio R. Prediction of contact maps with neural networks and correlated mutations. Protein Eng 2001;14:835–843.
16. Fariselli P, Olmea O, Valencia A, Casadio R. Progress in predicting inter-residue contacts of proteins with neural networks and correlated mutations. Proteins Suppl 2001;5:157–162.
17. Pollastri G, Baldi P. Prediction of contact maps by GIOHMMs and recurrent neural networks using lateral propagation from all four cardinal corners. Bioinformatics Suppl 2002;18:62–70.
18. Shao Y, Bystroff C. Predicting interresidue contacts using templates and pathways. Proteins Suppl 2003;53:497–502.
19. Vullo A, Frasconi P. Prediction of protein coarse contact maps. J Bioinform Comput Biol 2003;1:411–431.
20. MacCallum RM. Striped sheets and protein contact prediction. Bioinformatics Suppl 2004;20:224–231.
21. Gupta N, Mangal N, Biswas S. Evolution and similarity evaluation of protein structures in contact map space. Proteins 2005;59: 196–204.
22. Punta M, Rost B. PROFcon: novel prediction of long-range contacts. Bioinformatics 2005;21:2960–2968.
23. Lesk AM. CASP2: Report on ab initio predictions. Proteins Suppl 1997;1:151–166.
24. Orengo CA, Bray JE, Hubbard T, LoConte L, Sillitoe I. Analysis and assessment of ab initio three-dimensional prediction, secondary structure, and contacts prediction. Proteins Suppl 1999;3:149–170.
25. Lesk AM, Lo Conte L, Hubbard T. Assessment of novel fold targets in CASP4: predictions of three-dimensional structures, secondary structures, and interresidue contacts. Proteins Suppl 2001;5:98–118.
26. Aloy P, Stark A, Hadley C, Russell RB. Predictions without templates: new folds, secondary structure, and contacts in CASP5. Proteins Suppl 2003;6:436–456.
27. Dunbrack et al. Targets. Proteins 2005;Suppl 7:8–18.
28. Fischer D, Elofsson A, Rychlewski L, Pazos F, Valencia A, Rost B, Ortiz AR, Dunbrack RL Jr. CAFASP2: the second critical assessment of fully automated structure prediction methods. Proteins Suppl 2001;5:171–183.
29. Eyrich VA, Przybylski D, Koh IY, Graña O, Pazos F, Valencia A, Rost B. CAFASP3 in the spotlight of EVA. Proteins Suppl 2003;6:548–560.
30. Graña O, Eyrich VA, Pazos F, Rost B, Valencia A. EVAcon: a protein contact prediction evaluation service. Nucleic Acid Res 2005;33(Web Server issue):W347–W351.
31. Holm L, Sander C. Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and detection of co-ordinate errors. J Mol Biol 1991;218:183–194.
32. Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. Correlated mutations contain information about protein–protein interaction. J Mol Biol 1997;271:511–523.
33. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25: 3389–3402.
34. Lundstrom J, Rychlewski L, Bujnicki J, Elofsson A. Pcons: a neural-network-based consensus predictor that improves fold recognition. Protein Sci 2001;10:2354–2362.
35. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 2003;19:1015–1018.
36. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 1997;268:209–225.
37. Bonneau R, Ruczinski I, Tsai J, Baker D. Contact order and ab initio protein structure prediction. Protein Sci 2002;11:1937–1944.
38. Bradley P, Baker D. Rosetta in CASP6. in prep.
39. Rychlewski L, Fischer D. LiveBench-8: the large-scale, continuous assessment of automated protein structure prediction. Protein Sci 2005;14:240–245.
40. Meiler J, Baker D. Coupled prediction of protein secondary and tertiary structure. Proc Natl Acad Sci USA 2003;100:12105–12110.
41. Meiler J, Müller M, Zeidler A, Schmäschke F. Generation and evaluation of dimension reduced amino acid parameter representations by artificial neural networks. J Mol Model 2001;7:360–369.
42. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 1993;232:584–599.
43. Rost B. PHD: predicting one-dimensional protein structure by profile based neural networks. Methods Enzymol 1996;266:525–539.
44. Rost B, Yachdav G, Liu J. The Predict Protein server. Nucleic Acids Res Suppl 2004;32:321–326.
45. Rost B. Protein secondary structure prediction continues to rise. J Struc Biol 2001;134:204–218.
46. Rost B. How to use protein 1D structure predicted by PROFphd. In: Walker JE, editor. The proteomics protocols handbook, methods in molecular biology. Totowa NJ: Humana 2005. p 875–901.
47. Creighton T. Proteins: structures and molecular properties. New York: Company WHFa; 1992.
48. Wootton JC, Federhen S. Analysis of compositionally biased regions in sequence databases. Methods Enzymol 1996;266:554–571.
49. Andreeva A, Howorth D, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. SCOP database in 2004: refinements integrate structure and sequence family data. Nucleic Acids Res Suppl 2004;32: 226–229.
50. Punta M, Rost B. Protein folding rates estimated from contact predictions. J Mol Biol 2005;348:507–512.
51. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. Acta Crystallogr D Biol Crystallogr 2002;58:899–907.
52. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:27–28.